

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE CIENCIAS EXACTAS NATURALES Y AMBIENTALES

CARRERA DE CIENCIA DE DATOS

Aprendizaje profundo para el reconocimiento de polinizadores: una revisión sistemática de arquitecturas, métricas y factores de desempeño

Monografía previa a la obtención del título de Ingeniero en Ciencia de Datos

Melani Nayely Loor Pérez

Quito, 2026

Certifico que la Monografía de Ciencia de Datos de **Melani Nayely Loor Pérez** ha sido concluida de conformidad con las normas establecidas; por lo tanto, puede ser presentada para la calificación correspondiente.



Andrés Esteban Merino Toapanta

Tutor de la monografía

Quito, 25 de junio de 2026

DEDICATORIA

A mis padres, Héctor y Rosita, por su esfuerzo constante, su apoyo incondicional y por haberme dado las oportunidades que hicieron posible llegar hasta este momento.

AGRADECIMIENTOS

A mis padres, Héctor y Rosita, quienes han sido el pilar fundamental en cada etapa de mi formación. Gracias por su apoyo incondicional, por su esfuerzo constante y por creer en mí en todo momento. Todo lo que he alcanzado hasta ahora es, en gran parte, resultado de su dedicación, sus valores y el amor con el que me han acompañado siempre.

A mi hermano, Andrés, por su apoyo, compañía y por estar presente a lo largo de este proceso. Gracias por los momentos compartidos y por tu respaldo constante.

A mi tutor de monografía, Andrés Merino, por su guía, paciencia y compromiso durante el desarrollo de este trabajo. Sus observaciones, recomendaciones y acompañamiento académico fueron esenciales para orientar este proceso y fortalecer la calidad de la investigación.

Al equipo dinamita, Sofía y Martín, por su apoyo, compañía y por hacer más llevadero este camino. Gracias por estar presentes tanto en los momentos de estrés como en los de alegría y por motivarme a seguir adelante.

Finalmente, a mi gatito, Itzaé, por acompañarme en cada madrugada de trabajo, por su presencia constante durante los momentos más exigentes y por ser una fuente silenciosa de calma y compañía.

TABLA DE CONTENIDOS

1. Resumen **1**

2. Abstract **2**

3. Introducción **3**

4. Objetivos **7**

4.1. Objetivo general 7

4.2. Objetivos específicos 7

5. Fundamento teórico **8**

5.1. Polinizadores y su monitoreo 8

5.2. Fundamentos del aprendizaje profundo 9

5.2.1. Diferencias entre aprendizaje automático tradicional y aprendizaje profundo 10

5.3. Visión computacional 11

5.4. Redes Neuronales 13

5.4.1. Arquitectura general de una red neuronal profunda 13

5.4.2. Redes Neuronales Convolucionales 17

5.5. Métricas de desempeño 28

5.5.1. Matriz de confusión 29

5.5.2. Métricas para clasificación 29

5.5.3. Métricas para detección 30

5.5.4. Métricas para segmentación 30

5.6. Factores que influyen en el rendimiento 31

5.6.1. Aprendizaje por transferencia 32

5.6.2. Aumento de datos 33

5.6.3. Métodos de optimización 37

6. Metodología **39**

6.1. Tipo de estudio 39

6.2. Protocolo de revisión 39

6.3. Preguntas de investigación 39

6.4. Fuente de información y estrategia de búsqueda 40

6.5. Criterios de inclusión y exclusión 42
6.6. Proceso de selección de estudios 43
6.7. Resumen del proceso de selección 45
6.8. Extracción de datos 46
6.9. Análisis de datos 47

7. Resultados y discusión **48**
7.1. RQ1. Arquitecturas y tipos de modelos más utilizados 48
7.2. RQ2. Métricas de evaluación más reportadas 54
7.3. RQ3. Factores técnicos asociados 58
7.4. Discusión 63

8. Conclusiones **67**

9. Referencias **69**

ÍNDICE DE TABLAS

5.1. Comparación de modelos de CNN. 28
5.2. Métricas de clasificación 29
5.3. Métricas por tarea 31

6.1. Palabras clave utilizadas para la estrategia de búsqueda. 40
6.2. Criterios de inclusión y exclusión 43
6.3. Preguntas de elegibilidad 45

7.1. Tipos de tarea reportados por estudio. 49
7.2. Arquitecturas generales de aprendizaje profundo reportadas por estudio. 51
7.3. Arquitecturas específicas reportadas en los estudios incluidos. 53
7.4. Métricas de evaluación reportadas por estudio. 54
7.5. Relación comparativa entre tarea, tipo de modelo y métricas predominantes . . . 57
7.6. Factores técnicos generales reportados por estudio. 58
7.7. Estrategias técnicas específicas reportadas en los estudios incluidos. 60

ÍNDICE DE FIGURAS

5.1. Representación esquemática de tareas de visión por computador. 12
5.2. Arquitectura general de una red neuronal profunda. 15
5.3. Esquema general de una CNN 22
5.4. Estructura general de EfficientNet. 23
5.5. Estructura general de MobileNet. 24
5.6. Estructura general de Inception. 24
5.7. Estructura general de ResNet. 25
5.8. Estructura general de DenseNet. 26
5.9. Estructura general del modelo YOLO. 26
5.10. Estructura general del modelo Faster R-CNN. 27
5.11. Estructura general del modelo Mask R-CNN. 27
5.12. Esquema general del aprendizaje por transferencia. 32
5.13. Ejemplo de transformaciones geométricas. 34
5.14. Ejemplo de transformaciones fotométricas. 34
5.15. Ejemplo de ruido y desenfoque. 35
5.16. Ejemplo de técnicas avanzadas de aumento de datos. 36
5.17. Ejemplo de segmentación y aumento basado en objetos. 37

6.1. Proceso de selección de estudios basado en PRISMA. 46

7.1. Frecuencia de tipos de tarea reportados en los estudios analizados. 50
7.2. Frecuencia de arquitecturas de aprendizaje profundo reportadas en los estudios analizados. 52
7.3. Frecuencia de métricas de evaluación reportadas en los estudios analizados. . . . 56
7.4. Frecuencia general de factores técnicos reportados en los estudios analizados. . . . 60

1. RESUMEN

En este trabajo se realizó una revisión sistemática de literatura sobre el reconocimiento de imágenes de polinizadores mediante aprendizaje profundo, con énfasis en arquitecturas, métricas de evaluación y factores técnicos asociados al rendimiento. La búsqueda se efectuó en Scopus mediante una estrategia estructurada, siguiendo el enfoque PRISMA para la identificación, selección y depuración de los estudios, junto con criterios de inclusión y exclusión previamente definidos. El corpus final estuvo conformado por 53 estudios publicados desde 2020, centrados en tareas de clasificación, detección y segmentación. Los resultados evidencian un predominio de la detección y una presencia destacada de la familia YOLO, junto con arquitecturas como ResNet, MobileNet, EfficientNet, Faster R-CNN y Mask R-CNN. Las métricas más reportadas fueron exactitud, precisión y sensibilidad, mientras que en detección destacaron mAP, IoU y FPS. Además, se identificó que el rendimiento depende de la arquitectura y de estrategias como aumento de datos, aprendizaje por transferencia y métodos de optimización. La evidencia revisada muestra avances relevantes en el reconocimiento automatizado de polinizadores e indica que la comparación por tipo de tarea, modelo y métrica permite ordenar una literatura metodológicamente heterogénea.

Palabras clave: Aprendizaje profundo, Métricas de desempeño, Polinizadores, Reconocimiento de imágenes, Visión por computador.

2. ABSTRACT

This study conducted a systematic literature review on pollinator image recognition using deep learning, with emphasis on architectures, evaluation metrics, and technical factors associated with model performance. The search was carried out in Scopus through a structured search strategy, following the PRISMA approach for the identification, screening, and selection of studies, together with previously defined inclusion and exclusion criteria. The final corpus consisted of 53 studies published since 2020, focused on classification, detection, and segmentation tasks. The results show a predominance of detection tasks and a notable presence of the YOLO family, along with architectures such as ResNet, MobileNet, EfficientNet, Faster R-CNN, and Mask R-CNN. The most frequently reported metrics were accuracy, precision, and recall, while mAP, IoU, and FPS were prominent in detection tasks. In addition, model performance was found to depend on the architecture and on strategies such as data augmentation, transfer learning, and optimization methods. The reviewed evidence shows relevant advances in the automated recognition of pollinators and indicates that comparison by task type, model, and metric makes it possible to organize a methodologically heterogeneous body of literature.

Keywords: Computer vision, Deep learning, Image recognition, Performance metrics, Pollinators.

3. INTRODUCCIÓN

El aprendizaje profundo se ha consolidado como una de las líneas más influyentes de la inteligencia artificial aplicada al análisis de imágenes. Su desarrollo reciente ha estado acompañado por avances en capacidad computacional, disponibilidad de datos y diseño de arquitecturas cada vez más eficaces para procesar información visual de alta complejidad (Herath, 2025; Mienie y Svart, 2024). En este marco, la visión por computador ha experimentado una expansión sostenida, especialmente en tareas como clasificación, detección y segmentación, en las que los modelos pueden aprender representaciones jerárquicas a partir de los datos sin necesidad de diseñar manualmente todas las características relevantes.

Dentro de este campo, las redes neuronales convolucionales han ocupado un lugar central. Su importancia radica en que permiten capturar patrones espaciales, reconocer estructuras visuales locales y combinarlas progresivamente en representaciones más abstractas (Khan et al., 2020; Raj y Yos, 2025). Gracias a ello, las redes neuronales convolucionales (*Convolutional Neural Networks*, CNN) y sus variantes se han consolidado como una base técnica dominante en numerosos problemas de reconocimiento visual, especialmente cuando los objetos de interés presentan variaciones de forma, tamaño, posición, iluminación o fondo.

La relevancia de estas arquitecturas no se limita a contextos industriales o comerciales. También ha alcanzado dominios científicos en los que el reconocimiento visual constituye una tarea clave para la generación de conocimiento. Entre ellos se encuentra el estudio de la biodiversidad, ámbito en el que la identificación de organismos a partir de imágenes ha ganado importancia como apoyo al monitoreo ecológico, la documentación de especies y la observación sistemática de cambios en comunidades biológicas (Borowiec et al., 2022; Høye et al., 2021). En este escenario, el uso de aprendizaje profundo representa una oportunidad para fortalecer procesos de análisis que tradicionalmente han dependido de trabajo manual intensivo.

Uno de los casos más relevantes dentro de esta intersección entre visión por computador y ecología es el reconocimiento de polinizadores a partir de imágenes. Los polinizadores, particularmente abejas, abejorros y otros insectos asociados a la polinización, cumplen una función

esencial en la reproducción de numerosas plantas y en el mantenimiento de la diversidad biológica (Ghosh y Hasan, 2025). Por esta razón, su observación e identificación resultan valiosas para el seguimiento de comunidades ecológicas, la evaluación de cambios en poblaciones y la comprensión de procesos vinculados con la conservación (Hellwig et al., 2024).

A pesar de esta importancia, la identificación de polinizadores sigue presentando dificultades prácticas y metodológicas. Gran parte del reconocimiento de especies continúa dependiendo de observación experta, análisis morfológico y revisión manual de imágenes. Estos enfoques requieren tiempo, entrenamiento especializado y condiciones de captura relativamente favorables para distinguir organismos que, en muchos casos, presentan una alta similitud visual entre sí (Spiesman et al., 2021). Además, la variabilidad de iluminación, postura, escala, movimiento y fondo puede dificultar aún más la diferenciación entre especies o grupos taxonómicos (Ratnayake et al., 2021).

Estas limitaciones hacen evidente la necesidad de contar con herramientas computacionales capaces de asistir o automatizar parte del proceso de reconocimiento. En este punto, el aprendizaje profundo aparece como una alternativa especialmente prometedora. En particular, las redes neuronales convolucionales permiten aprender directamente a partir de imágenes rasgos visuales que serían difíciles de formalizar mediante reglas manuales. Esta capacidad resulta especialmente relevante cuando se trabaja con organismos pequeños, con alta variación intraespecífica o con entornos visuales complejos, características frecuentes en el reconocimiento de polinizadores (Borowiec et al., 2022; Høye et al., 2021).

El potencial de estas técnicas ya se refleja en estudios recientes que reportan aplicaciones de aprendizaje profundo para clasificación de especies, detección de insectos en imágenes y monitoreo automatizado en tiempo real (Bjerge et al., 2022; Stark et al., 2023). Asimismo, se han propuesto sistemas orientados a contextos específicos de observación, lo que evidencia que el reconocimiento automatizado de polinizadores mediante imágenes ha dejado de ser una posibilidad teórica para convertirse en una línea activa de investigación aplicada (Stark et al., 2025; Urbano Espinosa et al., 2025). Sin embargo, este crecimiento no ha ido acompañado de una síntesis suficientemente clara que permita comparar enfoques, métricas y decisiones técnicas dentro del campo.

Al revisar la literatura disponible, se observa que los avances no se presentan de forma homogénea ni fácilmente comparable. Los estudios difieren en los grupos de polinizadores analizados, en el tipo de imágenes empleadas, en las tareas abordadas y en las arquitecturas implementadas. Mientras algunos trabajos se concentran en clasificación, otros priorizan detección, seguimiento automatizado o segmentación; además, las condiciones de captura, el tamaño de los conjuntos de datos y el nivel de detalle taxonómico también varían de manera importante (Proença et al., 2025; Spiesman et al., 2021). Los estudios son heterogéneos, por lo que requieren una síntesis organizada que permita compararlos dentro de grupos metodológicamente semejantes.

La comparación entre estudios también se ve limitada por la diversidad de métricas reportadas y por la descripción desigual de las decisiones técnicas que acompañan el entrenamiento de los modelos. No todos los trabajos informan las mismas medidas de desempeño, ni lo hacen con el mismo nivel de detalle; además, factores como la disponibilidad limitada de datos etiquetados, el desbalance entre clases, la variabilidad morfológica intra e interespecífica, la transferencia de aprendizaje, el aumento de datos y los métodos de optimización pueden influir de manera importante en los resultados (Khanzhina et al., 2022).

En este contexto, la ausencia de una síntesis estructurada constituye un problema académico concreto, ya que dificulta responder con claridad qué arquitecturas se utilizan con mayor frecuencia, qué métricas predominan y qué factores técnicos se asocian con mejores niveles de desempeño reportado.

La pertinencia de esta investigación radica en la necesidad de organizar y comparar una literatura reciente que, aunque valiosa, permanece dispersa. Una revisión sistemática permite reunir estudios bajo criterios explícitos de selección, sintetizar sus enfoques y ofrecer una visión integrada de los componentes más relevantes del campo. En áreas afines, este tipo de revisión ha resultado útil para clarificar el desarrollo de arquitecturas, patrones de evaluación y desafíos persistentes (Pacal et al., 2024; S. Wang et al., 2025).

En el caso del reconocimiento de polinizadores mediante aprendizaje profundo, una síntesis de esta naturaleza puede servir para describir el estado del arte y orientar futuras decisiones

metodológicas en torno a la selección de modelos, métricas y estrategias técnicas.

Desde esta perspectiva, la presente monografía tiene como propósito caracterizar el estado del arte del reconocimiento de imágenes de polinizadores mediante aprendizaje profundo, a partir de estudios científicos publicados desde 2020. El análisis se centra en tareas de clasificación, detección y segmentación, con énfasis en el papel de las redes neuronales convolucionales y de otras arquitecturas relacionadas, por constituir uno de los enfoques más representativos en problemas de reconocimiento visual.

De manera más específica, el estudio busca identificar las arquitecturas y tipos de modelos más utilizados, sintetizar las métricas de evaluación reportadas en la literatura y analizar los principales factores técnicos asociados al rendimiento de los modelos. En conjunto, estas dimensiones permiten organizar la evidencia reciente desde una perspectiva comparativa y ofrecer una visión integrada de un campo que ha crecido de manera notable, pero cuya consolidación metodológica aún presenta limitaciones importantes.

4. OBJETIVOS

En esta sección se presentan los objetivos que orientan el desarrollo de la revisión sistemática. Estos permiten delimitar el alcance del estudio y organizar el análisis de la literatura científica.

4.1 OBJETIVO GENERAL

Caracterizar el estado del arte en el reconocimiento de imágenes de polinizadores en aprendizaje profundo, mediante una revisión sistemática de estudios científicos desde 2020 sobre redes neuronales en clasificación, detección o segmentación.

4.2 OBJETIVOS ESPECÍFICOS

A partir del objetivo general, se establecen los siguientes objetivos específicos, los cuales permiten descomponer el análisis en dimensiones concretas relacionadas.

- Identificar los modelos y arquitecturas de aprendizaje profundo en el reconocimiento de imágenes de polinizadores y las configuraciones de redes neuronales.
- Comparar las métricas de desempeño de la literatura científica, tales como precisión, exactitud, sensibilidad y medida F1, de acuerdo con la tarea y el tipo de modelo.
- Analizar los factores técnicos que se asocian con el rendimiento de los modelos, como la transferencia de aprendizaje, el aumento de datos y los métodos de optimización.

5. FUNDAMENTO TEÓRICO

5.1 POLINIZADORES Y SU MONITOREO

Los polinizadores son esenciales para la reproducción de plantas y el mantenimiento de la biodiversidad global. Se estima que aproximadamente el 87 % de las plantas con flores y el 75 % de los cultivos agrícolas dependen de la polinización para su reproducción (Ghosh y Hasan, 2025). Este servicio ecológico es crucial para los ecosistemas naturales y también la seguridad alimentaria humana, sin embargo, su declive, debido a la destrucción del hábitat, el cambio climático y la contaminación, representa una amenaza directa para la seguridad alimentaria y la biodiversidad global. Este problema ha escalado en importancia en la última década, haciendo urgente la implementación de métodos más eficaces y precisos para su monitoreo (Hellwig et al., 2024).

Tradicionalmente, los métodos de monitoreo de polinizadores han consistido en observación directa y recolección de muestras, los cuales dependen en gran medida de la experiencia del observador y del tiempo disponible para la recolección de datos. Aunque son efectivos, estos métodos tienen limitaciones significativas: son propensos a errores humanos, especialmente cuando se trata de especies morfológicamente similares; dependen de expertos, lo que reduce la consistencia y calidad de los datos obtenidos; y son intensivos en trabajo, lo que limita la frecuencia y extensión de los monitoreos (Batz et al., 2023).

En la actualidad, los avances en aprendizaje profundo *(Deep Learning)* ofrecen soluciones innovadoras para superar estas limitaciones ya que, muestran una gran precisión y eficiencia en la detección automatizada de polinizadores a partir de imágenes aéreas, superando a los métodos tradicionales en términos de rapidez y exactitud. Este enfoque permite automatizar la clasificación de polinizadores, reducir la dependencia de expertos y minimizar los errores humanos, mejorando la fiabilidad de los datos (Z. Wang et al., 2025).

Además, la integración de tecnologías como drones con modelos de aprendizaje profundo facilita el monitoreo no invasivo a gran escala, lo cual mejora la calidad y frecuencia del monitoreo

y permite la recolección de datos en diversas condiciones y en lugares de difícil acceso, sin interferir en el comportamiento de los polinizadores (Kim et al., 2025).

No obstante, el monitoreo automatizado en condiciones reales de campo introduce desafíos técnicos que pueden degradar el desempeño de los modelos de visión por computador. Entre los principales se encuentran la variabilidad de iluminación natural, los fondos complejos que dificultan separar el insecto del entorno, la oclusión parcial por vegetación y el desenfoque por movimiento debido a la rapidez de los polinizadores.

Adicionalmente, el tamaño reducido de muchas especies, la variación de escala y punto de vista en capturas aéreas o con cámaras fijas, y la alta similitud visual entre taxones incrementan la probabilidad de confusión en la clasificación. A ello se suman dificultades propias de los datos ecológicos, como el desbalance de clases, la disponibilidad limitada de imágenes etiquetadas y el *domain shift*, que ocurre cuando un modelo entrenado en un escenario pierde capacidad de generalización al aplicarse en ambientes, estaciones o configuraciones de captura distintas (Ștefan et al., 2025).

En cuanto a la conservación de los polinizadores, el uso de modelos de aprendizaje profundo en combinación con programas de monitoreo a gran escala y la integración de la ciencia se presentan como herramientas clave para comprender mejor las dinámicas de las poblaciones de polinizadores y mitigar su declive (Hellwig et al., 2024).

5.2 FUNDAMENTOS DEL APRENDIZAJE PROFUNDO

El aprendizaje profundo es una subdisciplina del aprendizaje automático que se centra en modelar datos a través de múltiples capas de neuronas, utilizando estructuras complejas y transformaciones no lineales (Bhargavi, 2020). Este enfoque está inspirado en el funcionamiento del cerebro humano, permitiendo procesar información a través de niveles jerárquicos para entender representaciones y características a partir de datos sin procesar.

A través de esta jerarquización, el aprendizaje profundo ha sido capaz de abordar problemas de alta complejidad, como el reconocimiento de imágenes, el procesamiento del lenguaje

natural y la predicción de patrones en grandes volúmenes de datos. Una de las principales características del aprendizaje profundo es su uso de modelos de redes neuronales profundas, las cuales sirven como puente para representar grandes cantidades de datos y aprender relaciones complejas dentro de ellos (D. Sharma, 2020).

El aprendizaje profundo ha mostrado un rendimiento destacado en varias áreas, lo que demuestra su versatilidad y eficacia. En visión por computadora, se ha utilizado para tareas como la clasificación de imágenes y la detección de objetos, mientras que en el reconocimiento de voz ha facilitado el procesamiento y reconocimiento de patrones vocales (Zulkernine et al., 2020), permitiendo avances en sistemas de asistencia virtual. En el ámbito de la biología, se utiliza para tareas como la identificación de especies, el monitoreo ambiental y el modelado ecológico, contribuyendo a la conservación de la biodiversidad y la mejora de la sostenibilidad ambiental (Borowiec et al., 2022).

5.2.1 DIFERENCIAS ENTRE APRENDIZAJE AUTOMÁTICO TRADICIONAL Y APRENDIZAJE PROFUNDO

El aprendizaje automático tradicional suele depender de ingeniería de características, es decir, del diseño manual de descriptores antes de entrenar el modelo. Esto exige conocimiento del dominio y puede limitar el desempeño cuando la variabilidad visual es alta (J.-J. Chen et al., 2024; Rajasree et al., 2025).

El aprendizaje profundo, en cambio, se apoya en aprendizaje de representaciones, donde redes neuronales aprenden automáticamente características jerárquicas a partir de datos sin procesar. Este enfoque ha mostrado alta eficacia en reconocimiento, detección y segmentación de objetos (Khastavaneh y Ebrahimpour-Komleh, 2020).

Los métodos tradicionales pueden ser competitivos con menos datos cuando las características son informativas, mientras que el aprendizaje profundo suele requerir más datos etiquetados, aunque su rendimiento puede sostenerse en escenarios de datos reducidos mediante configuraciones adecuadas (Edwin Selva Rex et al., 2022; Rajasree et al., 2025).

En términos prácticos, el aprendizaje profundo demanda mayor cómputo y memoria, por lo que se beneficia del uso de GPU y análisis de eficiencia. Además, suele presentar menor interpretabilidad directa, pese a que generalmente logra mayor desempeño en problemas visuales complejos; en sistemas con restricciones energéticas, la eficiencia de las representaciones es un factor crítico (J.-J. Chen et al., 2024; Štaka et al., 2025).

5.3 VISIÓN COMPUTACIONAL

La visión computacional es un campo dentro de la inteligencia artificial que permite a las máquinas interpretar y comprender el contenido visual de imágenes y videos, de manera similar a como lo haría un ser humano. En el contexto de la identificación de polinizadores, la visión computacional utiliza redes neuronales convolucionales (CNN, por sus siglas en inglés *Convolutional Neural Networks*) para realizar tareas como clasificación, detección y segmentación de imágenes.

La clasificación implica asignar una etiqueta a una imagen completa, como identificar la especie de un insecto en una fotografía. En el ámbito de la identificación de polinizadores, la clasificación es crucial para distinguir entre diferentes especies de insectos. Un estudio reciente utilizó CNN para clasificar insectos en imágenes de alta resolución capturadas por trampas de cámara, logrando una precisión media (mAP, por sus siglas en inglés *Mean Average Precision*) del 54,23 % en la clasificación a nivel de orden (Sa'doun et al., 2024).

En cuanto a la detección, esta implica identificar y localizar objetos específicos dentro de una imagen. En el caso de los polinizadores, la detección se refiere a localizar la presencia de abejas en imágenes de un campo o una colmena. En esta línea, (Sa'doun et al., 2024) desarrollaron un estudio aplicado de monitoreo ambiental de alta resolución, en el que emplearon la red YOLOv8 para detectar insectos polinizadores en imágenes obtenidas mediante trampas de cámara macro. Sus resultados mostraron el potencial de este enfoque para recolectar datos precisos a nivel de especie.

Por su parte, la segmentación es una técnica que permite dividir una imagen en regiones significativas, como separar un polinizador del fondo (Indu y Priyadharsini, 2023). Existen dos

tipos principales de segmentación: segmentación semántica y segmentación de instancias.

La segmentación semántica asigna una etiqueta a cada píxel de la imagen, un estudio utilizó U-Net, una arquitectura popular para la segmentación de imágenes, para segmentar imágenes de epífitas adquiridas con drones, evaluando los resultados con el índice de Jaccard y el puntaje Dice (Raheesh et al., 2022). Por otro lado, la segmentación de instancias distingue entre diferentes objetos de la misma clase, se aplica en redes como Mask R-CNN, que ha sido utilizada para identificar automáticamente lesiones en hojas de tomate, logrando una precisión promedio del 73,00 % (P. Kaur et al., 2024).

Figura 5.1

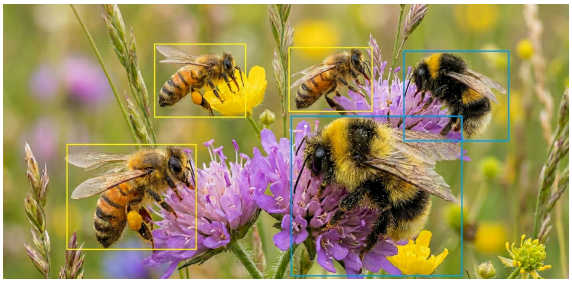
Representación esquemática de tareas de visión por computador.

(a) *Clasificación*



Identificación de la clase principal en la imagen: Abeja.

(b) *Detección de objetos*



Localización y clasificación de abejas y abejorros mediante cajas delimitadoras.

(c) *Segmentación semántica*



Asignación de una categoría a cada píxel según la clase del objeto.

(d) *Segmentación de instancia*



Separación individual de cada polinizador como una instancia distinta.

Nota. Elaboración propia.

La Figura 5.1 presenta una representación esquemática de las principales tareas de visión por computador aplicadas al reconocimiento de polinizadores. En clasificación, el modelo asigna

una única etiqueta al objeto principal de la imagen. En detección, identifica y localiza múltiples objetos mediante cajas delimitadoras. En segmentación semántica, clasifica regiones de la imagen de acuerdo con su clase, sin distinguir individuos de una misma categoría. Finalmente, en segmentación de instancia, separa cada objeto de forma individual, incluso cuando varios pertenecen a la misma clase.

5.4 REDES NEURONALES

Las redes neuronales son sistemas computacionales inspirados en el cerebro humano, compuestos por múltiples capas de neuronas artificiales que procesan información y aprenden a realizar tareas específicas mediante el ajuste de sus conexiones internas. Existen varias arquitecturas de redes neuronales, siendo las redes neuronales convolucionales las más utilizadas en el reconocimiento de imágenes debido a su capacidad para aprender y generalizar patrones complejos (Stark et al., 2025).

En el reconocimiento de imágenes, las técnicas de extracción de características juegan un papel crucial, ya que permiten identificar y representar elementos visuales importantes de una imagen, facilitando su clasificación y análisis (Sajib et al., 2025). Los beneficios de las redes neuronales en el reconocimiento de imágenes incluyen alta precisión en tareas de clasificación, la automatización de procesos que antes requerían intervención humana, y su adaptabilidad a diferentes tipos de datos y tareas, lo que las hace muy versátiles en diversas aplicaciones.

5.4.1 ARQUITECTURA GENERAL DE UNA RED NEURONAL PROFUNDA

Una red neuronal profunda se organiza en una jerarquía de capas que transforma progresivamente los datos de entrada hasta producir una salida, usualmente una predicción de clase o una estimación continua. En visión por computador, esta organización se concreta en bloques que extraen y combinan características, lo que permite construir representaciones internas cada vez más informativas para la tarea (Khan et al., 2020).

En este tipo de modelos, las capas iniciales aprenden representaciones de bajo nivel, como

bordes, texturas y patrones simples, mientras que las capas más profundas integran dichas señales para formar representaciones de mayor nivel asociadas con formas y estructuras complejas, facilitando la discriminación de objetos o categorías visuales. Esta propiedad constituye la base del aprendizaje jerárquico de representaciones en redes profundas (Zakeri et al., 2023; Zurowietz y Nattkemper, 2020).

La capacidad de una red para modelar relaciones complejas depende, entre otros aspectos, de su profundidad y su ancho. Aumentar la profundidad incrementa el número de transformaciones sucesivas y favorece la construcción de representaciones más abstractas, mientras que aumentar el ancho incrementa la cantidad de unidades por capa y, con ello, la diversidad de patrones que pueden ser capturados en una misma etapa de procesamiento (Khan et al., 2020).

En aplicaciones de clasificación de imágenes, es común que la arquitectura incluya capas que extraen características, mecanismos de reducción de dimensionalidad y capas finales que integran la información para producir la predicción. Esta organización permite que el modelo aprenda representaciones compactas que preservan información relevante y reduzcan redundancias, mejorando la capacidad de generalización cuando se diseña y ajusta adecuadamente (Chandraprakash et al., 2024; Khan et al., 2020).

CAPAS

Una red neuronal profunda se estructura en capas con funciones diferenciadas dentro del proceso de propagación hacia adelante. Cada capa transforma progresivamente la representación de los datos, permitiendo que el modelo construya descripciones internas cada vez más adecuadas para la tarea de clasificación.

La capa de entrada recibe los datos en forma numérica y establece la representación inicial sobre la cual operará el modelo. Su función principal es garantizar que la información sea correctamente incorporada al sistema, sin realizar transformaciones complejas. (Joselin et al., 2024).

Las capas ocultas constituyen el núcleo del proceso de aprendizaje. A través de transfor-

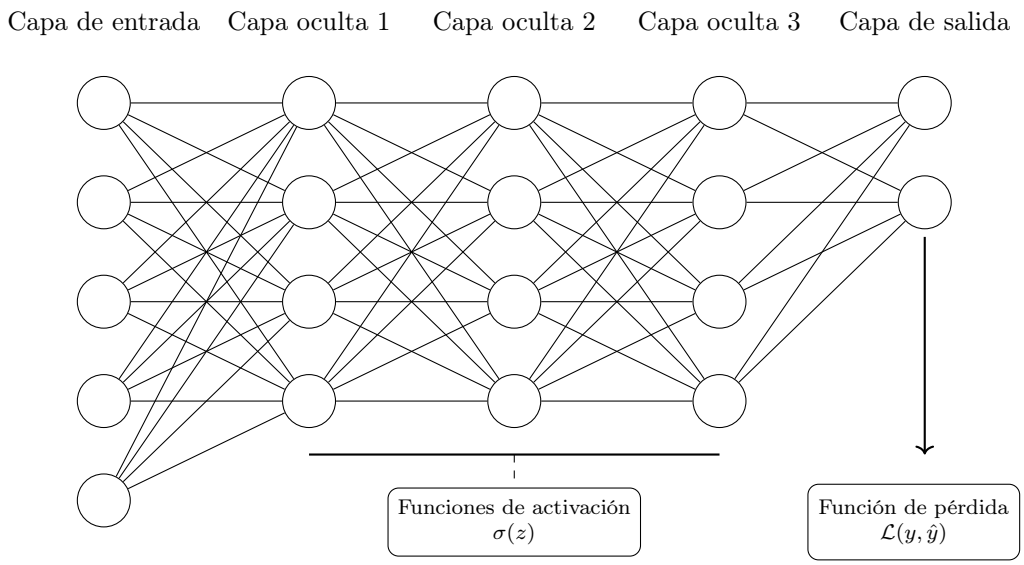
maciones sucesivas, estas capas modifican la representación de los datos, permitiendo capturar patrones, dependencias y estructuras internas. La presencia de múltiples capas favorece la construcción de representaciones jerárquicas, en las cuales la información es refinada progresivamente hasta volverse más adecuada para la decisión final. La profundidad y la anchura de la red influyen directamente en su capacidad de representación y en la complejidad de las funciones que puede aproximar. (Apicella et al., 2024; Kam Ho, 2022).

La profundidad y el ancho de la red influyen directamente en su capacidad de representación. Una arquitectura suficientemente profunda permite modelar transformaciones complejas mediante etapas sucesivas, mientras que una anchura adecuada contribuye a la formación de regiones de decisión más flexibles en el espacio de características (Chaturvedi et al., 2023).

Finalmente, la capa de salida produce la respuesta final del modelo. Su estructura depende de la naturaleza de la tarea, pudiendo generar una etiqueta única, múltiples etiquetas, coordenadas espaciales o mapas estructurados. En todos los casos, esta capa sintetiza la información transformada por las capas previas y la convierte en una predicción interpretable. (Y. Liu et al., 2024).

Figura 5.2

Arquitectura general de una red neuronal profunda.



Nota. Elaboración propia.

En la Figura 5.2 se detalla la arquitectura general de una red neuronal profunda, en la cual la información fluye desde la entrada hacia la salida mediante transformaciones sucesivas. En cada capa oculta se aplican funciones de activación después de las transformaciones lineales, introduciendo no linealidad y permitiendo modelar relaciones complejas. En la capa de salida también se emplea una función de activación adecuada según la tarea. La función de pérdida se define sobre las salidas finales del modelo y mide la discrepancia entre las predicciones y los valores reales, generando la señal de error utilizada durante el entrenamiento.

FUNCIONES DE ACTIVACIÓN Y PÉRDIDA

Las funciones de activación introducen no linealidad en las redes neuronales profundas, permitiendo modelar relaciones complejas en tareas de clasificación, detección y segmentación. En las capas ocultas, estas funciones se aplican después de cada transformación lineal, mientras que en la capa de salida se seleccionan en función del tipo de predicción requerida. Su elección influye directamente en la estabilidad del entrenamiento y en la propagación del gradiente (Khalid et al., 2020; Too et al., 2020; Zhu y Chen, 2020).

Entre las activaciones más empleadas se encuentra la unidad lineal rectificada (ReLU), definida como:

$$\mathrm{ReLU}(x) = \mathrm{máx}(0, x),$$

la cual favorece la eficiencia computacional y mitiga el problema del gradiente desvanecido, aunque puede presentar unidades inactivas (Khalid et al., 2020; Too et al., 2020).

Se han propuesto variantes como Leaky ReLU y funciones híbridas diseñadas para mejorar la estabilidad y evitar neuronas muertas, especialmente en escenarios de segmentación médica y tareas complejas (Biswas et al., 2024; Kiliçarslan, 2023; Waqas et al., 2025).

En la capa de salida, para clasificación multiclase se utiliza comúnmente la función *softmax*, que transforma los puntajes en una distribución de probabilidad:

$$\mathrm{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}, \quad i = 1, \ldots, K.$$

En clasificación binaria o problemas multi-etiqueta se emplea frecuentemente la sigmoide, que

produce probabilidades independientes por clase (Tian et al., 2022).

En cuanto a las funciones de pérdida, estas se definen sobre la salida del modelo y cuantifican la discrepancia entre la predicción y la verdad de referencia. Para clasificación, la entropía cruzada categórica se expresa como:

$$\mathcal{L}_{CE}(\mathbf{y}, \hat{\mathbf{p}}) = -\sum_{i=1}^{K} y_i \log(\hat{p}_i),$$

mientras que la versión binaria se formula como:

$$\mathcal{L}_{BCE}(y, \hat{p}) = -\big[y \log(\hat{p}) + (1-y)\log(1-\hat{p})\big],$$

siendo ambas ampliamente utilizadas en tareas de reconocimiento visual (Ito et al., 2023; Tian et al., 2022).

En segmentación, donde existe frecuente desbalance entre clases, la pérdida Dice se basa en la superposición entre la predicción y la región verdadera:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_i p_i y_i}{\sum_i p_i + \sum_i y_i},$$

y ha demostrado mejorar la sensibilidad en regiones pequeñas (Yeung et al., 2022; Y. Zhang et al., 2021). Asimismo, se han desarrollado pérdidas unificadas que combinan entropía cruzada y Dice para manejar desbalance severo (Yeung et al., 2022).

En detección de objetos, además de pérdidas de clasificación, se emplean pérdidas basadas en la intersección sobre unión (IoU) y funciones de regresión para ajustar las coordenadas de las cajas delimitadoras, optimizando tanto la localización como la categorización del objeto (Fu et al., 2021; Tian et al., 2022).

5.4.2 REDES NEURONALES CONVOLUCIONALES

Las redes neuronales convolucionales son una clase de redes neuronales profundas diseñadas específicamente para procesar datos con una estructura de cuadrícula, como las imágenes. Las CNN se destacan por su capacidad para capturar tanto características locales como globales de los datos de entrada mediante el uso de capas convolucionales, de agrupamiento (*pooling*) y completamente conectadas (Yi, 2022).

Las capas convolucionales permiten detectar patrones específicos dentro de la imagen de entrada aplicando diversos filtros. Las capas de agrupamiento reducen las dimensiones espaciales de los mapas de características, lo que ayuda a disminuir la carga computacional y a controlar el sobreajuste y las capas completamente conectadas se utilizan al final de la red para realizar la clasificación final basada en las características extraídas por las capas convolucionales y de agrupamiento (Raj y Yos, 2025).

Una de las principales características de las CNN es su conectividad local, donde cada neurona en las capas convolucionales está conectada solo a una pequeña región de la capa anterior, permitiendo detectar características locales como bordes o texturas. También destacan por su invariancia espacial, ya que las operaciones de agrupamiento reducen la resolución espacial de las características, manteniendo las más importantes y proporcionando invariancia a traslaciones y deformaciones (J. Wang et al., 2022), lo que permite reconocer objetos, aunque estén desplazados.

Las CNN siguen una jerarquía de características, donde las primeras capas capturan características simples, mientras que las capas profundas extraen patrones más complejos, mejorando así el reconocimiento y clasificación de objetos en las imágenes (J. Wang et al., 2022).

El proceso de entrenamiento de las CNN consiste en recopilar y etiquetar imágenes para ajustar los pesos de los filtros y minimizar el error de clasificación. Luego, se usan conjuntos de datos separados para validar y probar la precisión del modelo, garantizando que la red generalice bien a nuevas imágenes. Finalmente, una vez entrenada, la CNN puede clasificar imágenes en tiempo real, facilitando su monitoreo y estudio.

CAPA CONVOLUCIONAL

La operación de convolución constituye el mecanismo central para la extracción automática de características en redes neuronales profundas aplicadas al análisis de imágenes. Esta operación consiste en aplicar un filtro o *kernel* de tamaño reducido que se desplaza sobre la entrada, realizando productos elemento a elemento y sumas locales para generar una nueva representación (Farag, 2023; S. Kaur y Agrawal, 2021). La convolución bidimensional puede

expresarse como:

$$Y(i,j) = \sum_m \sum_n X(i+m, j+n)\, K(m,n),$$

donde X representa la entrada, K el kernel y Y el mapa de salida.

Los kernels son matrices de parámetros entrenables que permiten detectar patrones específicos en distintas regiones espaciales. Cada kernel produce un mapa de características (*feature map*), el cual resalta la presencia de una característica particular a lo largo de la imagen (Farag, 2023; Linse et al., 2023). La utilización de múltiples filtros en una misma capa permite capturar diferentes tipos de estructuras simultáneamente.

El concepto de campo receptivo local implica que cada neurona de una capa convolucional procesa únicamente una región limitada de la entrada, favoreciendo la detección de patrones espaciales y reduciendo el número de parámetros respecto a capas densas (S. Kaur y Agrawal, 2021).

A través de múltiples capas, la red construye representaciones jerárquicas: las capas iniciales capturan patrones simples, mientras que las capas profundas integran dichas representaciones en estructuras más complejas (Ma et al., 2022; Ye y Yu, 2021).

El *stride* determina el desplazamiento del kernel en cada paso, influyendo en la resolución espacial del mapa de salida. Un *stride* mayor reduce la dimensión espacial y el costo computacional, aunque puede disminuir el nivel de detalle. Por su parte, el *padding* consiste en añadir bordes artificiales a la entrada para controlar el tamaño de la salida y preservar información en los límites de la imagen (Farag, 2023; Zhou et al., 2024).

Dada una entrada de tamaño $H \times W$, un filtro de tamaño $K \times K$, *padding* P y *stride* S, la dimensión de salida viene dada por:

$$H_{out} = \left\lfloor \frac{H - K + 2P}{S} \right\rfloor + 1, \qquad W_{out} = \left\lfloor \frac{W - K + 2P}{S} \right\rfloor + 1.$$

Cuando $P = 0$ (*valid padding*), la salida reduce su tamaño espacial, lo que puede implicar pérdida de información en los bordes y afectar la extracción de características relevantes en

tareas de clasificación y detección (Shahnaz y Mollah, 2023; Zou et al., 2021). En contraste, el *same padding* introduce valores adicionales para mantener la resolución, facilitando arquitecturas profundas sin degradación progresiva de la representación espacial (Shahnaz y Mollah, 2023).

El *zero padding* es la estrategia más utilizada por su simplicidad computacional, aunque puede introducir discontinuidades artificiales en los bordes, debilitando la información estructural crítica para la detección precisa de contornos (Yu et al., 2023; Zou et al., 2021). Estudios recientes muestran que el impacto del padding en el rendimiento no es trivial, especialmente cuando se combina con operaciones de *pooling* o redimensionamiento (Shahnaz y Mollah, 2023; Yu et al., 2023).

Como alternativa, técnicas como *reflect* y *replication padding* buscan preservar de forma más natural la continuidad de los bordes, reduciendo artefactos inducidos por ceros constantes. En tareas sensibles a objetos pequeños o información posicional, la codificación explícita de información espacial y el tratamiento adecuado de bordes han demostrado mejorar la precisión del modelo (Feng et al., 2024).

Más recientemente, se han propuesto enfoques adaptativos y basados en aprendizaje. El *learning-based padding* modela dinámicamente los valores de borde mediante mecanismos convolucionales y de conectividad contextual, optimizando la coherencia estructural en clasificación y segmentación (Ning et al., 2023).

CAPA DE AGRUPACIÓN

El *pooling* es una operación de reducción espacial que resume regiones locales de un mapa de características para disminuir la dimensionalidad y aumentar la invariancia traslacional. Formalmente, para una región $\mathcal{R}$ de tamaño $k \times k$, el *max pooling* se define como $y = \max_{x \in \mathcal{R}} x$, mientras que el *average pooling* se expresa como $y = \frac{1}{|\mathcal{R}|} \sum_{x \in \mathcal{R}} x$ (Liawatimena et al., 2022; Nirthika et al., 2022).

El *max pooling* favorece la invariancia frente a pequeñas traslaciones, pero puede descartar información relevante, afectando la detección de objetos pequeños o detalles finos en

segmentación (Hao et al., 2021; Nirthika et al., 2022). Por su parte, el *average pooling* preserva información global, aunque tiende a suavizar características discriminativas (Nirthika et al., 2022). Estudios recientes evidencian que la elección de tamaño y stride influye directamente en métricas de desempeño como precisión y *Dice score* (Shadoul et al., 2025).

El *global pooling* (GAP y GMP) colapsa cada mapa de características en un único valor, reduciendo drásticamente parámetros y favoreciendo la regularización en clasificación, aunque sacrifica estructura espacial crítica para detección y segmentación (Nirthika et al., 2022). En contraste, métodos de *generalized* o *adaptive pooling* aprenden parámetros de agregación, equilibrando retención de información e invariancia, con mejoras reportadas en tareas de segmentación médica y clasificación especializada (Hao et al., 2021; Kozlova y Kitov, 2024).

Más recientemente, mecanismos de *attention-based pooling* integran pesos adaptativos que priorizan regiones relevantes, incrementando robustez bajo ruido y condiciones adversas, con impacto positivo en exactitud y sensibilidad (Chung et al., 2020; Zhong et al., 2023).

CAPA TOTALMENTE CONECTADA

La capa totalmente conectada (FC) transforma el vector de características $\mathbf{z}$ en una representación apta para decisión mediante $\mathbf{a} = f(\mathbf{W}\mathbf{z} + \mathbf{b})$, donde f suele ser una no linealidad (p. ej., ReLU) o una activación de salida (p. ej., softmax). En CNN, este bloque actúa como agregador global de evidencia aprendida en etapas convolucionales, especialmente en clasificación (Mahmoudi et al., 2020; Saleem et al., 2022).

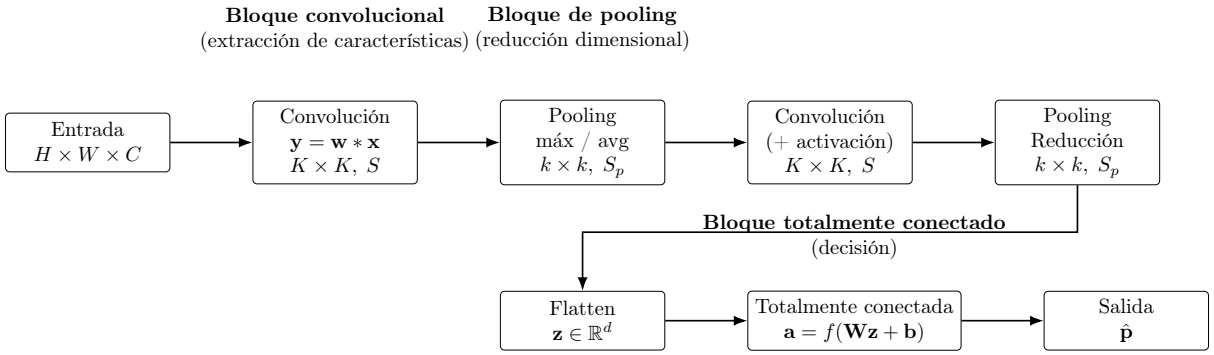
El principal costo de FC es su alta parametrización, lo que incrementa memoria, tiempo de entrenamiento y riesgo de sobreajuste; por ello se exploran reducciones de conexiones o variantes más compactas para mantener desempeño con menor complejidad (Ju et al., 2022; Saleem et al., 2022). En detección, aún se emplean capas densas en cabezales tipo R-CNN, pero se proponen formulaciones tensoriales para reducir parámetros sin degradar la predicción de clases y cajas (EL ICHI et al., 2025).

A nivel práctico, el diseño de FC se justifica cuando se requiere modelar interacciones

no lineales entre características de alto nivel, aunque su beneficio depende del tamaño de datos y regularización. En aplicaciones específicas, se han propuesto capas densas kernelizadas para mejorar la separabilidad en reconocimiento facial, sugiriendo que el bloque FC puede especializarse según el dominio (Mahmoudi et al., 2020).

Figura 5.3

Esquema general de una CNN



Nota. Elaboración propia. H = altura de la imagen, W = ancho de la imagen, C = número de canales, K = tamaño del *kernel*, S = *stride*, k = tamaño de la ventana, S_p = *stride* de la ventana, $*$ = convolución, d = dimensionalidad del vector tras *flatten*. $\mathbf{W}$ y $\mathbf{b}$ = pesos y sesgos; $f(\cdot)$ = función de activación y $\hat{\mathbf{p}}$ = probabilidades predichas.

En la Figura 5.3 se identifica la estructura general de una red neuronal profunda. En la primera fila, los bloques de convolución extraen características locales mediante filtros, mientras que los bloques de pooling reducen la dimensionalidad espacial conservando información relevante. Posteriormente, el resultado se aplana (*flatten*) y pasa a la capa totalmente conectada, donde se combinan las características aprendidas para generar la salida final en forma de probabilidades de clase.

MODELOS CNN

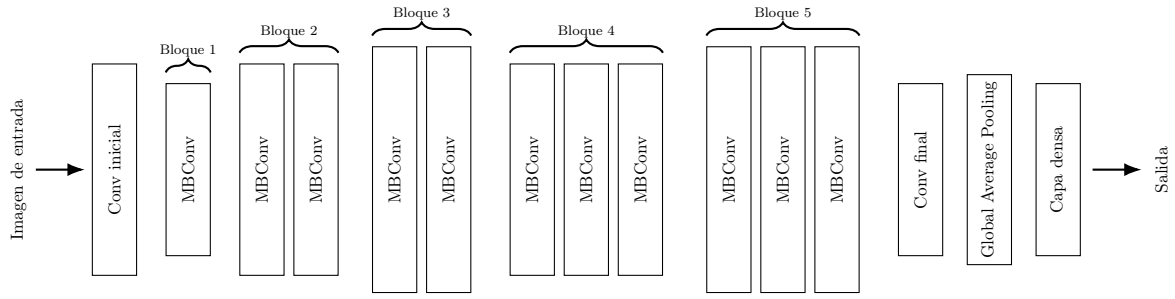
En el reconocimiento de imágenes mediante aprendizaje profundo no existe una única arquitectura CNN aplicable a todas las tareas. Por el contrario, la literatura reporta distintos modelos diseñados para mejorar la extracción de características, reducir el costo computacional, facilitar el entrenamiento de redes profundas o resolver tareas específicas como clasificación, detección y segmentación. En esta sección se describen los principales modelos identificados,

entre ellos EfficientNet, MobileNet, Inception, ResNet, DenseNet, YOLO, Faster R-CNN y Mask R-CNN.

- **EfficientNet:** emplea un escalamiento compuesto que equilibra profundidad, ancho y resolución, optimizando la relación precisión–FLOPs mediante bloques MBConv con *squeeze-and-excitation*, lo que favorece aplicaciones de clasificación con recursos limitados (Anumol, 2023; Mambile et al., 2025). Por ejemplo, EfficientNetB3 ha demostrado un rendimiento muy bueno, logrando una precisión general del 98 % en un conjunto de datos de 10,000 imágenes de diez clases de insectos, incluidas las abejas (A. Sharma et al., 2024).

Figura 5.4

Estructura general de EfficientNet.



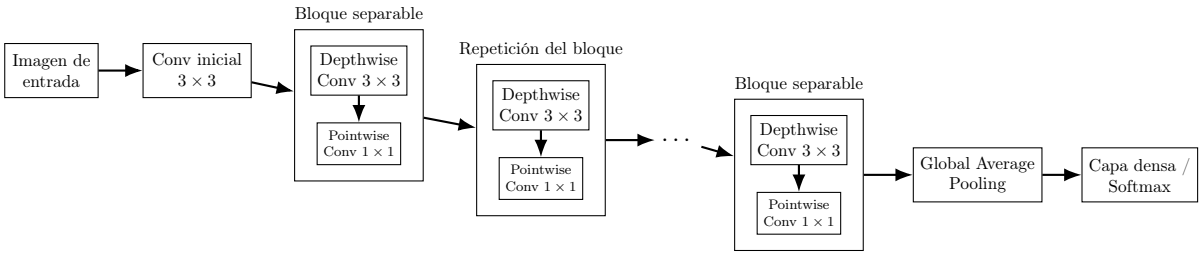
Nota. Elaboración propia.

El modelo de la Figura 5.4, inicia con una convolución inicial, seguida de una secuencia de bloques MBConv organizados por etapas, una convolución final, una operación de global average pooling y una capa densa para generar la salida. Las distintas variantes de EfficientNet se obtienen mediante un escalamiento compuesto que ajusta de forma equilibrada la profundidad, el ancho y la resolución de entrada.

- **MobileNet:** reduce el costo computacional mediante convoluciones separables en profundidad y bloques residuales invertidos. Por esta razón, se considera una arquitectura adecuada para aplicaciones en dispositivos embebidos o sistemas de monitoreo en campo, donde la eficiencia del modelo es un factor relevante (Mambile et al., 2025; Y. Srinivas y Kumar, 2025). MobileNetV3 ha mostrado un rendimiento superior en el reconocimiento de insectos invasores, alcanzando una precisión del 92,8 % y un tiempo de reconocimiento de solo 0,012 segundos por imagen (Huang et al., 2024).

Figura 5.5

Estructura general de MobileNet.



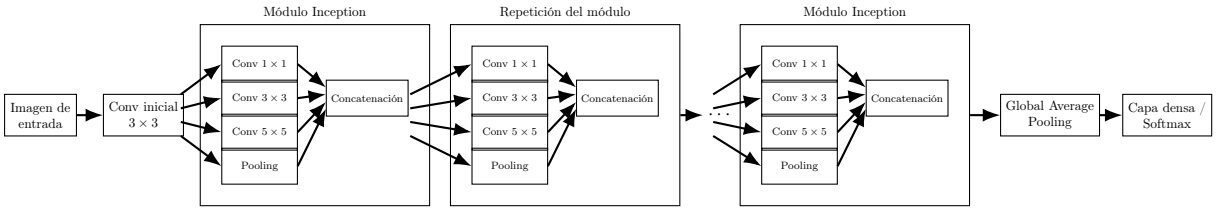
Nota. Elaboración propia.

El modelo de la Figura 5.5 consiste en sustituir convoluciones estándar por bloques de convolución separable en profundidad. Cada bloque divide el procesamiento en una convolución *depthwise*, que actúa sobre cada canal por separado, y una convolución *pointwise* de tamaño 1×1, encargada de combinar la información entre canales. Esta estrategia permite reducir significativamente el número de parámetros y el costo computacional del modelo.

- **Inception:** es una arquitectura CNN basada en módulos paralelos con distintos tamaños de filtro, permitiendo capturar patrones multi-escala relevantes en imágenes con variabilidad morfológica (Anumol, 2023; Kukkadapu, 2025). InceptionV3 ha logrado un equilibrio notable entre precisión y velocidad, con una precisión del 91,6 % y un tiempo de clasificación promedio de 3,34 ms al identificar 36 especies de abejorro (Spiesman et al., 2021).

Figura 5.6

Estructura general de Inception.



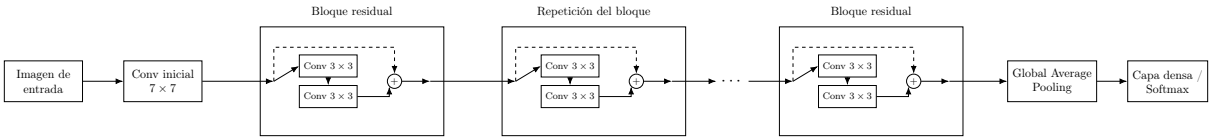
Nota. Elaboración propia.

En la Figura 5.6 se indica que Inception puede extraer características a múltiples escalas dentro de un mismo módulo y posteriormente integrarlas mediante concatenación, lo que mejora su capacidad de representación sin depender de una única resolución espacial.

- **ResNet:** incorpora conexiones residuales que estabilizan el entrenamiento de redes profundas, mejorando la extracción jerárquica de características en tareas de clasificación y segmentación (Anumol, 2023; Y. Srinivas y Kumar, 2025). Los modelos ResNet, incluyendo ResNet50 y ResNet101, se han utilizado eficazmente para diversas tareas, como la detección de ácaros Varroa en abejas y la clasificación de subespecies de abejas melíferas (A. Sharma et al., 2024).

Figura 5.7

Estructura general de ResNet.



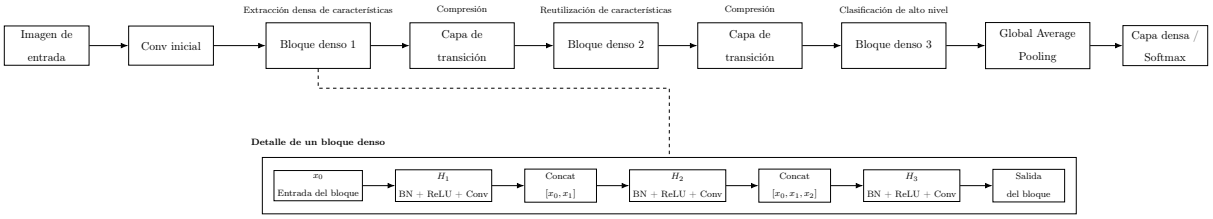
Nota. Elaboración propia.

En la Figura 5.7 se observa la característica distintiva de ResNet: el uso de bloques residuales con conexiones de atajo. Estas conexiones permiten sumar la entrada original con la salida de las convoluciones del bloque, facilitando el flujo de información y de gradientes a través de redes profundas y mitigando el problema de degradación del rendimiento.

- **DenseNet:** conecta cada capa con las capas anteriores dentro de bloques densos, favoreciendo reutilización de características y eficiencia paramétrica, lo que resulta ventajoso en escenarios con fondos complejos (Anumol, 2023; Zhukov e Ivanovych, 2023). Por ejemplo, DenseNet201 se ha utilizado para clasificar 19 clases de insectos con una precisión promedio del 87 % (Akter et al., 2023).

Figura 5.8

Estructura general de DenseNet.



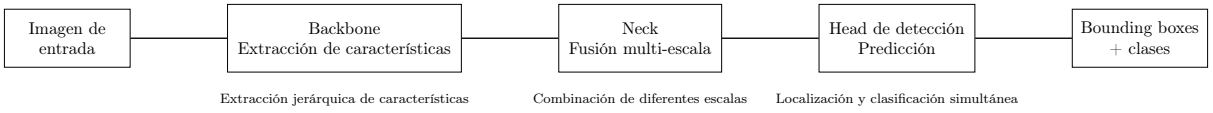
Nota. Elaboración propia.

En la Figura 5.8 se indica la arquitectura de DenseNet, donde la información fluye a través de múltiples bloques densos separados por capas de transición. A diferencia de otras redes, en cada bloque denso las capas no solo reciben la salida inmediata anterior, sino la concatenación de todas las salidas previas, lo que favorece la reutilización de características y mejora la propagación de la información.

- **YOLO:** es una familia de modelos de detección de objetos de una sola etapa. A diferencia de los enfoques que primero proponen regiones y luego clasifican objetos, YOLO realiza localización y clasificación de manera simultánea, lo que favorece aplicaciones en tiempo real. (Losiński y Czarnowski, 2025; Takimoto et al., 2021). Un estudio utilizó el modelo YOLOv5 para rastrear e informar el comportamiento de las abejas a partir de imágenes. El modelo se perfeccionó con un conjunto de datos de 9664 imágenes, logrando una precisión de reconocimiento eficaz y capacidades de detección en tiempo real (Alex et al., 2025).

Figura 5.9

Estructura general del modelo YOLO.



Nota. Elaboración propia.

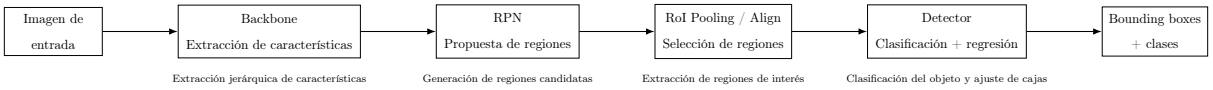
La Figura 5.9 presenta la arquitectura de YOLO, donde la imagen de entrada es procesada por un *backbone* para extraer características, un *neck* para fusionarlas a múltiples escalas y un *head* que realiza simultáneamente la localización y clasificación, generando como salida las

cajas delimitadoras y sus clases.

- **Faster R-CNN:** es un modelo de detección de dos etapas ampliamente reconocido por su equilibrio entre eficiencia y precisión. Su arquitectura se compone de dos módulos principales: una Red de Propuesta de Regiones (RPN), que es una red totalmente convolucional encargada de escanear la imagen para proponer áreas candidatas donde podrían encontrarse objetos, y un detector que procesa dichas regiones para clasificarlas y ajustar sus coordenadas mediante cuadros delimitadores o *bounding boxes* (Stiemer et al., 2023) tal como se indica en la Figura 5.10.

Figura 5.10

Estructura general del modelo Faster R-CNN.

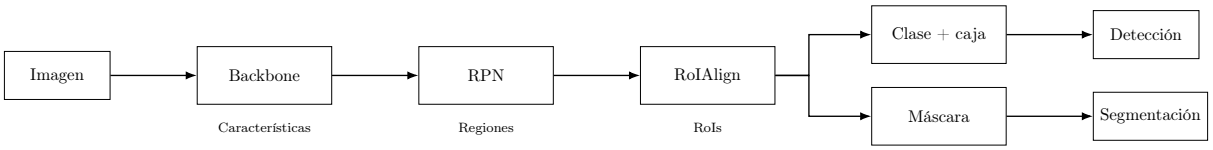


Nota. Elaboración propia.

- **Mask R-CNN:** es una extensión de la arquitectura Faster R-CNN que añade una rama paralela dedicada a predecir una máscara de segmentación para cada objeto detectado. Esta modificación permite que el sistema evolucione de una detección basada únicamente en cuadros delimitadores a una segmentación de instancias a nivel de píxel, logrando delinear la forma exacta del objeto de interés (Nizam et al., 2020).

Figura 5.11

Estructura general del modelo Mask R-CNN.



Nota. Elaboración propia.

En la Figura 5.11 se observa que Mask R-CNN extiende Faster R-CNN añadiendo una rama paralela de segmentación que, a partir de las regiones de interés (RoIs), predice máscaras binarias a nivel de píxel. Mientras la rama principal realiza clasificación y regresión de *bounding*

boxes, la nueva rama opera en paralelo usando RoIAlign para preservar la precisión espacial, permitiendo segmentación de instancias además de detección.

La comparación entre arquitecturas CNN permite identificar las diferencias principales en su diseño y aplicación. Mientras algunos modelos priorizan la eficiencia computacional, otros se enfocan en mejorar la precisión, la detección en tiempo real o la segmentación de objetos. La Tabla 5.1 resume estas características.

Tabla 5.1

Comparación de modelos de CNN.

Modelo	**Estrategia arquitectónica**	**Aplicación principal**
EfficientNet	Escalamiento compuesto equilibrado de profundidad, ancho y resolución	Clasificación eficiente con bajo costo computacional
MobileNet	Convoluciones separables en profundidad y bloques residuales invertidos	Clasificación en dispositivos embebidos
Inception	Módulos paralelos de convolución multi-escala	Clasificación con variabilidad morfológica
ResNet	Conexiones residuales para facilitar el entrenamiento de redes profundas	Clasificación y segmentación robusta
DenseNet	Conectividad densa y reutilización de características entre capas	Clasificación en entornos complejos
YOLO	Detección de una sola etapa orientada al procesamiento en tiempo real	Localización y conteo de polinizadores
Faster R-CNN	Detector de dos etapas con red de propuesta de regiones para clasificación y regresión	Detección precisa de objetos
Mask R-CNN	Extensión de Faster R-CNN con una rama adicional para segmentación de instancias a nivel de píxel	Detección y segmentación precisa de objetos

Nota. Elaboración propia.

5.5 MÉTRICAS DE DESEMPEÑO

Las tareas de clasificación, detección y segmentación de imágenes con redes neuronales convolucionales se evalúan mediante métricas específicas que cuantifican el desempeño desde distintas perspectivas. Estas medidas permiten comparar modelos de reconocimiento de imágenes y seleccionar configuraciones con mejor equilibrio entre precisión, robustez y eficiencia computacional.

5.5.1 MATRIZ DE CONFUSIÓN

La matriz de confusión resume el desempeño de un clasificador comparando predicciones vs. etiquetas reales. En clasificación binaria, organiza los conteos de aciertos y errores en cuatro casos: verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Esta representación permite identificar si el modelo tiende a confundir sistemáticamente una clase con otra, y constituye la base para calcular métricas como exactitud (*accuracy*), precisión (*precision*), sensibilidad (*recall*) y medida F1 (*F1-score*).

5.5.2 MÉTRICAS PARA CLASIFICACIÓN

En clasificación, se emplean con frecuencia la exactitud (*accuracy*), la precisión (*precision*), la sensibilidad (*recall*) y la medida F1 (*F1-score*) (Gupta et al., 2025). En la Tabla 5.2 se definen cada una de ellas.

Tabla 5.2

Métricas de clasificación

Métrica	**Fórmula**
Exactitud	$\frac{TP+TN}{TP+TN+FP+FN}$
Precisión	$\frac{TP}{TP+FP}$
Sensibilidad	$\frac{TP}{TP+FN}$
F1	$2 \cdot \frac{\text{Precisión} \cdot \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$

Nota. Elaboración propia.

Adicionalmente, el área bajo la curva ROC (*Area Under the Curve*, AUC) evalúa la capacidad del modelo para discriminar entre clases, siendo especialmente informativa en clasificación binaria o con desbalance de clases (Parvin y Mehedi Hasan, 2020). Si TPR es la tasa de verdaderos positivos y FPR la tasa de falsos positivos, entonces:

$$TPR = \frac{TP}{TP+FN}, \qquad FPR = \frac{FP}{FP+TN}, \qquad AUC = \int_0^1 TPR(FPR)\, d(FPR).$$

La matriz de confusión complementa este análisis al desglosar aciertos y errores por clase, facilitando la identificación de patrones sistemáticos de confusión.

5.5.3 MÉTRICAS PARA DETECCIÓN

En detección, la métrica base es la intersección sobre unión (*Intersection over Union*, IoU), que mide la superposición entre la caja predicha B_p y la caja real B_{gt}:

$$IoU = \frac{|B_p \cap B_{gt}|}{|B_p \cup B_{gt}|}$$

Una métrica agregada ampliamente usada en detección de objetos es la precisión media (*mean Average Precision*, mAP), definida como el promedio de la *Average Precision* (AP) calculada para todas las clases evaluadas. Para una clase c, la AP_c corresponde al área bajo la curva precisión–sensibilidad, donde $\text{Precision}_c(\text{Recall})$ indica cómo varía la precisión del modelo para esa clase en función de distintos niveles de sensibilidad:

$$AP_c = \int_0^1 \text{Precision}_c(\text{Recall})\, d(\text{Recall}), \qquad mAP = \frac{1}{C}\sum_{c=1}^{C} AP_c.$$

Donde c es una clase específica, C es el número total de clases, AP_c es la precisión promedio de la clase c y mAP es el promedio de las precisiones promedio obtenidas en todas las clases.

En aplicaciones de tiempo real, también se reporta *Frames Per Second* (FPS) como indicador de velocidad de inferencia y viabilidad operativa del sistema (Saetchnikov et al., 2021).

5.5.4 MÉTRICAS PARA SEGMENTACIÓN

En segmentación, se emplea IoU a nivel de máscara (por clase o por instancia), donde A representa el conjunto de píxeles predichos y B el conjunto de píxeles reales:

$$IoU = \frac{|A \cap B|}{|A \cup B|}.$$

Asimismo, se utilizan precisión y F1 a nivel de píxel: la precisión refleja la proporción de

píxeles positivos predichos que son correctos, mientras que F1 resume el equilibrio entre precisión y sensibilidad en la asignación pixel a pixel.

Un indicador estrechamente relacionado es el coeficiente *Dice*, que enfatiza el solapamiento relativo y resulta particularmente útil cuando las regiones de interés son pequeñas, ya que penaliza con mayor fuerza las discrepancias en áreas reducidas (Varadarajan et al., 2021). Dicho coeficiente está definido como:

$$Dice = \frac{2|A \cap B|}{|A| + |B|}.$$

Las métricas de desempeño utilizadas en visión por computador dependen del tipo de tarea que se evalúa. La Tabla 5.3 resume las métricas más utilizadas.

Tabla 5.3

Métricas por tarea

Tarea	**Métricas**
Clasificación	Exactitud, Precisión, Sensibilidad, Medida F1, AUC (ROC)
Detección de objetos	IoU, AP/mAP, Precisión, Sensibilidad, FPS
Segmentación	Pixel Exactitud, IoU/mIoU, Dice, Precisión, Sensibilidad, Medida F1

Nota. Elaboración propia.

5.6 FACTORES QUE INFLUYEN EN EL RENDIMIENTO

El rendimiento de los modelos de aprendizaje profundo en la identificación de polinizadores depende de la arquitectura seleccionada y de la implementación de estrategias técnicas que optimizan su capacidad de generalización. Debido a la complejidad de los entornos naturales y la limitada disponibilidad de bases de datos biológicas etiquetadas, el uso de técnicas como la transferencia de aprendizaje, el aumento de datos y algoritmos de optimización, resulta determinante.

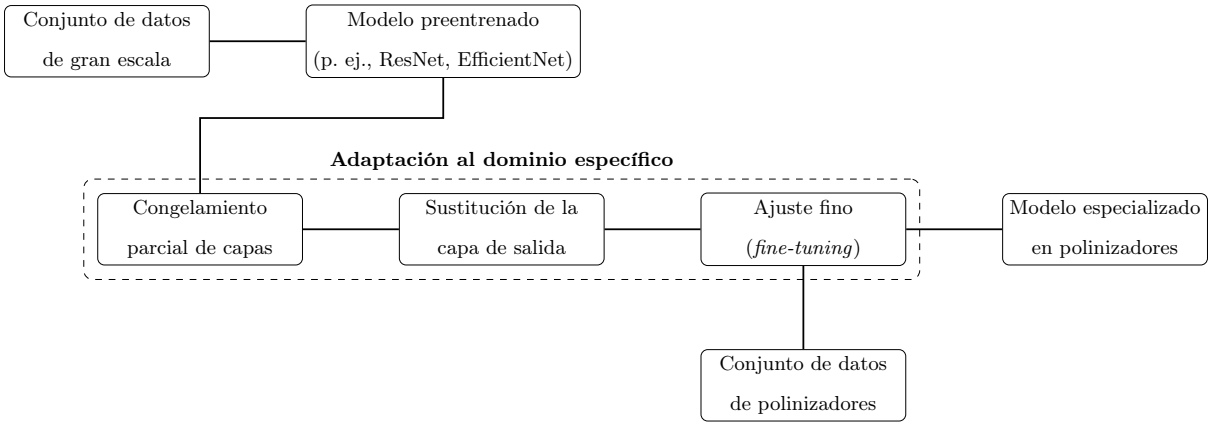
5.6.1 APRENDIZAJE POR TRANSFERENCIA

El aprendizaje por transferencia consiste en reutilizar pesos preentrenados en grandes conjuntos de datos y adaptarlos a dominios específicos mediante *fine-tuning* parcial o total. Este enfoque reduce el tiempo de entrenamiento y mejora la generalización cuando los datos son limitados (Burri et al., 2023; Jagtap y Yadav, 2025; Zheng et al., 2021). En la Figura 5.12 se presenta su esquema general.

Por ejemplo, en el reconocimiento de polinizadores, permite entrenar modelos con conjuntos de datos más pequeños, aprovechando el conocimiento de modelos pre entrenados para mejorar la precisión y generalización en la identificación de especies, incluso en condiciones variables (Minervini et al., 2021).

Figura 5.12

Esquema general del aprendizaje por transferencia.



Nota. Elaboración propia.

En clasificación de insectos y especies relacionadas, es común congelar las capas iniciales y ajustar únicamente las finales, o realizar ajuste completo si existe alta divergencia de dominio. Arquitecturas como ResNet, EfficientNet y DenseNet son utilizadas como *backbones* por su estabilidad y eficiencia (Jian et al., 2023; Sajib et al., 2025; Yang et al., 2020).

En detección, el aprendizaje por transferencia se implementa reutilizando modelos preen-

trenados como Faster R-CNN, lo que mejora la localización de organismos en entornos naturales con menor cantidad de datos etiquetados (Tseng et al., 2022). La reutilización del extractor de características reduce el costo computacional y acelera la convergencia.

En segmentación, el ajuste fino permite adaptar representaciones aprendidas a patrones morfológicos específicos de polinizadores. Esto mitiga el impacto del cambio de dominio entre imágenes genéricas y escenarios ecológicos reales (Hassanie et al., 2025; Zheng et al., 2021).

5.6.2 AUMENTO DE DATOS

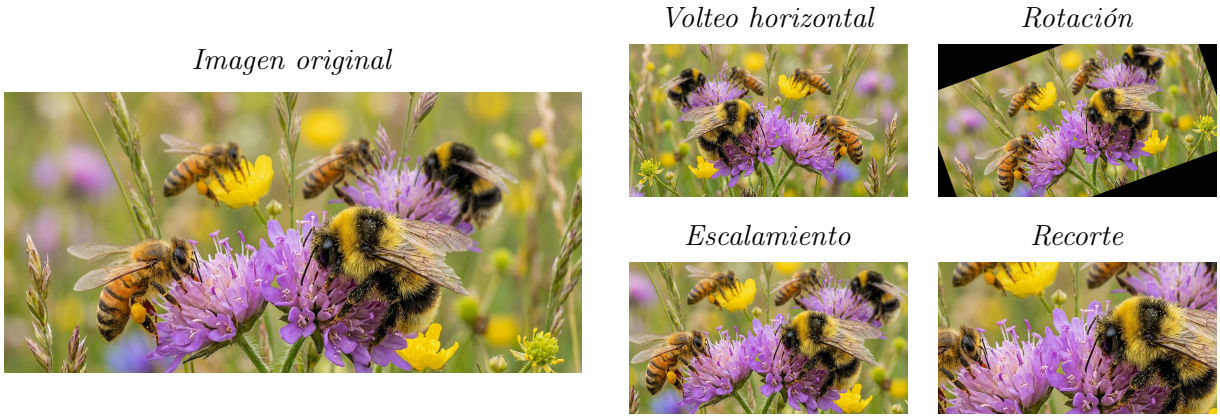
El aumento de datos genera nuevas muestras a partir de transformaciones aplicadas sobre los datos originales, como rotaciones, traslaciones, recortes, cambios de escala, ajustes de brillo, contraste, ruido o desenfoque (Z. Liu et al., 2024). En visión por computador, esta técnica se emplea para ampliar la variabilidad del conjunto de entrenamiento, mejorar la robustez del modelo frente a cambios visuales y disminuir el sobreajuste. En el reconocimiento de polinizadores, su aplicación es relevante porque las imágenes suelen presentar variaciones de iluminación, postura, escala, fondo y movimiento, condiciones que pueden afectar la precisión y confiabilidad del modelo.

TRANSFORMACIONES GEOMÉTRICAS

Las transformaciones geométricas incluyen volteo (*flip*), rotación (*rotate*), recorte (*crop*) y escalado (*scale*). Estas operaciones inducen invariancias útiles cuando el insecto aparece con distintas orientaciones o posiciones, favoreciendo la generalización en clasificación y detección (M. Kaur et al., 2022; Madaan y Sharma, 2024; Wu et al., 2024). En la Figura 5.13 se presenta un ejemplo de este tipo de transformaciones.

Figura 5.13

Ejemplo de transformaciones geométricas.



Nota. Elaboración propia.

TRANSFORMACIONES FOTOMÉTRICAS

Las transformaciones fotométricas modifican brillo (*brightness*), contraste (*contrast*) y variaciones de color (*color jitter*). Su objetivo es disminuir la sensibilidad del modelo a cambios de iluminación en entornos naturales (Agarwal et al., 2025; M. Kaur et al., 2022; Qian et al., 2023). En la Figura 5.14 se indica un ejemplo con las distintas transformaciones.

Figura 5.14

Ejemplo de transformaciones fotométricas.



Nota. Elaboración propia.

RUIDO Y DESENFOQUE

La adición de ruido (*noise*) y desenfoque (*blur*) aproxima imperfecciones reales del sensor o movimiento, lo cual incrementa la tolerancia del modelo a imágenes degradadas (Agarwal et al., 2025; Andriyanov y Andriyanov, 2020; Madaan y Sharma, 2024). En la Figura 5.15 se indica un ejemplo con este tipo de imperfecciones.

Figura 5.15

Ejemplo de ruido y desenfoque.

Imagen original



Ruido gaussiano



Desenfoque gaussiano



Desenfoque por movimiento



Nota. Elaboración propia.

MÉTODOS AVANZADOS Y ESTRATEGIAS AUTOMÁTICAS

Métodos como MixUp, CutMix y Mosaic combinan muestras para aumentar la diversidad del conjunto de entrenamiento y reducir el sobreajuste, tal como se indica en la Figura 5.16. Además, las estrategias automáticas de aumento de datos permiten seleccionar o aprender automáticamente combinaciones de transformaciones, como recorte, rotación, cambio de escala, brillo o contraste, de acuerdo con la tarea y las características del conjunto de datos. A diferencia de una política manual fija, estas estrategias buscan optimizar qué transformaciones aplicar y con qué intensidad para mejorar la generalización del modelo, especialmente en tareas de detec-

ción (Agarwal et al., 2025; Qian et al., 2023).

Figura 5.16

Ejemplo de técnicas avanzadas de aumento de datos.



Nota. Elaboración propia.

SEGMENTACIÓN Y AUMENTO BASADO EN OBJETOS

En segmentación, el aumento basado en objetos inserta o compone instancias para reforzar el aprendizaje de bordes y regiones, siendo útil cuando la clase objetivo ocupa un área pequeña o existe alta variabilidad de fondo (S. Chen et al., 2024; Illarionova et al., 2021). En la Figura 5.17 se indica un ejemplo con estas variaciones.

Figura 5.17

Ejemplo de segmentación y aumento basado en objetos.

Imagen original *Aumento basado en objetos*



Nota. Elaboración propia.

5.6.3 MÉTODOS DE OPTIMIZACIÓN

Los métodos de optimización, como el descenso de gradiente y sus variantes ajustan los parámetros de las redes neuronales para minimizar la función de pérdida (Rudenko et al., 2020). En el reconocimiento de polinizadores, estos métodos son esenciales para un entrenamiento eficiente y para evitar el sobreajuste mediante técnicas como regularización y abandono, garantizando un buen rendimiento con datos nuevos.

En esta sección se describen los principales métodos y estrategias asociados al entrenamiento de redes neuronales profundas. Además de los algoritmos de optimización propiamente dichos, como SGD, Adam, RMSProp y AdamW, se consideran estrategias complementarias que influyen en la estabilidad y generalización del modelo, tales como el muestreo estratificado, las políticas de tasa de aprendizaje, la regularización y la estabilidad.

- **Descenso de Gradiente Estocástico:** también llamado SGD, es un referente por su comportamiento estable, aunque puede converger más lentamente; su variante con *momentum* incorpora una acumulación de gradientes que acelera el avance en direcciones consistentes (Postalcloğlu, 2020; Sudewo et al., 2025).

- **Optimizadores adaptativos:** ajustan tasas de aprendizaje por parámetro y suelen mejorar la rapidez de convergencia; sin embargo, su desempeño final depende del dominio y de la configuración de hiperparámetros (Azis et al., 2025; Postalcloğlu, 2020). RMSProp adapta la

tasa de aprendizaje mediante un promedio móvil de los gradientes cuadrados, lo que ayuda a estabilizar el entrenamiento cuando existen variaciones fuertes en los gradientes. Adam combina esta adaptación con el uso de momentos del gradiente, mientras que AdamW desacopla el *weight decay* de la actualización del gradiente, lo que puede favorecer la generalización frente a Adam en prácticas modernas de entrenamiento (Azis et al., 2025; Sudewo et al., 2025).

- **Optimizadores recientes y enfoques híbridos:** métodos como DiffGrad y AdaNorm introducen correcciones dinámicas basadas en el comportamiento o norma del gradiente para estabilizar el entrenamiento y mejorar el ajuste en CNN (Dubey et al., 2020, 2023). En paralelo, enfoques híbridos exploran combinar optimizadores adaptativos y tradicionales para conservar convergencia rápida sin degradar la generalización (Pattanaik et al., 2025).

- **Muestreo estratificado:** es una técnica que divide la población en subgrupos para mantener proporciones representativas y evitar sesgos por desbalance. En estudios de polinizadores, permite incluir adecuadamente clases raras como parásitos o anomalías, mejorando el aprendizaje del modelo y su capacidad de generalizar en distintos entornos (Lee et al., 2025)

- **Políticas de tasa de aprendizaje:** regulan η_t durante el entrenamiento. Esquemas como *step decay*, *cosine annealing* y *warmup* pueden mejorar la convergencia y evitar estancamiento en tareas de clasificación visual (Xue et al., 2022).

- **Regularización y estabilidad:** el *weight decay* controla el crecimiento de pesos y contribuye a la generalización, mientras que *early stopping* detiene el entrenamiento cuando la validación deja de mejorar, reduciendo sobreajuste y costo computacional (Azis et al., 2025; Loganathan y Latha, 2025). El *gradient clipping* limita $\|\nabla_\theta \mathcal{L}\|$ para evitar explosiones de gradiente en entrenamientos inestables (Xue et al., 2022).

6. METODOLOGÍA

6.1 TIPO DE ESTUDIO

La presente investigación se desarrolla bajo un enfoque cualitativo-descriptivo, mediante una revisión sistemática de la literatura orientada a caracterizar el estado del arte del reconocimiento de imágenes de polinizadores mediante aprendizaje profundo. Para ello, se adoptan las directrices PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), debido a que proporciona una estructura explícita, reproducible y trazable para la identificación, cribado e inclusión de estudios científicos.

6.2 PROTOCOLO DE REVISIÓN

Antes de iniciar la revisión sistemática se definió un protocolo de trabajo para orientar el proceso de búsqueda, selección y análisis de los estudios. Este protocolo tuvo como finalidad asegurar que la revisión siguiera un procedimiento ordenado, transparente y reproducible, de acuerdo con las directrices PRISMA.

El protocolo incluyó la delimitación del tema de estudio, la selección de Scopus como fuente de información, el diseño de una estrategia de búsqueda mediante operadores booleanos, la definición de criterios de inclusión y exclusión, y la estructuración del proceso de cribado. Este proceso contempló la revisión de duplicados, el análisis por título, resumen y palabras clave, y la evaluación de texto completo. Además, se definieron previamente las variables que serían extraídas de cada artículo, entre ellas las arquitecturas utilizadas, el tipo de tarea, las métricas de desempeño y los factores técnicos asociados al rendimiento de los modelos.

6.3 PREGUNTAS DE INVESTIGACIÓN

Con el fin de estructurar el análisis de la literatura revisada, se plantearon preguntas de investigación que orientan la identificación, comparación e interpretación de los estudios incluidos. Estas preguntas permiten vincular directamente los objetivos específicos con el proceso

de revisión sistemática, asegurando que la extracción de datos y la discusión de resultados se centren en aspectos relevantes para caracterizar el estado del arte.

- **RQ1:** ¿Cuáles son las arquitecturas y tipos de modelos de aprendizaje profundo más utilizados para el reconocimiento de imágenes de polinizadores en artículos publicados desde el año 2020?
- **RQ2:** ¿Qué métricas de evaluación se reportan con mayor frecuencia en los estudios sobre reconocimiento de polinizadores mediante aprendizaje profundo?
- **RQ3:** ¿Qué factores técnicos se asocian con mayores niveles de precisión en los modelos de reconocimiento de polinizadores?

6.4 FUENTE DE INFORMACIÓN Y ESTRATEGIA DE BÚSQUEDA

La identificación de estudios se realiza en la base de datos Scopus, seleccionada por su cobertura multidisciplinaria y su relevancia en áreas como visión por computador, aprendizaje automático y ciencias ambientales.

Para construir la estrategia de búsqueda se definieron palabras clave agrupadas según las principales dimensiones de la revisión: población de interés, métodos de aprendizaje profundo, tareas de visión por computador y métricas o factores de desempeño. Esta organización permitió combinar términos relacionados con polinizadores e insectos, modelos de reconocimiento visual y criterios de evaluación del rendimiento.

Tabla 6.1

Palabras clave utilizadas para la estrategia de búsqueda.

Dimensión	**Palabras clave y sinónimos**	**Relación con la revisión**
Polinizadores e insectos	*bee*, *honey bee*, *wild bee*, *bumblebee*, *pollinator*, *hymenoptera*, *insect*, *pollinating insects*	Población o grupo biológico de interés

Reconocimiento de imágenes	*image recognition*, *image classification*, *object detection*, *computer vision*, *automatic recognition*, *image processing*	Tareas de visión por computador aplicadas al reconocimiento visual
Modelos de aprendizaje profundo	*deep learning*, *neural network*, CNN, *vision transformer*, EfficientNet, ResNet, YOLO, *machine learning*	Métodos y arquitecturas utilizadas para el reconocimiento de imágenes
Métricas y desempeño	*metrics*, AUC, *accuracy*, *precision*, *recall*, F1-score, mAP, IoU, FPS, *performance*, *computational cost*	Evaluación y comparación del rendimiento de los modelos
Conjuntos de datos	*bee dataset*, *pollinator dataset*, *insect image dataset*	Fuentes de imágenes empleadas para entrenamiento o evaluación

Nota. Elaboración propia.

A partir de estas dimensiones, se construyó una cadena de búsqueda en Scopus mediante operadores booleanos. La búsqueda se aplicó en los campos de título, resumen y palabras clave, con el propósito de recuperar estudios relacionados con reconocimiento de imágenes de polinizadores mediante aprendizaje profundo:

```
TITLE-ABS-KEY ( ( "deep learning" OR "neural network*" OR cnn OR
"vision transformer" OR "efficientnet" OR "resnet" OR yolo OR
"object detection" OR "image classification" OR "machine learning"
OR metrics ) AND ( "bee image" OR "honey bee image" OR
"wild bee image" OR "bumblebee image" OR "pollinator image" OR
"hymenoptera image" OR "insect image dataset" OR "bee dataset" OR
"pollinator dataset" OR "insect image classification" OR
"bee recognition" OR "insect recognition" OR
( "pollinator" AND "image" ) OR ( "bee" AND "image" ) ) )
AND PUBYEAR > 2019
AND ( LANGUAGE ( english ) OR LANGUAGE ( spanish ) )
AND ( LIMIT-TO ( OA , "all" ) )
```

La estrategia incorpora términos asociados con aprendizaje profundo, redes neuronales, arquitecturas relevantes y tareas de visión por computador, combinados con términos relacionados con imágenes de abejas, polinizadores e insectos. Además, se restringen los resultados a publicaciones en inglés o español, de acceso abierto y publicadas a partir del año 2020.

6.5 CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN

Los criterios de inclusión y exclusión se definieron con el propósito de seleccionar estudios pertinentes, recientes y comparables con el objetivo de la revisión sistemática. Estos criterios permitieron delimitar el corpus a investigaciones que emplearan métodos de aprendizaje profundo para el reconocimiento de imágenes de polinizadores, que reportaran métricas cuantitativas de desempeño y que estuvieran disponibles para revisión completa.

La inclusión de estudios publicados desde el año 2020 se justificó por el interés de caracterizar el estado reciente del campo, considerando que las arquitecturas de visión por computador y las técnicas de entrenamiento han evolucionado rápidamente en los últimos años. Asimismo, se restringió la búsqueda a publicaciones en inglés o español debido a la disponibilidad de lectura y análisis dentro del alcance de la investigación. El criterio de acceso abierto se estableció para garantizar la posibilidad de revisar el texto completo y extraer información metodológica de manera verificable.

También se consideró indispensable que los estudios utilizaran imágenes como fuente principal de datos, ya que el enfoque de esta monografía se centra en tareas de visión por computador, como clasificación, detección y segmentación. De igual forma, se exigió que los trabajos aplicaran aprendizaje profundo y reportaran métricas de desempeño, pues estos elementos son necesarios para comparar arquitecturas, evaluar resultados y analizar factores técnicos asociados al rendimiento de los modelos. Los criterios establecidos se presentan en la Tabla 6.2.

Tabla 6.2

Criterios de inclusión y exclusión

Criterios de inclusión	**Criterios de exclusión**
Estudios que emplean métodos de aprendizaje profundo.	Estudios que no emplean métodos de aprendizaje profundo.
Estudios escritos en inglés o español.	Estudios escritos en idiomas distintos al inglés o español.
Estudios que reportan métricas cuantitativas de desempeño.	Estudios que no reportan métricas de evaluación comparables.
Estudios con acceso abierto al texto completo.	Estudios sin acceso abierto o sin disponibilidad de texto completo.
Estudios enfocados en insectos polinizadores o grupos directamente relacionados.	Estudios sin enfoque en insectos polinizadores o no relacionados con el monitoreo de polinizadores.
Estudios que utilizan imágenes para reconocimiento, clasificación, detección o segmentación.	Estudios que no utilizan imágenes como fuente principal de datos.
Estudios publicados desde el año 2020.	Estudios publicados antes del año 2020.

Nota. Elaboración propia.

6.6 PROCESO DE SELECCIÓN DE ESTUDIOS

El proceso de selección de estudios se desarrolló de acuerdo con la lógica de las directrices PRISMA, considerando las fases de identificación, eliminación de duplicados, cribado inicial, evaluación de elegibilidad por texto completo e inclusión final. Esta organización permitió documentar de manera transparente la reducción progresiva de registros hasta conformar el corpus definitivo de análisis.

En la fase de identificación se recuperaron 281 registros a partir de la aplicación de la estrategia de búsqueda en Scopus y de los criterios establecidos previamente en la Tabla 6.2. Estos registros constituyeron el conjunto inicial de artículos potencialmente relevantes para la revisión sistemática.

Posteriormente, se realizó la revisión de duplicados. En esta etapa no se detectaron registros repetidos, por lo que los 281 artículos identificados se mantuvieron para el cribado por título, resumen y palabras clave. Este primer filtro permitió evaluar la correspondencia inicial de cada estudio con el tema de investigación, considerando especialmente el uso de imágenes de polinizadores, la aplicación de métodos de aprendizaje profundo, el reporte de métricas de desempeño y la presencia de factores técnicos vinculados con el rendimiento de los modelos.

Como resultado del cribado inicial, se aceptaron 54 artículos y se excluyeron 227 registros que no cumplían suficientemente con el alcance temático o metodológico de la revisión. Los artículos preseleccionados fueron evaluados posteriormente mediante lectura de texto completo, con el fin de determinar su elegibilidad final.

Para la evaluación de texto completo se utilizó una lista de verificación compuesta por cuatro preguntas de elegibilidad, presentadas en la Tabla 6.3. Cada pregunta fue calificada según el grado de cumplimiento del criterio: $Yes = 1$, $Partial = 0.5$ y $No = 0$. La suma de las cuatro puntuaciones permitió obtener un puntaje total de elegibilidad para cada artículo.

Se estableció un umbral mínimo de 3 puntos para incluir un estudio en la revisión final. Los artículos con puntaje inferior a dicho umbral fueron excluidos por no presentar suficiente pertinencia metodológica o analítica para los objetivos de la investigación. En esta etapa se excluyó 1 artículo, por lo que el corpus final quedó conformado por 53 estudios.

Tabla 6.3

Preguntas de elegibilidad

Código	Pregunta de elegibilidad
Q1	¿El estudio utiliza imágenes de polinizadores como base para reconocimiento, clasificación, detección o segmentación?
Q2	¿El estudio aplica métodos de aprendizaje profundo para el reconocimiento de imágenes de polinizadores?
Q3	¿El estudio reporta métricas cuantitativas de desempeño para evaluar el modelo propuesto, tales como *accuracy*, *precision*, *recall*, *F1-score*, *IoU*, *mAP*, *Dice*, *AUC* o *FPS*?
Q4	¿El estudio describe factores técnicos relevantes que influyen en el rendimiento del modelo, tales como transferencia de aprendizaje, aumento de datos o métodos de optimización?

Nota. Elaboración propia.

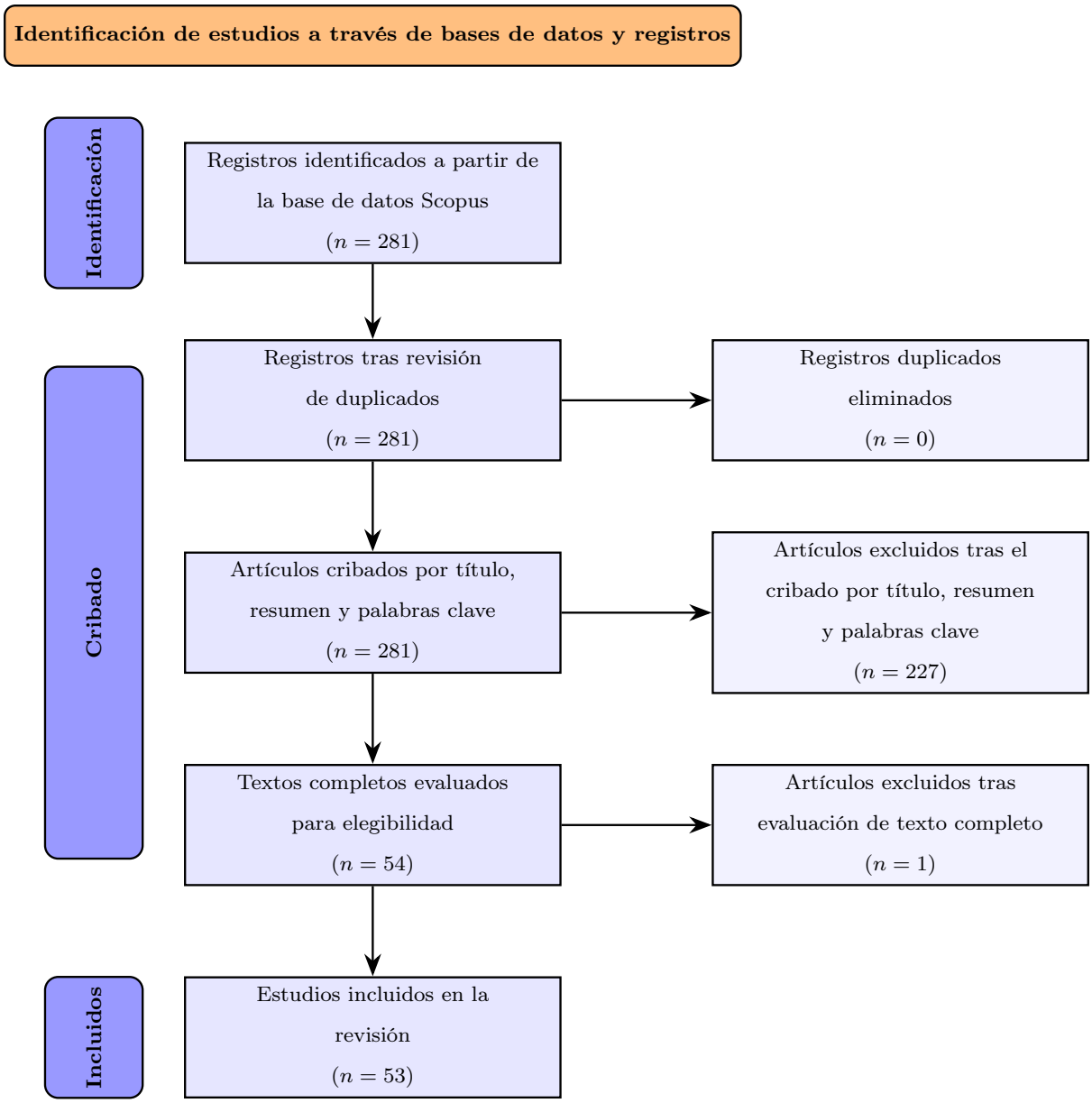
Finalmente, se incluyeron 53 artículos que cumplieron con los criterios de inclusión y alcanzaron el umbral de elegibilidad establecido. Estos estudios constituyen la base documental para la extracción de datos y el análisis descriptivo-comparativo de la revisión sistemática. El PDF `art0166` fue revisado en texto completo, pero se mantuvo fuera del corpus analítico por corresponder a un artículo de datos cuyo aporte principal era la descripción de un conjunto de imágenes, sin una evaluación de modelos suficientemente comparable para las tablas de síntesis.

6.7 RESUMEN DEL PROCESO DE SELECCIÓN

El flujo de selección de estudios se sintetiza en la Figura 6.1, donde se presentan las etapas de identificación, cribado, evaluación de elegibilidad e inclusión final del corpus analizado.

Figura 6.1

Proceso de selección de estudios basado en PRISMA.



Nota. Elaboración propia.

6.8 EXTRACCIÓN DE DATOS

Una vez definido el corpus final, se aplica un formulario de extracción de datos estructurado con el fin de registrar de manera homogénea la información relevante de cada estudio. La extracción se orienta a responder de forma directa las preguntas de investigación y a organizar posteriormente la presentación de resultados por categorías temáticas. Las variables extraídas se presentan a continuación:

- **Modelo o arquitectura de aprendizaje profundo**: se registran las arquitecturas o familias arquitectónicas reportadas en el estudio. Las categorías consideradas incluyen *DenseNet*, *EfficientNet*, *Inception*, *MobileNet*, *ResNet*, *YOLO*, *Faster R-CNN*, *Mask R-CNN* y *Other*.

- **Tipo de tarea**: se identifica si el estudio aborda una o múltiples tareas de *Classification*, *Detection* o *Segmentation*.

- **Métricas reportadas**: se registran las métricas de evaluación cuantitativa utilizadas en cada estudio. Entre ellas se consideran *AUC*, *Accuracy*, *Dice*, *F1-score*, *FPS*, *IoU*, *Precision*, *Recall* y *mAP*.

- **Uso de transferencia de aprendizaje**: se registra si el estudio reporta o no el empleo de *transfer learning*.

- **Uso de aumento de datos**: se registra si el estudio reporta o no la aplicación de técnicas de *data augmentation*.

- **Uso de métodos de optimización**: se registra si el estudio reporta explícitamente el uso de un método de optimización durante el entrenamiento.

La codificación de estas variables se realiza a partir de la terminología explícita utilizada por los autores en cada artículo. Cuando un estudio reporta múltiples arquitecturas, métricas o tareas, todas ellas son registradas.

6.9 ANÁLISIS DE DATOS

El análisis de los datos se realiza de manera descriptiva y comparativa, permitiendo identificar tendencias, enfoques predominantes y factores técnicos que influyen en el desempeño de los modelos reportados en la literatura. Dado que la investigación se basa en una revisión sistemática de estudios previos, no se aplican pruebas estadísticas inferenciales propias, limitándose el análisis a la comparación de métricas reportadas por los autores con el fin de sintetizar y contrastar los resultados existentes. Para RQ3, la asociación entre factores técnicos y mejor desempeño se interpreta en términos comparativos dentro de bloques metodológicamente semejantes, sin pretender inferencia causal.

7. RESULTADOS Y DISCUSIÓN

Los resultados se organizan en función de las preguntas de investigación, de modo que cada una de ellas estructura una parte específica de la síntesis.

Para responder la RQ1, se analizan las arquitecturas y tipos de modelos de aprendizaje profundo reportados con mayor frecuencia, así como su relación con el tipo de tarea abordada. Para responder la RQ2, se identifican y comparan las métricas de evaluación más empleadas en los estudios incluidos. Finalmente, para responder la RQ3, se examinan los factores técnicos descritos en la literatura, particularmente el uso de transferencia de aprendizaje, aumento de datos y métodos de optimización, con el fin de identificar su asociación con niveles superiores de desempeño reportado. En esta monografía, dicha asociación se interpreta a partir de la recurrencia de estos factores en estudios con mejores resultados dentro de bloques metodológicamente comparables.

7.1 RQ1. ARQUITECTURAS Y TIPOS DE MODELOS MÁS UTILIZADOS

Para responder la primera pregunta de investigación, primero se identificó el tipo de tarea abordada por los estudios incluidos, debido a que la selección de una arquitectura de aprendizaje profundo depende directamente del problema que se busca resolver. En este sentido, los trabajos revisados fueron organizados según tres tareas principales: detección, clasificación y segmentación. Posteriormente, se analizaron las arquitecturas generales y específicas empleadas en cada estudio, con el fin de reconocer los modelos más frecuentes.

En primer lugar, la Tabla 7.1 presenta la distribución de los estudios según el tipo de tarea abordada. Esta clasificación permite establecer el contexto metodológico desde el cual deben interpretarse las arquitecturas utilizadas, ya que no todos los modelos responden al mismo propósito: algunos se orientan a localizar polinizadores en imágenes, otros a clasificarlos por especie o grupo, y otros a delimitar regiones específicas mediante segmentación.

Tabla 7.1

Tipos de tarea reportados por estudio.

Estudio	Detección	Clasificación	Segmentación
Alves et al., 2020	×	×	×
Nizam et al., 2020	×		×
Dembski y Szymański, 2020	×	×	
J. Zhang et al., 2021		×	
Braga et al., 2021	×	×	
Spiesman et al., 2021		×	
Kelley et al., 2021		×	
Bożek et al., 2021	×		×
Voudiotis et al., 2021	×		
Voudiotis et al., 2022	×		
Suzuki-Ohno et al., 2022		×	
Rodrigues et al., 2022	×	×	×
Klasen et al., 2022		×	
Balmaki et al., 2022		×	
Bjerge et al., 2022	×	×	
García et al., 2022		×	
Kulyukin y Kulyukin, 2023	×		
Bjerge, Alison et al., 2023	×	×	
Gernat et al., 2023	×		
Yoo et al., 2023	×	×	
Stiemer et al., 2023	×		
Bjerge, Frigaard y Karstoft, 2023	×		
Stark et al., 2023	×	×	
Leocádio et al., 2024	×		
Lei et al., 2024	×		
Bjerge et al., 2024	×	×	×
Rohafauzi et al., 2024	×		
Spiesman et al., 2024		×	
Liang, 2024	×		
Santiago-Plaza et al., 2024		×	
Nguyen et al., 2024	×	×	
Kongsilp et al., 2024	×		×
Sledevič et al., 2024	×		
Sledevič y Matuzevičius, 2024	×		×
Rozenbaum et al., 2024	×		
Wittmann et al., 2024	×		
Ho et al., 2024	×		
Kriouile et al., 2024	×		×
Padubidri et al., 2024	×		
Dadgostar et al., 2025		×	
Ostwald et al., 2025			×
Taniguchi et al., 2025	×	×	
Lee et al., 2025	×		
Alex et al., 2025	×		
Calisan y Olgun, 2025	×	×	
C. L. Srinivas et al., 2025		×	
Copeland et al., 2025		×	
Oliveira et al., 2025		×	
Chiranjeevi et al., 2025		×	
Gomez-Jaime et al., 2025	×	×	
Ştefan et al., 2025	×	×	
Stark et al., 2025		×	
Sledevič et al., 2025	×		×

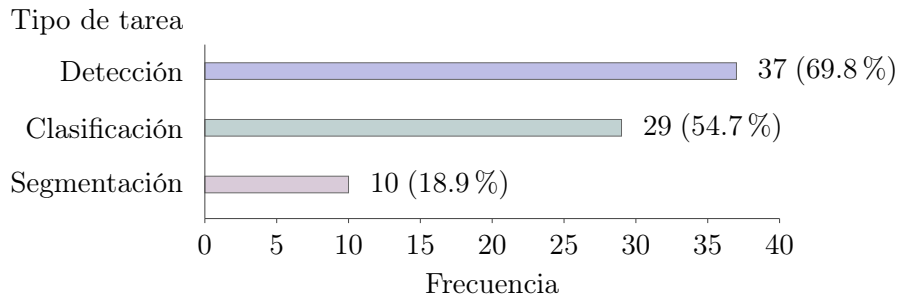
Total	**37**	**29**	**10**

Nota. Elaboración propia.

A partir de la identificación individual de las tareas abordadas por cada estudio, se sintetizó la frecuencia agregada de cada tipo de tarea dentro del corpus analizado. Las categorías no son mutuamente excluyentes, ya que un mismo estudio puede abordar simultáneamente detección, clasificación y segmentación. Por ello, la Figura 7.1 presenta las ocurrencias de cada tarea y su porcentaje respecto del total de estudios incluidos.

Figura 7.1

Frecuencia de tipos de tarea reportados en los estudios analizados.



Nota. Elaboración propia.

La tarea más frecuente fue la detección, presente en 37 estudios (69.8 %), seguida de la clasificación con 29 estudios (54.7 %) y, en menor medida, la segmentación con 10 estudios (18.9 %). Este resultado indica que la literatura reciente se ha concentrado principalmente en localizar polinizadores dentro de imágenes, lo cual explica la alta presencia posterior de arquitecturas orientadas a detección de objetos.

Una vez establecido el tipo de tarea predominante, se analizaron las arquitecturas generales y específicas de aprendizaje profundo reportadas en los 53 estudios incluidos. En la Tabla 7.2 se indica la presencia de los modelos en cada estudio analizado.

Tabla 7.2

Arquitecturas generales de aprendizaje profundo reportadas por estudio.

Estudio	Efficient Net	Mobile Net	Inception	ResNet	DenseNet	YOLO	Faster R-CNN	Mask R-CNN	Otros
Alves et al., 2020	×		×	×	×				×
Nizam et al., 2020							×	×	
Dembski y Szymański, 2020									×
J. Zhang et al., 2021		×			×				×
Braga et al., 2021									×
Spiesman et al., 2021			×		×				×
Kelley et al., 2021				×	×	×	×		
Bożek et al., 2021									×
Voudiotis et al., 2021			×	×			×		×
Voudiotis et al., 2022						×			×
Suzuki-Ohno et al., 2022									×
Rodrigues et al., 2022				×					×
Klasen et al., 2022									×
Balmaki et al., 2022									×
Bjerge et al., 2022						×			×
García et al., 2022									×
Kulyukin y Kulyukin, 2023						×			
Bjerge, Alison et al., 2023						×			
Gernat et al., 2023									×
Yoo et al., 2023					×				×
Stiemer et al., 2023						×	×		×
Bjerge, Frigaard y Karstoft, 2023						×	×		
Stark et al., 2023						×			
Leocádio et al., 2024						×			
Lei et al., 2024						×			
Bjerge et al., 2024		×				×			×
Rohafauzi et al., 2024	×					×			
Spiesman et al., 2024		×							
Liang, 2024			×	×		×			×
Santiago-Plaza et al., 2024					×				
Nguyen et al., 2024						×	×		
Kongsilp et al., 2024					×			×	
Sledevič et al., 2024						×			
Sledevič y Matuzevičius, 2024						×			
Rozenbaum et al., 2024						×			
Wittmann et al., 2024						×			
Ho et al., 2024						×	×	×	×
Kriouile et al., 2024					×	×	×	×	
Padubidri et al., 2024									×
Dadgostar et al., 2025					×				
Ostwald et al., 2025									×
Taniguchi et al., 2025						×			×
Lee et al., 2025						×			
Alex et al., 2025						×			
Calisan y Olgun, 2025					×				
C. L. Srinivas et al., 2025				×					

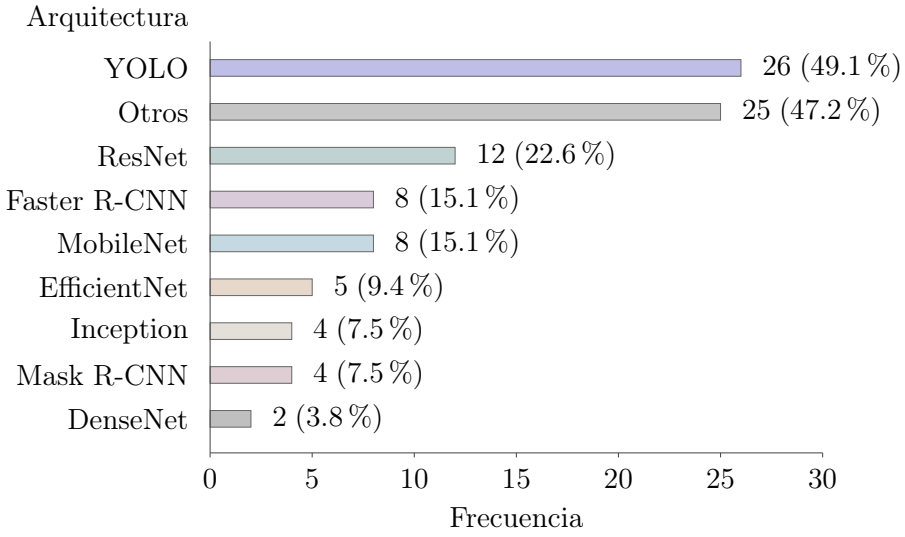
Copeland et al., 2025					×				
Oliveira et al., 2025		×		×					×
Chiranjeevi et al., 2025						×			×
Gomez-Jaime et al., 2025						×			
Ștefan et al., 2025						×			
Stark et al., 2025		×		×	×				
Sledevič et al., 2025						×			
Total	**5**	**8**	**4**	**12**	**2**	**26**	**8**	**4**	**25**

Nota. Elaboración propia.

De la misma manera que en la clasificación de estudios por tipo de tarea, las frecuencias reportadas en la Figura 7.2 representan ocurrencias de uso y no categorías mutuamente excluyentes. Por tanto, los porcentajes deben interpretarse como la proporción de estudios en los que se identificó cada arquitectura respecto del total de artículos analizados.

Figura 7.2

Frecuencia de arquitecturas de aprendizaje profundo reportadas en los estudios analizados.



Nota. Elaboración propia.

En la Figura 7.2 se observa un predominio claro de la familia YOLO, presente en 26 estudios (49.1 % del corpus). En segundo lugar se ubica el grupo de otras arquitecturas con 25 estudios (47.2 %), seguido de ResNet con 12 estudios (22.6 %). Las arquitecturas Faster R-CNN y MobileNet aparecen en 8 estudios cada una (15.1 %), mientras que EfficientNet se identifica en 5 estudios (9.4 %). Inception y Mask R-CNN se reportan en 4 estudios cada una (7.5 %) y DenseNet en 2 estudios (3.8 %).

En relación con el predominio de la detección identificado previamente en la distribución de tareas, la frecuencia de arquitecturas confirma una correspondencia metodológica entre el tipo de problema abordado y los modelos seleccionados, especialmente en variantes de YOLO, mientras que arquitecturas clásicamente asociadas a clasificación profunda, como EfficientNet, MobileNet o ResNet, aparecen con una presencia relevante, aunque menor, dentro del corpus revisado.

Sin embargo, no es suficiente con reportar las principales familias de arquitecturas; por ello se establece qué modelos específicos aparecen con más frecuencia. En la Tabla 7.3 se determina la frecuencia de las arquitecturas en los diferentes estudios.

Tabla 7.3

Arquitecturas específicas reportadas en los estudios incluidos.

Arquitectura específica	**Frecuencia**	**Porcentaje (%)**	**Arquitectura específica**	**Frecuencia**	**Porcentaje (%)**
YOLO	16	30.2	DenseNet	1	1.9
YOLOv5	8	15.1	DenseNet201	1	1.9
Faster R-CNN	7	13.2	EfficientNet	1	1.9
ResNet	7	13.2	EfficientNetB0	1	1.9
CNN	4	7.5	EfficientNetB1	1	1.9
Mask R-CNN	4	7.5	EfficientNetV2L	1	1.9
MobileNetV2	4	7.5	GoogLeNet	1	1.9
SSD	4	7.5	InsectNet	1	1.9
YOLOv8	4	7.5	Keypoint detector	1	1.9
InceptionV3	3	5.7	MnasNet	1	1.9
ResNet50	3	5.7	MnasNet0_5	1	1.9
U-Net	3	5.7	MnasNet1_3	1	1.9
VGG	3	5.7	Pose estimation	1	1.9
YOLOv7tiny	3	5.7	RegNetX400MF	1	1.9
EfficientNetB4	2	3.8	RegNetY1_6GF	1	1.9
Faster R-CNN-InceptionV2	2	3.8	RegNetY400MF	1	1.9
MobileNet	2	3.8	RegNetY800MF	1	1.9
MobileNetV3	2	3.8	ResNet18	1	1.9
ResNet101	2	3.8	RetinaNet	1	1.9
Transformer	2	3.8	SSD-InceptionV2	1	1.9
VGG16	2	3.8	SSD-MobileNetV1	1	1.9
Xception	2	3.8	SSD-MobileNetV2	1	1.9
YOLOv3	2	3.8	Swarm-engine	1	1.9

YOLOv5nano	2	3.8	VGG11	1	1.9
YOLOv5small	2	3.8	VGG19	1	1.9
AlexNet	1	1.9	Wide ResNet	1	1.9
CAMNN	1	1.9	YOLOv4	1	1.9
DNN-CNN	1	1.9	YOLOv4Tiny	1	1.9
DeepLabv3	1	1.9	YOLOv5m	1	1.9
DeepWings	1	1.9	YOLOv5s	1	1.9

Nota. Elaboración propia.

Entre las arquitecturas específicas, YOLO fue la más reportada, seguida por YOLOv5, Faster R-CNN y ResNet, lo que confirma el predominio de enfoques orientados a detección y clasificación. Además, la presencia de múltiples variantes de una misma familia, especialmente en YOLO, ResNet y MobileNet, sugiere que la literatura no solo se concentra en arquitecturas generales, sino también en versiones adaptadas para mejorar el desempeño según la tarea y el contexto de aplicación.

7.2 RQ2. MÉTRICAS DE EVALUACIÓN MÁS REPORTADAS

Para responder la segunda pregunta de investigación, se identificaron las métricas de evaluación empleadas en los estudios incluidos, con el fin de reconocer cuáles son los criterios más utilizados para valorar el desempeño de los modelos de aprendizaje profundo aplicados al reconocimiento de polinizadores. La Tabla 7.4 indica las métricas reportadas por estudio.

Tabla 7.4

Métricas de evaluación reportadas por estudio.

Estudio	Exactitud	Precisión	Recall	mAP	F1	IoU	FPS	AUC	Dice
Alves et al., 2020	×	×	×		×				
Nizam et al., 2020	×			×			×		
Dembski y Szymański, 2020		×	×			×			
J. Zhang et al., 2021	×	×		×				×	
Braga et al., 2021	×	×	×		×	×	×		
Spiesman et al., 2021	×	×	×						
Kelley et al., 2021	×	×							
Bożek et al., 2021	×			×			×		
Voudiotis et al., 2021	×					×			
Voudiotis et al., 2022	×	×	×	×	×	×			
Suzuki-Ohno et al., 2022	×	×	×		×				
Rodrigues et al., 2022	×	×		×					
Klasen et al., 2022	×								
Balmaki et al., 2022	×		×						

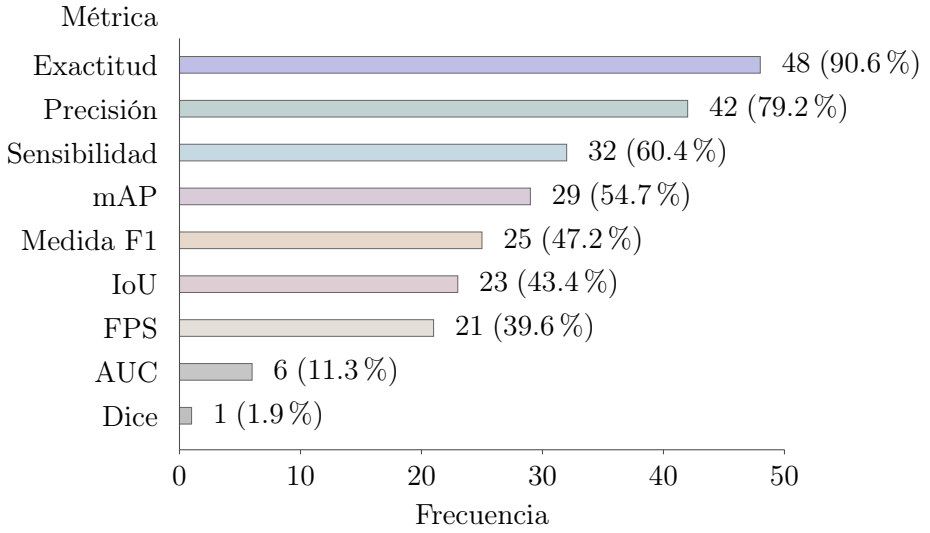
Bjerge et al., 2022	×	×	×	×		×	×			
García et al., 2022	×									
Kulyukin y Kulyukin, 2023	×	×			×	×	×			
Bjerge, Alison et al., 2023	×	×	×	×	×	×	×			
Gernat et al., 2023			×	×	×					
Yoo et al., 2023	×	×	×	×	×					
Stiemer et al., 2023	×	×		×	×	×	×			
Bjerge, Frigaard y Karstoft, 2023	×	×	×		×	×	×			
Stark et al., 2023	×	×	×	×		×				
Leocádio et al., 2024		×					×			
Lei et al., 2024	×	×	×	×	×	×	×			
Bjerge et al., 2024	×	×		×	×			×		
Rohafauzi et al., 2024	×	×								
Spiesman et al., 2024	×	×	×	×	×					
Liang, 2024	×	×	×	×	×		×			
Santiago-Plaza et al., 2024	×						×			
Nguyen et al., 2024	×	×	×	×	×		×			
Kongsilp et al., 2024	×	×	×	×		×	×			
Sledevič et al., 2024	×	×	×	×		×	×			
Sledevič y Matuzevičius, 2024							×			
Rozenbaum et al., 2024	×	×		×		×	×			
Wittmann et al., 2024	×	×	×	×			×			
Ho et al., 2024	×	×		×		×				
Kriouile et al., 2024	×	×	×	×		×				
Padubidri et al., 2024		×				×				
Dadgostar et al., 2025	×	×	×	×	×			×		
Ostwald et al., 2025	×	×	×		×	×				×
Taniguchi et al., 2025	×	×	×		×	×		×		
Lee et al., 2025	×	×	×	×	×	×				
Alex et al., 2025	×	×	×	×		×	×			
Calisan y Olgun, 2025	×	×	×	×	×					
C. L. Srinivas et al., 2025	×	×	×		×					
Copeland et al., 2025	×	×	×		×					
Oliveira et al., 2025	×	×	×	×	×			×		
Chiranjeevi et al., 2025	×									
Gomez-Jaime et al., 2025	×	×		×		×				
Ștefan et al., 2025	×	×	×		×	×	×	×		
Stark et al., 2025	×									
Sledevič et al., 2025	×	×	×	×	×		×			
Total	**48**	**42**	**32**	**29**	**25**	**23**	**21**	**6**		**1**

Nota. Elaboración propia.

A partir del registro individual presentado en la Tabla 7.4, se sintetizaron las frecuencias generales de uso de cada métrica con el propósito de identificar los criterios de evaluación predominantes en los estudios analizados, como se indica en la Figura 7.3.

Figura 7.3

Frecuencia de métricas de evaluación reportadas en los estudios analizados.



Nota. Elaboración propia.

En la Figura 7.3 se observa que la métrica más reportada es la exactitud, presente en 48 estudios (90.6 %), seguida de precisión con 42 estudios (79.2 %) y sensibilidad con 32 estudios (60.4 %). Entre las métricas más vinculadas a detección destaca mAP, reportada en 29 estudios (54.7 %), mientras que IoU aparece en 23 estudios (43.4 %) y FPS en 21 estudios (39.6 %). La medida F1 se reporta en 25 estudios (47.2 %), en tanto que AUC y Dice presentan una presencia mucho más limitada, con 6 y 1 estudios, respectivamente.

Este patrón indica que la literatura privilegia métricas de uso general, como exactitud, precisión y sensibilidad, pero al mismo tiempo incorpora indicadores específicos de detección, como mAP, IoU y FPS, cuando la tarea lo requiere. En contraste, las métricas más propias de segmentación aparecen con menor frecuencia, lo que también es consistente con la menor presencia de esta tarea en el corpus revisado.

Aunque la frecuencia general de las métricas ofrece una primera aproximación al patrón de evaluación, su interpretación requiere considerar el tipo de tarea y la familia de modelos utilizada. En este sentido, la Tabla 7.5 resume dicha relación.

Tabla 7.5

Relación comparativa entre tarea, tipo de modelo y métricas predominantes

Tarea	**Tipo de modelo predominante**	**Métricas predominantes**	**Comparativa**
Clasificación	ResNet, EfficientNet, MobileNet, Inception y otras CNN	Exactitud, precisión, recall, medida F1 y, en menor medida, AUC	La evaluación se centra en la capacidad de discriminar correctamente entre clases y resumir el desempeño global del clasificador.
Detección	YOLO, Faster R-CNN y Mask R-CNN	mAP, IoU, FPS, además de precisión y recall	La evaluación combina calidad de localización, capacidad de detección y velocidad de inferencia, especialmente en escenarios de monitoreo automatizado.
Segmentación	Mask R-CNN, YOLO con extensión de segmentación y otras arquitecturas de segmentación	IoU, Dice y medida F1	La evaluación se orienta a la precisión espacial y al grado de solapamiento entre la predicción y la región de referencia.

Nota. Elaboración propia.

La Tabla 7.5 permite observar que las métricas cambian en frecuencia y en función del objetivo analítico del modelo. En clasificación predominan métricas asociadas al acierto global y al equilibrio entre clases, mientras que en detección adquieren mayor relevancia aquellas que combinan localización, precisión y velocidad. En segmentación, por su parte, las métricas se orientan a evaluar superposición espacial, lo que explica la presencia de IoU y Dice, aunque esta última aparezca con menor frecuencia debido a la reducida proporción de estudios de segmentación en el corpus.

En consecuencia, la respuesta a RQ2 indica que la evaluación del reconocimiento de polinizadores mediante aprendizaje profundo se apoya principalmente en exactitud, precisión y sensibilidad a nivel global; sin embargo, cuando los estudios se comparan de acuerdo con la tarea

y el tipo de modelo, se observa una diferenciación más clara: la clasificación privilegia exactitud, precisión, sensibilidad y medida F1; la detección incorpora de forma más marcada mAP, IoU y FPS; y la segmentación se apoya en métricas de superposición espacial como IoU, Dice y medida F1.

7.3 RQ3. FACTORES TÉCNICOS ASOCIADOS

Para responder la tercera pregunta de investigación, se identificaron tres factores técnicos recurrentes en los estudios incluidos: el aprendizaje por transferencia, el aumento de datos y los métodos de optimización. En primer lugar, se presenta la Tabla 7.6 para reconocer qué estrategias fueron reportadas en cada investigación.

Tabla 7.6

Factores técnicos generales reportados por estudio.

Estudio	**Aprendizaje por transferencia**	**Aumento de datos**	**Método de optimización**
Alves et al., 2020	×	×	×
Nizam et al., 2020	×	×	×
Dembski y Szymański, 2020			×
J. Zhang et al., 2021	×		
Braga et al., 2021	×	×	
Spiesman et al., 2021	×	×	×
Kelley et al., 2021	×	×	×
Bożek et al., 2021		×	
Voudiotis et al., 2021	×		
Voudiotis et al., 2022	×	×	
Suzuki-Ohno et al., 2022		×	
Rodrigues et al., 2022	×	×	×
Klasen et al., 2022	×	×	×
Balmaki et al., 2022	×	×	
Bjerge et al., 2022		×	
García et al., 2022			×
Kulyukin y Kulyukin, 2023		×	×
Bjerge, Alison et al., 2023	×	×	
Gernat et al., 2023		×	×
Yoo et al., 2023	×	×	×
Stiemer et al., 2023	×	×	×
Bjerge, Frigaard y Karstoft, 2023	×	×	
Stark et al., 2023	×	×	×
Leocádio et al., 2024		×	

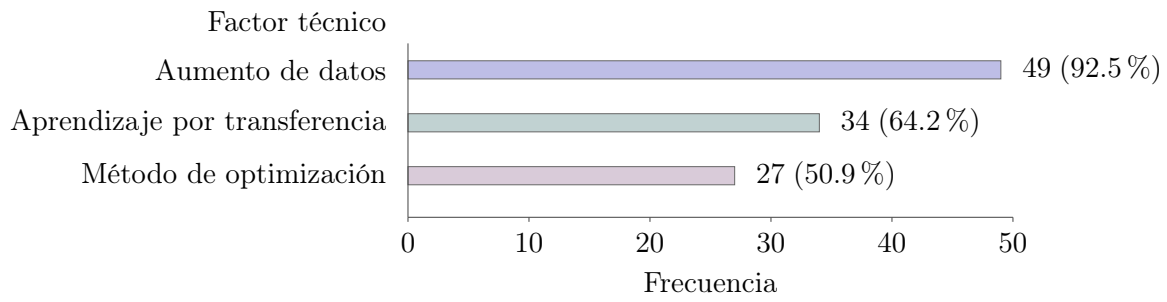
Lei et al., 2024		×	
Bjerge et al., 2024	×	×	×
Rohafauzi et al., 2024	×	×	
Spiesman et al., 2024		×	×
Liang, 2024	×	×	×
Santiago-Plaza et al., 2024	×	×	×
Nguyen et al., 2024	×	×	
Kongsilp et al., 2024	×	×	×
Sledevič et al., 2024		×	×
Sledevič y Matuzevičius, 2024		×	
Rozenbaum et al., 2024	×	×	
Wittmann et al., 2024		×	
Ho et al., 2024	×	×	
Kriouile et al., 2024	×	×	
Padubidri et al., 2024		×	
Dadgostar et al., 2025		×	×
Ostwald et al., 2025		×	×
Taniguchi et al., 2025	×	×	×
Lee et al., 2025		×	
Alex et al., 2025	×	×	
Calisan y Olgun, 2025	×	×	×
C. L. Srinivas et al., 2025	×	×	×
Copeland et al., 2025	×	×	×
Oliveira et al., 2025	×	×	×
Chiranjeevi et al., 2025	×	×	
Gomez-Jaime et al., 2025		×	
Ștefan et al., 2025	×	×	
Stark et al., 2025	×	×	×
Sledevič et al., 2025		×	
Total	**34**	**49**	**27**

Nota. Elaboración propia.

A partir de la Tabla 7.6, se calcularon las frecuencias generales de aparición de cada factor técnico, como se indica en la Figura 7.4.

Figura 7.4

Frecuencia general de factores técnicos reportados en los estudios analizados.



Nota. Elaboración propia.

El aumento de datos fue el factor técnico más frecuente, con 49 apariciones, equivalente al 92,5 % de los estudios. Le siguen el aprendizaje por transferencia, con 34 apariciones, y los métodos de optimización, con 27. Este patrón sugiere que la ampliación y diversificación artificial de los datos constituye una estrategia central en el reconocimiento de polinizadores, probablemente debido a la limitada disponibilidad de imágenes etiquetadas, la variabilidad de condiciones de captura y la necesidad de mejorar la capacidad de generalización de los modelos.

Para profundizar en el análisis de RQ3, se revisaron las estrategias específicas incluidas dentro de cada factor técnico general. Esta revisión permitió identificar cuáles fueron las prácticas más recurrentes en la literatura y cómo se distribuyeron entre estudios metodológicamente comparables. La Tabla 7.7 resume estas menciones.

Tabla 7.7

Estrategias técnicas específicas reportadas en los estudios incluidos.

Factor técnico	**Detalle específico**	**Frecuencia**	**Porcentaje (%)**
Aprendizaje por transferencia	Preentrenado	24	45.3
	Transfer learning	15	28.3
	Ajuste fino	10	18.9
Aumento de datos	Aumento de datos	27	50.9
	Recorte	27	50.9
	Rotación	14	26.4
	Volteo	6	11.3
	Mosaico	3	5.7

	Muestreo estratificado	2	3.8
	Datos sintéticos	1	1.9
Método de optimización	Tasa de aprendizaje	21	39.6
	Optimizador	12	22.6
	Adam	10	18.9
	Momentum	6	11.3
	SGD	6	11.3
	Decaimiento de peso	3	5.7
	AdamW	2	3.8

Nota. Elaboración propia.

En aprendizaje por transferencia predominó el uso de modelos preentrenados, seguido de las menciones explícitas a *transfer learning* y ajuste fino. En aumento de datos se repitieron con mayor frecuencia las menciones generales a *data augmentation* y recorte, seguidas por rotación y volteo. En optimización, la estrategia más reportada fue el ajuste de la tasa de aprendizaje, seguida por la mención general del optimizador y del algoritmo Adam.

En la misma línea, la asociación con mayores niveles de precisión no se interpreta como causalidad estadística, sino como recurrencia comparativa de factores técnicos en estudios con tarea, métricas y condiciones de evaluación semejantes.

Se identificaron cuatro bloques metodológicos comparables que permiten organizar la discusión de los resultados. Esta agrupación busca reconocer patrones comunes entre investigaciones que comparten objetivos técnicos similares.

En el primer bloque se ubican los estudios orientados a tareas de clasificación. En esta línea, Alves et al. (2020), Spiesman et al. (2021), Spiesman et al. (2024), Calisan y Olgun (2025) y Stark et al. (2025) reportan el uso de métricas como exactitud, precisión, sensibilidad y medida F1, junto con arquitecturas basadas principalmente en redes convolucionales, como ResNet, EfficientNet, Inception y MobileNet. En estos trabajos, el aprendizaje por transferencia, el aumento de datos, el recorte, la rotación y métodos de optimización como Adam, AdamW o descenso de gradiente estocástico aparecen como estrategias asociadas con la mejora de la generalización y la estabilidad del entrenamiento.

Un segundo bloque corresponde a estudios de detección evaluados mediante métricas espaciales y de eficiencia, como mAP, IoU y FPS. En este grupo, Stark et al. (2023), Alex et al. (2025), Lei et al. (2024), Sledevič et al. (2024), Ștefan et al. (2025) y Sledevič et al. (2025) emplean principalmente arquitecturas como YOLO, Faster R-CNN y Mask R-CNN. Estos estudios enfatizan estrategias como el ajuste fino, el aumento de datos, el recorte, el mosaico, la rotación, el aprendizaje previo y el ajuste de la tasa de aprendizaje, lo que sugiere que el rendimiento en detección depende tanto de la arquitectura como de la preparación del entrenamiento.

Un tercer bloque reúne estudios de detección que, aunque trabajan con localización de objetos, reportan métricas más cercanas a la clasificación, como exactitud, precisión, sensibilidad y medida F1. En este caso, Kongsilp et al. (2024), Nguyen et al. (2024) y Bjerge, Alison et al. (2023) muestran combinaciones entre arquitecturas como YOLO, Faster R-CNN, Mask R-CNN y ResNet, junto con aprendizaje por transferencia, ajuste fino, aumento de datos, recorte y métodos de optimización como momentum, Adam o descenso de gradiente estocástico. Este bloque evidencia que algunos estudios de detección evalúan el desempeño desde una lógica general de acierto del modelo, lo que limita la comparación directa con investigaciones que priorizan métricas espaciales como mAP o IoU.

Finalmente, el cuarto bloque corresponde a los estudios enfocados en segmentación. En esta categoría, Nizam et al. (2020), Kongsilp et al. (2024), Rodrigues et al. (2022) y Sledevič et al. (2025) reportan métricas como IoU, coeficiente Dice y medida F1, junto con arquitecturas como Mask R-CNN, YOLO y otros enfoques de segmentación. En estos trabajos, el aumento de datos, el recorte, el aprendizaje por transferencia, el ajuste fino y las configuraciones específicas de entrenamiento aparecen como factores relevantes para mejorar la precisión espacial del modelo.

La respuesta a RQ3 indica que los factores técnicos asociados con mayores niveles de precisión reportada corresponden a combinaciones metodológicas recurrentes. En clasificación destacan la transferencia de aprendizaje y el aumento de datos; en detección, la asociación más clara se observa entre aumento de datos, ajuste fino y calibración de la tasa de aprendizaje, especialmente en estudios basados en YOLO; y en segmentación, aunque la evidencia es menor, vuelven a repetirse el aumento de datos y la adaptación del modelo mediante aprendizaje por transferencia.

7.4 DISCUSIÓN

Los resultados de esta revisión muestran que el reconocimiento de imágenes de polinizadores mediante aprendizaje profundo constituye una línea activa de investigación, pero todavía presenta limitaciones en su consolidación metodológica. Aunque los estudios revisados evidencian avances importantes, no es posible afirmar que exista una arquitectura superior para todos los casos. El desempeño de los modelos depende del tipo de polinizador analizado, la tarea abordada, las condiciones de captura, el nivel taxonómico, el conjunto de datos y las métricas utilizadas (Høye et al., 2021; Khanzhina et al., 2022; Ratnayake et al., 2021). Por ello, la comparación entre estudios se realizó dentro de grupos metodológicamente semejantes.

Una primera tendencia relevante es el predominio de la detección como tarea principal. Esto resulta coherente con las necesidades del monitoreo de polinizadores, donde muchas aplicaciones buscan localizar individuos en imágenes para estimar presencia, actividad, visitas florales, tránsito en colmenas o patrones de comportamiento. En estos escenarios, detectar al polinizador dentro de una escena compleja puede ser más útil que clasificar una imagen completa, ya que permite trabajar con registros visuales obtenidos en condiciones reales de campo (Alex et al., 2025; Bjerge et al., 2022; Bjerge, Alison et al., 2023; Stark et al., 2023).

La alta presencia de modelos YOLO también se entiende desde esta lógica aplicada. En estudios de polinizadores, donde los insectos suelen ser pequeños, móviles y aparecen sobre fondos variables, se requieren modelos capaces de localizar objetos con rapidez. YOLO resulta atractivo porque combina localización y clasificación en una sola etapa, lo que favorece sistemas de monitoreo casi en tiempo real. Sin embargo, su predominio no implica que sea el mejor modelo para cualquier tarea. Su ventaja se observa principalmente cuando el objetivo es detectar polinizadores de forma eficiente en imágenes (Alex et al., 2025; Kulyukin y Kulyukin, 2023; Stark et al., 2023; Ștefan et al., 2025).

En clasificación, las arquitecturas convolucionales profundas siguen siendo relevantes para diferenciar especies, familias o grupos funcionales de polinizadores. Esta tarea es especialmente exigente porque muchas especies presentan similitudes morfológicas, variación intraespecífica y rasgos visuales sutiles, como patrones en alas, coloración, forma corporal o estructuras especí-

ficas. Por tanto, el desempeño en clasificación depende del modelo, la calidad de las imágenes, la precisión de las etiquetas, el balance entre clases y el nivel taxonómico evaluado (Calisan y Olgun, 2025; Spiesman et al., 2021; Spiesman et al., 2024; C. L. Srinivas et al., 2025).

La segmentación aparece con menor frecuencia, pero tiene un valor importante para el análisis detallado de polinizadores. Esta tarea permite separar al insecto del fondo, delimitar individuos superpuestos o identificar regiones corporales específicas. Su baja presencia puede deberse a que requiere anotaciones más costosas y precisas, especialmente cuando se trabaja con organismos pequeños y fondos naturales complejos. Aun así, la segmentación puede ser clave en estudios donde se necesita analizar al polinizador en cuanto a: su forma, posición o interacción con flores y otros individuos (Kriouile et al., 2024; Nizam et al., 2020; Rodrigues et al., 2022; Sledevič et al., 2025).

Otro aspecto central es que las métricas condicionan la interpretación del rendimiento. En reconocimiento de polinizadores, una exactitud alta puede ser útil para clasificación, pero no indica necesariamente que el modelo localice bien al insecto dentro de la imagen. Por eso, en detección son más informativas métricas como mAP, IoU y FPS, ya que evalúan la calidad de localización y la viabilidad operativa del sistema. En segmentación, indicadores como IoU, Dice y F1 permiten valorar mejor el grado de coincidencia entre la región predicha y la región real del polinizador (Gupta et al., 2025; Saetchnikov et al., 2021; Varadarajan et al., 2021).

Esta diversidad de métricas evidencia una limitación importante del campo, ya que los resultados no siempre pueden compararse de forma directa entre estudios con tareas, datos y protocolos distintos. Sin embargo, esta revisión permite ordenar esa heterogeneidad mediante bloques metodológicamente comparables. Así, un modelo orientado a clasificar especies de abejas en imágenes controladas no se contrasta de la misma manera que un modelo diseñado para segmentar individuos en fondos florales complejos. Desde esta perspectiva, el análisis por bloques aporta una forma más rigurosa de interpretar el rendimiento reportado en la literatura y evita establecer comparaciones generales entre estudios que responden a objetivos técnicos diferentes.

Los factores técnicos asociados al entrenamiento son especialmente relevantes en este dominio. Las imágenes de polinizadores suelen presentar variaciones de iluminación, escala, pos-

tura, movimiento, fondo, oclusión y calidad de captura. Además, los conjuntos de datos suelen ser limitados o desbalanceados, porque no todas las especies aparecen con la misma frecuencia ni son igual de fáciles de fotografiar. En este contexto, el aumento de datos, el aprendizaje por transferencia, el ajuste fino y la calibración de hiperparámetros son condiciones que pueden marcar la diferencia entre un modelo útil y uno que solo funciona en condiciones demasiado controladas (M. Kaur et al., 2022; Lee et al., 2025; Nguyen et al., 2024; Sajib et al., 2025).

El aumento de datos adquiere especial importancia porque permite simular parte de la variabilidad visual propia del monitoreo de polinizadores. Transformaciones como recorte, rotación, volteo, cambios de brillo o mosaico pueden ayudar al modelo a reconocer insectos en diferentes posiciones, tamaños y condiciones de iluminación. Sin embargo, estas técnicas deben aplicarse con cuidado, ya que una transformación excesiva podría alterar rasgos morfológicos relevantes para diferenciar especies o grupos taxonómicos (Agarwal et al., 2025; M. Kaur et al., 2022; Lee et al., 2025).

El aprendizaje por transferencia también resulta pertinente porque muchos estudios trabajan con datos reducidos o difíciles de etiquetar. Aprovechar modelos preentrenados permite partir de representaciones visuales generales y adaptarlas al reconocimiento de polinizadores mediante ajuste fino. Esto puede reducir el costo de entrenamiento y mejorar la generalización, especialmente cuando se dispone de pocas imágenes por especie. No obstante, su efectividad depende de qué tan distinto sea el dominio de entrenamiento original respecto a las imágenes reales de polinizadores, que suelen incluir fondos naturales, flores, colmenas, trampas de cámara o especímenes capturados (Burri et al., 2023; Jagtap y Yadav, 2025; Nguyen et al., 2024; J. Zhang et al., 2021).

La optimización del entrenamiento también influye en el desempeño final. El uso de optimizadores como Adam, AdamW o SGD, junto con la configuración de la tasa de aprendizaje, la regularización y las políticas de entrenamiento, puede explicar diferencias entre estudios que emplean arquitecturas similares. Dos modelos basados en YOLO o ResNet pueden comportarse de manera distinta si fueron entrenados con datos, aumentos, hiperparámetros y protocolos de validación diferentes (Azis et al., 2025; Postalcloğlu, 2020; Sudewo et al., 2025).

Desde una perspectiva ecológica, estos avances son relevantes porque pueden apoyar el monitoreo de polinizadores en un contexto de creciente preocupación por su declive. La automatización del reconocimiento visual puede reducir la dependencia exclusiva de observación manual, facilitar el procesamiento de grandes volúmenes de imágenes y permitir seguimientos más frecuentes. En este sentido, su aplicación en monitoreo ecológico exige que los modelos sean evaluados más allá de bases controladas y demuestren robustez en escenarios reales, donde los polinizadores se mueven, se ocultan parcialmente, cambian de orientación y aparecen sobre fondos visualmente complejos (Bjerge et al., 2022; Hellwig et al., 2024; Stark et al., 2023; Wittmann et al., 2024).

La evidencia revisada muestra que el reconocimiento automatizado de polinizadores debe evaluarse por la precisión alcanzada por un modelo y por su capacidad para responder a una necesidad ecológica concreta. Detectar, clasificar o segmentar polinizadores implica objetivos distintos dentro del monitoreo, por lo que la utilidad de cada arquitectura depende de su ajuste con el tipo de imagen, el nivel taxonómico, las condiciones de captura y el protocolo de evaluación. Desde esta perspectiva, el avance del campo requiere fortalecer la validación en escenarios reales y ampliar tareas menos exploradas, especialmente la segmentación. Así, los modelos de aprendizaje profundo podrán aportar de manera más robusta al monitoreo ecológico y no limitarse a resultados favorables en entornos experimentales controlados.

8. CONCLUSIONES

A partir de la revisión sistemática realizada, se concluye que el reconocimiento de imágenes de polinizadores mediante aprendizaje profundo constituye una línea relevante dentro de la visión por computador aplicada al monitoreo ecológico. La organización de los estudios por bloques comparables permitió interpretar la evidencia de manera ordenada, relacionando el tipo de tarea, la arquitectura utilizada, las métricas de evaluación y las estrategias técnicas reportadas en cada investigación.

1. Respecto a la primera pregunta de investigación, se concluye que la elección de arquitecturas en el reconocimiento de imágenes de polinizadores está directamente vinculada con la tarea abordada. Los modelos de detección, especialmente la familia YOLO, resultan pertinentes cuando el objetivo es localizar polinizadores en imágenes y apoyar escenarios de monitoreo automatizado. En cambio, arquitecturas como ResNet, EfficientNet, MobileNet e Inception son más relevantes en tareas de clasificación, donde se busca diferenciar especies, familias o grupos funcionales a partir de rasgos visuales. La segmentación cumple una función complementaria, al permitir una delimitación espacial más detallada del organismo y su separación respecto del fondo.

2. La segunda pregunta de investigación permite concluir que las métricas de evaluación adquieren sentido según la tarea y el tipo de modelo analizado. En clasificación, exactitud, precisión, sensibilidad y medida F1 permiten valorar la asignación correcta de categorías. En detección, mAP, IoU y FPS son más adecuados porque integran localización, precisión del reconocimiento y velocidad de inferencia. En segmentación, IoU, Dice y medida F1 permiten evaluar la coincidencia espacial entre la predicción del modelo y la región real del polinizador. Por ello, la comparación del desempeño requiere considerar la correspondencia entre tarea, arquitectura y métrica empleada.

3. Para la tercera pregunta de investigación, se concluye que el rendimiento de los modelos se fortalece mediante estrategias técnicas orientadas a mejorar la generalización, la adaptación al dominio y la estabilidad del entrenamiento. El aumento de datos permite enfrentar

variaciones frecuentes en imágenes de polinizadores, como cambios de iluminación, postura, escala, fondo y movimiento. El aprendizaje por transferencia facilita el uso de modelos preentrenados en conjuntos de datos biológicos con disponibilidad limitada de imágenes etiquetadas. Los métodos de optimización y el ajuste de hiperparámetros contribuyen a mejorar el proceso de aprendizaje y el desempeño final de las arquitecturas utilizadas.

4. La comparación por bloques metodológicos constituye un aporte central de la revisión porque permite interpretar el estado del arte sin reducir el análisis a una única métrica o a una sola familia de modelos. Esta organización relaciona clasificación, detección y segmentación con sus arquitecturas, métricas y estrategias de entrenamiento correspondientes. De este modo, el desempeño de los modelos se interpreta dentro del propósito técnico para el cual fueron diseñados, evitando comparaciones generales entre estudios con objetivos, datos y condiciones de evaluación distintas.

5. El reconocimiento automatizado de polinizadores mediante aprendizaje profundo aporta una base técnica útil para el monitoreo ecológico, al facilitar la identificación visual de organismos y ampliar la capacidad de análisis de imágenes en contextos biológicos. La revisión permite establecer que la utilidad de estos modelos se fortalece cuando existe correspondencia entre la tarea de reconocimiento, la arquitectura seleccionada, la métrica empleada y la estrategia de entrenamiento aplicada. Esta relación constituye un criterio central para orientar aplicaciones futuras en función de las necesidades del monitoreo, la calidad de los datos disponibles y el nivel de detalle requerido en el análisis de polinizadores.

9. REFERENCIAS

Agarwal, A., Dash, A., Galbale, G. y Singh, S. (2025). Analyzing Impact of Data Augmentation Techniques in Computer Vision. *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, 733-737. https://doi.org/10.1109/CICTN64563.2025.10932445

Akter, R., Islam, M. S., Sohan, K. y Ahmed, M. I. (2023). Insect recognition and classification using optimized densely connected convolutional neural network. *Lecture Notes in Networks and Systems*, *624*, 251-264. https://doi.org/10.1007/978-3-031-25344-7_23

Alex, A. J., Barnes, C. M., Machado, P., Ihanlie, I., Markó, G., Bencsik, M. y Bird, J. J. (2025). Enhancing pollinator conservation: Monitoring of bees through object recognition. *Computers and Electronics in Agriculture*, *228*, 109665. https://doi.org/10.1016/j.compag.2024.109665

Alves, T. S., Pinto, M. A., Ventura, P., Neves, C. J., Biron, D. G., Junior, A. C., de Paula Filho, P. L. y Rodrigues, P. J. (2020). Automatic detection and classification of honey bee comb cells using deep learning. *Computers and Electronics in Agriculture*, *170*. https://doi.org/10.1016/j.compag.2020.105244

Andriyanov, N. y Andriyanov, D. (2020). Pattern Recognition on Radar Images Using Augmentation. *Proceedings - 2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, 289-291. https://doi.org/10.1109/USBEREIT48449.2020.9117669

Anumol, C. (2023). Advancements in CNN Architectures for Computer Vision: A Comprehensive Review. *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*. https://doi.org/10.1109/AICERA/ICIS59538.2023.10420413

Apicella, A., Isgrò, F. y Prevete, R. (2024). Hidden classification layers: Enhancing linear separability between classes in neural networks layers. *Pattern Recognition Letters*, *177*, 69-74. https://doi.org/10.1016/j.patrec.2023.11.016

Azis, H., Andini, M., Herman, Syafie, L., Manga, A. e Ilmawan, L. (2025). A Comparative Study of SGD, Adam, AdamW, and RMSprop Optimizers for VGG19-Based Rupiah Banknote Classification. *2025 9th International Conference on Electrical, Electronics*

and Information Engineering (ICEEIE). https://doi.org/10.1109/ICEEIE66203.2025.11251942

Balmaki, B., Rostami, M. A., Christensen, T., Leger, E. A., Allen, J. M., Feldman, C. R., Forister, M. L. y Dyer, L. A. (2022). Modern approaches for leveraging biodiversity collections to understand change in plant-insect interactions. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.924941

Batz, P., Will, T., Thiel, S., Ziesche, T. M. y Joachim, C. (2023). From identification to forecasting: The potential of image recognition and artificial intelligence for aphid pest monitoring. *Frontiers in Plant Science*, *14*, 1150748. https://doi.org/10.3389/fpls.2023.1150748

Bhargavi, K. (2020). Deep learning architectures and tools: A comprehensive survey. En *Deep Learning Applications and Intelligent Decision Making in Engineering* (pp. 55-75). IGI Global. https://doi.org/10.4018/978-1-7998-2108-3.ch002

Biswas, K. et al. (2024). Adaptive Smooth Activation Function for Improved Organ Segmentation. *LNCS*.

Bjerge, K., Mann, H. M. R. y Høye, T. T. (2022). Real-time insect tracking and monitoring with computer vision and deep learning. *Remote Sensing in Ecology and Conservation*, *8*(3), 315-327. https://doi.org/10.1002/rse2.245

Bjerge, K., Alison, J., Dyrmann, M., Frigaard, C. E., Mann, H. M. R. y Høye, T. T. (2023). Accurate detection and identification of insects from camera trap images with deep learning. *PLOS Sustainability and Transformation*, *2*. https://doi.org/10.1371/journal.pstr.0000051

Bjerge, K., Frigaard, C. E. y Karstoft, H. (2023). Object Detection of Small Insects in Time-Lapse Camera Recordings. *Sensors*, *23*. https://doi.org/10.3390/s23167242

Bjerge, K., Karstoft, H., Mann, H. M. R. y Høye, T. T. (2024). A deep learning pipeline for time-lapse camera monitoring of insects and their floral environments. *Ecological Informatics*, *84*. https://doi.org/10.1016/j.ecoinf.2024.102861

Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G. y White, A. E. (2022). Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, *13*, 1640-1660. https://doi.org/10.1111/2041-210X.13901

Bożek, K., Hebert, L., Portugal, Y. y Stephens, G. J. (2021). Markerless tracking of an entire honey bee colony. *Nature Communications*, *12*. https://doi.org/10.1038/s41467-021-21769-1

Braga, D., Madureira, A. M., Scotti, F., Piuri, V. y Ajith, A. P. (2021). An Intelligent Monitoring System for Assessing Bee Hive Health. *IEEE Access*, *9*, 89009-89019. https://doi.org/10.1109/ACCESS.2021.3089538

Burri, S., Ahuja, S., Kumar, A. y Baliyan, A. (2023). Exploring the Effectiveness of Optimized Convolutional Neural Network in Transfer Learning for Image Classification: A Practical Approach. *2023 International Conference on Advancement in Computation and Computer Technologies (InCACCT).*

Calisan, M. y Olgun, N. (2025). Evaluation of Classification Performance of Honey Bee Species With Deep Learning Models. *Electrica*, *25*. https://doi.org/10.5152/electrica.2025.24157

Chandraprakash, D., Pallavi, D. R., Ramachandran, M., Sivaji, C. y Amudha, M. (2024). Convolutional Neural Networks for Image Classification Using the SPSS Method. *Lecture Notes in Networks and Systems*, *1167*, 405-416. https://doi.org/10.1007/978-3-031-73318-5_42

Chaturvedi, A., Apoorva, N., Awasthi, M. S. et al. (2023). Analyzing the Performance of Novel Activation Functions on Deep Learning Architectures. *Lecture Notes in Electrical Engineering*, *928*, 903-915. https://doi.org/10.1007/978-981-19-5482-5_76

Chen, J.-J., Husnain, A. y Cheng, W.-W. (2024). Exploring the Trade-Off Between Performance and Cost in Facial Recognition: Deep Learning Versus Traditional Computer Vision. *Lecture Notes in Networks and Systems*, *823*, 400-412. https://doi.org/10.1007/978-3-031-47724-9_27

Chen, S., Gao, Q. y He, Y. (2024). A pest image recognition method for long-tail distribution problem. *Frontiers in Environmental Science*, *12*, 1391770. https://doi.org/10.3389/fenvs.2024.1391770

Chiranjeevi, S., Saadati, M., Deng, Z., Koushik, J., Jubery, T. Z. N., Mueller, D. S., O'Neal, M. E., Merchant, N. C., Singh, A., Singh, A. K., Sarkar, S. H., Singh, A. y Ganapathysubramanian, B. (2025). InsectNet: Real-time identification of insects using an end-to-end machine learning pipeline. *PNAS Nexus*, *4*. https://doi.org/10.1093/pnasnexus/pgae575

Chung, J., Kim, D., Kang, T. y Lim, M. (2020). Traffic Sign Recognition in Harsh Environment Using Attention Based Convolutional Pooling Neural Network. *Neural Processing Letters*, *51*(3), 2551-2573. https://doi.org/10.1007/s11063-020-10211-0

Copeland, D. C., Mott, B. M., Kortenkamp, O. L., Erickson, R. J., Allen, N. O. y Anderson, K. E. (2025). Image-based honey bee larval viral and bacterial diagnosis using machine learning. *Scientific Reports*, *15*. https://doi.org/10.1038/s41598-025-16261-5

Dadgostar, S., Mobini Dehkordi, P., Khodarahmi, S. y Ghorbani, M. M. (2025). Automated Honey Bee Subspecies Identification Using Advanced Wing Venation Analysis and Adaptive Hierarchical Clustering. *Ecology and Evolution*, *15*. https://doi.org/10.1002/ece3.72101

Dembski, J. y Szymański, J. (2020). Weighted clustering for bees detection on video images. *Lecture Notes in Computer Science*, *12141 LNCS*, 453-466. https://doi.org/10.1007/978-3-030-50426-7_34

Dubey, S., Chakraborty, S., Roy, S., Mukherjee, S., Singh, S. y Chaudhuri, B. (2020). DiffGrad: An Optimization Method for Convolutional Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(11), 4500-4511. https://doi.org/10.1109/TNNLS.2019.2955777

Dubey, S., Kumar Singh, S. y Chaudhuri, B. (2023). AdaNorm: Adaptive Gradient Norm Correction based Optimizer for CNNs. *2023 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 5273-5282. https://doi.org/10.1109/WACV56688.2023.00525

Edwin Selva Rex, C. R., Annrose, J. y Jenifer Jose, J. (2022). Comparative analysis of deep convolution neural network models on small scale datasets. *Optik*, *271*, 170238. https://doi.org/10.1016/j.ijleo.2022.170238

EL ICHI, A., KADDAH, W., EL BOUZ, M. y BADOC, I. (2025). TER-FCN: Tensor Einstein Reduced Fully Connected Network applied to Region-based Convolutional Neural Networks for object detection. *Signal, Image and Video Processing*, *19*(8), 621. https://doi.org/10.1007/s11760-025-04283-5

Farag, M. M. (2023). Design and Analysis of Convolutional Neural Layers: A Signal Processing Perspective. *IEEE Access*.

Feng, D., Zhuang, X., Chen, Z., Zhong, S., Qi, Y., Chen, H. y Ma, H. (2024). Position information encoding FPN for small object detection in aerial images. *Neural Computing and Applications*, *36*(26), 16023-16035. https://doi.org/10.1007/s00521-024-09917-2

Fu, Y. et al. (2021). Structure-aware Loss Function for Ultrasound Image Segmentation. *IEEE IUS*.

García, C. A. Y., Rodrigues, P. J., Tofilski, A., Elen, D., McCormak, G. P., Oleksa, A., Henriques, D., Ilyasov, R. A., Kartashev, A., Bargain, C., Fried, B. y Pinto, M. A. (2022). Using the Software DeepWings© to Classify Honey Bees across Europe through Wing Geometric Morphometrics. *Insects*, *13*. https://doi.org/10.3390/insects13121132

Gernat, T., Jagla, T., Jones, B. M., Middendorf, M. y Robinson, G. E. (2023). Automated monitoring of honey bees with barcodes and artificial intelligence reveals two distinct social networks from a single affiliative behavior. *Scientific Reports*, *13*. https://doi.org/10.1038/s41598-022-26825-4

Ghosh, S. y Hasan, M. N. (2025). Plant-pollinator interaction. En *Pollinators: Life, Biology and Threats* (pp. 63-73). Springer. https://doi.org/10.1007/978-981-95-0359-9_4

Gomez-Jaime, A., Meyers, L., Rodríguez-Cordero, J. A., Agosto-Rivera, J. L., Giray, T. y Mégret, R. (2025). Paint Blob Detection and Decoding for Identification of Honey Bees. *Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, *2*, 242-250. https://doi.org/10.5220/0013385800003912

Gupta, S., Kishan, B. y Mittal, P. (2025). Evaluating the efficacy of convolutional neural networks across diverse datasets. *Proceedings of the Pakistan Academy of Sciences: Part A*, *62*(1), 53-65. https://doi.org/10.53560/PPASA(62-1)681

Hao, K., Lin, S., Qiao, J. y Tu, Y. (2021). A Generalized Pooling for Brain Tumor Segmentation. *IEEE Access*, *9*, 159283-159290. https://doi.org/10.1109/ACCESS.2021.3130035

Hassanie, S., Gohar, A., Hashim Ali, R. y Faiz, S. (2025). A Scalable AI Approach to Bird Species Identification for Conservation and Ecological Planning. *IEEE Access*.

Hellwig, N., Sommerlandt, F. M. J., Graben, S., Lindemann, L., Sickel, W., Krüger, L. y Dieker, P. (2024). Six steps towards a spatial design for large-scale pollinator surveillance monitoring. *Insects*, *15*(4), 229. https://doi.org/10.3390/insects15040229

Herath, H. M. N. (2025). Evolution and advancements from neural network to deep learning. *ASEAN Journal of Educational Research and Technology*, *4*(1), 59-80. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105011167851

Ho, C. H., Gilchrist, I. R., Amponsah, K. A., Stoltz, A., Harpur, B. A., Cantley, S. y Reibman, A. R. (2024). Motion-based domain randomization for detecting honey bees inside a hive. *IS and T International Symposium on Electronic Imaging Science and Technology*, *36*. https://doi.org/10.2352/EI.2024.36.8.IMAGE-240

Høye, T. T., Arje, J., Bjerge, K., Hansen, O. L. P., Iosifidis, A., Leese, F., Mann, H. M. R., Meissner, K., Melvad, C. y Raitoharju, J. (2021). Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences*, *118*(2), e2002545117. https://doi.org/10.1073/pnas.2002545117

Huang, Y., Lu, L., Shen, H., Wang, F. y Qiao, X. (2024). Research on invading insect recognition based on convolutional neural network. *Journal of Chinese Agricultural Mechanization*, *45*(7), 222-227, 260. https://doi.org/10.13733/j.jcam.issn.2095-5553.2024.07.033

Illarionova, S., Nesteruk, S., Shadrin, D., Ignatiev, V., Pukalchik, M. y Oseledets, I. (2021). Object-Based Augmentation for Building Semantic Segmentation: Ventura and Santa Rosa Case Study. *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, 1659-1668. https://doi.org/10.1109/ICCVW54120.2021.00191

Indu, V. T. y Priyadharsini, S. S. (2023). Optimized deep feature selective concatenation of recurrent encoder model and pretrained convolutional neural network model for leaf disease identification. *Journal of Electronic Imaging*, *32*(1), 013046. https://doi.org/10.1117/1.JEI.32.1.013046

Ito, T. et al. (2023). Analyzing the Accuracy, Representations, and Explainability of Various Loss Functions. *IJCNN*.

Jagtap, V. y Yadav, R. (2025). Transfer Learning for Enhancing Computer Vision. En *Lecture Notes in Networks and Systems*.

Jian, W., Xingjian, L. y Gang, L. (2023). Image recognition of rice diseases and insect pests based on deep residual network and transfer learning. *Journal of Chinese Agricultural Mechanization*.

Joselin, J., Sofia, V., Dhanalakshmi, S. e Indhumathi, S. (2024). Deep Learning for Designing Heuristic Methods for Healthcare Data Analytics. En *Deep Learning for Smart Healthcare: Trends, Challenges and Applications* (pp. 90-106). https://doi.org/10.1201/9781003469605-6

Ju, R.-Y., Chiang, J.-S., Chen, C.-C. y Lin, Y.-S. (2022). Connection Reduction of DenseNet for Image Recognition. *2022 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. https://doi.org/10.1109/ISPACS57703.2022.10082842

Kam Ho, T. (2022). Complexity of Representations in Deep Learning. *Proceedings - International Conference on Pattern Recognition*, 2657-2663. https://doi.org/10.1109/ICPR56361.2022.9956594

Kaur, M., Ardekani, I., Sharifzadeh, H. y Varastehpour, S. (2022). A CNN-Based Identification of Honeybees' Infection using Augmentation. *International Conference on Electrical,*

Computer, Communications and Mechatronics Engineering (ICECCME). https://doi.org/10.1109/ICECCME55909.2022.9988275

Kaur, P., Hamal, S., Gautam, V., Singh, M. P. y Singh, S. P. (2024). Hybrid deep learning model for multi-bird lesions detection in Solanum lycopersicum leaves. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-15940-7

Kaur, S. y Agrawal, R. (2021). Convolutional networks. En *Deep Learning Approaches to Cloud Security*.

Kelley, W., Valova, I., Bell, D., Ameh, O. y Bader, J. (2021). Honey sources: Neural network approach to bee species classification. *Procedia Computer Science*, *192*, 650-657. https://doi.org/10.1016/j.procs.2021.08.067

Khalid, M. et al. (2020). Empirical Evaluation of Activation Functions in Deep Convolution Neural Network. *TSP 2020*.

Khan, A., Sohail, A., Zahoora, U. y Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*(8), 5455-5516. https://doi.org/10.1007/s10462-020-09825-6

Khanzhina, N., Filchenkov, A., Minaeva, N., Novoselova, L., Petukhov, M., Kharisova, I., Pinaeva, J., Zamarin, G., Putin, E., Zamyatina, E. y Shalytko, A. (2022). Combating data incompetence in pollen images detection and classification for pollinosis prevention. *Computers in Biology and Medicine*, *140*, 105064. https://doi.org/10.1016/j.compbiomed.2021.105064

Khastavaneh, H. y Ebrahimpour-Komleh, H. (2020). Representation Learning Techniques: An Overview. *Lecture Notes on Data Engineering and Communications Technologies*, *45*, 89-104. https://doi.org/10.1007/978-3-030-37309-2_8

Kiliçarslan, S. (2023). A novel nonlinear hybrid HardSReLUE activation function. *Multimedia Tools and Applications*.

Kim, B., Pak, J. y Son, H. (2025). Wasp-hive candidate site search system using a small drone. *Entomological Research*, *55*(3), e70034. https://doi.org/10.1111/1748-5967.70034

Klasen, M., Ahrens, D., Eberle, J. y Steinhage, V. (2022). Image-Based Automated Species Identification: Can Virtual Data Augmentation Overcome Problems of Insufficient Sampling? *Systematic Biology*, *71*, 320-333. https://doi.org/10.1093/sysbio/syab048

Kongsilp, P., Taetragool, U. y Duangphakdee, O. (2024). Individual honey bee tracking in a beehive environment using deep learning and Kalman filter. *Scientific Reports*, *14*. https://doi.org/10.1038/s41598-023-44718-y

Kozlova, O. y Kitov, V. (2024). Automatic tuning of the pooling operation in convolutional neural networks for coal rock classification. *Gornaya Promyshlennost*, (6), 125-134. https://doi.org/10.30686/1609-9192-2024-6-125-134

Kriouile, Y., Ancourt, C., Wgrzyn-Wolska, K. M. y Bougueroua, L. (2024). Nested object detection using mask R-CNN: application to bee and varroa detection. *Neural Computing and Applications*, *36*, 22587-22609. https://doi.org/10.1007/s00521-024-10393-x

Kukkadapu, S. (2025). Image Recognition and Extraction of Deep Features Using GoogLeNet. *2025 8th International Symposium on Big Data and Applied Statistics (ISBDAS)*, 346-353. https://doi.org/10.1109/ISBDAS64762.2025.11116982

Kulyukin, V. A. y Kulyukin, A. V. (2023). Accuracy vs. Energy: An Assessment of Bee Object Inference in Videos from On-Hive Video Loggers with YOLOv3, YOLOv4-Tiny, and YOLOv7-Tiny. *Sensors*, *23*. https://doi.org/10.3390/s23156791

Lee, H., Shin, J.-y., Kim, S., Kim, M., Kim, M.-s., Lee, H. y Mo, C. (2025). Enhancing Bee Mite Detection with YOLO: The Role of Data Augmentation and Stratified Sampling. *Agriculture (Switzerland)*, *15*. https://doi.org/10.3390/agriculture15111221

Lei, C., Lu, Y., Xing, Z., Zhang, J., Li, S., Wu, W. y Liu, S. (2024). A Honey Bee In-and-Out Counting Method Based on Multiple Object Tracking Algorithm. *Insects*, *15*. https://doi.org/10.3390/insects15120974

Leocádio, R. R. V., Rêgo Segundo, A. K. y Pessin, G. (2024). A Brazilian native bee (Tetragonisca angustula) dataset for computer vision. *Data in Brief*, *55*. https://doi.org/10.1016/j.dib.2024.110659

Liang, A. (2024). Developing a Multimodal System for Bee Object Detection and Health Assessment. *IEEE Access*, *12*, 158703-158713. https://doi.org/10.1109/ACCESS.2024.3464559

Liawatimena, S., Abdurahman, E., Trisetyarso, A., Wibowo, A., Edbert, I., Ario, M. y Effendi, F. (2022). Performance Optimization of Maxpool Calculation Using 4D Rank Tensor. *ICIC Express Letters, Part B: Applications*, *13*(7), 767-776. https://doi.org/10.24507/icicelb.13.07.767

Linse, C., Barth, E. y Martinetz, T. (2023). Convolutional Neural Networks Do Work with Pre-Defined Filters. *International Joint Conference on Neural Networks*.

Liu, Y., Yang, S. y Bao, Y. (2024). Deep Neural Network Optimization Based on Binary Method for Handling Multi-Class Problems. *IEEE Access*, *12*, 46881-46890. https://doi.org/10.1109/ACCESS.2024.3382195

Liu, Z., Zhang, J., He, Q. y Wang, C. (2024). Understanding data augmentation from a robustness perspective. *ICASSP 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6760-6764. https://doi.org/10.1109/ICASSP48485.2024.10447081

Loganathan, R. y Latha, S. (2025). The Empirical Comparison of Deep Neural Network Optimizers for Binary Classification of OCT Images. *Proceedings on Engineering Sciences*, *7*(1), 547-554. https://doi.org/10.24874/PES07.01D.011

Losiński, J. y Czarnowski, I. (2025). Comparison of CNN architectures for INTERCO. *Procedia Computer Science*, *270*, 861-870. https://doi.org/10.1016/j.procs.2025.09.206

Ma, T., Dalca, A. V. y Sabuncu, M. R. (2022). Hyper-Convolution Networks for Biomedical Image Segmentation. *WACV*.

Madaan, V. y Sharma, N. (2024). Satellite Image Classification Using Convolutional Neural Networks. *2024 3rd International Conference for Advancement in Technology (ICONAT)*. https://doi.org/10.1109/ICONAT61936.2024.10774671

Mahmoudi, M. A., Chetouani, A., Boufera, F. y Tabia, H. (2020). Kernelized Dense Layers for Facial Expression Recognition. *Proceedings - International Conference on Image Processing (ICIP)*, 2226-2230. https://doi.org/10.1109/ICIP40778.2020.9190694

Mambile, C., Leo, J. y Kaijage, S. (2025). Comparative analysis of CNN architectures for satellite-based forest fire detection: A mobile-friendly approach using Sentinel-2 imagery. *Remote Sensing Applications: Society and Environment*, *40*, 101739. https://doi.org/10.1016/j.rsase.2025.101739

Mienie, I. D. y Svart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information*, *15*(12), 755. https://doi.org/10.3390/info15120755

Minervini, M., Hausman, S. y Frosini, L. (2021). Transfer learning technique for automatic bearing fault diagnosis in induction motors. *2021 IEEE 13th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*, 186-192. https://doi.org/10.1109/SDEMPED51010.2021.9605514

Nguyen, D. T., Le, T. N., Phung, T. H., Nguyen, D. M., Nguyen, H. Q., Pham, H. T., Phan, T. T. H., Vu, H. y Le, T. L. (2024). Improving pollen-bearing honey bee detection from videos captured at hive entrance by combining deep learning and handling imbalance techniques. *Ecological Informatics*, *82*. https://doi.org/10.1016/j.ecoinf.2024.102744

Ning, C., Gan, H., Shen, M. y Zhang, T. (2023). Learning-based padding: From connectivity on data borders to data padding. *Engineering Applications of Artificial Intelligence*, *121*, 106048. https://doi.org/10.1016/j.engappai.2023.106048

Nirthika, R., Manivannan, S., Ramanan, A. y Wang, R. (2022). Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Computing and Applications*, *34*(7), 5321-5347. https://doi.org/10.1007/s00521-022-06953-8

Nizam, A., Mohd-Isa, W. N. y Ali, A. B. (2020). Image segmentation of meliponine bee using Mask-RCNN. *International Journal of Engineering Trends and Technology*, 17-21. https://doi.org/10.14445/22315381/CATI2P203

Oliveira, T. M., Monteiro, J. M., Franco, J. W. y Machado, J. (2025). Influence of quantization of convolutional neural networks on image classification of pollen-bearing bees. *Proceedings of the International Conference on Enterprise Information Systems*, *1*, 147-157. https://doi.org/10.5220/0013475700003929

Ostwald, M. M., Chen, K., Alexander, N., Ding, L., González, V. H. y Seltmann, K. C. (2025). Climate explains global functional trait variation in bees. *Functional Ecology*, *39*, 1748-1760. https://doi.org/10.1111/1365-2435.70051

Pacal, I., Kunduracıoğlu, I., Alma, M. H., Deveci, M., Kadry, S., Nedoma, J., Slaný, V. y Martínek, R. (2024). A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review*, *57*(11), 304. https://doi.org/10.1007/s10462-024-10944-7

Padubidri, C. G., Kamilaris, A., Charalambous, A., Lanitis, A. y Constantinides, M. (2024). The Be-Hive Project—Counting Bee Traffic Based on Deep Learning and Pose Estimation. *Lecture Notes in Networks and Systems*, *823 LNNS*, 531-545. https://doi.org/10.1007/978-3-031-47724-9_35

Parvin, F. y Mehedi Hasan, M. A. (2020). A comparative study of different types of convolutional neural networks for breast cancer histopathological image classification. *2020 IEEE Region 10 Symposium (TENSYMP)*, 945-948. https://doi.org/10.1109/TENSYMP50017.2020.9230787

Pattanaik, Y., Pattanayak, P., Dash, A., Darshana, S., Pani, S., Biswal, B., Mohanty, S. y Balabantray, A. (2025). Exploring Optimization Dynamics: Hybrid Approaches Combining Adaptive and Traditional Techniques for Deep Learning Models. *4th IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*. https://doi.org/10.1109/ICDCECE65353.2025.11035942

Postalcloğlu, S. (2020). Performance Analysis of Different Optimizers for Deep Learning-Based Image Recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, *34*(2), 2051003. https://doi.org/10.1142/S0218001420510039

Proença, M. D. C., Rebelo, M. T., Valent, R., Mateus, R., Gaspar, P. D., Lopes, C. M. y Franco, J. C. (2025). Identification of green leafhoppers (Cicadellidae) in vineyards through an automatic image acquisition system from yellow sticky traps associated with deep-learning. *Ciencia e Técnica Vitivinícola*, *40*(1), 1-9. https://doi.org/10.1051/ctv/ctv202400101

Qian, S., Du, J., Zhou, J., Xie, C., Jiao, L. y Li, R. (2023). An effective pest detection method with automatic data augmentation strategy in the agricultural field. *Signal, Image and Video Processing*, *17*(2), 563-571. https://doi.org/10.1007/s11760-022-02261-9

Raheesh, R., Sajith Variyar, V. V., Sivanpillai, R., Sowmya, V., Soman, K. P. y Brown, G. K. (2022). Segmentation of epiphytes in grayscale images using a CNN-transformer hybrid architecture. *Lecture Notes in Networks and Systems*, *446*, 119-129. https://doi.org/10.1007/978-981-19-1559-8_13

Raj, R. y Yos, A. (2025). An extensive study of convolutional neural networks: Applications in computer vision for improved robotics perceptions. *Sensors*, *25*(4), 1033. https://doi.org/10.3390/s25041033

Rajasree, R., Geetha, R., Vidhya, S., Mukunthan, M. A., Kumar, S. H. y Jose Anand, A. (2025). Comparative Analysis of Feature Engineering-Based Learning and Deep Learning Approaches for Classification Problems on Imbalanced Datasets. *2025 4th OPJU International Technology Conference on Smart Computing for Innovation and Advancement in Industry 5.0 (OTCON)*. https://doi.org/10.1109/OTCON65728.2025.11070444

Ratnayake, M. N., Dyer, A. G. y Dorin, A. (2021). Towards computer vision and deep learning facilitated pollination monitoring for agriculture. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2915-2924. https://doi.org/10.1109/CVPRW53098.2021.00327

Rodrigues, P. J., Gomes, W. y Pinto, M. A. (2022). DeepWings©: Automatic Wing Geometric Morphometrics Classification of Honey Bee (Apis mellifera) Subspecies Using Deep Learning for Detecting Landmarks. *Big Data and Cognitive Computing*, *6*. https://doi.org/10.3390/bdcc6030070

Rohafauzi, S., Kassim, M. B., Ja'afar, H. B., Rustam, I. y Miskon, M. T. (2024). A review on internet of things-based stingless bee's honey production with image detection framework. *International Journal of Electrical and Computer Engineering*, *14*, 2282-2292. https://doi.org/10.11591/ijece.v14i2.pp2282-2292

Rozenbaum, E., Shrot, T., Daltrophe, H., Kunya, Y. y Shafir, S. H. (2024). Machine learning-based bee recognition and tracking for advancing insect behavior research. *Artificial Intelligence Review*, *57*. https://doi.org/10.1007/s10462-024-10879-z

Rudenko, O., Berzonov, O. y Uylink, K. (2020). First-order optimization (training) algorithms in deep learning. *CEUR Workshop Proceedings*, *2604*, 921-935. https://www.scopus.com/inward/record.uri?eid=2-s2.0-850851731868

Sa'doun, M., Dalton, D., Berger, V., Paulus, G., Karl-Heinrich, A., Svetnik, I., Schultz, J. y Unglaub, P. (2024). High resolution environmental monitoring of pollinator insects through macro camera trapping and AI. *CEUR Workshop Proceedings*, *3786*, 39-45. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85207518767

Saetchnikov, I. V., Tchernyavskaia, E. A. y Skakun, V. V. (2021). Object detection for unmanned aerial vehicle camera via convolutional neural networks. *IEEE Journal on Miniaturization for Air and Space Systems*, *2*(2), 98-103. https://doi.org/10.1109/JMASS.2020.3040976

Sajib, M. T. A., Badhon, N. H., Salehin, I., Rifat, M. S. H., Ahmed, F., Saha, P. y Moon, N. N. (2025). A comparative deep learning methodology for plant insect image classification: Assessment of CNN architectures and augmentation techniques. *MethodsX*, *15*, 103680. https://doi.org/10.1016/j.mex.2025.103680

Saleem, M. A., Senan, N., Wahid, F., Aamir, M., Samad, A. y Khan, M. (2022). Comparative Analysis of Recent Architecture of Convolutional Neural Network. *Mathematical Problems in Engineering*, *2022*, 7313612. https://doi.org/10.1155/2022/7313612

Santiago-Plaza, G. A., Meyers, L., Gomez-Jaime, A., Meléndez-Ríos, R., Noel, F., Agosto, J. L., Giray, T., Rodríguez-Cordero, J. A. y Mégret, R. (2024). Identification of Honeybees with Paint Codes Using Convolutional Neural Networks. *Proceedings of the International*

Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, *2*, 772-779. https://doi.org/10.5220/0012460600003660

Shadoul, I., Al-Hmouz, R., Hossen, A., Mesbah, M. y Deveci, M. (2025). The effect of pooling parameters on the performance of convolution neural network. *Artificial Intelligence Review*, *58*(9), 271. https://doi.org/10.1007/s10462-025-11273-z

Shahnaz, M. y Mollah, A. F. (2023). On the Performance of Convolutional Neural Networks with Resizing and Padding. *Lecture Notes in Networks and Systems*, *404*, 51-62. https://doi.org/10.1007/978-981-19-0105-8_6

Sharma, A., Varastehpour, S., Ardekani, I. y Sharifzadeh, H. (2024). Bee disease Varroa prediction: Utilizing convolutional neural networks with augmentation for robust detection and identification of honeybee infection. *1st International Conference on Innovative Engineering Sciences and Technological Research (ICIESTR)*. https://doi.org/10.1109/ICIESTR60916.2024.10798343

Sharma, D. (2020). Deep learning without tears: A simple introduction. *Resonance*, *25*(1), 15-32. https://doi.org/10.1007/s12045-019-0919-9

Sledevič, T. y Matuzevičius, D. (2024). Labeled dataset for bee detection and direction estimation on entrance to beehive. *Data in Brief*, *52*. https://doi.org/10.1016/j.dib.2024.110060

Sledevič, T., Serackis, A., Matuzevičius, D., Plonis, D. y Andriukaitis, D. (2024). Keypoint-Based Bee Orientation Estimation and Ramp Detection at the Hive Entrance for Bee Behavior Identification System. *Agriculture (Switzerland)*, *14*. https://doi.org/10.3390/agriculture14111890

Sledevič, T., Serackis, A., Matuzevičius, D., Plonis, D. y Vdoviak, G. (2025). Visual recognition of honeybee behavior patterns at the hive entrance. *PLOS ONE*, *20*. https://doi.org/10.1371/journal.pone.0318401

Spiesman, B. J., Gratton, C., Hatfield, R. G., Hsu, W. H., Jepsen, S., McCormack, B., Patel, K. y Wang, G. (2021). Assessing the potential for deep learning and computer vision to identify bumble bee species from images. *Scientific Reports*, *11*, 7580. https://doi.org/10.1038/s41598-021-87210-1

Spiesman, B. J., Gratton, C., Gratton, E. M. e Hines, H. M. (2024). Deep learning for identifying bee species from images of wings and pinned specimens. *PLOS ONE*, *19*. https://doi.org/10.1371/journal.pone.0303383

Srinivas, C. L., Sreerag, M., Mukherjee, R., Devy, S. M., Mishra, S. y Deb, R. (2025). Image-based South Asian bee species identification: a machine learning approach. *Journal of Insect Conservation*, *29*. https://doi.org/10.1007/s10841-025-00691-7

Srinivas, Y. y Kumar, M. (2025). Exploring Prominent Convolutional Neural Network Frameworks to Identify COVID-19 Deceases by Using Medical Images. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, *597*, 184-203. https://doi.org/10.1007/978-3-031-77075-3_16

Štaka, Z., Mišić, M. y Tomašević, M. (2025). CPU vs. GPU: Performance Evaluation of Classical Machine and Deep Learning Algorithms. *2025 24th International Symposium INFOTEH-JAHORINA (INFOTEH)*. https://doi.org/10.1109/INFOTEH64129.2025.10959248

Stark, T., Štefan, V., Wurm, M., Spanier, R., Taubenböck, H. y Knight, T. M. (2023). YOLO object detection models can locate and classify broad groups of flower-visiting arthropods in images. *Scientific Reports*, *13*, 16364. https://doi.org/10.1038/s41598-023-43482-3

Stark, T., Wurm, M., Štefan, V., Wolf, F., Taubenböck, H. y Knight, T. M. (2025). Utilizing CNNs for classification and uncertainty quantification for 15 families of European fly pollinators. *PLOS ONE*, *20*(9), e0323984. https://doi.org/10.1371/journal.pone.0323984

Ștefan, V., Stark, T., Wurm, M., Taubenböck, H. y Knight, T. M. (2025). Successes and limitations of pretrained YOLO detectors applied to unseen time-lapse images for automated pollinator monitoring. *Scientific Reports*, *15*. https://doi.org/10.1038/s41598-025-16140-z

Stiemer, L. N., Thoma, A. y Braun, C. (2023). MBT3D: Deep learning based multi-object tracker for bumblebee 3D flight path estimation. *PLOS ONE*, *18*. https://doi.org/10.1371/journal.pone.0291415

Sudewo, E., Kunta Biddinika, M., Umar, R. y Fadlil, A. (2025). Evaluating the Impact of Optimizer Hyperparameters on ResNet in Hanacaraka Character Recognition. *Preservation, Digital Technology and Culture*, *54*(2), 113-123. https://doi.org/10.1515/pdtc-2024-0061

Suzuki-Ohno, Y., Westfechtel, T., Yokoyama, J., Ohno, K., Nakashizuka, T., Kawata, M. y Okatani, T. (2022). Deep learning increases the availability of organism photographs taken by citizens in citizen science programs. *Scientific Reports*, *12*. https://doi.org/10.1038/s41598-022-05163-5

Takimoto, H., Sato, Y., Nagano, A., Shimizu, K. y Kanagawa, A. (2021). Using a two-stage convolutional neural network to rapidly identify tiny herbivorous beetles in the field. *Ecological Informatics*, *66*, 101466. https://doi.org/10.1016/j.ecoinf.2021.101466

Taniguchi, H., Tsukuda, Y., Motoki, K., Goto, T., Yoshida, Y. y Yasuba, K. I. (2025). Development of an AI-based Image Analysis System to Calculate the Visit Duration of a Green Blow Fly on a Strawberry Flower. *Horticulture Journal*, *94*, 64-72. https://doi.org/10.2503/hortj.QH-159

Tian, Y., Su, D., Lauria, S. y Liu, X. (2022). Recent advances on loss functions in deep learning. *Neurocomputing*.

Too, E. C., Yujian, L., Gadosey, P. et al. (2020). Performance analysis of nonlinear activation function in convolution neural network for image classification. *International Journal of Computational Science and Engineering*.

Tseng, H.-H., Yang, M.-D., Saminathan, R. y Wu, D.-H. (2022). Rice Seedling Detection in UAV Images Using Transfer Learning and Machine Learning. *Remote Sensing*.

Urbano Espinosa, A. F., Vásquez Valencia, C. I. y Solis Pino, A. F. (2025). An embedded deep learning system for real-time detection of pollen-bearing bees in Colombian agroecological contexts. *2025 IEEE Colombian Conference on Communications and Computing*. https://doi.org/10.1109/COLCOM66267.2025.11185800

Varadarajan, V., Garg, D. y Kotecha, K. (2021). An efficient deep convolutional neural network approach for object detection and recognition using a multi-scale anchor box in real-time. *Future Internet*, *13*(12), 307. https://doi.org/10.3390/fi13120307

Voudiotis, G., Kontogiannis, S. I. y Pikridas, C. (2021). Proposed smart monitoring system for the detection of bee swarming. *Inventions*, *6*. https://doi.org/10.3390/inventions6040087

Voudiotis, G., Moraiti, A. y Kontogiannis, S. I. (2022). Deep Learning Beehive Monitoring System for Early Detection of the Varroa Mite. *Signals*, *3*, 506-523. https://doi.org/10.3390/signals3030030

Wang, J., Zeng, X., Duan, S., Zhou, Q. y Peng, H. (2022). Image target recognition based on improved convolutional neural network. *Mathematical Problems in Engineering*, *2022*, 2213295. https://doi.org/10.1155/2022/2213295

Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., Su, C. y Wei, W. (2025). Advances in deep learning applications for plant disease and pest detection: A review. *Remote Sensing*, *17*(4), 698. https://doi.org/10.3390/rs17040698

Wang, Z., Su, X., Lu, Y. y Liu, L. (2025). Deep learning for insect detection and classification: A review. *Proceedings of SPIE - The International Society for Optical Engineering*, *13793*, 137930X. https://doi.org/10.1117/12.3077589

Waqas, N. et al. (2025). Med-ReLU: A Parameter-Free Hybrid Activation Function. *Computers, Materials and Continua*.

Wittmann, K., Gamal Ibrahim, M., Straw, A. D., Klein, A. M. y Staab, M. (2024). Monitoring fast-moving animals—Building a customized camera system and evaluation toolset. *Methods in Ecology and Evolution*, *15*, 836-842. https://doi.org/10.1111/2041-210X.14322

Wu, C.-Y., Ke, S.-M., Liao, H.-S., Chen, C.-H., Hung, Y.-H. y Lai, S.-C. (2024). Insect Classification Using Spatial Pyramid Pooling and Convolution Neural Network. *2024 IEEE 20th Asia Pacific Conference on Circuits and Systems (APCCAS) and IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics Electronics (PrimeAsia)*, 325-329. https://doi.org/10.1109/APCCAS62602.2024.10808507

Xue, M., Li, J. y Luo, Q. (2022). Toward Optimal Learning Rate Schedule in Scene Classification Network. *IEEE Geoscience and Remote Sensing Letters*, *19*. https://doi.org/10.1109/LGRS.2020.3040359

Yang, M., Zhang, Y. y Liu, T. (2020). Corn disease recognition based on the Convolutional Neural Network with a small sampling size. *Chinese Journal of Eco-Agriculture*.

Ye, Z. y Yu, J. (2021). AKSNet: A novel convolutional neural network with adaptive kernel width. *Journal of Manufacturing Systems*.

Yeung, M. et al. (2022). Unified Focal loss. *Computerized Medical Imaging and Graphics*.

Yi, C. (2022). Application of convolutional networks in clothing design from the perspective of deep learning. *Scientific Programming*, 6173981. https://doi.org/10.1155/2022/6173981

Yoo, J., Siddiqua, R., Liu, X., Ahmed, K. A. y Zakir Hossain, M. Z. (2023). BeeNet: An End-To-End Deep Network For Bee Surveillance. *Procedia Computer Science*, *222*, 415-424. https://doi.org/10.1016/j.procs.2023.08.180

Yu, C., Hung, P.-H., Hong, J.-H. y Chiang, H.-Y. (2023). Efficient Max Pooling Architecture with Zero-Padding for Convolutional Neural Networks. *2023 IEEE 12th Global Conference on Consumer Electronics (GCCE)*, 747-748. https://doi.org/10.1109/GCCE59613.2023.10315268

Zakeri, A., Xia, Y., Ravikumar, N. y Frangi, A. F. (2023). Deep learning for vision and representation learning. En *Medical Image Analysis* (pp. 451-474). https://doi.org/10.1016/B978-0-12-813657-7.00044-3

Zhang, J., Guo, H. y Chen, Z. (2021). A concatenated approach based on transfer learning and PCA for classifying bees and wasps. *Journal of Physics: Conference Series*, *1873*. https://doi.org/10.1088/1742-6596/1873/1/012058

Zhang, Y. et al. (2021). Rethinking the Dice Loss for Deep Learning Lesion Segmentation. *Journal of Shanghai Jiaotong University*.

Zheng, Y., Huang, J., Chen, T. y Zhou, W. (2021). Transfer of Learning in the Convolutional Neural Networks on Classifying Geometric Shapes Based on Local or Global Invariants. *Frontiers in Computational Neuroscience*.

Zhong, S., Wen, W. y Qin, J. (2023). SPEM: Self-adaptive Pooling Enhanced Attention Module for Image Recognition. *Lecture Notes in Computer Science*, *13834*, 41-53. https://doi.org/10.1007/978-3-031-27818-1_4

Zhou, H., Zhou, H. y Dong, J. (2024). Advances and Challenges of Optical Convolution Computation. *Acta Optica Sinica*.

Zhu, J. y Chen, Z. (2020). Comparative analysis of various new activation functions based on convolutional neural network. *Journal of Physics: Conference Series*.

Zhukov, O. e Ivanovych, H. (2023). A Comparative Study of Deep Convolutional Neural Network Architectures to Identify Full Bee Body in Images. *International Conference on Information and Digital Technologies (IDT)*, 21-24. https://doi.org/10.1109/IDT59031.2023.10194440

Zou, G., Luo, Y. y Feng, Z. (2021). Research on the Extraction of Image Edge Information in Convolutional Neural Networks. *Journal of Physics: Conference Series*, *2083*(3), 032015. https://doi.org/10.1088/1742-6596/2083/3/032015

Zulkernine, F., Gasmallah, M., Isah, H., Lam, J., Mahfuz, S. y Khan, S. (2020). A hands-on tutorial on deep learning for object and pattern recognition. *CASCON 2019 Proceedings - Conference of the Centre for Advanced Studies on Collaborative Research*, 386-387. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087416922

Zurowietz, M. y Nattkemper, T. W. (2020). An Interactive Visualization for Feature Localization in Deep Neural Networks. *Frontiers in Artificial Intelligence*, *3*, 49. https://doi.org/10.3389/frai.2020.00049