UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas

Development and validation of an innovative system for weed species identification and mapping by combining UAV-based imagery and deep learning techniques

T E S I S D O C T O R A L

Presentada para optar al título de Doctor por:

Gustavo Adolfo Mesías Ruiz

Ingeniero Electrónico
Máster en Automática y Robótica

Madrid, 2024

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Agronómica,
Alimentaria y de Biosistemas

Doctorado en Tecnología Agroambiental para una Agricultura Sostenible

Development and validation of an innovative system for weed species identification and mapping by combining UAV-based imagery and deep learning techniques

T E S I S D O C T O R A L

Presentada para optar al título de Doctor por:

Gustavo Adolfo Mesías Ruiz

Ingeniero Electrónico
Máster en Automática y Robótica

Bajo la dirección de:

Dr. José Dorado Gómez
Dr. José Manuel Peña Barragán

Madrid, 2024

Título: Development and validation of an innovative system for weed species identification and mapping by combining UAV-based imagery and deep learning techniques

Autor: Gustavo Adolfo Mesías Ruiz

Programa de Doctorado: Tecnología Agroambiental para una Agricultura Sostenible

Dirección de Tesis:

Dr. José Dorado Gómez, Investigador Científico, Instituto de Ciencias Agrarias - Consejo Superior de Investigaciones Científicas(Director)

Dr. José Manuel Peña Barragán, Científico Titular, Instituto de Ciencias Agrarias - Consejo Superior de Investigaciones Científicas(Co-Director)

Revisores Externos:

Tribunal de Tesis:

Fecha de Defensa de Tesis:

"Esta tesis ha sido financiada por la ayuda FPI del Ministerio de Ciencia e Innovación de España (PRE2018-083227), la Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033) y la Unión Europea NextGenerationEU/PRTR a través del proyecto nhTAI (AGL2017-83325-C4-1-R)."

A mi familia, por su amor incondicional y apoyo en cada paso de este largo camino.

Agradecimientos

Al concluir esta tesis doctoral, quiero dedicar un momento especial para reconocer a todas las personas que, con su apoyo, comprensión y aliento, han sido fundamentales para la culminación de este proyecto. Sin ellas, este logro no hubiera sido posible. A cada uno de ustedes, les extiendo mi más sincero y profundo agradecimiento.

En primer lugar, expreso mi gratitud más sincera a mis directores de tesis, el Dr. José Dorado y el Dr. José Manuel Peña. Su guía experta y su continuo apoyo han sido esenciales para llevar este trabajo a buen puerto. Agradezco profundamente la confianza depositada en mí al integrarme en proyectos de investigación claves para el desarrollo de esta tesis. Su paciencia inquebrantable durante el proceso de redacción, y su dedicación tanto profesional como personal, han dejado una marca imborrable en mi formación, no solo académica, sino también humana.

De igual forma, quiero destacar la labor de David Campos y José Manuel Martín, quienes desempeñaron un papel crucial en la creación de la base de datos de este estudio. Su esfuerzo incansable, reflejado en largas horas de trabajo y dedicación frente al monitor, ha sido indispensable para sentar los cimientos sólidos sobre los cuales se ha construido este proyecto. A ambos, mi más sincero agradecimiento por su compromiso con la excelencia técnica y científica.

Mi reconocimiento también va para mis compañeros del grupo Tech4Agro: César, Dionisio, Ana, Héctor, Vicente, Irene, Benjamín, José, Juan Diego y Miguel. Haber trabajado junto a ustedes ha sido una experiencia enriquecedora en todos los aspectos. Su talento, su apoyo constante y su espíritu colaborativo han sido una fuente de inspiración que me ha impulsado a seguir adelante. Gracias por cada conversación, cada idea compartida y por su amistad invaluable.

Por último, pero no menos importante, a mi familia, a quienes debo todo lo que soy y lo que he logrado. A mi madre Alicia, por su amor incondicional, ese que no conoce límites ni condiciones. Cada sacrificio que has hecho y cada esfuerzo que has invertido son un tesoro invaluable que guardo en mi corazón. Gracias por enseñarme a soñar en grande. A mi padre, Eduardo, por su apoyo constante. A mi hermana, Diana, y a mi hermano, Marlon, por su cariño inquebrantable, por recordarme el valor de los lazos familiares y por ser una constante fuente de apoyo emocional. Este logro es tanto suyo como mío.

A todos ustedes, desde lo más profundo de mi corazón, gracias por haber sido parte de este viaje. Sin su apoyo, nada de esto hubiera sido posible.

Abstract

Crop protection is essential to ensure agricultural production against threats such as weeds, diseases and pests. The evolution towards Agriculture 5.0 has incorporated advanced technologies such as artificial intelligence (AI), robotics and deep learning (DL), which enable smart and sustainable crop management. One of the most critical challenges in crop protection is the early and accurate identification of weed species, which allows for effective control during the most sensitive stage of these weeds. In addition, site-specific weed management in precision agriculture requires detection of these species for targeted mapping. However, this process faces difficulties due to the morphological similarities between crops and weeds in early growth stages. Technologies such as unmanned aerial vehicles (UAVs), advanced sensors and DL are essential in precision agriculture as they improve both, the efficiency and sustainability of farming practices.

The main objective of this research was to develop an automatic system for early identification of weed species using UAV-captured imagery and cutting-edge AI algorithms. To achieve this, advances in weed identification using AI were first reviewed, creating a taxonomy that includes both traditional and advanced neural network-based approaches. In addition, representative datasets of high quality UAV imagery were developed, in which weed species were labeled in their early growth stages and under various conditions.

Using these databases, the study focused on the multiclass classification of weeds through convolutional neural networks (CNNs). To this end, frameworks such as TensorFlow and PyTorch were used, optimizing the handling of large volumes of data and improving computational efficiency. Convolutional neural networks, architectures such as VGG16, ResNet152, Inception-ResNet-v2, EfficientNet-B0 and YOLOv8, as well as vision transformers (ViT) such as ViT-Base and Swin-T were analyzed in the classification task. These models demonstrated their effectiveness in identifying weeds under various environmental conditions, achieving accuracies above 90%. However, the variability in performance with unbalanced datasets complicated the identification of less represented species.

Comparison between CNN-based models and ViT revealed that architectures such as YOLOv8 and Swin-T offer balanced performance in the accurate classification of weeds at the species level in early stages. Additionally, the integration of DL models for object detection, such as Faster R-CNN, YOLOv8m, DETA and DETR with geographic information systems enabled the automated generation of geo-referenced weed distribution maps. The use of techniques such as generative adversarial networks (GANs) to generate synthetic images has shown improvements in model generalization across different phenological states of weeds, significantly increasing accuracy while also raising computational costs.

The integration of UAV imagery and cutting-edge DL algorithms enables early and accurate identification of weed species, as well as the generation of georeferenced maps essential for informed decision-making. These maps facilitate the visualization of weed distribution, contributing to sustainable management strategies and reducing herbicide use. This research, pioneering the detection, classification and mapping of weeds using low-altitude UAV imagery, highlights the need to develop efficient classifiers and detectors for targeted control using advanced technologies.

Resumen

La protección de cultivos es esencial para garantizar la producción agrícola frente a amenazas como malas hierbas, enfermedades y plagas. La evolución hacia la Agricultura 5.0 ha incorporado tecnologías avanzadas como la inteligencia artificial (IA), la robótica y aprendizaje profundo (DL), permitiendo una gestión inteligente y sostenible de los cultivos. Uno de los desafíos más críticos en la protección de cultivos es la identificación temprana y precisa de especies de malas hierbas, lo cual permite realizar un control efectivo en la etapa más sensible de las especies arvenses. Además, el manejo localizado de malas hierbas en el marco de la agricultura de precisión requiere la detección de estas especies para la elaboración de mapas específicos. Sin embargo, este proceso enfrenta dificultades debido a las similitudes morfológicas entre cultivos y malas hierbas en etapas tempranas de crecimiento. Tecnologías como los vehículos aéreos no tripulados (UAV), sensores avanzados y DL son fundamentales en la agricultura de precisión, ya que mejoran tanto la eficiencia como la sostenibilidad de las prácticas agrícolas.

El objetivo principal de esta investigación fue desarrollar un sistema automático para la identificación temprana de especies de malas hierbas utilizando imágenes capturadas por UAV y algoritmos de IA de última generación. Para ello, en primer lugar, se revisaron los avances en la identificación de malas hierbas mediante IA, creando una taxonomía que abarca tanto métodos tradicionales como avanzados basados en redes neuronales. Además, se desarrollaron bases de datos representativas con imágenes de alta calidad obtenidas por UAV, en las cuales se etiquetaron especies de malas hierbas en las primeras etapas de crecimiento y en diversas condiciones.

Utilizando estas bases de datos, el estudio se centró en la clasificación multiclase de malas hierbas mediante redes neuronales convolucionales (CNN). Para ello, se utilizaron *frameworks* como TensorFlow y PyTorch, lo que permitió optimizar el manejo de grandes volúmenes de datos y mejorar la eficiencia computacional. Se evaluaron modelos de CNN como VGG16, ResNet152, Inception-ResNet-v2, EfficientNet-B0 y YOLOv8, así como transformadores de visión (ViT) como ViT-Base y Swin-T. Estos modelos demostraron su eficacia en la identificación de malas hierbas en diversas condiciones ambientales, alcanzando exactitudes superiores al 90%. Sin embargo, la variabilidad del rendimiento en bases de datos desbalanceadas complicó la identificación de especies menos representadas.

La comparación entre modelos basados en CNN avanzadas y ViT reveló que arquitecturas como YOLOv8 y Swin-T ofrecen un rendimiento equilibrado en la clasificación precisa de malas hierbas a nivel de especie en etapas tempranas. Además, la integración de modelos de DL para la detección de objetos, como Faster R-CNN, YOLOv8m, DETA y DETR, con sistemas de información geográfica, permitió la generación automatizada de mapas georreferenciados de distribución de malas hierbas. El uso de técnicas como redes generativas adversarias (GANs) para generar imágenes sintéticas ha mostrado mejoras en la generalización de modelos en distintos estados fenológicos de las malas hierbas, incrementando significativamente la exactitud, aunque también aumentando el costo computacional.

Esta investigación, pionera en la detección, clasificación y mapeo de malas hierbas a partir de imágenes de UAV, destaco la necesidad de desarrollar clasificadores y detectores eficientes para un control específico mediante tecnologías avanzadas. La integración de imágenes UAV y algoritmos

de DL de última generación permitió la identificación temprana y precisa de especies de malas hierbas, así como la generación de mapas georeferenciados esenciales para una toma de decisiones informadas. Estos mapas facilitan la visualización de la distribución de malas hierbas, contribuyendo a estrategias de manejo sostenible y a la reducción del uso de herbicidas.

Tabla de Contenido

Lista de Figuras

Lista de Tablas

Abreviaturas

UPM Universidad Politécnica de Madrid

CSIC Consejo Superior de Investigaciones Científicas

ICA Instituto de Ciencias Agrarias

GNSS *Global Navigation Satellite System*

IoT *Internet de las cosas*

Ag5.0 Agricultura 5.0

IA Inteligencia Artificial

ML *Machine Learning*

SSWM *Site-Specific Weed Management*

DSS *Decision Support System*

UAV *Unmanned Aerial Vehicle*

RGB *Red-Green-Blue*

CPU *Central Processing Unit*

GPU *Graphics Processing Unit*

TPU *Tensor Processing Unit*

OBIA *Object-Based Image Analysis*

ANN *Artificial Neural Network*

DL *Deep Learning*

CNN *Convolutional Neural Network*

RNN *Recurrent Neural Network*

GAN *Generative Adversarial Network*

R-CNN *Region-based Convolutional Neural Network*

YOLO *You Only Look Once*

SSD *Single Shot Detector*

FPN *Feature Pyramid Network*

CC Computación Cognitiva

ViT *Visual Transformers*

FCN *Fully Convolutional Network*

TL *Transfer Learning*

RVR *Relative Variation Rate*

CAGR *Compound Annual Growth Rate*

GSD *Ground Sample Distance*

ReLU *Rectified Linear Unit*

NLP *Natural Language Processing*

MHSA *Multi-Head Self-Attention*

DeiT *Data-efficient image Transformer*

RPN *Region Proposal Network*

STN *Spatial Transformer Network*

IoU *Intersection over Union*

mAP *Mean Average Precision*

LCS *Local Coordinate System*

GCS *Global Coordinate System*

WGS84 *World Geodetic System* 84

API *Application Programming Interface*

VGGNet *Visual Geometry Group Network*

ResNet *Residual Network*

GFPGAN *Generative Facial Prior GAN*

MSE *Mean Squared Error*

DETR *Detection Transformer*

DETA *Detection Transformers with Assignment*

Capítulo 1

Introducción General

1.1 Avances tecnológicos en la protección de cultivos

La protección de cultivos es un conjunto integral de prácticas y tecnologías orientadas a salvaguardar la producción agrícola frente a diversas amenazas, como malas hierbas, enfermedades y plagas del cultivo. Este campo abarca desde métodos tradicionales, como el control mecánico y biológico, hasta la aplicación de plaguicidas de síntesis. La importancia de la protección de cultivos radica en su capacidad para minimizar las pérdidas potenciales de rendimiento, que pueden oscilar entre el 50 % y el 80 % de la producción agrícola mundial, según el tipo de cultivo y la región (Oerke y Dehne, 2004).

A lo largo de la historia, la humanidad ha desarrollado continuamente nuevos métodos y prácticas para proteger sus cultivos. Desde la antigüedad hasta la década de 1950, la llamada Agricultura 1.0 dependía en gran medida de la mano de obra para el control manual de malas hierbas, enfermedades y plagas, lo que resultaba en rendimientos bajos pero suficientes para alimentar a la población. A finales de la década de 1950, la Agricultura 2.0 trajo consigo el uso de plaguicidas sintéticos y maquinaria especializada, marcando el inicio de un enfoque más industrializado y orientado a la producción a gran escala y a menor costo por unidad de producto. A finales del siglo XX, con la Agricultura 3.0, emergió un enfoque disruptivo basado en el uso de nuevas tecnologías y modelos basados en datos. Este concepto, conocido como agricultura de precisión, introdujo herramientas como la telemática, los sistemas de navegación global por satélite (GNSS, del inglés *Global Navigation Satellite System*), maquinaria guiada y dispositivos de detección, cuyo objetivo era optimizar las tareas de protección de cultivos, reducir costos y minimizar el impacto ambiental de los plaguicidas, mejorando al mismo tiempo la calidad de los alimentos. Posteriormente, la integración de tecnologías geoespaciales, ciencias computacionales y digitalización en la agricultura dio paso a la Agricultura 4.0. En esta etapa, sensores, telefonía móvil, sistemas embebidos, computación en la nube, internet de las cosas (IoT, del inglés *Internet of Things*) y *big data*, incorporados en maquinaria autónoma y pulverizadores inteligentes, permitieron implementar el paradigma de protección de cultivos de precisión (Zhai et al., 2020). Actualmente, la Agricultura 5.0 (Ag5.0) representa un salto hacia la gestión inteligente de cultivos, caracterizada por la automatización de procesos de toma de decisiones, operaciones no tripuladas y una intervención humana cada vez menor. Este avance está respaldado por sistemas de inteligencia artificial (IA), robótica avanzada y complejos algoritmos de Aprendizaje Automático (ML, del inglés *Machine Learning*) (Saiz-Rubio y Rovira-Más, 2020).

En las próximas décadas, la agricultura moderna se enfrentará a dos desafíos sin precedentes. El primero es el impacto del cambio climático en los sistemas agrícolas (Hoegh-Guldberg et al., 2019), que genera desestabilización en las prácticas agrícolas (Mulla et al., 2020) y temporadas de cultivo irregulares debido a eventos climáticos extremos (e.g., altas termperaturas, sequías prolongadas, lluvias torrenciales, etc.) en grandes regiones productivas (Falkland y White, 2020; Piao et al., 2019), lo que inevitablemente conllevará la aparición de nuevas especies invasoras o el aumento de la gravedad de las ya existentes. El segundo reto es alimentar a una población humana y animal en crecimiento, garantizando a su vez la seguridad alimentaria y, al mismo tiempo, utilizando menos agroquímicos y aplicando controles estrictos a lo largo de toda la cadena de suministro agrícola (van Dijk et al., 2020). Frente a este panorama, la Ag5.0 deberá ofrecer soluciones innovadoras basadas en IA, algoritmos de ML y otras tecnologías emergentes que interactúen continuamente con los cultivos y su entorno. Esto requerirá un enfoque transdisciplinario, donde la protección de cultivos de precisión se posicionará como una disciplina clave en la revolución de la Ag5.0, mediante

la implementación de nuevas estrategias para reducir de manera drástica el uso de agroquímicos en el control de malas hierbas, enfermedades y plagas.

1.2 Protección de cultivos de precisión: manejo localizado de las malas hierbas

Las malas hierbas representan una de las mayores amenazas para la productividad agrícola, siendo responsables del mayor porcentaje de pérdidas potenciales en los cultivos a nivel mundial. Según los estudios de Oerke y Dehne (2004), y Radicetti y Mancinelli (2021), las pérdidas por las malas hierbas oscilan entre el 32 % y el 34 %, respectivamente, superando a los efectos de otros grupos de patógenos y plagas agrícolas. Esta elevada incidencia se debe a su alta capacidad competitiva y adaptabilidad en condiciones adversas, compitiendo de manera efectiva con los cultivos por nutrientes, humedad, luz y espacio. Aunque los herbicidas han sido la principal herramienta para su control debido a su alta eficacia, la dependencia excesiva de estos productos ha impulsado la búsqueda de enfoques más sostenibles y de menor impacto ambiental, como el manejo lozalizado de malas hierbas (SSWM, del inglés *Site-Specific Weed Management*). El SSWM es una estrategia agronómica que se centra en la ubicación precisa, identificación y manejo dirigido de poblaciones de malas hierbas en áreas específicas, evitando tratar todo el campo de manera uniforme (Fernández-Quintanilla et al., 2018; Lati et al., 2021).

La identificación temprana y exacta de especies de malas hierbas es un componente clave del SSWM, dado que diferentes especies requieren métodos de control específicos (Fernandez-Quintanilla et al., 2022; Sa et al., 2018; Søgaard y Lund, 2007). No obstante, esta tarea es compleja y requiere mucho tiempo debido a las similitudes morfológicas entre muchas malas hierbas y las plantas de cultivo, lo que demanda herramientas especializadas y conocimientos expertos (Fernández-Quintanilla et al., 2018). La identificación incorrecta puede llevar a decisiones de tratamiento erróneas y, en consecuencia, a una reducción en la efectividad del control. Además, la variabilidad morfológica de una misma especie de mala hierba a lo largo de sus diferentes etapas de crecimiento hace imperativo el desarrollo de sistemas capaces de reconocer y distinguir con exactitud estos cambios (Wang et al., 2019a).

El uso de nuevas tecnologías en la protección de cultivos tiene como objetivo detectar e identificar los síntomas o problemas causados por las malas hierbas, enfermedades y plagas de cultivos (Behmann et al., 2015), seguido de una aplicación localizada mediante un control químico o mecánico. Este proceso se organiza en tres etapas principales dentro de la estrategia de protección de cultivos de precisión: 1) percepción, 2) análisis y toma de decisiones, y 3) actuación. La etapa de percepción implica la inspección del campo y la adquisición de información de las plantas (por ejemplo, imágenes de cultivos y/o malas hierbas) mediante sensores o cámaras montadas en plataformas terrestres o aéreas. La etapa de actuación consiste en la aplicación precisa del tratamiento con equipos inteligentes, usualmente asistidos por receptores GNSS. El puente entre la percepción y la actuación es la etapa de análisis, que implica una evaluación exhaustiva de los datos obtenidos, seguida de la generación de mapas de tratamiento o prescripción basados en sistemas de apoyo a la toma de decisiones (DSS, del inglés *Decision Support System*).

Recientes revisiones destacan a los drones, también conocidos como vehículos aéreos no tripulados

(UAV, del inglés *Unmanned Aerial Vehicle*), los algoritmos de ML, diversos robots y equipos autónomos como las tecnologías más disruptivas en cada etapa del proceso (Cardim Ferreira Lima et al., 2020; Dainelli et al., 2021; Filho et al., 2020). Los UAVs desempeñan un papel crucial en la fase de percepción gracias a su capacidad para capturar datos sobre grandes extensiones de terreno en poco tiempo, utilizando diferentes tipos de cámaras y sensores, tales como cámaras RGB (del inglés, *Red-Green-Blue*), sensores multiespectrales e hiperespectrales, cámaras térmicas y sensores activos como LiDAR, radar o sonar. Estas tecnologías han permitido avances significativos en la monitorización, tanto por observación directa de la posición de rodales de malas hierbas, como por el diagnóstico de síntomas asociados a enfermedades (por ejemplo, decaimiento de las hojas o estrés térmico) o a plagas de los cultivos (por ejemplo, daños foliares).

La etapa de análisis constituye uno de los mayores desafíos para la protección de cultivos de precisión y ,probablemente, representa el principal cuello de botella para su aplicación. El objetivo final de esta etapa es la detección precisa y oportuna de cada especie de mala hierba, enfermedad o plaga, en un contexto caracterizado por la gran diversidad de escenarios de cultivo-plaga y la variabilidad de síntomas asociados. Además, factores ambientales y culturales, como las condiciones climáticas, las propiedades del suelo y las decisiones de manejo del agricultor, también influyen en su tipología, gravedad e impacto en el cultivo (Oerke et al., 2012; Pätzold et al., 2020). La complejidad de este proceso puede abordarse mediante técnicas de ML, que permiten aprender de grandes volúmenes de datos y de la interacción entre múltiples factores. Los avances en hardware, con procesadores centrales (CPU, del inglés *Central Processing Unit*), gráficos (GPU, del inglés *Graphics Processing Unit*) y tensores (TPU, del inglés *Tensor Processing Unit*), cada vez más potentes (Wang et al., 2019c), han hecho posible el análisis de cada vez mayores cantidades de datos a lo largo del tiempo. Esto permite la construcción de modelos predictivos y generativos que facilitan tareas analíticas críticas, como la clasificación de imágenes, la detección de objetos, el reconocimiento de patrones y la geolocalización, entre otros con el objetivo de proponer soluciones a los complejos desafíos de la protección de cultivos.

Finalmente, la actuación es el componente clave que permite la implementación a gran escala de las estrategias de protección de cultivos de precisión. En la última década, se ha realizado un gran esfuerzo científico y tecnológico para desarrollar maquinaria autónoma, pulverizadores inteligentes y robots agrícolas capaces de implementar de manera eficaz las estrategias de SSWM (Lowenberg-DeBoer et al., 2021; Shafi et al., 2019). Estos avances permiten aplicar tratamientos directos en tiempo real (Pérez-Ruiz et al., 2015) o mediante mapas de prescripción (Fernández-Quintanilla et al., 2018), siguiendo los principios establecidos por la Sociedad Internacional de Agricultura de Precisión (ISPA, 2021).

1.3 Visión artificial aplicada a la detección de malas hierbas

El desarrollo de sistemas de visión artificial ha permitido optimizar la automatización de procesos agrícolas, entre ellos, la detección exacta de malas hierbas, mediante técnicas avanzadas de procesamiento de imágenes. Dos de las herramientas más relevantes en este ámbito son los índices de color y de vegetación, que permiten la diferenciación entre cultivos y malas hierbas. Los índices de color son fórmulas matemáticas que combinan valores de bandas roja, verde y azul, dentro del espacio de color RGB. Esta combinación resalta las características cromáticas de las plantas, facilitando su

segmentación del fondo. Por otro lado, los índices de vegetación, basados en bandas espectrales visibles e infrarrojas, destacan las diferencias de reflectividad o reflectancia entre tipos de cobertura del suelo (e.g., vegetación, suelo desnudo, etc.). Ambos índices se caracterizan por su simplicidad matemática y su eficiencia computacional, lo que permite una implementación rápida en sistemas de procesamiento de imágenes. Son particularmente útiles para separar vegetación de suelo desnudo, generando un contraste elevado que facilita la discriminación entre cultivos y malas hierbas.

Sin embargo, el uso de índices de color para la identificación de malas hierbas presenta diversas limitaciones que pueden afectar su exactitud y eficacia (Hasan et al., 2021). Estas limitaciones incluyen:

- Sensibilidad a las condiciones de iluminación (Golzarian et al., 2012): Los índices de color son especialmente sensibles a variaciones en la iluminación, lo que puede alterar los colores percibidos en la imagen y causar errores en la segmentación. Esto dificulta la detección precisa de vegetación y suelo en condiciones cambiantes de luz.
- Interferencia de residuos y elementos del suelo: La presencia de residuos, piedras, o materiales reflectivos en el suelo puede generar falsos positivos, llevando a una incorrecta clasificación de elementos no vegetales como plantas.
- Variabilidad en la apariencia de las plantas (Sancho-Adamson et al., 2019): Factores como la etapa de crecimiento, el estado de salud, o las diferencias entre especies pueden modificar los valores de los índices de color, limitando su aplicabilidad en escenarios agrícolas diversos.
- Dificultades en la diferenciación entre cultivos y malas hierbas (Argüelles y March, 2022): Los índices de color suelen fallar al intentar diferenciar entre cultivos y malas hierbas que comparten colores similares, restringiendo su utilidad en aplicaciones con cultivo presente.
- Dependencia del espacio de color RGB (Sachin et al., 2018): Muchos índices de color se basan en el espacio de color RGB, que no es perceptualmente uniforme y es más susceptible a variaciones de iluminación que otros espacios de color como el modelo de color de la Comisión Internacional de Iluminación (CIE, del inglés *Commission Internationale d'Eclairage*), CIELAB, lo que reduce la robustez de estos índices en condiciones de campo.
- Necesidad de ajuste de umbrales (Meyer y Neto, 2008): La conversión de índices de color en imágenes binarias requiere la selección de umbrales adecuados, un proceso que puede ser inconsistente y demandar ajustes manuales o algoritmos adicionales, lo que incrementa la posibilidad de segmentaciones inexactas.
- Sobrecarga computacional en métodos combinados: Aunque la combinación de varios índices de color puede mejorar la precisión, también incrementa la complejidad computacional, lo que representa un desafío para aplicaciones en tiempo real o con grandes volúmenes de datos.

La similitud en las propiedades de reflectancia entre los cultivos y las malas hierbas durante las etapas tempranas de crecimiento constituye un desafío importante para una adecuada discriminación, lo que subraya la necesidad de técnicas más robustas y adaptativas. En este contexto, el análisis de imágenes basado en objetos (OBIA, del inglés *Object-Based Image Analysis*) se ha propuesto como un enfoque complementario y efectivo (Blaschke, 2010). A diferencia de los enfoques basados en índices o píxeles, OBIA segmenta la imagen en objetos significativos, como plantas individuales

u órganos vegetales, utilizando la similitud espectral y la proximidad espacial de los píxeles. Esto permite extraer y clasificar características descriptivas como textura, tamaño, forma, orientación y contexto espacial. El enfoque OBIA ha demostrado ser particularmente útil para la cartografía de malas hierbas en cultivos en hileras durante las primeras etapas de crecimiento. Estudios basados en imágenes adquiridas desde UAV han demostrado su efectividad en cultivos como maíz (López-Granados et al., 2016a; Peña et al., 2013), algodón (de Castro et al., 2018) y girasol (de Castro et al., 2018; López-Granados et al., 2016b; Torres-Sánchez et al., 2021), entre otros.

1.4 Inteligencia Artificial para el análisis de imágenes dirigido a clasificar malas hierbas

La IA se define como el campo que busca desarrollar máquinas capaces de realizar tareas que normalmente requieren inteligencia humana. Estas tareas incluyen la resolución de problemas complejos, la toma de decisiones, el reconocimiento de patrones y el aprendizaje autónomo. Arthur Samuel[a], considerado uno de los pioneros en el campo del ML, definió esta disciplina como la capacidad de las computadoras para aprender y mejorar su desempeño sin necesidad de ser programadas explícitamente para cada tarea. Posteriormente, Rich (1985) describió la IA como la rama de la informática que se enfoca en diseñar sistemas capaces de realizar tareas que, hasta ese momento, requerían habilidades humanas superiores. En este contexto, John McCarthy, reconocido como uno de los fundadores de la IA, planteó que el objetivo fundamental de esta disciplina es desarrollar máquinas capaces de comportarse de manera inteligente (McCarthy et al., 2007).

En la actualidad el ML, es un subcampo de la IA, se centra en el uso de algoritmos para extraer información de datos y construir modelos que puedan hacer predicciones o tomar decisiones basadas en nuevos datos no previamente modelados. El ML permite que una máquina aprenda y mejore su rendimiento en una tarea específica a partir de la experiencia, sin ser explícitamente programada para ello, lo que es fundamental en el desarrollo de sistemas inteligentes que se adaptan y mejoran con el tiempo, basándose en los datos procesados y la retroalimentación recibida del entorno (Patterson y Gibson, 2017).

1.4.1 Algoritmos tradicionales de *Machine Learning*

Los algoritmos tradicionales de ML abordan el aprendizaje analizando e interpretando datos de entrada mediante arquitecturas bien establecidas, que están optimizadas para recursos computacionales estándar. Si bien estos algoritmos logran resultados satisfactorios en muchas aplicaciones, a menudo carecen de la exactitud y versatilidad de los modelos más modernos y complejos. Entre las tareas más comunes del ML, la clasificación destaca por su uso generalizado en diversas disciplinas. Los algoritmos más conocidos para clasificación incluyen las máquinas de vectores de soporte (SVM, del inglés *Support Vector Machines*), los árboles de decisión (DT, del inglés *Decision Trees*), los bosques aleatorios (RF, del inglés *Random Forest*) y el algoritmo de los k-vecinos más cercanos (kNN, del inglés *k-Nearest Neighbors*).

En la clasificación supervisada, el objetivo es categorizar datos estructurados o no estructurados en

[a] http://infolab.stanford.edu/pub/voy/museum/samuel.html

clases predeterminadas. Esto puede implicar clasificaciones binarias, donde se predice un estado de verdadero o falso, o clasificaciones multi-categoría, cuando hay más de dos clases objetivo (Djafri y Gafour, 2022; Sen et al., 2020). Estos algoritmos son ampliamente utilizados en análisis de imágenes, reconocimiento de objetos y detección de malas hierbas, enfermedades de plantas y plagas (Mesías-Ruiz et al., 2023). Las fases previas, como el preprocesamiento de imágenes, la segmentación y la extracción de características, son esenciales para mejorar la eficacia de la clasificación. Esto a menudo implica el uso de otros tipos de algoritmos, como los de regresión, agrupamiento y reducción de dimensionalidad.

Los algoritmos de regresión, que también forman parte del aprendizaje supervisado, se utilizan para modelar la relación entre variables de entrada y salida continuas mediante funciones paramétricas (como la regresión lineal) o no paramétricas (como el uso de funciones kernel). La regularización, como la técnica LASSO o la regresión Ridge, es fundamental para evitar el sobreajuste en los modelos, mejorando la exactitud tanto en la fase de entrenamiento como en la de prueba (Zou y Hastie, 2005).

Los algoritmos de agrupamiento (*ensemble*) combinan las predicciones de varios modelos de ML para mejorar el rendimiento predictivo, y son particularmente efectivos en tareas de clasificación y regresión. Ejemplos de estos algoritmos incluyen *AdaBoost*, *Bagging*, *CatBoost* y máquinas de incremento de gradiente (GBM, del inglés *Gradient Boosting Machines*), que ayudan a equilibrar el rendimiento y el costo computacional (Telikani et al., 2022). Por otro lado, los algoritmos de agrupamiento son esenciales en el aprendizaje no supervisado y semi-supervisado, permitiendo organizar los datos en grupos con objetos similares. Los métodos más comunes incluyen el agrupamiento particional, que agrupa los datos en un número fijo de clusters (Nanda y Panda, 2014), y el agrupamiento jerárquico, que crea una estructura en forma de árbol (dendrogramas) que permite identificar relaciones jerárquicas entre los datos sin necesidad de definir previamente el número de clusters (Murtagh y Contreras, 2012).

La reducción de dimensionalidad es otra técnica clave en el preprocesamiento de datos, ya que simplifica las bases de datos al eliminar características irrelevantes, reduciendo la complejidad computacional (Xu et al., 2019). Esta técnica es especialmente útil para el preprocesamiento de datos en problemas complejos, permitiendo conservar las características más relevantes del conjunto original y eliminar aquellas menos útiles (Chhikara et al., 2020).

Finalmente, los algoritmos de detección de anomalías (e.g. *Isolation Forest, One-Class SVM*, Detección de Anomalías Basada en Análisis de Componentes Principales (PCA)) y los de aprendizaje de reglas de asociación (e.g. *Apriori*, *Eclat*) juegan un papel importante en la identificación de patrones irregulares o en la generalización de reglas dentro de las bases de datos, siendo útiles para la identificación y control de malas hierbas y otras amenazas agrícolas (Chandola et al., 2009; Fürnkranz y Kliegr, 2015).

1.4.2 Redes Neuronales Artificiales y modelos de *Deep Learning*

Las redes neuronales artificiales (ANN, del inglés *Artificial Neural Network*) son modelos computacionales altamente flexibles y personalizables, inspirados en las redes neuronales biológicas. Estas redes consisten en múltiples capas de neuronas artificiales que procesan información y, mediante el entrenamiento con grandes bases de datos, aprenden patrones complejos para resolver problemas

no lineales. Las ANN actúan como aproximadores universales de funciones matemáticas, lo que les permite realizar tareas como la clasificación, la regresión, la identificación de patrones y la generación de nuevos datos (Schmidhuber, 2015). Su capacidad para aprender de manera autónoma las características relevantes de los datos sin intervención humana directa es crucial en aplicaciones como la protección de cultivos, donde se requiere analizar grandes volúmenes de datos provenientes de sensores y cámaras.

El desarrollo de ANN con múltiples capas ha dado lugar al aprendizaje profundo (DL, del inglés *Deep Learning*), una extensión que ha transformado múltiples disciplinas, incluida la protección de cultivos de precisión (Allmendinger et al., 2022; Farooq et al., 2019; Ferentinos, 2018; Hasan et al., 2021; Kamilaris y Prenafeta-Boldú, 2018; Rai et al., 2023; Rakhmatulin et al., 2021; Tugrul et al., 2022; Xia et al., 2018). El DL permite la creación de estructuras jerárquicas de aprendizaje, donde las características más simples son combinadas en niveles superiores para formar conceptos más complejos (LeCun et al., 2015). Este enfoque ha revolucionado campos como la visión artificial, permitiendo un análisis más rápido y eficiente de imágenes. Además, la posibilidad de ajustar pequeñas bases de datos a modelos preentrenados con datos diversos optimiza los tiempos de entrenamiento y reduce el consumo de recursos computacionales (Kamilaris y Prenafeta-Boldú, 2018).

El desarrollo del DL ha sido progresivo desde los años 40, destacando hitos como el Perceptrón de Frank Rosenblatt, que, a pesar de sus limitaciones iniciales, estableció las bases de las redes neuronales. Sin embargo, el algoritmo de retropropagación (Arbib, 1969), retomado en los años 80 permitió avances significativos al optimizar el entrenamiento de redes multicapa (Patterson y Gibson, 2017). En la última década, el DL ha crecido exponencialmente gracias a la disponibilidad de grandes volúmenes de datos y al aumento de la capacidad computacional, particularmente con el uso de GPUs. Actualmente, el DL supera a los métodos tradicionales de ML en diversas tareas, destacándose especialmente en la visión artificial y el reconocimiento de imágenes (LeCun et al., 2015). El uso de DL en la agricultura de precisión representa un avance crucial. Los modelos predictivos avanzados que generan estos algoritmos permiten optimizar la toma de decisiones basada en datos promoviendo un uso más eficiente de los recursos. La capacidad del DL para identificar patrones complejos y no lineales en datos multiespectrales, geoespaciales y climáticos es esencial para implementar estrategias de manejo agrícola más adaptativas y sostenibles, posicionándose como una herramienta indispensable en la agricultura inteligente.

Dentro del DL, destacan enfoques especializados, como las Redes Neuronales Convolucionales (CNN, del inglés *Convolutional Neural Network*), aplicadas en visión artificial y clasificación de imágenes; las Redes Neuronales Recurrentes (RNN, del inglés *Recurrent Neural Network*), utilizadas para la predicción temporal; y las Redes Generativas Adversarias (GAN, del inglés *Generative Adversarial Network*), empleadas en la generación de imágenes sintéticas (Sarker, 2021). Las CNN han revolucionado la clasificación de imágenes en la protección de cultivos al extraer automáticamente características de los datos de imagen, lo que contrasta con los métodos tradicionales de ML que requieren la selección manual de dichas características (Hong et al., 2020). El desempeño de las CNN depende del número de capas y parámetros, siendo más efectivas con mayor profundidad, aunque esto implica mayores demandas computacionales.

Una extensión de las CNN es su aplicación en la detección de objetos, que mejora el reconocimiento visual en contextos multicategoría. Entre las arquitecturas de CNN más utilizadas en la protección de

cultivos están la Red Neuronal Convolucional Basada en Regiones (R-CNN, del inglés *Region-based Convolutional Neural Network*) (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2016a), You Only Look Once (YOLO) (Redmon et al., 2016), Single Shot Detector (SSD) (Liu et al., 2016), Redes de Pirámide de Características (FPN, del inglés *Feature Pyramid Network*) (Lin et al., 2017a), RetinaNet (Lin et al., 2017b) y Mask R-CNN (He et al., 2017).

1.4.3 Computación Cognitiva

La computación cognitiva (CC) representa una nueva era en la tecnología de la información, donde los sistemas computacionales no solo procesan datos, sino que también aprenden a simular funciones cognitivas humanas como la percepción, la memoria y el razonamiento. Esta capacidad para analizar grandes cantidades de datos y reconocer patrones complejos tiene un enorme potencial en la agricultura de precisión, especialmente en la detección y clasificación de malas hierbas, plagas y enfermedades (Aghav-Palwe y Gunjal, 2021). Los sistemas de CC pueden acelerar el análisis de datos, mejorando la toma de decisiones en tiempo real y optimizando los procesos de producción (Sreedevi et al., 2022).

Uno de los principales desafíos de la CC es la necesidad de grandes bases de datos para entrenar los modelos, algo que no siempre está disponible en la protección de cultivos. Para mitigar esta limitación, se han desarrollado técnicas de aumento de datos, como las GANs, que generan datos sintéticos a partir de imágenes existentes, mejorando así el rendimiento y la capacidad de generalización de los modelos (Mumuni y Mumuni, 2022). Estas técnicas permiten entrenar modelos con mayor precisión, reduciendo el riesgo de sobreajuste y mejorando la detección de anomalías en los cultivos.

El uso de arquitecturas avanzadas como los transformadores de visión (ViT, del inglés *Visual Transformer*), que emplean mecanismos de atención dinámica, ha revolucionado tareas como la visión artificial y el procesamiento del lenguaje natural (Dosovitskiy et al., 2021; Vaswani et al., 2017). De hecho, la aparición de ViTs ha dado lugar a modelos comparables a las CNN en el dominio de las tareas de visión artificial (Dosovitskiy et al., 2021; Liu et al., 2023; Yang et al., 2022). Esta arquitectura funciona de manera similar al cerebro humano, utilizando conocimientos previos para tomar decisiones en nuevas situaciones, facilitada por su mecanismo de atención dinámica. Los ViT permiten una integración eficiente de datos multiespectrales y contextuales, facilitando decisiones precisas en escenarios agrícolas complejos.

La CC también facilita la adaptación a nuevos entornos sin necesidad de grandes volúmenes de datos de entrenamiento, minimizando los riesgos de sesgo y sobreajuste (Lytras y Visvizi, 2021). Esta capacidad es crucial para gestionar complejidades en la producción agrícola moderna, integrando datos de diversas fuentes para apoyar decisiones basadas en evidencia (Lonij y Fiot, 2016). La CC está revolucionando la agricultura moderna mediante la integración de técnicas de *Soft Computing*, como la lógica difusa, las ANN y los mapas cognitivos difusos. Estas tecnologías permiten una gestión más eficiente y sostenible de los recursos, optimizando los rendimientos de los cultivos y reduciendo el impacto ambiental.

1.5 Mapas de prescripción para el tratamiento localizado de malas hierbas

Los mapas de prescripción se han consolidado como herramientas fundamentales para optimizar el uso de insumos en la gestión de cultivos. Estos mapas permiten la zonificación detallada de parcelas agrícolas, basada en parámetros específicos como el umbral de tratamiento o la malla de tratamiento, facilitando la aplicación de dosis variables de herbicidas según las necesidades de cada área de la parcela. Este enfoque resulta particularmente eficaz para el control de malas hierbas, plagas y enfermedades, mejorando no solo la eficiencia de los tratamientos, sino también reduciendo significativamente el impacto ambiental y los costos económicos asociados (de Castro et al., 2018).

Una de las principales limitaciones en la aplicación de herbicidas es su baja eficiencia, ya que se estima que hasta un 90 % de los productos aplicados no alcanzan su objetivo (Kudsk y Streibig, 2003). Esto se debe, en gran medida, a la aplicación uniforme de los tratamientos en toda la parcela, sin considerar que solo áreas específicas pueden estar afectadas por malas hierbas. En este sentido, el uso de UAVs combinados con DL ha mostrado una gran precisión en la clasificación y mapeo de infestaciones de malas hierbas a nivel de píxel. Estos mapas, georreferenciados, permiten la elaboración de mapas de prescripción detallados, los cuales permiten ajustar la aplicación de herbicidas de acuerdo con la cobertura de malas hierbas, lo que resulta en ahorros de herbicidas de entre el 58 % y el 70 % (Huang et al., 2018a). La integración de estas tecnologías con equipos de dosificación variable y DSS proporciona un manejo más eficiente y sostenible de los insumos, minimizando el solapamiento en la aplicación de productos y mejorando tanto los resultados económicos como ambientales (Pérez-Ortiz et al., 2015).

Estudios recientes en cultivos como el maíz y la remolacha azucarera han demostrado que la aplicación de herbicidas basada en mapas generados por imágenes UAV permite una reducción significativa de la superficie tratada sin afectar el rendimiento del cultivo (Barreto et al., 2020). Las estrategias SSWM basadas en mapas de prescripción han logrado reducir el uso de herbicidas en hasta un 39 % en comparación con aplicaciones uniformes, postulándose como una solución eficaz para el control de malas hierbas en la agricultura moderna (Castaldi et al., 2017).

1.6 Estudios previos en malas hierbas

En el ámbito específico de la clasificación de malas hierbas mediante imágenes capturadas con UAV, investigaciones recientes han reportado un excelente rendimiento usando diversas arquitecturas CNN en campos de soja (dos Santos Ferreira et al., 2017), remolacha azucarera (Sa et al., 2018), espinaca y judías (Bah et al., 2018), trigo (Zhang et al., 2020a), arroz (Huang et al., 2020), guisantes y fresas (Khan et al., 2021), y cultivos mixtos de caña de azúcar, espinaca, plátano y pimientos (Ajayi y Ashi, 2023). Estos estudios han logrado una buena discriminación entre dos y cuatro clases generales, como cultivos, malas hierbas, suelo desnudo, y ocasionalmente, malas hierbas de hoja ancha y gramíneas. La eficacia de los clasificadores utilizados, como versiones de ResNet, SegNet, AlexNet, YOLO, VGGNet, GoogLeNet, LeNet, EfficientNet, se ha evaluado en una variedad de imágenes de UAV tomadas a altitudes de vuelo que varían entre 2 y 30 metros sobre el nivel del suelo, y en diferentes configuraciones de bases de datos (e.g. balanceadas versus desbalanceadas, variaciones en el número de etiquetas, etc). Sin embargo, pocos estudios se han centrado en la

clasificación de diversas especies de malas hierbas, un objetivo crucial para aplicar estrategias de SSWM (Wang et al., 2019a), y aquellos que lo han hecho aún están lejos de los avances alcanzados con imágenes terrestres (Chen et al., 2022; Espejo-Garcia et al., 2023; Olsen et al., 2019).

El uso de arquitecturas CNN para resolver la tarea de detección de objetos ha permitido el mapeo de cultivos y malas hierbas. Huang et al. (2018b) utilizaron redes totalmente convolucionales (FCN, del inglés *Fully Convolutional Network*) para mapear *Cyperus iria* L. y *Leptochloa chinensis* (L.) Ness en campos de arroz, generando mapas de cobertura y prescripción. de Camargo et al. (2021) optimizaron el modelo ResNet-18 para clasificar *Matricaria chamomilla* L., *Papaver rhoeas* L., *Veronica hederifolia* L. y *Viola arvensis* ssp. *arvensis* en cultivos de trigo de invierno, estableciendo una base para un mapeo rápido. Asimismo, Fraccaro et al. (2022) emplearon la arquitectura U-Net para detectar *Alopecurus myosuroides* Hudson en cultivos de trigo, mientras que Gallo et al. (2023) evaluaron el detector de objetos YOLOv7 en plantaciones de achicoria, demostrando su efectividad en la detección precisa de *Mercurialis annua* L.

La comparación entre CNNs y ViTs en la clasificación de malas hierbas es un área de investigación emergente, con literatura reciente que abarca tanto imágenes capturadas con UAV (Reedha et al., 2022) como con sensores próximos (Espejo-Garcia et al., 2023). El primer estudio evaluó el rendimiento de ViT-B32, ViT-B16, EfficientNet-B0, EfficientNet-B1 y ResNet50 en el reconocimiento de plantas de cultivo y malas hierbas (como categorías generales) en imágenes de UAV, concluyendo que ViT superaba ligeramente a las CNNs. El segundo estudio probó variantes de Swin y EfficientNet-v2 para clasificar nueve especies de malas hierbas australianas en la base de datos DeepWeeds reportando una exactitud top-1 del 98.61 % con un modelo SVM entrenado con características extraídas por Swin-v2 (Olsen et al., 2019).

dos Santos Ferreira et al. (2017) usaron la arquitectura CaffeNet, un derivado de AlexNet, para clasificar malas hierbas en cultivos, obteniendo una exactitud promedio de clasificación de 99.1 % y 99.5 % para bases de datos equilibradas y desbalanceadas, respectivamente. Bah et al. (2018) utilizaron la arquitectura ResNet18 para analizar una base de datos de imágenes capturadas por UAV, logrando un valor notable del área bajo la curva de 0.957. Huang et al. (2020) emplearon cuatro CNN preentrenadas (AlexNet, VGG16, GoogLeNet y ResNet-101) para identificar y clasificar malas hierbas en campos de arroz infestados principalmente por *L. chinensis*, *C. iria*, *Digitaria sanguinalis* (L.) Scop. y *Echinochloa crus-galli* (L.) Beauv., alcanzando exactitudes del 86.8 % al 88.4 %. Khan et al. (2021) desarrollaron un marco semisupervisado utilizando GANs para clasificar malas hierbas y cultivos, logrando una exactitud cercana al 90 %.

En cuanto a la identificación de especies de malas hierbas, Chen et al. (2022) emplearon aprendizaje por transferencia (TL, del inglés *Transfer Learning*) con 27 modelos de DL en una base de datos que incluía 15 especies distintas, destacando el modelo ResNet101 con una *F1-score* de 99.1 %. La identificación precisa de especies de malas hierbas es fundamental para mejorar las técnicas de control, como lo demuestra el trabajo de Valente et al. (2022), donde enfatizan la importancia de mapear con exactitud *Rumex obtusifolius* L. para implementar estrategias de control efectivas. Utilizando imágenes UAV, lograron una *F1-score* de 78.4 % con el modelo preentrenado MobileNet. Además, la integración de diversas arquitecturas de CNN ha permitido avances significativos en la resolución de varias tareas de visión artificial. Un ejemplo es el estudio de Shahi et al. (2023), quienes combinaron modelos de segmentación como SegNet, DeepLabV3 y UNet con arquitecturas de clasificación como VGG16, ResNet50, DenseNet121, EfficientNetB0 y MobileNetV2. El resultado

más destacado fue una *F1-score* de 88.2 %, lograda mediante la combinación de EfficientNetB0 y UNet. El estudio automatizó con éxito la segmentación semántica de malas hierbas como *Sorghum halepense* (L.) Pers., *Convolvulus arvensis* L. y *Portulaca oleracea* L. dentro de un cultivo de algodón.

1.7 Objetivos

El **objetivo general** de la tesis fue desarrollar un sistema automático para la identificación de especies de malas hierbas en estado temprano, utilizando imágenes capturadas por UAV y algoritmos de IA de última generación. Este sistema integra la agricultura de precisión, la IA y tecnologías emergentes como estrategia clave para optimizar la aplicación de herbicidas, reducir su uso y minimizar sus impactos negativos, fomentando prácticas agrícolas más sostenibles. Para lograr este objetivo general, se plantearon los siguientes **objetivos específicos**:

A. Revisar los avances más recientes en la identificación de especies de malas hierbas:

- **A1.** Proponer una taxonomía que organice los métodos publicados en la literatura para la identificación en estado temprano de desarrollo.
- **A2.** Crear bases de datos de referencia específicos y representativos.
- **A3.** Evaluar el rendimiento y la eficacia de los métodos con datos de obtenidos en diversos escenarios agrícolas reales.

B. Desarrollar una metodología para el procesamiento de grandes volúmenes de datos:

- **B1.** Derivar un método general que contemple la diversidad de imágenes RGB y la necesidad de algoritmos de alta eficiencia computacional.
- **B2.** Optimizar el procesamiento de datos mediante técnicas de paralelización y reducción de dimensionalidad.

C. Determinar y analizar los desafíos críticos en la identificación temprana de especies de malas hierbas:

- **C1.** Seleccionar y analizar las técnicas de DL más adecuadas para la identificación de especies de malas hierbas bajo diferentes condiciones ambientales.
- **C2.** Desarrollar una metodología robusta para la identificación temprana de especies.

D. Implementar algoritmos de IA de última generación:

- **D1.** Mejorar el rendimiento de los algoritmos seleccionados en comparación con los métodos tradicionales.
- **D2.** Explorar soluciones específicas para mejorar la exactitud de la identificación.

E. Desarrollar y validar métodos automatizados para la generación de mapas de malas hierbas georreferenciados, evaluando su eficacia en la gestión de cultivos y la implementación de técnicas SSWM.

1.8 Estructura de la tesis

La tesis doctoral está organizada en los siguientes capítulos:

- En el capítulo 2 se presenta la **Metodología General** implementada, describiendo los conceptos teóricos de las técnicas utilizadas en las distintas fases de la investigación realizada.
- El capítulo 3 incluye una revisión bibliográfica de los **Últimos avances en protección de cultivos de precisión**, particularmente la integración del ML y las tecnologías emergentes en esta disciplina. Esta revisión contextual constituye un valioso recurso para investigadores, profesionales y responsables políticos en el campo de la agricultura de precisión y esta recogida en el artículo:
    - Mesías-Ruiz G.A., Pérez-Ortiz M., Dorado J., de Castro A.I. and Peña J.M. (2023). Boosting precision crop protection towards agriculture 5.0 *via* machine learning and emerging technologies: A contextual review. *Frontiers in Plant Science*. 14:1143326.
- En el capítulo 4 se presenta la **Base de datos** creada y utilizada para el entrenamiento, validación y generalización de los modelos de DL implementados en esta tesis. Este *dataset* constituye un recurso de alta calidad para las tareas de visión artificial, tal como se detalla en el artículo.
    - Mesías-Ruiz G.A., Peña J.M., de Castro A.I., and Dorado J. (2024). DRONE imagery dataset for early-season WEED detection and classification in maize and tomato crops. *Data in Brief*. (Under review)
- En el capítulo 5 se realiza una evaluación de las fases de entrenamiento e inferencia durante la **Clasificación de malas hierbas mediante imágenes UAV y modelos CNN estándar**, donde se explora el tamaño de la base de datos de entrenamiento necesario para obtener métricas de clasificación superiores al 80 %. Además, se analizó la relación óptima entre las especies minoritarias y mayoritarias en bases de datos desbalanceadas en la etapa de inferencia. Este trabajo se ha publicado en:
    - Mesías-Ruiz G.A., Borra-Serrano I., Peña J.M., de Castro A.I., Fernández-Quintanilla C., & Dorado J. (2024). Weed species classification with UAV imagery and standard CNN models: Assessing the frontiers of training and inference phases. *Crop Protection*, 182: 106721.
- En el capítulo 6 se realiza la **Comparación de detectores multiobjeto basados en arquitecturas CNN y ViT**, evaluando además su capacidad para la creación de mapas de prescripción. Este trabajo forma parte del artículo:
    - Mesías-Ruiz G.A., Dorado J., de Castro A.I., Borra-Serrano I., & Peña J.M. (2024). Classification, detection and mapping of early-season weed species in UAV images with cutting-edge CNN and Vision Transformer models. *Computers and Electronics in Agriculture*, (Under Review).
- En el capítulo 7 se describen **Avances en computación cognitiva dirigida al monitoreo de malas hierbas mediante imágenes UAV**, en particular para la clasificación de especies arvenses en estados fenológicos diferentes. Este trabajo utiliza el modelo de IA generativa

GFP-GAN para mejorar e incrementar la resolución espacial de la base de datos, el cual ha sido publicado en el artículo:

  - Mesías-Ruiz G.A., Peña J.M., de Castro A.I., Borra-Serrano I., & Dorado J. (2024). Cognitive Computing Advancements: Improving Precision Crop Protection through UAV Imagery for Targeted Weed Monitoring. *Remote Sensing*, 16, 3026.

- En el capítulo 8 se presenta la **Discusión General** de los trabajos descritos anteriormente.
- En el capítulo 9, se resumen las **Conclusiones Generales** así como las líneas de investigación futuras que se derivan tras la realización de esta tesis.
- Finalmente se exponen la bibliografía utilizada durante la realización del presente documento.

1.9 Contribuciones al avance científico

Los siguientes artículos publicados en revistas internacionales recopilan y expanden algunas de las ideas desarrolladas en esta investigación:

A1 Mesías-Ruiz G.A., Pérez-Ortiz M., Dorado J., de Castro A.I. & Peña J.M. Boosting precision crop protection towards agriculture 5.0 via machine learning and emerging technologies: A contextual review. *Front. Plant Sci*, 14:1143326, doi: 10.3389/fpls.2023.1143326, 2023, Factor de Impacto (2023): 4.6 (Q1).

A2 Mesías-Ruiz G.A., Borra-Serrano I., Peña J.M., de Castro A.I., Fernández-Quintanilla C., & Dorado J. Weed species classification with UAV imagery and standard CNN models: Assessing the frontiers of training and inference phases. *Crop Protection*, 182: 106721, https://doi.org/10.1016/j.cropro.2024.106721, 2024, Factor de Impacto (2023): 2.5 (Q1).

A3 Mesías-Ruiz G.A., Peña J.M., de Castro A.I., Borra-Serrano I., & Dorado J. Cognitive Computing Advancements: Improving Precision Crop Protection through UAV Imagery for Targeted Weed Monitoring. *Remote Sens.*, 16, 3026, https://doi.org/10.3390/rs16163026, 2024, Factor de Impacto (2023): 4.2 (Q1)

Actualmente se encuentran en proceso de revisión los siguientes artículos:

A4 Mesías-Ruiz G.A., Dorado J., de Castro A.I., Borra-Serrano I., & Peña J.M. Classification, detection and mapping of early-season weed species in UAV images with cutting-edge CNN and Vision Transformer models. *Computers and Electronics in Agriculture*, (Under Review), 2024, Factor de Impacto (2023): 7.7 (Q1).

A5 Mesías-Ruiz G.A., Peña J.M., de Castro A.I., and Dorado J. Drone imagery dataset for early-season weed detection and classification in maize and tomato crops. *Data in Brief*, (Under review), 2024, Factor de Impacto (2023): 0.21 (Q3)

Las bases de datos creadas y utilizadas en las publicaciones derivadas de esta investigación se encuentran publicadas bajo la modalidad de acceso abierto en:

D1 Mesías-Ruiz G.A., Borra-Serrano I., Peña J.M., de Castro A.I., Fernández-Quintanilla C., & Dorado J. Unmanned Aerial Vehicle Imagery for Early Stage Weed Classification and

Detection in Maize and Tomato Crops [Data set]. DIGITAL.CSIC, 2024, http://doi.org/10.20350/DIGITALCSIC/16131.

D2 Mesías-Ruiz G.A., Peña J.M., de Castro A.I., & Dorado J. (2024). DRONEWEED: DRONE imagery dataset for early-season WEED classification [Data set]. DIGITAL.CSIC. http://doi.org/10.20350/DIGITALCSIC/16559

Además, también se han publicado algunos trabajos en congresos nacionales e internacionales:

C1 Mesías-Ruiz G.A., Peña J.M., de Castro A.I., Borra-Serrano I., & Dorado J. Detección y clasificación de malas hierbas mediante drones y redes neuronales profundas: creación de mapas para tratamiento localizado. En *IV Encontro Nacional de Herbologia / XIX Congreso de la Sociedad Española de Malherbología (SEMh) 2024* . ISBN 978-84-09-40561-9, Pp. 175–179, 2024. Beja, Portugal.

C2 de Castro A.I., Mesías-Ruiz G.A., Llenes J.M., Borra-Serrano I., Rueda-Ayala C., Dorado J., Recasens J., & Peña J.M. DIGINVASIVE: Sistema de alerta y control de especies invasoras. Caso de estudio: *Amaranthus palmeri*. En *IV Encontro Nacional de Herbologia / XIX Congreso de la Sociedad Española de Malherbología (SEMh) 2024*. ISBN 978-84-09-40561-9, Pp. 195–200. 2024, Beja, Portugal.

C3 Mesías-Ruiz G.A., Dorado J., de Castro A.I. & Peña J.M. Cognitive computing for classification of six weed species in tomato and maize crops. En *14^{th} European Conference on Precision Agriculture* (ECPA). ISBN 978-90-8686-114-9, Pp. 209–210. 2023, Bolonia, Italia.

C4 de Castro A.I., Mesías-Ruiz G.A., Llenes-Espigares J.M., Dorado J., Recasens J. & Peña J.M. DIGINVASIVE: A digital system to map the presence of invasive weed plants. *14^{th} European Conference on Precision Agriculture* (ECPA). ISBN 978-90-8686-114-9, Pp. 129–130, 2023, Bolonia, Italia.

C5 Belissent N., Peña J.M., Mesías-Ruiz G.A., Shawe-Taylor J., & Pérez-Ortiz M. Transfer and zero-shot learning for weed species detection with small datasets and unseen classes. *14^{th} European Conference on Precision Agriculture* (ECPA). ISBN 978-90-8686-114-9, Pp. 213–214. 2023, Bolonia, Italia.

C6 Mesías-Ruiz G.A., Dorado J., de Castro A.I., Martín J.M., Campos D., Fernández-Quintanilla C., & Peña J.M. Especificaciones óptimas para la identificación de especies arvenses en maíz mediante imágenes tomadas con drones y aprendizaje automático. En *XVIII Congreso de la Sociedad Española de Malherbología (SEMh) 2022*. ISBN 978-84-09-40561-9, Pp. 421–426. 2022, Mérida-España.

C7 Peña J.M., de Castro A.I., Mesías-Ruiz G.A., Fernández-Quintanilla C, & Dorado J. 10 years of UAV technology for weed mapping: overview and future trends for site-specific weed management. En *19^{th} European Weed Research Society Symposium* (EWRS), Pp. 77. 2022, Atenas, Grecia.

C8 Castro S., Iñacasha J., Mesías-Ruiz G.A., & Oñate W. Prototype based on a LoraWAN network for storing multivariable data, oriented to agriculture with limited resources. En *Proceedings of Seventh International Congress on Information and Communication Technology* (ICICT). doi.org/10.1007/978-981-19-1610-6, Pp. 245-255. 2022, Londres, Inglaterra.

C9 de Castro A.I., Mesías-Ruiz G.A., Dorado J., & Peña J.M. Tecnologías geo-espaciales y digitalización en sanidad vegetal. En *La sanidad vegetal en cultivos mediterráneos y subtropicales. Retos ante una transición agroecológica*. 2021, Tenerife, España.

Otras publicaciones realizadas durante el doctorado:

- Belissent N., Peña J.M., Mesías-Ruiz G.A., Shawe-Taylor J. & Pérez-Ortiz M. Transfer and Zero-Shot Learning for Scalable Weed Detection and Classification in UAV Images. *Knowledge-Based Systems*, 292: 111586, https://doi.org/10.1016/j.knosys.2024.111586, 2024, Factor de Impacto (2023): 7.2 Q1.

Durante la realización de la tesis se realizó una estancia predoctoral en el extranjero:

- **Grupo de Investigación en Electrónica y Telemática (GIETEC), Universidad Politécnica Salesiana del Ecuador, Quito**. Con una duración de tres meses comprendidos entre noviembre de 2021 a febrero de 2022. Tutor de la estancia: **Dr. German Arévalo Bermeo**, Director de la carrera de Ingeniería en Telecomunicaciones. Dentro de las actividades realizadas constan la utilización de equipos de medición de radio frecuencia, para antenas de transmisión y recepción de 2,4Ghz y 5GHz. Capacitación sobre el equipo de radio por Software (NI USRP-2944, National Instruments Corporation, Austin, Texas, EE. UU.), e introducción a la comunicación vehículo a vehículo (V2V). Las actividades desarrolladas permitieron definir la plataforma tecnológica que permita obtener menor perdida en la transmisión de una señal de video en tiempo real. Así como, la caracterización de la calidad de video cuando el transmisor se encuentra en movimiento y el receptor permanece estático. Estos estudios y análisis dieron lugar a las propuestas tecnológicas que se realizaron en el apartado *Telecommunications for precision crop protection* del artículo **A1** y congreso **C8**.

Capítulo 2

Metodología General

2.1 Revisión bibliográfica

El primer paso llevado a cabo en esta tesis doctoral fue una revisión contextual sobre el papel de la IA, particularmente ML, y otras tecnologías emergentes, para resolver los retos actuales y futuros de la protección de cultivos (capítulo 3). Esta revisión bibliográfica incluyó un análisis bibliométrico, un enfoque de investigación que utiliza técnicas cuantitativas para examinar datos bibliográficos, como publicaciones científicas y citas, con el objetivo de explorar y analizar grandes volúmenes de información científica. Esta metodología permite a los investigadores identificar las tendencias evolutivas de un campo específico y destacar áreas emergentes dentro del mismo (Donthu et al., 2021).

La metodología que se empleó para realizar el análisis bibliométrico consistió en:

a) Definir los objetivos:

   · Identificar los algoritmos de ML más utilizados y sus contribuciones en las disciplinas del conocimiento: Se realizó un análisis exhaustivo de la base de datos de Scopus[a] para identificar la producción científica basada en los algoritmos de ML comprendida entre 2010 y 2022.

   · Analizar el impacto científico de ML en la protección de cultivos, particularmente los algoritmos tradicionales de ML, modelos de ANNs y su aplicación en la detección y control de malas hierbas, enfermedades y plagas.

   · Explorar el desarrollo temporal y las tendencias de publicación de algoritmos de ML en el contexto de la protección de cultivos.

b) Analizar el alcance:

   · Identificación de los algoritmos de ML aplicados en diversas disciplinas del conocimiento que presentan un enfoque específico en su implementación para la protección de cultivos.

   · Aplicación en los tres pilares de la protección de cultivos: las malas hierbas, enfermedades y plagas de cultivos. Para cada pilar, se discuten aplicaciones relevantes de algoritmos de ML y modelos de ANNs, lo que proporciona una visión detallada del estado del arte y los desafíos específicos.

   · Revisión de tecnologías emergentes aplicadas a la protección de cultivos en el contexto de la Agricultura 5.0. Particularmente las que están diseñadas para mejorar la capacidad de detección temprana de malas hierbas, enfermedades y plagas, dirigidas hacia una toma de decisiones autónoma y en tiempo real.

c) Seleccionar las métricas para caracterizar la distribución temporal del número de publicaciones:

   · Tasa de Variación Relativa (RVR, del inglés *Relative Variation Rate*): Proporciona una medida de la intensidad de la fluctuación anual en cada categoría de ML, lo que refleja la variabilidad en el crecimiento anual del número de documentos publicados (Ec. 2.1).

   · Tasa de Crecimiento Anual Compuesto (CAGR, del inglés *Compound Annual Growth*

[a] https://www.scopus.com/

Rate): Cuantifica el aumento porcentual acumulado del número de documentos en cada categoría de ML durante el período de estudio (2010-2022), proporcionando una visión general del crecimiento sostenido a lo largo del tiempo (Ec. 2.2).

$$RVR = \frac{x_t}{x_{t-1} - 1} \tag{2.1}$$

$$CAGR = \left(\frac{x_t}{x_0}\right)^{\frac{1}{n}} - 1 \tag{2.2}$$

Donde: x_t : número de artículos en el último período comparado
x_{t-1} : número de artículos en el período anterior al comparado
x_0 : número de artículos en el período inicial
n : número de períodos comparados

d) Recopilar datos:

· Definición de los términos de búsqueda, aplicando el criterio de búsqueda en el título, resumen y palabras clave de los artículos, y utilizando operadores booleanos para refinar los resultados.

· Uso combinado de los términos "machine learning" AND "category", donde "category" se refiere a las tareas específicas que abordan los algoritmos de ML, como la clasificación, la regresión, entre otras, con el fin de evaluar el impacto del ML en la producción científica.

· Uso combinado de los términos "precision agriculture" OR "precision farming" AND "algorithm" AND "crop protection". Donde "algorithm" hace referencia a los algoritmos identificados y "crop protection" abarca los tres dominios clave de la protección de cultivos, con el objetivo de determinar el impacto científico de los algoritmos de ML en este campo.

e) Ejecutar el análisis y presentar resultados:

· El análisis bibliométrico se realizó utilizando herramientas y software especializados, como VOSviewer (Visualizing Scientific Landscapes, Leiden University, The Netherlands), para generar visualizaciones de redes y crear mapas de co-ocurrencia de palabras clave, y OriginPro (OriginLab Corporation, Northampton, MA, USA), para el análisis gráfico de tendencias en la producción científica.

2.2 Plan de trabajo

2.2.1 Esquema general

El plan de trabajo de esta tesis doctoral contempló el desarrollo de modelos de IA y DL para optimizar el análisis de imágenes capturadas por UAV, con el fin de generar mapas automatizados de especies de malas hierbas. Este desarrollo seguió un esquema general (Figura 2.1) riguroso y sistemático, lo cual garantizó la precisión y efectividad de los modelos (Coulibaly et al., 2022; Sharma et al., 2020). El esquema incluyó una serie de fases críticas, cada una de las cuales resultó esencial para alcanzar los objetivos de la tesis.

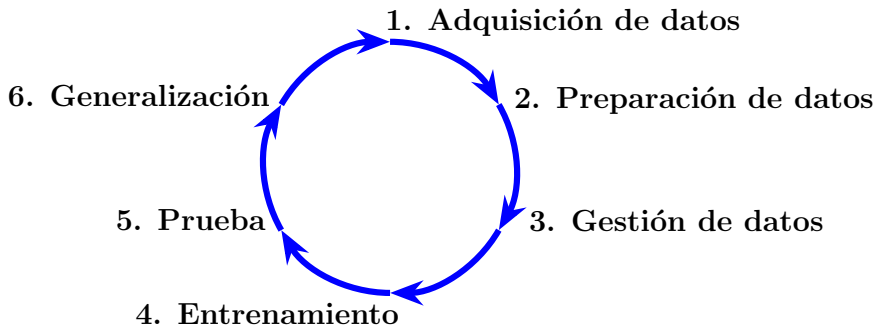


Figura 2.1: Esquema general del plan de trabajo de la tesis.

La primera fase del plan de trabajo consistió en la adquisición y exploración de datos. En esta etapa, se recopilaron los datos relevantes y se aseguró que estuvieran en el formato adecuado para el análisis, con la colaboración de expertos en la identificación de especies de malas hierbas. A continuación, se procedió a la preparación de los datos, que implicó el procesamiento necesario para el entrenamiento del modelo. Esta fase incluyó la exploración de las características, la evaluación de la calidad y el formato de los datos, así como su preprocesamiento para el análisis posterior.

A continuación, se llevó a cabo la gestión de datos, un proceso que implicó la limpieza y transformación de los datos brutos para eliminar inconsistencias como valores perdidos, duplicados, datos inválidos y ruido, con el fin de garantizar la fiabilidad del modelo.

La modelización de datos constituyó el núcleo del plan de trabajo, orientándose hacia la clasificación y detección de objetos mediante la aplicación de técnicas avanzadas de DL. Esta fase se dividió en dos etapas fundamentales: la fase de entrenamiento, en la cual el modelo fue entrenado para identificar patrones y características relevantes, y la fase de prueba, donde se evaluó la exactitud del modelo utilizando una base de datos no utilizada previamente.

Finalmente, se abordó la interpretación y comunicación de resultados, una tarea crucial en los modelos de IA. La capacidad de explicar los resultados de manera generalizable es esencial para mantener el rigor científico y demostrar el valor del modelo. El plan de trabajo culminó con la generalización, documentación y mantenimiento del modelo, garantizando su operatividad y su capacidad de evolución en función de las necesidades del sector de aplicación.

2.2.2 Conocimientos fundamentales y competencias para el desarrollo de modelos de Inteligencia Artificial

Para el desarrollo de esta tesis doctoral, se ha requerido el dominio de una serie de conocimientos y habilidades fundamentales, que constituyen los requisitos esenciales para construir un modelo agrícola basado en DL. Estos conocimientos han sido integrados de manera sistemática a lo largo del proceso de investigación, lo que ha permitido una implementación rigurosa y precisa de los modelos propuestos.

En primer lugar, ha sido indispensable contar con un sólido conocimiento en ciencias de la computación, lo que incluye una comprensión profunda de los principios fundamentales de las tecnologías de la información y la computación. Esta base ha sido crucial para el manejo adecuado de datos y la implementación eficiente de algoritmos de DL.

El álgebra lineal ha sido una competencia fundamental adquirida durante este proceso, dado que las operaciones con vectores, matrices y tensores son cruciales para la manipulación de grandes bases de datos y la formulación de modelos matemáticos complejos. Del mismo modo, la estadística y la probabilidad han desempeñado un papel central en el análisis de datos y en la toma de decisiones basada en modelos probabilísticos, permitiendo una evaluación robusta de la incertidumbre inherente en la identificación automática de especies de malas hierbas.

Además, se ha profundizado en el cálculo diferencial e integral, conocimientos esenciales para comprender y aplicar los métodos de optimización que sustentan los algoritmos de entrenamiento de modelos de DL. La competencia en programación ha sido otro pilar fundamental, específicamente en lenguajes como Python y C++, los cuales se han utilizado para desarrollar e implementar los algoritmos de DL.

Por último, se ha adquirido una experiencia avanzada en la gestión de datos y hardware, necesaria para manejar y procesar grandes volúmenes de datos agrícolas, así como para optimizar los recursos computacionales involucrados en la ejecución de los modelos. Este conocimiento ha asegurado que el modelo desarrollado tenga una buena exactitud, sea escalable y eficiente en su aplicación práctica.

A lo largo del desarrollo de esta tesis, se ha alcanzado un alto nivel de competencia en cada una de estas áreas, lo que ha permitido llevar a cabo una investigación exhaustiva y rigurosa. Este conjunto de conocimientos ha sido indispensable para enfrentar y superar los desafíos técnicos y científicos, resultando en la creación de varios modelos basados en DL que cumplen con estándares académicos y de investigación.

2.3 Base de datos

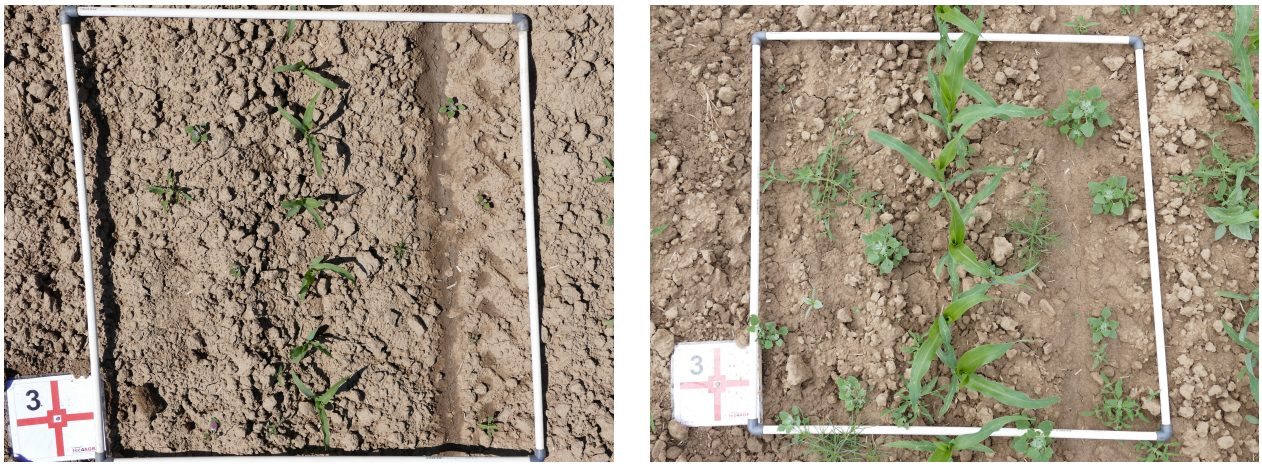
2.3.1 Zona de estudio

Los estudios realizados en esta tesis se llevaron a cabo en diversas ubicaciones geográficas de España, abarcando diferentes tipos de cultivos y condiciones agrícolas. Los cultivos y zonas de estudio fueron:

a) **Maíz:** en la Finca Experimental La Poveda (Instituto de Ciencias Agrarias, Agencia Estatal Consejo Superior de Investigaciones Científicas, ICA-CSIC), ubicada en Arganda del Rey, Madrid, España (40°18'59,25"N, 3°29'21,53"W). La parcela de estudio tenía una superficie de aproximadamente 7.400 m^2. Los vuelos de UAV se realizaron los días 18 y 27 de mayo de 2020, correspondientes a 44 y 52 días después de la siembra (DAS), coincidiendo con el estado fenológico del cultivo BBCH14 (4 hojas desplegadas) y BBCH17 (4 hojas desplegadas), respectivamente, según la escala BBCH (Meier, 2018).

b) **Tomate:** en dos fincas comerciales ubicadas en Santa Amalia, Badajoz, España (38°59'15,58"N, 6°02'57,71"W y 38°59'40,19"N, 5°57'17,54"W). Las parcelas estudiadas tenían una superficie de 12.000 m^2 y 14.000 m^2 , encontrándose en el estado fenológico BBCH501 (primer botón floral visible) y BBCH509 (noveno botón floral visible), respectivamente, los días 1 y 2 de junio de 2021, coincidiendo con las fechas en que se realizaron los vuelos de UAV.

En cada parcela de estudio, se instalaron 30 marcos georreferenciados de 1 m $\times$ 1 m distribuidos aleatoriamente. En cada uno de estos marcos (unidades de verdad-terreno), se capturó manualmente una imagen utilizando una cámara digital LEICA V-LUX Typ 114 (Leica Camera AG, Wetzlar, Alemania). Estas unidades de verdad-terreno se emplearon para la validación *in situ* de las especies de malas hierbas (Figura 2.2), ya que permitían observar las infestaciones naturales en los campos de estudio en el mismo momento en que se realizaron los vuelos con UAV.



(a) **(b)**

Figura 2.2: Marco de muestreo utilizado como verdad-terreno del cultivo de maíz en los estados fenológicos: (a) BBCH14 y (b) BBCH17.

2.3.2 Adquisición de imágenes

La adquisición de imágenes se realizó con una cámara RGB de bajo coste, modelo Sony ILCE-6300L (Sony Group Corporation, Tokio, Japón), montada en un UAV cuadricóptero modelo md4-1000 (microdrones GmbH, Siegen, Alemania) (Figura 2.3). La cámara cuenta con un sensor CMOS Exmor® de tipo APS-C (23,5 × 15,6 mm) con 24,2 megapíxeles efectivos. La configuración del plan de vuelo del UAV se diseñó utilizando la aplicación mdCockpit, estableciendo una velocidad horizontal constante de 2 m/s a una altitud fija de 11 m sobre el nivel del suelo. Esta configuración resultó en un tamaño de pixel en el terreno (GSD, del inglés: *Ground Sample Distance*) de 0,17 cm por píxel. Las imágenes fueron capturadas con un solapamiento del 70 % tanto lateral como frontal, y con una resolución de 6000 × 3376 píxeles. El número total de imágenes obtenidas en cada parcela durante los vuelos se muestra en la Tabla 2.1:

Figura 2.3: Vuelo del UAV md4-1000 equipado con un sensor RGB Sony ILCE-6300L sobre un cultivo de tomate en estado fenológico BBCH50.

Tabla 2.1: Número de imágenes obtenidas para la creación de la base de datos

Cultivo	Estado fenológico	Imágenes
Maíz	BBCH14	568
Maíz	BBCH17	565
Tomate	BBCH501	895
Tomate	BBCH509	950

2.3.3 Procesamiento de imágenes

Ortomosaico

A partir de las imágenes capturadas por el UAV en una parcela determinada, se generó un ortomosaico, que consiste en una imagen única, uniforme y georreferenciada con alta precisión. Este proceso se llevó a cabo utilizando el software especializado Agisoft PhotoScan (Agisoft LLC, San Petersburgo, Rusia), mediante las fases principales de alineación de imágenes y construcción de la geometría del terreno. Además, se realizó la georreferenciación manual de las imágenes empleando las coordenadas de varios puntos de control terrestre, estratégicamente distribuidos en los campos de estudio y obtenidos con un receptor GNSS Trimble® R2 (Trimble Inc, Westminster, EE. UU.). En la Tabla 2.2 se detallan las propiedades de los ortmosaicos generados.

Tabla 2.2: Propiedades los ortomosaicos obtenidos

Cultivo	Estado fenológico	Dimensiones (píxeles)	Tamaño en disco (GB)
Maíz	BBCH14	52,122 × 57,404	5.3
Maíz	BBCH17	63,627 × 67,170	7.8
Tomate	BBCH501	72,304 × 34,574	6.3
Tomate	BBCH509	70,632 × 67,204	9.8

Partición

Cada ortomosaico fue dividido en secciones más pequeñas mediante un proceso de partición de imágenes, lo que permitió su análisis individual. Esta estrategia facilitó el procesamiento de las imágenes y redujo el coste computacional del análisis. Dado que los ortomosaicos de los campos de maíz y tomate eran de gran tamaño, se particionaron automáticamente en imágenes de 1000 × 1000 píxeles utilizando un programa desarrollado específicamente para esta tarea (Algorithm 2.1), el cual fue implementado en Python.

Algorithm 2.1 Orthomosaic partitioning

Input: Orthomosaic directory ($orthomosaic\_directory$), Output directory ($output\_directory$), Partition size (m_size)
Output: Partitioned images
function GETCOORDINATESTOPLEFT($orthomosaic\_directory$)
    open the orthomosaic using Rasterio
    $transf \leftarrow$ dataset transformation
    calculate the coordinates of the upper left-hand corner
    return coordinates of the upper left corner
end function
function PARTITIONORTOMOSAIC($orthomosaic\_directory$, $output\_directory$, m_size)
    create an empty list for storing the partition data
    $width$, $height \leftarrow$ width and height of the dataset
    $coordinates(i, j) \leftarrow$ GETCOORDINATESTOPLEFT($orthomosaic\_directory$)

Algorithm 2.1 Orthomosaic partitioning (Part 2)

```
    for each i from 0 to width with a step of m_size do
        for each j from 0 to height with a step of m_size do
            create a window for the partition[coordinates]
            read partition using Rasterio
            obtain ← partition[coordinates]
            create a PIL image from the partition
            save the image in the output_directory with a name based on the coordinates.
        end for
    end for
end function
```

Etiquetado

Con la ayuda de expertos en identificación de malas hierbas se etiquetaron las diferentes especies sobre las imágenes particionadas, marcando de forma manual recuadros delimitadores con la herramienta gráfica de software libre labelImg (Tzutalin, 2015). Los archivos de las anotaciones realizadas sobre las imágenes particionadas se guardan con extensión XML en formato PASCAL VOC. En este proceso se identificó el número total de individuos (e.g. etiquetas) de cada especie de mala hierba que se detalla en la Tabla 2.3.

Tabla 2.3: Número de etiquetadas para cada especie de mala hierba y cultivo, por cada etapa fenológica.

	Etapa de crecimiento temprano	**Etapa de cremiento posterior**
	MAIZ (BBCH14)	**MAIZ (BBCH17)**
Atriplex patula	1.000	1.459
Chenopodium album	1.200	2.175
Convolvulus arvensis	1.200	1.102
Datura ferox	683	589
Lolium rigidum	1.000	80
Salsola kali	1.200	1.216
Sorghum halepense	1.600	103
Maíz	12.364	24.614
	TOMATE (BBCH501)	**TOMATE (BBCH509)**
Cyperus rotundus	3.090	134
Portulaca oleracea	1.875	177
Solanum nigrum	1.900	2.175
Tomate	3.890	2.732
Total etiquetas	31.002	36.556

2.4 Clasificación en imágenes

2.4.1 Redes Neuronales Convolucionales

Las CNN son una clase especializada de redes neuronales profundas diseñadas para procesar datos que tienen una estructura en forma de cuadrícula, como las imágenes. Su arquitectura está inspirada en la organización de la corteza visual de los animales (Eickenberg et al., 2017), donde neuronas individuales responden a estímulos en regiones específicas del campo visual.

Funcionamiento

El funcionamiento de una CNN se basa en la aplicación de convoluciones sobre la base de datos de entrada. Una convolución es una operación matemática que combina dos conjuntos de información: los datos de entrada (imagen, I) y un filtro (o kernel, K)(Ec. 2.3). En el contexto de las CNNs, este proceso permite a la red extraer características de los datos de entrada que son relevantes para la tarea de clasificación en imágenes, como bordes, texturas o formas de objetos.

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n) \tag{2.3}$$

El flujo de trabajo de una CNN involucra varias capas secuenciales (Figura 2.4):

a) Capa de convolución: Es el componente central de la CNN. Esta capa aplica un conjunto de filtros a la entrada, ejecutando una operación de convolución para generar los mapas de activación. Cada filtro está diseñado para detectar características específicas de la imagen, como bordes horizontales o verticales.

b) Capa de activación: Tras la convolución, se aplica una función de activación, comúnmente la ReLU (*Rectified Linear Unit*), que introduce no linealidades en la red, permitiendo que aprenda patrones más complejos.

c) Capa de *pooling*: También conocida como submuestreo o *downsampling*, esta capa reduce las dimensiones espaciales (ancho y altura) de los mapas de activación, conservando las características más relevantes y reduciendo la carga computacional. La operación de *pooling* más común es el *max pooling*.

d) Capas completamente conectadas (*Fully Connected Layers*): Al final de la CNN, se encuentran capas completamente conectadas que convierten los mapas de activación en vectores de características y realizan la clasificación en función de las características extraídas.

Características

a) Localidad espacial y conectividad parcial: Las CNN explotan la estructura local de los datos al conectar cada neurona de una capa solo a una pequeña región de la capa anterior (campo receptivo). Esto permite que la red aprenda características locales, como bordes, que son esenciales para interpretar el contenido de una imagen.

b) Compartición de parámetros: Los filtros en las capas de convolución se comparten a lo largo de toda la imagen, lo que permite a la red detectar las mismas características en diferentes

posiciones. Esto reduce significativamente el número de parámetros en comparación con una red completamente conectada.

c) Invariancia a la traducción: Gracias a la estructura de convolución y *pooling*, las CNNs pueden identificar características clave en distintas partes de la imagen, lo que las hace robustas frente a pequeñas variaciones o traslaciones en la entrada.

d) Profundidad y jerarquía de características: Al agregar más capas de convolución y *pooling*, las CNN pueden aprender características jerárquicas, desde bordes simples en las capas iniciales hasta representaciones más abstractas en capas más profundas.

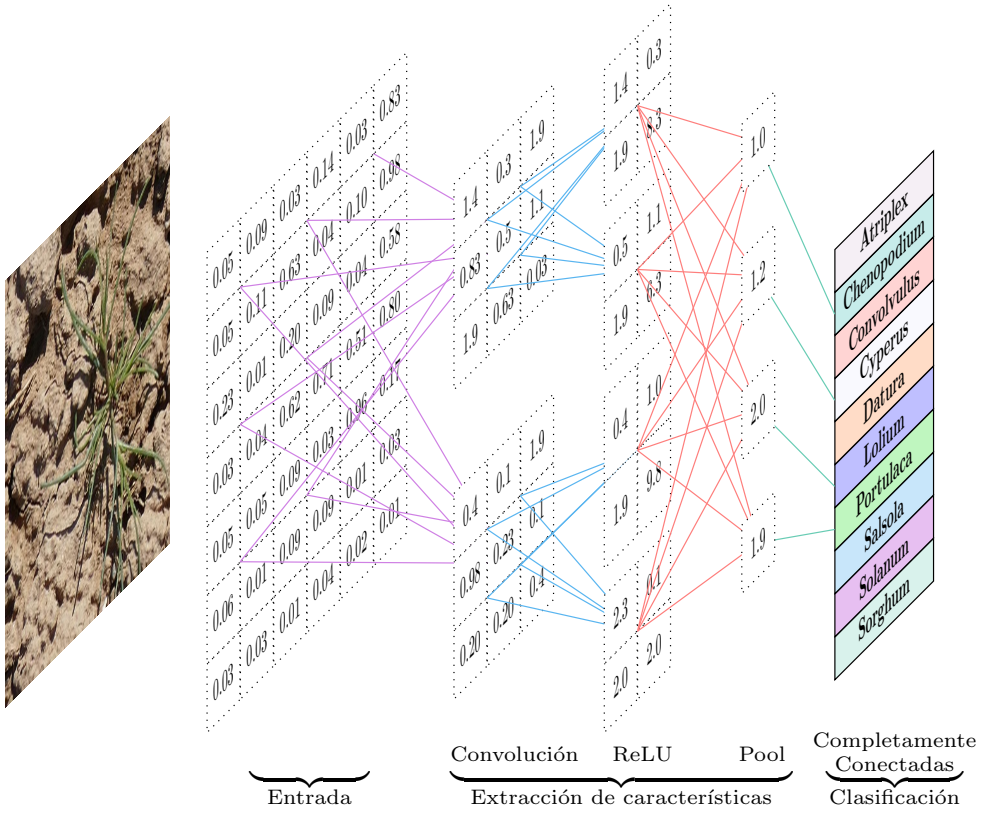


Figura 2.4: Arquitectura general de alto nivel de la CNN.

2.4.2 Visual Transformers

Los ViT representan un avance innovador en el campo de la visión artificial, ya que adaptan la arquitectura *Transformer* (Vaswani et al., 2017), originalmente diseñada para el procesamiento natural del lenguaje (NLP, del inglés *Natural Language Processing*), para procesar datos visuales. El notable éxito de los *Transformers* en NLP, especialmente en tareas que requieren modelar dependencias de largo alcance y comprensión contextual, ha inspirado su aplicación en visión artificial, donde están mostrando un gran potencial para superar a las CNN tradicionales en varios puntos de referencia.

Funcionamiento

Los ViTs funcionan descomponiendo una imagen en fragmentos más pequeños, que luego se procesan de forma similar a los tokens de palabras en las tareas de NLP. Su funcionamiento sigue los siguiente pasos:

a) Incorporación de parches de imagen: La imagen de entrada se divide en parches de tamaño fijo, cada uno de los cuales se aplana en un vector. Estos vectores se proyectan linealmente en un espacio de dimensiones superiores, creando lo que se conoce como una "incrustación de parches". Cada parche, ahora representado como un vector, actúa como un token, análogo a una palabra en NLP.

b) Codificación posicional: A diferencia de las palabras en una frase, los parches de imagen no tienen un orden inherente. Para proporcionar al *Transformer* información espacial, se añaden codificaciones posicionales a cada incrustación de parche, lo que permite al modelo conservar la estructura espacial de la imagen.

c) Codificador del *Transformer*: La secuencia de incrustaciones de parches, complementada con sus codificaciones posicionales, se introduce en un codificador *Transformer*. El codificador utiliza mecanismos de autoatención multicabezal (MHSA, del inglés *Multi-Head Self-Attention*) para capturar las relaciones entre las distintas partes de la imagen, considerando tanto el contexto local y como global. Esto es crucial para las tareas visuales donde es esencial entender las interacciones entre las distintas partes de la imagen.

d) Token de clasificación: En las tareas de clasificación, se añade a la secuencia de parches un token especial de clasificación que se puede aprender. Tras el procesamiento a través de las capas del *Transformer*, este token sintetiza la información de todos los parches y se utiliza para predecir la etiqueta de clase final.

Características

a) Mecanismo de autoatención: El núcleo de los ViTs es el mecanismo de autoatención, que permite al modelo centrarse en diferentes partes de la imagen con intensidad variable según su relevancia para la tarea. A diferencia de las CNNs, que dependen de los campos receptivos locales y de las operaciones de convolución, los *Transformers* pueden modelar directamente las dependencias de largo alcance, lo que les permite captar el contexto global de forma eficiente.

b) Codificación posicional: Para manejar la falta de orden inherente a los parches de imagen, las codificaciones posicionales son fundamentales en los ViT. Estas codificaciones permiten al modelo comprender la disposición espacial de los parches, lo que es vital para tareas como la detección de objetos.

c) Flexibilidad en los tamaños de entrada: Los ViT son menos sensibles al tamaño y la resolución de entrada en comparación con las CNNs. Debido a su mecanismo de atención no local, los ViTs pueden procesar imágenes de distintos tamaños sin necesidad de cambios arquitectónicos significativos, lo que los hace muy versátiles.

d) Escalabilidad: Los ViT pueden escalar en función de la disponibilidad de datos. Si bien

los primeros experimentos mostraron que necesitaban grandes bases de datos para obtener buenos resultados, modelos más recientes como el *Transformer* de imágenes eficiente en datos (DeiT, del inglés *Data-efficient image Transformer* han demostrado que, con estrategias de entrenamiento adecuadas, los ViT pueden lograr resultados competitivos incluso con bases de datos limitadas.

e) Modelos jerárquicos e híbridos: Los avances recientes en los ViT incluyen estructuras jerárquicas y modelos híbridos que combinan las fortalezas de las CNN y los *Transformers*. Estos modelos aprovechan las operaciones convolucionales para la extracción temprana de características, seguidas de capas de *Transformers* para el modelado del contexto global, lo que mejora el rendimiento en diversas tareas de visión artificial.

f) Desafíos y compensaciones: A pesar de sus ventajas, los ViT enfrentan desafíos como mayores costes computacionales y una tendencia al sobreajustarse cuando se trabaja con datos limitados. Se están explorando diversas estrategias para mitigar estos problemas, como la incorporación de sesgos inductivos convolucionales o el uso de la tokenización jerárquica.

2.4.3 Redes Generativas Adversarias

Las GAN son un tipo de modelo de DL que se entrenan utilizando un enfoque basado en la teoría de juegos (Cao et al., 2019). En los últimos años, han tenido un impacto significativo en el campo de la visión artificial, logrando grandes avances en desafíos como la generación de imágenes realistas.

Funcionamiento

El modelo GAN (Goodfellow et al., 2014) se compone de dos redes neuronales llamadas Generador (G) y Discriminador (D), que se entrenan mediante un proceso adversarial (Figura 2.5). La entrada para G es un vector de ruido aleatorio (z), y produce datos sintéticos $G(z)$. Por otro lado, el D recibe tanto la muestra de datos generada como una muestra de datos reales (x) del conjunto de entrenamiento. El D discrimina entre las muestras reales y las generadas. Durante el entrenamiento, se optimiza la función descrita en la Ec. 2.4:

$$min_G max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1-D(G(z)))] \tag{2.4}$$

Donde $p_{data}(x)$ y p_z representan las distribuciones de datos reales y de datos generados, respectivamente.

El objetivo de D es maximizar la probabilidad de asignar las etiquetas correctas tanto a las muestras falsas generadas por G como a las muestras reales del conjunto de entrenamiento. Simultáneamente, G se entrena para minimizar la pérdida $log(1-D(G(z)))$. Esta pérdida se calcula generando primero una muestra z a partir de la distribución a priori de la variable de ruido de entrada p_z. A continuación, el modelo G genera una muestra $x = G(z)$ en el espacio de datos. Posteriormente, el modelo D estima la probabilidad de que x provenga de los datos reales en lugar de haber sido generada por G. La función objetivo $log(1 - D(G(z)))$ se utiliza para entrenar a G a generar muestras más cercanas a los datos reales. Este objetivo incentiva a G a minimizar dicha pérdida produciendo muestras que tengan una mayor probabilidad de ser clasificadas como reales por D (Goodfellow et al., 2014).

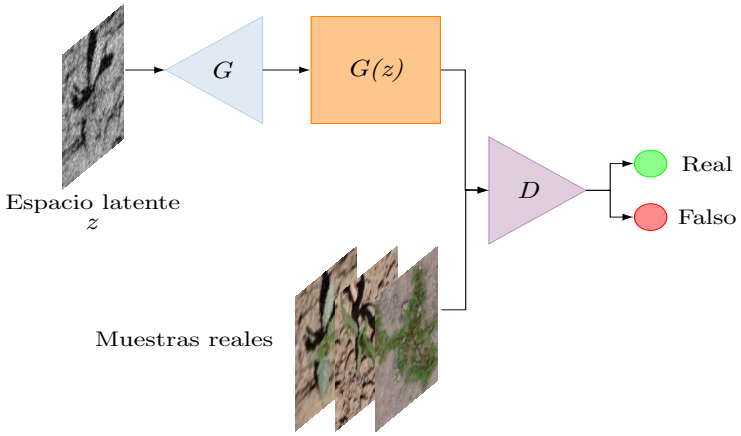


Figura 2.5: Arquitectura de un modelo GAN.

En otras palabras, se busca que G genere datos cada vez más realistas y difíciles de distinguir de los datos reales. Una vez que los modelos G y D están entrenados, G puede utilizarse para generar nuevas muestras que imitan las características y patrones presentes en el conjunto de entrenamiento (Mirza y Osindero, 2014). Esto significa que G es capaz de crear datos sintéticos que siguen los patrones de los datos reales utilizados durante el entrenamiento.

Características

a) Arquitectura dual: Las GAN se componen de dos redes neuronales (G y D), que se entrenan simultáneamente en un juego de suma cero. Esta estructura dual permite que las GAN generen datos sintéticos de alta calidad.

b) Capacidad de generación: El G tiene la capacidad de producir datos completamente nuevos que, aunque son sintéticos, siguen la distribución de los datos reales. Esto es especialmente útil en aplicaciones como la generación de imágenes, la creación de contenido multimedia y la simulación de escenarios complejos.

c) Problemas Comunes:

- Colapso de modo: Ocurre cuando el generador produce una base limitada de datos sintéticos similares, en lugar de reflejar la diversidad de la distribución real.
- Desvanecimiento de gradiente: Sucede cuando las actualizaciones de los parámetros son demasiado pequeñas, lo que impide el aprendizaje efectivo del G.
- Inestabilidad en el entrenamiento: Debido a la naturaleza competitiva del proceso de entrenamiento, las GAN pueden sufrir oscilaciones y encontrar dificultades para converger a un estado estable.

d) Aplicaciones: Las GAN han demostrado ser excepcionalmente útiles en áreas como la síntesis de imágenes, la traducción de imágenes entre dominios, la generación de videos, la mejora de imágenes y la creación de datos sintéticos para la ampliación de bases de datos en aprendizaje supervisado.

2.5 Detección de objetos en imágenes

La detección de objetos es una tarea fundamental en el campo de la visión artificial, diseñada para identificar y localizar instancias de objetos dentro de una imagen o un vídeo. Los objetivos principales de los algoritmos de detección de objetos son dibujar recuadros delimitadores alrededor de estos objetos y clasificarlos en categorías específicas.

Los algoritmos de detección de objetos basados en el DL se clasifican en dos tipos:

a) Detectores de objetos en dos etapas: Estos algoritmos dividen el proceso de detección en dos fases principales. En la primera, se generan cajas delimitadoras candidatas que podrían contener objetos, a menudo mediante una Red de Propuesta de Regiones (RPN, del inglés Region Proposal Network). En la segunda etapa, cada caja candidata se clasifica en una categoría específica y se calcula la probabilidad asociada a esta clasificación.

   · CNN basadas en regiones (R-CNN): La familia R-CNN, que incluye R-CNN, Fast R-CNN y Faster R-CNN (Girshick, 2015), ha sido fundamental en el desarrollo de detectores de dos etapas. R-CNN extrae un gran número de propuestas de regiones de la imagen utilizando métodos como la búsqueda selectiva, y cada propuesta se clasifica de forma independiente utilizando una CNN. Fast R-CNN mejora este método al compartir el cálculo entre las propuestas, reduciendo así la redundancia y el coste computacional. Faster R-CNN introduce la RPN, reemplazando la búsqueda selectiva por un componente de red aprendible, lo que hace que la generación de propuestas sea más rápida y precisa.

   · Redes piramidales de características FPN: Se integran a menudo con detectores de dos etapas para mejorar el rendimiento, especialmente en la detección de objetos de diferentes escalas. Al utilizar una pirámide de características multiescala, las FPN mejoran la capacidad del modelo para detectar objetos pequeños combinando mapas de características de alta resolución con otros de baja resolución y semánticamente ricos.

   Un ejemplo notable de un algoritmo de detección de objetos en dos etapas utilizado en esta tesis es Faster R-CNN.

b) Detectores de objetos de una etapa: Estos algoritmos combinan la localización y clasificación de objetos en un proceso unificado. Utilizan cajas de anclaje predefinidas, distribuidas densamente a lo largo de la imagen, para predecir simultáneamente la ubicación y la clase de los objetos.

   · YOLO: Es uno de los detectores de una sola etapa más populares, conocido por su capacidad para la detección de objetos en tiempo real. YOLO (Redmon et al., 2016) divide la imagen en una cuadrícula y predice los recuadros delimitadores y las probabilidades de clase directamente a partir de la imagen. Este enfoque elimina la necesidad de propuestas de regiones, acelerando considerablemente el proceso de detección. YOLOv3 y sus sucesores han perfeccionado este proceso incorporando técnicas avanzadas como la predicción multiescala, mejorando así su robustez y precisión.

   · SSD (Detector multicaja de un solo disparo): Es otro conocido detector de una sola etapa que utiliza un enfoque similar al de YOLO pero con algunas diferencias clave. SSD (Liu et al., 2016) predice múltiples cuadros delimitadores con diferentes relaciones de

aspecto para cada ubicación en un mapa de características. Combina las predicciones de múltiples mapas de características a diferentes escalas para manejar mejor objetos de distintos tamaños.

Los algoritmos de detección de objetos de una etapa analizados en esta tesis incluyen YOLOv8 (Jocher et al., 2023), DETR (Carion et al., 2020) y DETA (Ouyang-Zhang et al., 2022).

Aunque los modelos de dos etapas ofrecen mayor exactitud, su uso está limitado por sus elevados requerimientos computacionales (Xiuling et al., 2024), lo que dificulta su implementación en dispositivos con recursos restringidos. Por otro lado, los detectores de una etapa, destacan por su velocidad, lo que los hace ideales para aplicaciones en tiempo real.

Esta tesis cobra importancia al explorar un campo con aplicaciones prácticas significativas: la detección y clasificación de especies de malas hierbas en estado temprano utilizando imágenes capturadas desde UAV. Este problema presenta un desafío particular para los algoritmos de detección de objetos, ya que las malas hierbas en etapas de crecimiento temprano son pequeñas, estan dispersas de manera irregular y se camuflan en el entorno natural, condiciones bajo las cuales los métodos convencionales de visión artificial suelen fallar. Es por ello que se plantea la necesidad de desarrollar o adaptar técnicas de detección más robustas que puedan detectar con exactitud objetos pequeños como las malas hierbas en estado temprano, en escenas complejas.

2.6 Entrenamiento

El entrenamiento de un modelo de DL, específicamente una red neuronal, se basa en el ajuste iterativo de los pesos y sesgos de la red para minimizar una función de coste, que cuantifica el error entre las predicciones de la red y los valores reales observados. Este ajuste se realiza mediante el algoritmo de retropropagación, que utiliza la regla de la cadena para calcular el gradiente de la función de coste con respecto a cada peso y sesgo de la red.

En cada iteración del entrenamiento, se evalúa cómo un pequeño cambio en un peso o sesgo afecta el coste total de la red. El proceso de retropropagación implica calcular tres componentes clave: 1) la derivada del coste respecto a la activación de la neurona; 2) la derivada de la activación respecto a la entrada ponderada, que depende de la función de activación utilizada (como la sigmoid o ReLU); y 3) la derivada de la entrada ponderada respecto al peso. Estos cálculos permiten determinar la dirección y magnitud de los ajustes necesarios para cada peso y sesgo, con el objetivo de reducir el error global. La retropropagación ajusta los parámetros de cada capa comenzando desde la capa de salida y avanzando hacia atrás, capa por capa, actualizando gradualmente la red entera.

Aunque el concepto de retropropagación puede aplicarse a una red simple con una sola neurona por capa, se extiende de manera efectiva a redes más complejas con múltiples neuronas y capas, siguiendo los mismos principios fundamentales de cálculo del gradiente. Este enfoque iterativo y basado en gradientes es el corazón del DL, permitiendo a las redes neuronales aprender patrones complejos y generalizar mejor a datos nuevos.

2.6.1 Hiperparámetros

Los hiperparámetros son parámetros ajustables que permiten controlar el proceso de optimización del modelo. La configuración de estos hiperparámetros puede influir significativamente en el rendimiento del modelo y en las tasas de convergencia durante el entrenamiento. En la presente tesis, se ajustaron los siguientes hiperparámetros para todos los modelos implementados:

- Número de épocas (*Epochs*): Cantidad total de iteraciones completas sobre la base de datos de entrenamiento. Cada época representa una pasada completa por todos los datos.
- Tamaño del lote (*Batch size*): Número de muestras de datos procesadas simultanemente en cada iteración antes de actualizar los parámetros del modelo. El tamaño del lote afecta la estabilidad y velocidad del entrenamiento.
- Tasa de aprendizaje (*Learning rate*): Magnitud de los ajustes realizados en los parámetros del modelo durante cada actualización. Una tasa de aprendizaje baja puede hacer que el proceso de entrenamiento sea más lento, mientras que una tasa de aprendizaje alta puede llevar a una convergencia inestable o a un comportamiento impredecible del modelo.

2.6.2 Bucle de optimización

Después de establecer los hiperparámetros, se entrena y optimiza el modelo con un bucle de optimización. Cada iteración del bucle de optimización se denomina época. Cada época consta de dos fases principales:

- Bucle de entrenamiento (*train loop*): En esta fase, se recorre la base de datos de entrenamiento para ajustar los parámetros del modelo con el fin de que converjan hacia los valores óptimos. Durante el bucle de entrenamiento, se calculan los gradientes y se actualizan los pesos del modelo para minimizar la función de coste.
- Bucle de validación/prueba (*validation/test loop*): Tras completar el bucle de entrenamiento, se evalúa el rendimiento del modelo en la base de datos de validación o prueba. Esta fase permite comprobar si el modelo está mejorando y ajustar los hiperparámetros si es necesario. El objetivo es verificar que el modelo generalice bien a datos no vistos y evitar problemas como el sobreajuste.

2.6.3 Función de pérdida

La función de pérdida cuantifica el grado de disimilitud entre las predicciones realizadas por el modelo y los valores reales objetivo. Es la métrica que se busca minimizar durante el proceso de entrenamiento del modelo. Para calcular la pérdida, se realiza una predicción utilizando las entradas de una muestra de datos dada y se compara el resultado con el valor verdadero de la etiqueta asociada a esos datos. La función de pérdida proporciona una medida de cuán lejos están las predicciones del modelo de los valores reales, y su minimización guía el ajuste de los parámetros del modelo para mejorar su exactitud.

2.6.4 Optimizador

La optimización es el proceso de ajuste iterativo de los parámetros del modelo con el objetivo de minimizar el error de predicción en cada paso del entrenamiento. Los algoritmos de optimización utilizados en la tesis fueron Adam y AdamW. Ambos algoritmos mejoran la convergencia y la estabilidad del entrenamiento al adaptar las tasas de aprendizaje para cada parámetro, considerando tanto el primer momento (la media de los gradientes) como el segundo momento (la varianza de los gradientes).

2.6.5 Aumento de datos

El aumento de datos (*Data Augmentation*) es una técnica fundamental en el campo de la visión artificial y DL que busca ampliar y diversificar la base de datos de entrenamiento (Mumuni y Mumuni, 2022). Dado que la cantidad y calidad de los datos disponibles son críticos para el rendimiento de los modelos de DL, el aumento de datos se utiliza para mitigar problemas de sobreajuste, mejorar la generalización y hacer que los modelos sean más robustos frente a variaciones que no se observan directamente en los datos de entrenamiento.

El principio básico del aumento de datos consiste en aplicar transformaciones a los datos existentes, como imágenes o secuencias de video, para generar nuevas instancias de entrenamiento sin cambiar las etiquetas asociadas. Estas transformaciones pueden incluir técnicas tradicionales, como rotaciones y escalados, así como métodos modernos, como la generación de datos sintéticos. El aumento de datos puede aplicarse en el espacio de entrada, en el espacio de características, o mediante la síntesis de datos completamente nuevos.

Técnicas y métodos

a) Transformaciones geométricas:

   · Rotación, traslación y escalado: Estas técnicas alteran la estructura geométrica de la imagen sin modificar el contenido semántico. Son simples pero efectivas para aumentar la diversidad de la base de datos.

   · Transformaciones afines y no afines: Incluyen métodos más complejos como proyección y la deformación no lineal, que son útiles para simular variaciones en la perspectiva o deformaciones no rígidas.

b) Transformaciones fotométricas:

   · Ajustes de brillo, contraste y saturación: Estas modificaciones alteran las propiedades visuales de la imagen, lo que ayuda a los modelos a volverse más robustos frente a variaciones de iluminación y condiciones ambientales.

c) Métodos avanzados en el espacio de entrada:

   · Redes de transformación espacial (STN, del inglés *Spatial Transformer Network*): Estas redes aprenden transformaciones útiles directamente a partir de los datos, lo que aumenta la diversidad en las representaciones aprendidas (Jaderberg et al., 2015).

   · Métodos de región: Incluyen técnicas como *CutOut* (DeVries y Taylor, 2017), *Random*

Erasing (Zhong et al., 2020), y *CutMix* (Yun et al., 2019), que eliminan o reemplazan partes de la imagen para simular oclusiones o mezclar información entre distintas imágenes.

d) Técnicas de síntesis de datos:

- Modelado gráfico y renderizado neural: Utilizan herramientas de gráficos por computadora o redes neuronales para generar escenas 2D y 3D realistas, proporcionando datos sintéticos para entrenar modelos en tareas como la conducción autónoma.
- GAN: Se emplean para crear imágenes sintéticas que imitan la distribución de los datos reales. Esto resulta beneficioso en tareas de adaptación de dominio y generación de datos en condiciones adversas. Las GAN permiten la expansión masiva de bases de datos mediante la generación de nuevas imágenes que, aunque conservan la coherencia con las características fundamentales de los datos originales, introducen variaciones controladas (Olaniyi et al., 2022).

Ventajas

a) Mejora de la generalización: Al introducir variaciones adicionales en los datos, los modelos aprenden a ser más robustos frente a las diferencias en los datos de prueba, lo que mejora su capacidad de generalización.

b) Reducción de sobreajuste: Incrementar la diversidad en la base de datos de entrenamiento disminuye la probabilidad de que el modelo se ajuste excesivamente a patrones específicos de la base de datos original, reduciendo el riesgo de sobreajuste.

c) Mayor eficiencia de modelos simples: Con una base de datos más diversa, incluso los modelos menos complejos pueden alcanzar un rendimiento comparable al de modelos más sofisticados.

Desventajas

a) Incremento en la complejidad computacional: La aplicación de técnicas avanzadas, especialmente aquellas que implican CNNs, puede requerir un aumento significativo en los recursos computacionales.

b) Generación de datos irrelevantes o espurios: Algunas técnicas de aumento, como la mezcla de imágenes o características, pueden introducir datos que no son representativos de la distribución de datos reales, lo que puede llevar a una disminución del rendimiento del modelo.

c) Dependencia en la heurística humana: La selección y diseño de transformaciones adecuadas todavía dependen en gran medida del conocimiento humano sobre el dominio y el problema específico.

En esta tesis, las imágenes adquiridas tenían un tamaño máximo de 85 píxeles, lo que representa una resolución espacial baja. Para algunos modelos, resulta desafiante aprender características útiles a partir de imágenes con tan baja resolución. Para abordar este problema, se podría considerar la posibilidad de entrenar un modelo más complejo con una capacidad superior para extraer características

de imágenes pequeñas. Sin embargo, esto implica un mayor tiempo de entrenamiento, un mayor consumo de recursos informáticos y, por ende, un incremento en el gasto de energía.

2.6.6 Aprendizaje por transferencia

El TL es una técnica de ML en la que un modelo desarrollado para una tarea se reutiliza como punto de partida para desarrollar un modelo sobre una segunda tarea. Resulta especialmente útil en el DL cuando la base de datos inicial no es lo suficientemente grande o no está debidamente etiquetada como para entrenar un modelo desde cero (Hasan et al., 2021). En lugar de partir de cero, el TL aprovecha los conocimientos adquiridos a partir de un modelo preentrenado sobre un problema relacionado. El TL se utiliza principalmente para:

a) Reducir la necesidad de grandes bases de datos etiquetados: Al utilizar modelos preentrenados, se puede reducir significativamente la cantidad de datos etiquetados necesarios para el entrenamiento.

b) Acelerar el tiempo de entrenamiento: Dado que el modelo ya ha aprendido las características de la tarea de origen, requiere menos tiempo para aprender nuevas tareas.

c) Mejorar el rendimiento en tareas con datos limitados: Los modelos entrenados con TL a menudo superan a los modelos entrenados desde cero, especialmente en situaciones en las que los datos son escasos.

Las técnicas de TL empleadas en el desarrollo de la tesis fueron:

a) Ajuste fino: Consistió en tomar un modelo preentrenado y actualizar ligeramente los pesos con los nuevos datos específicos para la tarea de clasificación. Esto se implementó con los modelos estudiados en el capítulo 5, cuyos sus pesos preentrenados en ImageNet, se afinaron para una tarea de clasificación diferente.

b) Extracción de características: En este enfoque, la base convolucional de un modelo preentrenado se utiliza como extractor de características fijo. Sólo la capa de clasificación final se entrena con los nuevos datos. Esta técnica fue empleada en la implementación del detector descrito en el capítulo 5.

c) Adaptación al dominio: Este proceso implica ajustar un modelo entrenado en un dominio específico para que funcione adecuadamente en un dominio diferente pero relacionado. Esta metodología se exploró en profundidad en el capítulo 6, donde se llevó a cabo una adaptación al dominio en su forma más extrema.

d) Aprendizaje multitarea: En este enfoque, un único modelo se entrena para realizar múltiples tareas con la expectativa de que el aprendizaje de una de ellas beneficie a las demás. Este enfoque se aplicó en el capítulo 7, donde se incorpora implícitamente el TL, dado que el conocimiento compartido entre las tareas puede mejorar el rendimiento en todas ellas.

2.7 Métricas de evaluación

2.7.1 Matriz de confusión

En esta tesis, se utilizó la matriz de confusión como herramienta para evaluar el rendimiento de los modelos de clasificación. Esta matriz, de dimensiones $N \times N$, donde N representa el número de clases en la base de datos, establece una correlación directa entre las etiquetas reales y las predicciones del modelo. Es fundamental en el ámbito del ML, ya que facilita una comparación precisa entre las predicciones generadas por el modelo y los valores observados. Este análisis no solo permite identificar el nivel de acierto del modelo, sino también comprender sus debilidades, lo cual es crucial para mejorar su precisión y robustez en aplicaciones futuras. La matriz de confusión resume el rendimiento del modelo en términos de cuatro categorías principales:

a) Verdaderos Positivos (TP): Representan los casos en los que el modelo clasifica correctamente una instancia como perteneciente a la clase positiva. Es decir, son las predicciones positivas correctas.

b) Falsos Positivos (FP): Corresponden a los casos en los que el modelo clasifica incorrectamente una instancia como perteneciente a la clase positiva cuando, en realidad, pertenece a la clase negativa. Este tipo de error se conoce como error tipo I.

c) Verdaderos Negativos (TN): Son los casos en los que el modelo clasifica correctamente una instancia como perteneciente a la clase negativa. Es decir, son las predicciones negativas correctas.

d) Falsos Negativos (FN): Representan los casos en los que el modelo clasifica incorrectamente una instancia como perteneciente a la clase negativa cuando, en realidad, pertenece a la clase positiva. Este tipo de error se conoce como error tipo II.

Estas categorías permiten calcular diferentes métricas de rendimiento que son fundamentales para evaluar la efectividad de un modelo de clasificación. Entre las métricas más comunes derivadas de la matriz de confusión se incluyen:

· ***Accuracy***: Fracción de predicciones correctas con respecto el total de predicciones (Ec. 5.1).

· ***Precision***: Fracción de instancias correctamente clasificadas como positivas con respecto al total de instancias clasificadas como positivas (Ec. 5.2).

· ***Recall***: Fracción de instancias positivas correctamente identificadas (Ec. 5.3).

· ***Specificity***: Fracción de instancias negativas correctamente identificadas (Ec. 5.4).

· ***F1-Score***: Media armónica entre *precision* y *recall*, que ofrece un balance entre ambas métricas (Ec. 2.9).

$$Accuracy = \frac{T_P + T_N}{T_P + F_P + F_N + T_N} \tag{2.5}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{2.6}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{2.7}$$

$$\textit{Specificity} = \frac{T_N}{T_N + F_P} \tag{2.8}$$

$$\textit{F1-Score} = \frac{2T_P}{2T_P + F_P + F_N} \tag{2.9}$$

Cada una de estas métricas proporciona una visión diferente del rendimiento del modelo. Es fundamental seleccionar la métrica adecuada según el problema específico en el que se esté trabajando. Por ejemplo, la *accuracy* solo es válida cuando la distribución de clases es simétrica (todas las clases tiene una cantidad similar de muestras).

Para evaluar los modelos de detección de objetos, se utilizaron las siguientes métricas:

- ***Intersection over Union (IoU)***: Evalúa la exactitud de las predicciones del modelo determinando en qué medida una región predicha se solapa con el objeto real. *IoU* se calcula como el área de intersección dividida por el área de unión entre la caja de la verdad-terreno y la caja predicha por el modelo (Ec. 2.10).
- ***Average Precision (AP)***: Representa la media de las precisiones a lo largo de diferentes umbrales de recuperación (n), proporcionando una evaluación integral del rendimiento del modelo, teniendo en cuenta tanto la precisión como la capacidad de recuperación en un solo valor numérico. La AP se calcula mediante un bucle que recorre todos los pares de valores de *precision* y *recall*, determinando la diferencia entre el valor de *recall* actual y el siguiente, y luego multiplicando por la *precision* correspondiente al umbral actual. (Ec. 2.11).
- ***Mean Average Precisión (mAP)***: Compara el cuadro delimitador de la verdad-terreno con el cuadro detectado y devuelve una puntuación. Cuanto mayor sea la puntuación, más preciso será el modelo en sus detecciones (Ec. 2.12).

$$\textit{IoU} = \frac{\textit{Intersection Area}}{\textit{Union Area}} \tag{2.10}$$

$$AP = \sum_{k=0}^{k=n-1} [Recall(k) - Recalls(k+1) * Precision(k)] \tag{2.11}$$

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \tag{2.12}$$

Para evaluar de manera objetiva si el modelo ha predicho correctamente la ubicación de una caja delimitadora, se emplea un umbral basado en la puntuación de la IoU. Si la predicción del modelo genera una caja con una puntuación de IoU mayor o igual al umbral establecido, se considera que existe una alta superposición entre la caja predicha y una de las cajas de verdad-terreno. En este caso, el modelo se considera exitoso en la detección del objeto, y la región detectada se clasifica como positiva, es decir, que contiene un objeto. Por el contrario, si la puntuación de IoU es menor que el umbral, se interpreta que la predicción del modelo es deficiente, ya que la superposición entre la caja predicha y la caja de verdad-terreno es insuficiente. En este escenario, la región detectada se clasifica como negativa, lo que indica que no contiene un objeto.

2.8 Mapa de prescripción: Transformación de coordenadas: locales a globales

Para visualizar de manera detallada y georreferenciada la información de interés generada por los modelos de DL, se implementó una transformación geométrica que incluía escalamiento para transformar coordenadas locales a coordenadas globales. El sistema de coordenadas locales (LCS, del inglés *Local Coordinate System*) corresponde al sistema de referencia local (imágenes particionadas a partir de un ortomosaico), donde las coordenadas son relativas a un origen local. Por otro lado, el sistema de coordenadas globales (GCS, del inglés *Global Coordinate System*) es un sistema de referencia global, como el sistema geodésico WGS84, donde las coordenadas son absolutas y están referenciadas a la Tierra. Para transformar las coordenadas de un punto desde el sistema local al sistema global, se aplica la Ec. 2.13.

$$P_{GCS} = \Gamma \cdot P_{LCS} \tag{2.13}$$

Donde Γ es la matriz de transformación. Por lo tanto, la expresión para localizar un punto en el GCS se expresa como:

$$\begin{bmatrix} x_{GCS} \\ y_{GCS} \end{bmatrix} = \begin{bmatrix} Ax_{GCS} \\ Ay_{GCS} \end{bmatrix} + \Gamma \cdot \begin{bmatrix} x_{LCS} \\ y_{LCS} \end{bmatrix} \tag{2.14}$$

Donde $(Ax_{GCS}; Ay_{GCS})$ corresponden a las coordenadas que se utilizaron para realizar la partición del ortomosaico. Además, se estableció la matriz de transformación (Γ), que contiene el escalado necesario para aplicar a las coordenadas de los puntos y realizar esta conversión. Este escalado se define como la relación entre la GSD y el tamaño de la imagen particionada (m_{size}).

$$\begin{bmatrix} x_{GCS} \\ y_{GCS} \end{bmatrix} = \begin{bmatrix} Ax_{GCS} \\ Ay_{GCS} \end{bmatrix} + \begin{bmatrix} \frac{GSD}{m_{size}} & 0 \\ 0 & -\frac{GSD}{m_{size}} \end{bmatrix} \cdot \begin{bmatrix} x_{LCS} \\ y_{LCS} \end{bmatrix}$$

$$\begin{bmatrix} x_{GCS} \\ y_{GCS} \end{bmatrix} = \begin{bmatrix} Ax_{GCS} \\ Ay_{GCS} \end{bmatrix} + \frac{GSD}{m_{size}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} x_{LCS} \\ y_{LCS} \end{bmatrix} \tag{2.15}$$

En el caso particular de la representación del punto central de la mala hierba en el ortomosaico, con la información de las coordenadas locales de las cajas delimitadoras proporcionadas por los modelos de DL, se calculan las coordenadas del centroide (Ec. 2.16) para luego ser proyectadas al GCS utilizando la Ec. 2.15. Como se muestra en la Figura 2.6, la imagen contenida en el plano LCS $(x_{LCS}; y_{LCS})$ corresponde a la inferencia realiza por el modelo de DL. El punto $(x_m; y_m)$ muestra la ubicación de la mala hierba identificada y georreferenciada en el mundo real.

$$(x_c; y_c) = \left(\frac{x_{min} + x_{max}}{2}; \frac{y_{min} + y_{max}}{2} \right) \tag{2.16}$$

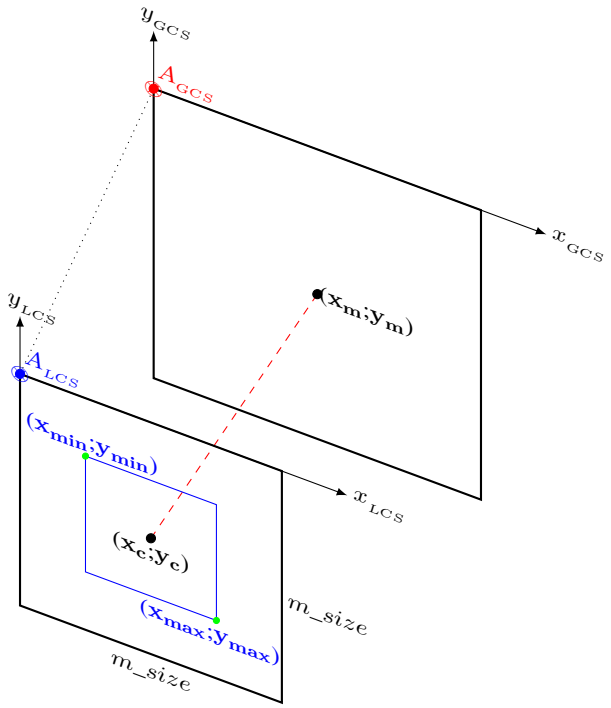


Figura 2.6: Relación entre la imagen evaluada en el modelo de DL y su correspondiente ubicación en el ortomosaico.

Para automatizar la georreferenciación de las coordenadas de los centroides de las especies de malas hierbas detectadas y clasificadas dentro de los ortomosaicos, se implementó el Algoritmo 2.2.

Algorithm 2.2 Georreferenciación de malas hierbas (centroide)

Input: Directory with images
Output: CSV file with calculated values
Initialize: Models
procedure OBJECTDETECTIONANDCENTROIDCALCULATION
    Images ← ListFilesInDirectory()
    for each ImageName in Images **do**
        Class, BoundingBox, Confidence ← EvaluateModel(Models, ImageName)
        Centroid(x_c, y_c) ← CalculateCentroid(BoundingBox) ▷ Ec. 2.16
        OrthoCentroid(x_m, y_m) ← TransformToOrthomosaicCoordinates(Centroid) ▷ Ec. 2.15
        SaveToCSV(Class, OrthoCentroid, Confidence)
    end for
end procedure

2.9 Software y Hardware

2.9.1 Software

El desarrollo y entrenamiento de modelos de DL para la identificación de especies de malas hierbas basado en imágenes se realizó en esta tesis mediante los frameworks de ML Keras con TensorFlow[b] y PyTorch[c], ambos ampliamente reconocidos por su flexibilidad, robustez y soporte para una amplia variedad de arquitecturas de redes neuronales. La elección de estos frameworks se fundamenta en su capacidad para manejar operaciones tensoriales altamente optimizadas, su integración con hardware acelerado y su facilidad para desarrollar modelos personalizados.

Keras - TensorFlow

Keras es una API (del inglés, *Application Programming Interface*) de alto nivel para redes neuronales que se ejecuta sobre TensorFlow, proporcionando una interfaz intuitiva y de fácil de usar para la construcción de modelos. TensorFlow (Abadi et al., 2016), desarrollado por Google Brain, es un marco de trabajo de código abierto que permite la ejecución eficiente de algoritmos de ML y otras operaciones numéricas intensivas en GPU y TPU. Para el entrenamiento de los modelos de clasificación y detección de especies de malas hierbas, se utilizaron varias bibliotecas y extensiones de TensorFlow, tales como:

- TensorFlow GPU: Esta extensión permite la utilización de la GPU para acelerar el cálculo de operaciones tensoriales, fundamental para el entrenamiento de modelos a gran escala.
- TensorFlow Hub: Un repositorio de modelos preentrenados que facilitó la transferencia de aprendizaje, acelerando el proceso de entrenamiento al utilizar pesos iniciales optimizados.

PyTorch

PyTorch, desarrollado por Facebook AI Research (FAIR), es otro marco de trabajo de ML que ha ganado popularidad gracias a su enfoque dinámico de definición de gráficos computacionales (*define-by-run*), que facilita la depuración y modificación en tiempo real de los modelos. En esta tesis, PyTorch se utilizó para implementar y entrenar los modelos de detección de objetos utilizando técnicas avanzadas de DL. Las bibliotecas clave utilizadas con PyTorch incluyeron:

- TorchVision: Un conjunto de herramientas que contiene modelos preentrenados, funciones de transformación de datos y cargas de datos específicas para visión artificial, lo que facilita la preparación y normalización de bases de datos.
- PyTorch Lightning: Utilizado para simplificar el proceso de entrenamiento de modelos, proporcionando una estructura clara y reduciendo el código repetitivo, facilitando así la experimentación rápida.

[b] https://www.tensorflow.org/guide/keras
[c] https://pytorch.org

2.9.2 Hardware

El entrenamiento y prueba de los modelos de DL se realizaron en el sistema operativo Ubuntu 20.04.4 LT y 22.04.4 LTS, ampliamente utilizado en la comunidad científica debido a su estabilidad, soporte de software y compatibilidad con controladores de GPU. El sistema estuvo equipado con una GPU NVIDIA GeForce RTX 3070 Ti, basada en la arquitectura Ampere, que cuenta con 8 GB de memoria GDDR6X, 6144 núcleos CUDA y soporte para el entrenamiento de precisión mixta a través de Tensor Cores. El sistema también contó con un procesador de 12ª generación Intel® Core™ i7-12700K a 3.61GHz, que posee 12 núcleos (8 de alto rendimiento y 4 de alta eficiencia), soportando un rendimiento elevado en tareas que requieren procesamiento paralelo intensivo.

Capítulo 3

Boosting precision crop protection towards agriculture 5.0 *via* machine learning and emerging technologies: A contextual review.

Publicación asociada a este capítulo

- Mesías-Ruiz GA, Pérez-Ortiz M, Dorado J, de Castro AI and Peña JM (2023) Boosting precision crop protection towards agriculture 5.0 *via* machine learning and emerging technologies: A contextual review. *Front. Plant Sci.*, 14:1143326. doi: 10.3389/fpls.2023.1143326

Abstract

Crop protection is a key activity for the sustainability and feasibility of agriculture in a current context of climate change, which is causing the destabilization of agricultural practices and an increase in the incidence of current or invasive pests, and a growing world population that requires guaranteeing the food supply chain and ensuring food security. In view of these events, this article provides a contextual review in six sections on the role of artificial intelligence (AI), machine learning (ML) and other emerging technologies to solve current and future challenges of crop protection. Over time, crop protection has progressed from a primitive agriculture 1.0 (Ag1.0) through various technological developments to reach a level of maturity closely in line with Ag5.0 (section 3.1), which is characterized by successfully leveraging ML capacity and modern agricultural devices and machines that perceive, analyze and actuate following the main stages of precision crop protection (section 3.2). Section 3.3 presents a taxonomy of ML algorithms that support the development and implementation of precision crop protection, while section 3.4 analyses the scientific impact of ML on the basis of an extensive bibliometric study of $>$120 algorithms, outlining the most widely used ML and deep learning (DL) techniques currently applied in relevant case studies on the detection and control of crop diseases, weeds and plagues. Section 3.5 describes 39 emerging technologies in the fields of smart sensors and other advanced hardware devices, telecommunications, proximal and remote sensing, and AI-based robotics that will foreseeably lead the next generation of perception-based, decision-making and actuation systems for digitized, smart and real-time crop protection in a realistic Ag5.0. Finally, section 3.6 highlights the main conclusions and final remarks.

3.1 Linking Crop Protection to the technological evolution of agriculture

Crop protection involves a large number of critical farming activities with a decisive impact on the viability and sustainability of agriculture. Throughout history, humans have developed new methods and practices to protect their crops. From ancient times to about 1950, agriculture 1.0 employed a large workforce to manually control crop pests (i.e., plant diseases, weeds and other plagues, both vertebrate and invertebrate), which produced low yields but in sufficient quantity to feed the population. In the late 1950s, agriculture 2.0 began with the use of synthetic pesticides and specialized machines to control the common crop pests. At that stage, agriculture evolved towards the economic edge, aiming to produce more food at a cheaper price, i.e., towards a more industrialized agriculture. At the end of the 20th century, agriculture 3.0 emerged with the idea of using new technologies and data-driven modeling as essential tools to take decisions and manage cropping systems. This disruptive concept led to the origin of precision agriculture, in which telematics, global navigation satellite systems (GNSS), machinery guidance, and sensing devices aimed to optimize the crop protection tasks, to reduce costs and environmental impacts of pesticides, and to improve food quality. What followed was a further step in the integration of geo-spatial technologies, computer sciences and digitization into the agricultural process, where sensors, mobile telephony, embedded systems, cloud computing, internet of things (IoT) and big data were incorporated on board of autonomous machinery, smart sprayers and actuators to facilitate the application of the precision crop protection paradigm within the concept of agriculture 4.0 (Zhai et al., 2020). Continuing this evolution, Agriculture 5.0 (Ag5.0) will promote a new era of intelligent crop management

with automatized decision making processes, unmanned operations and progressively less human intervention supported by the latest Artificial Intelligence (AI) systems, advanced robotics, and powerful Machine Learning (ML) algorithms (Saiz-Rubio y Rovira-Más, 2020).

Modern agriculture will face in the next decades two immense challenges never seen in previous generations. The first one is the impact of climate change in agricultural systems (Hoegh-Guldberg et al., 2019), which causes destabilization of farming practices (Mulla et al., 2020) and irregular crop seasons due to excessive heat and water scarcity in large productive areas (Piao et al., 2019); (Falkland y White, 2020), which inevitably leads to the emergence of new invasive pests or the increased severity of existing ones. The second one is to produce food for a growing human and animal population, while ensuring food security by using fewer agrochemicals and imposing strict controls at all stages of the agricultural supply chain (van Dijk et al., 2020). In view on this imminent future, Ag5.0 must offer creative solutions based on AI, ML algorithms and other technological innovations that continuously interact with the crop and its environment, which will require undoubtedly transdisciplinary studies and interdisciplinary collaborations, where precision crop protection becomes a key discipline in the Ag5.0 revolution by implementing new procedures and strategies to drastically reduce the use of agrochemicals in the control of diseases, weeds and plagues.

3.2 The stages of precision crop protection: perception, analysis and actuation

The use of new technologies in crop protection aims at detecting and identifying the symptoms or problems caused by crop pests (Behmann et al., 2015), followed by a site-specific application of a chemical or mechanical control action. This process comprises the three main stages for pursuing a precision crop protection strategy, as follows (Figure 3.1): 1) perception, 2) analysis and, optionally (but recommendable) decision-making, and 3) actuation. The perception stage involves field inspection and acquisition of plant information (e.g., crop and/or weed imaging) through a sensor or camera mounted on an on-ground or a remotely-sensed platform, while the actuation stage consists on the application of a prescribed site-specific treatment with a smart equipment usually assisted by a GNSS receiver. The necessary link between perception and actuation is the analysis stage, which consists of in-depth evaluation of digital crop data by using diverse data-driven techniques and identifying targeting areas of crops with problems associated to diseases, weeds and plagues. The analysis stage also often includes the generation of management zones and treatment/prescription maps following a decision-making process, e.g. based on the outcomes of a decision support system (DSS).

Recent bibliographic reviews point out to Unmanned Aerial Vehicles (UAVs), innovative ML algorithms, and various robots and autonomous equipment as the most disruptive technology for each stage, respectively (Cardim Ferreira Lima et al., 2020; Dainelli et al., 2021; Filho et al., 2020). UAVs are playing an important role in the perception stage due to their capability to capture crop data from large areas in a short time and with diverse types of cameras and sensors (e.g., RGB cameras, multi- and hyper- spectral sensors, thermal cameras, active sensors such as LiDAR, radar or sonar), which have led to significant progress in pest monitoring with the help of powerful analysis procedures, either by direct observation of the pest (e.g., weed patches), by diagnosis of the main

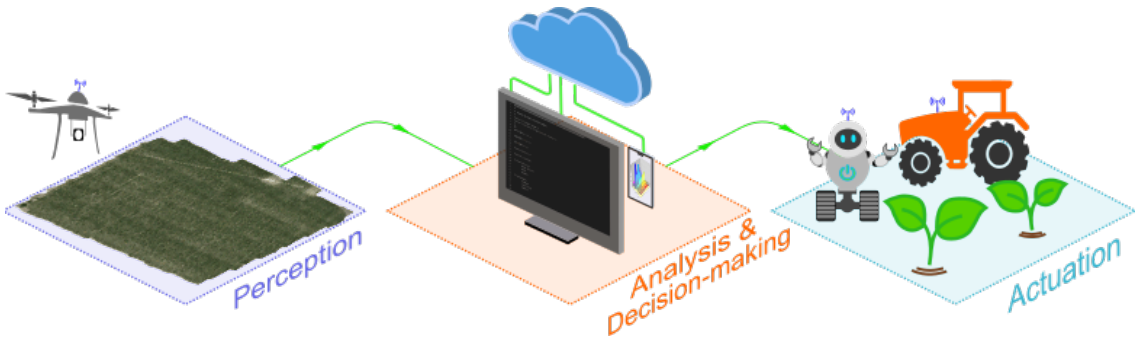


Figura 3.1: The main stages of precision crop protection.

symptoms of the disease (e.g., leave decay or thermal stress), or by detection of damages caused in the crop leaves and canopy (e.g., foliar losses due to a plague attack).

The analysis stage is the major challenge for many crops, probably being the bottleneck for the progress of precision crop protection. The ultimate objective of this stage is the accurate and timely detection of each crop-specific disease, weed or plague, whose complexity lies in the vast number of possible crop-pest scenarios with a diverse typology of associated symptoms, in addition to other environmental and cultural factors such as different weather conditions, soil properties, and farmers' decisions on crop field management, which impact the type and degree of severity of pest occurrences (Oerke et al., 2012; Pätzold et al., 2020). This diversity of variables and factors can be addressed by ML methods with the ability to learn from experience (i.e., data) and integrate information from multiple sources. ML enables the analysis of massive amounts of crop and pest data over time by taking advantage of the continuous evolution of the hardware with increasingly powerful central (CPU), graphics (GPU) and tensor (TPU) processing units (Wang et al., 2019a). As a result, ML can study the behavior of natural crop-pest systems by capturing and exploiting the underlying patterns in the data and build predictive/generative models accordingly for critical analytical tasks such as image classification, object detection, pattern recognition, geo-location, etc., aimed to propose solutions for complex crop protection challenges.

Finally, actuation is the task that leveraged large-scale viability of precision crop protection strategies, leading to great scientific and technological effort in the last decade to develop autonomous machinery, smart sprayers and agricultural robots that effectively implement site-specific crop management (Lowenberg-DeBoer et al., 2021; Shafi et al., 2019), either by direct treatment in real-time (Pérez-Ruiz et al., 2015) or, eventually, assisted by a prescription map (Fernández-Quintanilla et al., 2018) according to the principles established by the International Society of Precision Agriculture (ISPA, 2021).

3.3 ML taxonomy based on the tasks to be solved

The ML algorithms have been conventionally classified according to different criteria, based on: i) the nature of the model (full or partial probabilistic/generative model *vs*. discriminant model), ii) the type of reasoning applied (inductive or transductive, depending on whether the model performs a reasoning from observed training cases to general rules or the other way around, respectively), or iii) the data availability and the supervision process (unsupervised, supervised, semi-supervised and

reinforcement learning). However, the extent of ML within the scope of precision crop protection is best described by an alternative criterion based on the task to be solved, which leads to an expanded taxonomy of six categories, as follows: classification, regression, clustering, anomaly detection, dimensionality reduction, and association rule learning. These six tasks can be addressed with traditional ML algorithms or, for some specific tasks mainly classification and regression, with the more advanced artificial neural network (ANN) models, which in turn also include Deep Learning (DL) algorithms (Figure 3.2).

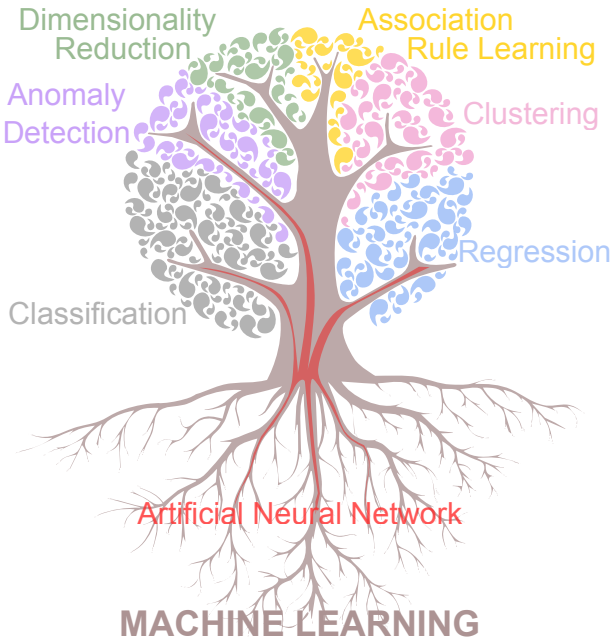


Figura 3.2: Taxonomy of Machine Learning according to the type of task to be solved.

3.3.1 Traditional ML algorithms

Traditional ML algorithms usually approach learning tasks by analyzing and interpreting input data with well-established architectures optimized for common computing resources, thus often achieving satisfactory results but with less accuracy and versatility than sophisticated ANN algorithms. Of the tasks listed above, classification is the most common in many disciplines, with well-known algorithms such as support vector machine (SVM), decision trees (DT), random forest (RF), K-Nearest Neighbor (k-NN), etc. Classification algorithms are part of the supervised learning type, aiming to categorize a certain set of structured or unstructured data in classes, being a binary classification when the objective is to predict the state of true or false, and a multi-category classification when there are more than two objective classes (Djafri y Gafour, 2022; Sen et al., 2020). These algorithms are used for predictive tasks in the fields of image analysis, video, object recognition, data mining, etc. (Kowsari et al., 2019), all of which are relevant to deal with the challenge of automatic identification, detection or classification of plant diseases, weeds and plagues. This objective also usually requires previous phases such as image/video preprocessing, segmentation and feature extraction that imply the use of other algorithms of the regression, clustering and dimensionality reduction typology.

The regression algorithms are also part of the supervised learning type and consist in relating continuous input and output variables through a function, which can be set by parametric or non-parametric approaches. In the former case, the output values are predicted by an explicit analytical formula that adjusts the known points by establishing and minimizing a cost function (e.g. linear regression) that link the input and output variables (Gaitán, 2020; Wei et al., 2015). In the latter case, a kernel function is defined to determine the prediction for the output based on similar experiences of the inputs, hence it depends on the correlation between the output and the known points surrounding the input (Čížek, Pavel and Sadıkoğlu, Serhan, 2020). A form of regression that allows correction of overfitting is the regularization algorithms, which avoid generating low error (i.e., high accuracy) in the training but high error during the testing (Zou y Hastie, 2005). Common algorithms in this group are LASSO Regularization, Ridge Regularization and Elastic Net Regularization.

Within the two previous categories, the ensemble algorithms are the combination of predictions from various ML techniques applied to a single model improving predictive performance (Sagi y Rokach, 2018). In classification, an ensemble of classifiers is generally more accurate than the individual classifiers that compose it. Individual decisions are combined by weighted or unweighted votes in the classification of new examples (Hooftman et al., 2022), which allows a good balance between performance and computational cost (Telikani et al., 2022). The ensemble algorithms in regression improve accuracy while reduce bias and variance errors, avoiding over-adjustment when results deserve extra training (Ren et al., 2016b). Some outstanding algorithms in this group are adaBoost, bootstrap aggregation (Bagging), category boosting (CatBoost), extremely randomized trees, gradient boosting machines (GBM), RF, stacked generalization (Stacking).

The clustering algorithms are part of the unsupervised and semi-supervised learning methods, which allow grouping the data into sets of similar objects to maximize the intra-cluster similarity and minimize inter-cluster similarity (Ezugwu et al., 2022). The partitional clustering applies techniques to obtain a single partition by an objective function of the input data, in a fixed number of clusters, using iterative relocation clusters and resulting in the best configuration of the total number of executions (Nanda y Panda, 2014). The hierarchical clustering performs the division of data (root node) by a sequence of nested partitions, known as tree type structures (dendrograms). This approach follows a type of pattern agglomerated (from bottom to top) or by divisive clustering (from top to bottom), with no need to define the number of clusters in advance (Murtagh y Contreras, 2012).

The dimensionality reduction algorithms transform a high-dimensional data set into a representative lower-dimensional subset, as not all data features may be equally relevant for the problem at hand, greatly reducing computational complexity (Xu et al., 2019). This technique is widely used for data preprocessing, by two different ways: i) feature selection, in which the input features are combined to obtain a new dataset with a smaller number of new variables that retain the original information based on the input components and projection, and ii) feature extraction, in which the most relevant features of the original dataset are kept by removing those features that contribute little or nothing to the output features (Chhikara et al., 2020).

The anomaly detection algorithms try to find patterns, outliers or some kind of exception in the data that do not conform to the expected behavior (Chandola et al., 2009), by mean of a function that decide about the detection of an unknown or heterogeneous novelty present in the datasets with a class imbalance (Guansong et al., 2022). Isolation Forest, One-Class SVM, and PCA-Based Anomaly Detection are the most common algorithms to detect anomalies with application in crop

protection.

The association rule learning algorithms serve to find regularities present in parts of the dataset (descriptive rules) and generalize the dataset to enable predictions on new data (predictive rules) (Fürnkranz y Kliegr, 2015). These algorithms can identify an association rule in the form $A \rightarrow B$, based on the indicators support, confidence and lift. Support from $A \rightarrow B$ is the percentage of all items in A and B. Confidence is the percentage of A and B by the percentage of A. Lift indicates the probability of B occurring since A has occurred (Hashimoto et al., 2018). Within this category, the algorithms Apriori and Eclat are the most popular.

3.3.2 Artificial neural networks and deep learning models

The ANN algorithms are highly customizable and flexible computing models roughly inspired by biological neural networks, based on creating connected networks of simple processing units (neurons) that together can learn complex patterns and solve undefined problems. The ANNs works as universal approximators for any mathematical function, whose learning process is based on training from large datasets through sequential computations until accurate patterns are obtained. Then, when new patterns are presented, ANNs are able to predict them. These algorithms are mainly applied in tasks of classification and regression, e.g. in approximation functions (i.e. mapping multiple inputs to a single output), pattern classification (i.e. identification of new patterns through association and pattern recognition), associative memories (i.e. pattern recognition from limited information in the subset of data), and generation of new significant patterns, which can help in the reconstruction of patterns with greater characteristics (Schmidhuber, 2015).

Neural networks with two or more layers are the conceptual basis to generate DL models, whose progress has been spectacular in recent years in all disciplines, even in precision crop protection (Allmendinger et al., 2022; Farooq et al., 2019; Ferentinos, 2018; Hasan et al., 2021; Kamilaris y Prenafeta-Boldú, 2018; Rai et al., 2023; Rakhmatulin et al., 2021; Tugrul et al., 2022; Xia et al., 2018). DL algorithms transform data to construct complex concepts in a hierarchical structure with several levels of abstraction, so that the higher levels are composed of the characteristics of the lower levels (LeCun et al., 2015). The great potential of DL in many fields employing image analysis is allowing small data sets to be fitted to pre-trained models with different data, reducing training time and optimizing hardware resources (Kamilaris y Prenafeta-Boldú, 2018). DL covers different approaches suited to specific problems, for example, convolutional neural networks (CNNs) are used in computer vision and image classification, recurrent neural networks (RNNs) are used for prediction and language modelling, autoencoder is used in dimensionality reduction, and generative adversarial networks (GANs) are used in the generation of new images (Sarker, 2021).

CNN architectures for image classification is the most common application of DL in precision crop protection. The CNN algorithms find the features of objects of interest by self-learning from the image data, in contrast to traditional ML algorithms that require the user to establish such features (Hong et al., 2020). Performance of CNNs varies depending of number of parameters and convolutional layers (network depth), which in turn is directly constrained by the power of the available computing resources (Table 3.1). A broader application of CNN-based classifiers is object detection, which overcomes the issue of visual recognition in multi-class domains and object labelling in computer vision. Examples of CNN architectures for object detection and classification

implemented in crop protection include Region-based Convolutional Neural Network (R-CNN) (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2016a), You Only Look Once (YOLO) (Redmon et al., 2016), Single Shot Detector (SSD) (Liu et al., 2016), Feature Pyramid Networks (FPN) (Lin et al., 2017a), RetinaNet (Lin et al., 2017b) and Mask R-CNN (He et al., 2017).

Tabla 3.1: Characteristics of the Deep Learning architectures most commonly used in Crop Protection.

CNN Architecture	Depth (layers)	Million parameters	Top-5 Accuracy % *
LeNet-5 (LeCun et al., 1998)	5	0,06	-
AlexNet (Krizhevsky et al., 2017)	8	60	84.6
VGG-Net (Simonyan y Zisserman, 2014)	16	138.4	90.1
GoogLeNet (Szegedy et al., 2015)	22	4	92.2
ResNet (He et al., 2016)	152	60.4	93.1
Xception (Chollet, 2017)	126	22.8	94.5
DenseNet (Huang et al., 2017)	402	20.2	93.6
MobileNet (Howard et al., 2017)	55	4.3	89.5

* ImageNet validation dataset

3.4 Scientific impact and relevant contributions of ML in precision crop protection

An extensive bibliometric study of the Scopus database (www.scopus.com) revealed 107 traditional ML algorithms and 18 ANN models applied in all disciplines between 2010 and 2022, of which 105 and 17 algorithms, respectively, have been implemented in precision crop protection objectives with diverse degree of contribution in the domains of crop diseases, weeds and plagues (Table 3.2). SVM topped the list of traditional algorithms applied in precision crop protection objectives with >1,700 publications, followed by linear regression (LR) and Stacking with >1,500 publications each one. Principal Component Analysis (PCA), RF and DT are other algorithms with high impact reaching more than 1,100 publications each. A four group of relevant algorithms is formed by Bagging, logistic regression (LoR), k-NN and k-means clustering, which appear in more than 500 publications of precision crop protection. Some algorithms rank relatively high in terms of their use in precision crop protection in comparison to all disciplines (PCP/All), such as k-NN, simple linear iterative clustering (SLIC), stacking and stepwise discriminant analysis (SDA) (>10 % PCP/All), or in comparison to precision agriculture (PCP/PA), such as Gaussian Mixture Regression (GMR) (>70 % PCP/PA). Among the ANN models, convolutional neural networks (CNNs) are by far the most widely used in precision crop protection with >1,200 publications, mainly focused on detecting and classifying crop diseases, weeds or plagues with image-based technology, with ResNet, GoogLeNet and VGGNet being the most applied models, and to a lesser extent LeNet and Xception models (Figure 3.3).

Tabla 3.2: Numbers publications of machine learning algorithms according to the proposal taxonomy (source Scopus)

Algorithm	Task to be solved (†)						Number of ML Publications				PCP/PA(‡)
							In crop protection (PCP)			In PA	
	Clas	Regr	Clus	Anom	Dim	Asso	Diseases	Weeds	Plagues		
Traditional:											
Support Vector Machine (SVM)	✓	✓					612	560	540	>1,000	***
Linear Regression (LR)		✓					287	699	693	>10,000	***
Stacked Generalization (Stacking)	✓	✓					292	393	812	>1,000	***
Principal Component Analysis (PCA)	✓				✓		383	458	518	>10,000	***
Random Forest (RF)	✓	✓					374	437	395	>1,000	***
Decision Trees (DT)	✓						311	380	414	>1,000	***
Bootstrap Aggregation (Bagging)	✓	✓					195	356	414	>1,000	***
Logistic Regression (LoR)	✓						129	171	451	>1,000	***
k-Nearest Neighbours (k-NN)	✓						276	195	247	>1,000	***
K-Means Clustering			✓				210	185	152	>1,000	***
Hierarchical Clustering			✓				114	143	182	>1,000	***
Linear Discriminant Analysis (LDA)	✓						158	125	129	>1,000	***
Naïve Bayes	✓						146	95	169	>1,000	****
Regression Trees		✓					94	134	124	>1,000	***
Factor Analysis					✓		37	80	208	>1,000	***
Stochastic Gradient Descent	✓						117	84	74	>100	****
Partial Least Squares Regression (PLSR)		✓					101	127	40	>1,000	***
Support Vector Regression (SVR)		✓					75	75	84	>1,000	***
Expectation Maximization			✓				45	40	131	>1,000	***
Singular Value Decomposition (SVD)					✓		24	37	151	>1,000	***
LASSO		✓					39	44	118	>100	***
AutoEncoder					✓		53	46	102	>100	****
Multi Dimensional Scaling (MDS)					✓		58	68	70	>1,000	***
Self-Organizing Maps	✓						57	63	71	>1,000	***
Extreme Learning Machine (ELM)		✓					59	64	64	>1,000	***
Gaussian Mixture Model (GMM)			✓				49	46	91	>100	****
AdaBoost	✓	✓					58	46	74	>100	***
Fuzzy c-Means (FCM)			✓				61	59	46	>100	***
Partial Least Squares Discriminant Analysis					✓		63	53	19	>1,000	**
Fuzzy Clustering			✓				33	56	43	>100	***
Independent Component Analysis (ICA)					✓		22	22	85	>100	****
Ridge Regression (RR)		✓					15	25	77	>100	***
Extreme Gradient Boosting (xGBoost)	✓	✓					24	25	64	>100	***
Stepwise Regression		✓					24	40	39	>1,000	**
Quadratic discriminant analysis	✓						51	26	25	>100	****
Gaussian Process Regression (GPR)		✓					31	19	40	>100	***
Polynomial Regression		✓					15	33	41	>100	***
Principal Component Regression (PCR)		✓					33	31	22	>100	***
Boosted Trees (BoT)	✓	✓					22	20	25	>100	***
Simple Linear Iterative Clustering (SLIC)			✓				31	30	4	>100	****
Apriori						✓	8	14	42	>100	***
Subset Selection					✓		18	21	23	>100	***
Quantile Regression		✓					6	15	39	>100	***
Ordinary Least Squares (OLS) Regression		✓					6	13	41	>100	***
DBSCAN			✓				16	22	17	>100	***
Model Trees		✓					13	19	21	>100	**
Spectral Clustering			✓				5	12	35	>100	***
Gradient Boosting Machines (GBM)	✓	✓					10	9	28	>100	***
Poisson Regression		✓					2	14	30	>100	***
Multivariate Adaptive Regression Splines (MARS)		✓					10	16	20	>100	**
Minimum Spanning Trees			✓				3	11	27	>100	***
t-Distributed Stochastic Neighbor Embedding (t-SNE)					✓		13	6	21	>100	***
Stepwise Multiple Linear Regression (SMLR)		✓					17	16	5	>100	***
Stepwise Discriminant Analysis (SDA)							20	13	4	>100	****
Generalized Regression Neural Network (GRNN)		✓					12	12	12	>100	***

(Continued)

(†) Clas=Classification; Regr=Regression; Clus=Clustering; Anom=Anomaly Detection; Dim=Dimensionality Reduction; Asso=Association Rule Learning

(‡) ***** >50 %; **** >25 %; *** >10 %; ** >5 %; *>0.5 %; –No cases

Tabla 3.2: Continued

Algorithm	Task to be solved (†)						Number of ML Publications				PCP/PA(‡)
							In crop protection (PCP)			In PA	
	Clas	Regr	Clus	Anom	Dim	Asso	Diseases	Weeds	Plagues		
Maximum likelihood classifier (MLC)	✓						7	24	3	>100	**
One Rule	✓						6	6	21	>100	***
Kernel Principal Component Analysis (k-PCA)					✓		10	6	13	>10	****
One Class SVM				✓			8	5	16	>10	****
Gradient Boosted Regression Trees	✓	✓					11	11	4	>100	***
Quality Threshold			✓				6	7	12	>100	***
Gaussian Naive Bayes	✓						10	7	7	>10	****
Fisher's linear discriminant analysis	✓						10	4	9	>10	****
Fuzzy K-Means			✓				4	14	5	>100	***
Bagging Trees (BaT)	✓	✓					10	6	7	>100	***
Multiple-Kernel Learning (MKL)	✓						2	6	14	>10	****
Isomap					✓		5	1	15	>100	***
Kernel Ridge Regression (KRR)		✓					3	5	13	>10	***
Extremely Randomized Trees	✓	✓					6	7	7	>10	***
Rotation Forest	✓	✓					7	7	3	>100	***
Isolation Forest				✓			5	2	10	>10	****
Multinomial Naive Bayes	✓						2	2	13	>10	****
Laplacian Eigenmaps					✓		2	1	13	>10	***
Elastic Net Regression		✓					2	3	10	>10	***
LASSO Regularization		✓					1	3	11	>10	****
K-Medoids Clustering			✓				1	3	9	>10	***
Least-Angle Regression (LAR)		✓					2	3	7	>10	***
Mean Shift Clustering			✓				1	7	4	>10	****
Locally Weighted Regression (LWR)		✓					1	2	9	>100	**
FP-growth						✓	2	3	6	>10	****
Elastic Net Regularization		✓					1	2	8	>10	***
Zero-Shot Learning							3	2	6	>10	****
Locality Preserving Projections					✓		3	1	6	>10	***
Bayesian Network Classifier	✓						2	4	4	>10	****
Forward Feature Selection					✓		4	3	2	>10	***
Voting Classifier	✓	✓					2	3	3	>10	****
Decision Stump	✓						1	4	3	>10	***
Local Linear Embedding (LLE)					✓		5	1	2	>10	***
Ordinal Regression		✓					–	1	6	>10	**
Local Outlier Factor (LOF)				✓			–	–	6	>10	***
Gaussian Mixture Regression (GMR)		✓					–	–	6	>1	*****
Random Subspace Methods	✓	✓					1	2	2	>10	***
Category Boosting (CatBoost)	✓	✓					–	–	5	>10	**
Clustering Large Applications (CLARA)			✓				2	1	2	>10	***
DENCLUE			✓				2	1	2	>10	****
Ridge Regularization		✓					–	2	2	>10	***
Bayesian Linear Regression		✓					1	1	2	>10	**
Sammon Mapping					✓		1	1	2	>10	***
Eclat						✓	1	1	1	>10	***
Relevance Vector Regression		✓					–	–	2	>10	***
Bernoulli Naive Bayes	✓						–	–	2	>10	***
K-Modes Clustering			✓				–	–	2	>1	****
Regularized Linear Discriminant Analysis (RLDA)					✓		–	2	–	>10	***
Zero Rule	✓						–	–	1	>1	***
Gradient Descent Regression		✓					–	–	1	>1	****
Fast-MCD				✓			–	–	–	>1	–
PCA-Based Anomaly Detection				✓			–	–	–	>1	–
Artificial Neural Networks:											
Convolutional Neural Network (CNN)	✓	✓					528	395	339	>1,000	****
Back Propagation	✓	✓					190	176	189	>1,000	***
Radial Basis Function (RBF)	✓						149	135	167	>1,000	***
Recurrent Neural Network (RNN)	✓	✓					92	80	159	>1,000	***
Multi-Layer Perceptron (MLP)	✓	✓					65	66	83	>1,000	***

(†) Clas=Classification; Regr=Regression; Clus=Clustering; Anom=Anomaly Detection; Dim=Dimensionality Reduction; Asso=Association Rule Learning *(Continued)*

(‡) ***** >50 %; **** >25 %; *** >10 %; ** >5 %; *>0.5 %; –No cases

Tabla 3.2: Continued

Algorithm	Task to be solved (†)						Number of ML Publications				PCP/PA(‡)
							In crop protection (PCP)			In PA	
	Clas	Regr	Clus	Anom	Dim	Asso	Diseases	Weeds	Plagues		
Generative Adversarial Network (GAN)		✓					86	43	85	>100	****
Deep Belief Network (DBN)	✓	✓					46	35	45	>100	****
Probabilistic Neural Network (PNN)	✓						52	22	21	>100	****
Boltzmann Machine		✓					24	15	36	>100	****
Restricted Boltzmann Machine (RBM)		✓					18	9	29	>100	****
Stacked Autoencoder	✓			✓			8	12	13	>100	****
Learning Vector Quantization (LVQ)	✓						14	8	3	>100	***
Kohonen's Self-Organising Map (SOM)	✓						4	4	5	>10	***
Single-Layer Perceptron (SLP)	✓	✓					4	4	6	>10	***
Hopfield Networks		✓					3	2	8	>10	****
Bayesian Regularized Neural Networks	✓						–	–	4	>10	**
Supervised Kohonen Network (SKN)	✓						6	9	–	>10	*****
Counter-Propagation ANNs (CP-ANNs)	✓						–	–	–	>1	–

(†) Clas=Classification; Regr=Regression; Clus=Clustering; Anom=Anomaly Detection; Dim=Dimensionality Reduction; Asso=Association Rule Learning
(‡) ***** >50 %; **** >25 %; *** >10 %; ** >5 %; *>0.5 %; –No cases

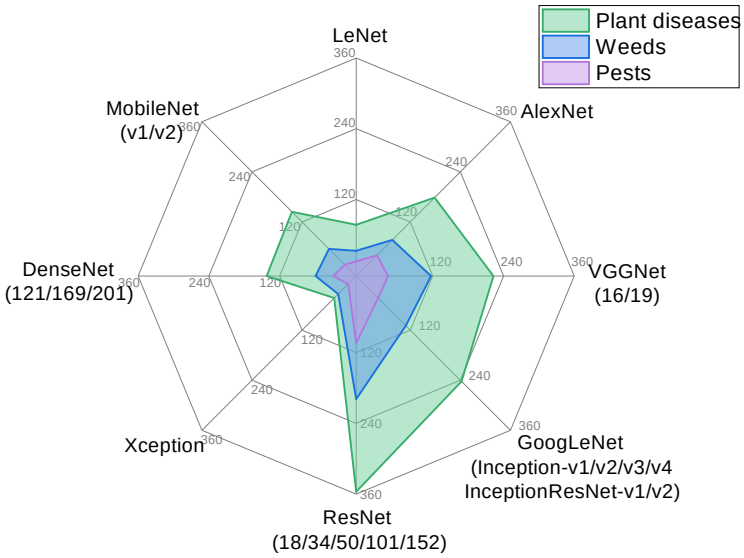


Figura 3.3: Number of publications of CNN architectures commonly used in the three domains of precision crop protection (crop diseases, weeds and crop plagues) from 2010 to 2022 (source: Scopus). Figure compiled with the conjunction of "CNN architecture" and each of the three crop protection domains (crop diseases, weeds and crop plagues) as search criteria within the article title, abstract and keywords.

A temporal analysis on ML-based publications shows that the adoption of ML algorithms has increased steadily year on year across all disciplines over the last decade (Figure 3.4a), which in turn is boosting the development of precision crop protection strategies (Figure 3.4b). Comparing the trends in both figures, peak values were reached in the last year in all cases, with classification and regression tasks being the most common by far in the group of traditional ML algorithms (55 % and 29 % across all cases and 47 % and 41 % in precision crop protection, respectively), followed by clustering, anomaly detection and dimensionality reduction tasks in the case of all disciplines, with

considerably less impact (11 %, 3 % and 2 %, respectively), and a negligible value for association rule learning. However, the dimensionality reduction algorithms were much more widely used in precision crop protection (11 %) than the other three categories. In the case of ANN algorithms, their use has increased significantly in the last five years, counting 29,956 (Figure 3.4a) and 759 new publications (Figure 3.4b) in 2022 across all disciplines and in precision crop protection, respectively. Compared to the traditional ML algorithms, ANN algorithms remain at the highest rates since 2018 across all disciplines, but still do not exceed traditional classification algorithms in precision crop protection, although they did overcome dimensionality reduction algorithms in 2019 and regression algorithms in 2022.

These positive indicators on the growing impact of ML in precision crop protection are supported by numerous applications and case studies outlined in detail in quite a few recent scientific reviews (Behmann et al., 2015; Chadha et al., 2021; Liakos et al., 2018; Muppala y Guruviah, 2020; Saleem et al., 2021; Wang et al., 2019a). An in-depth analysis of some relevant publications reveals key challenges addressed by diverse image-based or sensor technology together with ML algorithms in the specific domains of crop diseases (Table 3.3), weeds (Table 3.4) and plagues (Table 3.5), as discussed hereunder.

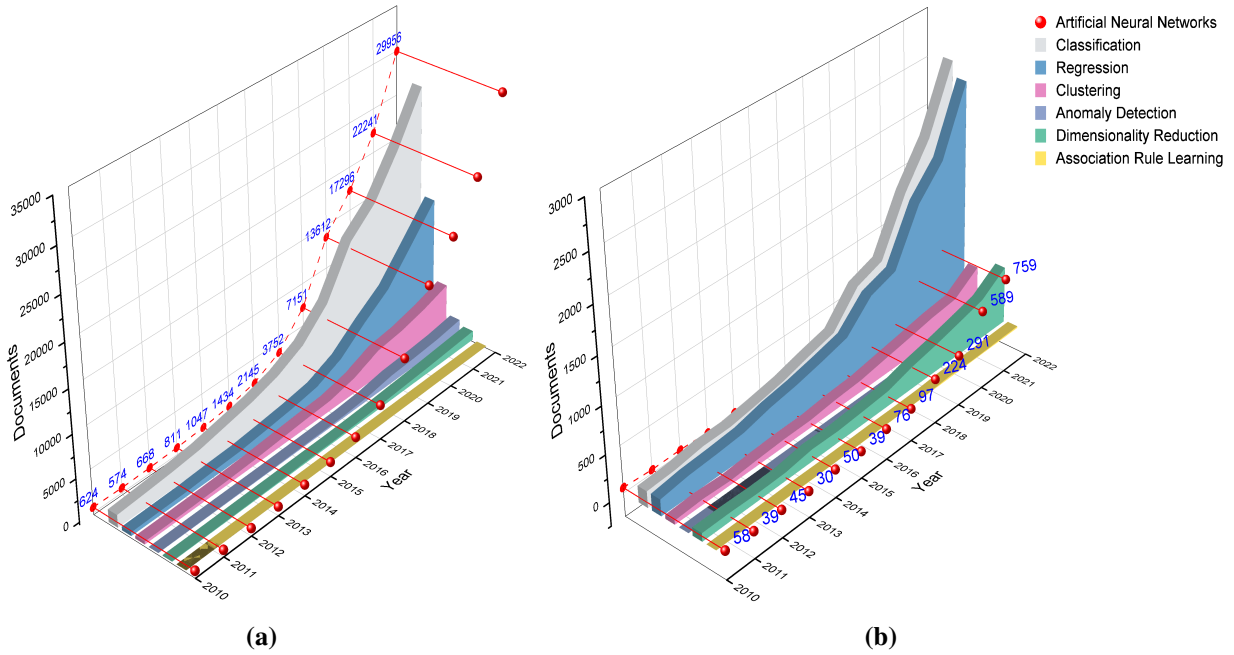


Figura 3.4: Publications trends (2010 – 2022) of traditional ML algorithms (colored solid areas) and ANNs (dashed red line) in all disciplines (A), and for precision crop protection applications (B), according to the proposed taxonomy (source: Scopus).

Tabla 3.3: Relevant investigations on ML algorithms in the domain of crop diseases.

Image/sensor technology	Crop / Pathogen type	Main objective	Task to be solved	ML Algorithm	Reference
Field spectroradiometer	Wheat / fungal	Detection and monitoring of powdery mildew (*Erysiphe graminis*)	Regression, ensemble	PLSR, SVM, RF	(Feng et al., 2022)
	Potato / fungal	Pre- and post-symptomatic detection of late blight (*Phytophthora infestans*) in leaves	Classification, ensemble	RF, PLS-DA	(Gold et al., 2020)
	Avocado / fungal, nutrient deficiency	Early and late detection of laurel wilt (*Raffaelea lauricola*), N defiency and Fe deficiency in leaves	Classification	DT, MLP	(Abdulridha et al., 2018)
	Tomato / bacterial, fungal	Discrimination of bacterial spots (*Xanthomonas vesicatoria*) among others fungal diseases (e.g. Late blight and target) with similar symptoms	Dimensionality reduction, classification	PCA, k-NN	(Lu et al., 2018)
	Strawberry / fungal	Asymptomatic and symptomatic detection of anthracnose crown rot (*Colletotrichum*)	Classification, regression	FDA, SDA, k-NN	(Lu et al., 2017)
	Avocado / fungal	Early and late detection of laurel wilt (*Raffaelea lauricola*) & phytophthora root rot	Classification	MLP, RBF	(de Castro et al., 2015)
On-ground hyperspectral camera	X Sugar beet / fungal	Early detection of rhizoctonia root and crown rot (*Rhizoctonia solani*) in leaves	Classification, regression, ensemble	PLS, RF, k-NN, Linear SVM, Radial SVM	(Barreto et al., 2020)
	Seed potatoes / viral	Real-time detection of potato virus y (*pvy, genus potyvirus*, family *potyviridae*) in tractor-mounted imagery	Classification	Fully CNN	(Polder et al., 2019)
	Wheat / fungal	Early detection of head blight (*Fusarium*)	Classification	VGG, RNN	(Jin et al., 2018)
	Tobacco / viral	Early (pre-symptomatic) detection of tobacco mosaic virus (tmv) in tobacco leaves	Classification, regression, ensemble	PLS-DA, RF, SVM, BPNN, ELM, LS-SVM	(Zhu et al., 2017)
Satellite multispectral and thermal images	Coffee / bacterial	Detection and progress of bacterial blight (*Pseudomonas syringae pv. Garcae*)	Classification, ensemble	RF, SVM, Naïve Bayes	(de Carvalho Alves et al., 2022)
Airborne hyperspectral and thermal images	Olive and almond trees / bacterial, fungal	Detection of *Xylella fastidiosa* (bacteria) and *Verticillium dahlia* (fungus) symptoms across species and pathogens	Classification, clustering	SVM, RF	(Zarco-Tejada et al., 2021)
	Olive trees / bacterial	Previsual symptoms detection of *Xylella fastidiosa* infection	Classification, ensemble	LDA, SVM, RBF, neural network ensemble	(Zarco-Tejada et al., 2018)
	Olive trees / fungal	Early detection and quantification of Verticillium wilt (*Verticillium dahlia*)	Classification	LDA, SVM	(Calderón et al., 2015)
Airborne hyperspectral & UAV-based multispectral images	Citrus trees / bacterial	Identification of Huanglongbing (HLB) with two aerial imaging systems	Regression, Classification	Stepwise regression, SVM, LDA, QDA	(Garcia-Ruiz et al., 2013)
UAV-based hyperspectral images	Wheat / fungal	Detection of yellow rust (*Puccinia striiformis* f. Sp. Tritici (pst)) across crop cycle	Classification, regression	ResNet, RF	(Zhang et al., 2019a)
UAV-based multispectral images	Apple trees / bacterial	Detection of apple fire blight (*Erwinia amylovora*)	Dimensionality reduction, anomaly detection, classification	mRMR, Isolation forest, DT, RF, SVM	(Xiao et al., 2022)
	Banana / bacterial, viral	Discrimination between Banana Xanthomonas wilt (BXW) and Bunchy top virus (BBTV) diseases	Classification, dimensionality reduction	VGG16, ResNet50	(Selvaraj et al., 2020)
	Pear trees / bacterial	Detection of fire blight (*Erwinia amylovora*)	Classification	SVM, RBF	(Bagheri, 2020)
Repository of RGB images of leaves	Grapes / fungal	Diagnosing *black rot*, *black measles* (esca) and *leaf blight* diseases in leaves for potential use in mobile devices	Classification	AlexNet, MobileNets, ShuffleNet	(Tang et al., 2020)
	Corn / fungal	Real-time detection of common rust and northern leaf blight damages in leaves	Classification	CNN	(Mishra et al., 2020)

Continued

Tabla 3.3: Continued

Image/sensor technology	Crop / Pathogen type	Main objective	Task to be solved	ML Algorithm	Reference
	Tomato / bacterial, fungal, viral	Real-time detection of tomato mosaic virus in leaves	Classification	AlexNet, SqueezeNet	(Durmuş et al., 2017)
On-ground RGNIr for leaves	Pear trees / bacterial	Detection of fire blight (*Erwinia amylovora*)	Classification	SVM, RBF	(Bagheri, 2020)

Tabla 3.4: Relevant investigations on ML algorithms in the domain of crop weeds.

Image/sensor technology	Crop / Weed species	Main objective	Task to be solved	ML Algorithm	Reference
Field spectroradiometer	No crop / *Sorghum halepense*	Differentiating glyphosate- resistant and susceptible Johnsongrass plants	Classification, regression, ensemble	k-NN, RF, SVM with FLDA	(Huang et al., 2022)
	No crop / *Amaranthus species*	Spectral discrimination of six Amaranthus species	Classification	SVM, Generalized Linear Model, DT, Naïve Bayes	(Sohn et al., 2021)
	No crop / *Cyperaceae* weeds	Spectral discrimination of *Cyperus esculentus* clones and morphologically similar weeds	Classification, dimensionality reduction	RF, regularized LoR, PLS-DA	(Lauwers et al., 2020)
	Wheat, broad bean / Cruciferous weeds	Selecting optimal spectral bands for image-based weed detection	Classification	MLP, RBF	(de Castro et al., 2012)
	Wheat / *Avena sterilis*, *Phalaris* spp.	Selecting suitable timeframe and spectral regions for discriminating wheat and two grass weeds	Classification, Dimensionality reduction	Stepwise discriminant analysis	(Gómez-Casero et al., 2010)
On-ground hyperspectral camera	Spring wheat, barley / *Kochia scoparia*	Differentiating glyphosate- and dicamba-resistant and susceptible Kochia plants	Classification	SVM with RBF kernel	(Nugent et al., 2018)
	No crop / *Amaranthus palmeri*	Differentiating glyphosate- resistant and susceptible Palmer amaranth plants	Classification, dimensionality reduction	MLC, FLDA	(Reddy et al., 2014)
	Rice / *Echinochloa crusgalli*, *Oryza sativa*	Discrimination of two weed species (Barnyard grass and weedy rice) with similar spectral signatures	Classification, regression, ensemble	RF, SVM, feature selection: successive projection algorithm (SPA).	(Zhang et al., 2019b)
	Maize / *Convolvulus arvensis*, *Rumex*, *Cirsium arvense*	Discrimination of three weed species	Classification, dimensionality reduction, ensemble	k-NN, RF, PCA	(Gao et al., 2018)
Satellite multi-spectral images	Winter wheat / Cruciferous weeds	Mapping cruciferous weed patches in multiple fields at broad scale	Classification	MLC	(de Castro et al., 2013)
UAV-based multi-spectral and/or RGB images	Wheat / blackgrass weed	Spectral analysis and mapping of blackgrass weed	Classification, dimensionality reduction	Feature selection, RF with Bayesian optimization	(Su et al., 2022)
	Sunflower, cotton / broad-leaved & grass weeds	Discrimination between broad-leaved and grass weeds	Classification	ANN-based MLP	(Torres-Sánchez et al., 2021)
	Vineyard / *Cynodon dactylon*	Detection of bermudagrass in complex scenarios with cover crop, bare soil and vines	Classification	DT	(de Castro et al., 2020)
	Sunflower, cotton / Several weeds	Early-season weed mapping between and within crop rows	Classification, ensemble	RF	(de Castro et al., 2018)
	Sunflower, maize / Several weeds	Selecting patterns and features for between and within crop-row weed mapping	Classification, clustering	K-means clustering, SVM	(Pérez-Ortiz et al., 2016)
	Sunflower / Several weeds	Comparing several ML paradigms to distinguish both weeds outside and within crop rows	Classification, clustering	k-means clustering, Linear SVM-based approximation, k-NN, SVM	(Pérez-Ortiz et al., 2015)

Continued

Tabla 3.4: Continued

Image/sensor technology	Crop / Weed species	Main objective	Task to be solved	ML Algorithm	Reference
On-ground RGB imagery	Tomato / Several weeds	Object detection and classification of five weed species	Classification	RetinaNet, Faster RCNN, YOLOv7	(López-Correa et al., 2022)
	Potato / *Chenopodium album*	Comparing CNN-based method to detect Chenopodium album in the crop field	Classification	GoogLeNet, VGG-16, EfficientNet	(Hussain et al., 2021)

Tabla 3.5: Relevant investigations on ML algorithms in the domain of crop plagues.

Image/sensor technology	Crop / Plague type	Main objective	Task to be solved	ML Algorithm	Reference
VNIR-SWIR spectroradiometer	Cotton / Worm	Modeling the spectral response of cotton plants under the *Fall armyworm* attacks	Classification	RF, DT, MLP, XGBoost, SVM, Naïve Bayes, LoR, k-NN	(Ramos et al., 2022)
Portable NIR spectroscopy & e-nose sensors	Wheat / Aphid	Detecting level of *Oat aphids* infestation and predicting insect number	Classification, regression	ANN-based regression models, Bayesian Regularization, SVM	(Fuentes et al., 2021)
UAV-based multispectral imagery	Cotton / Spider mite	Detection of two-spotted spider mite in crop fields	Classification	SVM, AlexNet	(Huang et al., 2018b)
RGB imagery from traps	No crop / Pest moth	Detecting *Helicoverpa assulta*, *Spodoptera litura* and *Spodoptera exigua* in pheromone trap images	Classification	Faster-RCNN ResNet, Faster RCNN Inception, R-FCN ResNet, RetinaNet ResNet, RetinaNet Mobile, SSD Inception	(Hong et al., 2020)
	No crop / Multi-class plagues	Detection and classification of multi-class plague species in trap images	Classification	VGG16, ZF, ResNet50, ResNet101	(Liu et al., 2019)
Repository of insect images	No crop / Multi-class plagues	Detection and classification of multi-class plague species in insect images	Classification	VGG19, SSD, Fast RCNN	(Xia et al., 2018)
On-ground RGB imagery	Tomato and pepper / Pest	Vision-based automated detection and identification of *Bemisia tabaci & Trialeurodes vaporariorum*	Classification	k-NN, MLP, SSD, Faster-RCNN	(Gutierrez et al., 2019)
	Strawberry / Thrips	Real-time detection of thrips (Thysanoptera) in flower images	Classification	SVM	(Ebrahimi et al., 2017)

In recent literature, one major goal is to study slight alterations in crop spectral information or other sensory components (e.g., odors or flavors) associated with pathogen infestations or with damages caused by a plague attack (Tables 3.3 and 3.5). This is generally done with on-ground measurements of plant leaves or canopies collected by hyperspectral cameras, field spectroradiometers or other portable sensors (e.g., e-nose sensor), and analyzing the spectral signatures or sensor data with ML classification and/or regression algorithms, aiming to discriminate between healthy and infested plants at the earliest possible stages or to model/predict the spectral response of infested plants. Dimensionality reduction algorithms (e.g., PCA, PLS-DA) is also often applied to transform large datasets into a lower dimensional space to facilitate further analysis. This approach was used at the disease domain, e.g., for early stage classification of anthracnose crown rot disease (by *Colletotrichum fungus*) in strawberry crop with SDA, FLDA and k-NN algorithms (Lu et al., 2017), classifying pre- and post- symptomatic fungal infestations of late blight (*Phytophthora infestans*) in potato leaves with PLS-DA and RF algorithms (Gold et al., 2020), monitoring the rate of fungal

powdery mildew (*Erysiphe graminis*) disease in wheat with PLSR, SVR and RFR algorithms (Feng et al., 2022), and pre-symptomatic detection of tobacco mosaic virus in tobacco leaves with PLS-DA, RF, SVM, BPNN, ELM and LS-SVM (Zhu et al., 2017); while at the plague domain was used, e.g., for predicting and classifying oat aphids (*Rhophalosiphum padi*) number in wheat cultivation with ANNs models applied to NIR and e-nose data (Fuentes et al., 2021), and spectral modelling of cotton plants against fall armyworm (*Spodoptera frugiperda*) attacks with RF, XGBoost, Naïve Bayes, LoR, SVM, MLP and k-NN algorithms (Ramos et al., 2022). These tools have also shown effective in other more complex scenarios dealing with hyperspectral discrimination of various diseases or other stresses/deficiencies that may cause similar symptomatology, such as fungal *Rhizoctonia* root and crown rot (*Rhizoctonia solani*) diseases in sugar beet leaves with PLS, RF, k-NN, and SVM (Barreto et al., 2020), bacterial spots (*Xanthomonas vesicatoria*) disease among other fungal diseases (late blight and target) in tomato leaves with PCA and k-NN algorithms (Lu et al., 2018), fungal laurel wilt (*Raffaelea lauricola*) and *Phytophthora* root rot diseases in avocado trees with ANN-based MLP and RBF models (de Castro et al., 2015), and laurel wilt disease against N and Fe nutrient deficiencies in avocado leaves with DT and MLP (Abdulridha et al., 2018).

At the domain of weed science (Table 3.4), field hyperspectral technology have been routinely tested to find the best spectral regions or vegetation indices to discriminate between weeds and crops at different phenological stages (Basinger et al., 2020; Peña-Barragán et al., 2006), generally with the aim of extrapolating results for remote sensing applications (de Castro et al., 2012; Gómez-Casero et al., 2010) in the context of site-specific weed management. Moreover, ML algorithms have recently dealt with challenging issues such as: 1) discrimination of multiple weed species with similar spectral response, such as Barnyard grass (*Echinochloa crusgalli*) and weedy rice (*Oryza sativa*) in rice crops with RF, SVM and SPA (Zhang et al., 2019b), *Convolvulus arvensis*, *Rumex*, and *Cirsium arvense* in maize crops with PCA, k-NN and RF (Gao et al., 2018), six *Amaranthus* species with SVM, DT and Naïve Bayes (Sohn et al., 2021) and *Cyperus esculentus* clones and morphologically similar weeds with RF, regularized LoR and PLS-DA (Lauwers et al., 2020), or 2) differentiation of herbicide- resistant and susceptible Palmer amaranth (*Amaranthus palmeri*) plants, Kochia (*Kochia scoparia*) plants or Johsongrass (*Sorghum halepense*) plants with MLC and FLDA (Reddy et al., 2014), SVM with RBF kernel (Nugent et al., 2018), and k-NN, RF and SVM with FLDA (Huang et al., 2022), respectively.

Disease, weed and plague detection and mapping with remote sensing have been particularly benefited from the adoption of ML algorithms (de Castro Megías et al., 2021; Lassalle, 2021; Roslim et al., 2021). In this context, proper selection of spectral and spatial image resolutions, as well as the optimal timing, is crucial to achieve satisfactory results (Khanal et al., 2017; Peña et al., 2015), which promotes the use of UAVs or manned aircrafts to the detriment of satellites in precision crop protection. Nonetheless, ML and satellite imagery can be useful in broad-scale applications, e.g., for evaluating integrated bacterial blight disease management in coffee plantations with several ecological variables (*Landsat-8* surface reflectance values and VIs, relief morphometry and hydrological attributes) by using RF, SVM and Naïve Bayes (de Carvalho Alves et al., 2022), or for mapping cruciferous weed patches in multiple winter wheat fields with *QuickBird* satellite imagery by using MLC (de Castro et al., 2013). Thermal and hyper-spectral aerial images with capability to capture slight variations in crop temperature and in narrow spectral bands associated to certain physiological indicators, respectively are commonly used in early detection of crop diseases, such as for identifying bacterial *Huanglongbing* (HLB) disease in citrus trees with stepwise regression, SVM, LDA and

QDA (Garcia-Ruiz et al., 2013), fungal Verticillium wilt (*Verticillium dahlia*) disease in olive trees with LDA and SVM (Calderón et al., 2015), bacterial *Xylella fastidiosa* infections in olive trees (Zarco-Tejada et al., 2018), and fungal yellow rust (*Puccinia striiformis*) across crop cycle in wheat with RF and CNN-based Inception-ResNet blocks (Zhang et al., 2019a). SVM with a Gaussian kernel and RF algorithms also helped to diminish the uncertainty of distinguishing trees affected by diverse biotic (i.e., infections by *Xylella fastidiosa* and *Verticillium dahlia* pathogens) and abiotic (i.e., water status) stressors that produce analogous symptoms on spectral traits in olive and almond orchards (Zarco-Tejada et al., 2021).

Most recent research in precision crop protection relies on analyzing UAV images collected with low-cost RGB cameras or multispectral imaging systems, which compromise image spectral resolution in favour of much higher spatial resolution. This ultra-high spatial resolution is particularly relevant to detect very small weed seedlings in their earliest stages, which is generally the optimal time for implementing SSWM strategies. In these scenarios, ML algorithms tackled previously unsolved challenging tasks such as: 1) distinguishing weeds outside and inside crop rows with k-NN, SVM or k-means clustering in sunflower (Pérez-Ortiz et al., 2015) and in maize (Pérez-Ortiz et al., 2016), or with an ensemble of RF trees in sunflower and cotton (de Castro et al., 2018); 2) discriminating between broad-leaved and grass weeds in sunflower and cotton by using ANN-based MLP Torres-Sánchez et al., 2021; 3) mapping bermudagrass patches in vineyards with cover crops by using DT (de Castro et al., 2020); and 4) spectral analysis and mapping of blackgrass weed in wheat parcels by using feature selection and RF with Bayesian optimization, respectively (Su et al., 2022). In the domains of crop diseases and plagues, relevant studies with UAV multispectral imagery are mainly focused on classifying crop/tree area damaged by a disease infestation or a plague attack, e.g., detecting bacterial fire blight (*Erwinia amylovora*) disease in apple or in pear trees with a combination of dimensionality reduction (mRMR), anomaly detection (isolation forest) and classification (DT, RF, SVM) algorithms (Xiao et al., 2022), or by using SVM classifier with RBF (Bagheri, 2020), respectively, discriminating bacterial (banana xanthomonas wilt - BXW) and viral (banana bunchy top virus - BBTV) diseases in banana plantations with the RetinaNet model based on the ResNet50 architecture as detector and the VGG16 architecture pre-trained with the ImageNet dataset as classifier (Selvaraj et al., 2020), and classifying cotton pixels affected by two-spotted spider mite attacks with the CNN-based AlexNet algorithm (Huang et al., 2018b).

Advances in CNN algorithms have greatly promoted the use of field imaging systems and proximal sensing for precision crop protection applications in the last years, as a tool to improve classification accuracy in complex crop/pest scenarios (Barbedo, 2020) and to implement real-time applications (Rakhmatulin et al., 2021). In fact, recent innovations in agricultural robotics and weeding systems are based on CNN classifiers for pest detection and classification (Allmendinger et al., 2022; Gerhards et al., 2022; Li et al., 2022; Oberti y Schmilovitch, 2021). Some recent studies in the weed domain are the classification of *Chenopodium album* in potato fields by comparing CNN-based GoogLeNet, VGG-16 and EfficientNet (Hussain et al., 2021) and of five different weed species in tomato fields with CNN-based RetinaNet, Faster RCNN and YOLOv7 (López-Correa et al., 2022), in the disease domain are the early detection of fungal head blight (*Fusarium*) disease in wheat by applying CNN-based VGG and RNN classifiers to on-ground hyperspectral images (Jin et al., 2018) and diagnosing of fungal *black rot*, *black measles* (esca) and *leaf blight* diseases by applying CNN-based AlexNet, MobileNets and ShuffleNet to a repository of RGB images of grape leaves for potential use in mobile devices (Tang et al., 2020), while in the plague domain are the detection

and classification of multi-class plague species in trap images by using CNN-based ZF, VGG16, ResNet50 and ResNet101 (Liu et al., 2019) and the detection of *Helicoverpa assulta*, *Spodoptera litura* and *Spodoptera exigua* moths in pheromone trap images by comparing Faster RCNN, R-FCN ResNet, Retinanet and SSD Inception classifiers (Hong et al., 2020), among many other case studies in the three crop protection domains.

3.5 Emerging technologies of precision crop protection in line with Ag5.0

Crop protection has used technology to reinvent itself over time, with AI tools and ML algorithms being the main drivers in the last decade towards the implementation of automated, smart and precise tasks following the precision agriculture and digital Ag4.0 paradigms. While AI involves the scientific and technological research of machines that are able to perceive, reason, learn, adapt, make decisions and act rationally to meet objectives in a given environment, the advances in ML are behind the recent rise of AI in primary, industrial and service sectors. As discussed above, many of the ML algorithms have already been successfully applied in agriculture and other disciplines (Table 3.2), while others unprecedented in agriculture are now reaching the level of maturity needed to address new precision crop protection goals in line with emerging Ag5.0.

These goals will primarily focus on developing and exploiting two issues: 1) early detection of crop pests, and 2) autonomous real-time multitasking systems. On the one side, the former will enable the application of more effective control measurements at the optimal time before the damage provoked by a disease, weed or plague becomes too severe. The development of data-driven early detectors is particularly urgent considering the adverse effects that current climate change scenario are causing on cropping systems due to the spread of newly emerging or invasive pests (IPPC Secretariat, 2021; Juroszek et al., 2020). To this end, the implementation of Ag5.0 technologies will facilitate data fusion from various sources and tools (e.g. climate data, proximal and remote sensing, crop and soil sensors, farm management information systems, etc.) and assess the spatio-temporal occurrence and severity of the pests (Shankar et al., 2020), which will lead to improve early detectors and diagnostic algorithms (Picon et al., 2019; Ramcharan et al., 2019). On the other side, the latter aims the design of powerful autonomous systems capable of simultaneously doing the three main stages of precision crop protection in real time (see section 2), i.e. identifying occurrences of crop diseases, weeds or plagues at different spatial and temporal scales, analyzing crop and pest information, and make the decision of applying a customized site-specific management adjusted to each crop-pest scenario (Birrell et al., 2020; Lottes et al., 2020; Pretto et al., 2019).

Ag5.0 technology will tackle these and other future challenges with a multidisciplinary domain that relies on powerful ML algorithms (Coulibaly et al., 2022; Liakos et al., 2018), along with the latest technological solutions on hardware (Bustio-Martínez et al., 2022), telecommunications (Chopra et al., 2017; Ejaz et al., 2020), and robotics (Albiero et al., 2022; Ren et al., 2020), which may contribute now or in the short, medium or long term given their different degrees of maturity and use (Figure 3.5), as discussed below.

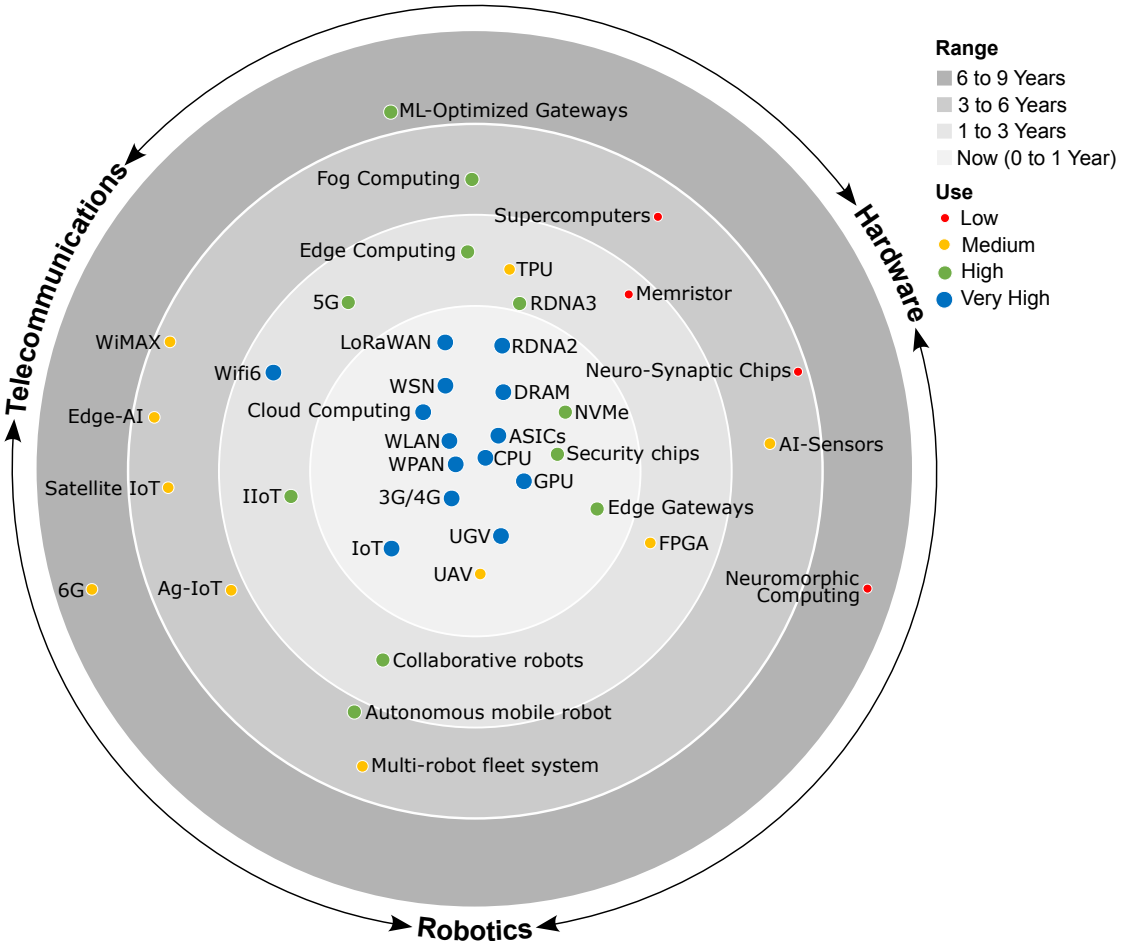


Figura 3.5: Multidisciplinary technological domain of Ag5.0 with a different degree of maturity and use ranging from mature technologies in the core circle to future technologies in the peripheral circle. CPU: central processing unit, GPU: graphics processing unit, TPU: tensor processing unit, DRAM: dynamic random-access memory, RDNA: radeon DNA, NVMe: non-volatile memory express, ASIC: application-specific integrated circuit, FPGA: field programable gate array, LPWAN: low power wide area network, WLAN: wireless local area network, WPAN: wireless personal area network, WSN: wireless sensor network, IoT: internet of things, IIoT: industrial IoT, Ag-IoT: agricultural IoT, LiFi: light fidelity, WiMAX: worldwide interoperability for microwave access, TSN: time-sensitive networking, xG: cellular network generation.

3.5.1 Hardware solutions for precision crop protection

Hardware tools are moving agriculture disciplines into digitization with innovative smart sensors, IoT ecosystems, architectures for specialized graphics processing, multicore embedded systems, and a number of new electronic devices, focused on the acquisition and use of crop data (Muhammad et al., 2019). The convergence of technologies is enabling to turn traditional agricultural sensors into smart sensors with built-in AI processing, that is, AI-Sensors with a dedicated chip embedded in the same sensor that can process ML tasks and, for example, may simultaneously perform object perception and analysis. Sony IMX500 and IMX501 (Sony Group Corporation, Tokyo, Japan) are two commercial image-based AI-sensors (Sony, 2020), in which the acquired signals are executed

with a digital image signal processor at high-speed by the logic chip (i.e., 3.1 millisecond processing by the MobileNet V1). This processing speed is feasible as the sensor generates semantic information belonging to the image metadata instead of the image information, reducing data volume. In crop protection, these AI-sensors would facilitate the detection, recognition and control of targeting areas of crops with specific pest problems in real-time (e.g., weed species identification) and following optimized DSS prescriptions.

Advances in architectures for specialized graphics processing, such as GPU, TPU, radeon DNA (RDNA), in dynamic random-access memories (DRAM) and in communication and storage access protocols (e.g., non-volatile memory express, NVMe) are enabling greater programmability, opening up a wide range of Ag5.0 applications based on virtual modeling, the creation of digital twins and the use of supercomputers. A digital twin is a multi-physics, multi-scale, probabilistic simulation of a complex system that uses the best available physical models and sensor updates to reflect the life of its corresponding twin (Glaessgen y Stargel, 2012). While simulation-based analysis within a digital twin will lead to the development of innovative and more powerful DSS tools for precise pest management, the use of supercomputers will enable the study of crop-pest models in less time and drastically improve the performance of ML detectors and classifiers of crop diseases, weeds or plagues. Currently, the bottleneck to implement CNN-based architectures with high capacity for knowledge generalization is the training stage with large training datasets, but supercomputers will assist in overcoming this weakness by increasing the input data (data augmentation) and decreasing the computational time for model creation.

Performance of CNN-based architectures can be also improved with the use of Field Programmable Gate Array (FPGA), which enables the implementation of logic functions and is the basis for the creation of multicore embedded systems (Qiu et al., 2016; Shawahna et al., 2019). This technology will benefit precision crop protection with the development of new software applications running in operating systems used in agriculture IoT (Ag-IoT) solutions (Zhang et al., 2020b) and the adaptation of customized pest detectors to mobile devices.

The devices connected to IoT systems are potentially risky in the absence of security elements or algorithms (Ibrahim y Gebali, 2022), reason why the devices performing edge gateway functions have improved designs with the use of application-specific integrated circuits (ASICs), just as security chips are essential in the implementation of Industrial IoT (IIoT) (Oñate y Sanz, 2023) and Ag-IoT systems. Wide evolution and adaptability of ML algorithms lead to their employ in optimizing gateway equipment tasks (ML-Optimized Gateways), making the performance of these tasks efficient even with resource limitations. The use of ML-Optimized Gateways in Ag5.0 will allow optimizing edge computing devices, reduce latency and increase privacy, which will result to create more efficient and safe models.

The future of both hardware and software solutions may reach a turning point in the medium term with the application of the computing principles derived from the memristors (Strukov et al., 2008). These devices are composed of two terminals with three layers, i.e. two electrodes for the communication of electrical signals and one storage layer that can be dynamically reconfigured when the inputs are stimulated, enabling data storage and direct processing (Zidan et al., 2018). The functioning of memristive elements is similar to that of neuronal synapses, becoming the technological basis of neuromorphic computing (Xia y Yang, 2019) and Spiking Neural Networks (SNNs) research (Jeong y Shi, 2019), which relies on a new neuron that is characterized by having a

time-varying internal state, known as spiking neuron (Brette et al., 2007; Ghosh-Dastidar y Adeli, 2009). SNNs are the artificial representation that most closely emulate the brain, differing from ANNs in the incorporation of time as an explicit dependency in computations (Davies et al., 2018). Comparing to ANNs, SNNs achieve lower latency classifications, shorter computation times in the training phase, high accuracies and low energy consumption (Diehl et al., 2015; Esser et al., 2016), which can foresee that neuromorphic computing and SNNs will be the future tools to develop computational systems and create new electronic devices with a high impact on Ag5.0 technology.

3.5.2 Telecommunications for precision crop protection

Precision crop protection is increasingly heading towards a system-of-systems approach with multiple connected practices to achieve an integrated crop management strategy, in which on-ground, proximal and remote sensing are key technologies to assess and monitor all the biotic and abiotic factors that might affect crop health. In this framework, telecommunications are essential to connect devices (i.e., platforms, processors, actuators) and transfer data acquired by sensors, creating a networking environment that adds value in the tasks of data processing, pest prediction, decision-making, and crop management.

Wireless Sensor Networks (WSN) are leading communication systems in agriculture with various technologies that differ from each other mainly in their operating mode and specifications in terms of frequency range, transfer rate and power consumption (Thakur et al., 2019). Bluetooth and Zigbee (developed under IEEE 802.15.1 and 802.15.4 standards, respectively) are characterized by open specification, short range operation, high level data transmission with low latencies and low power consumption (Khanji et al., 2019; Zeadally et al., 2019). Zigbee covers larger distance (<100 m) than Bluetooth (<10 m), although data transfer is faster in Bluetooth (1-24 Mbps) than in Zigbee (40-240 Kbps). The alternatives to increase the range of operation and data transfer are the wireless fidelity (Wi-Fi) system, generally used for local area networks with a range of 50-100 m or even several hundred meters, and the worldwide inter-operability for microwave access (WiMAX) system used as a long-distance communication solution (up to 50 km). The development of IoT and the advance of low power wide area networks (LPWAN) are promoting the Long Range (LoRa) radio communication system and the LoRaWAN protocol as the most promising technology in agricultural disciplines (Castro et al., 2023), because of its long-range data transmission (dozens of kilometers, very useful in rural areas), low power consumption and secure connectivity (Gu et al., 2020). LoRaWAN uses a modified frequency modulation, operates in the Industrial, Scientific and Medical frequency band defined according to the geographical area (Asia 433 MHz, Europe 868 MHz and America 915 MHz), hence the sensors can operate in the license-free bandwidth (Lavric, 2019).

High-speed and efficient telecommunications are essential to implement real-time operations in actuator platforms (i.e., tractors, self-propelled sprayers, unmanned ground vehicles (UGVs), UAVs, etc.) that are focused to simultaneously percept, analysis and treat pest occurrences. In engineering and computer science, the concept of real-time is given to those processes whose execution, measured as the ratio between the input and output of a variable, occurs at very low time values (<milliseconds), there being a difference between real-time system and real-time computer system (Poniszewska-Maranda et al., 2020). Cloud computing, edge computing and edge AI are the three technologies to implement real-time actions on actuator platforms for precision crop protection in line with Ag5.0.

Cloud computing is the convergence of information technology and business activity to provide services over the Internet. Companies such as Amazon, Google and Microsoft compete in the continuous improvement of infrastructures, hardware, computer security and high information processing (Mahmoud y Xia, 2019). To perform precision crop protection operations in real-time using cloud computing, the information collected with a sensorized platform must first be transmitted to the Internet, then processed and analyzed on any ML-based cloud service, and finally the prescription returned to the same platform to implement the actuation. These interactive operations need access times as short as possible, very close to real-time, to meet users' demands, for which network architectures for wireless connections enabling Internet access such as 5G are already underway, with a view to the upcoming development of 6G. For example, the integration of 5G and future 6G with UAVs has enormous potential to apply precise aerial treatments of weed patches and eventually other pest occurrences following real-time detection (Ullah et al., 2020). Technical aspects aside, security and privacy issues are of particular concern in cloud computing systems, as infrastructures and applications may be subject to malicious attacks, as reported by (Maniah et al., 2019) and (Sun, 2020). Indeed, privacy-sensitive reasons together with the progressive increase in data volume due to the connection of more devices has led to the introduction of fog computing, which allows decentralized processing, low latency and high bandwidth (Bonomi et al., 2012).

As mentioned before, current research is focused to platforms that detect, process and treat at the same time, which require a high computational cost in the limiting conditions of an equipment located on the farm, using the encoding method for signal transmission and taking into account the latency time of the radio transmission equipment. In this scenario, edge computing systems is a viable option as they allow the computing process to be performed close to the data source without the need for an Internet connection, thus avoiding data transmission problems and providing superior privacy and security, as well as reducing communication costs and energy consumption given the huge number of computations performed in ML modeling (García-Valls et al., 2018). A further step in the development of this computational architecture is offered by the devices for AI on the edge (Edge AI), which are embedded systems equipped with ML algorithms. Edge Intelligence is still at an early stage of research (Zhou et al., 2019), but is attracting great interest across all technological disciplines, with enormous potential in the development of agricultural robotics and autonomous crop protection treatments, since AI chips have achieved a high calculation capacity in the implementation of CNNs (Gao y Zhou, 2019).

3.5.3 Robotics for precision crop protection

Autonomous mobile robots (AMRs) allows the industry to increase productivity by doing more with fewer people, having great potential for boosting precision crop protection strategies in line with Ag5.0. AMRs have the ability to navigate with little or no human intervention under their control, in partially unknown environments (Alatise y Hancke, 2020). Therefore, their locomotion, perception, cognition and navigation systems must be able to address dynamic crops in position and time; in addition to: i) providing solutions to labor shortages, and ii) acquire real-time data for data-driven decision making, with the aim of significantly increasing yields within sustainable production(Shamshiri et al., 2018).

Several research projects have been developed to link robotic platforms to agricultural activities (Wolfert et al., 2017). In order to have completely robotized agricultural fields, robots must be

able to adapt to the external environment and to the different types of land surface. Due to the great technological advances implemented in recent years, some robotic agricultural activities are already becoming commercially available (Botta et al., 2022; Lowenberg-DeBoer et al., 2020; Saiz-Rubio y Rovira-Más, 2020; Santos Valle y Kienzle, 2020; Sparrow y Howard, 2021), being the use of UGVs and UAVs that detect weeds and act in real-time with high precision the most popular robotic system to implement a precision crop protection activity (Li et al., 2022; Oberti y Schmilovitch, 2021). Some AMRs with great potential are: 1) RIPPA (Australian Centre for Field Robotics, The University of Sydney, Austria), based on the design of their previous robot LADYBIRD, uses an intelligent perception system and is equipped with a variable injection precision applicator, with an operating autonomy of twenty one continuous hours (Bogue, 2016); 2) AgBot-II (Queensland University of Technology, Brisbane, Australia) with a vision system not only detecting but also classifying weed species in real time, then using the Inception-v3 architecture as its DDS, which allows to decide the weed management method to apply, either mechanical, chemical or a combination of both, weeds on accuracies over 90 % (McCool et al., 2018); 3) Robotti (Agrointelli, Aarhus, Denmark), whose module-based construction allows it to operate in various soil environments, adapting to different types of crops (Grimstad y From, 2017); 4) AVO (Ecorobotix, Yverdon, Switzerland) that uses CNNs algorithms for the detection and selective control of weeds by herbicide spraying in real time, obtaining a detection rate of 85 % (https://ecorobotix.com/en/avo/); 5) BONIROB (AMAZONE Technology Leeden GmbH & Co. KG, Germany) (https://info.amazone.de/DisplayInfo.aspx?id=29417) with an integrated system using camera-based machine vision, image processing to detect the plants and a sprayer with individually controlled valves, allows selective and precise control of weeds, thus achieving both ecological and economic advantages; 6) Kilter AX1 (Kilter AS, Norway) (https://www.kiltersystems.com/ax1) uses machine vision combined with AI and a novel nozzle technology that applies a micro-drop (6×6mm resolution), which allows to reduce the amount of herbicides up to 95 %; 7) DINO (Naïo Technologies, France) (https://www.naio-technologies.com/en/dino/), a weeding robot with an accuracy of 2 cm achieved by the RTK GPS system that has a vision system to detect the crop rows and adjust the position of the mechanical weeding tools in row, allowing high precision weeding and hoeing; 8) Odd.bot (Odd.Bot B.V., The Netherlands) (https://www.odd.bot/), a mechanical in-row weeding robot that relies on machine vision and AI-based seedling recognition; 9) Titan FT35 (FarmWise Labs Inc., USA) (https://farmwise.io/) uses machine vision and ML algorithms trained to learn the characteristics of crops such as broccoli, lettuce, cauliflower and tomatoes to differentiate between the crop and weeds; it has six internal weeders with blades that eliminate weeds with centimeter accuracy; and 10) FARMING GT (Farming revolution GmbH, Germany) (https://farming-revolution.com/) distinguishes weed seedlings with 99 % reliability in different crops (e.g., cabbage, lettuce varieties, onions, corn, sugar beet, pumpkin, field bean, potato, canola, soybean, wheat), then carrying out in-row and inter-row mechanical weeding.

Collaborative or cooperative robots will support the future development of Ag5.0 (Lytridis et al., 2021). These robots are designed to complement the routine activities by improving their ergonomics (Pauline et al., 2019) and also sharing the workspace. An advanced application of collaborative robots is in organic food production, particularly in pest control with nonchemical methods by using robotic mechanical control (Machleb et al., 2020) and viable handling systems for harvesting (De-An et al., 2011; Zhang et al., 2021), which has been shown as a solution to increase the benefits of organic crop management (Giampieri et al., 2022; Pérez-Ruíz et al., 2014). The development of Ag5.0 will allow

the convergence of UGV and UAV systems, for their collaborative and cooperative operation under a unified control, giving rise to Multi-robot Fleet Systems (MFS). Workload performed by several small robots composing a MFS is equivalent to that developed by a larger machine, highlighting that the MFS have a more precise positioning (Gonzalez-de-Santos et al., 2017).

Current technology has allowed the development and maturation of sensory-motor autonomy, reactive autonomy and cognitive autonomy in UAVs (Floreano y Wood, 2015), making them a great tool that together with RGB, multispectral, and hyperspectral sensors facilitate the acquisition of information on plant diseases, weeds, and plagues. That is why in Ag5. 0, detection and actuation systems based on ML algorithms and implemented in embedded systems will be part of the UAVs. ML techniques within Ag5.0 will allow the integral management of fleets of autonomous vehicles (UAV and UGV) decentralized in real time, besides being the basis for the implementation of robust navigation systems, such as the redundant system developed by (Belhajem et al., 2016) where they used ANNs in conjunction with genetic algorithms and the Extended Kalman Filter to reliably estimate the position of a vehicle in real time in the absence of GPS signal. The objective of having fleets of autonomous vehicles is the application of specific treatments for the detection and action on weeds and others pests (Emmi et al., 2014), which will finally reduce production costs and reduce the environmental impact of the use of herbicides and pesticides.

3.6 Conclusions

This article provides a framework on the future direction of precision crop protection, with a focus to scientific, agronomic and industrial applications of traditional ML algorithms and recent advances in the ANNs models. In the period 2010-2022, 125 algorithms applied in all disciplines were identified, of which 122 were used in the domains of crop diseases, weeds and plagues, with the aims of solving tasks on classification, regression, clustering, anomaly detection, dimensionality reduction, and association rule learning, and moving precision crop protection closer to the emerging concept of Ag5.0. This process should be accompanied by innovations and dedicated solutions in the areas of hardware, telecommunications and robotics, some of which are already being implemented in agriculture and others are still unprecedented, as this article outlines by introducing 39 emerging technologies and citing some 80 scientific and technical references. The transition from current Ag4.0 to future Ag5.0 strategies in the field of precision crop protection will be driven mainly by their focus and level of automation. Ag5.0 will promote a new era of intelligent crop management with a greater emphasis on solving complex crop protection objectives (e.g. early detection of crop pests) and enhancing management practices (e.g. autonomous real-time multitasking) as a whole, with a main focus to automatized decision-making processes, unmanned operations and progressively less human intervention supported by the latest AI systems, advanced robotics, and powerful ML algorithms.

Capítulo 4

Drone imagery dataset for early-season weed classification in maize and tomato crops

Publicación asociada a este capítulo

- Mesías-Ruiz G.A., Peña J.M., de Castro A.I., and Dorado J. (2024). Drone imagery dataset for early-season weed detection and classification in maize and tomato crops. *Data in Brief*. (Under review)

Abstract

Identifying weed species at early-growth stages is critical for precision agriculture. Accurate classification at the species-level enables targeted control measures, significantly reducing pesticide use. This paper presents a dataset of RGB images captured with a Sony ILCE-6300L camera mounted on an unmanned aerial vehicle (UAV) flying at an altitude of 11 meters above ground level. The dataset covers various agricultural fields in Spain, focusing on two summer crops: maize and tomato. It is designed to enhance early-season weed classification accuracy by including images from two phenological stages. Specifically, the dataset contains 31,002 labeled images from the early-growth stage—maize with four unfolded leaves (BBCH14) and tomato with the first flower bud visible (BBCH501)—as well as 36,556 images from a more advanced-growth stage—maize with seven unfolded leaves (BBCH17) and tomato with the ninth flower bud visible (BBCH509). In maize, the weed species include *Atriplex patula*, *Chenopodium album*, *Convolvulus arvensis*, *Datura ferox*, *Lolium rigidum*, *Salsola kali* and *Sorghum halepense*. In tomato, the weed species include *Cyperus rotundus*, *Portulaca oleracea* and *Solanum nigrum*. The images, stored in JPG format, were labeled in orthomosaic partitions, with each image corresponding to a specific plant species. This dataset is ideally suited for developing advanced deep learning models, such as CNNs and ViTs, for early classification of weed species in maize and tomato crops using UAV imagery. By providing this dataset, we aim to advance UAV based weed detection and mapping technologies, contributing to precision agriculture with more efficient, accurate tools that promote sustainable and profitable farming practices.

Tabla 4.1: Specifications Table

Specifications Table	
Subject	Computer Science. Agricultural Sciences.
Specific subject area	Computer Vision and Pattern Recognition. Agronomy and Crop Science.
Type of data	Raw.
Data collection	Images were acquired using a Sony ILCE-6300L camera (Sony Group C Tokyo, Japan) mounted on a UAV quadcopter model md4-1000 (microdrones GmbH, Siegen, Germany) flying at an altitude of 11 meters above ground level. This was conducted during the early-season of maize and tomato crops that were naturally infested with weeds, on clear, sunny days during midday hours. Orthophotos were generated and partitioned into 1000 × 1000 pixel images for easier handling using Phyton (Algorithm 4.1). Experts in weed science then visually identified and labeled the weed species with the labelImg software, and the dataset was stored according to the PASCAL VOC convention (Algorithm 4.2).
Data source location	Maize: CSIC Experimental Farm La Poveda, located in Arganda del Rey, Madrid, Spain (40°18'57.59"N, 3°29'22.57"W; datum WGS84). Tomato: Two commercial fields located in Santa Amalia, Badajoz, Spain (38°59'15.58"N, 6°02'57.71"W and 38°59'40.19"N, 5°57'17.54"W; datum WGS84).
Data accessibility	Repository name: DRONEWEED: DRONE imagery dataset for early-season WEED classification [Data set]; DIGITAL.CSIC Data identification number: doi.org/10.20350/digitalCSIC/16559 Direct URL to data: https://doi.org/10.20350/digitalCSIC/16559
Related research article	G.A. Mesías-Ruiz, J.M. Peña, A.I. de Castro, I. Borra-Serrano, J. Dorado, Cognitive computing advancements: Improving precision crop protection through UAV imagery for targeted weed monitoring. *Remote Sensing* 16 (2024) 3026. https://doi.org/10.3390/rs16163026

Value of the Data

- DRONEWEED dataset supports machine learning researchers in developing technological solutions to enhance the early classification of weed species affecting maize and tomato productivity.
- DRONEWEED dataset can be used to train, validate and test deep learning models, such as CNNs and ViTs, enhancing the robustness and accuracy of weed classification systems.
- The dataset supports computer vision tasks like multiclass classification, enabling exploration of synthetic image generation (e.g., GANs) for data augmentation and model improvement.

· The dataset includes all possible instances and, to our knowledge, is one of the largest publicly accessible datasets on weed species in maize and tomato crops in Spain.

4.1 Background

DRONEWEED dataset was created to provide an open, accessible, high-quality resource for early-season weed classification using UAV imagery. Accurately labeled datasets are crucial for the development of effective and practical deep learning applications (Mesías-Ruiz et al., 2023). This dataset annotated in JPG format, is particularly suited for developing advanced models like Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) for the multiclass classification of weed species at early-growth stages in maize and tomato crops (Mesías-Ruiz et al., 2024a; Mesías-Ruiz et al., 2024c). Other datasets, such as CoFly-WeedDB (Krestenitis et al., 2022) and DeepWeeds (Olsen et al., 2019) also include various weeds but in other domains and image settings. The CoFly-WeedDB dataset focuses on RGB imagery captured at 5 meters above a cotton crop in Greece and is limited to three common weed species (Johnsongrass, Purslane and Field bindweed), while the DeepWeeds dataset focuses on eight relevant weed species from grazing areas in Australia. In contrast, DRONEWEED dataset aims to enhance UAV-based weed classification and mapping in two major agricultural crops, driving precision agriculture in tomato and maize through more efficient, accurate and sustainable farming practices (Lati et al., 2021).

4.2 Data Description

This article presents a comprehensive dataset of imagery featuring weed seedlings and crops, collected during the early-season of maize and tomato in Spain. The dataset comprises 67,558 labeled images, organized into two primary folders: one for maize ('MAIZE') and one for tomato ('TOMATO'). Each folder is further divided into subfolders corresponding to two phenological growth stages: early (e.g. 'MAIZE_1' and 'TOMATO_1') and late (e.g. 'MAIZE_2' and 'TOMATO_2'). Within each phenological stage, there are additional subfolders for the weed species associated with each crop, as well as specific folders for the crops themselves. For instance, each 'MAIZE' folder contains eight subfolders: seven for the associated weed species (e.g. 'MAIZE_#_atriplex' for *Atriplex patula* L.; 'MAIZE_#_chenopodium' for *Chenopodium album* L.; 'MAIZE_#_convolvulus' for *Convolvulus arvensis* L.; 'MAIZE_#_datura' for *Datura ferox* L.; 'MAIZE_#_lolium' for *Lolium rigidum* Gaud; 'MAIZE_#_salsola' for *Salsola kali* L.; and 'MAIZE_#_sorghum' for *Sorghum halepense* (L.) Pers.) and one for maize images (e.g. 'MAIZE_#_maize'). Similarly, the 'TOMATO' folders contain four subfolders: three for the associated weed species (e.g. 'TOMATO_#_cyperus' for *Cyperus rotundus* L.; 'TOMATO_#_portulaca' for *Portulaca oleracea* L.; and 'TOMATO_#_solanum' for *Solanum nigrum* L.) and one for tomato images (e.g. 'TOMATO_#_tomato'). These subfolders (organized by crop, phenological stage and species) are provided in ZIP format to facilitate easy downloading. The labeled images, saved in JPG format, have varying resolutions dimensions encompassing the entire specimen. Each file is named according to the species (whether weed or crop), the phenological stage (1 for early, 2 for more advanced) and a consecutive number within each subfolder. Fig. 4.1 shows samples of weed species and crops at both phenological stages.

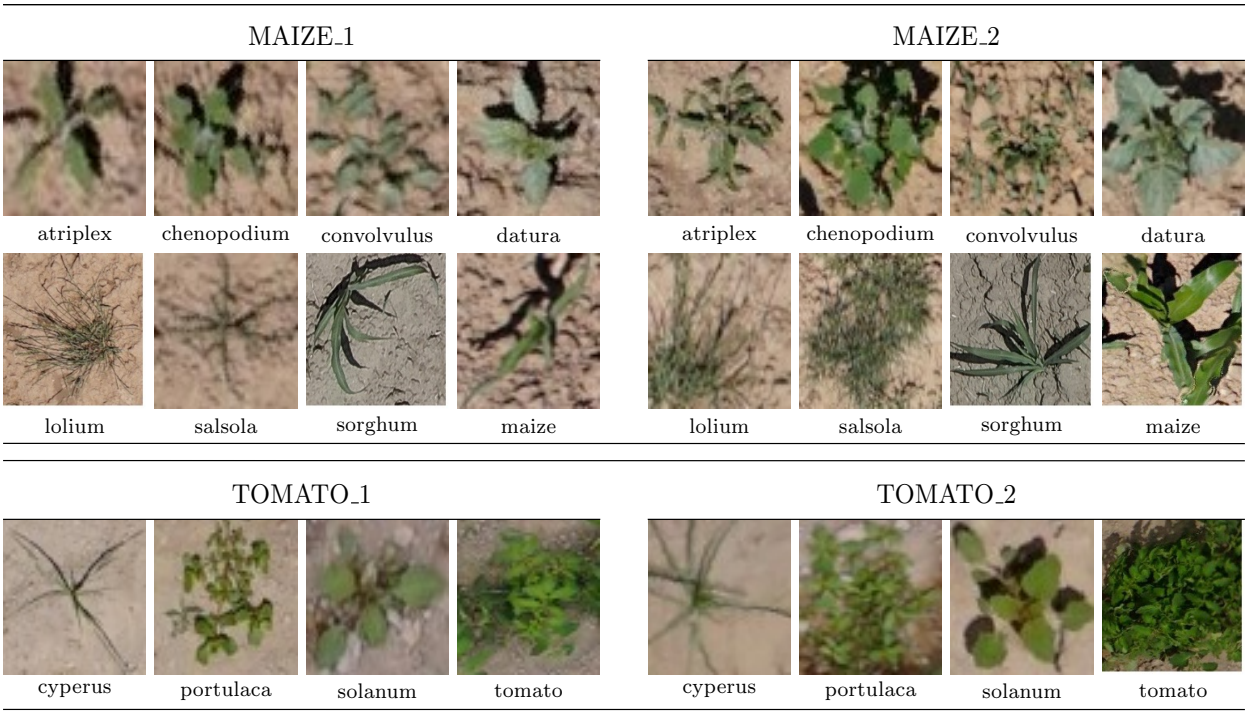


Figura 4.1: Examples of labeled images in maize crop in BBCH14 (MAIZE_1) and BBCH17 (MAIZE_2) phenological stages, as well as in tomato crop in BBCH501 (TOMATO_1) and BBCH509 (TOMATO_2) phenological stages. Species in MAIZE are: (atriplex) *Atriplex patula*, (chenopodium) *Chenopodium album*, (convolvulus) *Convolvulus arvensis*, (datura) *Datura ferox*, (lolium) *Lolium rigidum*, (salsola) *Salsola kali*, (sorghum) *Sorghum halepense*, and (maize) *Zea mays* L. Species in TOMATO are: (cyperus) *Cyperus rotundus*, (portulaca) *Portulaca oleracea*, (solanum) *Solanum nigrum*, and (tomato) *Solanum lycopersicum* L.

4.3 Experimental Design, Materials and Methods

4.3.1 Field data collection

Images were acquired from two distinct locations in Spain (Fig. 4.2): the CSIC Experimental Farm La Poveda in Arganda del Rey, Madrid (40°18'57.59"N, 3°29'22.57"W; datum WGS84), where maize was cultivated, and two commercial tomato fields in Santa Amalia, Badajoz (38°59'15.58"N, 6°02'57.71"W and 38°59'40.19"N, 5°57'17.54"W; datum WGS84). Sampling was conducted during the crop early-season at two different phenological stages of maize: the four unfolded leaves (BBCH14) on 18/05/2020, and the seven unfolded leaves (BBCH17) on 27/05/2020. Similarly, tomato crops were sampled in two fields with different growth stages: first flower bud visible, (BBCH501) and ninth flower bud visible, (BBCH509), on 01/06/2021 and 02/06/2021, respectively.

A commercial RGB camera, the Sony ILCE-6300L (SonyGroup Corporation, Tokyo, Japan), was mounted on a vertical take-off and landing (VTOL) quadcopter UAV, model md4-1000 (microdrones GmbH, Siegen, Germany), to capture images at a fixed altitude of 11 meters above ground level, resulting in a ground sampling distance (GSD) of 0.17 cm per pixel. This high resolution made it possible to capture fine details in the morphological differences of weed species, even in early

stages of growth or in conditions of competition with the crop. The resolution setting was calibrated according to the desired geographic coverage and topographic complexity of the study area, considering that the crops analyzed were located in flat and homogeneous areas. A 70 % overlap was maintained both laterally and frontally during image capture, ensuring full coverage of the study area. The camera was equipped with an APS-C Exmor® CMOS sensor (23.5 × 15.6 mm) with 24.2 effective megapixels, producing images with a resolution of 6,000 × 3,376 pixels. In all cases, image acquisition was performed during zenithal sunlight, which provided optimal illumination conditions by minimizing shadows and maximizing homogeneous illumination on surfaces.

Figura 4.2: Sampling locations in Spain included the maize crop at the CSIC Experimental Farm in Arganda del Rey, Madrid (yellow label); and the tomato crop in commercial fields in Santa Amalia, Badajoz (red label).

4.3.2 Generation of orthomosaics

The generation of an orthomosaic involved stitching together multiple overlapping UAV images to produce a single high-resolution image that covered the entire area of interest. For maize, 568 and 565 images were used for the BBCH14 and BBCH17 growth stages, respectively, while 895 and 950 images were used for the BBCH501 and BBCH509 growth stages in tomato, respectively, to create orthomosaics of the study fields. This process was carried out using Agisoft PhotoScan software (Agisoft LLC, St. Petersburg, Russia), which utilized information extracted from the RGB channels. Geospatially referenced control points (six to seven, depending on the field) were strategically placed in the experimental fields to ensure accuracy. The orthomosaic corrected distortions caused by the camera angle and varying image perspectives, enabling precise analysis of the objects within the image and facilitating accurate identification and classification of weed species.

4.3.3 Image partitioning

Due to the large size of the generated orthomosaics (e.g. 61,000 × 41,175 pixels for maize at BBCH14, and 72,304 × 34,574 pixels for tomato BBCH501), the images were partitioned into smaller, more manageable sections for individual analysis. This partitioning was carried out using Python and the 'rasterio' library (Algorithm 4.1). The algorithm automatically subdivided the orthomosaics into smaller fragments of 1000 × 1000 pixels, facilitating species recognition and labeling. This approach enhanced processing efficiency and reduced computational complexity for subsequent analyses.

Algorithm 4.1 Orthomosaic partitioning

Input: Orthomosaic directory ($orthomosaic\_directory$), Output directory ($output\_directory$), Partition size (m_size)
Output: Partitioned images
function GETCOORDINATESTOPLEFT($orthomosaic\_directory$)
    open the orthomosaic using Rasterio
    $transf \leftarrow$ dataset transformation
    calculate the coordinates of the upper left-hand corner
    return coordinates of the upper left corner
end function
function PARTITIONORTOMOSAIC($orthomosaic\_directory$, $output\_directory$, m_size)
    create an empty list for storing the partition data
    $width$, $height \leftarrow$ width and height of the dataset
    $coordinates(i, j) \leftarrow$ GETCOORDINATESTOPLEFT($orthomosaic\_directory$)
    for each i from 0 to $width$ with a step of m_size **do**
        for each j from 0 to $height$ with a step of m_size **do**
            create a window for the $partition[coordinates]$
            read partition using Rasterio
            obtain $\leftarrow partition[coordinates]$
            create a PIL image from the partition
            save the image in the $output\_directory$ with a name based on the $coordinates$.
        end for
    end for
end function

4.3.4 Species labeling

To ensure reliability and consistency in the labeling process, several key measures were implemented. Prior to UAV image acquisition, a field identification phase was conducted by a team of weed experts, who characterized the weed species present in the maize and tomato crops. Subsequently, image labeling was performed by several experts, which reduced individual bias and improved consistency in species identification. In cases where the identification of weed species in the partitioned images was complex or ambiguous, the experts chose to omit the labeling of these images, thus ensuring the quality and accuracy of the labeled data included in the DRONEWEED dataset. Annotations were made manually. This process involved drawing bounding boxes around each plant and labeling the various objects visible in each subdivided image (i.e. 1000 × 1000 pixels). The labeling was

performed using the open-source graphical tool labelImg (Tzutalin, 2015). Only whole plants were labeled, while those divided by image boundaries were excluded (Fig. 4.3). The final output was a collection of individual images, each linked to a specific label corresponding to the identified plant species. Each label was saved in PASCAL VOC convention format (Fig. 4.4), facilitating structured storage (Algorithm 4.2) and easy retrieval for use in modeling.

Figura 4.3: Example of labeling using labelImg software.

```xml
<annotation>
    <folder>tech4Agro</folder>
    <filename>zona_f1_1792.jpg</filename>
    <path>/Volumes/Datos_GME/Datasets/data_mh_tomato/zona_f1_1792.jpg<
    <source>
        <database>DRONEWEED</database>
    </source>
    <size>
        <width>1000</width>
        <height>1000</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>cyperus</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>571</xmin>
            <ymin>345</ymin>
            <xmax>663</xmax>
            <ymax>404</ymax>
        </bndbox>
    </object>
```

Figura 4.4: Example of file generated during labeling following the PASCAL VOC convention format.

Algorithm 4.2 Cropping and saving labels by species

Input: XML files
Output: processed data stored in different directories
Define: the directories for each class (species)
species_directories $\leftarrow$ { 'lolium' $\rightarrow$ $path\_lolium$, ...}
for each PASCAL VOC file **do**
    read coordinates of bounding box
    $coordinates \leftarrow [ymin : ymax, xmin : xmax]$
    read the class value
    $class \leftarrow$ *class value*
    crop the image
    $imagen\_cut \leftarrow imagen[coordinates]$
    save the cropped image
    if $class$ exists in *species_directories* **then**
        save $imagen\_cut$ in directory *species_directories*
    end if
end for

4.3.5 Organization and Storage of the Dataset

The dataset was organized into separate folders based on crop type, phenological stage and weed species (Table 4.2). This structure organization facilitates easy reuse by researchers aiming to test different deep learning model architectures or develop new image classification techniques for

precision agriculture. The dataset compilation process included the following automated steps: i) reading the coordinates of weed bounding boxes from the PASCAL VOC format file; ii) extracting weeds from the subdivided images; iii) assigning a unique identifier to each weed corresponding to its specific image section; iv) classifying weed species according to their labels; and v) storing the images in separate directories based on their respective labels. This process was meticulously executed due to the high sensitivity of DL models to both the quality and consistency of the input data. Ensuring that the dataset is diverse, representative, and of sufficient volume is crucial for achieving robust and accurate model performance.

Tabla 4.2: Number of labeled images for each crop, phenological stage and species included in the database

	Early-growth stage	**Late-growth stage**
	MAIZE_1 (BBCH14)	**MAIZE_2 (BBCH17)**
Atriplex patula	1,000	1,459
Chenopodium album	1,200	2,175
Convolvulus arvensis	1,200	1,102
Datura ferox	683	589
Lolium rigidum	1,000	80
Salsola kali	1,200	1,216
Sorghum halepense	1,600	103
Maize	12,364	24,614
	TOMATO_1 (BBCH501)	**TOMATO_2 (BBCH509)**
Cyperus rotundus	3,090	134
Portulaca oleracea	1,875	177
Solanum nigrum	1,900	2,175
Tomato	3,890	2,732
Total labels	31,002	36,556

4.3.6 Utility of the dataset

In (Mesías-Ruiz et al., 2024c), synthetic images were generated using GANs from the DRONEWEED dataset, obtaining representations of weeds in various morphological configurations and emulating the variations present in advanced stages of growth. This approach not only increased the volume of the dataset, but also introduced new combinations of visual patterns, which improved the accuracy of the models in identifying weeds in complex and varied situations.

4.3.7 Future perspectives

The DRONEWEED dataset was designed to serve as a basis for future research aimed at implementing weed detection models capable of dynamically adapting to seasonal and environmental

variations. This will facilitate continuous and accurate monitoring in crop protection using artificial intelligence platforms and implementation of precision agriculture strategies.

Limitations

The dataset is constrained by the geographical and climatic conditions specific to the study areas in Spain, which may limit the generalizability of the trained models to other regions or crops. Additionally, the dataset primarily focuses on early-growth stages, potentially limiting its applicability for later-growth stages. To overcome this limitation, a useful strategy would be to collect supplementary data from regions with contrasting climatic and soil conditions. For example, integrating data from areas with Mediterranean, semi-arid, or temperate climates could increase the robustness and adaptability of trained models, allowing them to recognize patterns in a wider range of agroecological environments. This would not only improve classification accuracy in other geographic areas, but also enhance the model's ability to adapt to diverse agricultural conditions.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms

Capítulo 5

Weed species classification with UAV imagery and standard CNN models: Assessing the frontiers of training and inference phases

Publicación asociada a este capítulo Weed species classification with UAV imagery and standard CNN models: assessing the frontiers of training and inference phases

- Mesías-Ruiz G.A., Borra-Serrano I., Peña J.M., de Castro A.I., Fernández-Quintanilla C., & Dorado J. (2024). Weed species classification with UAV imagery and standard CNN models: Assessing the frontiers of training and inference phases. *Crop Protection*, doi : 10.1016/j.cropro.2024.106721.

Abstract

Accurate weed species identification is crucial for effective site-specific weed management (SSWM), enabling targeted and timely control measures for each weed in crop field. This study advanced the current approach to species-level weed identification during the early growth stage by integrating unmanned aerial vehicles (UAVs) imagery with standard convolutional neural networks (CNNs) models such as VGG16, Resnet152 and Inception-Resnet-v2. For this, a robust dataset was created with 33,467 labels of weeds (*Atriplex patula*, *Chenopodium album*, *Convolvulus arvensis*, *Cyperus rotundus*, *Lolium rigidum*, *Portulaca oleracea*, *Salsola kali*, *Solanum nigrum*) and crops (maize, tomato), which was subjected to different training, validation and test scenarios. Model inputs were adjusted in order to align them with the information represented by the UAV images. Initially, models were developed in balanced scenarios, gradually increasing label numbers to assess their performance. Inception-ResNet-v2 achieved over 90 % accuracy with 400 labels, while ResNet152 and VGG16 required 600 and 800 labels, respectively, for similar accuracy. In a more complex and realistic scenarios with unbalanced datasets, Inception-ResNet-v2 outperformed, likely due to its deeper architecture and enhanced capability to capture intricate features and patterns within UAV images. The study emphasized the importance of the minority-to-majority species ratio in unbalanced datasets, which affects minority species classification. To prevent misclassification, it is crucial to determine the right number of labels for CNN model training and validation. Weed maps were generated after species classification using the Faster R-CNN algorithm as an object detector. This advancement in methodology facilitates the precise and efficient implementation of SSWM techniques.

5.1 Introduction

Site-specific weed management (SSWM) is an agronomic strategy that focuses on the precise location, identification and targeted management of weed populations in specific aereas, rather than treating the entire field uniformly (Fernández-Quintanilla et al., 2018; Lati et al., 2021). Accurate and quick identification of weed species in an early-stage is one of the key components of SSWM, since different species require distinct control measures (Fernandez-Quintanilla et al., 2022; Sa et al., 2018). Weed species identification is often a challenging and time-consuming process that demands skilled personnel. Since, many weed species possess similar physical characteristics, incorrect identification and treatment decisions may be made, leading to a decrease in control efficacy.

The digitization of agriculture is a contemporary approach that employs technology to enhance efficiency and productivity. This approach integrates various technologies, such as sensors, unmanned aerial vehicles (UAVs), artificial intelligence (AI), and big data analysis, to enhance crop yields while minimizing the environmental impact of crop production through a reduction in pesticide use. UAVs have great potential in agriculture owing to their capacity to rapidly and effectively gather large amounts of highly detailed imagery over crops and fields, thus improving spatial and temporal data that facilitates crop scouting, particularly the detection of weeds, pests and diseases (Rejeb et al., 2022). Deep Learning (DL), a subfield of Machine Learning (ML), has demonstrated particular success in image processing, improving the accuracy of classification tasks and reducing variability in regression problems (Mesías-Ruiz et al., 2023). Currently, DL techniques such as convolutional neural networks (CNNs) outperform traditional ML techniques in object recognition

(Ioffe y Szegedy, 2015). CNNs are a type of algorithm inspired by the neural mechanisms underlying the visual system, making them well-suited for computer vision and image recognition tasks. CNNs have been particularly successful in these areas because they are specifically designed to handle high-dimensional input data, such as images, and are able to automatically learn relevant features from the data.

DL algorithms have been recently applied to UAV-based image analysis with some restrictions regarding the number of classes, i.e. aiming to generate a segmented database containing only general classes such as crop, soil, shadows, and two groups of grass and broadleaf weeds (dos Santos Ferreira et al., 2017; Torres-Sánchez et al., 2021). Indeed, (dos Santos Ferreira et al., 2017) utilized the CaffeNet architecture, a derivative of AlexNet, to classify weeds among crops, achieving an average classification accuracy of 99.1 % and 99.5 % for balanced and unbalanced datasets, respectively, using a dataset of 15,336 segmented crop labels, including 3,249 soil samples, 7,376 soybean samples, 3,520 grass samples and 1,191 broadleaf weed samples captured at a height of four meters. (Bah et al., 2018) used a supervised dataset of UAV-captured images at 20 meters height, consisting of 30,105 weed images (thistles and young shoots) and 46,736 crop images (beans and spinach), employing the ResNet18 architecture for discriminatory analysis, resulting in a notable area under the curve value of 0.957. (Huang et al., 2020) focused on weed identification, classification and mapping in two rice fields mainly infested by *Leptochloa chinensis*, *Cyperus iria*, *Digitaria sanguinalis* and *Echinochloa crus-galli*, using a UAV dataset captured at 10 meters height, employing four pre-trained CNNs (AlexNet, VGG-16, GoogLeNet, and ResNet-101) as classifiers. The study demonstrated accuracy performances of 86.8 %, 88.4 %, 87.2 %, and 86.6 %, respectively, for discriminating between weeds and rice plants. (Khan et al., 2021) devised an optimized semi-supervised framework using generative adversarial networks to classify weeds and crops, utilizing a dataset of 2,800 weed images and 2,600 pea and strawberry crop images captured at a flight altitude of two meters, achieving classification accuracy nearing 90 %.

Some advancements have been achieved in the classification of weed species. For instance, (Chen et al., 2022) employed transfer learning with 27 DL models on a dataset of 5,187 images representing 15 distinct weed species at various growth stages, with the ResNet101 model demonstrating superior performance in weed species identification by achieving the highest *F1-score* of 99.1 %. Precise identification of weed species is fundamental to improve control techniques, as evidenced in the work of (Valente et al., 2022), where they emphasize the importance of accurately mapping of *Rumex obtusifolius* for implementing effective control strategies. By utilizing images captured at a 10-meter altitude, they achieved an *F1-score* of 78.4 % using the pre-trained MobileNet model. Additionally, the integration of diverse CNN architectures has enabled significant advances in solving various computer vision tasks. An example is the study by (Shahi et al., 2023), where segmentation models such as SegNet, DeepLabV3+ and UNet, were combined with classification architectures such as VGG16, ResNet50, DenseNet121, EfficientNetB0 and MobileNetV2. The most noteworthy outcome was an *F1-score* of 88.2 %, achieved through the combination of EfficientNetB0 and UNet. The study successfully automated the semantic segmentation of weeds like *Sorghum halepense*, *Convolvulus arvensis* and *Portulaca oleracea* within a cotton crop, utilizing a dataset (Krestenitis et al., 2022) acquired by flights at 5 meter height.

After reviewing the literature and evaluating the performance of various models in image classification tasks, the present study focused on three standard CNN architectures, selected based on their

performance and wide adoption in the scientific community: VGG16 (Simonyan y Zisserman, 2014), ResNet152 (He et al., 2016), and Inception-ResNet-v2 (Szegedy et al., 2017). These architectures offer a diverse set of features and depths, enabling a comprehensive evaluation of their performance.

The main objective of this research was to investigate the constraints and limitations associated with the training and testing phases when developing a methodology for classifying weed species in early-stage field conditions using CNN architectures and UAV imagery. Indeed, the development of an automated and accurate method to detect and classify weed species at an early growth stage can support SSWM systems by enabling specific crop protection measures, reducing crop management costs, and minimizing herbicide-related environmental impact. In particular, the study addressed four specific objectives: 1) Determine the minimum number of labels, i.e. optimal dataset size, required to achieve an acceptable CNN model performance with an accuracy, precision and recall value above 90 %. 2) Evaluate the performance, efficiency and suitability of three CNN models in classifying weed species using a balanced dataset. 3) Examine the impact of an unbalanced dataset, specifically focusing on scenarios where a) there is a higher number of crop labels, and b) there are varying numbers of labels for both crop and weed species, on the accuracy of weed species classification and the overall performance of the CNN models. And 4) Implement an object detector integrated with the most effective classification model to assess its performance in accurately locating and identifying weed species for mapping purposes within a field. This research contributes to the progress of CNN models for UAV image classification and detection of weed species, offering valuable insights for future studies in this field.

5.2 Materials and methods

Figure 5.1 provides an overview of the methodology to achieve the proposed objectives. Firstly, UAV flight plans were designed to allow automated flights, ensuring a consistent approach to image capture. Secondly, the preprocessing stage involved the creation of orthomosaics, image splitting and species labeling. Thirdly, three CNN-based models were trained using both the training and validation datasets to perform the classification task. Subsequently, the generalization of knowledge to new data was assessed using widely-used classification metrics. Finally, the best classification model was integrated with an object detection architecture to automatically identify weed species and locate them in the orthomosaics.

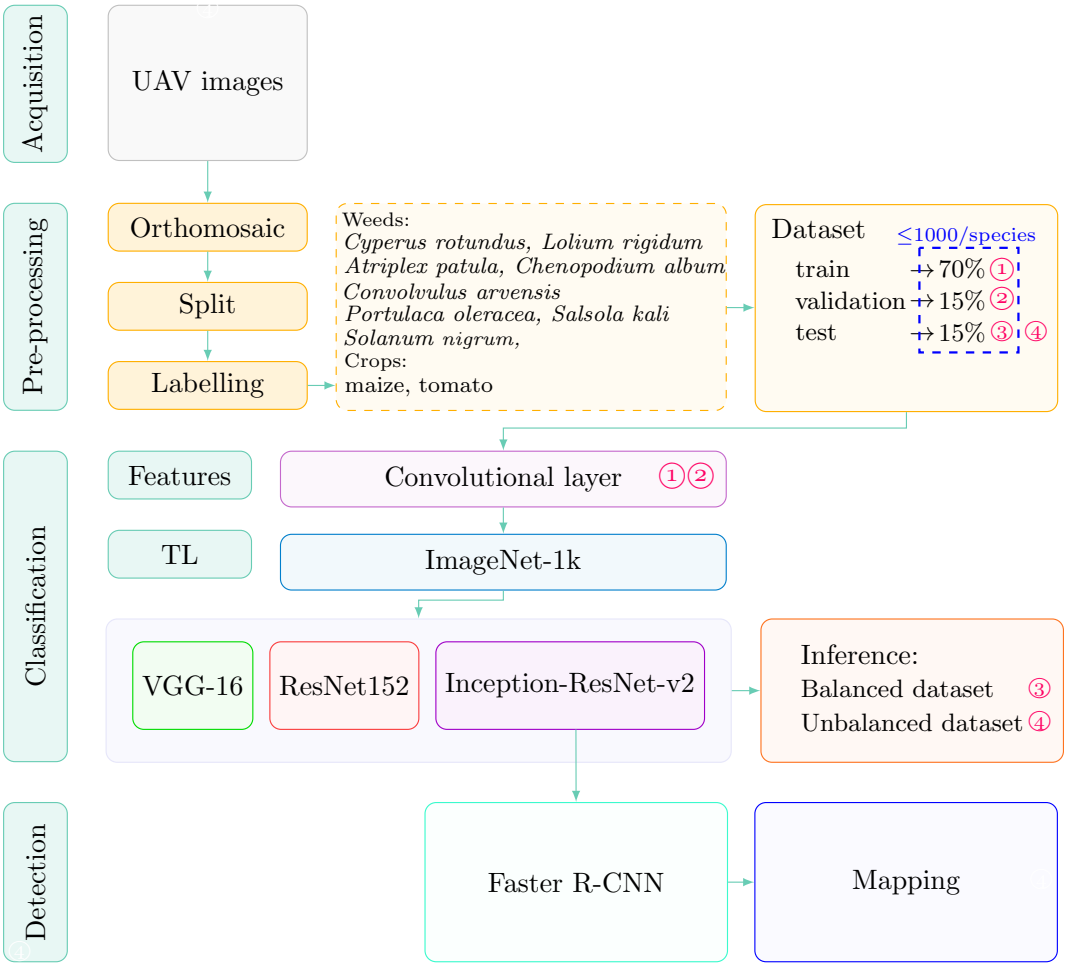


Figura 5.1: Outline of the proposed methodology for the classification of weed species in early growth stage of maize and tomato crops.

5.2.1 Acquisition of UAV images

Two different locations were used in this study. On the one side, UAV images of an irrigated maize crop were acquired at the CSIC experimental farm La Poveda, Arganda del Rey, Madrid, Spain (40°18'59.25"N 3°29'21. 53"W), at an early stage of crop growth (4 leaves unfolded, BBCH 14. (Meier, 2018). The area has a Mediterranean Continental climate characterized by cold winters, hot summers, and an average yearly rainfall of around 400 mm. The soil type is sandy loam, with a plow layer containing 1.6 % organic matter. An area of around 0.75 hectares was sampled, which consisted of conventionally tilled maize. The maize was grown with rows spaced 0.75 m apart and a density of 85,000 plants per hectare. On the other side, UAV images an irrigated tomato crop were acquired from a commercial farm located in Santa Amalia, Badajoz, Spain (38°59'39.0"N 6°03'31.7"W), also at an early stage of the crop growth (first flower bud visible, BBCH 501. (Meier, 2018). Soil in this farm is predominantly sandy and has low organic matter content. Approximately 1.2 hectares were surveyed, comprising conventionally tilled tomato plants. The tomato crop was cultivated with 1.5-meter row spacing and a plant density of 25,000 per hectare.

A low-cost RGB camera, model Sony ILCE-6300L (Sony Group Corporation, Tokyo, Japan), was mounted on the quadcopter UAV, model md4-1000 (microdrones GmbH, Siegen, Germany) for image acquisition. The camera featured an APS-C type Exmor® CMOS sensor (23.5 × 15.6 mm) with 24.2 effective megapixels. The UAV flight plan parameters were set using the mdCockpit application (microdrones GmbH, Siegen, Germany). The configuration specified a constant horizontal speed of 2 m/s at a fixed altitude of 11 m above ground level, resulting in a ground sampling distance of 0.17 cm per pixel. The acquired images had an overlap ratio of 70 % laterally and 70 % frontally, with each image measuring $6,000 \times 3,376$ pixels.

5.2.2 Preprocesing

Orthomosaic

Using the acquired imagery together with geospatial data extracted from strategically placed control points in the experimental fields (seven and six for the maize and tomato, respectively), we systematically generated geomatics products with Agisoft PhotoScan software (Agisoft LLC, St. Petersburg, Russia). These products enabled the creation of orthomosaics, incorporating the information extracted from the RGB channels. Consequently, the generation of an orthomosaic involved stitching together multiple overlapping UAV images to produce a single high-resolution image that covers the entire area of interest.

Split

The orthophotos were subdivided into smaller sections to enhance processing efficiency and reduce computational complexity in subsequent analysis. This subdivision was prompted by the substantial size of the orthomosaics, which size were $61,000 \times 41,175$ pixels (5.3 GB) for maize and $72,304 \times 34,574$ pixels (6.3 GB) for tomato, making them unwieldy for direct processing. This procedure was executed by implementing an algorithm that automatically partitioned the orthophoto into split images of $1,000 \times 1,000$ pixels. The algorithm was developed in Python, using the necessary libraries, such as rasterio for geospatial data management and PIL (Pillow) for image manipulation. After accessing the orthomosaics, the geospatial coordinates of each upper left corner were obtained. Iteration was performed on the partitions defined on the *x* and *y* axes. Subsequently, an image was generated and saved as a JPG file. This partitioning process facilitated the recognition and labelling of distinct species.

Labelling

Experts labelled the weed species by manually drawing bounding boxes and annotating several visible objects each split image. This procedure was carried out with the free software graphical tool labelImg (Tzutalin, 2015). The following weed species were labelled to compile the dataset: 1) *Solanum nigrum* (4,175 labels); 2) *Cyperus rotundus* (3,500); 3) *Portulaca oleracea* (1,900); 4) *Chenopodium album* (1,200); 5) *Convolvulus arvensis* (1,200); 6) *Salsola kali* (1,200); 7) *Atriplex patula* (1,017); and 8) *Lolium rigidum* (1,014). In addition, 12,364 maize plants and 6,622 tomato plants were labeled. The annotations made on the split images were saved with XML extension in PASCAL VOC format. Dataset generation included the following steps: 1) reading the XML file coordinates that generate a square delimiting the weed; 2) extraction of the weeds on the split

images; 3) assignment of a distinct identifier to designate its corresponding split area; 4) classifying the weed species according to their labels; and 5) storing them in separate directories based on their respective labels. The final result is a collection of individual images corresponding to each label, i.e. each plant of every species. This generated dataset is available at (Mesías-Ruiz et al., 2024b).

Dataset

DL models are highly sensitive to quality and consistency of the source dataset, which should be diverse, representative, and in sufficient quantity to ensure robust and accurate performance. Since there is no prior research on the optimal size of training, validation, and test subsets for developing a DL model, we addressed this by dividing the dataset into multiple subsets to test the models across various scenarios. Initially, we relied on the number of labels for the least frequent weed species in the source dataset, namely *Lolium rigidum* and *Atriplex patula*, each with slightly more than 1,000 labels. Several balanced datasets of different sizes were created: 100, 200, 400, 600, 800 and 1,000 labels per weed species. This strategy was implemented to avoid bias towards the majority species. A constant distribution of 70 % for training, 15 % for validation and 15 % for test was maintained for the implementation of each selected classification model.

Tabla 5.1: Distribution of labels of the unbalanced datasets: Test I with crop as the predominant species, and Test II with a varying number of labels for each species.

	Unbalanced dataset	
Species	**Test I**	**Test II**
Cyperus rotundus	150	2,500
Lolium rigidum	150	14
Atriplex patula	150	17
Chenopodium album	150	200
Convolvulus arvensis	150	200
Portulaca oleracea	150	900
Salsola kali	150	200
Solanum nigrum	150	3,175
Maize	[150; 4,500]	11,364
Tomato	[150; 4,500]	5,622

Additionally, two tests (Tests I and II) were conducted using unbalanced datasets (Table 5.1) to evaluate the generalization ability of the models in scenarios that better reflect real-world conditions. These tests were performed on the models developed using a balanced dataset, where the maximum number of labels per weed species (i.e. 700 for training and 150 for validation) was utilized. Test I was designed to simulate a scenario where crop labels dominate the dataset. To conduct this analysis, the relationship between the number of crop and weed labels was analyzed by adjusting the number of crop labels at various ratios: 1:1, 1:2, 1:10, 1:20, and 1:30. This enabled us to achieve a range of crop labels from 150 to 4,500. Alternatively, Test II aimed to assess the models' ability to generalize

on an imbalanced dataset with varying numbers of labels for each crop and weed species. Based on the results of Tests I and II, the optimal correlation between species imbalance was identified to achieve classification metrics that align with our objective. This involved evaluating the frontiers at which a decrease in model performance occurred as a result of the number of species labels. In this study, separate datasets were analyzed for the maize-weed and tomato-weed cropping systems.

5.2.3 Classification

Models

The three models used for classification task, namely VGG16, ResNet152, and Inception-ResNet-v2, have unique architectures characterized by varying numbers of layers and connectivity patterns. This study explored the key features of these models, evaluating their strengths and limitations, and the factors that differentiate them from other CNN architectures.

The VGG16 model is known for its architectural simplicity and performance improvements due to the implementation of the VGG block, which increases the depth of CNNs. Specifically, the VGG16 model comprised of 16 convolution layers with a 3×3 kernel convolution, five Max pooling layers with a 2×2 size, and three fully connected layers after the last Max pooling layer. ReLu activation was used in all layers, and the final layer employed a softmax classifier.

The ResNet152 model is an architecture that introduces residual connections to facilitate the training of very deep networks. By adding layers to the network, its complexity and expressiveness are increased. The introduction of the residual block in ResNet152 addresses a common issue encountered in other architectures, where shallower convolutional networks tend to perform better than their deeper counterparts. This is due to problems such as vanishing gradients and the curse of dimensionality, where increasing the number of layers may not always improve performance. The residual block overcomes this challenge by enabling information to skip certain layers and directly pass to the next layer through a residual connection, thereby improving the performance of the network with deeper layers (He et al., 2016).

The Inception-ResNet-v2 model was created by experimenting with different convolution kernel sizes ranging from 1×1 to 11×11 in order to achieve optimal performance. Through this process, it was discovered that utilizing a combination of kernels of varying sizes was most effective, leading to the development of the Inception block. Compared to other blocks, the Inception block is more computationally efficient due to its reduced parameter count, while also generating more expressive features by combining features produced by kernels of differing sizes (Szegedy et al., 2015).

Training, validation and test

Training, validation and test of the CNN models were submitted to the Keras-Tensorflow DL framework (Abadi et al., 2016). The Ubuntu 20.04.4 LTS operating system and Python 3.9 were used on a workstation with an Intel® Core (TM) i7-12700K CPU, a NVIDIA GeForce RTX 3070 Ti Graphic Processing Unit (GPU) with 8GB of memory. To optimize the classifiers parameters, the expected input of the models was modified by adjusting it to the average size of our label set. For the VGG16 and ResNet152 models an image size of 64×64 pixels was used, while for Inception-ResNet-v2 the image size was 75×75 pixels. After loading the models, the transfer learning technique

(Pan y Yang, 2010) was employed with the pre-trained ImageNet weights, replacing the last fully connected layer by a new fully connected layer with 9 neurons. The hyper-parameters for all three models were identical, with a batch size of 32, 30 epochs, adaptive moment estimation (Adam) like as optimizer with a learning rate of 10^{-5}, and categorical cross-entropy serving as the loss function. Leveraging the advantages of using transfer learning in image pre-training, we hypothesized that having a sufficient number of images for each class (e.g., up to 1,000 images) within a balanced dataset could obviate the need for data augmentation (Khalifa et al., 2021; Shawky et al., 2020) in our study. Such an approach would facilitate our research objective of evaluating and comparing the performance of the CNN-based models while preserving the inherent quality of the original dataset derived from our UAV imagery. However, exploring data augmentation techniques is a potential area for future research. The optimal training accuracies for all case studies were achieved between 22 and 27 epochs.

Evaluation

The evaluation of the classification was carried out through a confusion matrix to better understand the performance of the models. The matrix includes four predictions categorized as follows: T_P = true positive, T_N = true negative, F_P = false positive, and F_N = false negative. The criteria of "type I error"(false positive) and "type II error"(false negative) are used to perform a detailed analysis on the performance of the models. Additionally, four different metrics commonly used in ML were derived from the four predictions of the confusion matrix.

1. *Accuracy:* determines the degree of approximation of the measurements of a quantity to the real value, it can lead to errors in the quality of the model when the class imbalance is high.

$$Accuracy = \frac{T_P + T_N}{T_P + F_P + F_N + T_N} \tag{5.1}$$

2. *Precision:* represents the positive prediction value, which measures the quality of the model.

$$Precision = \frac{T_P}{T_P + F_P} \tag{5.2}$$

3. *Recall:* measures the ability of the model to identify true positives.

$$Recall = \frac{T_P}{T_P + F_N} \tag{5.3}$$

4. *F1-score:* Calculates the harmonic mean between precision and recall, providing a balance between the two metrics. It is a general score that represents the model overall performance, indicating its ability to retrieve relevant results. *F1-score* is especially useful for unbalanced classes.

$$F1 - score = \frac{2T_P}{2T_P + F_P + F_N} \tag{5.4}$$

5.2.4 Object detection

The detection and localization of weed species in the two fields was carried out by applying the Faster R-CNN algorithm as an object detector. This algorithm is composed of two fundamental modules: a fully convolutional network in charge of proposing regions of interest (RPN) within the image, and the Fast R-CNN detector that executes the detection of objects within these proposed regions (Ren et al., 2016a). The implementation of the Faster R-CNN detector was carried out using the TensorFlow object detection API. The Adam optimizer was used, with an image input size of 640×640 pixels, and a minimum confidence threshold of 0.4 was set to validate the detections.

Model performance evaluation was performed individually for maize and tomato crops, using the images acquired in subsection 5.2.2, which were adjusted to the model input size. After obtaining the predictions and locations of the objects of interest, the local coordinates of the bounding boxes were extracted and aligned on the orthophoto. These translated coordinates were exported to a CSV file, which simplified the generation of weed maps and their visualization on the orthomosaic.

For the evaluation of maize, 30 frames of $1m^2$ with their corresponding georeferencing, randomly distributed in the field, were located on the orthophoto. These frames were used to carry out the field verification of the weed species detected. For the tomato crop, the model results were compared with approximately 200 plants of each weed species (i.e. *Cyperus rotundus*, *Portulaca oleracea* and *Solanum nigrum*), previously labeled on the orthophoto by experts. The percentage detected and correctly classified was calculated. These approaches allowed a rigorous evaluation of the discriminative capacity of DL architectures applied to agricultural fields.

5.3 Results

5.3.1 Inference of the CNN models for classification as a function of dataset size

As expected, performance of the studied CNN models improved as the number of labels increased, with the highest accuracy obtained with 1,000 labels for each weed species in both crops (Figure 5.2). Inception-ResNet-v2 achieving the highest mean accuracy of 98.6 % for the both crops types, while ResNet152 and VGG16 had mean accuracies of 95.6 % and 92.6 %, respectively. These models achieved high accuracy with minimal variance (less than $\pm 0{,}13\,\%$) in error means. An acceptable threshold over 80 % accuracy was achieved with only 200 labels, although the three CNN models exceeded 90 % accuracy with 800 labels per weed species. Inception-ResNet-v2 consistently outperformed the other models, achieving over 90 % accuracy with just 400 labels, while ResNet152 and VGG16 required 600 and 800 labels, respectively, to reach similar accuracy.

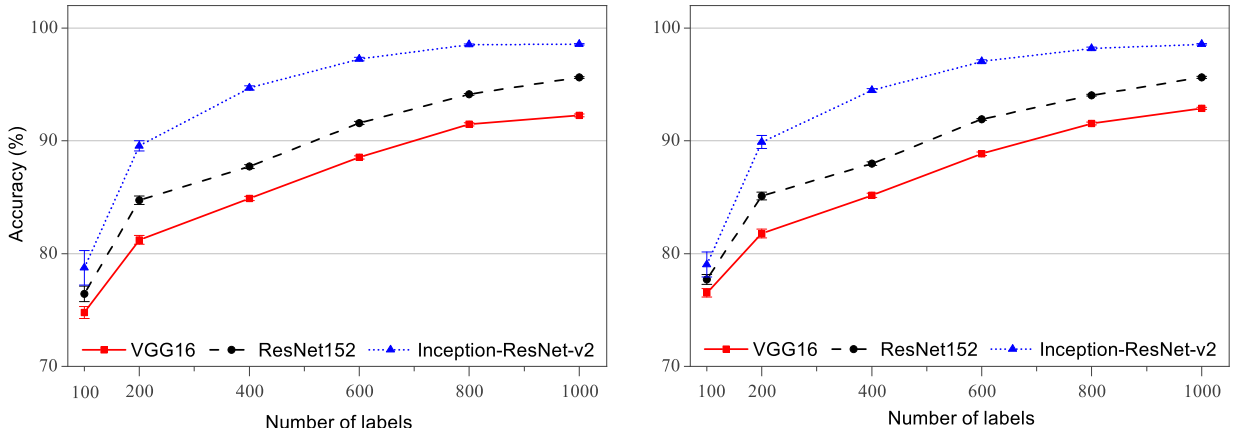


Figura 5.2: Relationship between the number of labels and model accuracy in the balanced data set for the VGG16, ResNet152 and Inception-ResNet-v2 models in two different crops: (a) Maize crop, and (b) Tomato crop.

Figure 5.2 illustrates that the relationship between the increase in labels and the improvement in model accuracy was not linear. When the number of labels was increased from 100 to 200, there was a larger relative variation in accuracy compared to increasing labels from 400 and above. In the maize crop, the VGG16, ResNet152, and Inception-ResNet-v2 models showed accuracy improvements of 8.6 %, 10.8 %, and 13.7 % respectively, when the number of labels increased from 100 to 200. Similarly, in the tomato crop, the corresponding accuracy gains were 6.9 %, 9.5 %, and 13.7 %. However, the Inception-ResNet-v2 model approached an asymptote, as the relative variation between 800 and 1,000 labels did not exceed 0.4 % for both maize and tomato crops.

5.3.2 Inference of the CNN models with balanced datasets

Figure 5.3 shows the analysis of the performance of the three models for each weed species using a balanced dataset with 150 labels of each species. In maize crop, the precision consistently resulted in a mean above 86 %. Overall, Inception-ResNet-v2 achieved the highest precision values for overall species, whereas VGG16 had the lowest values. The highest mean precision was achieved for *Portulaca oleracea* (99.9 % with Inception-ResNet-v2; and 95.7 % with VGG16) and for *Cyperus rotundus* (97.7 %) with ResNet152, while the lowest was observed for *Solanum nigrum* (86.2 % with VGG16; 92.8 % with ResNet152; and 96.2 % with Inception-ResNet-v2).

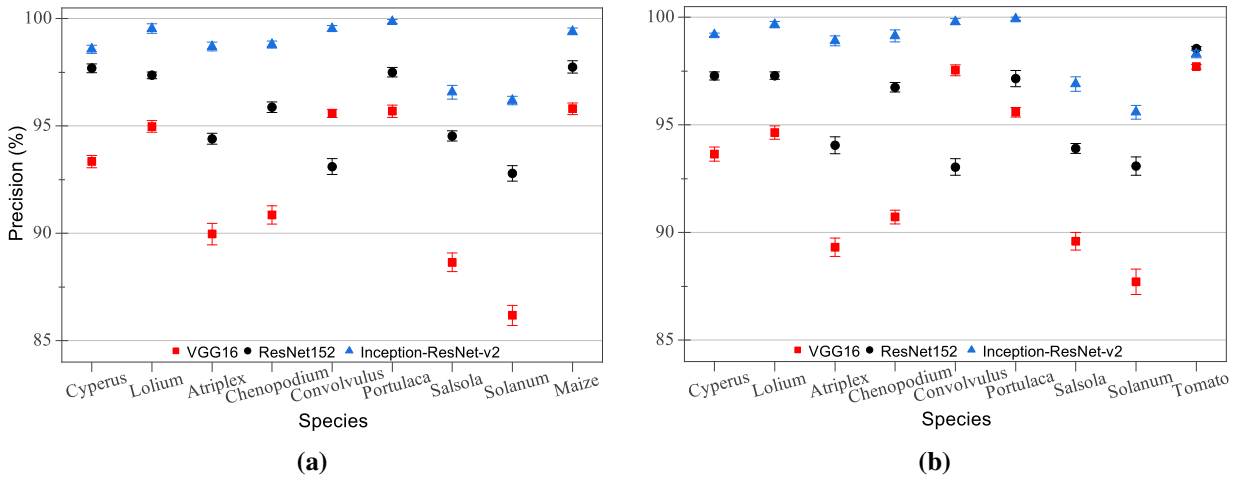


Figura 5.3: Mean precision (bars indicating standard error among ten executions for each model) using the maximum number of labels (1,000 for weed species) in the balanced dataset for VGG16, ResNet152 and Inception-ResNet-v2 models as affected by weed species in two different crops: (a) Maize crop, and (b) Tomato crop.

In the case of tomato, the performance analysis of the three models across diferent weed species consistently yielded a mean precision above 87 %. Similar to maize crop, Inception-ResNet-v2 generally achieved the highest precision values, while VGG16 had the lowest values. The highest mean precision was attained for *Portulaca oleracea* (99.9 %) with Inception-ResNet-v2, *Convolvulus arvensis* (97.5 %) with VGG16, and *Cyperus rotundus* and *Lolium rigidum* (97.3 %) with ResNet152. The lowest mean precision was observed for *Solanum nigrum* (87.7 %) with VGG16 and (95.7 %) with Inception-ResNet-v2, as well as for *Convolvulus arvensis* (93.0 %) with ResNet152.

5.3.3 Inference of CNN models with unbalanced dataset: crop as predominant class (Test I)

Test I was applied to analyze the performance of the different classification models in real-life scenarios, i.e. in which the source dataset is composed by an unbalanced number of labels with the crop as predominant species. Figure 5.4 illustrates the evolution of the *F1-score* metric by weed species and the number of maize labels increasing in different ratios. In general, most weed species (e.g. *Cyperus rotundus*, *Lolium rigidum*, *Portulaca oleracea*, *Salsola kali* and *Solanum nigrum*) consistently displayed mean *F1-score* values above 86 %, irrespective of the weed-to-crop label ratio. Nevertheless, certain weed species showed a declining *F1-score* value with an increasing number of crop labels. *Atriplex patula* consistently demonstrated a decrease in *F1-score* value across all three models, dropping below 80 % in both VGG16 and ResNet152 models when a large number of maize labels were present. The decline was less pronounced but still noticeable in *Convolvulus arvensis* for these two models. In line with the declining trend observed in some weed species, the analysis of the confusion matrices in maize crop (Figure 5.5) revealed a remarkable number of false positives in *Atriplex patula* ($F_P = 0{,}8\,\%$ and $0{,}9\,\%$ for VGG16 and ResNet152, respectively) and *Convolvulus arvensis* ($F_P = 0{,}9\,\%$ and $0{,}6\,\%$ for VGG16 and ResNet152, respectively). In addition, the percentage of false negatives generated between *Atriplex patula* and *Solanum nigrum* may have

contributed to decrease the *F1-score*, with $F_N = 0{,}2\,\%$ and $0{,}1\,\%$ for VGG16 and ResNet152, respectively.

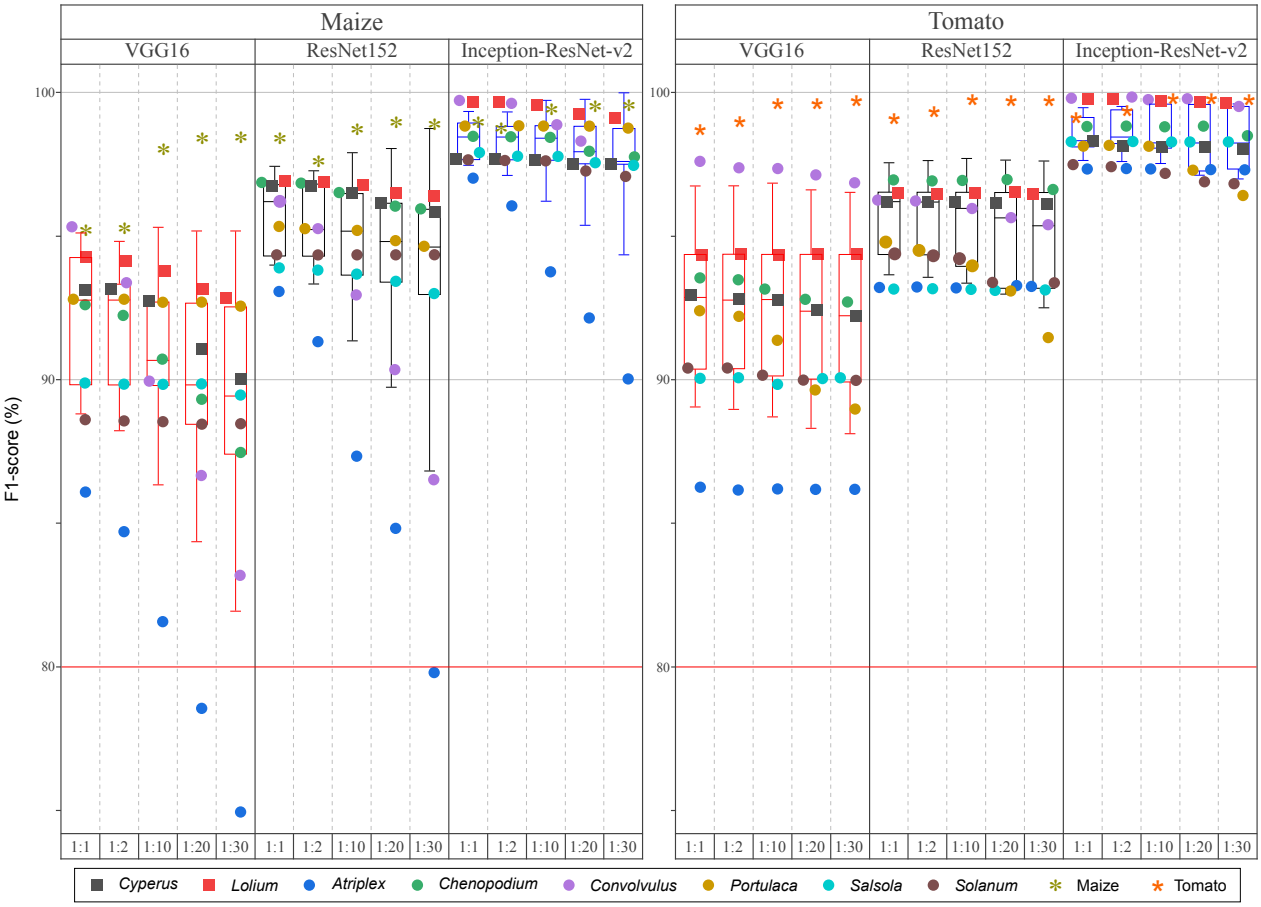


Figura 5.4: Box (mean *F1-score*) and whisker (± 1 SD) diagram showing Test I results across the range of weed-to-crop label ratios (from 1:1 to 1:30), for maize and tomato, using ten executions for the VGG16, ResNet152 and Inception-ResNet-v2 models.

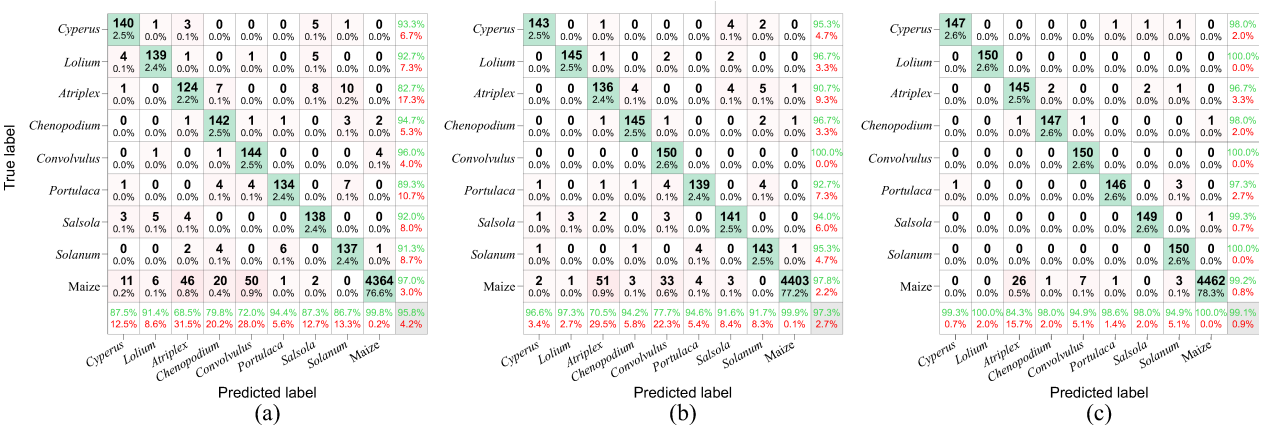


Figura 5.5: Confusion matrices of CNN models using Test I on maize crop and individual weed species. (a) VGG16, (b) ResNet152, (c) Inception-ResNet-v2.

Test I showed improved results in the three models for the tomato crop in comparison to the maize crop (Figure 5.4), with consistent *F1-score* values observed across all weed species, regardless of the weed-to-crop label ratio. Remarkably, the Inception-ResNet-v2 and ResNet152 models consistently achieved *F1-score* values above 91 % in the tomato crop. In contrast, the VGG16 model produced slightly lower values (though still above 86 %). Among the weed species using the VGG16 model, only *Atriplex patula* displayed the lowest *F1-score* value. However, this species maintained a stable *F1-score* value above 86.2 % across the entire range of weed-to-crop label ratios.

5.3.4 Inference of CNN models with unbalanced dataset: unequal number of labels for each species (Tets II)

Based on the *F1-score* values (Table 5.2), the results of Test II indicated that Inception-ResNet-v2 model performed the best in both crops. Indeed, the *F1-scores* for most weed species were consistently above 93 %, except for *Atriplex patula* and *Lolium rigidum*. The ResNet152 model showed values above 83 % in both crops, except for the aforementioned weed species and *Convolvulus arvensis* in the maize crop. The VGG16 model showed the lowest results, with values below 80 % in *Atriplex patula*, *Lolium rigidum* and *Chenopodium album* in both crops, in addition to *Convolvulus arvensis* in maize. These findings confirm that all models had a higher generalization ability with respect to the most commonly represented species (e.g. maize and tomato crops) compared to the weed species that had lower representation in the evaluated dataset (e.g. *Atriplex patula* and *Lolium rigidum*).

Tabla 5.2: *F1-score* metrics for three CNN models using Test II with an unbalanced dataset of different weed species and crops.

Species	Maize crop (%)			Tomato crop (%)		
	VGG16	ResNet152	Inception-ResNet-v2	VGG16	ResNet152	Inception-ResNet-v2
Cyperus rotundus	95.17±0.09	97.30±0.05	98.43±0.03	95.74±0.06	97.53±0.04	98.37±0.04
Lolium rigidum	**25.46±0.72**	**55.01±1.19**	**58.53±1.56**	**27.94±0.47**	**64.36±1.23**	**61.25±1.74**
Atriplex patula	**11.21±0.28**	**13.57±0.34**	**27.26±1.06**	**20.80±0.44**	**27.73±0.53**	**54.61±1.53**
Chenopodium album	**61.22±0.44**	85.18±0.26	94.28±0.47	**66.43±0.51**	85.98±0.29	94.62±0.43
Convolvulus arvensis	**74.56±0.61**	**79.21±0.58**	93.18±0.52	94.79±0.27	95.26±0.19	98.94±0.22
Portulaca oleracea	90.14±0.14	93.30±0.12	95.96±0.14	88.77±0.13	91.89±0.13	95.25±0.07
Salsola kali	82.49±0.46	83.45±0.33	95.78±0.23	83.70±0.33	85.29±0.25	95.54±0.26
Solanum nigrum	92.95±0.10	96.09±0.05	97.51±0.10	92.85±0.10	95.90±0.05	97.67±0.05
Maize	98.44±0.04	98.84±0.04	99.50±0.03			
Tomato				99.41±0.01	99.49±0.01	99.72±0.01

The analysis of the confusion matrices in tomato crop (Figure 5.6) showed false negatives generated between *Atriplex patula* and *Solanum nigrum*, with $F_N = 0{,}4\,\%$ and 0.5 % for VGG16 and ResNet152, respectively. Other false negatives were found between *Solanum nigrum* and *Chenopodium album* ($F_P = 1{,}4\,\%$ and 0.4 % for VGG16 and ResNet152, respectively), as well as between *Solanum nigrum* and *Portulaca oleracea* ($F_P = 0{,}9\,\%$ and 0.6 % for VGG16 and ResNet152, respectively). In addition, some false positives between *Cyperus rotundus* and *Lolium rigidum* ($F_P = 0{,}4\,\%$ for

VGG16), as well as between *Cyperus rotundus* and *Salsola kali* ($F_P = 0{,}4\,\%$ for ResNet152) were observed.

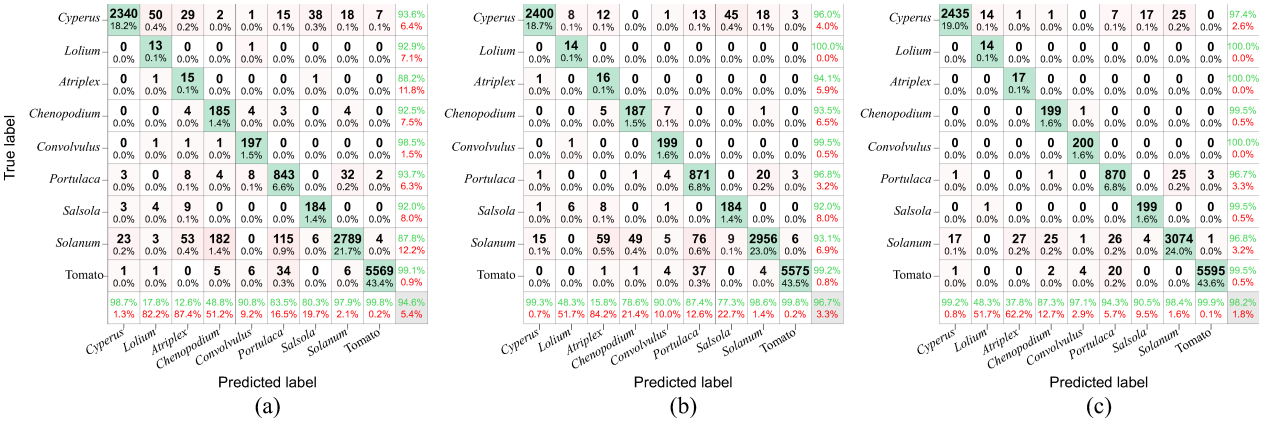


Figura 5.6: Confusion matrices of CNN models using Test II on tomato crop and individual weed species. (a) VGG16, (b) ResNet152, (c) Inception-ResNet-v2

These findings lead to an inquiry concerning the appropriate ratio of species imbalance necessary to attain acceptable classification metrics, specifically with a rate exceeding 80 %. Based on the results from Test I, where the decrease in *F1-score* was mainly attributed to the prevalence of maize labels in the dataset, the VGG16 and ResNet152 models achieved species-specific performance greater than 80 % when the ratio of minority to majority species ($R = S_{minority}/S_{majority}$ was 1:10 and 1:20, respectively (Figure 5.4). On the other hand, Inception-ResNet-v2 model consistently achieved a *F1-score* higher than 90 % when trained on 4,500 maize labels (e.g. R= 1:30); moreover, Test II results (Table 2) showed the lowest *F1-score* for *Atriplex patula* when tested on 11,364 maize labels (e.g. R= 1:667). In this study, we assessed the suitable ratio in the Inception-ResNet-v2 model by varying the numbers of *Atriplex patula* labels (17, 50, 75, 150, and 167) while adjusting the number of maize labels in increments (e.g. from 200 to 10,000) constant. Figure 5.7 establishes that an R value of approximately 1:107 or higher is required to attain an *F1-score* higher than 80 %.

Figura 5.7: Assessment of the optimal ratio of minority to majority species using *Atriplex patula* labels and maize labels (from n=17 to n=150) and maize labels (e.g. from 200 to 10,000) to achieve an *F1-score* above 80 % with the Inception-ResNet-v2 model.

Following an examination of the results across various scenarios for the two selected crop types, it was discerned that the Inception-ResNet-v2 model consistently demonstrated superior performance. This model stood out in terms of accuracy and efficiency, thereby establishing it as the best choice for advancing to the subsequent phase of our study.

5.3.5 Detection and mapping of weed species

After integrating the RPN and other Faster R-CNN modules with the output of the Inception-ResNet-v2 feature extractor, the inference process was performed on the dataset of real images corresponding to each field (subsection 5.2.2). The inference times resulted in 120.7 seconds for the maize field and 141.6 seconds for the tomato field. During this period, the algorithm detected 70,134 individual plants in the maize field (including *Atriplex patula* 676, *Chenopodium album* 8,373, *Convolvulus arvensis* 8,881, *Lolium rigidum* 71, *Salsola kali* 2,293, and maize 49,840), and 14,377 individual plants in the tomato field (*Cyperus rotundus* 9,248, *Portulaca oleracea* 3,164, and *Solanum nigrum* 1,965). Upon acquiring the predictions and weed locations in each individual image, these coordinates were subsequently transferred to the orthophoto coordinates, enabling the generation of weed maps and their visualization in the orthomosaics (Figure 5.8).

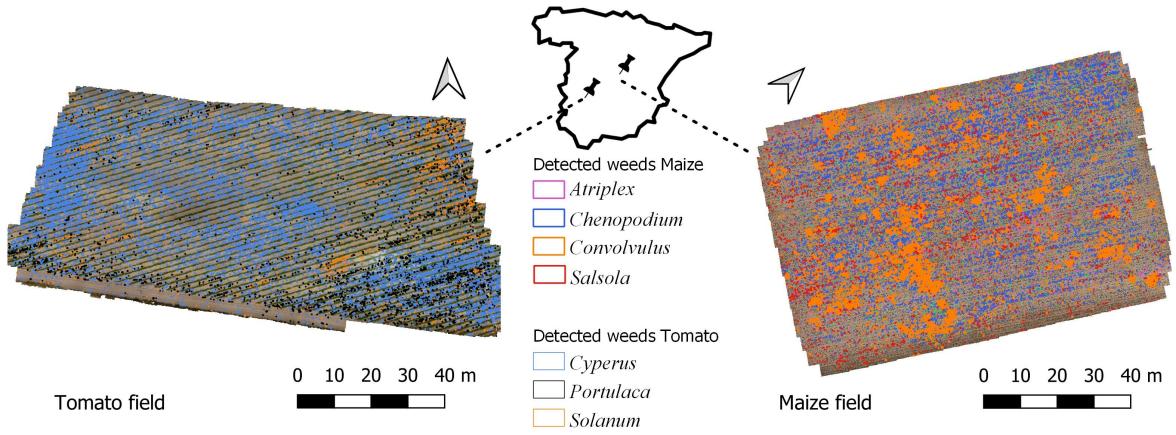


Figura 5.8: Weed maps created using the integrated architecture of Faster R-CNN and Inception-ResNet-v2 for the tomato and maize fields.

Table 5.3 shows the evaluation of the detection and classification results for various species provided by the architecture integrating Faster R-CNN and Inception-ResNet-v2, contrasting with ground-truth data in both maize and tomato crops. The ground-truth data consisted of a total of 970 labels, distributed between 356 and 614 for maize and tomato crops, respectively. The model successfully detected and classified 794 labels (237 for maize and 557 for tomato crops), with only eight instances of misclassification. Furthermore, *Cyperus rotundus* and *Portulaca oleracea* demonstrate outstanding performance, with *F1-score* values above 99 %. Conversely, species like *Chenopodium album* and *Solanum nigrum* exhibit high accuracy but lower recall, indicating proper classification when detected, albeit with some instances missed. Notably, *Salsola kali* displays comparatively lower overall performance, with an *F1-score* of 36.4 %.

Tabla 5.3: Confusion matrix and metrics for the evaluation of classification and detection by species, derived from the comparison of model results and ground-truth results in maize and tomato crops.

Species	*Cyperus rotundus*	*Lolium rigidum*	*Atriplex patula*	*Chenopodium album*	*Convolvulus arvensis*	*Portulaca oleracea*	*Salsola kali*	*Solanum nigrum*	Maize	Tomato	Support	Precision (%)	Recall (%)	F1-score (%)	Detection (%)
Cyperus rotundus	199										203	100.0	98.0	**99.0**	98.0
Lolium rigidum *		-									-	-	-	-	-
Atriplex patula			2								2	66.7	100.0	80.0	100.0
Chenopodium album				52	2						125	96.3	41.6	**58.1**	43.2
Convolvulus arvensis			1	2	33		1				49	94.3	67.3	78.6	75.5
Portulaca oleracea						204					205	99.0	99.5	**99.3**	99.5
Salsola kali							2				8	66.7	25.0	**36.4**	25.0
Solanum nigrum						2		154			206	100.0	74.8	85.6	75.7
Maize									140		172	100.0	81.4	89.7	81.4
Tomato ⁕										-	-	-	-	-	-
Total											970				

* *Lolium rigidum* was not found in the true ground frames.
⁕ The tomato category was excluded from evaluation due to its nature as a continuous row crop.

Figure 5.9 illustrates cases of evaluation of the model's discrimination performance. This analysis involved comparing the output images generated by our Faster R-CNN Inception-ResNet-v2 model with real images of maize and tomato fields.

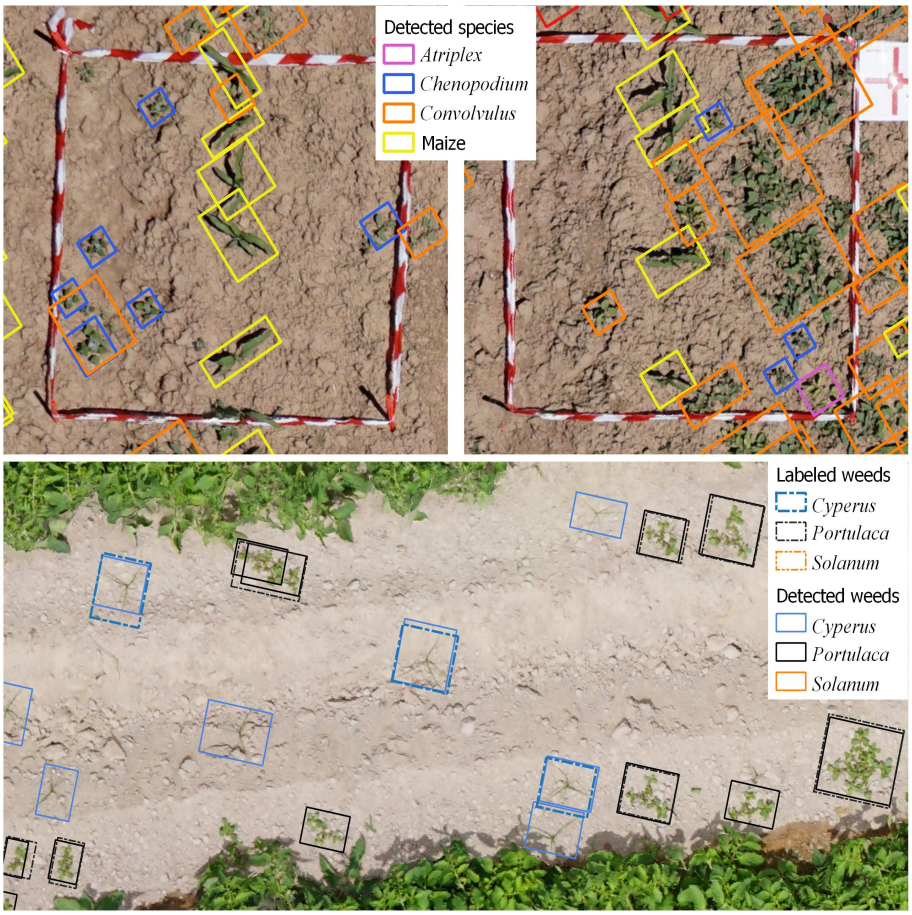


Figura 5.9: Detection and identification of crop and weed species in maize (upper images), as well as weed species in tomato (lower image), using the integrated architecture of Faster R-CNN with the Inception-ResNet-v2 classifier on split orthophoto images.

5.4 Discussion

The contribution of this study allows the detection, classification and location of weeds at the species level in an early growth stage, which enables tailored management and control strategies per species. This is particularly important because significant morphological and physiological variations exist within each weed species and they required different treatments. Identifying weed species at an early-stage using DL in UAV imagery poses a significant challenge (Rai et al., 2023). The development of an automated method for detecting and classifying weed species entails constraints to be solved and limitations, notably during training and testing phases. In this sense, we assessed the minimum dataset size required to attain 90 % accuracy. The results revealed that increasing the number of labels in the training and validation phase enhances the accuracy of CNN-based models in the classification task.

In our study, we found that using a UAV to capture images at an altitude of 11 meters resulted in orthophotos with a pixel size of 0.17 cm, which is a higher altitude and lower spatial resolution compared to similar research. For instance, (dos Santos Ferreira et al., 2017) and (Shahi et al., 2023) conducted weed classification research using UAV imagery at altitudes of 4 and 5 meters, respectively, achieving relatively higher spatial resolutions of 0.11 and 0.14 cm/pixel. Consequently, extracting useful features from our images poses a greater challenge. To mitigate this issue, one approach could involve training a more complex model capable of learning from smaller images. However, such a solution entails longer training time, increased consumption of computing resources, and hence increased energy expenditure. The CNN models selected for this study, exhibited varying levels of performance. Specifically, in balanced datasets, the minimum dataset size required to train CNN models and achieve accuracies exceeding 90 % ranged from just 400 labels for Inception-ResNet-v2 to 800 labels for VGG16, with intermediate label requirements for the ResNet152 model.

Due to the robustness of our dataset, we were able to train the CNN models with equal numbers of labels for each species, thereby preventing any bias towards the dominant class (Liu et al., 2022). The CNN models, VGG16, ResNet152 and Inception-ResNet-v2, achieved noteworthy classification accuracies of 92.6 %, 95.6 % and 98.6 %, respectively, when evaluated using models created with the maximum number of balanced labels per species. These results demonstrate the ability of CNN models to generalize knowledge by adapting to new data from natural environments at early growth stages, showing promising potential for implementing SSWM systems at an early stage of crop growth.

Previous research has effectively integrated remote sensing methods with artificial neural networks to differentiate weed groups, such as monocotyledons versus dicotyledons (Torres-Sánchez et al., 2021). Employing a similar methodology, (dos Santos Ferreira et al., 2017) utilized CaffeNet model on their segmented database, assessing its performance against traditional ML classifiers (e.g. Support Vector Machines, AdaBoost, and Random Forests). The evaluation involved feature extraction techniques such as shape, color and texture. Their results showed an accuracy of over 98 % in classifying all classes using the CNN model, thus concluding that the CNN model did not rely on the choice of feature extractor, unlike other models. Similar to prior studies (Hasan et al., 2021), our research aimed to take a step further by demonstrating the capability of CNN architectures in discriminating between multiple weed species in crop fields. We utilized a UAV imagery-based dataset, which proved robust to training across various CNN architectures through transfer learning techniques. The results were remarkable in accurately classifying all the weed species present in the two crop fields under investigation.

Although we acknowledge the limitations of our evaluation dataset in representing the full extent of field conditions, the analysis performed with Test II considering an unbalanced dataset provided results that closely approximate real-life scenarios. Under these more realistic conditions, the Inception-ResNet-v2 model produced the best results among the applied metrics, likely due to its greater depth and ability to capture more complex features and patterns in the UAV imagery. Indeed, the Inception-ResNet-v2 model achieved a performance (mean *F1-score*) ranging from 27.3 % to 99.7 %, while the ResNet152 model exhibited a range of 13.6 % to 99.5 %, and the VGG16 model showed a range of 11.2 % to 99.4 %. It is worth noting that the low values in the three CNNs corresponded to the results for the minority weed species, *Atriplex patula*. This is likely attributed to its confusion with another weed, *Solanum nigrum*, as indicated by the high

percentages of false negatives observed in the confusion matrices, irrespective of the crop. When calculating the mean value of the whole species, Inception-ResNet-v2 model showed a performance (mean *F1-score*) above 86 %, ResNet152 nearly 80 % and VGG16 nearly 74 %. Based on these results, the Inception-ResNet-v2 architecture proved to be the most effective model of weedy species, regardless of whether the dataset was balanced or unbalanced. In previous studies on individual species classification using UAV imagery, (dos Santos Ferreira et al., 2019) reported a mean accuracy of 83.4 % for four species using VGG16, while (Chew et al., 2020) achieved *F1-score* ranging from 49 % to 96 % for six species. Considering the lower number of species classified in previous works, some authors have suggested that the performance of CNN-based models tends to deteriorate when training on multiple species (Dyrmann et al., 2016; Olsen et al., 2019), making our work interesting as our dataset includes ten species.

A significant contribution of this study was the finding that the accuracy of weed species classification relies on identifying the ratio between minority and majority species, as it can have a substantial impact on the efficacy of SSWM systems. Neglecting this ratio could result in errors in decision-making and unintended consequences that could affect productivity weed control, hence crop productivity. Based on our results, in order to ensure an *F1-score* greater than 80 % for each species using the models VGG16, ResNet152 and Inception-ResNet-v2 on an unbalanced dataset, a ratio of 1:12, 1:24, and 1:107 between the majority and minority species, respectively, should be maintained. Often, control methods are tailored to target a specific species. However, if the relationship between the target species and the other species for which a CNN-based identification system can accurately identify is not considered, there is a possibility of misapplying control methods to the crop.

This article introduces inference time data, 120.7 seconds for maize and 141.6 seconds for tomato. This information not only provides insights into the model's computational efficiency (which clearly depends on the hardware used and the complexity of the orthomosaics), but also offers a general insight of how quickly the model can analyze and process large datasets. It is important to note that the efficiency demonstrated by the model is not comparable to that of human experts who could manually undertake the same task. Regarding individual plant detection, the findings demonstrate notable efficacy. For maize, the algorithm successfully identified a total of 70,134 individual plants, providing a detailed distribution between different weed species and the maize crop itself. Likewise, for tomato, the model detected 14,377 individual plants underscoring its capability to discriminate both the primary crop and the weed population within the field. Indeed, the evaluation of prediction accuracy by Faster R-CNN Inception-ResNet-v2 model against ground-truth data revealed generally satisfactory species detection rates, albeit with a few exceptions, such as *Chenopodium album* in maize and *Solanum nigrum* in tomato fields. The lower detection rate for these species could possibly be attributed to the small size of their seedlings during the early stages, typically less than 5 cm^2. Addressing this challenge might necessitate higher resolution UAV imagery, enabling more accurate identification and classification.

In summary, by employing 850 labels for each species in the training and validation phases of the Inception-ResNet-v2 classifier, along with the Faster R-CNN object detector, our approach has successfully identified all weed species within the maize and tomato experiments. The precise identification of individual weed classes and crops has enabled the generation of detailed and customized maps, thereby facilitating accurate and site-specific weed management strategies.

5.5 Conclusions

This research represents a significant advance in early-growth stage weed species classification using UAV imagery and standard CNN models, such as VGG16, ResNet152 and Inception-ResNet-v2. We introduce a methodological innovation addressing the challenge of training a CNN model with images of varying sizes, deviating from its original design specifications and causing performance decline. To counteract this, we adapt the model inputs to align with the information conveyed by low spatial resolution images, providing an effective solution to mitigate model performance degradation in scenarios with image size variability. This approach offers valuable insights for the practical application of CNN in diverse environmental settings. In addition, we assessed the constraints and limitations related to the development of a weed species classification method during both the training and testing phases. Our findings reveal that, in balanced datasets, the minimum dataset size necessary to train CNN models and attain accuracies exceeding 90 % varied from 400 to 800 labels, depending on the particular model used. Inception-ResNet-v2 consistently outperformed other models with accuracy values of 96 % in both maize and tomato crops, likely due to its deep architecture and greater ability to capture intricate features and patterns within UAV images.

In complex and realistic scenarios with unbalanced datasets, Inception-ResNet-v2 also showed the best performance. The study revealed the critical role of the minority-to-majority species ratio on the classification accuracy, particularly affecting less represented species, like *Atriplex patula* and *Lolium rigidum* weeds. Unequal label distribution penalizes minority species, decreasing accuracy, while favoring more abundant species, such as crops. Therefore, determining the appropriate number of labels for CNN model training and validation is crucial for the precise and effective implementation of SSWM techniques. Our results suggest maintaining ratios of 1:12, 1:24, and 1:107 between the majority and minority species for VGG16, ResNet152 and Inception-ResNet-v2 models, respectively, to ensure *F1-score* exceeding 80 % per species.

The integration of Faster R-CNN algorithms with Inception-ResNet-v2 shows promise for detecting and classifying various weed species in maize and tomato crops. While these architectures present opportunities for optimization in agriculture, it is worth recognizing their limitations. Factors such as image resolution, plant overlap and the presence of other structures or elements in the field can affect detection accuracy. Generalization to different environmental conditions and crop types requires further evaluation. Despite the challenges, the results obtained support the potential usefulness of these architectures in agriculture and provide valuable information for crop management. Further research is recommended to address limitations and refine the applicability of the technology in diverse agricultural settings.

Capítulo 6

Classification, detection and mapping of early-season weed species in UAV images with cutting-edge CNN and Vision Transformer models

Publicación asociada a este capítulo Classification, detection and mapping of early-season weed species in UAV images with cutting-edge CNN and Vision Transformer models

- Mesías-Ruiz GA, Borra-Serrano I, Dorado J, de Castro AI, Peña JM. Classification, detection and mapping of early-season weed species in UAV images with cutting-edge CNN and Vision Transformer models. *Computers and Electronics in Agriculture*. 2024. (Under review)

Abstract

Early-season weed species identification plays a pivotal role in the context of precision agriculture. Accurate weed classification, detection and mapping at species level enables the implementation of site-specific control measures, potentially resulting in a significant reduction in herbicide use. This study addresses the challenge of creating geo-referenced maps of weed species during the early season using imagery acquired by unmanned aerial vehicles (UAVs). Three convolutional neural network models (Inception-ResNet-v2, EfficientNet-B0, and YOLO-v8) and two vision transformer-based models (ViT-Base and Swin-T) were evaluated. Training and inference of the models was performed on a large dataset consisting of 16,051 images including nine weed species of two summer crops: maize and tomato. The models, evaluated on an unbalanced dataset, achieved a weighted average *F1-score* between 85.5 % and 98.3 %, with YOLO-v8 and Swin-T showing better performance in weed species classification. Next, the object detectors YOLO-v8m and DETA were implemented to generate weed density maps based on species-specific economic weed thresholds. The performance of the weed detection and classification models was evaluated in maize and tomato fields. The YOLO-v8m model achieved 80 % accuracy in maize and 94 % accuracy in tomato. In terms of computational costs, the YOLO-v8m model proved to be faster during the inference tests, suggesting its suitability for real-time applications and great potential to boost UAV-based weed species detection and mapping procedures.

6.1 Introduction

Technological advances, such as unmanned aerial vehicles (UAVs) and machine learning (ML), are transforming agriculture by providing frequent crop data and techniques to help farmers and agronomists to make informed decisions and optimize field operations, remarkably boosting the implementation of precision agriculture (PA) strategies. On the one hand, the UAVs are able to capture high resolution images over large cropping areas, giving diverse uses such as surveying, environmental monitoring, crop scouting and mapping (de Castro et al., 2021; Rejeb et al., 2022). On the other hand, ML and, particularly, some specific deep learning (DL) techniques such as convolutional neural networks (CNNs), have demonstrated particular success in image processing, improving the accuracy of detection and classification tasks on diverse agricultural goals (Mesías-Ruiz et al., 2023).

Among the most common applications of PA is site-specific weed management (SSWM), which involves identifying and treating weeds according to their spatial location, rather than managing the entire field uniformly (Lati et al., 2021). Over the last decade, much progress has been made in the area of digital image processing for weed scouting and mapping (Fernandez-Quintanilla et al., 2022). Focusing on UAV imagery, most studies essentially aimed at: 1) mapping the weeds as a general class, e.g. in maize (Peña et al., 2013), sunflower (Pérez-Ortiz et al., 2015), soybean (dos Santos Ferreira et al., 2017), sugar beet (Sa et al., 2018), rice (Huang et al., 2020), etc., 2) mapping only a single weed species predominant in the study crop-field, e.g. *Cynodon dactylon* in vineyards (Jiménez-Brenes et al., 2019), *Alopecurus myosuroides* in wheat fields (Fraccaro et al., 2022), *Mercurialis annua* in chicory plantations (Gallo et al., 2023), or 3) mapping between broad- and narrow- leaved weeds (Panduangnat et al., 2024; Torres-Sánchez et al., 2021).

Based on this background, current research is moving towards more challenging tasks aimed at discriminating multiple weed species at early growth stages. The purpose is to facilitate timely weed control and increase the effectiveness of SSWM, since weeds compete with crops for vital resources from the first growth stages and, also, different weed species require distinct control measures (Montull et al., 2014). Although this task is complex due to the small sizes and morphological similarities of weed and crop plants in early season (Fernández-Quintanilla et al., 2018), some studies have already reached good results with standard CNN models applied to UAV imagery in some crops. For example, Huang et al. (2018a) applied fully convolutional networks (FCNs) to map two weed species in rice fields, de Camargo et al. (2021) optimized the ResNet-18 model for classifying four weed species in winter wheat crops, while Mesías-Ruiz et al. (2024a) mapped three and four weed species in tomato and maize crops by using the Faster R-CNN and Inception-ResNet-v2 models, respectively.

However, to increase the scope of the UAVs for operational SSWM applications, efforts should be focused on exploring new tools capable of mapping a larger number of weed species with UAV images, which will require powerful models and expert knowledge. The vision transformers (ViT), a new paradigm in the domain of image recognition, has emerged to compete with the supremacy of CNNs in these tasks (Dosovitskiy et al., 2021). The ViT architecture uses attention mechanisms to establish robust interconnections between input and output features and highlight those of utmost relevance, yielding improved parallelization and decreasing computational process required in conventional convolution operations. Consequently, the ViT models generally facilitates faster training times, increased parallel processing capabilities and potentially higher performance than CNNs (Liu et al., 2023). Comparison between CNNs and ViTs for weed classification tasks has only been assessed in particular scenarios, e.g. to classify nine Australian native weed species in on-ground images (Espejo-Garcia et al., 2023) or in recognizing weed and crop plants (as general classes) in UAV images collected over beet, parsley and spinach fields (Reedha et al., 2022).

This view motivated our research aiming to evaluate the performance of five cutting-edge DL models (both CNNs and ViTs) for detecting, classifying and mapping nine weed species in early-season maize and tomato crops by using UAV-based ortomosaics. Additionally, the specific objectives of this study were: 1) analyzing the models´ efficiency in terms of computational costs during the training and inference phases, 2) selecting the two best performing CNN and ViT architectures and implementing multi-object weed detectors, and 3) generating gridded treatment maps based on plant density of weed species for practical SSWM operations in both studied crops.

6.2 Materials and methods

Two summer crops of global economic relevance, i.e. maize and tomato, together with nine of their major weed species were the agricultural scenarios studied. The procedure consisted of five main steps (Figure 6.1): 1) UAV imagery acquisition, 2) imagery pre-processing involving ortho-mosaic building, image splitting and plants' labelling to generate the crop-weed dataset, 3) classification, which included different techniques of feature extraction (convolutional layer vs. window transformer), transfer learning (ImageNet-1k *vs*. ImageNet-21k) and inference for each CNN and ViT model, respectively, 4) Weed detection and geolocation by using the YOLO-v8m and DETA architectures applied to the CNN and ViT models that best performed in the previous

classification step, respectively, and 5) generation of gridded treatment maps based on plant density for each weed species.

Independent training/validation, test and generalization datasets were used for models' training (i.e. feature extraction), comparison and outputs evaluation, respectively, using confusion matrices and other spatial-based metrics (see section 6.2.6). All the computing operations were conducted on a workstation running the Ubuntu 20.04.4 LTS operating system and Python 3.9. The hardware included an Intel® Core (TM) i7-12700K CPU and an NVIDIA GeForce RTX 3070 Ti Graphic Processing Unit (GPU) featuring 8 GB of memory. Next, each stage is described in detail.

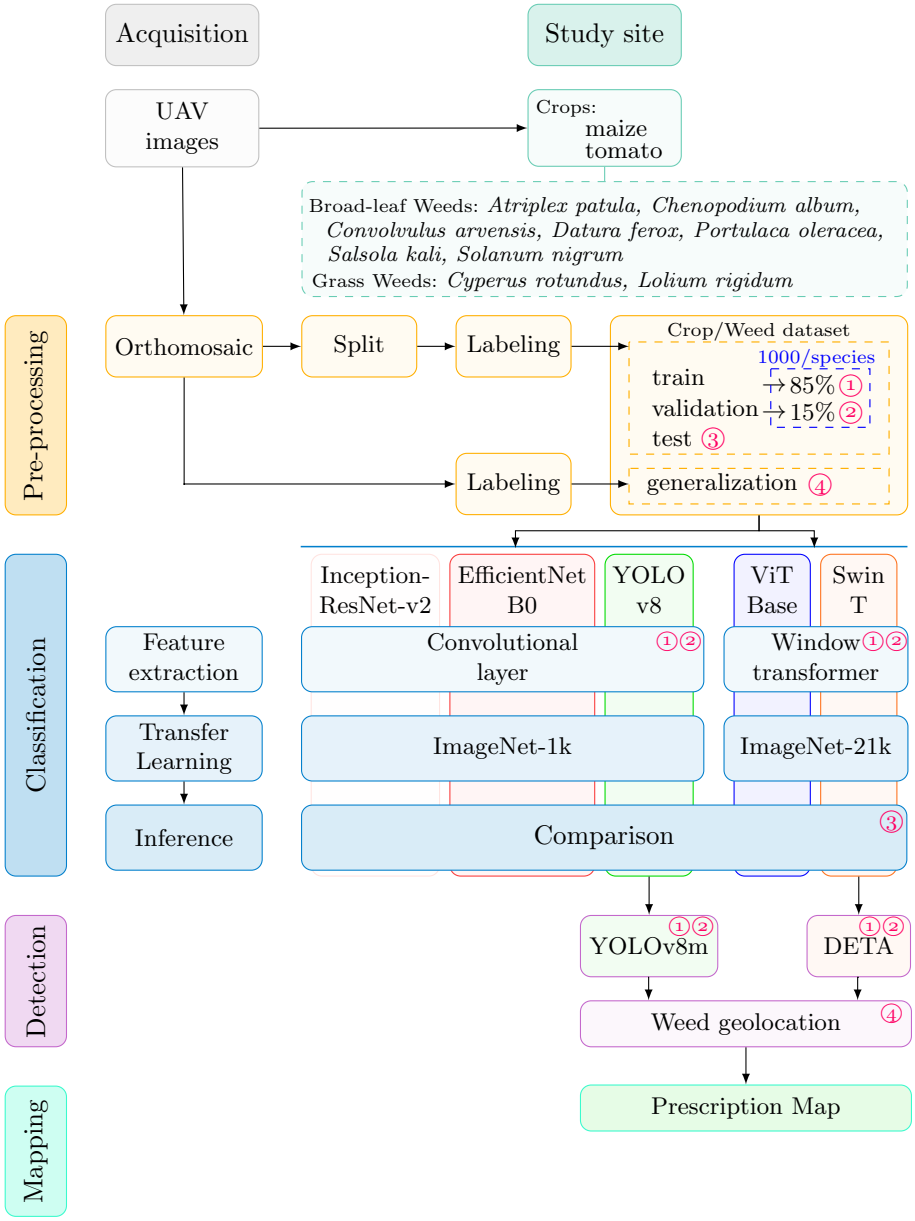


Figura 6.1: Pipeline of the procedure applied to classify, detect and map weed species. The red numbers refer to the division of the dataset in training (1), validation (2), test/inference (3), and generalization (4).

6.2.1 Study site and acquisition of UAV images

The UAV images were collected in an 0.75 ha experimental maize field located in Arganda del Rey (Madrid, Spain) and in a 1.2 ha commercial tomato field located in Santa Amalia (Badajoz, Spain), with central coordinates 40°18'57.59"N, 3°29'22.57"W and 38°59'15.58"N, 6°02'57.71"W (Lat/Lon coordinate system, datum WGS84), respectively. Soils in both fields were predominantly sandy loam and had low organic matter content, being conventionally tilled and frequently irrigated during crop development. Maize was cultivated in rows spaced 0.75 m apart and at a density of 85,000 plants per hectare, while tomato was in rows spaced 1.5 m apart at a density of 25,000 plants per hectare. During the UAV flights, the maize plants were at the early growth stage of four unfolded leaves (BBCH 14), and the tomato plants at the early stage of first flower bud visible (BBCH 501) (Meier, 2018). Nine common weed species were naturally present in the fields, as follows: *Atriplex patula*, *Chenopodium album* L., *Convolvulus arvensis* L., *Datura ferox*, *Portulaca oleracea* L., *Salsola kali* L., and *Solanum nigrum* L. as broad-leaf weeds, and *Cyperus rotundus* L. and *Lolium rigidum* Gaud as grass weeds (Figure 6.2).

The images were acquired with a low-cost commercial red-green-blue (RGB) camera, model Sony ILCE-6300L (Sony Group Corporation, Tokyo, Japan) mounted in a quadcopter UAV, model md4-1000 (microdrones GmbH, Siegen, Germany), flying at 11 m above ground level (AGL). The camera featured an APS-C type Exmor® CMOS sensor (23.5×15.6 mm) that captures images of 6000×3376 pixels with 24.2 effective megapixels. The UAV route was designed to fly at a speed of 2 m/s, forward and side overlaps of 70 %, resulting in 565 and 895 images of the maize and tomato fields, respectively, with a ground sampling distance (GSD) of 0.17 cm/pixel.

6.2.2 Image pre-processing

The imagery pre-processing stage encompassed the generation of the field orthomosaics, image splitting, and labeling of the crop and weed species. Firstly, the orthomosaics were built by stitching together the multiple overlapping UAV images of each field to produce a single high-resolution image covering the entire area of interest. This process involved the main phases of image alignment and field geometry building that were automatically done with the Agisoft PhotoScan software (Agisoft LLC, St. Petersburg, Russia), plus manual image georeferencing by using the coordinates of several ground control points of the study fields taken with a global navigation satellite system (GNSS) receiver. Orthomosaicking rectified distortions caused by varying camera angles and perspectives, thereby facilitating precise analysis of the image objects (Peña et al., 2015).

The orthomosaics of the maize and tomato fields were 2.60×10^9 and 4.15×10^9 pixels in size, respectively, which are too huge to be directly processed by conventional computers or workstations. Thus, image splitting was applied to automatically partition the orthomosaics into smaller sections of 1000×1000 pixels using a program specifically developed in Python. This partitioning significantly reduced the computational cost of the analysis, allowing more efficient handling of the images. As a result, 1699 and 2010 images of the maize and tomato fields were generated, respectively. Finally, experts on weed science used the free graphical annotation tool labelImg (Tzutalin, 2015) to manually draw bounding boxes and label the name of the weed species observed in every split image (Figure 6.2).

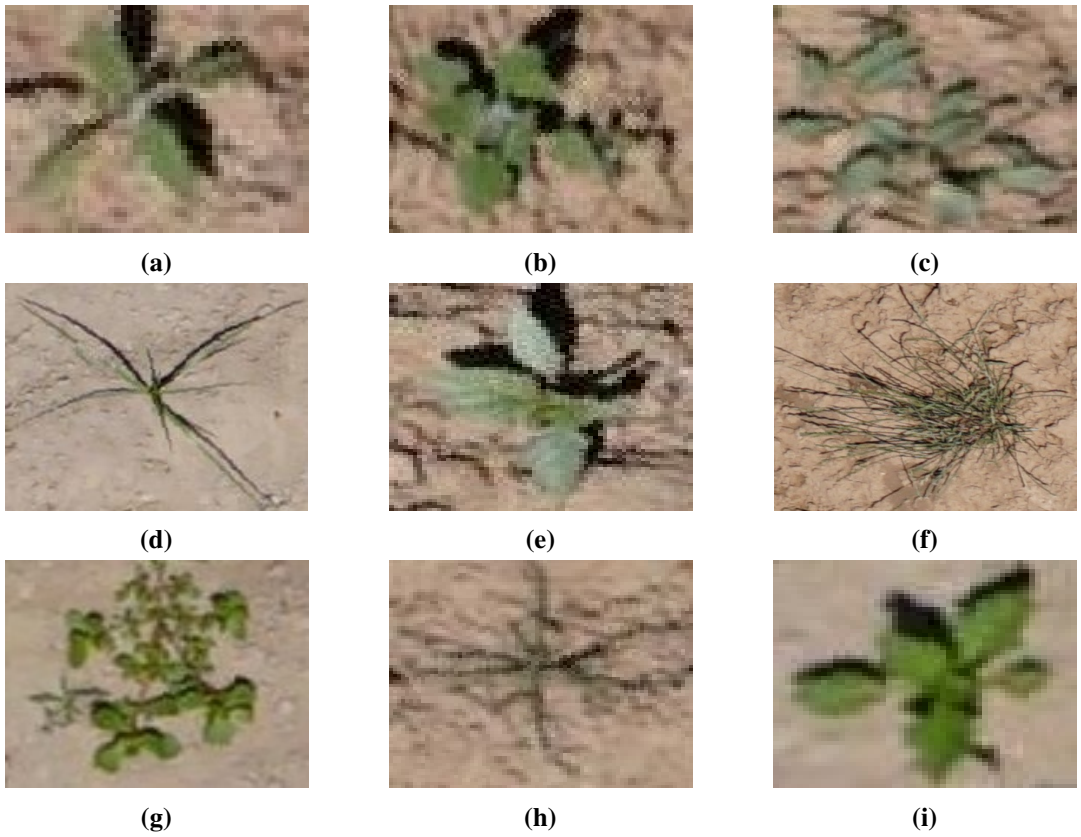


Figura 6.2: Examples of the UAV-based imaging labels of the nine weed species observed in the study fields: *Atriplex patula* (a), *Chenopodium album* L. (b), *Convolvulus arvensis* L. (c), *Cyperus rotundus* L. (d), *Datura ferox* (e), *Lolium rigidum* Gaud (f), *Portulaca oleracea* L. (g), *Salsola kali* L. (h), and *Solanum nigrum* L. (i).

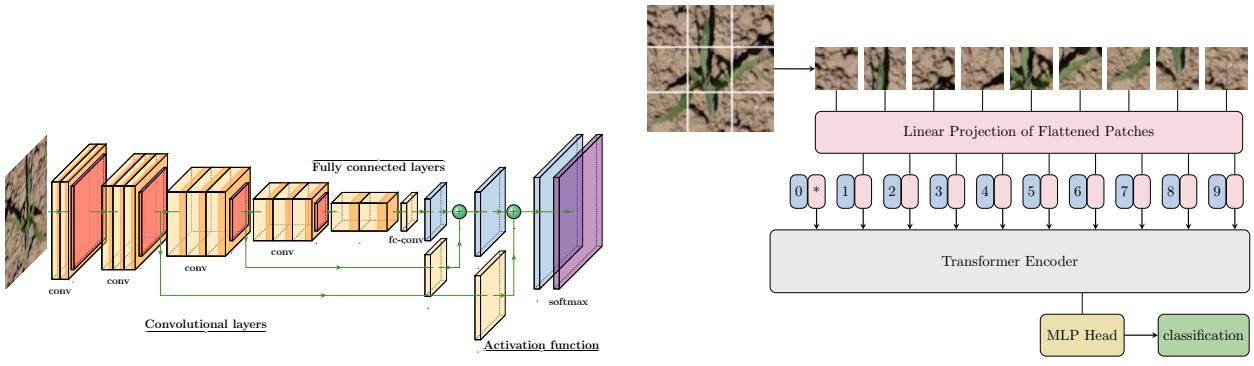
The annotations made on the split images were saved with XML extension. Algorithm 6.1 (Appendix 6.6.1) was implemented to compile the labelling dataset and obtain one weed per image. This procedure was meticulously executed to generate a dataset with sufficient diversity, representativeness and quantity to guarantee robust and accurate outcomes, thus avoiding potential overfitting of the supervised learning framework (Santos y Papa, 2022). A source dataset of 8967 weed labels were selected, which were split into 85 % and 15 % to apply a training and validation procedure, respectively. Other 7084 additional weed labels were used to generate an unbalanced dataset to apply the inference phase and compare the performance of the five studied models. This full dataset is freely available at (Mesías-Ruiz et al., 2024b).

6.2.3 Weed classification

Three cutting-edge CNN models, i.e. Inception-ResNet-v2, EfficientNet-B0 and YOLO-v8, and two vision transformer models, i.e. ViT-Base and Swin-T, were evaluated and compared for the classification task. The five models have specific characteristics that provide insights into their superior ability to fulfil the classification tasks (Appendix 6.6.2, Table 6.9). Each CNN model possesses a unique architecture characterized by varying numbers of layers and connectivity patterns

to address complex computer vision challenges with extremely high precision. The CNN models are composed of several layers, each with a specific function of extracting and processing features from the input data (Figure 6.3a). It starts with an input layer that takes the data, followed by multiple convolutional layers that apply filters to detect relevant patterns and features in the data. After each convolutional layer, an activation function is usually applied to introduce nonlinearity into the model, and the pooling layers are used to reduce dimensionality and retain the most important features. Lastly, fully connected layers combine the extracted features to perform the desired prediction or classification. In the case of the ViT models, their drive is by leveraging its architecture's prowess from natural language processing. They are based on the idea of multimodal attention, where images are decomposed into small patches, which are treated as input sequences (Figure 6.3b). Each patch is embedded in a latent space that can be processed by the Transformer model.

The classification stage was executed within the Keras-TensorFlow DL framework (Abadi et al., 2016). This framework included feature extraction with convolutional layers (for CNNs) or local windows (for ViT), transfer learning by incorporating weights previously trained on the ImageNet-1k (Krizhevsky et al., 2017) or ImageNet-21K (Ridnik et al., 2021) datasets, respectively, and inference evaluation on an unbalanced test dataset. Transfer learning was performed by replacing the last layer with a global mean clustering layer using the softmax activation function and adjusting it to the number of classes in our study. The Inception-ResNet-v2 and YOLO-v8 models shared the same optimizer with a learning rate set to 10^{-5}, while the EfficientNet-B0, ViT-Base and Swin-T models were trained with learning rates of 10^{-4} and 10^{-3}, respectively. All classifiers were trained using a batch size of 32 over 30 epochs. The adaptive moment estimation optimizer (Adam) was employed, with categorical cross-entropy serving as the loss function (Kingma y Ba, 2017). The weed label image sizes were normalized to 224×224 pixels during model training. Data augmentation techniques (Khalifa et al., 2021; Shawky et al., 2020) were purposefully omitted at the classification step, aiming to assess and compare classifiers' performance based only on the original source dataset and without the influence of synthetic data.



(a) **(b)**

Figura 6.3: Graphical scheme of the CNN (a) and ViT (b) architectures. Figure (b) adapted from (Dosovitskiy et al., 2021)

6.2.4 Weed detection and geolocation

Two multi-object detectors, i.e. YOLO-v8m and Detection Transformers with Assignment (DETA), were implemented for simultaneously locating and identifying the weed species present within an image. The detectors were selected according to the classifiers that best performed in the previous classification phase, i.e. the CNN YOLO-v8 and the ViT SwinT models, respectively (see section 6.3.2).

The YOLO-v8m is a cutting-edge detector that has shown an exceptional balance between speed and accuracy (Redmon et al., 2016), occupying an intermediate position in terms of model size and computational complexity between the resource-constrained "nano" and "small" models and the more computationally demanding larger architectures (Jocher et al., 2023). Alternatively, DETA, based on the strong two-stage DETR framework (Zhu et al., 2021), generates overlapping object detections and employs the Non-Maximum Suppression (NMS) method for post-processing. DETA also introduces an object balancing technique to improve performance on small objects. DETA features a more expressive approach in its heads, i.e. it uses a 6-layer transformer encoder on the convolutional backbone in the first stage, unlike traditional CNN detectors that typically employ 2-4 convolutional layers. In the second stage, DETA uses a 6-layer transformer decoder, while a conventional CNN uses only 2 linear layers (Ouyang-Zhang et al., 2022).

Different strategies were applied to train and validate both detectors. On the one hand, the training images for YOLO-v8m were resized to 640×640 pixels and, additionally, a data augmentation technique was randomly implemented to such images to increase the diversity of the training data. This process included horizontal flipping with a probability of 50 %, Gaussian blur with a kernel size of 3×3 and a sigma between 0.5 and 0.5, as well as contrast adjustment by multiplying each pixel value by a random factor between 0.1 and 0.9. Image resizing was also performed to a target size of 640 pixels, with a scale factor between 0.75 and 1.3. This approach was employed to minimize the risk of reducing the model's capacity for generalization in real-world scenarios. A learning rate of $1\text{x}10^{-4}$ was used in the hyperparameter settings of YOLO-v8m, with batches of size 8 and a total of 200 training epochs, and the Adam algorithm as the optimization function. The binary cross entropy was selected as the classification loss function, while the complete intersection over union (CIoU) metric was chosen for the box loss function. This comprehensive approach provides a more detailed evaluation of the model's ability to accurately locate objects during training. On the other hand, to train the DETA model, the PyTorch data loaders was employed to resize the training dataset to 800×800 pixels, which is the minimum size supported by DETR. A learning rate of $1\text{x}10^{-5}$, weight decay of $1\text{x}10^{-4}$, with batch sizes of 2 and a total of 10 training epochs were used for the hyperparameter settings, and the AdamW algorithm for the optimization function (Loshchilov y Hutter, 2019). PyTorch Lightning was the DL framework used to train the DETA model, with the trainer parameters set to a gradient clipping value of 0.1, a gradient accumulation of 8, and a row record aggregation frequency of 5.

Finally, a customized algorithm was developed to geolocate the weed species detected and classified within the partitioned images into the field orthomosaics (Algorithm 6.2, Appendix 6.6.1). Once the inference is performed on the partitioned (1000×1000 pixels) images, the algorithm calculated the centroid's local coordinates $(x_{cr}; y_{cr})$ of the bounding boxes of the detected weeds (equation 6.1) and converted them to global geographic coordinates $(x_m; y_m)$, thus translating from local to global coordinates and allowing to pinpoint the local weed locations to the real-world reference

frame (Figure 6.4). This process was also defined by a transformation matrix (Γ) in the form $P_{global} = \Gamma \cdot P_{local}$ (equation 6.2), which contains the scaling values defined as the ratio between the GSD and the size of the partitioned image (l_r), needed to perform this conversion from local to global coordinates of every detected weed (equation 6.3).

$$(x_c; y_c) = \left(\frac{x_{min} + x_{max}}{2}; \frac{y_{min} + y_{max}}{2}\right) \tag{6.1}$$

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} A_{mx} \\ A_{my} \end{bmatrix} + \Gamma \cdot \begin{bmatrix} x_{c_r} \\ y_{c_r} \end{bmatrix} \tag{6.2}$$

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} A_{mx} \\ A_{my} \end{bmatrix} + \begin{bmatrix} \frac{GSD}{l_r} & 0 \\ 0 & -\frac{GSD}{l_r} \end{bmatrix} \cdot \begin{bmatrix} x_{c_r} \\ y_{c_r} \end{bmatrix} \tag{6.3}$$

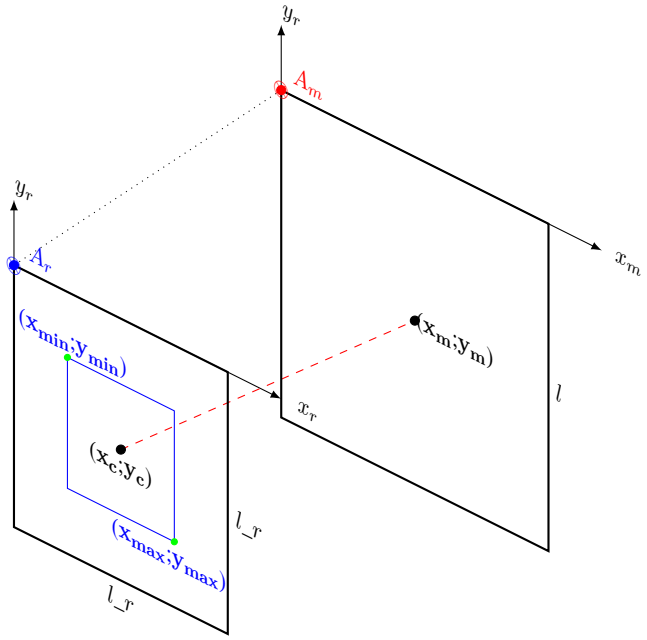


Figura 6.4: Scheme of the weed geolocation process, showing the relationship between the local coordinates $(x_{cr}; y_{cr})$ corresponding to the weed position in the partitioned images and their global coordinates $(x_m; y_m)$ corresponding to the weed position in the field orthomosaics.

6.2.5 Gridded treatment maps for SSWM

The classifier+detector model with the highest accuracy was used to create the weed species maps in the two studied fields, followed by gridded treatment maps according to weed density (#weeds/m^2). A search on scientific literature the economic weed thresholds (EWT) for controlling each weed species detected (Coble y Mortensen, 1992; López-Granados et al., 2016b; Norris, 1999; Oerke, 2006; San Martín et al., 2016), which was used to classify three zones in the treatment maps, as follows: 1) no weeds, so no treatment is needed (zone 1), 2) weed density below EWT, so no treatment

is economically recommended (zone 2), and 3) weed density above EWT, so weed treatment is economically recommended (zone 3). These species-specific treatment maps are potentially useful to calculate in advance critical variables for SSWM operations such as total area of weed treatment, herbicide type and doses, and operational savings. Thus, a GIS software (i.e. QGIS, version 3.26) was used to visualize and evaluate the treatment maps, and to compute operational metrics for SSWM implementation.

6.2.6 Models' evaluation

Evaluation of the classification step

Models' performance on weed species classification was evaluated in terms of: 1) training epochs and computational cost, which was mostly sensitive to the training and validation phases, and specifically to the number of trainable parameters, model size and time devoted to each phase, and 2) model overall accuracy, which was focused on the inference phase and specifically on the comparison between the metrics obtained from the confusion matrices in an unbalanced test dataset. Accuracy is a reliable metric of the overall model performance in scenarios with a balanced data distribution, however it might be misleading in unbalanced datasets where one class predominates. In such cases, a model might achieve high overall accuracy by mainly predicting the dominant class, disregarding insights from minority classes. Therefore, the recall and *F1-score* metrics are particularly recommended to assess the impact of minority classes in model performance. The performance metrics on Table 6.1 were analyzed on a species-specific basis, showing the capacity to identify specific classes in which the models exhibited notable performance disparities across the dataset.

Tabla 6.1: Confusion matrix metrics used to evaluate model performance.

Metric	Definition	Equation*
Overall Accuracy (OA)	Percentage of correct predictions respected to the total number of elements in the dataset.	$OA = \frac{T_P+T_N}{T_P+F_P+F_N+T_N}$
Precision (P)	Represents the positive prediction values.	$P = \frac{T_P}{T_P+F_P}$
Recall (R)	Measures the ability of the model to identify true positives.	$R = \frac{T_P}{T_P+F_N}$
F1-score ($F1$)	Calculates the harmonic mean between P and R, providing a balance between the two metrics. It is especially useful for unbalanced dataset.	$F1 = \frac{2T_P}{2T_P+F_P+F_N}$

* T_P = true positive, T_N = true negative, F_P = false positive, F_N = false negative.

Evaluation of the detection and mapping steps

The detectors were evaluated with several metrics for computing object classification and location accuracy, and the ability to identify multiple objects in an image (Table 6.2). The loss function is a weighted sum of three individual loss components, i.e. bounding box regression loss, classification loss, and confidence loss. Notably, the detectors utilize the CIoU loss function for bounding box regression. Specifically, CIoU considers the IoU, the normalized distance between the predicted and ground truth bounding box centers, and the aspect ratio penalty (Zheng et al., 2022). This comprehensive approach to bounding box regression loss contributes to model's overall detection performance (Shixin et al., 2024).

Next, the species-specific weed maps were evaluated by comparing the mapping outputs to ground-truth weed plants observed and classified in both field orthophotos by weed experts (i.e. generalization dataset), which accounted to 50 and 200 labels by weed species in the studied maize and tomato fields, respectively. The results were analyzed in terms of true positive, false positive and false negative ratios of each weed species, which showed the number of detected plants and those correctly or incorrectly identified for further designing an effective weed treatment.

Tabla 6.2: Metrics used to assess object detection.

Metric	Definition	Equation*
Mean Average Precision (mAP)	Compare the bounding boxes of detected *vs*. ground-truth.	$mAP = \frac{1}{\lvert classes \rvert} \sum_{c \in classes} \frac{T_P}{F_P + T_P}$
Intersection over Union (IoU)	Determines whether a region has an object or not.	$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$
Complete-IoU ($CIoU$)	Loss function that improves the accuracy of object detection by considering the geometry of the bounding boxes.	$CIoU = 1 - IoU + D + \alpha V$

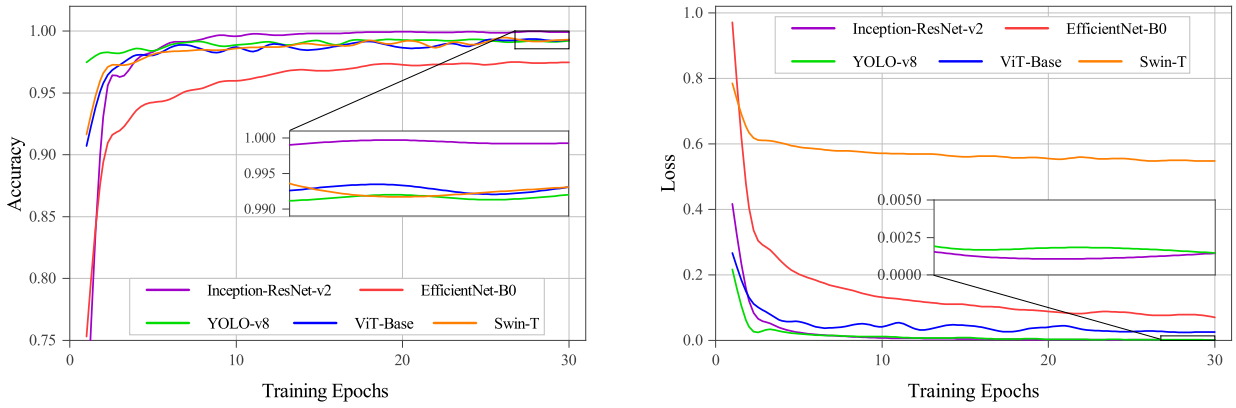
* $D = \frac{\rho^2(b, b^{gt})}{c^2}$, $V = \frac{4}{\pi^2}\left(\arctan\left(\frac{w^{gt}}{h^{gt}}\right) - \arctan\left(\frac{w}{h}\right)\right)^2$ where w and h are the width and height of the predicted box, and w^{gt} and h^{gt} represent the width and height of the ground box, respectively.

6.3 Results

6.3.1 Weed classification: Training epochs and computational cost

Training of the CNN and ViTs models studied was conducted iteratively up to 30 epochs. The goal was to allow the models to effectively capture relevant species features as the training process progressed. Variations in learning progress across iterations led to improved training accuracy and lower loss values (Figure 6.5). An evident trend was observed as accuracies of the five models gradually converged toward a stable level after the first epochs. The greatest effect occurred after the 15^{th} epoch, in which training accuracy exceeded 96 % in all cases following a consistent horizontal pattern. This empirical finding revealed the progressive improvement of the models towards better understanding the dataset complexities, resulting in higher prediction accuracy as the training process advances.

Regarding the loss function, all models converged rapidly with generally small decrease in loss values, with the exception of the Swin-T model which converged with loss values in the range of 0.55-0.60 and, to a lesser extent, the EfficientNet-B0 with loss values in the range of 0.10-0.15. The lowest loss values were observed for the Inception-ResNet-v2 and YOLO-v8 models, reaching 0.0011 in the epoch 28 and 0.0015 in the epoch 30, respectively.



(a) **(b)**

Figura 6.5: Evolution of training accuracy (a) and loss (b) as affected by the training epochs for the CNN and ViT models studied.

The training phase impacted the computational cost of the entire classification process, mainly as a result of the indirect connection between the number of trainable parameters and the size occupied on disk (Table 6.3). As a consequence, the time spent by the models was comparatively much higher in the training and validation phases, and significantly lower in the implementation phase, a vital issue especially in real-time applications. In this regard, the EfficientNet-B0 and Swin-T architectures required significantly fewer trainable parameters (0.3×10^6 and 0.4×10^6 parameters, respectively) than YOLO-v8 (1.5×10^6 parameters) and, particularly, than Inception-Resnet-v2 y ViT-Base (54.3×10^6 and 85.7×10^6 parameters, respectively), which in turn resulted in a time efficiency between 3 and 12 times higher than the other models studied. Particularly remarkable

values were observed for the YOLO-v8 model, which despite its moderate number of trainable parameters and having the smallest size on disk (3 Mb), required a notably longer inference time, especially when performing the unbalanced test.

Tabla 6.3: Computational cost of the studied models in terms of the number trainable parameters, size on disk and time dedicated to each classification phase.

Model	Nº of trainable parameters ($\times 10^6$)	Size on disk (Mb)*	Time (s)*	
			Training/ validation	Inference
CNN				
Inception-Resnet-v2	54.3	654	1302	17
EfficientNet-B0	0.3	21	236	7
YOLO-v8	1.5	3	1282	70
Visual Transformers				
ViT-Base	85.7	346	2528	28
Swin-T	0.4	113	468	14

*Rounded values.

6.3.2 Weed classification: Accuracy and *F1-scores*

The five models showed high overall accuracy (OA) in the classification of weed species, with values ranging from 83.8 % to 98.1 % (Table 6.4). On average, the YOLO-v8 and Swin-T models achieved the best performance from the CNNs and ViTs architectures, respectively, so these two models were selected for further steps regarding weed detection and mapping.

These models were also superior in classifying individual weed species, reaching *F1-score* values above 98 % accuracy for *C. album*, *C. arvensis*, *C. rotundus*, and above 90 % accuracy for the other species, excepting for *A. patula* and *D. ferox*. The impact of class imbalance in the model performance explained the worst results in these two minor species, which is a common issue in ML when some classes are represented much less frequently than others. In fact, an unbalanced inference test was used to assess the models' capability to classify weed species in real-world scenarios, where weeds are unevenly distributed in the crop fields. The varying sample size for each weed species increased the difficulty of the classification task, as the models faced large differences in the number of labels per class.

6.3.3 Weed detection: Training

The performance obtained during training of the two object detectors selected, i.e. YOLO-v8m for the YOLO-v8 model and DETA for the Swin-T model, is shown in Table 6.5. The YOLO-v8m model reported superior average accuracy, outperforming DETA in terms of overall mAP and at an IoU threshold of 50 %, suggesting that YOLOv8m was more effective at detecting objects with lower location accuracy. However, at a more stringent IoU threshold of 75 %, DETA was more robust in detecting the object locations. According to the object size, YOLO-v8m had superior performance in

Tabla 6.4: *F1-score* of the studied weed species classifiers.

		F1-score *%*				
		CNN models			**ViT models**	
Weeds	**Sample Size**	**Inception-ResNet-v2**	**Efficient Net-B0**	**YOLO-v8**	**ViT-Base**	**Swin-T**
A. patula	17	33.7	22.1	35.8	40.5	44.2
C. album	200	98.5	90.8	99.8	99.5	99.0
C. arvensis	200	99.8	99.3	99.7	99.8	100.0
C. rotundus	2500	87.9	96.0	99.2	95.4	98.9
D. ferox	1	6.7	8.7	4.0	2.5	33.3
L. rigidum	14	73.7	93.3	90.3	100.0	100.0
P. oleracea	900	91.7	89.4	95.8	88.5	97.4
S. kali	200	42.9	79.7	96.3	82.8	94.2
S. nigrum	3052	83.2	95.7	96.9	94.3	98.4
OA		83.8	93.6	97.0	93.2	98.1
Macro avg		68.7	75.0	79.8	78.1	85.0
Weighted avg		85.5	94.3	97.5	93.8	98.3

detecting objects of any size. Recall analysis indicated that YOLO-v8m had a better ability to detect at least one object in the image. On the other hand, DETA showed superior performance in detecting up to 10 objects, which may be indicative of a greater ability to handle multi-object scenarios. DETA also notably outperformed YOLO-v8m in detecting up to 100 objects, suggesting greater robustness in highly dense environments. On the other hand, the analysis of the DELTA Recall by object size showed opposite results than mAP for the three sizes. DETA had a significant advantage in terms of recall, which can be crucial in applications where it is important not to miss medium-sized objects.

It is should be noted that the hardware used, specifically the GPU, limited the batch size during object detection models training. This batch size restriction was the main reason to explain the low values observed in the validation metrics. The limitation of the batch size affects the ability of the model to generalize properly, resulting in suboptimal validation performance.

Tabla 6.5: Training performance of the YOLO-v8m and DETA object detectors in terms of mAP and Recall for different model settings.

Metric		YOLO-v8m	DETA
mAP			
	@[IoU=0.50]	0.423	0.255
	@[IoU=0.75]	0.049	0.073
	@[area=small]	0.051	0.038
	@[area=medium]	0.190	0.109
	@[area= large]	0.256	0.185
Recall			
	@[maxDets=1]	0.120	0.096
	@[maxDets=10]	0.229	0.354
	@[maxDets=100]	0.231	0.439
	@[area= small]	0.117	0.215
	@[area=medium]	0.304	0.432
	@[area=large]	0.366	0.580

6.3.4 Weed detection: Inference

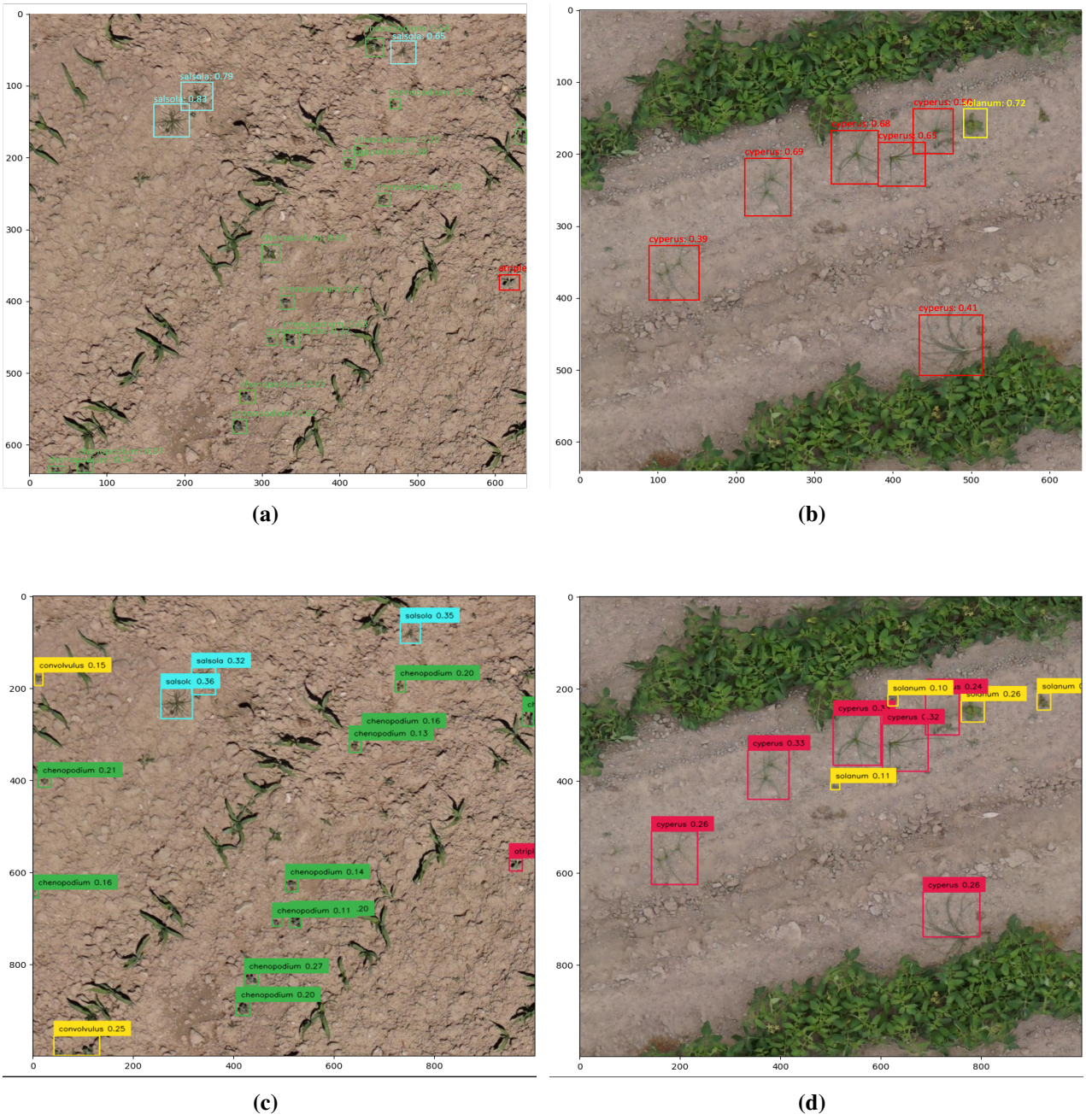
The inference for the object detectors was performed on the partitioned orthomosaics. Computational processing performed by the object detectors on both crops included the number of trained parameters, disk size, training times and inference 6.6.

Tabla 6.6: Training performance of YOLO-v8m and DETA object detectors in terms of model size and computational cost.

Model	Nº of trainable parameters ($\times 10^6$)	Size on disk (Mb)*	Time (s)*		
			Training/ validation	Inference	
				Maize	Tomato
YOLO-v8m	25.9	312	7.592	38	39
DETA	48.1	194	6.238	203	238

* Rounded values (the results shown are less precise, but easier to compare).

The results of YOLO-v8m showed higher probability values compared to those achieved by the DETA model (Figure 6.6).



(a) **(b)** **(c)** **(d)**

Figura 6.6: Detections by the YOLO-v8m (a, b) and DETA (c, d) models in maize (a, c) and tomato (b, d) images showing the boxes and prediction values of the weed species detected.

The number of ground-truth weed plants labelled in the maize field were 263, equally distributed among the five species present. When using YOLO-v8m, 212 were correctly detected and classified (true positive, TP), translated to an 80.6 % success. In the case of DETA, only 125 plants fulfilled both conditions, therefore a 47.5 % success (Table 6.7). For tomato, using YOLO-v8m 600 of the 618 labelled plants were correctly detected and classified (97 % TP), for DETA only 387 plants (62.6 %). The false positive (FP) values can be explained by duplicates, i.e. one single weed plant is detected two or three times, or by an error in the classification. False negative (FN) values were not

detected plants. It was observed that in the case of *D. ferox* some plants were classified as *C. album* when YOLO-v8m was used, and in the case of *P. oleracea* all plants detected were classified as *S. nigra* by the DETA model.

Tabla 6.7: Degree of agreement between species detected and classified using YOLO-v8m and DETA multi-object detectors in maize and tomato fields.

Crop field /	Nº ground-true	YOLO-v8m			DETA		
Weed species	plants	TP	FP	FN	TP	FP	FN
Maize field							
A. patula	52	43	3	9	16	6	36
C. album	52	51	2	1	39	3	13
C. arvensis	52	51	8	1	32	8	20
D. ferox	54	47	6	7	12	3	42
S. kali	53	52	2	1	36	6	17
Total	263	212	21	51	125	26	128
Tomato field							
P. oleracea	207	206	16	1	0	207	1
C. rotundus	204	202	13	2	200	31	4
S. nigrum	207	192	13	15	185	2	22
Total	618	600	42	18	385	240	27
Total	881	812	63	69	510	266	155

TP: True Positive. Plants detected and correctly classified.

FP: False Positive. Plants detected but incorrectly classified or duplicates.

FN: False Negative. Plants that were not detected (i.e. missing plants).

YOLO-v8m obtained a higher detection and correct classification capability (higher TP) in both fields. However, it had a considerable number of FP. DETA showed a significant deficiency in detection and correct classification in both fields, which resulted in DETA not being as reliable than YOLO-v8m for accurate classification in this context.

6.3.5 Weed species mapping and Gridded weed treatment maps

The best model derived from the previous step, i.e. YOLO-v8m, was used to generate the weed species treatment maps according to the three categories and theirs economic weed threshold (Table 6.8, Figure 6.7). In both fields, one species was dominant and had a major presence, i.e. requiring a larger area to be treated. Nevertheless, less than 30 % of the area even considering the area $<$ EWT. The relatively uniform distribution of *C. rotundus* is typical of perennial species that reproduce through tubers, and the *C. album* distribution is in patches. Most of the area did not require any treatment as it was free of weeds, thus ensuring large herbicide savings.

Tabla 6.8: Percentage area per category and species

	No weeds	**< EWT**	**> EWT**
Maize field			
A. patula	98.79	1.21	0
C. album	**73.48**	**23.23**	**3.29**
C. arvensis	86.91	9.03	4.06
D. ferox	99.96	0.04	0
S. kali	98.29	1.71	0
Tomato field			
P. oleracea	95.08	4.71	0.21
C. rotundus	**78.90**	**18.73**	**2.37**
S. nigrum	95.54	3.44	1.04

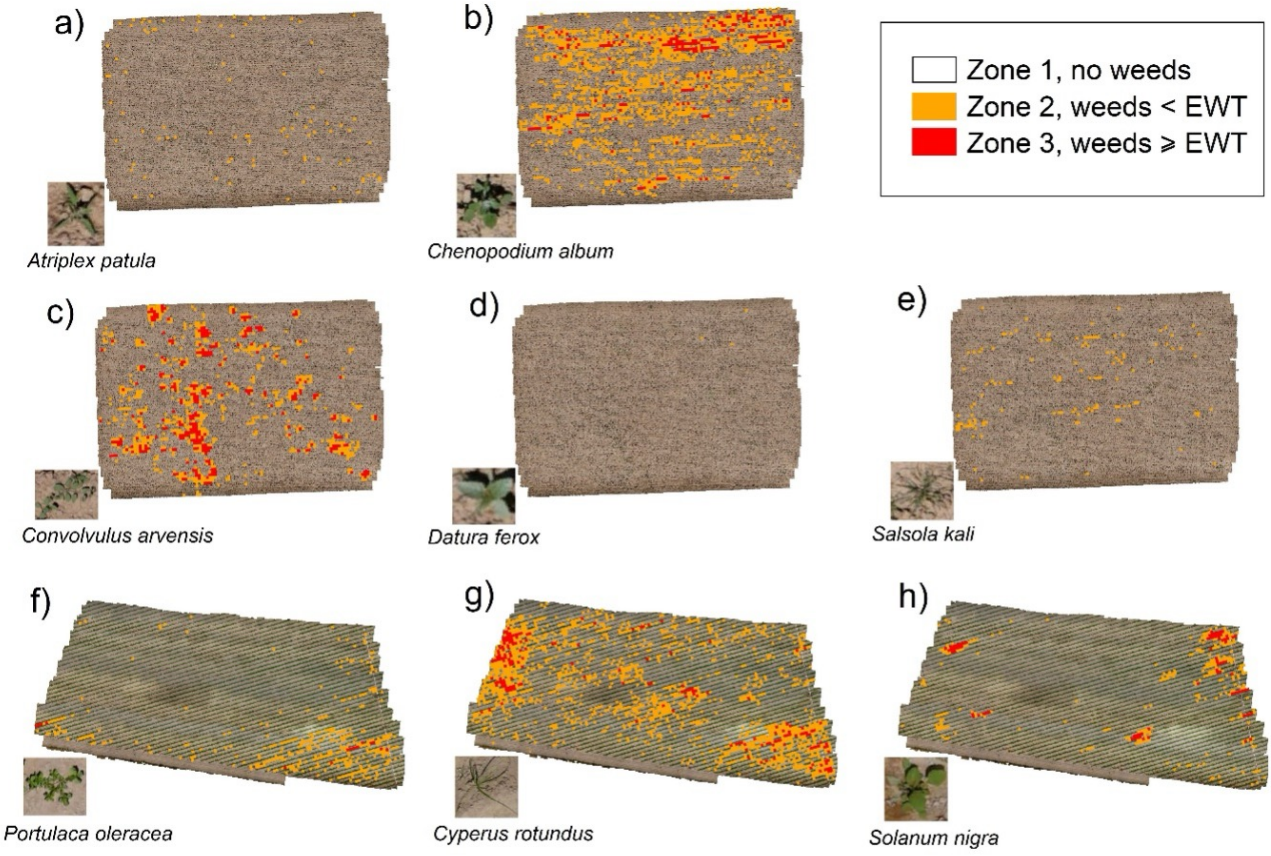


Figura 6.7: Treatment maps according to the weed density of each weed species detected and mapped in the tomato (a-e) and maize (f-h) fields using the YOLO-v8m model.

6.4 Discussion

This research reported reported high accuracy in classifying up to nine weed species in UAV imagery for all five CNN and ViT models, obtaining weighted average *F1-score* between 85.5 % and 98.3 % (section 6.3.2, Table 6.4). Only a few recent studies on weed classification using UAV images reported results in this range (Ajayi y Ashi, 2023; dos Santos Ferreira et al., 2017; Reedha et al., 2022). However, these previous studies did not aim to discriminate between weed species, but rather to discriminate generic classes, i.e. weed *vs*. crops, or weed groups (broadleaf *vs*. grass weeds). One of the exceptions was Wang et al. (2019b), who applied popular AlexNet CNN model to classify three weed species (*C. album*, *Humulus scandens* and *Xanthium sibiricum*) in wheat, peanut and maize and reported an *OA* rate of 99.7 %. Precision of UAV-based weed classification is affected by the difficulty of the task at hand, although the results generally are improving as complexity and opacity of the classifiers progresses. For example, Peña et al. (2013) used object-based image analysis (OBIA), a white-box technique mainly based on image segmentation and feature extraction, to detect weed coverage in maize fields with *OA* 86.0 %. Other authors combined OBIA with black-box classifiers to address more complex tasks, such as detect weeds within herbaceous crop rows using a random forest (RF) classifier with 59.0 % - 88.0 % weed detection accuracy (de Castro et al., 2018), discriminate between broadleaf and grass weeds in sunflower and cotton using multilayer perceptron (MLP) neural networks with 75.0 % and 65.0 % user accuracy, respectively (Torres-Sánchez et al., 2021), and classify four weed species in rice using four different CNN-based classifiers with 87.0 % average accuracy (Huang et al., 2020). Much better results were reported with the Inception-v2 model (*OA* 95.0 %) in detecting weeds in a mixed crop farmland of sugarcane, spinach, banana and pepper (Ajayi y Ashi, 2023), with the CNN-based CaffeNet architecture (*OA* 98.0 % - 99.0 %), a replication of AlexNet model, in classifying broadleaf and grass weeds against soil and soybean classes (dos Santos Ferreira et al., 2017), and with ViT-B32, ViT-B16, EfficientNet-B0, EfficientNet-B1 and ResNet50 (*OA* 98.0 % - 99.0 %) in classifying weeds emerged in beet, parsley and spinach fields (Reedha et al., 2022).

Progress in this domain is also being made with imagery collected with on-ground platforms, with accuracies of more than 95.0 % in weed species classification (Chen et al., 2022; Espejo-Garcia et al., 2023; Olsen et al., 2019). However, the models applied to on-ground imagery may not be well suited to those obtained with UAV platforms, as the resolution and quality of the UAV images is notably inferior than those obtained on the ground (Figure 6.8).

In the task of classifying unbalanced multiclass YOLO-v8 and the two ViTs models slightly outperformed the other two CNN-based models in most weed species. ViTs captured long-range relationships in the data, which is especially beneficial in complex classification tasks, whereas Inception-ResNet-v2 and EfficentNet-B0 may have struggled to capture relevant patterns in some of the weed images that were inherently variable or less representative. Similarly, Dyrmann et al. (2016) and Olsen et al. (2019) reported a tendency of CNN-based models to perform poorly with multiple unbalanced classes, whose hypothesis was also validated by Reedha et al. (2022) by increasing the number of images for both the training and inference phases. In our case, training with an equal number of elements per class avoided the need to apply specific techniques to mitigate the imbalance, such as data augmentation, species weighting, under- or over- sampling, or advanced synthetic sample generation methods.

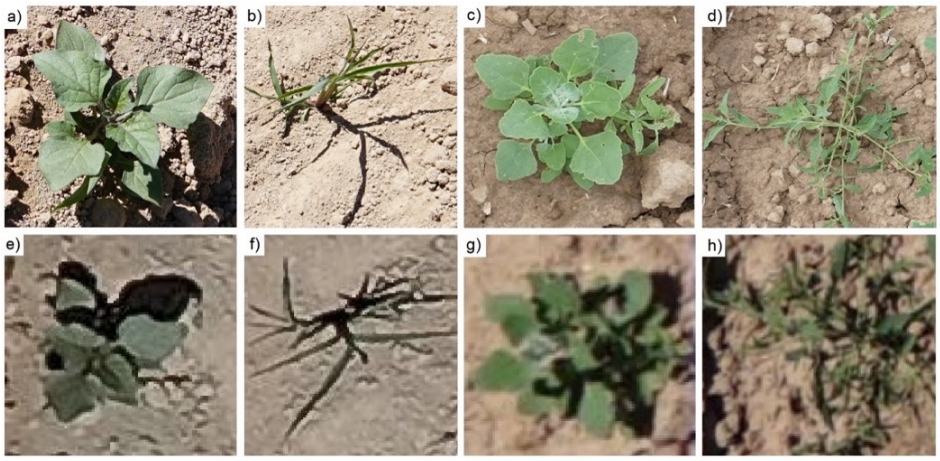


Figura 6.8: Two views of the same weed species taken on the ground (a-d) and with a UAV platform flying at 11-m above ground level (e-h), using a RGB camera model Sony ILCE-6300L in both cases. The weed species are *Solanum nigrum* L. (a, e), *Cyperus rotundus* L. (b, f), *Chenopodium album* L. (c, g), and *Salsola kali* L. (d, h).

Beyond issue about model accuracy so often analyzed, discussion should also focus on relevant details less addressed in previous research, such as the results on training convergence and model size (Figure 6.9), which could have a major impact on transferring these models to actual precision weed management activities.

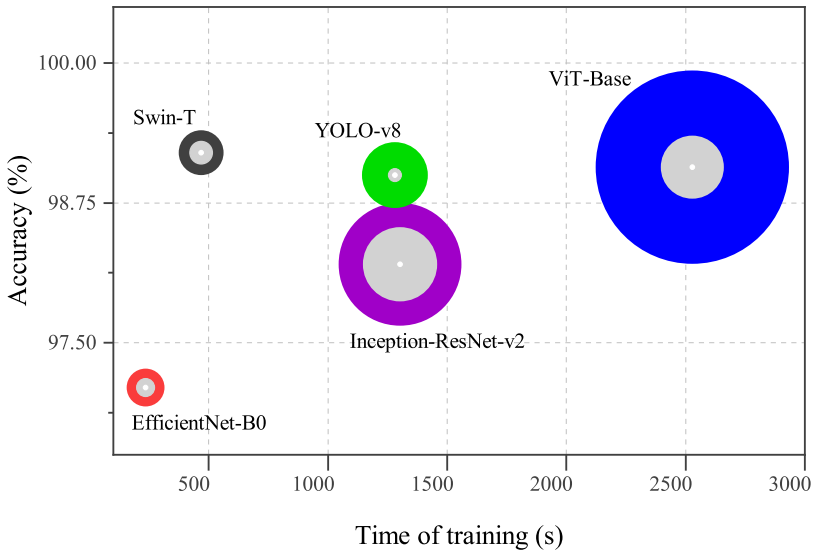


Figura 6.9: Relationship between overall accuracy (%), training time (s), number of trainable parameters (in color circles) and model size (in gray circles). For numbers in graphic circles refer to Table 6.3.

Regarding training convergence (refers to Figure 6.5), Inception-ResNet-v2 and YOLO-v8 achieved maximum accuracy and minimum loss values with 10 epochs, converging faster in the training phase than the other models. This suggests that these models effectively captured essential features from the training data, leading to robust generalization in a relatively short time. ViT-Base and

Swin-T showed slower convergence due, in part, to discerning more subtle data patterns, while EfficientNet-B0 exhibited the lowest rate but with a gradual and consistent convergence trajectory, suggesting that these three models spent more time exploring the training data and could acquire more intricate patterns. Variability of the learning progress across iterations also provided insights into the loss patterns during the training phase. All the models consistently reduced loss values, with YOLO-v8 and Swin-T showing the lowest and highest loss values, respectively. The former had exceptionally low loss rate from the beginning of training, suggesting a possible saturation point in terms of enhancement, while the latter converged with a high loss rate. Future research should identify the optimal parameters for training the Swin-T model while minimizing loss, specifically in the use of UAV-based images.

The relationship between the number of trainable parameters and model performance suggested a complex interaction between the models' ability to learn sophisticated patterns and the time required to process those patterns. Inception-ResNet-v2 and ViT-Base had the longest run times, while EfficientNet-B0 and Swin-T achieved a balance between robust performance and more efficient run times due to their relatively small number of trainable parameters. YOLO-v8 stood out as an interesting case due to its considerably longer run time in the unbalanced test, possibly as a result of the complexity of its trainable parameters.

The training results of the object detectors (validation) revealed that YOLO-v8m has a higher overall average accuracy and is more effective in detecting objects of varying sizes. However, DETA performed better when considering the total number of objects detected (Recall), especially in dense environments and for small and medium-sized objects. This contradiction suggests that the choice between these models should depend on the specific use case: for applications that prioritize detection accuracy (mAP), YOLO-v8m would be more suitable, while for those requiring a high overall detection rate (Recall), DETA could be more beneficial.

YOLO-v8m, despite having fewer trainable parameters and a larger disk size, presented considerably shorter inference times, making it more suitable for applications requiring real-time processing. On the other hand, DETA, with a larger number of parameters and a smaller disk size, proved to be more efficient in terms of training/validation time. The choice between these models will depend on the specific requirements of the application, such as the need for speed versus training and storage capacity.

The results obtained when assessing the knowledge-generating capacity of the maps generated using the ground-truth approach for maize and tomato crops were 80.0 %, 79.1 %, 94.0 % and 61.6 % for the YOLO-v8m and DETA models, respectively. These results were lower compared to the accuracies obtained by (de Camargo et al., 2021; Huang et al., 2018a; Sapkota et al., 2020), which were 91.96 %, 89.16 % and 94 %, respectively. This discrepancy is due to the fact that the number of species used in our research is higher than that used in previous studies. In addition, our proposal focused on the creation of a general model to evaluate two crops.

YOLO-v8m showed superior performance in both detection and accurate classification of weeds in both fields compared to DETA, particularly in the tomato field. Significant differences were observed in the number of plants detected and correctly classified in the both crops, with YOLO-v8m reporting 34.8 % and 33.1 % higher performance in tomato and maize, respectively. This differences might be partially explained by the models' varying ability to handle small objects and morphological

variations. Gao et al. (2024) demonstrated that using deep CNNs, in combination with advanced preprocessing techniques and weighted loss normalization, improves the performance of segmenting both small and large weeds. However, model accuracy can be influenced by spatial resolution and weed variability. For instance, DETA had difficulty detecting and classifying weeds in high-density areas, as observed in tomato crop. This limitation may be due to its reduced ability to handle occlusions and heterogeneity in weed growth, a challenge also discussed by Gao et al. (2024), who noted that weeds present highly diverse patterns in terms of size and morphological characteristics, which makes automatic recognition more complex.

Related to weed treatment maps, the model developed works towards reducing the inputs needed and a sustainable crop production, as it allows major herbicide savings as supported by the results obtained.

As mentioned, our research was pioneering in addressing detection, classification and automatic georeferenced mapping of various early state weed species using UAV imagery acquired over 11 m altitude AGL and cutting-edge models based on CNNs and ViTs. Flight altitude impacts image resolution and quality, being a critical factor that could constraint model accuracy and would merit further research. Moreover, research on UAV-based weed classifiers and detectors is essential to select the most accurate and fastest models for implementing diverse tasks with UAVs equipped with powerful AI-based hardware and telecommunication systems (Mesías-Ruiz et al., 2023). These UAVs could simultaneous percept, analyses and treat weed emergences, thus enabling species- and site- specific weed control in precise real-time operations (i.e. by UAV-based herbicide spraying).

6.5 Conclusions

This study makes a significant contribution not only by showing the competitive performance of ViT-based architectures against complex cutting-edge CNN architectures, but also by providing an automatic technology for accurate detection, classification and mapping of weeds at the species level during the early growth stages. This is especially critical because, in the initial stage of development, morphological and physiological differences between species can be minimal. The models demonstrated remarkable classification performance, when evaluated on an unbalanced dataset, i.e. in a real-world scenario, all models consistently maintained a weighted average *F1-score* above 85 %. The variability in *F1-score* based performance underscores the significance of assessing not only the overall accuracy but also the performance within each individual class. In this context, it is noteworthy that the YOLO-v8 and Swin-T models demonstrated balanced performance across all classes, rendering it a favored option for practical applications. Nevertheless, the choice of the appropriate classification model should consider not only inference capability, but also computational costs. Models with a smaller number of trainable parameters, such as EfficientNet-B0 and Swin-T, demonstrated surprising efficiency in terms of performance and runtime.

The YOLO-v8m and DETA models allowed the creation of detailed geo-referenced weed species maps, which can be used to apply SSWM strategies and significantly reduce the total area treated with herbicides. This not only represents considerable economic savings in production, but also contributes to environmental sustainability by reducing the excessive use of chemicals on crops. In addition, the methodology implemented in this research offers a notable advantage by eliminating the need for traditional sampling-based estimation procedures. Instead of relying on indirect and

potentially inaccurate methods to create weed distribution maps, it allows the use of accurate and up-to-date data provided by deep learning models.

The variations observed in weed infestation levels between tomato and maize fields provide essential information into the dynamics of weed species in relation to different production systems. This highlights the importance of developing tailored weed management strategies for each field, considering not only the specific weed species, but also their density and spatial distribution. The use of advanced precision agriculture tools, such as remotely sensed density maps, enables more efficient and targeted management, optimizing weed control while minimizing herbicide use. This approach not only reduces production costs but also lessens the environmental impact.

In summary, our findings indicate that both ViT and CNN models possess a similar ability to adapt to new data and achieve optimal knowledge generalization, making them as promising tools for SSWM during early crop growth stages. Moreover, considering the importance of efficiency in real-time agriculture operations, this research would serve as a basis for future investigation in the field of precision agriculture and machine learning. This includes exploring model understanding techniques to achieve an optimal balance between accuracy and efficiency, as well as delving into advanced data augmentation techniques and semi-supervised training to improve the robustness and generalizability of model knowledge.

6.6 Appendices

6.6.1 Appendix 1. Customized algorithms

Algorithm 6.1 Cropping and saving labels by species

Input: XML files
Output: processed data stored in different directories
Define: the directories for each class (species)
species_directories $\leftarrow$ { 'lolium' $\rightarrow$ $path\_lolium$, ...}
for each XML file **do**
    read coordinates of bounding box
    $coordinates \leftarrow [ymin : ymax, xmin : xmax]$
    read the class value
    $class \leftarrow$ *class value*
    crop the image
    $imagen\_cut \leftarrow imagen[coordinates]$
    save the cropped image
    if $class$ exists in *species_directories* **then**
        save $imagen\_cut$ in directory *species_directories*
    end if
end for

Algorithm 6.2 Weed georeferencing

Input: Orthomosaic Image
Output: CSV file with calculated values
Initialize: Models
procedure OBJECTDETECTIONANDCENTROIDCALCULATION
    Orthomosaic ← ReadOrthomosaic()
    Chunks ← PartitionIntoChunks($Orthomosaic, l_r \times l_r$)
    for *Chunks* **in** *Chunks* **do**
        Class, BoundingBox, Confidence ← EvaluateModel(*Models, Chunk*)
        Centroid(x_c, y_c) ← CalculateCentroid(*BoundingBox*) ▷ Ec. 6.1
        OrthoCentroid(x_m, y_m) ← TransformToOrthomosaicCoordinates(Centroid) ▷ Ec. 6.2
        SaveToCSV(*Class, OrthoCentroid, Confidence*)
    end for
end procedure

6.6.2 Appendix 2. Classification models

The Inception-ResNet-v2 model (Szegedy et al., 2015) employs a CNN architecture that combines Inception modules and residual ResNet connections to capture complex features and improve accuracy It is noted for its high accuracy and performance in image classification tasks, although it may be less efficient for real-time object detection. It is widely used for image classification and object detection. The EfficientNet-B0 model (Tan y Le, 2020) was optimized for resource efficiency by adjusting network depth, width and resolution, and stands out for its scalability, allowing the model size to be adjusted according to the available computational resources. It offers competitive performance with computational efficiency and is used in various computer vision tasks. You Only Look Once (YOLO)-v8 (Jocher et al., 2023) was designed specifically for real-time object detection, and it is known for its speed and accuracy. It incorporates features from other object detection architectures to predict bounding boxes and class probabilities in a single step. It is suitable for applications where low latency is crucial, such as autonomous driving and surveillance.

ViT Transformers (Dosovitskiy et al., 2021) operates on a patch-based basis, dividing the input image into non-overlapping patches that are then treated as tokens. It stands out for its ability to capture global context in images and performs well in image classification tasks with large datasets. Swin Transformers (Liu et al., 2021): Uses a window-based approach instead of patches, with a hierarchical architecture that allows capturing local and global information efficiently. It stands out for its efficiency and scalability, outperforming other models in certain tasks and being more efficient in the use of resources.

Tabla 6.9: A comparison of the five classification models evaluated in this research.

Model	Characteristics
Inception-ResNet-v2 (Szegedy et al., 2015)	- Architecture: A CNN architecture that combines inception modules (a main branch) and residual connections inspired by ResNet (the residual branch) to capture more complex features and improve accuracy. - Main strengths: High accuracy and performance on image classification tasks. Effective at handling a wide range of image sizes and extracts features across multiple scales and levels of abstraction. - Performance: Performs well in image classification tasks but is less efficient for real-time object detection or instance segmentation. Training can be computationally intensive due to its depth and architectural complexity. - Usage: Primarily used for image classification and object detection tasks.
EfficientNet-B0 (Tan y Le, 2020)	- Architecture: A CNN architecture optimized for resource efficiency by scaling the network's depth, width, and resolution. - Main strengths: Designed to be scalable, allowing the model size and capacity to be adjusted based on available computational resources and specific tasks. - Performance: Provides competitive performance while being computationally efficient. - Usage: A wide range of computer vision tasks, including image classification, object detection, and segmentation.
YOLO-v8 (Jocher et al., 2023)	- Architecture: A CNN model for real-time object detection, known for its speed and accuracy. An improvement over previous YOLO versions, incorporating features from other object detection architectures to predict bounding boxes and class probabilities in a single pass. - Main strengths: Aims to improve the accuracy of object detection while maintaining real-time performance. Suitable for applications where low latency is critical. - Performance: Designed for real-time applications, being known for its high speed and accuracy in these specific tasks. - Usage: Primarily used for autonomous driving, surveillance, robotics, object tracking, and counting.
ViT Transformers (Dosovitskiy et al., 2021)	- Architecture: Operates in a patch-based manner, dividing the input image into non-overlapping patches, which are treated as tokens. These patches are linearly embedded and processed by the transformer encoder architecture. Typically uses absolute positional encodings to provide spatial information to the model. - Main strengths: Effective at capturing global context in images. Strong performance when trained on large datasets. - Performance: Has shown impressive results in image classification tasks, but less well in other tasks like object detection or segmentation. - Usage: Commonly used for image classification tasks, especially when dealing with large-scale datasets.
Swin Transformers (Liu et al., 2021)	- Architecture: Uses a window-based approach rather than patches, dividing the input image into overlapping windows. Introduces a hierarchical architecture where the window-based operations are performed in multiple stages. Uses a relative positional encoding scheme to capture the relative positions between tokens within a window. - Main strengths: Its hierarchical design allows it to capture both local and global information efficiently by reusing the same windows across stages and using a shifting mechanism. Improved efficiency and scalability compared to traditional transformers for vision tasks. - Performance: Has achieved competitive performance with ViT and other architectures, often surpassing them in certain tasks. Designed to be more resource-efficient, requiring fewer parameters than ViT while achieving comparable performance. - Usage: Widely used for various computer vision tasks, including image classification, object detection, and image segmentation, particularly when efficiency and scalability are important.