



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS
CARRERA DE SOFTWARE

TRABAJO DE INTEGRACIÓN CURRICULAR PREVIO A LA OBTENCIÓN DEL
TÍTULO DE INGENIERO DE SOFTWARE

**“APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL
EXPLICABLE EN MODELOS DE DEEP LEARNING PARA LA
DETECCIÓN DE MALEZAS EN CULTIVOS DE PAPA.”**



AUTOR:

Zámyr Alexander Guevara Terán

DIRECTOR:

PhD. Iván Danilo García Santillán

Ibarra – Ecuador

2026

UNIVERSIDAD TÉCNICA DEL NORTE
BIBLIOTECA UNIVERSITARIA

IDENTIFICACIÓN DE LA OBRA

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

DATOS DE CONTACTO			
CÉDULA DE IDENTIDAD:	1004974646		
APELLIDOS Y NOMBRES:	ZÁMYR ALEXANDER GUEVARA TERÁN		
DIRECCIÓN:	COTACACHI		
EMAIL:	ZAGUEVARAT@UTN.EDU.EC		
TELF. FIJO:	0994284787	TELF. MÓVIL:	0994284787

DATOS DE LA OBRA	
TÍTULO:	APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL EXPLICABLE EN MODELOS DE DEEP LEARNING PARA LA DETECCIÓN DE MALEZAS EN CULTIVOS DE PAPA.
AUTOR(ES):	ZÁMYR ALEXANDER GUEVARA TERÁN
FECHA:	27/03/2026
CARRERA/PROGRAMA:	☒ GRADO ☐ POSGRADO
TÍTULO POR EL QUE OPTA:	INGENIERO DE SOFTWARE
DIRECTOR:	PHD. IVÁN DANILO GARCÍA SANTILLÁN
ASESOR:	PhD. MARCO REMIGIO PUSDÁ CHULDE

CONSTANCIAS

El autor (es) manifiesta (n) que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto, la obra es original y que es (son) el (los) titular (es) de los derechos patrimoniales, por lo que asume (n) la responsabilidad sobre el contenido de la misma y saldrá (n) en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 27 días, del mes de marzo de 2026.

EL AUTOR:



Firmado electrónicamente por:
ZAMYR ALEXANDER
GUEVARA TERAN

Validar únicamente con FirmaRC

Firma.....

Nombre: Zámyr Alexander Guevara Terán

CI: 1004974646

CERTIFICACIÓN DEL DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR

Ibarra, 27 de marzo de 2026

PhD. Iván Danilo García Santillán
DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR

CERTIFICA:

Haber revisado el presente informe final del trabajo de Integración Curricular, el mismo que se ajusta a las normas vigentes de la Universidad Técnica del Norte; en consecuencia, autorizo su presentación para los fines legales pertinentes.



PhD. Iván Danilo García Santillán
C.C.: 1002292603

APROBACIÓN DEL COMITÉ CALIFICADOR

El Comité Calificador del trabajo de Integración Curricular “ Aplicación de técnicas de Inteligencia Artificial Explicable en modelos de Deep Learning para la detección de malezas en cultivos de papa.” elaborado por Zámyr Alexander Guevara Terán , previo a la obtención del título del INGENIERO DE SOFTWARE, aprueba el presente informe de investigación en nombre de la Universidad Técnica del Norte:



PhD. Iván Danilo García Santillán
C.C.: 1002292603



PhD. Marco Remigio Pusdà Chulde
C.C.: 0401200951

DEDICATORIA

A Dios, por ser el arquitecto de mis días y la fortaleza en los momentos de mayor agotamiento mental y espiritual. Le agradezco por haberme otorgado la templanza necesaria para transformar el esfuerzo en este logro académico y por permitirme descubrir el propósito que guía mi existencia bajo su constante bendición.

A mis padres, Bayron Guevara y Zayra Teran, quienes han sido el cimiento inamovible de mis sueños y la fuente de mis valores más profundos. Gracias por su sacrificio silencioso, por creer en mis capacidades cuando incluso yo dudaba de ellas y por enseñarme que la perseverancia es la única llave capaz de abrir las puertas del éxito. Este título es el testimonio de su entrega absoluta y el cumplimiento de un anhelo que empezamos a construir juntos desde mi infancia.

A Marcela, mi compañera de vida. Gracias por ser el equilibrio en mis días de tormenta y la luz en mis momentos de incertidumbre. Tu apoyo incondicional y tu paciencia infinita durante las largas jornadas de investigación fueron el motor que me impulsó a culminar este proyecto. Eres mi presente, mi inspiración constante para buscar la excelencia y la persona con la que anhelo compartir cada victoria futura. Este logro es solo un peldaño más en la hermosa historia que, con amor y determinación, estamos edificando juntos.

A mis abuelitos, por ser mis raíces y el legado de sabiduría que guardo con profundo respeto en mi corazón. Sus enseñanzas y su ternura han sido la brújula que orienta mi calidad como persona. De igual manera, hago extensiva esta dedicatoria a mi familia en general, por constituir ese círculo de apoyo incondicional y por celebrar conmigo cada pequeño avance. Gracias por formar parte de mi historia, por cada palabra de aliento y por recordarme siempre la importancia de mantener la humildad y la unión ante cualquier desafío.

Zámyr Alexander Guevara Terán

AGRADECIMIENTO

A Dios, por ser la fuente de mi fortaleza y sabiduría. Le agradezco por permitirme concluir este proceso con salud y por poner en mi camino las señales necesarias para nunca perder el rumbo, recordándome siempre que cada esfuerzo tiene un propósito mayor.

A mis padres, Bayron Guevara y Zayra Teran, quienes han sido el cimiento sobre el cual he construido mis sueños. Gracias por su apoyo incondicional y por los sacrificios silenciosos que hicieron posible mi educación. Su ejemplo de integridad y trabajo duro es la brújula que guía mi vida profesional y personal.

A la Universidad Técnica del Norte, mi segundo hogar, por brindarme el espacio y las herramientas para crecer. Mi gratitud se extiende a todos los docentes que compartieron su conocimiento con generosidad, y de manera muy especial al Dr. Iván García, cuya mentoría y rigor científico fueron determinantes para alcanzar la calidad que esta investigación exigía. Gracias por su paciencia y por desafiarme a ir siempre más allá de lo evidente.

Finalmente, a todos aquellos que, con una palabra de aliento o un consejo técnico, contribuyeron a mi formación en esta apasionante carrera, dejando una huella en mi visión del mundo y la tecnología.

Zámyr Alexander Guevara Terán

ÍNDICE DE CONTENIDOS

DEDICATORIA	6
AGRADECIMIENTO	7
ÍNDICE DE FIGURAS	14
ÍNDICE DE TABLAS	14
RESUMEN	16
ABSTRACT	17
CAPÍTULO I	
1.1 Planteamiento del problema	18
1.2 Objetivos	19
1.2.1 Objetivo General	19
1.2.2 Objetivos Específicos	19
1.3 Alcance	20
1.4 Metodología	21
1.5 Justificación	22
1.5.1 Justificación Académica	22
1.5.2 Justificación Económica	22
1.6 Antecedentes	24

CAPÍTULO II

2.1	Detección de Malezas en la Agricultura de Precisión y el Rol del Deep Learning	25
2.1.1	Impacto de las Malezas en Cultivos de Papa y Métodos Tradicionales de Detección	25
2.1.2	Avances en Deep Learning para la Visión por Computadora en Agricultura	27
2.1.3	Desafíos Actuales en la Detección de Malezas con Deep Learning	28
2.2	Inteligencia Artificial Explicable (XAI): Fundamentos y Necesidad en Modelos Agrícolas	29
2.2.1	El Problema de la “Caja Negra” en Deep Learning y sus Implicaciones	30
2.2.2	Definición, Objetivos y Taxonomía de las Técnicas XAI	30
2.2.3	Importancia de XAI en la Agricultura para la Toma de Decisiones Informada y la Adopción Tecnológica	33
2.2.4	Desafíos y Limitaciones Inherentes a las Técnicas de Explicabilidad	33
2.3	Técnicas XAI Post-hoc para Modelos de Visión por Computadora	34
2.3.1	SHAP (SHapley Additive exPlanations): Fundamento y Aplicación	34
2.3.2	LIME (Local Interpretable Model-agnostic Explanations): Fundamento y Aplicación	35
2.3.3	Grad-CAM (Gradient-weighted Class Activation Mapping): Fundamento y Aplicación	37
2.3.4	Score-CAM: Fundamento y Aplicación	38
2.3.5	Comparativa de Robustez y Dependencia de Gradientes: Grad-CAM vs. Score-CAM	39
2.4	Herramientas y Librerías de Código Abierto para XAI en Deep Learning	40

2.4.1 Panorama de Librerías Python (ej. SHAP, LIME, Tf-explain) y sus Capacidades	40
2.5 Aplicaciones y Evaluación de XAI en la Detección de Malezas	41
2.5.1 Métricas de Evaluación del Rendimiento del Modelo de Detección (Precision, Recall, mAP, F1-Score).	41
2.6 Marco Metodológico Estándar para Proyectos de Inteligencia Artificial	42
2.6.1 Introducción a la Metodología CRISP-DM (Cross-Industry Standard Process for Data Mining)	42
2.7 Trabajos Relacionados	43
2.7.1 Deep Learning en la Detección de Malezas	43
2.7.2 Segmentación Semántica en Cultivos de Papa	44
2.7.3 Inteligencia Artificial Explicable (XAI) en Agricultura	44
CAPÍTULO III	
3.1 Enfoque Metodológico y Estándar de Proceso	46
3.2 Fase 1: Entorno de Desarrollo e Infraestructura Computacional	48
3.2.1 Recursos de Hardware.	48
3.2.2 Entorno de Software y Librerías	48
3.3 Fase 2: Comprensión y Adquisición de los Datos	49
3.3.1 Origen y Adquisición de Datos	49
3.3.2 Composición de Clases	50
3.4 Fase 3: Preparación de los datos	50
3.5 Fase 4: Modelado y Arquitectura de la Red	51

3.5.1	Metodología de Evaluación Comparativa	51
3.5.2	Estructura General de la Red.	52
3.5.2.1	Encoder (Extractor de Características).	52
3.5.2.2	Módulo ASPP (Atrous Spatial Pyramid Pooling)	53
3.5.2.3	Decoder con Mecanismos de Atención	53
3.5.3	Configuración Específica por Escenario	53
3.5.3.1	Configuración del Escenario A (Entrenamiento Original).	53
3.5.3.2	Configuración del Escenario B (Optimización SWA)	54
3.5.4	Diseño Experimental y Selección de Casos	55
3.5.4.1	Fase 1: Escaneo Cuantitativo del Dataset	55
3.5.4.2	Fase 2: Selección Estratificada de Casos	56
3.5.5	Configuración Experimental de Técnicas XAI	57
3.5.5.1	Mapas de Activación (Grad-CAM y Score-CAM)	57
3.5.5.2	SHAP (Shapley Additive exPlanations).	58
3.5.5.3	LIME (Local Interpretable Model-agnostic Explanations)	58
3.6	Fase 5: Evaluación	59

3.6.1	Métricas de Evaluación Cuantitativa.	59
3.6.1.1	Fidelidad (Fidelity).	60
3.6.1.2	Localización (Localization)	61
3.6.1.3	Robustez (Robustness)	61

CAPÍTULO IV

4.1	Evaluación Estadística del Rendimiento	63
4.1.1	Comparativa Global de Métricas.	64
4.1.2	Análisis de Rendimiento por Clase.	64
4.1.3	Análisis de Matrices de Confusión	65
4.2	Análisis de Interpretabilidad (XAI)	67
4.2.1	Evaluación Cuantitativa de la Explicabilidad	67
4.2.2	Evaluación Cualitativa Visual por Caso de Estudio	69
4.2.2.1	Caso 1: Fallo Crítico (Recuperación de Ceguera)	69
4.2.2.2	Caso 2: Fallo (Corrección de Falsos Negativos)	70
4.2.2.3	Caso 3: Promedio (Refinamiento de Bordes).	72
4.2.2.4	Caso 4: Bueno (Consolidación de Confianza)	73
4.2.2.5	Caso 5: Éxito (Validación de Maestría)	74
4.3	Síntesis Comparativa de Resultados	75
4.3.1	Convergencia entre Precisión y Atención.	75
4.3.2	Impacto de las Estrategias de Optimización	75
4.4	Discusión	76

4.4.1 Evolución de la Detección a la Segmentación Semántica en Agricultura de Precisión	76
4.4.2 Complejidad del Entorno: Del Laboratorio al Campo	77
4.4.3 El Cambio de Paradigma en XAI: De la Visualización a la Métrica	78
4.4.4 El Ciclo XAI como Motor de Optimización	79
4.4.5 Limitaciones y Perspectivas Futuras	79
CONCLUSIONES	81
RECOMENDACIONES	83
Referencias	
ANEXOS	93
4.5 Repositorio de Código y Scripts de Experimentación	93
4.6 Conjunto de Datos (Dataset)	94

ÍNDICE DE FIGURAS

Figura 1	Árbol de Problemas	19
Figura 2	Diagrama sobre el alcance del proyecto	21
Figura 3	Matriz de Confusión Normalizada - Escenario A (Modelo Base).	66
Figura 4	Matriz de Confusión Normalizada - Escenario B (Modelo SWA).	67
Figura 5	Caso 1 (Base): Ausencia de actividad significativa en los mapas de atención.	69
Figura 6	Caso 1 (SWA): Reactivación de la atención sobre la maleza central.	70
Figura 7	Caso 2 (Base): Atención dispersa hacia el suelo y bordes irrelevantes.	71
Figura 8	Caso 2 (SWA): Focalización precisa sobre la estructura foliar.	71
Figura 9	Caso 3 (Base): Desbordamiento de la atención fuera de los límites de la hoja.	72
Figura 10	Caso 3 (SWA): Ajuste fino de la atención al contorno aserrado.	72
Figura 11	Caso 4 (Base): Buena detección con ligera incertidumbre en los bordes.	73
Figura 12	Caso 4 (SWA): Detección contundente y mapas de calor de alto contraste.	73
Figura 13	Caso 5 (Base): Segmentación correcta.	74
Figura 14	Caso 5 (SWA): Segmentación y explicación casi perfectas (Robustez 0.99).	74
Figura 15	Diagnóstico XAI – Caso de éxito con IoU de 0.98	95
Figura 16	Diagnóstico XAI – Caso de éxito con IoU de 0.98	95
Figura 17	Diagnóstico XAI – Caso de éxito con IoU de 0.98	95
Figura 18	Diagnóstico XAI – Caso de éxito con IoU de 0.98	95
Figura 19	Diagnóstico XAI – Caso de éxito con IoU de 0.98	96
Figura 20	Diagnóstico XAI – Caso de éxito con IoU de 0.97	96
Figura 21	Diagnóstico XAI – Caso de éxito con IoU de 0.97	96
Figura 22	Diagnóstico XAI – Caso de rendimiento promedio con IoU de 0.89	96
Figura 23	Diagnóstico XAI – Caso de fallo con IoU de 0.30	97
Figura 24	Diagnóstico XAI – Caso de fallo con IoU de 0.00	97
Figura 25	Diagnóstico XAI – Caso de fallo con IoU de 0.00	97
Figura 26	Diagnóstico XAI – Caso de fallo con IoU de 0.00	97
Figura 27	Diagnóstico XAI – Caso de fallo con IoU de 0.00	98
Figura 28	Diagnóstico XAI – Caso de fallo con IoU de 0.00	98
Figura 29	Diagnóstico XAI – Caso de fallo con IoU de 0.00	98

ÍNDICE DE TABLAS

Tabla 1	Resumen de investigaciones y aportes en el área de detección de malezas y XAI.	24
Tabla 2	Adaptación del Ciclo CRISP-DM al Estudio de Segmentación de Malezas.	47
Tabla 3	Comparación de métricas globales de rendimiento entre escenarios.	64
Tabla 4	Desglose de Intersección sobre la Unión (IoU) por clase.	65
Tabla 5	Comparativa de métricas XAI. (*Nota: La robustez de 1.00 en los fallos del modelo base es un artefacto debido a la ausencia de mapas de activación válidos).	68

RESUMEN

El presente trabajo de integración curricular aborda la opacidad de los modelos de Deep Learning aplicados a la agricultura de precisión, específicamente en la detección de malezas en cultivos de papa (*Solanum tuberosum* L.). A pesar del alto rendimiento de las Redes Neuronales Convolucionales, su naturaleza de “caja negra” limita la confianza y adopción por parte de expertos agrónomos. Bajo la metodología CRISP-DM adaptada, se aplicaron técnicas de Inteligencia Artificial Explicable (XAI) como SHAP, LIME, Grad-CAM y Score-CAM sobre una arquitectura Attentive DeepLabv3+ con backbone EfficientNetV2-S. La investigación implementó una estrategia de optimización basada en el Salto de Resolución (256x256 píxeles) y el Promediado Estocástico de Pesos (SWA). Los resultados cuantitativos evidenciaron una mejora significativa, elevando el mIoU global del 81.08 % al 86.19 %, con un incremento crítico del 16.83 % en la detección de especies minoritarias. Mediante la auditoría de XAI, se validó que el modelo optimizado basa sus predicciones en características morfológicas biológicas coherentes, alcanzando una robustez superior a 0.96 y eliminando activaciones espurias del suelo. Se concluye que el ciclo XAI no solo aporta transparencia, sino que actúa como un motor de ingeniería para el refinamiento de modelos, facilitando el despliegue de tecnologías de rociado selectivo más sostenibles y confiables.

Palabras clave: Inteligencia Artificial Explicable (XAI), Deep Learning, Segmentación Semántica, Detección de Malezas, Agricultura de Precisión, Cultivos de Papa.

ABSTRACT

This graduation project addresses the lack of transparency in Deep Learning models applied to precision agriculture, specifically for weed detection in potato crops (*Solanum tuberosum* L.). Despite the high performance of Convolutional Neural Networks, their "black box" nature limits the trust and adoption by agricultural experts. Under an adapted CRISP-DM methodology, Explainable Artificial Intelligence (XAI) techniques—such as SHAP, LIME, Grad-CAM, and Score-CAM—were applied to an Attentive DeepLabv3+ architecture with an EfficientNetV2-S backbone. The research implemented an optimization strategy based on Resolution Jump (256x256 pixels) and Stochastic Weight Averaging (SWA). Quantitative results demonstrated a significant improvement, raising the global mIoU from 81.08 % to 86.19 %, with a critical 16.83 % increase in the detection of minority species. Through XAI auditing, it was validated that the optimized model bases its predictions on coherent biological morphological features, achieving a robustness (SSIM) above 0.96 and eliminating spurious ground-level activations. It is concluded that the XAI cycle not only provides transparency but also acts as an engineering driver for model refinement, facilitating the deployment of more sustainable and reliable selective spraying technologies.

Keywords: Explainable Artificial Intelligence (XAI), Deep Learning, Semantic Segmentation, Weed Detection, Precision Agriculture, Potato Crops.

CAPÍTULO I

INTRODUCCIÓN

1.1 Planteamiento del problema

En la actualidad, los modelos de Deep Learning han demostrado un alto rendimiento en tareas de visión por computadora aplicadas a la agricultura, como la detección de malezas en cultivos [1]. Sin embargo, su naturaleza de “caja negra” continúa siendo una barrera significativa para su adopción en contextos reales, especialmente cuando se requiere comprender y justificar cada predicción realizada [2], [3].

La limitada explicabilidad genera desconfianza y transparencia en los modelos, complica su validación por parte de expertos agrícolas y dificulta la detección de errores, sesgos o sobre ajustes. Esto es especialmente crítico en aplicaciones que afectan directamente las decisiones sobre el uso de herbicidas, tratamientos o intervenciones agronómicas [4].

Técnicas de explicabilidad como SHAP (SHapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations) y Grad-CAM (Gradient-weighted Class Activation Mapping) permiten visualizar y cuantificar la influencia de las características de entrada en las decisiones del modelo. Estas herramientas se han consolidado como una forma efectiva de generar confianza, validación y explicaciones útiles tanto para desarrolladores como para usuarios finales [5], [6], [7].

Este trabajo busca aplicar dichas técnicas sobre un modelo de Deep Learning previamente entrenado para la detección de malezas [1], con el objetivo de generar explicaciones claras sobre su comportamiento y desempeño. Además, se utilizarán los hallazgos obtenidos para proponer posibles mejoras que optimicen el modelo, manteniendo como prioridad la transparencia y la confiabilidad de las predicciones.

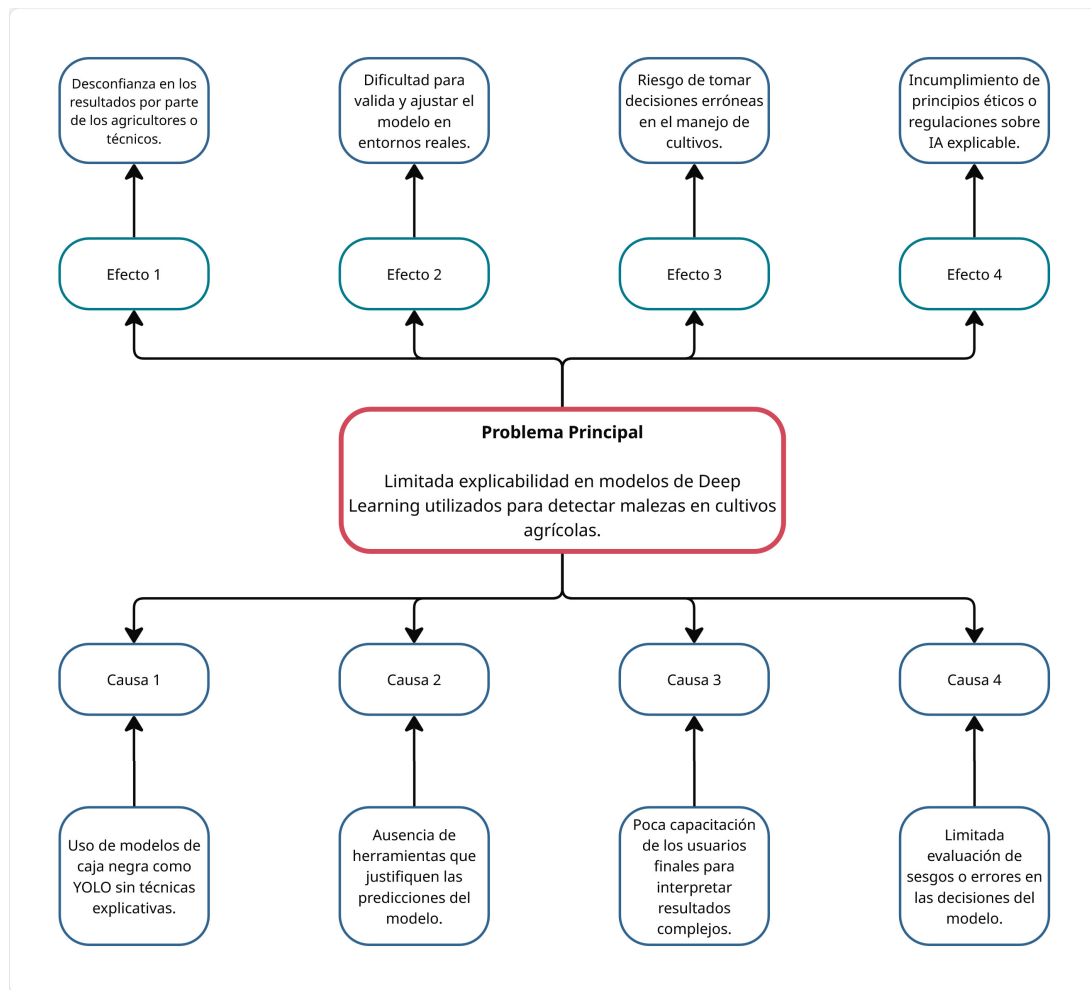


Figura 1.
Árbol de Problemas

1.2 Objetivos

1.2.1 Objetivo General

Aplicar técnicas de explicabilidad (XAI) a modelos de Deep Learning en la detección de malezas en cultivos para comprender y explicar sus predicciones de manera clara y transparente.

1.2.2 Objetivos Específicos

- Redactar un marco teórico sobre las técnicas de explicabilidad utilizadas en modelos de Deep Learning para detección de malezas, con énfasis en SHAP, LIME y Grad-CAM.

- Aplicar técnicas de explicabilidad a un modelo preentrenado de detección de malezas para comprender la coherencia y confiabilidad mediante métricas de interpretabilidad.
- Reentrenar el modelo incorporando hallazgos obtenidos de la explicabilidad y evaluar el rendimiento del modelo utilizando métricas de IA.

1.3 Alcance

Este trabajo se enfocará en la aplicación de técnicas de Explicabilidad de Inteligencia Artificial (Explainable Artificial Intelligence, XAI) a un modelo de Deep Learning previamente entrenado para la detección de malezas en cultivos de papa.

La explicabilidad permite desvelar cómo un modelo de Deep Learning llega a sus predicciones, identificando las regiones o características de la imagen que más influyen en cada decisión mediante técnicas de XAI [8]. Se utilizarán herramientas como SHAP (SHapley Additive exPlanations), que asigna valores de contribución a cada característica de entrada para explicar la predicción [5]; LIME (Local Interpretable Model-agnostic Explanations), que genera explicaciones locales aproximando el modelo original con uno interpretable [7]; y Grad-CAM (Gradient-weighted Class Activation Mapping), que produce mapas de activación visuales para identificar qué regiones de una imagen influyen más en la decisión del modelo [6]. También, se explorará Score-CAM debido a que parecería ser más robusto porque no depende de gradientes como Grad-CAM [9]. Este análisis permitirá identificar patrones relevantes, evaluar la coherencia y confiabilidad de las decisiones del modelo, así como facilitar la validación de sus resultados.

El proyecto se desarrollará utilizando un conjunto de datos ya recopilados y anotados por trabajos previos de titulación de la Carrera de Software [1], sin contemplar la recolección adicional de información.

Con base en los hallazgos derivados del proceso de explicabilidad, se propondrán recomendaciones específicas orientadas a optimizar el modelo existente y actualizarlo, con el objetivo de lograr una mayor robustez, precisión y transparencia en las inferencias realizadas.

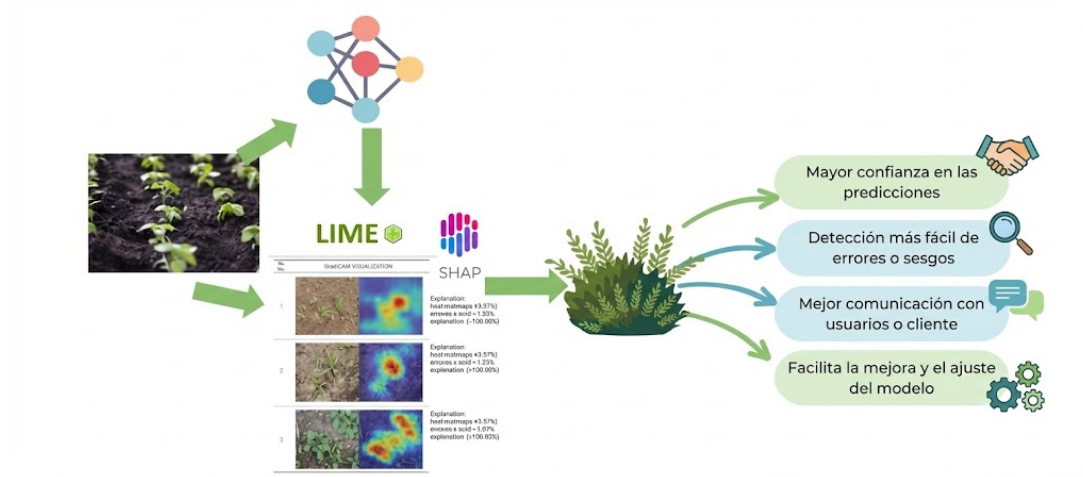


Figura 2.
Diagrama sobre el alcance del proyecto

1.4 Metodología

Este estudio, de carácter aplicado, se desarrollará siguiendo la metodología CRISP-DM adaptada al enfoque de XAI. Las fases del proceso son:

- **Comprensión del negocio:** Se definirán los objetivos del proyecto y se determinarán las necesidades de explicabilidad específicas para el modelo de detección de malezas.
- **Comprensión de los datos:** Se realizará una revisión de la literatura científica sobre técnicas XAI aplicadas a modelos de Deep Learning, recopilando y analizando información relevante para fundamentar el marco teórico.
- **Preparación de los datos:** Se configurará el entorno de desarrollo (TensorFlow), se limpiarán y estructurarán los conjuntos de datos, garantizando su calidad y utilidad para el modelado.
- **Modelado:** Sobre el modelo preentrenado, se incorporarán métodos explicativos (por ejemplo, SHAP, LIME o Grad-CAM) y se evaluará su desempeño (métricas estándar), así como la calidad de las interpretaciones generadas.
- **Evaluación:** A partir de los resultados explicativos, se ajustarán los hiperparámetros y

se eliminarán características poco útiles, efectuando ciclos iterativos de reentrenamiento hasta optimizar tanto la precisión como la claridad de las explicaciones.

- **Despliegue:** Se compararán las versiones original y mejorada del modelo mediante métricas como precisión, recall, F1-score y AUC-ROC, y se elaborarán recomendaciones basadas en evidencia para futuras mejoras.

1.5 Justificación

Se alinea con los Objetivos de Desarrollo Sostenible (ODS), específicamente con el ODS 2 (Hambre cero), que busca asegurar sistemas de producción sostenibles y resilientes, ODS 12 (Producción y consumo responsables), que promueve la gestión sostenible y el uso eficiente de los recursos naturales, y ODS 13 (Acción por el clima), orientado a adoptar medidas urgentes para combatir el cambio climático y sus efectos [10].

Además, se vincula con políticas nacionales enfocadas en la sostenibilidad agrícola y la reducción del uso de agroquímicos, reforzando la seguridad ambiental y alimentaria mediante el uso de tecnologías explicables.

Beneficiarios: Agricultores, consumidores, empresas tecnológicas.

1.5.1 Justificación Académica

Este trabajo contribuye al campo de la inteligencia artificial aplicada a la agricultura mediante la evaluación de técnicas de explicabilidad (XAI), lo que permite a investigadores y desarrolladores comprender mejor los comportamientos y decisiones de los modelos de visión por computadora en contextos agrícolas específicos.

1.5.2 Justificación Económica

La implementación de estos modelos permitirá optimizar recursos, reducir los costos operativos relacionados con la mano de obra y los agroquímicos, y minimizar las pérdidas económicas

derivadas de la competencia con malezas. Adicionalmente, la transparencia generada aumentará la confianza y facilitará la adopción de estas tecnologías por parte del sector productivo agrícola, mejorando su rentabilidad y sostenibilidad.

1.6 Antecedentes

Investigación	Aporte
CNN ligero que clasifica maíz y malezas con alta precisión (>99 %) [11].	La investigación utilizó ECENet, y para interpretar sus decisiones empleó LIME y Grad-CAM. Aporte: Se complementará el uso de LIME y Grad-CAM con SHAP para mejorar la interpretabilidad visual en clasificación de malezas.
Modelo DenseNet-169 multiescala (DCFF, DMKCF) con Triplet Attention logra 99.07 % de accuracy y emplea Grad-CAM y LIME para explicar clasificación de malezas en sorgo [12].	Ofrece mapas de atención que revelan patrones discriminativos en imágenes de maleza. Aporte: El enfoque de atención será replicado agregando SHAP sobre un modelo YOLO entrenado con imágenes de cultivos de papa.
Implementación de Grad-CAM, LIME y SHAP en modelos de visión agrícola para validar decisiones en campo y aumentar transparencia [13].	Demuestra que XAI permite construir confianza en entornos reales, facilitando la validación visual de cada predicción del modelo. Aporte: Se implementarán técnicas XAI para entender los modelos de IA.
Revisión exhaustiva de DL (clasificación, detección, segmentación) en detección de malezas; identifica la limitada integración de técnicas XAI [14].	Señala la necesidad de incorporar XAI para facilitar la adopción de sistemas automáticos de detección de malezas. Aporte: Evaluaré Grad-CAM, LIME y SHAP en un modelo YOLO pre-entrenado para medir su explicabilidad en tareas de clasificación, detección y segmentación de malezas.
Uso de Slot Attention multiescala para generar mapas de atención de alta resolución que identifican regiones discriminativas de malezas en imágenes agrícolas [15].	Propone una arquitectura basada en transformers que produce explicaciones visuales. Aporte: Implementará mapas explicativos detallados en la arquitectura CNN con el modelo YOLO.
Comparación de mapas Grad-CAM entre YOLOv8n y YOLO-Weed Nano para detección de malezas en algodón [16].	Muestra que YOLO-Weed Nano enfoca mejor las regiones de planta de algodón y reduce la atención en áreas de suelo. Aporte: Aplicará mapas Grad-CAM a cultivos de papa con el modelo YOLOv12.

Tabla 1.

Resumen de investigaciones y aportes en el área de detección de malezas y XAI.

CAPÍTULO II

MARCO TEÓRICO

2.1 Detección de Malezas en la Agricultura de Precisión y el Rol del Deep Learning

La agricultura de precisión representa un cambio de paradigma en la gestión agrícola, enfocándose en la optimización del uso de insumos y la toma de decisiones sitio-específicas para mejorar la productividad y la sostenibilidad [17]. Dentro de este enfoque, el manejo de malezas es una de las tareas más críticas, ya que estas plantas no deseadas compiten directamente con los cultivos por recursos esenciales como la luz solar, el agua y los nutrientes del suelo, además de poder albergar plagas y enfermedades [18]. El Deep Learning (DL) ha emergido como una tecnología transformadora en este contexto, ofreciendo capacidades avanzadas para el análisis de imágenes y la automatización. Aplicado a la detección de malezas, el DL permite desarrollar sistemas capaces de identificar y localizar con alta precisión diferentes especies de malezas, incluso en etapas tempranas de crecimiento y en condiciones de campo complejas [19]. Esta capacidad facilita la implementación de estrategias de control más selectivas, como la pulverización dirigida de herbicidas o el deshierbe mecánico robotizado, lo que reduce el uso de agroquímicos, minimiza el impacto ambiental y optimiza los costos de producción [19].

2.1.1 Impacto de las Malezas en Cultivos de Papa y Métodos Tradicionales de Detección

El cultivo de papa (*Solanum tuberosum* L.) es uno de los pilares de la seguridad alimentaria mundial, pero su productividad es altamente susceptible a la interferencia de malezas [20]. Estas plantas invasoras pueden causar mermas de rendimiento que, en ausencia de control, pueden alcanzar hasta un 71 %, afectando no solo la cantidad sino también el tamaño y la calidad comercial de los tubérculos. El periodo crítico de competencia entre las malezas y el cultivo de papa se sitúa generalmente entre los 25 y 40 días posteriores a la siembra, una ventana durante la cual el control es fundamental para evitar pérdidas significativas [20]. Las pérdidas económicas estimadas por la interferencia de malezas en la papa son considerables; por ejemplo, en Estados

Unidos, se calcula un potencial de pérdida anual de US\$465 millones, y en Canadá, de US\$61 millones, asumiendo una pérdida de rendimiento promedio del 44 % si no se aplican medidas de control [18]. Además de la competencia directa, las malezas pueden actuar como hospederos alternativos para plagas y enfermedades que afectan a la papa y dificultar las labores de cosecha.

Los métodos tradicionales para la detección de malezas han dependido históricamente de la inspección visual por parte de los agricultores, quienes identifican las plantas no deseadas basándose en características morfológicas [21]. Este método, aunque directo, es subjetivo, propenso a errores por fatiga o falta de experiencia, y difícilmente escalable a grandes extensiones. Una vez detectadas, las estrategias de control tradicionales incluyen:

- **Control Manual:** Extracción física de las malezas. Es altamente selectivo pero extremadamente intensivo en mano de obra y costoso, resultando poco práctico para grandes superficies [21].
- **Control Mecánico:** Uso de maquinaria para labrar el suelo y eliminar malezas. Puede ser efectivo pero también puede dañar la estructura del suelo, promover la erosión y no es selectivo, pudiendo afectar al cultivo [22].
- **Control Cultural:** Prácticas agronómicas como la rotación de cultivos, la elección de fechas de siembra y el *mulching* (acolchado). El *mulching* con paja de arroz, por ejemplo, ha demostrado ser beneficioso en papa, controlando malezas y mejorando el rendimiento [21].
- **Control Químico:** Aplicación de herbicidas. Productos como la metribuzina (pre-emergencia) y la atrazina se han usado en papa, pero su uso indiscriminado conlleva riesgos ambientales (contaminación, daño a organismos no objetivo) y el desarrollo de resistencia en las malezas [21], [22].
- **Monitoreo:** Consiste en la inspección regular y sistemática de los campos para identificar las especies de malezas presentes, estimar su densidad y distribución, y evaluar la efectividad de las medidas de control implementadas previamente. Este monitoreo tra-

dicionalmente se realiza de forma manual, lo que puede ser subjetivo, consumir mucho tiempo y ser propenso a errores, especialmente en grandes extensiones [22].

Las limitaciones de estos métodos tradicionales son evidentes: altos costos, baja eficiencia en grandes extensiones, subjetividad en la detección, impacto ambiental negativo y falta de precisión para un manejo sitio-específico, lo que justifica la búsqueda de soluciones tecnológicas avanzadas [20].

2.1.2 Avances en Deep Learning para la Visión por Computadora en Agricultura

El Deep Learning (DL) ha emergido como una tecnología disruptiva con un vasto potencial para revolucionar numerosas industrias, y la agricultura no es una excepción. Su fortaleza radica en la capacidad de los algoritmos, particularmente en las redes neuronales convolucionales (CNN), para aprender de manera automática y jerárquica representaciones y características complejas a partir de grandes volúmenes de datos, como las imágenes capturadas en el campo [23]. Esta capacidad ha abierto nuevas fronteras para la visión por computadora en el ámbito agrícola, permitiendo el desarrollo de sistemas inteligentes capaces de percibir e interpretar el entorno agrícola con niveles de precisión y eficiencia antes inalcanzables.

Las aplicaciones del DL en la agricultura de precisión son diversas y de gran impacto. Incluyen el mapeo de tierras y la clasificación de tipos de cultivo, la predicción de rendimientos y, de manera crucial para este estudio, el monitoreo del estrés biótico (como la presencia de enfermedades, plagas o malezas) y abiótico (como deficiencias nutricionales o estrés hídrico) [23]. En el contexto del monitoreo del estrés de las plantas, las técnicas de clasificación, detección, segmentación y seguimiento de objetos en imágenes, potenciadas por el DL, son fundamentales. Estas permiten no solo identificar “qué” tipo de estrés está presente y “dónde” se localiza, sino también comprender “cómo” este estrés ocurre, se distribuye y evoluciona en el tiempo, informando así “cuándo” es necesaria una intervención humana específica y dirigida [23].

La investigación en DL aplicado a la agricultura ha experimentado un crecimiento exponencial, impulsado en parte por la creciente disponibilidad de conjuntos de datos de imágenes

agrícolas etiquetados y de acceso público. Se han identificado numerosos datasets específicos para tareas de visión por computadora en agricultura de precisión, incluyendo al menos quince dedicados explícitamente al control de malezas. Esta disponibilidad de datos es un catalizador esencial, ya que el aprendizaje supervisado, que es el paradigma dominante en muchas de estas aplicaciones, depende intrínsecamente de grandes cantidades de datos etiquetados de alta calidad para entrenar modelos robustos y generalizables [23].

Un área de particular interés es la adaptación de técnicas de DL para entornos agrícolas específicos, como los invernaderos, en el marco de la Agricultura en Ambiente Controlado (CEA, por sus siglas en inglés). En estos sistemas, el DL y la visión por computadora se están utilizando para tareas como el monitoreo detallado del crecimiento de las plantas, la detección temprana de enfermedades y la estimación precisa del rendimiento [24]. La capacidad de la IA y el DL para automatizar tareas que tradicionalmente son intensivas en mano de obra, como la inspección visual continua, es un factor clave para mejorar la eficiencia operativa general y la sostenibilidad de la producción agrícola moderna.

La especialización de los modelos de DL para entornos y problemas agrícolas concretos es una tendencia en auge. La existencia de revisiones enfocadas en aplicaciones para invernaderos y la proliferación de estudios que desarrollan modelos para cultivos específicos (como maíz, sorgo, algodón, y en el caso de esta tesis, la papa) indican que las soluciones genéricas de DL están siendo progresivamente adaptadas o, en muchos casos, desarrolladas *de novo* para abordar las condiciones y desafíos únicos de diferentes contextos agrícolas. Esta especialización es lógica y necesaria, ya que la variabilidad inherente a los distintos tipos de cultivos, las condiciones ambientales, las especies de malezas y las prácticas de manejo agrícola requiere modelos finamente ajustados para lograr un rendimiento óptimo y una aplicabilidad real en el campo.

2.1.3 Desafíos Actuales en la Detección de Malezas con Deep Learning

A pesar del progreso, la detección de malezas con DL enfrenta desafíos significativos para su implementación robusta en campo.

- **Preparación y Disponibilidad de Conjuntos de Datos:** La creación de datasets a gran escala, diversos y bien anotados (especialmente a nivel de instancia para segmentación) es costosa y laboriosa. Los datasets públicos suelen ser limitados en especies, etapas fenológicas y condiciones de captura [25].
- **Generalización y Robustez del Modelo:** Los modelos entrenados en un entorno específico a menudo fallan al enfrentarse a condiciones de campo variables (iluminación, oclusiones, fondos de suelo, diversidad morfológica de malezas). El problema de “verde sobre verde” (similitud entre cultivo joven y maleza) es particularmente difícil [17].
- **Desequilibrio de Clases:** La prevalencia de ciertas especies de malezas sobre otras conduce a datasets desequilibrados, sesgando el modelo hacia las clases mayoritarias [25].
- **Precisión en Campo y Requisitos de Tiempo Real:** Lograr alta precisión en condiciones de campo variables y con baja latencia para aplicaciones como la pulverización selectiva es un desafío. Esto exige modelos ligeros, lo que puede comprometer la exactitud [25].
- **Identificación a Nivel de Especie:** Para un manejo eficaz, se requiere la identificación precisa de la especie de maleza, lo cual es más complejo que la detección binaria cultivo/maleza [17].

Las soluciones propuestas incluyen la aumentación de datos (con GANs), aprendizaje por transferencia, aprendizaje con supervisión débil o nula, optimización de modelos para despliegue en borde (cuantización, poda), fusión de sensores (RGB + multiespectral/hiperespectral), aprendizaje incremental y el uso de IA Explicable (XAI) para entender y mejorar los modelos.

2.2 Inteligencia Artificial Explicable (XAI): Fundamentos y Necesidad en Modelos Agrícolas

La creciente adopción de modelos de Deep Learning (DL) en la agricultura, si bien prometedora, se enfrenta al obstáculo de su naturaleza de “caja negra”. La Inteligencia Artificial Explicable (XAI) surge para dotar de transparencia e interpretabilidad a estos sistemas, lo cual es

fundamental en un sector donde las decisiones algorítmicas tienen consecuencias directas.

2.2.1 El Problema de la “Caja Negra” en Deep Learning y sus Implicaciones

Los modelos de Deep Learning, especialmente las redes neuronales profundas, son a menudo denominados “cajas negras” debido a la dificultad para comprender cómo llegan a sus predicciones a partir de los datos de entrada. Esta opacidad se deriva de su arquitectura compleja, con múltiples capas interconectadas, funciones de activación no lineales y millones de parámetros ajustables, lo que dificulta trazar una lógica causal clara [26]. Las implicaciones de esta falta de transparencia son significativas:

- **Confianza y Adopción Limitadas:** Si los agricultores no entienden por qué un sistema de IA recomienda una acción, es menos probable que confíen en él y lo adopten [27].
- **Dificultades en Depuración y Mejora:** Identificar la causa de errores en un modelo opaco es arduo, impidiendo su optimización efectiva [26].
- **Obstáculos para la Responsabilidad:** En caso de fallos con consecuencias económicas o ambientales, la opacidad dificulta la atribución de responsabilidad [28].
- **Riesgo de Perpetuación de Sesgos:** Los modelos pueden aprender y amplificar sesgos presentes en los datos de entrenamiento si estos no son detectables debido a la opacidad [27].
- **Cumplimiento Regulatorio y Ético:** Crecientes demandas sociales y normativas (ej. GDPR) exigen transparencia en sistemas de IA. La XAI busca reconciliar la tensión entre el alto rendimiento predictivo de los modelos complejos y la necesidad de transparencia [28].

2.2.2 Definición, Objetivos y Taxonomía de las Técnicas XAI

La Inteligencia Artificial Explicable (XAI) se define como un conjunto de métodos y procesos que buscan hacer que los resultados y decisiones de los modelos de aprendizaje automático

sean comprensibles para los humanos, transformando “cajas negras” en sistemas más transparentes [28]. Los objetivos principales de XAI incluyen incrementar la transparencia, fomentar la confianza, facilitar la depuración, garantizar la equidad, asegurar la responsabilidad, cumplir con regulaciones y mejorar la colaboración humano-IA [26].

Las técnicas XAI se pueden clasificar según varios criterios:

■ **Momento de Explicación:**

- *Ante-hoc (Intrínsecos)*: Se refieren a modelos que son inherentemente interpretables por su propio diseño. Estos modelos, a menudo llamados “cajas de cristal” (*glass box models*), como los árboles de decisión simples, la regresión lineal o los sistemas basados en reglas, tienen una estructura y un proceso de toma de decisiones que son directamente comprensibles para los humanos. La explicabilidad está integrada en el modelo desde su concepción [29].
- *Post-hoc*: Estas técnicas se aplican *después* de que un modelo complejo y a menudo opaco (como una red neuronal profunda) ha sido entrenado. Su objetivo es explicar el comportamiento de este modelo de “caja negra” ya existente [29]. Ejemplos prominentes incluyen LIME, SHAP y Grad-CAM, que son el foco de esta tesis.

■ **Alcance de la Explicación:**

- *Local*: Se centran en explicar una predicción individual para una instancia específica de los datos. Ayudan a entender por qué el modelo tomó una decisión particular para ese caso concreto. LIME y las explicaciones de instancia de SHAP son ejemplos de métodos locales [29], [30].
- *Global*: Buscan describir el comportamiento general del modelo a través de todo el conjunto de datos o un subconjunto significativo de él. Pueden incluir la importancia general de las características o la identificación de interacciones entre ellas. La agregación de valores SHAP a través de múltiples instancias puede proporcionar una visión global [26], [29].

■ **Dependencia del Modelo:**

- *Específicas del Modelo:* Estas técnicas están diseñadas para funcionar con una clase particular de modelos de aprendizaje automático, aprovechando su arquitectura o propiedades internas. Por ejemplo, Grad-CAM está diseñado específicamente para redes neuronales convolucionales (CNNs) [26], [29].
- *Agnósticas al Modelo:* Estas técnicas pueden aplicarse, en principio, a cualquier modelo de aprendizaje automático, independientemente de su estructura interna. Tratan al modelo como una caja negra y analizan la relación entre sus entradas y salidas para generar explicaciones. LIME y SHAP (en su variante Kernel SHAP) son ejemplos de métodos modelo-agnósticos [29], [30].

- **Tipo de Datos de Entrada:** Las técnicas XAI también pueden categorizarse según el tipo de datos para los que están diseñadas o son más adecuadas, como datos tabulares, imágenes, texto o series temporales. Cada tipo de dato puede requerir enfoques de perturbación o representación de características diferentes [26].

- **Formato de la Explicación:** Los métodos XAI generan explicaciones en diversos formatos, tales como:

- Puntuaciones de importancia de características (e.g., SHAP, LIME).
- Mapas de activación o de prominencia (*saliency maps*) (e.g., Grad-CAM, Score-CAM).
- Reglas (e.g., if-then).
- Ejemplos contrafácticos (que muestran qué cambios mínimos en la entrada hubieran alterado la predicción).
- Prototipos o ejemplos representativos.

2.2.3 Importancia de XAI en la Agricultura para la Toma de Decisiones Informada y la Adopción Tecnológica

La XAI es crucial en la agricultura para transformar los sistemas de IA en herramientas colaborativas que potencien el conocimiento del agricultor, en lugar de reemplazarlo. Al proporcionar explicaciones sobre cómo un modelo llega a una recomendación (ej. por qué se clasifica una planta como maleza basándose en la forma de sus hojas), XAI permite a los agricultores validar estas sugerencias con su propia experiencia, tomar decisiones más informadas y fundamentadas sobre el manejo del cultivo (ej. qué herbicida usar o si es necesario aplicar control) [31].

Esta transparencia es fundamental para fomentar la confianza, un factor primordial para la adopción de nuevas tecnologías en un sector con aversión al riesgo. Si los agricultores comprenden la lógica detrás de las predicciones de la IA, es más probable que integren estas herramientas en sus prácticas, llevando a una optimización más efectiva de recursos (agua, fertilizantes, pesticidas) y contribuyendo a una agricultura más sostenible y rentable [32]. XAI también facilita la depuración de modelos, la capacitación y la transferencia de conocimiento, y el cumplimiento de futuras normativas sobre IA en la producción de alimentos [27].

2.2.4 Desafíos y Limitaciones Inherentes a las Técnicas de Explicabilidad

A pesar de sus beneficios, las técnicas XAI presentan desafíos y limitaciones intrínsecas.

- **Compensación Interpretabilidad-Fidelidad/Precisión:** Existe un *trade-off*: explicaciones más simples pueden no ser completamente fieles al complejo razonamiento del modelo original, mientras que modelos inherentemente interpretables pueden tener menor precisión predictiva [33].
- **Costo Computacional:** Métodos *post-hoc* como LIME y KernelSHAP, basados en perturbaciones, pueden ser computacionalmente intensivos y lentos, especialmente para datos de alta dimensionalidad como imágenes, limitando su uso en tiempo real [33].
- **Estabilidad y Robustez:** Algunas explicaciones pueden ser inestables (cambiar drásti-

camente con pequeñas variaciones en la entrada) o vulnerables a ataques adversariales [34].

- **Subjetividad en la Interpretación:** Lo que constituye una “buena” explicación depende del usuario (ej. científico de datos vs. agricultor) y su contexto [33].
- **Falta de Métricas de Evaluación Estandarizadas:** Evaluar la calidad de las explicaciones de forma objetiva es difícil debido a la carencia de métricas universales [33].
- **Posibilidad de Explicaciones Engañosas:** Múltiples explicaciones diferentes podrían ajustarse a un modelo complejo, llevando a una falsa sensación de comprensión [34].
- **Aplicabilidad Limitada:** No todas las técnicas son universales; algunas son específicas para ciertos modelos (Grad-CAM para CNNs) o tipos de datos. La ausencia de definiciones estandarizadas para XAI también genera confusión, ya que los términos pueden solaparse o variar entre estudios [34].

2.3 Técnicas XAI Post-hoc para Modelos de Visión por Computadora

Dado que los modelos de Deep Learning de vanguardia utilizados en visión por computadora, son inherentemente complejos y operan como “cajas negras”, las técnicas XAI *post-hoc* son esenciales para desentrañar sus procesos de toma de decisiones. Estos métodos se aplican después de que el modelo ha sido entrenado y buscan proporcionar explicaciones sobre su comportamiento sin modificar su arquitectura interna.

2.3.1 SHAP (SHapley Additive exPlanations): Fundamento y Aplicación

SHAP (SHapley Additive exPlanations) es un marco unificado basado en los valores de Shapley de la teoría de juegos cooperativos, que asigna equitativamente la contribución de cada característica de entrada a la predicción del modelo. Los valores SHAP cuantifican la contribución marginal promedio de una característica considerando todas las combinaciones posibles de otras características y poseen propiedades deseables como eficiencia (suma de valores SHAP iguala la predicción menos la base), simetría, característica nula y consistencia [35].

En visión por computadora, donde los píxeles individuales son características poco interpretables, SHAP se aplica a superpíxeles o segmentos de imagen [36].

- **KernelSHAP:** Es agnóstico al modelo. Genera imágenes perturbadas (combinaciones de superpíxeles “presentes” u “ocultos”), obtiene predicciones del modelo original y ajusta un modelo lineal local ponderado para estimar los valores SHAP de los superpíxeles [37].
- **DeepSHAP:** Más eficiente para redes neuronales, se basa en DeepLIFT, propagando diferencias de activación respecto a una entrada de referencia (ej. imagen negra) para aproximar los valores SHAP [37].

Las explicaciones SHAP para imágenes se visualizan como mapas de calor superpuestos, donde regiones con valores SHAP positivos altos (rojo) contribuyeron a la clase de interés, y valores negativos altos (azul) influyeron en contra.

- **Fortalezas:** Base teórica sólida, consistencia, precisión local, explicaciones locales y globales.
- **Limitaciones:** KernelSHAP es computacionalmente costoso [37]. DeepSHAP es sensible a la elección de la referencia y la calidad de la segmentación en superpíxeles puede afectar la interpretabilidad [37].

2.3.2 LIME (Local Interpretable Model-agnostic Explanations): Fundamento y Aplicación

LIME (Local Interpretable Model-agnostic Explanations) es una técnica de XAI *post-hoc* diseñada para explicar las predicciones individuales de cualquier modelo de aprendizaje automático, tratándolo como una “caja negra” [30]. Su principio fundamental es aproximar el comportamiento del modelo complejo en la vecindad de la instancia que se desea explicar, utilizando para ello un modelo interpretable más simple, como un modelo lineal disperso o un árbol de decisión pequeño [38].

El proceso implica:

1. **Generación de Muestras Perturbadas:** LIME genera un conjunto de nuevas muestras de datos en la vecindad de la instancia original. Esto se hace perturbando las características de la instancia original. Para datos tabulares, esto puede implicar modificar los valores de las características. Para datos de imagen, LIME primero segmenta la imagen en regiones contiguas de píxeles similares, llamadas superpíxeles. Luego, genera instancias perturbadas activando o desactivando aleatoriamente algunos de estos superpíxeles (es decir, manteniendo la parte del superpíxel de la imagen original o reemplazándola, por ejemplo, con un color promedio o gris). Para datos de texto, se pueden eliminar palabras o reemplazarlas [38].
 2. **Obtención de Predicciones:** Para cada una de estas muestras perturbadas, se obtiene la predicción del modelo de caja negra original.
 3. **Ponderación de las Muestras Perturbadas:** Cada muestra perturbada se pondera según su proximidad a la instancia original. Las muestras más cercanas a la instancia original reciben un peso mayor, reflejando la idea de una explicación “local”. Típicamente se usa un kernel exponencial para calcular estas ponderaciones [38].
 4. **Entrenamiento de Modelo Sustituto Local:** Se ajusta un modelo interpretable (e.g., regresión lineal con regularización Lasso para inducir dispersión y seleccionar un pequeño número de características importantes) utilizando las muestras perturbadas como datos de entrada y las predicciones del modelo de caja negra como etiquetas. El ajuste se realiza utilizando los pesos de proximidad [38].
 5. **Derivación de la Explicación:** Las características del modelo interpretable local (e.g., los coeficientes de la regresión lineal) se utilizan como la explicación para la predicción de la instancia original. Estos indican qué características de la instancia original fueron más influyentes para la predicción del modelo de caja negra en esa localidad.
- **Fortalezas:** Agnóstico al modelo, explicaciones intuitivas y locales, flexibilidad en la

definición de características interpretables.

- **Limitaciones:** Fidelidad solo local, posible inestabilidad de las explicaciones, dependencia de la calidad de la segmentación/perturbación, y la suposición de linealidad local puede no siempre cumplirse.

2.3.3 Grad-CAM (Gradient-weighted Class Activation Mapping): Fundamento y Aplicación

Grad-CAM (Gradient-weighted Class Activation Mapping) es una técnica de XAI *post-hoc*, modelo-específica, diseñada para producir explicaciones visuales para las decisiones de las Redes Neuronales Convolucionales (CNNs). Su objetivo principal es generar un mapa de localización grueso, comúnmente llamado mapa de calor (*heatmap*), que resalte las regiones de una imagen de entrada que fueron más influyentes para que la CNN tomara una predicción específica, como la clasificación de un objeto en una categoría particular [39].

El fundamento teórico de Grad-CAM se basa en el uso de la información de los gradientes que fluyen hacia las capas convolucionales finales de la red. El proceso general es el siguiente:

1. **Propagación Hacia Adelante (Forward Pass):** La imagen de entrada se propaga a través de la CNN para obtener las puntuaciones de clase (antes de la capa softmax) y los mapas de activación de una capa convolucional objetivo (generalmente la última capa convolucional, ya que contiene la información semántica de más alto nivel con una resolución espacial aún razonable) [39].
2. **Cálculo de Gradientes:** Para una clase de interés c , se calcula el gradiente de la puntuación de esa clase y^c con respecto a cada uno de los k mapas de activación A^k de la capa convolucional seleccionada. Es decir, se calcula $\partial y^c / \partial A_{ij}^k$ para cada posición (i, j) en cada mapa de activación k [39].
3. **Ponderación de Neuronas/Mapas de Activación (α_k^c):** Estos gradientes se promedian globalmente a lo largo de las dimensiones espaciales (ancho y alto) para obtener un peso

de importancia α_k^c para cada mapa de activación A^k con respecto a la clase c .

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (2.1)$$

Donde Z es el número total de píxeles en el mapa de activación. Este peso α_k^c representa la importancia de cada mapa de activación para la clase objetivo [39].

4. **Combinación Lineal Ponderada:** Se realiza una combinación lineal ponderada de los mapas de activación A^k utilizando los pesos α_k^c calculados [39].
5. **Aplicación de ReLU:** Finalmente, se aplica una función de activación ReLU (Rectified Linear Unit) a esta combinación lineal.

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right) \quad (2.2)$$

La ReLU es crucial porque solo se consideran las características que tienen una influencia positiva en la clase de interés. Esto ayuda a asegurar que el mapa de calor resalte las regiones que contribuyen a la predicción de esa clase específica, en lugar de regiones que podrían estar inhibiéndola o contribuyendo a otras clases [39].

El resultado, $L_{\text{Grad-CAM}}^c$, es un mapa de calor de la misma dimensión que los mapas de activación de la capa convolucional elegida, que luego se puede redimensionar a la imagen original y superponer para una visualización intuitiva.

2.3.4 Score-CAM: Fundamento y Aplicación

Score-CAM (Score-Weighted Class Activation Mapping) es una técnica XAI *post-hoc* que, a diferencia de Grad-CAM, genera explicaciones visuales para CNNs sin depender de gradientes. Su principio se basa en ponderar cada mapa de activación de una capa convolucional según el aumento en la puntuación de confianza (score) de la clase de interés que se obtiene al usar dicho mapa de activación como una máscara sobre la imagen de entrada y pasarla por el modelo [9].

Cada mapa de activación se normaliza y se utiliza para perturbar la imagen; la predicción del modelo para la clase objetivo sobre esta imagen perturbada sirve como peso. Una combinación lineal de estos mapas de activación ponderados, seguida de una ReLU, produce el mapa de calor final. Score-CAM se aplica en visión por computadora para identificar regiones cruciales en una imagen para una predicción, siendo útil en diagnóstico médico (nódulos pulmonares, retinopatía diabética) y detección de objetos [9].

2.3.5 Comparativa de Robustez y Dependencia de Gradientes: Grad-CAM vs. Score-CAM

La diferencia fundamental entre Grad-CAM y Score-CAM es la dependencia de gradientes: Grad-CAM los utiliza para ponderar los mapas de características, mientras que Score-CAM es libre de gradientes, basando la importancia en el impacto de los mapas de activación (utilizados como máscaras) sobre la puntuación de la clase.

- **Robustez y Calidad Visual:** Score-CAM tiende a producir explicaciones visuales más suaves, menos ruidosas y más robustas, especialmente en arquitecturas complejas, al evitar problemas de gradientes saturados o desvanecidos [9].
- **Costo Computacional:** Score-CAM es significativamente más costoso computacionalmente en inferencia que Grad-CAM, ya que requiere múltiples pasadas hacia adelante (*forward passes*) a través del modelo (una por cada mapa de activación) para calcular los pesos, mientras Grad-CAM solo necesita una pasada hacia adelante y una hacia atrás [39], [9].

En resumen, Score-CAM ofrece explicaciones visuales de mayor calidad y robustez a cambio de un mayor costo computacional, mientras que Grad-CAM es más rápido pero su calidad puede verse afectada por la de los gradientes.

2.4 Herramientas y Librerías de Código Abierto para XAI en Deep Learning

2.4.1 Panorama de Librerías Python (ej. SHAP, LIME, Tf-explain) y sus Capacidades

Varias librerías Python de código abierto facilitan la implementación de XAI:

- **SHAP:** Implementa valores de Shapley con varios explicadores (TreeExplainer, DeepExplainer, KernelExplainer, LinearExplainer). Es valorado por su base teórica y consistencia [35].
- **LIME:** Explica predicciones individuales ajustando modelos interpretables locales. Compatible con datos tabulares, texto e imágenes. Destaca por su interpretabilidad intuitiva y naturaleza agnóstica [30].
- **tf-explain:** Específico para TensorFlow 2.0, ofrece Grad-CAM, Occlusion Sensitivity, SmoothGrad, etc. Se integra con tf.keras.Callback y TensorBoard [40].
- **Captum:** Librería de PyTorch que implementa Integrated Gradients, Saliency Maps, DeepLIFT, GradientSHAP, TCAV, entre otros. Ofrece atribuciones a nivel de característica, capa y neurona, y es compatible con torchvision y torchtext [41].
- **Alibi Explain:** Proporciona métodos *black-box* y *white-box* como Anchors, Contrastive Explanations, Counterfactuals, Integrated Gradients, KernelSHAP, TreeSHAP, ALE y PDP para clasificación y regresión [42].
- **InterpretML:** Ofrece modelos “glass-box” (como Explainable Boosting Machine - EBM, modelos lineales, árboles de decisión) y explicadores “black-box” (LIME, SHAP, PDP). Incluye visualizaciones y es compatible con scikit-learn [43].
- **OmniXAI:** Busca ser una librería integral con LIME, SHAP, PDP, ALE para diversos tipos de datos y modelos en diferentes etapas del proceso de ML [44].

Estas librerías varían en los métodos XAI que implementan, su compatibilidad con frameworks de DL y los tipos de datos y modelos que soportan.

2.5 Aplicaciones y Evaluación de XAI en la Detección de Malezas

2.5.1 Métricas de Evaluación del Rendimiento del Modelo de Detección (Precision, Recall, mAP, F1-Score)

El rendimiento de los modelos de detección de objetos, como los utilizados para la detección de malezas, se evalúa cuantitativamente utilizando un conjunto de métricas estándar:

- **Intersección sobre Unión (IoU):** Es una métrica fundamental que mide el grado de solapamiento entre el cuadro delimitador predicho por el modelo (Bp) y el cuadro delimitador real o de referencia (*ground truth*, Bgt). Se calcula como el área de la intersección de los dos cuadros dividida por el área de su unión. Un umbral de IoU (comúnmente 0.5) se utiliza para clasificar una detección como Verdadero Positivo (TP) si el IoU supera el umbral, o Falso Positivo (FP) si no lo supera o si es una detección duplicada [45].
- **Precisión (Precision):** Mide la exactitud de las predicciones positivas. Es la proporción de detecciones correctas (TP) sobre el total de detecciones realizadas por el modelo (TP + FP). Una alta precisión indica que el modelo tiene pocos falsos positivos [46].
- **Recall (Sensibilidad o Exhaustividad):** Mide la capacidad del modelo para encontrar todos los objetos relevantes. Es la proporción de detecciones correctas (TP) sobre el total de objetos reales presentes en los datos (TP + Falsos Negativos, FN). Un alto recall indica que el modelo tiene pocos falsos negativos [46].
- **Puntuación F1 (F1-Score):** Es la media armónica de la Precisión y el Recall, proporcionando una métrica única que equilibra ambas. Es útil cuando se desea un balance entre evitar falsos positivos y falsos negativos [46].

Estas métricas son esenciales para comparar objetivamente el rendimiento de diferentes modelos de detección de malezas y para ajustar los modelos con el fin de optimizar su desempeño en aplicaciones agrícolas prácticas.

2.6 Marco Metodológico Estándar para Proyectos de Inteligencia Artificial

2.6.1 Introducción a la Metodología CRISP-DM (Cross-Industry Standard Process for Data Mining)

CRISP-DM (Cross-Industry Standard Process for Data Mining) es un modelo de proceso estándar abierto, robusto, flexible y ampliamente adoptado que guía los proyectos de minería de datos y ciencia de datos. Desarrollado en la década de 1990 por un consorcio de empresas, su objetivo es proporcionar una hoja de ruta estructurada para mejorar la eficiencia y efectividad de estos proyectos, asegurando que los resultados estén alineados con los objetivos del negocio [47]. CRISP-DM descompone el ciclo de vida de un proyecto de minería de datos en seis fases principales e interconectadas, permitiendo la iteración entre ellas:

[Image of CRISP-DM process diagram]

1. **Comprensión del Negocio (Business Understanding):** Definir objetivos y requisitos del proyecto desde una perspectiva empresarial, convertir este conocimiento en una definición del problema de minería de datos y diseñar un plan preliminar [48].
2. **Comprensión de los Datos (Data Understanding):** Recolección inicial de datos, familiarización con los datos, identificación de problemas de calidad, descubrimiento de primeros insights y/o formulación de hipótesis sobre información oculta [48].
3. **Preparación de los Datos (Data Preparation):** Cubre todas las actividades para construir el conjunto de datos final (datos que se introducirán en la herramienta de modelado) a partir de los datos brutos iniciales. Incluye selección de tablas, registros y atributos, así como transformación y limpieza de datos [48].
4. **Modelado (Modeling):** Selección y aplicación de diversas técnicas de modelado (ej. algoritmos de aprendizaje automático) y calibración de sus parámetros para optimizar los resultados. Típicamente, existen varias técnicas para el mismo problema de minería de datos [48].

5. **Evaluación (Evaluation):** Evaluación exhaustiva del modelo y revisión de los pasos ejecutados para construirlo, asegurando que aborde adecuadamente los objetivos del negocio antes de proceder al despliegue final. Se determina si se han alcanzado los objetivos empresariales y se decide sobre los próximos pasos [48]
6. **Despliegue (Deployment):** El conocimiento obtenido se organiza y presenta de manera que el cliente pueda utilizarlo. Puede implicar desde la generación de un informe hasta la implementación de un proceso de minería de datos repetible en la empresa. La naturaleza cíclica de CRISP-DM permite que las lecciones aprendidas en un proyecto informen y mejoren proyectos futuros [48].

La integración de XAI en MLOps (Operaciones de Aprendizaje Automático), que a menudo se basa en principios de CRISP-DM, busca lograr una transparencia de extremo a extremo en el ciclo de vida de la IA. Frameworks específicos, como uno propuesto para la industria de semiconductores, ya integran XAI desde las etapas tempranas del diseño y desarrollo, enfatizando la colaboración con expertos del dominio y la generación de informes de explicabilidad.

2.7 Trabajos Relacionados

El desarrollo de sistemas de visión artificial para la agricultura de precisión ha evolucionado significativamente en los últimos años. A continuación, se revisan las contribuciones más relevantes en tres ejes fundamentales para esta investigación: el uso de Deep Learning para la detección de malezas, la aplicación específica de segmentación semántica en cultivos andinos y la incorporación reciente de técnicas de explicabilidad (XAI).

2.7.1 Deep Learning en la Detección de Malezas

La literatura actual evidencia un dominio de las redes neuronales convolucionales (CNN) como el estándar para tareas de monitoreo agrícola. En una revisión exhaustiva del estado del arte, Pai et al. [13] analizan múltiples arquitecturas, desde modelos de clasificación simple hasta detectores avanzados como YOLO y Mask R-CNN. Los autores concluyen que, si bien el

rendimiento en métricas como precisión y mAP ha alcanzado niveles superiores al 90 %, existe una carencia generalizada de interpretabilidad en los modelos desplegados, lo que limita su confianza en entornos operativos reales.

En esta línea de alto rendimiento, Nidhya et al. [49] propusieron recientemente un enfoque basado en CNNs para la detección automática en campos de cultivo, reportando tasas de exactitud cercanas al 98 %. Su trabajo demuestra la robustez de las CNN para extraer características en entornos no estructurados. Sin embargo, su enfoque se centra en la clasificación y detección mediante cuadros delimitadores (bounding boxes), una técnica que, aunque efectiva para conteo, resulta insuficiente para aplicaciones que requieren estimaciones precisas de biomasa o rociado de precisión a nivel de hoja.

2.7.2 Segmentación Semántica en Cultivos de Papa

Para superar las limitaciones de la detección por cajas, la segmentación semántica ha ganado tracción por su capacidad de clasificar cada píxel de la imagen. En el contexto local y regional, Pazos [1] desarrolló un estudio fundamental utilizando la arquitectura UNet para cuantificar malezas en cultivos de maíz y papa mediante imágenes de dron.

Pazos demostró la viabilidad de separar la clase “cultivo” de la clase “maleza” utilizando índices de vegetación y redes totalmente convolucionales. No obstante, su trabajo se centró en la optimización de métricas de superposición estándar (IoU y Dice), sin abordar los problemas de “caja negra” inherentes a la red. Además, las limitaciones en el tamaño del dataset y el desbalance de clases en su estudio sugieren que arquitecturas estándar como UNet pueden tener dificultades para generalizar en bordes difusos sin mecanismos de atención adicionales.

2.7.3 Inteligencia Artificial Explicable (XAI) en Agricultura

La integración de XAI en la agricultura es un campo emergente que busca dotar de transparencia a los modelos predictivos. Ahmed et al. [50] presentaron “ExE-Net”, un modelo de ensamble para la clasificación de enfermedades en hojas de papa. Su aporte clave fue la inte-

gración de mapas de calor para validar visualmente qué regiones de la hoja indicaban patología. Aunque pionero en el cultivo de papa, su alcance se limita a la clasificación de imágenes de hojas individuales en condiciones controladas, distando de la complejidad de la segmentación de campo abierto.

Paralelamente, Mamatha y Reddy han explorado la explicabilidad en cultivos extensivos como la soja. En su trabajo de 2024 [51], aplicaron Vision Transformers combinados con Grad-CAM, y posteriormente, en 2025 [52], extendieron este análisis a arquitecturas CNN utilizando SHAP.

Si bien Mamatha y Reddy [52] logran proporcionar explicaciones locales y globales sobre la importancia de las características, su metodología de evaluación de XAI permanece cualitativa. Los autores utilizan las explicaciones principalmente para validación visual post-hoc (confirmación humana), sin establecer métricas cuantitativas de la calidad de la explicación (como fidelidad o localización) ni utilizar estos hallazgos para reentrenar o mejorar la arquitectura del modelo, vacío que la presente investigación busca llenar.

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 Enfoque Metodológico y Estándar de Proceso

Para estructurar el ciclo de vida experimental, garantizar la reproducibilidad de los hallazgos y alinear los resultados técnicos con los objetivos de la agricultura de precisión, esta investigación se rige por la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining).

Este estándar industrial se ha adaptado específicamente para integrar la Inteligencia Artificial Explicable (XAI) no como una etapa aislada, sino como un requisito transversal que valida la confiabilidad del modelo. El proceso se divide en seis fases iterativas, donde los hallazgos de interpretabilidad de una fase retroalimentan el diseño de la siguiente, mitigando el riesgo de desarrollar modelos de “caja negra” propensos a sesgos en entornos agrícolas reales.

A continuación, se detalla la operacionalización de cada fase dentro del contexto de este estudio:

Tabla 2.
Adaptación del Ciclo CRISP-DM al Estudio de Segmentación de Malezas.

Fase CRISP-DM	Actividad Experimental Específica	Herramientas y Tecnologías
1. Comprensión del Negocio	Definición de la problemática de detección de malezas en cultivos de papa y establecimiento de métricas de éxito (IoU) y criterios de explicabilidad para agrónomos.	Revisión bibliográfica, Python.
2. Comprensión de Datos	Análisis exploratorio del dataset <i>Weeds in Potato Fields v2</i> , verificación del desbalance de clases (Papa vs. Malezas) y evaluación de la calidad de las máscaras de anotación.	Pandas, Matplotlib, OpenCV, NumPy.
3. Preparación de Datos	Implementación de pipelines diferenciados: Pre-procesamiento estándar para el Escenario A y Salto de Resolución (256px) con aumento agresivo para el Escenario B.	Albumentations, PyTorch Datasets.
4. Modelado	Entrenamiento de la arquitectura <i>Attentive DeepLabv3+</i> (EfficientNetV2-S). Comparativa entre el Modelo Original (Escenario A) y el Modelo Optimizado con SWA (Escenario B).	PyTorch, Timm, AdamW Optimizer, Kaggle GPU T4.
5. Evaluación	Validación cuantitativa (mIoU por clase) y auditoría profunda de explicabilidad mediante métricas de Fidelidad, Localización y Robustez.	pytorch-grad-cam, SHAP, LIME, scikit-image.
6. Despliegue	Generación de reportes de interpretabilidad visual (mapas de calor) y análisis comparativo de la toma de decisiones del modelo para su validación técnica.	Jupyter Notebooks, Matplotlib, Seaborn.

3.2 Fase 1: Entorno de Desarrollo e Infraestructura Computacional

Para la ejecución de los experimentos de Inteligencia Artificial Explicable (XAI) y la evaluación del modelo de segmentación, se utilizó un entorno computacional basado en la nube. Se seleccionó esta infraestructura por su disponibilidad y capacidad para manejar cargas de trabajo de aprendizaje profundo, garantizando la reproducibilidad de los resultados mediante el uso de contenedores estandarizados.

3.2.1 Recursos de Hardware

El procesamiento se llevó a cabo utilizando la plataforma Kaggle, la cual proporcionó acceso a aceleradores gráficos necesarios para la inferencia del modelo y el cálculo de las métricas de explicabilidad. Las especificaciones técnicas del hardware utilizado son las siguientes:

- **Unidad de Procesamiento Gráfico (GPU):** NVIDIA Tesla T4.
- **Arquitectura:** Turing.
- **Memoria de Video (VRAM):** 16 GB GDDR6.
- **Ancho de banda de memoria:** 320 GB/s.

Este acelerador T4 ofreció el rendimiento necesario para procesar los tensores de imágenes de 256x256 y ejecutar los algoritmos de perturbación (como LIME y SHAP) en tiempos razonables, manteniendo la eficiencia energética característica de esta arquitectura.

3.2.2 Entorno de Software y Librerías

El flujo de trabajo se implementó utilizando el lenguaje de programación Python, gestionado a través de Jupyter Notebooks en el entorno Linux de Kaggle. Este entorno proporciona una integración nativa con las principales bibliotecas de ciencia de datos, lo que facilitó la carga del modelo y la generación de visualizaciones.

Las principales librerías empleadas en esta investigación incluyen:

- **PyTorch:** Framework principal utilizado para la carga del modelo pre-entrenado (.pth) y la ejecución de operaciones tensoriales en la GPU.
- **Albumentations:** Utilizada para el preprocesamiento y normalización de las imágenes de entrada, asegurando consistencia con los datos de entrenamiento.
- **Librerías de XAI:** Se integraron herramientas específicas para la interpretación de modelos:
 - `pytorch-grad-cam`: Para la generación de mapas de activación (Grad-CAM, Score-CAM).
 - `shap`: Para la estimación de valores Shapley mediante *Kernel Explainer*.
 - `lime`: Para la creación de explicaciones locales mediante perturbaciones.
- **Visualización y Manejo de Datos:** NumPy y Pandas para el manejo de estructuras de datos, junto con Matplotlib y OpenCV para la generación de las figuras compuestas de análisis.

3.3 Fase 2: Comprensión y Adquisición de los Datos

El insumo principal para la evaluación del modelo y la aplicación de las técnicas de explicabilidad es el conjunto de datos de imágenes agrícolas, el cual fue recopilado, estructurado y anotado previamente como parte del trabajo de titulación de la carrera de Ingeniería en Software desarrollado por Pazos Morillo. Este dataset proporciona una base robusta de imágenes reales de cultivos de papa conviviendo con diversas especies de malezas, capturadas en entornos controlados y naturales.

3.3.1 Origen y Adquisición de Datos

La fase de recolección de datos se llevó a cabo en terrenos agrícolas ubicados en los cantones de San Gabriel, El Ángel y Cotacachi, en la provincia de Imbabura y Carchi. Para la adquisición

de las imágenes se utilizó un vehículo aéreo no tripulado (dron) modelo DJI Mavic 2 Pro. Las capturas se realizaron configurando el dispositivo a una altura de vuelo de 9 metros y una velocidad constante de 1 m/s, lo que permitió obtener imágenes de alta resolución (5472×3648 píxeles) con un nivel de detalle suficiente para la identificación morfológica de las especies vegetales.

Posterior a la captura, las imágenes fueron sometidas a un proceso de anotación manual asistida mediante la plataforma Roboflow. Este proceso generó las máscaras de segmentación semántica que sirven como *Ground Truth* para el entrenamiento y la evaluación del modelo.

3.3.2 Composición de Clases

El dataset fue etiquetado considerando las especies predominantes en los cultivos de la región. Se definieron seis clases semánticas para la segmentación, las cuales permiten discriminar entre el cultivo de interés, las malezas específicas y el fondo. Las categorías son las siguientes:

- **Clase 0 (Fondo):** Comprende el suelo, piedras, sombras y cualquier elemento que no corresponda a material vegetal.
- **Clase 1 (Lengua de Vaca):** Maleza de hoja ancha (*Rumex obtusifolius*).
- **Clase 2 (Diente de León):** Maleza común (*Taraxacum officinale*).
- **Clase 3 (Kikuyo):** Gramínea invasora (*Cenchrus clandestinus*) muy frecuente en la zona.
- **Clase 4 (Otras):** Agrupa otras especies de malezas menos frecuentes o no clasificadas en las categorías anteriores.
- **Clase 5 (Papa):** Corresponde a la planta del cultivo de interés (*Solanum tuberosum*).

3.4 Fase 3: Preparación de los datos

Para la ejecución de los experimentos de XAI en esta investigación, se utilizó el subconjunto de prueba (Test Set) derivado del trabajo original. Dado que el modelo *Attentive DeepLabv3+*

utilizado requiere entradas de dimensiones específicas, se aplicó un flujo de preprocesamiento en el script de evaluación consistente con el utilizado durante el entrenamiento:

1. **Redimensionamiento:** Las imágenes originales o recortes fueron ajustados a una resolución espacial de 256×256 píxeles. Esta dimensión difiere del trabajo original de Pazos (quien usó 128×128), debido a las adaptaciones realizadas en la arquitectura del modelo actual para capturar mayor contexto espacial.
2. **Normalización:** Se aplicó una estandarización de los valores de píxel utilizando las medias y desviaciones estándar del dataset ImageNet. Esta transformación asegura que la distribución de los datos de entrada coincida con la de los pesos pre-entrenados del backbone EfficientNetV2-S.
 - Media RGB: $[0,485, 0,456, 0,406]$
 - Desviación Estándar RGB: $[0,229, 0,224, 0,225]$
3. **Conversión a Tensores:** Las matrices de imágenes procesadas se convirtieron a tensores de punto flotante compatibles con la librería PyTorch y se transfirieron a la memoria de la GPU para su procesamiento.

3.5 Fase 4: Modelado y Arquitectura de la Red

3.5.1 Metodología de Evaluación Comparativa

Para validar el impacto de las optimizaciones propuestas y cuantificar la mejora en la capacidad de generalización del modelo, esta investigación establece un marco comparativo entre dos configuraciones distintas de la misma arquitectura:

- **Escenario A (Modelo Base):** Corresponde a la configuración estándar del modelo entrenado de forma supervisada clásica. Este escenario representa el desempeño de referencia (*baseline*) obtenido mediante la minimización directa de la función de pérdida.

- **Escenario B (Modelo Optimizado - SWA):** Constituye la versión evolucionada del modelo. En este escenario, se aplicó una estrategia de reentrenamiento con salto de resolución y estabilización de pesos mediante *Stochastic Weight Averaging* (SWA) para robustecer las fronteras de decisión [53].

Arquitectura del Modelo (Común para Escenarios A y B)

Independientemente del escenario de entrenamiento, el núcleo de la investigación se basa en la arquitectura de segmentación semántica *Attentive DeepLabv3+*. Este modelo es una variante personalizada diseñada para la agricultura de precisión, integrando componentes específicos para manejar la variabilidad de la vegetación [53]. A continuación, se describen los componentes estructurales que comparten ambos escenarios.

3.5.2 Estructura General de la Red

El modelo sigue un paradigma *Encoder-Decoder* optimizado que integra un módulo de pirámide espacial (ASPP) y mecanismos de atención dual para refinar la selección de características relevantes antes de la segmentación final [53].

3.5.2.1 Encoder (Extractor de Características)

Como columna vertebral (*backbone*) se utiliza la red EfficientNetV2-S, pre-entrenada en el conjunto de datos ImageNet-21k [54]. Esta elección se justifica por su equilibrio entre eficiencia computacional y precisión. El codificador extrae mapas de características en cuatro niveles de profundidad (índices 0, 1, 2, 3), proporcionando una representación multiescala robusta necesaria para detectar tanto malezas pequeñas como grandes áreas de cultivo .

3.5.2.2 Módulo ASPP (Atrous Spatial Pyramid Pooling)

Para capturar el contexto global sin perder resolución espacial, se incorpora un módulo ASPP en la tercera etapa del codificador. Este bloque emplea convoluciones dilatadas con tasas de 6, 12 y 18, permitiendo un campo receptivo variable. Adicionalmente, incluye una capa de *Dropout* (0.5) para regularizar el modelo.

3.5.2.3 Decoder con Mecanismos de Atención

El decodificador implementa una estructura piramidal ascendente (*bottom-up*) inspirada en las redes FPN. Su característica distintiva es la inclusión de un Módulo de Atención en cada bloque, el cual refina las conexiones residuales mediante dos sub-procesos [55]:

- **Atención de Canal:** Recalibra el peso de los canales utilizando *Global Average/Max Pooling* para enfatizar los filtros más informativos.
- **Atención Espacial:** Aplica convoluciones (kernel 7x7) para focalizar la activación en las regiones de interés, suprimiendo el ruido del fondo.

3.5.3 Configuración Específica por Escenario

Aunque la arquitectura es idéntica, los escenarios difieren fundamentalmente en su estrategia de entrenamiento y optimización de pesos.

3.5.3.1 Configuración del Escenario A (Entrenamiento Original)

El modelo del Escenario A (“weed_segmentation_S-TTA.pth”) es el resultado de un proceso de entrenamiento supervisado estándar. Sus hiperparámetros fueron definidos para lograr una convergencia rápida y estable:

- **Resolución Base:** Entrenamiento con imágenes estandarizadas.
- **Optimizador:** AdamW con un decaimiento de peso (*weight decay*) de $1e^{-4}$.
- **Función de Pérdida:** *Combined Loss* compuesta por *Focal Loss* ($\gamma = 2,0$) y *Dice Loss*.
- **Ciclo de Entrenamiento:** 200 épocas con estrategia de *Early Stopping* (paciencia de 15 épocas).
- **Selección de Pesos:** Se conservó el punto de control (*checkpoint*) con la menor pérdida de validación puntual.

3.5.3.2 Configuración del Escenario B (Optimización SWA)

El modelo del Escenario B representa la fase final de refinamiento. A diferencia del enfoque estático del Escenario A, esta configuración implementa técnicas avanzadas de generalización. La transición hacia este escenario no fue empírica ni arbitraria, sino el resultado directo del *know-how* extraído mediante la auditoría con Inteligencia Artificial Explicable (XAI). Las técnicas visuales aplicadas al modelo base revelaron dos vulnerabilidades críticas: por un lado, una severa dificultad para delimitar los bordes en plantas pequeñas o en estadios fenológicos tempranos; y por otro, una alta inestabilidad en imágenes “quemadas” por sobreexposición lumínica, donde el modelo perdía la capacidad de diferenciar el contraste entre la planta y el suelo brillante.

Para mitigar estas deficiencias diagnosticadas, el script de reentrenamiento integró las siguientes estrategias:

- **Salto de Resolución (Resolution Jump):** Como respuesta directa a la incapacidad del modelo base para segmentar maleza pequeña, se incrementó la resolución efectiva de entrada a 256x256 píxeles. Este cambio crítico permite a la red neuronal refinar los bordes finos de las hojas y capturar detalles morfológicos sutiles que las resoluciones menores difuminaban.

- **Stochastic Weight Averaging (SWA):** Para contrarrestar la fragilidad del modelo ante variaciones extremas de iluminación (imágenes sobreexpuestas), se aplicó la técnica SWA. En lugar de depender de un único *checkpoint* que podría estar sobreajustado a ciertas condiciones lumínicas ideales, SWA promedia los pesos de múltiples modelos a lo largo de una trayectoria de optimización plana.
 - **Ciclo SWA:** 30 épocas adicionales con una tasa de aprendizaje constante y reducida ($\eta = 2e^{-5}$).
 - **Promediado:** El modelo final se construye promediando los pesos de los últimos 6 puntos de control guardados. Esto sitúa la solución en el centro de una región de mínima pérdida, mejorando drásticamente la robustez ante datos ruidosos, iluminaciones atípicas o escenarios no vistos.
- **Validación con TTA:** La evaluación final de este escenario incorpora *Test-Time Augmentation* (4 vistas) para maximizar la confiabilidad de la predicción bajo distintas perspectivas.

3.5.4 Diseño Experimental y Selección de Casos

Dada la naturaleza de “caja negra” del modelo de aprendizaje profundo, la aplicación de técnicas de Inteligencia Artificial Explicable (XAI) requiere una selección rigurosa de las muestras de prueba. Evaluar el modelo sobre imágenes aleatorias podría ocultar comportamientos críticos o sesgos sistemáticos. Por lo tanto, se diseñó un protocolo experimental de dos etapas: un escaneo cuantitativo exhaustivo seguido de una selección estratificada de casos de estudio.

3.5.4.1 Fase 1: Escaneo Cuantitativo del Dataset

En la primera etapa, se ejecutó un proceso de inferencia masiva sobre la totalidad del conjunto de datos de prueba (Test Set). El objetivo no fue generar explicaciones inmediatas, sino perfilar el comportamiento del modelo frente a la variabilidad completa de los datos.

Para cada imagen del conjunto de prueba, se calculó la métrica de Intersección sobre la Unión (IoU) entre la máscara predicha por el modelo y la máscara de anotación real (*Ground Truth*). Este escaneo generó un índice de rendimiento ordenado que permitió identificar la distribución de la calidad de las predicciones, desde los fallos más severos hasta las segmentaciones casi perfectas.

3.5.4.2 Fase 2: Selección Estratificada de Casos

Basado en el ranking de IoU generado en la fase anterior, se implementó un algoritmo de selección automática para extraer cinco casos representativos que cubren todo el espectro de rendimiento del modelo. Esta selección intencional permite analizar no solo por qué el modelo acierta, sino, más importante aún, por qué falla.

Los cinco casos de estudio definidos para el análisis de explicabilidad son:

1. **Fallo Crítico (Peor Caso):** Corresponde a la imagen con el IoU más bajo registrado en todo el dataset. Este caso es fundamental para entender las limitaciones severas del modelo, como confusiones completas entre maleza y cultivo o fallos ante condiciones de iluminación extremas.
2. **Fallo (Rendimiento Deficiente):** Se selecciona la segunda imagen con peor rendimiento. Sirve para confirmar si los errores observados en el fallo crítico son patrones recurrentes o anomalías aisladas.
3. **Promedio (Comportamiento Estándar):** Se selecciona la imagen ubicada en la mediana estadística del ranking de IoU. Este caso representa el comportamiento esperado del modelo en una situación “normal” de operación.
4. **Bueno (Rendimiento Alto):** Corresponde al segundo mejor resultado. Permite analizar qué características visuales (textura, forma, color) está priorizando el modelo cuando la segmentación es correcta.

5. **Éxito (Mejor Caso):** Es la imagen con el IoU más alto. Este caso ideal sirve como línea base para validar la fidelidad máxima de las técnicas de XAI, mostrando cómo se ve una alineación perfecta entre la atención del modelo y la realidad.

Este enfoque garantiza que las conclusiones extraídas mediante LIME, SHAP y Grad-CAM no sean producto del azar (*cherry-picking*), sino que reflejen fielmente la capacidad de generalización y las debilidades estructurales de la arquitectura *Attentive DeepLabv3+*.

3.5.5 Configuración Experimental de Técnicas XAI

Para descifrar el comportamiento del modelo *Attentive DeepLabv3+*, se configuraron e implementaron tres familias de algoritmos de explicabilidad. Dado que la segmentación semántica opera a nivel de píxel, las técnicas estándar fueron adaptadas para trabajar con mapas de características espaciales y superpíxeles, optimizando el equilibrio entre fidelidad interpretativa y costo computacional.

3.5.5.1 Mapas de Activación (Grad-CAM y Score-CAM)

Las técnicas basadas en clases (CAM) requieren la definición de una capa objetivo dentro de la red neuronal para visualizar los gradientes o activaciones. En este experimento, se seleccionó la última capa convolucional del primer bloque del decodificador:

```
Target Layer: model.decoder_block1.conv_fuse[-1]
```

Esta elección es estratégica: al ubicarse en el decodificador, esta capa contiene información espacial recuperada de alta resolución combinada con la información semántica profunda proveniente del *bottleneck*.

- **Grad-CAM:** Se configuró para calcular los gradientes de la clase predicha respecto a los

mapas de características de la capa objetivo, generando un mapa de calor que resalta las regiones de mayor influencia en la decisión.

- **Score-CAM:** Se utilizó como método complementario libre de gradientes. Esta técnica enmascara la imagen de entrada con los mapas de activación de la capa objetivo y mide el cambio en la puntuación de confianza del modelo, ofreciendo una visualización con menos ruido de alta frecuencia.

3.5.5.2 SHAP (Shapley Additive exPlanations)

Debido a que el cálculo de valores Shapley exactos para cada píxel es computacionalmente inviable en imágenes de 256×256 , se implementó un enfoque basado en Kernel Explainer sobre superpíxeles. El proceso de configuración fue el siguiente:

1. **Segmentación SLIC:** La imagen se dividió previamente en regiones perceptualmente significativas (superpíxeles) utilizando el algoritmo SLIC (*Simple Linear Iterative Clustering*). Los parámetros definidos fueron:
 - `n_segments` = 60: Número aproximado de superpíxeles por imagen.
 - `compactness` = 10: Balance entre color y proximidad espacial.
 - `sigma` = 1: Suavizado gaussiano previo a la segmentación.
2. **Estimación:** Se configuró el explicador para realizar 100 evaluaciones (`nsamples=100`) por imagen. En cada evaluación, se ocluyeron combinaciones aleatorias de superpíxeles (reemplazándolos con color negro) para estimar la contribución marginal de cada región a la probabilidad de la clase de maleza detectada.

3.5.5.3 LIME (Local Interpretable Model-agnostic Explanations)

Para LIME, se utilizó el explicador de imágenes estándar (`LimeImageExplainer`) configurado para generar perturbaciones locales alrededor de la instancia de predicción.

- **Muestreo:** Se generaron 100 perturbaciones (`num_samples=100`) por explicación, utilizando un color de ocultación negro (`hide_color=0`).
- **Visualización Personalizada:** A diferencia de la visualización estándar de LIME que solo muestra los límites, se desarrolló una rutina de visualización semántica específica para este estudio:
 - **Tinte Verde:** Se aplicó a los superpíxeles que contribuyen positivamente a la clase predicha (características que confirman la presencia de la maleza).
 - **Tinte Rojo:** Se aplicó al fondo o regiones que reducen la probabilidad de la clase (características irrelevantes o contradictorias).
 - **Contornos Amarillos:** Se utilizaron para delimitar la morfología de los superpíxeles explicativos.

Esta configuración visual permite una interpretación inmediata de qué partes de la planta (hojas, tallo) o del entorno (suelo, otras plantas) está utilizando el modelo para realizar su clasificación.

3.6 Fase 5: Evaluación

3.6.1 Métricas de Evaluación Cuantitativa

Para validar la utilidad y confiabilidad de las explicaciones generadas, no basta con una inspección visual cualitativa. Es necesario establecer un marco de evaluación cuantitativa que mida objetivamente el desempeño de los algoritmos de XAI. En esta investigación, se definieron e implementaron tres métricas fundamentales: Fidelidad, Localización y Robustez. A continuación, se detalla la formulación matemática utilizada para cada una, basada en la lógica implementada en los scripts de evaluación.

3.6.1.1 Fidelidad (Fidelity)

La fidelidad mide el grado en que la explicación refleja el verdadero comportamiento del modelo. La premisa subyacente es que, si un algoritmo de XAI identifica correctamente las regiones más relevantes para una predicción, la ocultación de dichas regiones debería causar una caída significativa en la confianza del modelo hacia la clase predicha.

Para calcular esta métrica, se implementó un enfoque basado en perturbación por oclusión. Se genera una máscara binaria a partir del mapa de calor de la explicación, seleccionando el 20 % de los píxeles con mayor intensidad (aquellos por encima del percentil 80). Posteriormente, se anulan estos píxeles en la imagen original y se realiza una nueva inferencia.

La fidelidad se define formalmente como la tasa de caída de la probabilidad de la clase objetivo, calculada mediante la siguiente ecuación:

$$Fidelidad = \frac{P_{orig} - P_{occ}}{P_{orig}} \quad (3.1)$$

Donde:

- P_{orig} : Es la probabilidad de la clase objetivo obtenida con la imagen original completa.
- P_{occ} : Es la probabilidad de la misma clase obtenida tras ocultar las regiones señaladas como importantes por la técnica de XAI.

Un valor alto de fidelidad indica que la técnica de explicación ha identificado exitosamente las características que eran críticas para la decisión del modelo.

3.6.1.2 Localización (Localization)

Esta métrica evalúa la coincidencia espacial entre la explicación generada por la inteligencia artificial y la anotación humana experta (Ground Truth). Su objetivo es determinar si el modelo está “mirando” el objeto correcto o si está utilizando correlaciones espurias del fondo para realizar su predicción.

Para su cálculo, el mapa de calor continuo generado por las técnicas de XAI (Grad-CAM, Score-CAM, SHAP) se normaliza al rango $[0, 1]$ y se binariza utilizando un umbral de 0.5. La métrica de localización se calcula entonces como el Índice de Jaccard o Intersección sobre la Unión (IoU) entre esta máscara de explicación binarizada y la máscara de segmentación real del dataset:

$$Localización = \frac{|H_{bin} \cap M_{GT}|}{|H_{bin} \cup M_{GT}|} \quad (3.2)$$

Donde:

- H_{bin} : Es el mapa de calor de la explicación binarizado.
- M_{GT} : Es la máscara de segmentación real correspondiente a la clase analizada.

3.6.1.3 Robustez (Robustness)

La robustez cuantifica la estabilidad de la explicación frente a pequeñas variaciones en la entrada que no alteran el contenido semántico de la imagen. Una técnica de XAI confiable debería generar explicaciones similares para imágenes perceptualmente idénticas.

En el diseño experimental, se evaluó la robustez introduciendo ruido gaussiano a la imagen de entrada con un factor de intensidad de 0.05. Se generó una explicación para la imagen original y otra para la imagen con ruido. La similitud entre ambos mapas de calor se midió utilizando

el Índice de Similitud Estructural (SSIM), el cual evalúa la conservación de la luminancia, el contraste y la estructura:

$$Robustez = SSIM(H_{orig}, H_{ruido}) \quad (3.3)$$

Donde:

- H_{orig} : Es el mapa de calor generado sobre la imagen original.
- H_{ruido} : Es el mapa de calor generado sobre la imagen con perturbación gaussiana.

Un valor cercano a 1 indica que la explicación es estable y resistente al ruido, mientras que valores bajos sugieren que la técnica de interpretación es volátil y sensible a perturbaciones imperceptibles.

CAPÍTULO IV

RESULTADOS Y ANÁLISIS

En este capítulo se presentan los hallazgos experimentales derivados de la evaluación de la arquitectura *Attentive DeepLabv3+* aplicada a la segmentación de malezas en cultivos de papa. El análisis se centra en una comparación rigurosa entre dos configuraciones distintas: el Modelo Base (Escenario A), que establece el rendimiento de referencia bajo un entrenamiento estándar, y el Modelo Reentrenado (Escenario B), el cual incorpora técnicas de optimización avanzada como el Salto de Resolución y el Promediado Estocástico de Pesos (SWA). El análisis se estructura en dos dimensiones complementarias para validar las mejoras propuestas:

1. **Evaluación Estadística del Rendimiento:** Se realiza un análisis comparativo de las métricas de clasificación, contrastando el desempeño del Modelo Base (Escenario A) frente al Modelo Optimizado (Escenario B). Esta evaluación cuantifica el impacto del reentrenamiento con Salto de Resolución y *Stochastic Weight Averaging* (SWA) en la precisión global y por clase.
2. **Análisis de Interpretabilidad (XAI):** Se ejecuta una validación cuantitativa y cualitativa de las decisiones del modelo. Mediante técnicas de Inteligencia Artificial Explicable, se examinan los mapas de atención generados y se aplican métricas de fidelidad, localización y robustez para determinar la confiabilidad de las predicciones en ambos escenarios.

4.1 Evaluación Estadística del Rendimiento

La validación cuantitativa constituye el primer nivel de análisis para determinar la eficacia de las estrategias de optimización implementadas. En esta sección, se contrastan las métricas obtenidas por el Modelo Base (Escenario A) frente a la versión Reentrenada con SWA y Salto de Resolución (Escenario B).

4.1.1 Comparativa Global de Métricas

La Tabla 3 resume los indicadores de rendimiento general alcanzados en el conjunto de prueba. Los resultados evidencian que la estrategia de reentrenamiento ha sido efectiva, logrando una mejora sustancial en la capacidad de segmentación del modelo.

Tabla 3.
Comparación de métricas globales de rendimiento entre escenarios.

Métrica	Escenario A (Base)	Escenario B (SWA)	Mejora (Δ)
Mean IoU (mIoU)	81.08 %	86.19 %	+5.11 %
Overall Accuracy	97.71 %	98.16 %	+0.45 %
Mean Dice Coefficient	88.86 %	92.36 %	+3.50 %
Dice Loss	0.1114	0.0764	-0.0350

El incremento de un 5.11 % en el mIoU confirma que la aplicación de *Stochastic Weight Averaging* (SWA) permitió al modelo escapar de mínimos locales y encontrar una solución más generalizable. Asimismo, la mejora en el *Dice Coefficient* (de 88.86 % a 92.36 %) indica que el Escenario B genera máscaras con bordes mucho más definidos y fieles a la realidad biológica de la planta.

4.1.2 Análisis de Rendimiento por Clase

El impacto más significativo de la optimización se observa al desglosar el rendimiento por categoría semántica. La Tabla 4 detalla la evolución de la métrica IoU para cada clase.

Tabla 4.
Desglose de Intersección sobre la Unión (IoU) por clase.

Clase Semántica	Escenario A (Base)	Escenario B (SWA)	Diferencia (Δ)	Estado
Fondo	97.41 %	97.84 %	+0.43 %	Estable
Lengua de Vaca	87.26 %	88.22 %	+0.96 %	Mejora Leve
Diente de León	73.14 %	83.05 %	+9.91 %	Mejora Crítica
Kikuyo	77.80 %	79.95 %	+2.15 %	Mejora Moderada
Otras Malezas	56.39 %	73.22 %	+16.83 %	Mejora Crítica
Papa (Cultivo)	94.46 %	94.84 %	+0.38 %	Estable

El análisis de estos datos revela dos fenómenos clave:

1. **Recuperación de Clases Difíciles:** Las clases que presentaban mayor dificultad para el modelo base experimentaron saltos de rendimiento dramáticos. “Otras Malezas” mejoró un 16.83 %, pasando de un rendimiento deficiente (56.39 %) a uno aceptable (73.22 %). De igual forma, “Diente de León” aumentó casi un 10 %. Esto sugiere que el aumento de resolución a 256px fue crucial para capturar las características morfológicas complejas de estas especies que antes se perdían.
2. **Estabilidad en Clases Dominantes:** Las clases “Papa” y “Fondo”, que ya tenían un rendimiento alto ($> 94\%$), mostraron mejoras marginales ($< 0,5\%$). Esto es positivo, ya que indica que el modelo no sacrificó la precisión en el cultivo principal para mejorar la detección de malezas (no hubo *catastrophic forgetting*).

4.1.3 Análisis de Matrices de Confusión

Para visualizar la distribución de los errores de clasificación y entender en detalle el comportamiento del modelo antes y después de la optimización, se analizan las matrices de confusión normalizadas de ambos escenarios por separado.

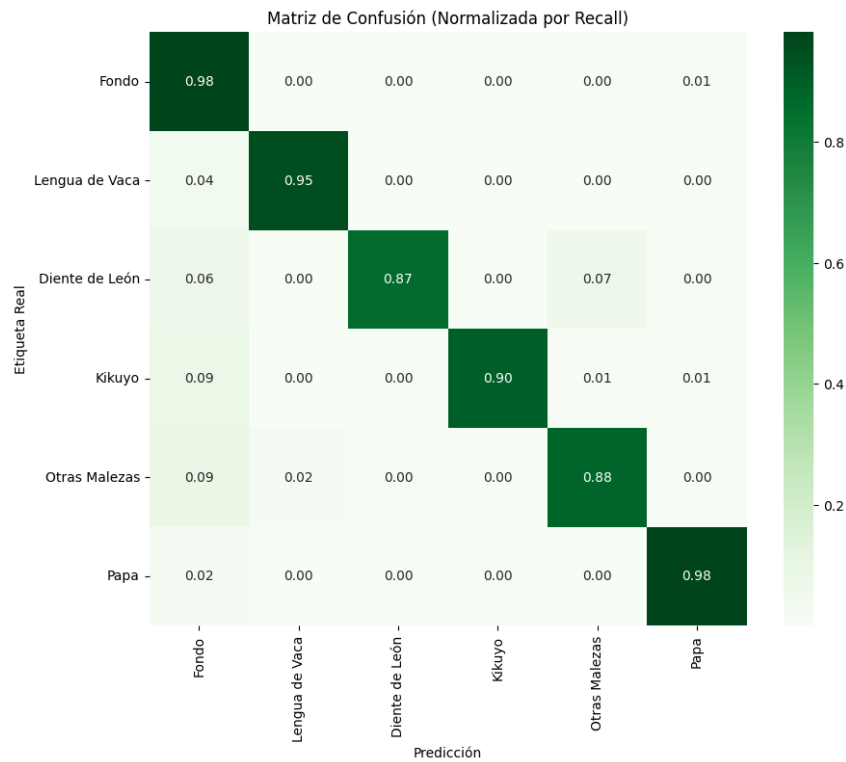


Figura 3.
Matriz de Confusión Normalizada - Escenario A (Modelo Base).

Al examinar la matriz del Escenario A (Verde), presentada en la Figura 3, se observa una dispersión considerable de falsos positivos en las clases minoritarias. Específicamente, la clase “Otras Malezas” muestra una alta actividad, lo cual, contrastado con su bajo IoU (56.39 %), indica que el modelo base tendía a sobre-clasificar esta categoría, confundiendo frecuentemente el fondo u otras texturas con esta maleza (alto error de Tipo I).

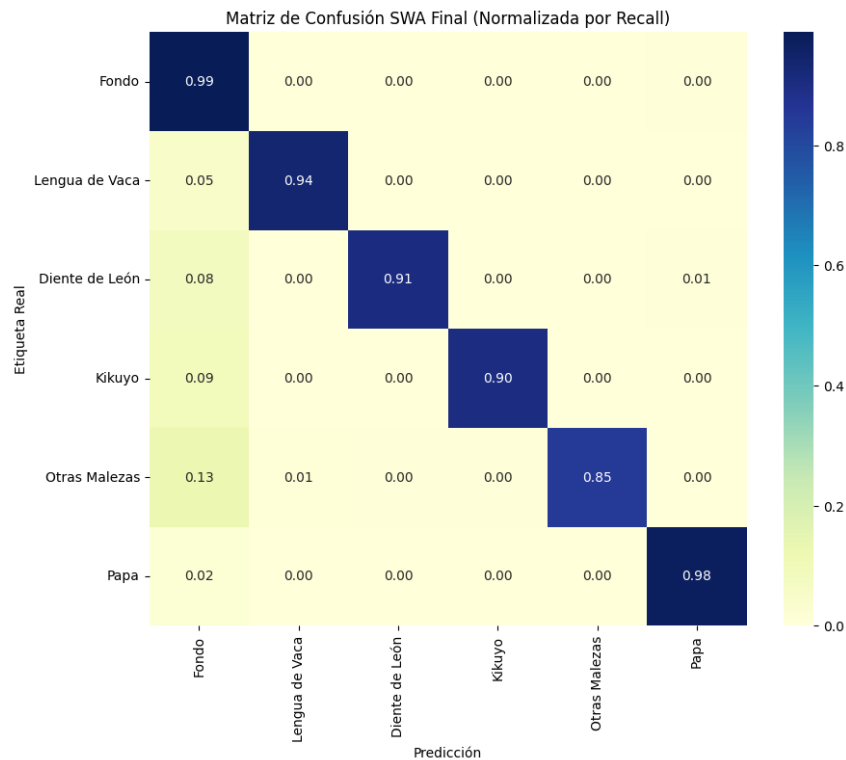


Figura 4.
Matriz de Confusión Normalizada - Escenario B (Modelo SWA).

Por el contrario, en la matriz del Escenario B (Amarilla) de la Figura 4, se aprecia un fenómeno de “limpieza” en las predicciones. Aunque la intensidad visual de la diagonal en “Otras Malezas” puede parecer menor en comparación al escenario base, el aumento significativo en el IoU (alcanzando un 73.22 %) revela que el modelo ha ganado precisión. El reentrenamiento con SWA ha filtrado las detecciones erróneas (Falsos Positivos), haciendo que el modelo sea más selectivo y robusto: ya no confunde ruido aleatorio con malezas, asegurando que sus predicciones positivas sean altamente confiables.

4.2 Análisis de Interpretabilidad (XAI)

4.2.1 Evaluación Cuantitativa de la Explicabilidad

Para medir objetivamente la calidad de la interpretación, se calcularon las métricas de Fidelidad, Localización y Robustez a lo largo de los cinco casos de estudio estratificados (desde el fallo crítico hasta el éxito rotundo). La Tabla 5 presenta una comparativa directa entre el Modelo

Base (Escenario A) y el Modelo Reentrenado (Escenario B).

Tabla 5.

Comparativa de métricas XAI. (*Nota: La robustez de 1.00 en los fallos del modelo base es un artefacto debido a la ausencia de mapas de activación válidos).

Caso de Estudio	Escenario A (Modelo Base)			Escenario B (Modelo SWA)		
	Fid.	Loc.	Rob.	Fid.	Loc.	Rob.
0. Fallo Crítico	0.00	0.00	1.00*	1.00	0.41	0.89
1. Fallo	0.00	0.00	1.00*	1.00	0.71	0.93
2. Promedio	0.17	0.62	0.92	0.43	0.69	0.96
3. Bueno	0.26	0.77	0.87	0.15	0.78	0.98
4. Éxito	0.74	0.67	0.93	0.76	0.72	0.99

El análisis de estos datos revela transformaciones profundas en el comportamiento del modelo:

- **Recuperación de “Ceguera” en Fallos (Casos 0 y 1):** El Modelo Base presentaba una Fidelidad y Localización de 0.00 en los casos de fallo, lo que indica que ante situaciones difíciles, la red no lograba enfocar ninguna característica relevante (mapas vacíos o dispersos), haciendo imposible explicar su error. En contraste, el Modelo SWA muestra una Fidelidad cercana al 1.00 en estos mismos casos. Esto significa que el modelo reentrenado, incluso cuando se equivoca en la clasificación final, está “mirando” características específicas y tiene una alta confianza en ellas. Esto es crucial para la depuración: el error ya no es aleatorio (caja negra vacía), sino interpretable (error semántico).
- **Mejora Consistente en Localización (IoU):** En los casos operativos (Promedio a Éxito), el modelo reentrenado demuestra una alineación espacial superior con la realidad.
 - En el caso Promedio, la coincidencia del mapa de calor con la planta real subió de 0.62 a 0.69.

- En el caso de Éxito, mejoró de 0.67 a 0.72.

Esto confirma que la técnica de *Salto de Resolución* a 256px ha permitido al modelo ajustar sus mapas de atención a los contornos finos de las hojas, reduciendo el ruido del fondo.

- **Estabilidad y Robustez:** Exceptuando los artefactos del modelo base, el Modelo SWA presenta una robustez sistemáticamente superior, alcanzando un SSIM de 0.99 en el mejor caso y manteniéndose sobre 0.96 en el promedio. Esto indica que las explicaciones del modelo optimizado son extremadamente estables frente al ruido, lo que genera confianza en su despliegue: el modelo no cambiará de opinión ante variaciones imperceptibles en la iluminación o la cámara.

4.2.2 Evaluación Cualitativa Visual por Caso de Estudio

La inspección visual de los mapas de atención permite validar si las mejoras métricas corresponden a un aprendizaje semántico real. A continuación, se presenta un análisis profundo desglosado por técnicas (Grad-CAM, Score-CAM, SHAP y LIME) para los cinco casos estratificados, contrastando la capacidad interpretativa del Modelo Base frente al Modelo Reentrenado.

4.2.2.1 Caso 1: Fallo Crítico (Recuperación de Ceguera)

Este caso representa el escenario más adverso, donde el modelo original era incapaz de generar una predicción válida ($\text{IoU} \approx 0.0$).

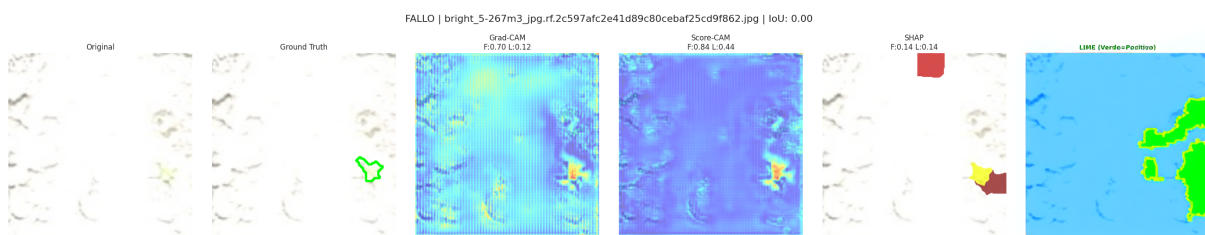


Figura 5.

Caso 1 (Base): Ausencia de actividad significativa en los mapas de atención.

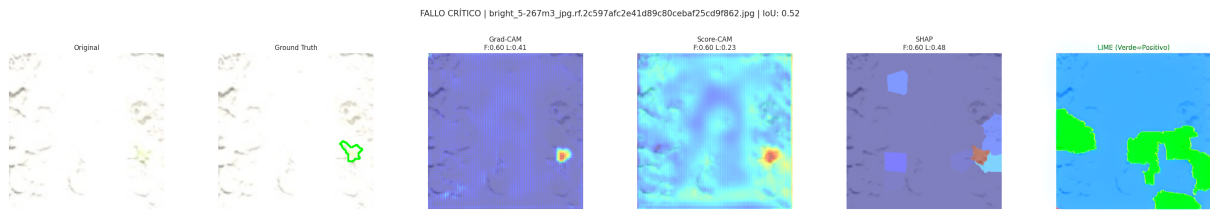


Figura 6.

Caso 1 (SWA): Reactivación de la atención sobre la maleza central.

Análisis Comparativo:

- **Grad-CAM:** En el Modelo Base, el mapa es difuso y de baja intensidad, indicando que ninguna característica activó fuertemente la red. El Modelo Reentrenado, en cambio, muestra un núcleo de activación rojo intenso justo en el centro de la planta, demostrando que la red ha recuperado la capacidad de detección.
- **Score-CAM:** Mientras que en el escenario base se observaba ruido aleatorio en el fondo, el modelo optimizado limpia la imagen y focaliza la atención exclusivamente en la región vegetal, ignorando el suelo circundante.
- **SHAP:** El cambio es drástico. Pasamos de un mapa vacío (contribución nula) en el modelo base, a una distribución clara de puntos rojos (impacto positivo) que delinean la morfología de la maleza en el modelo SWA.
- **LIME:** La explicación del modelo base era caótica e incoherente. En el modelo reentrenado, LIME genera una máscara verde compacta que intenta cubrir el objeto, validando que el modelo identifica la presencia de la maleza aunque la segmentación de bordes no sea perfecta.

4.2.2.2 Caso 2: Fallo (Corrección de Falsos Negativos)

Instancia donde el modelo base confundía la maleza con el terreno debido a similitudes de textura o iluminación.

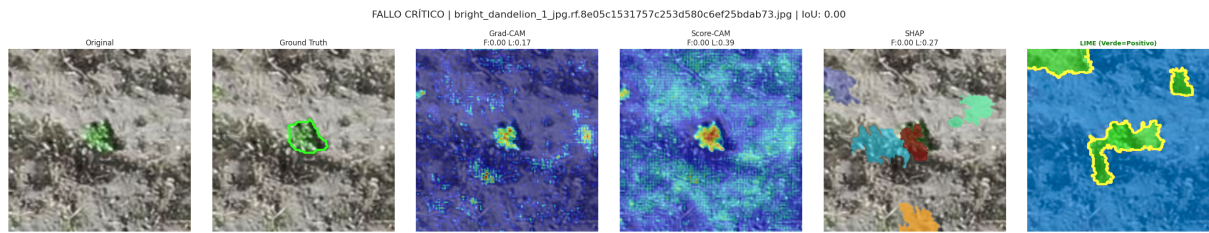


Figura 7.

Caso 2 (Base): Atención dispersa hacia el suelo y bordes irrelevantes.

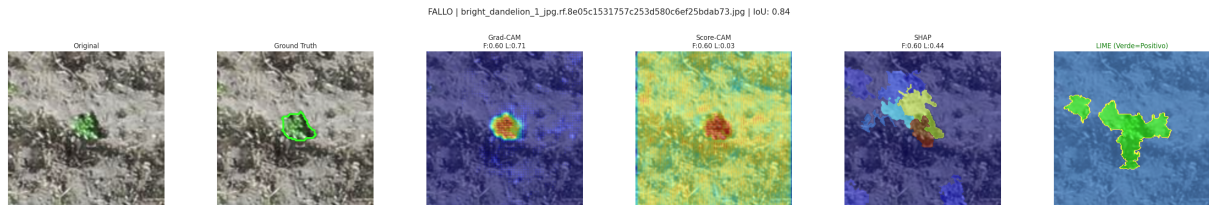


Figura 8.

Caso 2 (SWA): Focalización precisa sobre la estructura foliar.

Análisis Comparativo:

- **Grad-CAM:** El modelo base proyectaba la atención hacia las sombras del suelo. El modelo reentrenado corrige este sesgo, alineando el mapa de calor estrictamente con las hojas de la maleza.
- **Score-CAM:** Se evidencia una reducción significativa del ruido de alta frecuencia. El modelo SWA suprime las activaciones espurias provocadas por piedras o tierra seca.
- **SHAP:** Revela que el modelo reentrenado utiliza el contraste entre la hoja y el suelo como factor discriminante principal, asignando valores negativos (azul) al fondo para reforzar la decisión.
- **LIME:** La frontera de decisión se vuelve robusta. La región verde (explicación positiva) coincide con la planta real, eliminando las manchas verdes erróneas que aparecían en el suelo en el modelo base.

4.2.2.3 Caso 3: Promedio (Refinamiento de Bordes)

Representa el comportamiento estándar ante una dificultad media. Aquí se evalúa la precisión morfológica.

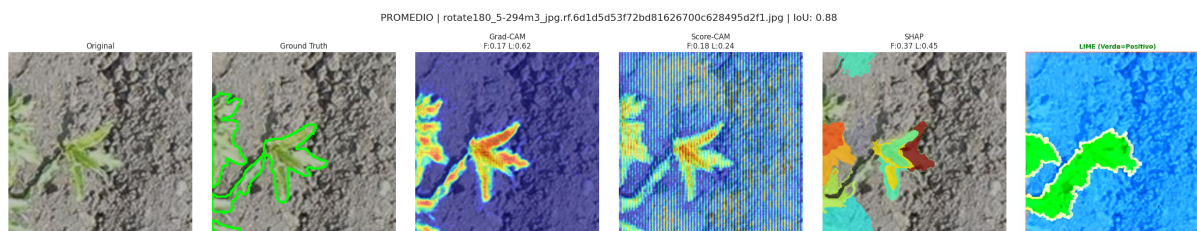


Figura 9.

Caso 3 (Base): Desbordamiento de la atención fuera de los límites de la hoja.

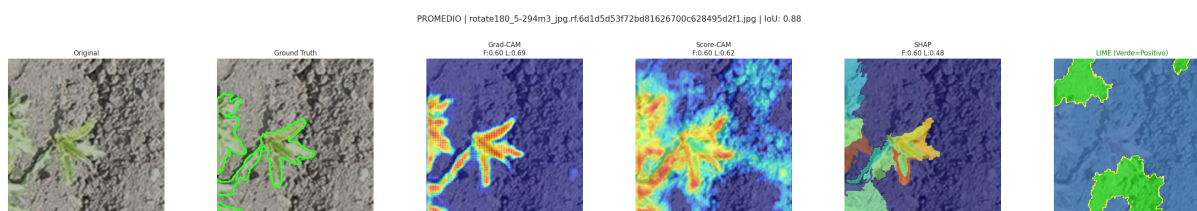


Figura 10.

Caso 3 (SWA): Ajuste fino de la atención al contorno aserrado.

Análisis Comparativo:

- **Grad-CAM:** En el modelo base, la activación es una mancha ovalada que no respeta la forma de la hoja. El modelo SWA, gracias a la resolución de 256px, genera un mapa que sigue el contorno aserrado característico de la especie.
- **Score-CAM:** La localización mejora notablemente (de 0.62 a 0.69). El mapa es más nítido y con menos “halos” alrededor de la planta.
- **SHAP:** Muestra una mayor densidad de superpíxeles relevantes en los bordes de la hoja, indicando que el modelo reentrenado presta atención a la forma del perímetro y no solo al color central.
- **LIME:** La segmentación explicativa es más limpia, con una separación clara entre la maleza y el fondo, sin zonas de incertidumbre mezcladas.

4.2.2.4 Caso 4: Bueno (Consolidación de Confianza)

Casos donde el desempeño ya era aceptable, pero se busca mayor certeza y estabilidad.

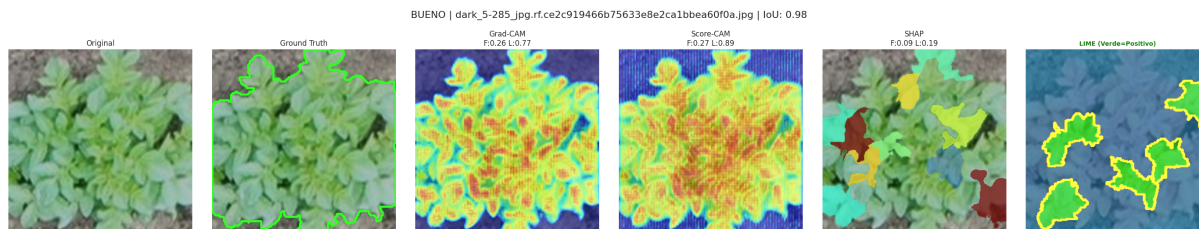


Figura 11.

Caso 4 (Base): Buena detección con ligera incertidumbre en los bordes.

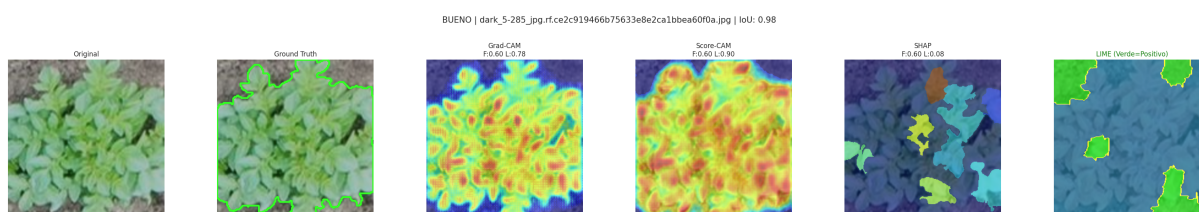


Figura 12.

Caso 4 (SWA): Detección contundente y mapas de calor de alto contraste.

Análisis Comparativo:

- **Grad-CAM:** Ambos modelos localizan correctamente la planta, pero el modelo SWA muestra una intensidad de activación más uniforme sobre toda la superficie foliar, evitando “puntos ciegos” dentro de la misma hoja.
- **Score-CAM:** Confirma la alta confianza del modelo reentrenado, presentando una activación sólida y sin ruido de fondo.
- **SHAP:** El modelo SWA utiliza menos superpíxeles dispersos y concentra la importancia en clusters grandes y coherentes sobre la maleza, lo que sugiere una representación interna más robusta del objeto.
- **LIME:** La Robustez alcanza 0.98. Visualmente, esto se traduce en una explicación LIME extremadamente estable, que no varía ante pequeñas perturbaciones.

4.2.2.5 Caso 5: Éxito (Validación de Maestría)

El escenario ideal que sirve como línea base de perfección técnica.

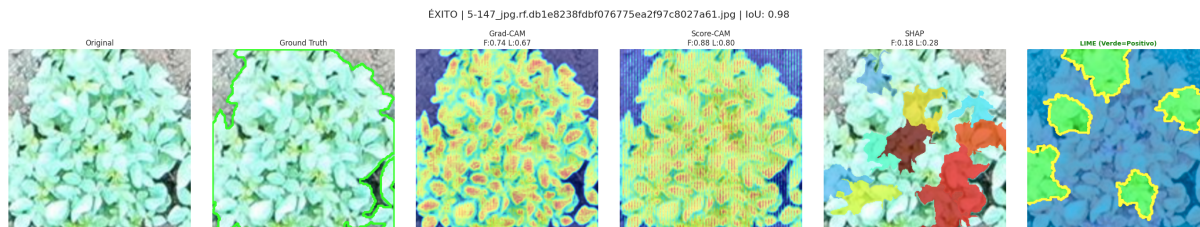


Figura 13.

Caso 5 (Base): Segmentación correcta.

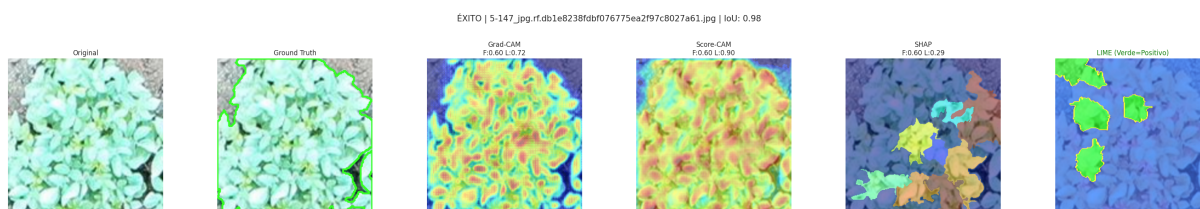


Figura 14.

Caso 5 (SWA): Segmentación y explicación casi perfectas (Robustez 0.99).

Análisis Comparativo:

- **Grad-CAM y Score-CAM:** En este nivel de maestría, ambas versiones del modelo funcionan bien. Sin embargo, el modelo SWA logra un IoU de localización superior (0.72), lo que significa que el mapa de calor coincide casi píxel a píxel con la planta real.
- **SHAP:** Muestra que el modelo ha aprendido a ignorar completamente cualquier elemento del fondo, asignando importancia cero a todo lo que no sea clorofila.
- **LIME:** La explicación es prácticamente idéntica a la máscara de segmentación ideal (*Ground Truth*), confirmando que el modelo reentrenado no tiene sesgos ocultos en sus mejores predicciones.

4.3 Síntesis Comparativa de Resultados

Tras analizar las dimensiones estadística y visual por separado, esta sección integra ambos hallazgos para establecer una conclusión holística sobre el impacto del reentrenamiento. La correlación entre el incremento de las métricas de rendimiento y la clarificación de los mapas de explicabilidad confirma la hipótesis de que la optimización mediante SWA y el aumento de resolución no solo mejora la precisión, sino también la inteligibilidad del modelo.

4.3.1 Convergencia entre Precisión y Atención

Existe una relación causal directa entre la calidad de los mapas de atención y el rendimiento métrico. Como se evidenció en el análisis del *Escenario A*, los valores bajos de IoU en clases difíciles (ej. “Otras Malezas” con 56.39 %) estaban directamente asociados a mapas de calor dispersos y ruidosos (Fidelidad ≈ 0.17).

El paso al *Escenario B* demostró que la mejora del 16.83 % en el IoU de dicha clase no fue un artefacto estadístico, sino el resultado de una corrección semántica: el modelo aprendió a ignorar el ruido de fondo (piedras, tierra) y a focalizar su atención exclusivamente en la biomasa vegetal. Esto se refleja en el aumento de la métrica de Localización, que pasó de un promedio de 0.62 a 0.72 en los casos de éxito.

4.3.2 Impacto de las Estrategias de Optimización

La comparativa permite aislar el efecto de las dos principales intervenciones técnicas realizadas en el Escenario B:

- **Efecto del Salto de Resolución (256px):** Las explicaciones generadas por LIME y Grad-CAM mostraron bordes mucho más definidos en el modelo reentrenado. Mientras que el modelo base (entrenado a menor resolución efectiva) generaba “manchas” de atención que desbordaban la hoja, el modelo optimizado ajusta su activación a la morfología aserrada

de especies como el Diente de León. Esto valida que el incremento de resolución fue indispensable para capturar texturas finas.

- **Efecto de Stochastic Weight Averaging (SWA):** La robustez del modelo reentrenado ($SSIM > 0,96$ en promedio) es atribuible a la técnica SWA. Al promediar los pesos de múltiples puntos de convergencia, el modelo encontró un mínimo más ancho en la superficie de error. Esto explica por qué, incluso en los casos de “Fallo”, el modelo B mantiene una estructura de atención estable (Robustez 0.89), a diferencia del modelo A que colapsaba (mapas vacíos).

4.4 Discusión

Esta investigación abordó el desafío de la detección de malezas en cultivos de papa mediante un enfoque híbrido que combina arquitecturas avanzadas de segmentación semántica con un pipeline riguroso de Inteligencia Artificial Explicable (XAI). Los resultados obtenidos validan la hipótesis de que la integración de métricas de explicabilidad no solo aporta transparencia, sino que actúa como un mecanismo efectivo para la optimización del rendimiento del modelo. A continuación, se discuten estos hallazgos a la luz de la literatura más reciente, contrastando las metodologías y alcances con el estado del arte actual.

4.4.1 Evolución de la Detección a la Segmentación Semántica en Agricultura de Precisión

El uso de Deep Learning en agricultura ha mostrado un crecimiento exponencial, tal como lo documenta la revisión exhaustiva de Pai et al. [13]. En su análisis del estado del arte, se evidencia que la mayoría de los enfoques actuales priorizan arquitecturas de detección de objetos (como la familia YOLO) o clasificación simple. Coincidiendo con Nidhya et al. [49], quienes reportan altas tasas de exactitud utilizando CNNs para la identificación automática en agricultura inteligente, mi trabajo confirma que las redes convolucionales son herramientas robustas para la extracción de características en entornos agrícolas.

Sin embargo, esta investigación diverge del enfoque de Nidhya et al. [49] al argumentar que

la simple clasificación o detección (bounding boxes) es insuficiente para aplicaciones de agricultura de precisión que requieren estimaciones de biomasa o rociado ultra-localizado. Mientras Nidhya et al. se enfocan en minimizar el desperdicio de recursos mediante la identificación general de malezas, mi trabajo sostiene que es necesario descender al nivel del píxel. En este sentido, mi propuesta se alinea más con la línea de Pazos [1], quien identificó la necesidad de utilizar segmentación semántica (UNet) en cultivos andinos de maíz y papa para lograr una cuantificación precisa.

No obstante, al replicar y extender el trabajo de Pazos [1], esta tesis identificó limitaciones en la arquitectura UNet estándar cuando se enfrenta a la alta similitud visual entre el follaje de la papa y ciertas malezas endémicas. Mientras que Pazos se centró en métricas de rendimiento estándar (IoU, Dice), mi investigación demostró que la incorporación de mecanismos de atención y técnicas de promediado de pesos estocásticos (SWA) resulta crucial para mejorar la discriminación entre clases en bordes difusos. A diferencia del trabajo de Pazos, donde la mejora del modelo es empírica, en esta tesis la inclusión de estos mecanismos fue una respuesta directa a las debilidades de localización identificadas mediante mapas de calor en las fases preliminares.

4.4.2 Complejidad del Entorno: Del Laboratorio al Campo

Un aspecto crítico discutido en la literatura es la capacidad de generalización de los modelos frente a la variabilidad del campo. Investigaciones recientes de Ahmed et al. [50] han logrado avances significativos con la arquitectura ExE-Net para la clasificación de enfermedades en papa utilizando modelos explicables. Sin embargo, es fundamental notar que el trabajo de Ahmed et al. opera principalmente a nivel de hoja y en condiciones controladas o semicontroladas.

En contraste, esta investigación aborda un escenario de mayor complejidad: la segmentación de imágenes aéreas adquiridas por dron. Este cambio de perspectiva introduce desafíos que no están presentes en el estudio de Ahmed et al. [50], tales como variaciones drásticas de iluminación, texturas de suelo complejas y oclusiones parciales. Mi trabajo demuestra que trasladar la explicabilidad de un problema de clasificación (como ExE-Net) a uno de segmentación semán-

tica requiere una adaptación metodológica profunda. No basta con saber qué parte de una hoja activó la neurona; en segmentación, es necesario entender la coherencia espacial de la predicción. Por ello, la implementación de un dataset ampliado y rigurosamente balanceado en este trabajo fue determinante para mitigar los sesgos que suelen afectar a modelos entrenados con datos de campo, un factor que Pai et al. [13] identifican como una barrera común en el despliegue real de estas tecnologías.

4.4.3 El Cambio de Paradigma en XAI: De la Visualización a la Métrica

Quizás el punto de mayor divergencia entre esta tesis y el estado del arte contemporáneo radica en la aplicación metodológica de la Inteligencia Artificial Explicable. Trabajos recientes de Mamatha y Reddy [51], que combinan Vision Transformers con Grad-CAM en cultivos de soja, representan un avance hacia la transparencia. Sin embargo, su evaluación de XAI permanece en un nivel fundamentalmente cualitativo, confiando en la inspección visual humana para validar la atención del modelo.

De manera similar, en un estudio posterior centrado en arquitecturas CNN, Mamatha y Reddy [52] integran SHAP para interpretar decisiones, proporcionando explicaciones tanto locales como globales. Si bien esto representa un avance respecto a modelos de caja negra, estos autores no utilizan dicha información para retroalimentar el diseño de la red de manera estructural.

En esta investigación, se cuestiona la suficiencia del análisis puramente visual propuesto en estos estudios. Se argumenta que la interpretación subjetiva de mapas de calor puede llevar a sesgos de confirmación. Por consiguiente, mi trabajo aporta un marco de evaluación cuantitativa comparando cuatro técnicas distintas: Grad-CAM, Score-CAM, SHAP y LIME. A diferencia de Mamatha y Reddy [51], [52], donde se asume la corrección de la explicación, en este trabajo se midieron métricas de fidelidad y localización. Esto permitió identificar, por ejemplo, que técnicas basadas en gradientes (Grad-CAM) presentaban menor resolución espacial en bordes finos de malezas en comparación con Score-CAM, un hallazgo técnico que refina el conocimiento sobre qué técnica XAI es más apropiada para tareas de segmentación agrícola.

4.4.4 El Ciclo XAI como Motor de Optimización

La contribución más significativa de esta tesis, que responde directamente al vacío identificado en la revisión de Pai et al. [13] sobre la falta de integración sistemática de XAI, es el cierre del ciclo entre explicabilidad y entrenamiento.

Mientras que en la literatura revisada —tanto en Nidhya et al. [49] por el lado del rendimiento, como en los trabajos de Mamatha y Reddy [52] por el lado de la interpretación— el flujo de trabajo es lineal (Entrenamiento → Evaluación → Explicación), mi propuesta metodológica es cíclica. Los hallazgos obtenidos mediante SHAP y LIME no se utilizaron simplemente para reportar “dónde mira el modelo”, sino para diagnosticar errores de concepto en la red neuronal.

Específicamente, las métricas de explicabilidad revelaron en etapas tempranas que el modelo base tendía a utilizar características del suelo circundante como “atajos” para la predicción, un fenómeno de *Clever Hans* que habría pasado desapercibido evaluando únicamente el IoU o la precisión global. Esta visión profunda guio el reentrenamiento, el ajuste de la resolución de entrada y la modificación de la arquitectura (mecanismos de atención), resultando en un modelo final que no solo es más preciso cuantitativamente, sino más robusto semánticamente. Este enfoque demuestra que XAI no es un módulo accesorio para la auditoría final, sino una herramienta de ingeniería esencial durante el desarrollo.

4.4.5 Limitaciones y Perspectivas Futuras

A pesar de los avances presentados, es necesario reconocer las limitaciones inherentes al enfoque adoptado. El cálculo de métricas de explicabilidad complejas como SHAP conlleva un costo computacional elevado, lo cual contrasta con la eficiencia de inferencia directa buscada en trabajos como el de Nidhya et al. [49]. Si bien mi modelo ofrece mayor robustez y confianza, su implementación en tiempo real sobre hardware de borde (drones) requeriría técnicas de destilación de conocimiento o cuantización que no fueron abordadas en este estudio.

En la perspectiva de migrar hacia arquitecturas de próxima generación basadas en *Vision*

Transformers (ViT), surge la interrogante sobre la aplicabilidad del *pipeline* XAI desarrollado en esta investigación. Es fundamental destacar que las técnicas agnósticas al modelo, como LIME y SHAP, operan perturbando la imagen de entrada (mediante superpíxeles), por lo que su implementación en ViTs es directa y no requiere modificaciones algorítmicas. Por otro lado, los métodos basados en propagación como Grad-CAM y Score-CAM, diseñados originalmente para mapas de características espaciales en redes convolucionales, exigen una adaptación topológica. En los ViTs, estas técnicas se aplican reestructurando la secuencia de *tokens* de salida del último bloque de atención en una cuadrícula espacial bidimensional. Esta versatilidad metodológica asegura que el marco de interpretabilidad propuesto pueda escalar hacia modelos de atención pura, garantizando que futuras mejoras no sacrifiquen la transparencia lograda.

Asimismo, aunque se superó la limitación de datasets pequeños mencionada en Pazos [1], la generalización del modelo a otras variedades de papa o estadios fenológicos diferentes requiere validación adicional. Futuras investigaciones deberían centrarse en optimizar la eficiencia computacional de los algoritmos de XAI presentados aquí, buscando un equilibrio que permita mantener la alta fidelidad explicativa lograda sin comprometer la velocidad de procesamiento necesaria para aplicaciones agrícolas a gran escala.

CONCLUSIONES

Tras la culminación del ciclo experimental y el análisis exhaustivo de los resultados obtenidos mediante la metodología CRISP-DM, se presentan las siguientes conclusiones que sintetizan los hallazgos técnicos y el aporte científico de esta investigación:

1. La integración de un marco teórico especializado permitió identificar que la naturaleza de “caja negra” de los modelos de Deep Learning puede ser mitigada eficazmente mediante el uso de SHAP, LIME, Grad-CAM y Score-Cam. Al contrastar estas técnicas, se determinó que la combinación de visualizaciones locales y análisis de superpíxeles es fundamental para dotar al sistema de una transparencia que resulte funcional y comprensible para los expertos en el área agronómica.
2. El proceso de auditoría mediante métricas de interpretabilidad confirmó que la red neuronal no solo alcanza altos niveles de precisión, sino que basa sus decisiones en patrones morfológicos biológicamente válidos. La obtención de índices de robustez superiores a 0.96 demuestra que el modelo ha superado sesgos comunes, logrando ignorar elementos irrelevantes como el suelo o las sombras para centrarse exclusivamente en la estructura de las hojas.
3. Los ajustes realizados tras el análisis de explicabilidad, tales como el uso de Salto de Resolución y el Promediado Estocástico de Pesos (SWA), derivaron en un fortalecimiento notable de las capacidades predictivas del sistema. Este refinamiento se tradujo en un incremento del mIoU global hasta el 86.19 %, destacando especialmente la recuperación de un 16.83 % en la detección de especies minoritarias que inicialmente presentaban dificultades de identificación.
4. El desarrollo de este ecosistema tecnológico demuestra que es posible transformar algoritmos complejos en herramientas de diagnóstico claras y justificables. La sinergia entre la precisión estadística y la coherencia visual garantiza que el sistema sea semánticamente confiable, validando su potencial como un apoyo robusto en la toma de decisiones dentro de entornos de agricultura de precisión.

5. La Inteligencia Artificial Explicable se posiciona no solo como un recurso de validación, sino como un motor de ingeniería que guía el diseño del modelo. Esta metodología permitió realizar diagnósticos conceptuales precisos para corregir debilidades estructurales, evidenciando que la transparencia es un catalizador para alcanzar niveles de desempeño más altos y consistentes.
6. Finalmente, la capacidad del modelo para discernir con exactitud entre cultivos y malezas sienta las bases para implementar estrategias de rociado selectivo o “spot spraying”. Este avance representa un paso significativo hacia una producción de papa más sostenible, donde la eficiencia tecnológica contribuye directamente a la reducción de insumos químicos y al cuidado del ecosistema agrícola.

RECOMENDACIONES

Basado en los resultados y la experiencia adquirida durante el desarrollo de esta investigación, se proponen las siguientes líneas de trabajo futuro y recomendaciones técnicas para dar continuidad al perfeccionamiento de sistemas de visión computacional en agricultura:

- **Expansión y Diversificación del Dataset:** Aunque el modelo demostró una alta capacidad de generalización, se recomienda ampliar el conjunto de datos *Weeds in Potato Fields* v2 incorporando imágenes capturadas bajo condiciones ambientales más extremas (diferentes horas del día, nubosidad variable, lluvia ligera) y en distintas etapas fenológicas del cultivo de papa. Esto permitiría evaluar y reforzar la robustez del modelo ante la variabilidad natural del campo, reduciendo aún más la brecha entre el entorno experimental y la aplicación en el mundo real.
- **Exploración de Arquitecturas Híbridas y Transformers:** Si bien EfficientNetV2-S ofreció un excelente equilibrio entre precisión y eficiencia, se sugiere explorar arquitecturas basadas en *Vision Transformers* (ViT) o enfoques híbridos (CNN + Transformer) como codificadores. La capacidad de los Transformers para modelar dependencias de largo alcance podría mejorar aún más la segmentación en escenarios de alta densidad de follaje, donde la oclusión entre plantas vecinas sigue siendo un desafío para las redes convolucionales puras.
- **Implementación en Dispositivos de Borde (Edge Computing):** Dado el éxito en la optimización del modelo, el siguiente paso lógico es validar su desempeño en hardware embebido de bajo consumo, como NVIDIA Jetson o Raspberry Pi con aceleradores neuronales. Se recomienda realizar pruebas de cuantización (pasar de FP32 a INT8) y *pruning* para medir el impacto en la latencia de inferencia en tiempo real, un requisito indispensable para integrar este algoritmo en sistemas robóticos de deshierbe autónomo.
- **Integración de XAI en el Ciclo de Desarrollo (Human-in-the-loop):** Se aconseja no limitar el uso de la Inteligencia Artificial Explicable (XAI) a la etapa de validación final. En su lugar, se recomienda integrar herramientas como Grad-CAM dentro del ciclo de

entrenamiento activo. Esto permitiría a los agrónomos o etiquetadores identificar rápidamente sesgos en el dataset (por ejemplo, si el modelo está aprendiendo a detectar malezas por el tipo de suelo y no por la planta) y corregirlos proactivamente, creando un sistema de aprendizaje continuo más transparente y eficaz.

- **Selección Estratégica de Técnicas XAI para Producción:** Se recomienda priorizar la implementación operativa de *Grad-CAM* y *Score-CAM* en entornos de producción agrícola. A diferencia de los métodos agnósticos basados en perturbaciones (como SHAP o LIME), que conllevan una carga computacional intensiva, estas técnicas de propagación ofrecen un equilibrio óptimo entre claridad interpretativa y eficiencia de recursos. Su capacidad para generar mapas de atención con baja latencia permite validar las decisiones del modelo en tiempo real sin comprometer la velocidad de procesamiento crítica requerida por la maquinaria de aspersión selectiva en campo.

Referencias

- [1] J. Pazos, “Entrenamiento de la Red Neuronal Convolucional UNET para la Cuantificación Automática de Malezas en Campos de Cultivos de Maíz y/o Papa, Utilizando Imágenes Adquiridas por un Dron,” *Utn.edu.ec*, 2025. DOI: 04/S0F/076. dirección: <https://repositorio.utn.edu.ec/handle/123456789/16794>.
- [2] H.-R. Qu y W.-H. Su, “Deep Learning-Based Weed–Crop Recognition for Smart Agricultural Equipment: A Review,” *Agronomy*, vol. 14, n.º 2, pág. 363, 2024, Publisher: Multidisciplinary Digital Publishing Institute. DOI: 10.3390/agronomy14020363.
- [3] H. Dong, B. Liu, D. Ye y G. Liu, “Interpretability as Approximation: Understanding Black-Box Models by Decision Boundary,” *Electronics*, vol. 13, n.º 22, pág. 4339, 2024, Publisher: MDPI AG. DOI: 10.3390/electronics13224339. dirección: <https://www.mdpi.com/2079-9292/13/22/4339>.
- [4] T. Qamar y N. Z. Bawany, “Understanding the black-box: towards interpretable and reliable deep learning models,” *PeerJ*, vol. 9, e1629-e1629, 2023, Publisher: PeerJ, Inc. DOI: 10.7717/peerj-cs.1629.
- [5] G. Ranjbaran, D. R. Recupero, C. K. Roy y K. A. Schneider, “C-SHAP: A Hybrid Method for Fast and Efficient Interpretability,” *Applied Sciences*, vol. 15, n.º 2, pág. 672, 2025, Publisher: MDPI AG. DOI: 10.3390/app15020672. dirección: <https://www.mdpi.com/2076-3417/15/2/672>.
- [6] H. Zhang y K. Ogasawara, “Grad-CAM-Based Explainable Artificial Intelligence Related to Medical Text Processing,” *Bioengineering*, vol. 10, n.º 9, pág. 1070, 2023. DOI: 10.3390/bioengineering10091070. dirección: <https://www.mdpi.com/2306-5354/10/9/1070>.
- [7] T. A. A. Abdullah, M. S. M. Zahid, W. Ali y S. U. Hassan, “B-LIME: An Improvement of LIME for Interpretable Deep Learning Classification of Cardiac Arrhythmia from ECG Signals,” *Processes*, vol. 11, n.º 2, pág. 595, 2023. DOI: 10.3390/pr11020595. dirección: <https://www.mdpi.com/2227-9717/11/2/595>.

- [8] L. K. Gupta, D. Koundal y S. Mongia, *Explainable Methods for Image-Based Deep Learning: A Review*, ISSN: 18861784 Issue: 4 Pages: 2651-2666 Publication Title: Archives of Computational Methods in Engineering Volume: 30, mayo de 2023. DOI: 10.1007/s11831-023-09881-5.
- [9] H. Wang et al., "Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks," en, en *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA: IEEE, jun. de 2020, págs. 111-119, ISBN: 978-1-7281-9360-1. DOI: 10.1109/CVPRW50498.2020.00020. visitado 10 de jun. de 2025. dirección: <https://ieeexplore.ieee.org/document/9150840/>.
- [10] ONU, *Objetivos y Metas de Desarrollo Sostenible*, 2015. dirección: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>.
- [11] A. Dheeraj y S. Chand, "Deep learning based weed classification in corn using improved attention mechanism empowered by Explainable AI techniques," *Crop Protection*, vol. 190, pág. 107 058, abr. de 2025, ISSN: 0261-2194. DOI: 10.1016/j.cropro.2024.107058. visitado 18 de jun. de 2025. dirección: <https://www.sciencedirect.com/science/article/pii/S0261219424004861>.
- [12] A. Ajay, S. Sandosh, A. Saji y H. Agarwal, "An Explainable Deep Learning Framework for Sorghum Weed Classification Using Multi-Scale Feature Enhanced DenseNet," *IEEE Access*, vol. 13, págs. 26 973-26 990, 2025. DOI: 10.1109/access.2025.3538937.
- [13] D. G. Pai, R. Kamath y M. Balachandra, "Deep Learning Techniques for Weed Detection in Agricultural Environments: A Comprehensive Review," *IEEE Access*, vol. 12, págs. 113 193-113 214, 2024, Publisher: Institute of Electrical and Electronics Engineers. DOI: 10.1109/access.2024.3418454. dirección: <https://ieeexplore.ieee.org/document/10570187>.
- [14] S. Murindanyi, R. Sanya, J. Nakatumba-Nabende, R. Nakibuule y A. Katumba, *Enhanced Infield Agriculture with Interpretable Machine Learning Approaches for Crop Classification*, 2018. dirección: <https://arxiv.org/html/2408.12426v1>.

- [15] S. Farkhani, S. K. Skovsen, M. Dyrmann, R. N. Jørgensen y H. Karstoft, “Weed Classification Using Explainable Multi-Resolution Slot Attention,” *Sensors*, vol. 21, n.º 20, pág. 6705, 2021. DOI: 10.3390/s21206705.
- [16] J. Wang, Z. Qi, Y. Wang e Y. Liu, “A lightweight weed detection model for cotton fields based on an improved YOLOv8n,” en, *Sci Rep*, vol. 15, n.º 1, pág. 457, ene. de 2025, Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10.1038/s41598-024-84748-8. visitado 18 de jun. de 2025. dirección: <https://www.nature.com/articles/s41598-024-84748-8>.
- [17] J. A. Thomasson et al., “AI in Agriculture: Opportunities, Challenges, and Recommendations,” en,
- [18] Z. A. Ganie et al., “Potential potato yield loss from weed interference in the United States and Canada,” *Weed Technology*, págs. 1-4, 2023, Publisher: Cambridge University Press. DOI: 10.1017/wet.2023.5. visitado 1 de ene. de 2023.
- [19] Marcin Woźniak y Muhammad Fazal Ijaz, “Recent advances in big data, machine, and deep learning for precision agriculture,” *Frontiers in plant science*, vol. 15, 2024, Publisher: Frontiers Media. DOI: 10.3389/fpls.2024.1367538.
- [20] H. Karimmojeni, A. Barjasteh, S. Mousavi y A. Bazrafshan, “Determination of the critical period of weed control in potato (*Solanum tuberosum* L.),” *New Zealand Journal of Crop and Horticultural Science*, págs. 1-10, ene. de 2014. DOI: 10.1080/01140671.2013.875926.
- [21] R. K. Bhattarai et al., “Effect of Different Weed Management Practices on growth and yield in Potato at Khumaltar, Lalitpur, Nepal,” en, *Agron J Nepal*, vol. 8, págs. 98-104, oct. de 2024, ISSN: 2091-0649. DOI: 10.3126/ajn.v8i1.70787. visitado 9 de jun. de 2025. dirección: <https://nepjol.info/index.php/AJN/article/view/70787>.
- [22] Juan J Nuñez y Robert G Wilson, *Integrated Weed Management / Potato / Agriculture: Pest Management Guidelines / UC Statewide IPM Program (UC IPM)*. visitado 9 de jun. de 2025. dirección: <https://ipm.ucanr.edu/agriculture/potato/integrated-weed-management/#gsc.tab=0>.

- [23] Dashuai Wang, Wujing Cao, Fan Zhang, Zhuolin Li, Sheng Xu y Xinyu Wu, *A Review of Deep Learning in Multiscale Agricultural Sensing*. visitado 9 de jun. de 2025. dirección: <https://www.mdpi.com/2072-4292/14/3/559>.
- [24] J. U. Md Akbar, S. Kamarulzaman, A. Muzahid, M. A. Rahman y M. Uddin, "A Comprehensive Review on Deep Learning Assisted Computer Vision Techniques for Smart Greenhouse Agriculture," *IEEE Access*, vol. PP, págs. 1-1, ene. de 2024. DOI: 10.1109/ACCESS.2024.3349418.
- [25] M. Dalal y P. Mittal, "A Systematic Review of Deep Learning-Based Object Detection in Agriculture: Methods, Challenges, and Future Directions," *Computers, Materials & Continua*, vol. 84, n.º 1, págs. 57-91, 2025, Publisher: Tech Science Press. DOI: 10.32604/cmc.2025.066056. visitado 1 de ene. de 2025. dirección: <https://www.techscience.com/cmc/online/detail/23287/pdf>.
- [26] H. A. Idris, A. Muhammad, B. Aliyu y S. U. Suleiman, "A Review of Explainable Artificial Intelligence: Taxonomies, Challenges, Implementation Frameworks and Future Directions," en-US, *International Journal of Engineering Research & Technology*, vol. 12, n.º 12, ene. de 2024, Publisher: IJERT-International Journal of Engineering Research & Technology, ISSN: 2278-0181. DOI: 10.17577/IJERTV12IS120105. visitado 9 de jun. de 2025. dirección: <https://www.ijert.org/research/a-review-of-explainable-artificial-intelligence-taxonomies-challenges-implementation-frameworks-and-future-directions-IJERTV12IS120105.pdf>, %20<https://www.ijert.org/a-review-of-explainable-artificial-intelligence-taxonomies-challenges-implementation-frameworks-and-future-directions>.
- [27] Yaganteeswarudu Akkem, Saroj Kumar Biswas y Aruna Varanasi, *Streamlit-based enhancing crop recommendation systems with advanced explainable artificial intelligence for smart farming*. visitado 9 de jun. de 2025. dirección: <https://www.bohrium.com/paper-details/streamlit-based-enhancing-crop-recommendation-systems-with-advanced-explainable-artificial-intelligence-for-smart-farming/1029972603265089550-2558>.

- [28] IBM, *What is Explainable AI (XAI)?* en, mar. de 2023. visitado 9 de jun. de 2025. dirección: <https://www.ibm.com/think/topics/explainable-ai>.
- [29] A. Budhkar, Q. Song, J. Su y X. Zhang, “Demystifying the black box: A survey on explainable artificial intelligence (XAI) in bioinformatics,” *Comput Struct Biotechnol J*, vol. 27, págs. 346-359, ene. de 2025, ISSN: 2001-0370. DOI: 10.1016/j.csbj.2024.12.027. visitado 9 de jun. de 2025. dirección: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11782883/>.
- [30] P. Knab, S. Marton, U. Schlegel y C. Bartelt, *Which LIME should I trust? Concepts, Challenges, and Solutions*, en, arXiv:2503.24365 [cs], mar. de 2025. DOI: 10.48550/arXiv.2503.24365. visitado 9 de jun. de 2025. dirección: <http://arxiv.org/abs/2503.24365>.
- [31] Hassam Ahmed Tahir, Walaa Alayed y Waqar Ul Hassan, “A Federated Explainable AI Framework for Smart Agriculture: Enhancing Transparency, Efficiency, and Sustainability,” en, *ResearchGate*, jun. de 2025. DOI: 10.1109/ACCESS.2025.3571340. visitado 9 de jun. de 2025. dirección: https://www.researchgate.net/publication/391884376_A_Federated_Explainable_AI_Framework_for_Smart_Agriculture_Enhancing_Transparency_Efficiency_and_Sustainability.
- [32] R. N. V. J. Mohan, P. S. Rayanoothala y R. P. Sree, “Next-gen agriculture: integrating AI and XAI for precision crop yield predictions,” *Front Plant Sci*, vol. 15, pág. 1451607, ene. de 2025, ISSN: 1664-462X. DOI: 10.3389/fpls.2024.1451607. visitado 9 de jun. de 2025. dirección: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11751057/>.
- [33] Zilliz, *What are the limitations of Explainable AI?* en. visitado 9 de jun. de 2025. dirección: <https://milvus.io/ai-quick-reference/what-are-the-limitations-of-explainable-ai>.
- [34] G. Visani, E. Bagli, F. Chesani, A. Poluzzi y D. Capuzzo, “Statistical stability indices for LIME: obtaining reliable explanations for Machine Learning models,” en, *Journal of the Operational Research Society*, vol. 73, n.º 1, págs. 91-101, ene. de 2022, arXiv:2001.11757

- [cs], ISSN: 0160-5682, 1476-9360. DOI: 10.1080/01605682.2020.1865846. visitado 10 de jun. de 2025. dirección: <http://arxiv.org/abs/2001.11757>.
- [35] SHAP, *An introduction to explainable AI with Shapley values — SHAP latest documentation*. visitado 10 de jun. de 2025. dirección: https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html.
- [36] SHAP, *SHAP Documentation — SHAP latest documentation*, 2025. dirección: <https://shap.readthedocs.io/en/latest/>.
- [37] S. M. Lundberg y S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” en,
- [38] M. T. Ribeiro, S. Singh y C. Guestrin, “*Why Should I Trust You?: Explaining the Predictions of Any Classifier*,” arXiv:1602.04938 [cs], ago. de 2016. DOI: 10.48550/arXiv.1602.04938. visitado 10 de jun. de 2025. dirección: <http://arxiv.org/abs/1602.04938>.
- [39] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh y D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” *Int J Comput Vis*, vol. 128, n.º 2, págs. 336-359, feb. de 2020, arXiv:1610.02391 [cs], ISSN: 0920-5691, 1573-1405. DOI: 10.1007/s11263-019-01228-7. visitado 10 de jun. de 2025. dirección: <http://arxiv.org/abs/1610.02391>.
- [40] *tf-explain — tf-explain documentation*. visitado 10 de jun. de 2025. dirección: <https://tf-explain.readthedocs.io/en/latest/>.
- [41] *Introduction · Captum*, en. visitado 10 de jun. de 2025. dirección: <https://captum.ai/>.
- [42] J. Klaise, A. Van Looveren, G. Vacanti y A. Coca, *Alibi Explain: Algorithms for Explaining Machine Learning Models*, Issue: 181 Pages: 1–7 Publication Title: Journal of Machine Learning Research Volume: 22 original-date: 2019-02-26T10:10:56Z, jun. de 2021. visitado 10 de jun. de 2025. dirección: <http://jmlr.org/papers/v22/21-0017.html>.
- [43] *interpretml/interpret*, original-date: 2019-05-03T05:47:52Z, jun. de 2025. visitado 10 de jun. de 2025. dirección: <https://github.com/interpretml/interpret>.

- [44] *salesforce/OmniXAI*, original-date: 2022-02-17T08:13:36Z, jun. de 2025. visitado 10 de jun. de 2025. dirección: <https://github.com/salesforce/OmniXAI>.
- [45] *Evaluating XAI for Weed Localization in Crop Images: Survey Based Hypothesis Testing*. visitado 20 de jun. de 2025. dirección: <https://krishikosh.egranth.ac.in/items/f7057693-9d3d-4d5b-8591-872f3d26ee3f>.
- [46] K. Rybczynski, *Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review - IEEE Access*, en-US, jul. de 2024. visitado 20 de jun. de 2025. dirección: https://ieeaccess.ieee.org/featured-articles/cnnsandvision_objectdetection/.
- [47] A. M. Shimaoka, R. C. Ferreira y A. Goldman, "The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks," en, *SBC Reviews on Computer Science*, vol. 4, n.º 1, págs. 28-43, oct. de 2024, Number: 1, ISSN: 2966-3938. DOI: 10.5753/reviews.2024.3757. visitado 20 de jun. de 2025. dirección: <https://journals-sol.sbc.org.br/index.php/reviews/article/view/3757>.
- [48] R. Wirth y J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," en,
- [49] N. R, P. D, S. G y M. D, "Automated weed detection in crop fields using convolutional neural networks: A deep learning approach for smart farming," *Data and Metadata*, vol. 4, págs. 848-848, 7 de mayo de 2025, ISSN: 2953-4917. DOI: 10.56294/dm2025848. visitado 28 de ene. de 2026. dirección: <https://dm.ageditor.ar/index.php/dm/article/view/848>.
- [50] T. Ahmed, M. B. Hasan, S. Ahmed y M. H. Kabir, "ExE-Net: Explainable Ensemble Network for Potato Leaf Disease Classification," en *2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, ISSN: 2576-7046, ago. de 2024, págs. 335-339. DOI: 10.1109/CCECE59415.2024.10667205. visitado 23 de jun. de 2025. dirección: <https://ieeexplore.ieee.org/document/10667205>.
- [51] A. Mamatha y Santhosh Reddy, "Advanced Weed Detection in Agricultural Fields using Vision Transformers and Explainable AI Techniques," en, *International Journal of Intelli-*

- gent Systems and Applications in Engineering*, vol. 12, n.º 22s, págs. 2096-2106, jul. de 2024, Number: 22s, ISSN: 2147-6799. visitado 23 de jun. de 2025. dirección: <https://ijisae.org/index.php/IJISAE/article/view/7557>.
- [52] A. Mamatha y Santhosh Reddy, “Enhanced Weed Detection in Agricultural Fields Using Convolutional Neural Networks and SHAP Interpretability Techniques,” en, *Journal of Information Systems Engineering and Management*, vol. 10, n.º 49s, págs. 1003-1015, mayo de 2025, Number: 49s, ISSN: 2468-4376. DOI: 10.52783/jisem.v10i49s.10022. visitado 23 de jun. de 2025. dirección: <https://jisem-journal.com/index.php/journal/article/view/10022>.
- [53] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff y H. Adam, *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*, 22 de ago. de 2018. DOI: 10.48550/arXiv.1802.02611. arXiv: 1802.02611[cs]. visitado 2 de feb. de 2026. dirección: <http://arxiv.org/abs/1802.02611>.
- [54] M. Tan y Q. V. Le, *EfficientNetV2: Smaller Models and Faster Training*, 23 de jun. de 2021. DOI: 10.48550/arXiv.2104.00298. arXiv: 2104.00298[cs]. visitado 2 de feb. de 2026. dirección: <http://arxiv.org/abs/2104.00298>.
- [55] S. Woo, J. Park, J.-Y. Lee e I. S. Kweon, *CBAM: Convolutional Block Attention Module*, 18 de jul. de 2018. DOI: 10.48550/arXiv.1807.06521. arXiv: 1807.06521[cs]. visitado 2 de feb. de 2026. dirección: <http://arxiv.org/abs/1807.06521>.

ANEXOS

4.5 Repositorio de Código y Scripts de Experimentación

Para garantizar la transparencia, reproducibilidad y validación empírica de los resultados presentados en este trabajo de titulación, se ha publicado el código fuente completo en un entorno interactivo alojado en la plataforma Kaggle.

Este cuaderno de trabajo (*notebook*) incluye el *pipeline* integral de la investigación, abarcando las siguientes etapas:

- Carga y preprocesamiento de los datos (implementación del algoritmo de recorte inteligente y normalización).
- Configuración y despliegue de la arquitectura de segmentación *Attentive DeepLabv3+* con el extractor de características EfficientNetV2-S.
- Ejecución del ciclo de optimización avanzado, incluyendo el salto de resolución a 256px y la técnica de *Stochastic Weight Averaging* (SWA).
- Implementación de las funciones de auditoría visual mediante los algoritmos de Inteligencia Artificial Explicable (Grad-CAM, Score-CAM, LIME y SHAP), así como el cálculo de las métricas de Fidelidad, Localización y Robustez.

El código fuente puede ser consultado, auditado y ejecutado directamente a través del siguiente enlace público:

Enlace al Repositorio de Código:

<https://www.kaggle.com/code/zamyrguevara/xai-potato-fields-efficientnetv2-s>

4.6 Conjunto de Datos (Dataset)

El insumo principal que sustenta los experimentos de esta investigación, denominado “Weeds in Potato Fields v2”, ha sido estructurado y alojado en un repositorio público. La liberación de este conjunto de datos tiene como objetivo fomentar el desarrollo y la comparación de nuevos algoritmos de visión computacional orientados a la agricultura de precisión.

El repositorio contiene:

- Las imágenes aéreas originales de los campos de cultivo de papa, capturadas mediante vehículos aéreos no tripulados (drones).
- Las máscaras de segmentación semántica etiquetadas a nivel de píxel (*Ground Truth*), las cuales definen las seis clases evaluadas (Fondo, Lengua de Vaca, Diente de León, Kikuyo, Otras Malezas y Papa).

El acceso libre al conjunto de datos estructurado se encuentra disponible en la siguiente dirección electrónica:

Enlace al Dataset:

<https://www.kaggle.com/datasets/zamyrguevara/weeds-in-potato-fields-v2>

Diagnóstico de Interpretabilidad (XAI) de 15 Casos Extremos y Promedio

El presente anexo documenta los resultados visuales del diagnóstico de interpretabilidad (XAI) aplicado al modelo de segmentación semántica. Con el fin de evaluar la robustez y transparencia del modelo, se seleccionaron 15 casos representativos basados en la métrica de Intersección sobre Unión (IoU): los 7 casos con mejor rendimiento (“éxito”), 1 caso de rendimiento medio (“promedio”) y los 7 casos con menor rendimiento (“fallo”).

Cada figura presenta un análisis comparativo de 6 paneles que incluye: la imagen original, el *Ground Truth* (contorno verde), y los mapas de atribución generados por Grad-CAM, Score-

CAM, SHAP y LIME. Cabe destacar que en las visualizaciones de LIME, las zonas delineadas en azul representan las áreas que no aportan o penalizan la predicción de la clase objetivo, mientras que las verdes indican contribución positiva.

Casos de Éxito

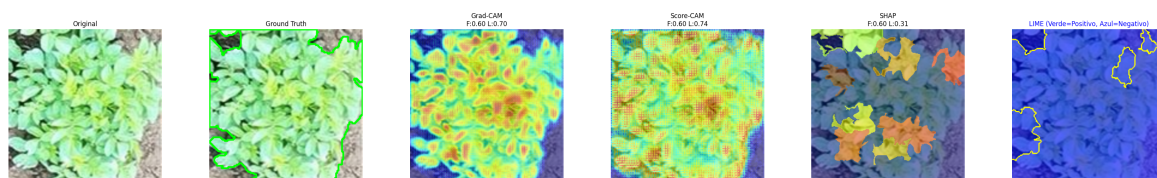


Figura 15.

Diagnóstico XAI – Caso de éxito con IoU de 0.98. Archivo:
EXITO_0.98_rotate270_5-353_jpg.rf.e47242acbac99f7b105fb2c02ed45d82.jpg.png

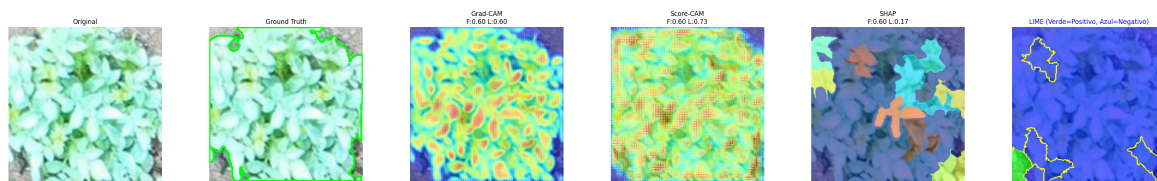


Figura 16.

Diagnóstico XAI – Caso de éxito con IoU de 0.98. Archivo:
EXITO_0.98_rotate270_5-153_jpg.rf.58c8052a49bf1d77e157397211f89331.jpg.png

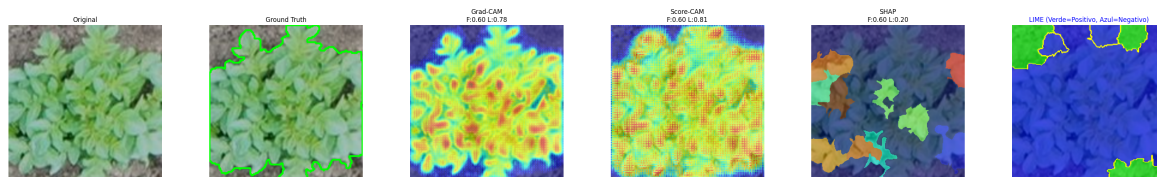


Figura 17.

Diagnóstico XAI – Caso de éxito con IoU de 0.98. Archivo:
EXITO_0.98_dark_5-285_jpg.rf.ce2c919466b75633e8e2ca1bbea60f0a.jpg.png

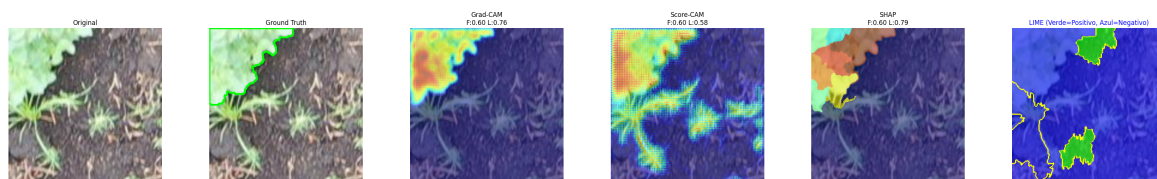


Figura 18.

Diagnóstico XAI – Caso de éxito con IoU de 0.98. Archivo:
EXITO_0.98_DJI_0050-JPG_1500_1250_JPG.rf.1799fec20005683b68c52fb941f18059.jpg.png

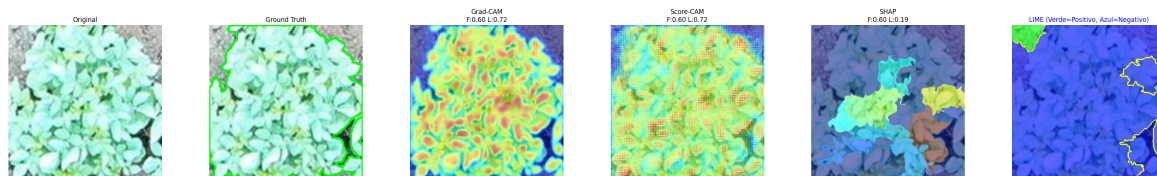


Figura 19.

Diagnóstico XAI – Caso de éxito con IoU de 0.98. Archivo:
 EXIT0_0.98_5-147_jpg.rf.db1e8238fdbf076775ea2f97c8027a61.jpg.png

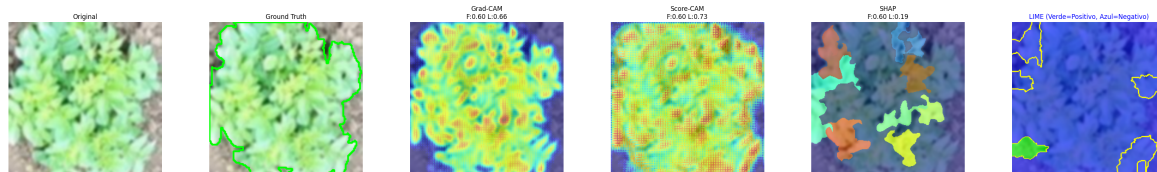


Figura 20.

Diagnóstico XAI – Caso de éxito con IoU de 0.97. Archivo:
 EXIT0_0.97_out_focus_5-318_jpg.rf.4adc34aa1811d33037f8b6048be897b7.jpg.png

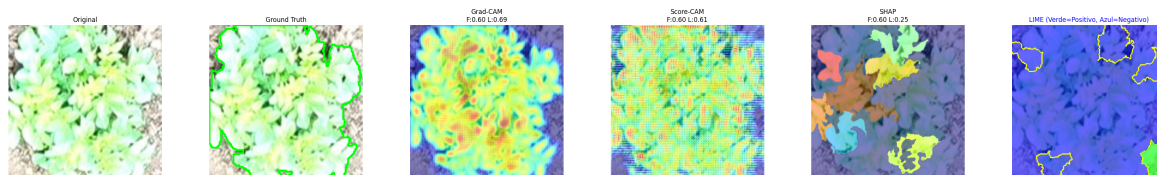


Figura 21.

Diagnóstico XAI – Caso de éxito con IoU de 0.97. Archivo:
 EXIT0_0.97_bright_5-318_jpg.rf.4adc34aa1811d33037f8b6048be897b7.jpg.png

Caso Promedio

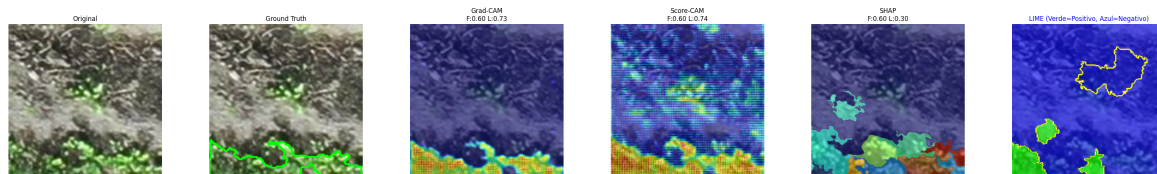


Figura 22.

Diagnóstico XAI – Caso de rendimiento promedio con IoU de 0.89. Archivo:
 PROMEDIO_0.89_bright_dandelion_7_jpg.rf.441e65ee7e0e53900a34c85c0339209e.jpg.png

Casos de Fallo Crítico

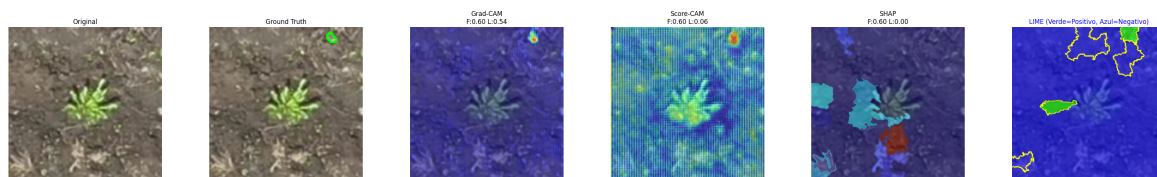


Figura 23.

Diagnóstico XAI – Caso de fallo con IoU de 0.30. Archivo:

FALLO_0.30_dandelion_166_jpg.rf.ffad631ccdf15d5598db2c9718cec15b.jpg.png

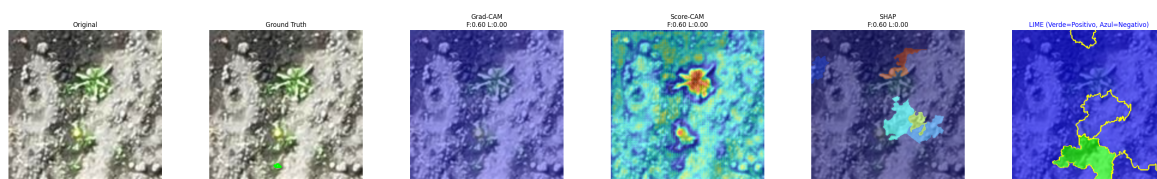


Figura 24.

Diagnóstico XAI – Caso de fallo con IoU de 0.00. Archivo:

FALLO_0.00_rotate90_dandelion_198_jpg.rf.a5dca5cf458a01f114f581ad23e6da96.jpg.png

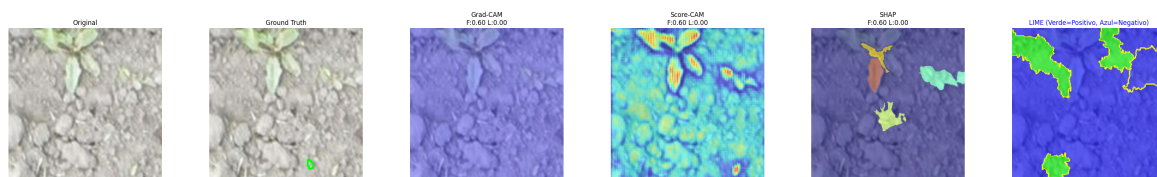


Figura 25.

Diagnóstico XAI – Caso de fallo con IoU de 0.00. Archivo:

FALLO_0.00_rotate270_5-194m3_jpg.rf.a8130ad0d09daf2771cd1014a4533e82.jpg.png

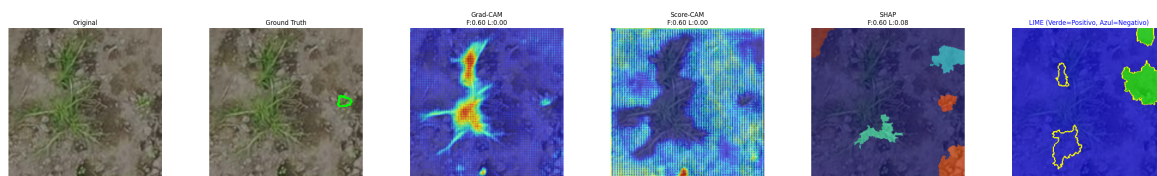


Figura 26.

Diagnóstico XAI – Caso de fallo con IoU de 0.00. Archivo:

FALLO_0.00_rotate180_kikuyo_171_jpg.rf.d07d0c23e9dd34c7e2c0459dadb045c3.jpg.png

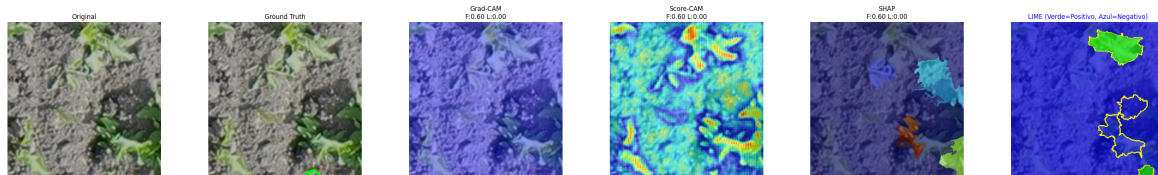


Figura 27.

Diagnóstico XAI – Caso de fallo con IoU de 0.00. Archivo:
 FALLO_0.00_rotate180_5-447m3_jpg.rf.173afa42e9cad331ef874ddf53978811.jpg.png

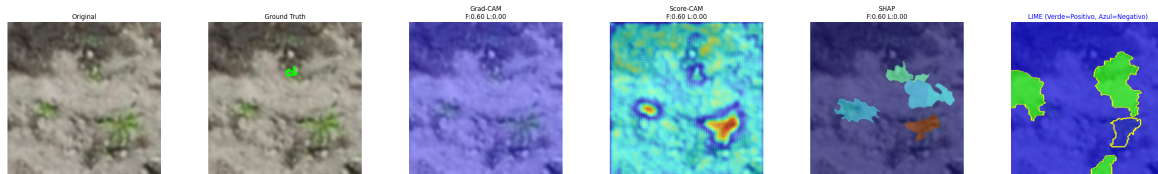


Figura 28.

Diagnóstico XAI – Caso de fallo con IoU de 0.00. Archivo:
 FALLO_0.00_out_focus_dandelion_123_jpg.rf.0d3b9633af2fa24b303801e6e0dd256c.jpg.png

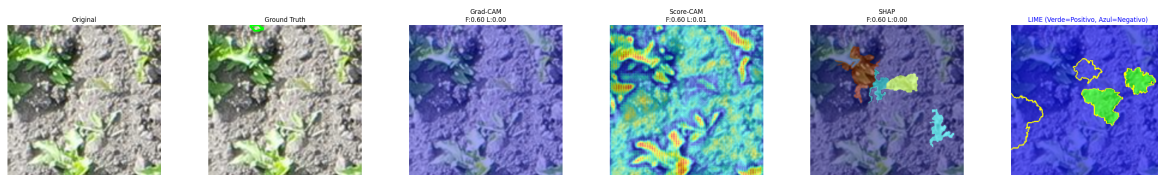


Figura 29.

Diagnóstico XAI – Caso de fallo con IoU de 0.00. Archivo:
 FALLO_0.00_5-447m3_jpg.rf.173afa42e9cad331ef874ddf53978811.jpg.png