

The logo consists of the lowercase letters "viu" in white, centered within a solid orange circle.

**viu**

**Universidad  
Internacional  
de Valencia**

# **Desarrollo de un sistema de diagnóstico de enfermedades en hojas de tomate mediante modelos de aprendizaje profundo**

||| Titulación:  
Máster en Big Data y  
Ciencia de Datos  
Curso Académico  
2024-2025

||| Alumno/a: Marín Lucas,  
Rubén  
DNI: 07272889J  
Director/a del TFT:  
Ricardo Lebrón Aguilar

||| Convocatoria:  
**SEGUNDA**

# Índice general

<b>Índice de figuras</b>	<b>3</b>
<b>Índice de cuadros</b>	<b>3</b>
<b>1. Introducción</b>	<b>9</b>
<b>2. Objetivos</b>	<b>14</b>
2.1. Objetivos específicos . . . . .	14
<b>3. Estado del arte</b>	<b>15</b>
<b>4. Implementación y desarrollo</b>	<b>18</b>
4.1. Herramientas usadas . . . . .	18
4.2. Procedencia y descripción de los datos . . . . .	19
4.3. Preprocesado de los datos . . . . .	20
4.4. Modelado . . . . .	24
4.4.1. Modelo <i>MobileNetV2</i> 1 . . . . .	26
4.4.2. Modelo <i>MobileNetV2</i> 2 . . . . .	27
4.4.3. Modelo <i>MobileNetV2</i> 3 . . . . .	28
4.4.4. Modelo <i>EfficientNetB0</i> 1 . . . . .	28
4.4.5. Modelo <i>EfficientNetB0</i> 2 . . . . .	29
4.4.6. Modelo <i>EfficientNetB0</i> 3 . . . . .	30
4.4.7. Modelo <i>NASNetMobile</i> 1 . . . . .	31
4.4.8. Modelo <i>NASNetMobile</i> 2 . . . . .	31
4.5. Arquitecturas . . . . .	32
4.5.1. <i>MobileNetV2</i> . . . . .	32
4.5.2. <i>EfficientNetB0</i> . . . . .	33
4.5.3. <i>NASNetMobile</i> . . . . .	34
<b>5. Evaluación y resultados</b>	<b>36</b>
5.1. Métricas . . . . .	36
5.2. Modelo <i>MobileNetV2</i> 1 . . . . .	37
5.2.1. Entrenamiento 1 . . . . .	37
5.2.2. Entrenamiento 2 . . . . .	40
5.2.3. Entrenamiento 3 . . . . .	43
5.3. Modelo <i>MobileNetV2</i> 2 . . . . .	46
5.4. Modelo <i>MobileNetV2</i> 3 . . . . .	48
5.5. Modelo <i>EfficientNetB0</i> 1 . . . . .	52

5.6. Modelo <i>EfficientNetB0</i> 2 . . . . .	53
5.7. Modelo <i>EfficientNetB0</i> 3 . . . . .	54
5.8. Modelo <i>NASNetMobile</i> 1 . . . . .	55
5.9. Modelo <i>NASNetMobile</i> 2 . . . . .	57
5.10. Comparativa de modelos . . . . .	60
<b>6. Conclusiones</b>	<b>62</b>
<b>A. Anexo: Código fuente</b>	<b>63</b>

# Índice de figuras

1.1. <a href="#">Origen del tomate. Fuente: (Silva, s.f.)</a> . . . . .	9
1.2. <a href="#">Tizón tardío en una planta de tomate (Robinson et al., 2013)</a> . . . . .	10
1.3. <a href="#">Tizón precoz en una planta de tomate (Renaissance Farms Editors, s.f.)</a> . .	11
1.4. <a href="#">Mancha bacteriana en una planta de tomate (P. Keinath, 2023)</a> . . . . .	11
1.5. <a href="#">Picadura de la araña roja de dos manchas en una planta de tomate (Inspired Pencil editors, 2021)</a>	
1.6. <a href="#">Virus de la hoja amarilla en una planta de tomate (Greenlife editors, 2025)</a>	12
4.1. <a href="#">Imagen aleatoria por clase</a> . . . . .	21
4.2. <a href="#">Resumen del modelo <i>MobileNetV2 1</i></a> . . . . .	26
4.3. <a href="#">Resumen del modelo <i>MobileNetV2 2</i></a> . . . . .	27
4.4. <a href="#">Resumen del modelo <i>MobileNetV2 3</i></a> . . . . .	28
4.5. <a href="#">Resumen del modelo <i>EfficientNetB0 1</i></a> . . . . .	29
4.6. <a href="#">Resumen del modelo <i>EfficientNetB0 2</i></a> . . . . .	29
4.7. <a href="#">Resumen del modelo <i>EfficientNetB0 3</i></a> . . . . .	30
4.8. <a href="#">Resumen del modelo <i>NASNetMobile 1</i></a> . . . . .	31
4.9. <a href="#">Resumen del modelo <i>NASNetMobile 2</i></a> . . . . .	32
5.1. <a href="#">Histórico de métricas del entrenamiento 1 del modelo <i>MobileNetV2 1</i></a> . . .	38
5.2. <a href="#">Resumen de métricas del modelo <i>MobileNetV2 1</i> en el entrenamiento 1</a> . .	39
5.3. <a href="#">Matriz de confusión del modelo <i>MobileNetV2 1</i> en el entrenamiento 1</a> . .	40
5.4. <a href="#">Histórico de métricas del entrenamiento 2 del modelo <i>MobileNetV2 1</i></a> . .	41
5.5. <a href="#">Resumen de métricas del modelo <i>MobileNetV2 1</i> en el entrenamiento 2</a> . .	42
5.6. <a href="#">Matriz de confusión del modelo <i>MobileNetV2 1</i> en el entrenamiento 2</a> . .	43
5.7. <a href="#">Histórico de métricas del entrenamiento 3 del modelo <i>MobileNetV2 1</i></a> . .	44
5.8. <a href="#">Resumen de métricas del modelo <i>MobileNetV2 1</i> en el entrenamiento 3</a> . .	45
5.9. <a href="#">Matriz de confusión del modelo <i>MobileNetV2 1</i> en el entrenamiento 3</a> . .	46
5.10. <a href="#">Histórico de métricas del entrenamiento del modelo <i>MobileNetV2 2</i></a> . .	47
5.11. <a href="#">Resumen de métricas del modelo <i>MobileNetV2 2</i></a> . . . . .	48
5.12. <a href="#">Matriz de confusión del modelo <i>MobileNetV2 2</i></a> . . . . .	49
5.13. <a href="#">Histórico de métricas del entrenamiento del modelo <i>MobileNetV2 3</i></a> . .	50
5.14. <a href="#">Resumen de métricas del modelo <i>MobileNetV2 3</i></a> . . . . .	50
5.15. <a href="#">Matriz de confusión del modelo <i>MobileNetV2 3</i></a> . . . . .	51
5.16. <a href="#">Histórico de métricas del del modelo <i>EfficientNetB0 1</i></a> . . . . .	53
5.17. <a href="#">Histórico de métricas del entrenamiento del modelo <i>EfficientNetB0 2</i></a> . .	54

5.18. Histórico de métricas del entrenamiento del modelo <i>EfficientNetB0</i> 1 . . . . .	55
5.19. Histórico de métricas del entrenamiento del modelo <i>NASNetMobile</i> 1 . . . . .	56
5.20. Histórico de métricas del entrenamiento del modelo <i>NASNetMobile</i> 2 . . . . .	57
5.21. Resumen de métricas del modelo <i>NASNetMobile</i> 2 . . . . .	58
5.22. Matriz de confusión del modelo <i>NASNetMobile</i> 2 . . . . .	59

# Índice de cuadros

1.1. <a href="#"><u>Top 10 países productores de tomates 2022. Fuente: (Wikipedia contributores, s.f.-a)</u></a> . . . . .	10
4.1. <a href="#"><u>Clases del conjunto de datos</u></a> . . . . .	20
4.2. <a href="#"><u>Los 5 tamaños de imágenes más comunes</u></a> . . . . .	20
4.3. <a href="#"><u>Número de imágenes por clases en el conjunto de entrenamiento</u></a> . . . . .	24
4.4. <a href="#"><u>Pesos asignados a cada clase en el entrenamiento del modelo</u></a> . . . . .	24
5.1. <a href="#"><u>Resúmenes de modelos</u></a> . . . . .	60
5.2. <a href="#"><u>Comparativa de modelos con métricas globales</u></a> . . . . .	60

El cultivo de tomate es uno de los cultivos hortícolas más importantes a nivel mundial. Sin embargo, su producción se ve amenazada por una amplia variedad de enfermedades causadas por hongos, bacterias, virus y nematodos. Estas enfermedades pueden provocar una bajada de rendimiento que van desde reducciones parciales hasta la pérdida completa de la cosecha. Dada la magnitud del impacto de estas enfermedades, la detección temprana y precisa de las mismas es crucial. Permite una correcta intervención que minimiza las pérdidas, permitiendo la reducción del uso innecesario de los agroquímicos y mejorando la sostenibilidad.

Los enfoques tradicionales para la detección de enfermedades suelen requerir mucha mano de obra, consumen mucho tiempo y son propensos a errores, lo que dificulta la intervención temprana. En este contexto el objetivo general de este proyecto consiste en desarrollar y evaluar un sistema inteligente basado en técnicas de aprendizaje profundo capaz de identificar el estado de salud y las principales enfermedades en hojas de tomate a partir de imágenes, asegurando un alto grado de precisión y capacidad de generalización.

Entre las tareas de preprocesamiento típicas se han transformado todas las imágenes a tamaños de 224x224 y se han normalizado. Además, se han usado las técnicas de *data augmentation* y fijación de pesos en las clases con el objetivo evitar sesgos hacia ciertas clases y aumentar la capacidad de generalización. Posteriormente para el modelado se han creado modelos de redes neuronales convolucionales basados en las arquitecturas *MobileNetV2*, *EfficientNetB0* y *NASNetMobile* preentrenadas con una base de datos de imágenes muy grande (*ImageNet*), dado que permite aprovechar representaciones visuales previamente aprendidas como bordes, texturas y patrones para iniciar el entrenamiento en un dominio específico con menor cantidad de datos.

Entre los modelos evaluados, *MobileNetV2* con 100 capas descongeladas (*fine tuning*), tasa de aprendizaje (*learning rate*) de  $1e-5$  y entrenado a 20 épocas, mostró el mejor rendimiento, alcanzando una precisión global del 94 % y un *F1-score* macro de 0.939, lo que refleja una elevada capacidad de precisión y una buena generalización entre las distintas enfermedades.

Este resultado confirma la viabilidad del uso de redes neuronales convolucionales para la detección automática de patologías en hojas de tomate, ofreciendo una herramienta potencialmente útil para la monitorización temprana de cultivos y la reducción del uso de productos fitosanitarios, contribuyendo así a una agricultura más precisa y sostenible.

**Palabras clave:** detección de enfermedades, hojas de tomate, redes neuronales convolucionales, *MobileNetV2*, *fine tuning*

## Agradecimientos

Quiero dar las gracias a toda mi familia y amigos por ayudarme y darme fuerzas para seguir adelante en todo momento.

Quiero agradecer a los amigos que he encontrado en el máster, ya que me han ayudado a seguir adelante desde el comienzo. Seguramente alguna asignatura hubiera sido más difícil de superar sin vosotros.

Por último, gracias a todos los profesores que me he encontrado a lo largo de este viaje, ya que me han formado como el futuro ingeniero informático que seré.

# 1. Introducción

Tomate o tomatera (*Solanum lycopersicum*) es una planta herbácea de la familia Solanaceae cultivada en todo el mundo para el cultivo de su fruto, el tomate o jitomate, uno de los ingredientes más universales de ensaladas y salsas en el mundo entero. (Wikipedia contributors, s.f.-b)

Según los últimos estudios filogenéticos, la planta silvestre de la cual surge el tomate doméstico actual tiene origen en la zona andina del norte de Perú y sur de Ecuador. Su domesticación y diversificación posterior se originó en México. Los pueblos aztecas y mayas lo usaban en su cocina y fue exportado al resto del mundo a partir de la llegada de los españoles que lo distribuyeron a lo largo de sus colonias en el Caribe y la península ibérica a partir de lo cual pudo llegar al resto de Europa. También lo llevaron a Filipinas y de allí pudo entrar al continente asiático. (Silva, s.f.)

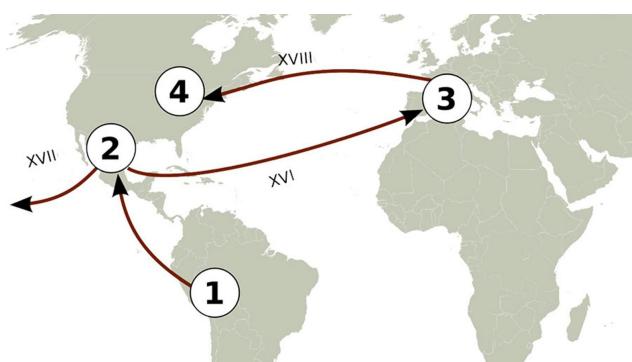


Figura 1.1.: [Origen del tomate](#). Fuente: (Silva, s.f.)

La producción mundial de tomate ascendió a más de 186 millones de toneladas en 2022 según los datos de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). Según esta misma organización la evolución de los 10 países que más han producido hasta 2022 es la mostrada en el Cuadro 1.1 (Wikipedia contributors, s.f.-a).

Como ya se ha mencionado, el cultivo de tomate es uno de los cultivos hortícolas más importantes a nivel mundial. Sin embargo, su producción se ve amenazada por una amplia variedad de enfermedades causadas por hongos, bacterias, virus y nematodos. Estas enfermedades pueden provocar una bajada de rendimiento que van desde reduc-

Cuadro 1.1.: [Top 10 países productores de tomates 2022](#). Fuente: (Wikipedia contributors, s.f.-a)

Top 10 países productores de tomates 2022				
País	2000	2010	2020	2022
China	22 200	46 760	64 680	68 242
India	7430	12 433	20 550	20 694
Turquía	8890	10 052	13 204	13 000
Estados Unidos	12 622	14 053	10 939	10 200
Egipto	6786	8545	6494	6275
Italia	7538	6025	6248	6136
México	2666	2998	4137	4208
Brasil	3005	4107	3757	3810
Nigeria	1261	1800	3390	3685
España	3766	4313	4313	3652

ciones parciales hasta la pérdida completa de la cosecha.

Entre las enfermedades más comunes se encuentran:

- Tizón tardío (*Late blight*): El agente causal es el oomiceto *Phytophthora infestans* que afecta a las plantas solanáceas. Se desarrolla mejor en condiciones de humedad alta, lluvias, rocío persistente o días nublados, y temperaturas frías (entre 10° y 20°). Su ciclo incluye esporangios, zoosporas (en condiciones húmedas) y puede sobrevivir mediante estructuras resistentes (oosporas) o restos vegetales infectados (Editors of Encyclopedia Britannica, 2024). Las primeras lesiones aparecen típicamente en hojas o tallos donde aparecen manchas de aspecto mojado que luego se tornan a manchas oscuras e irregulares. Bajo condiciones muy húmedas se observa una capa blanca o algodonosa de esporulación en el envés de las hojas u otras partes (Gevens & Wilbur, 2014) como se puede observar en la Figura 1.2.



Figura 1.2.: [Tizón tardío en una planta de tomate \(Robinson et al., 2013\)](#)

- Tizón precoz (*Early blight*): El agente causal principal es el hongo *Alternaria solani*.

*Iani* (y en algunos casos *Alternaria tomatophila*) que puede sobrevivir en residuos de plantas enfermas, en el suelo, sobre plantas de la familia de las solanáceas y también transmitirse por semillas o plántulas infectadas (McGrath, s.f.). Las condiciones favorables para su desarrollo incluyen: temperaturas cálidas (por encima de 25°), humedad elevada o periodos de hojas mojadas (Pacific Northwest contributors, 2025b). Los primeros síntomas suelen aparecer en las hojas más viejas, situadas en la base de la planta, aparecen manchas oscuras que a menudo presentan anillos concéntricos. Alrededor de esas manchas la hoja puede ponerse amarilla (clorosis) y luego necrosarse, lo que deriva en caída de hojas o defoliación parcial del vegetal (Baysal-Gurel & Miller, 2010) como puede observarse en la Figura 1.3.



Figura 1.3.: [Tizón precoz en una planta de tomate \(Renaissance Farms Editors, s.f.\)](#)

- Mancha bacteriana (*Bacterial spot*): Es causada por bacterias del género *Xanthomonas* que atacan al tomate. El patógeno puede sobrevivir en residuos vegetales, en semillas contaminadas, en plantas voluntarias o en malas hierbas de la familia de las solanáceas (Pacific Northwest contributors, 2025a). Las condiciones favorables incluyen alta humedad y temperaturas templadas o calientes (20° o más). En las hojas aparecen manchas húmedas, luego se tornan oscuras, ásperas y con margen amarillo en ocasiones. Las manchas suelen aparecer primero en hojas viejas o en el dosel inferior de la planta, desde donde pueden subir (Davis et al., s.f.) como muestra la Figura 1.4.



Figura 1.4.: [Mancha bacteriana en una planta de tomate \(P. Keinath, 2023\)](#)

- Picadura de la araña roja de dos manchas (*Twospotted spider mite*): El agente es la araña roja de dos manchas (*Tetranychus urticae*) un ácaro (ni hongo, ni bacteria) que ataca muchas hortalizas incluyendo el tomate. Se alimenta mediante piezas bucales que perforan las células de las hojas (generalmente en el envés) y succionan el contenido, provocando daño celular, pérdida de clorofila y debilitamiento de la planta. Su reproducción es rápida, especialmente bajo condiciones de calor y sequedad, prefiere hojas viejas, la parte baja de la planta, y condiciones de ambiente seco, polvoso, con buena radiación y temperatura elevadas. En las hojas aparecen pequeñas manchas puntiformes, amarillas o blanquecinas, que resultan del daño celular tipo "picoteado" (*stippling*) como se muestra en la Figura 1.5. (Hazard, 2022)



Figura 1.5.: [Picadura de la araña roja de dos manchas en una planta de tomate \(Inspired Pencil editors, 2022\)](#)

- Virus de la hoja amarilla (*Yellow leaf curl virus*): El agente causal es un virus los begomovirus, que afecta principalmente al tomate y ocasionalmente otras solanáceas. El vector principal es la mosca blanca, *Bemisia tabaci*, que adquiere el virus al alimentarse de plantas infectadas y luego lo transmite a plantas sanas. Las hojas muestran claramente: amarilleamiento entre las venas (clorosis interveinal), márgenes de la hoja amarillos, hojas enrolladas o curvadas hacia arriba y hacia dentro, y foliolos más pequeños de lo normal como muestra la Figura 1.6. (Díaz-Pendón et al., 2010)

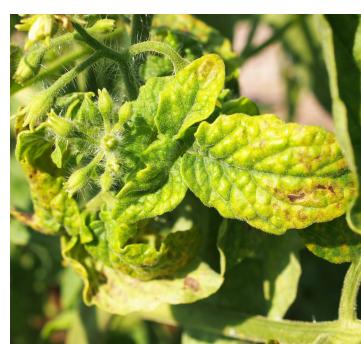


Figura 1.6.: [Virus de la hoja amarilla en una planta de tomate \(Greenlife editors, 2025\)](#)

La manifestación simultánea o sucesiva de estas enfermedades es una de las principales causas en la disminución en la productividad del cultivo a escala global. Además, muchas de estas enfermedades no solo viven en la planta sino que persisten en el suelo, semillas o herramientas que hayan interactuado con la planta, lo que dificulta su erradicación y aumenta los costos del tratamiento (Silva, s.f.).

Dada la magnitud del impacto de estas enfermedades, la detección temprana y precisa de las mismas es crucial. Permite una correcta intervención que minimiza las pérdidas, permitiendo la reducción del uso innecesario de los agroquímicos y mejorando la sostenibilidad. En este contexto, las tecnologías basadas en visión por computadora, sensores remotos e inteligencia artificial ofrecen soluciones eficaces para mejorar el seguimiento y el control sanitario de este cultivo clave.

## 2. Objetivos

El objetivo general de este proyecto consiste en desarrollar y evaluar un sistema inteligente basado en técnicas de aprendizaje profundo capaz de identificar el estado de salud y las principales enfermedades en hojas de tomate a partir de imágenes, asegurando un alto grado de precisión y capacidad de generalización.

### 2.1. Objetivos específicos

1. Revisar el estado del arte relacionado con la detección de enfermedades en plantas, con especial énfasis en el cultivo del tomate.
2. Analizar y preparar el conjunto de datos aplicando las técnicas necesarias de preprocesamiento de imágenes para garantizar la calidad y homogeneidad del material de entrenamiento.
3. Implementar y entrenar distintos modelos utilizando arquitecturas preentrenadas con alguna base de datos para adaptarlas a la tarea de clasificación de enfermedades en hojas de tomate.
4. Evaluar los modelos desarrollados comparando su rendimiento mediante métricas como precisión, sensibilidad o capacidad de generalización.

### 3. Estado del arte

En los últimos años la aplicación de técnicas de inteligencia artificial en la agricultura ha cobrado un papel relevante, especialmente en tareas de diagnóstico temprano de enfermedades en cultivos. El uso de aprendizaje profundo permite automatizar la detección de patrones en imágenes, lo cual puede ayudar a los agricultores a tomar decisiones más rápidas y eficientes.

Inicialmente, el enfoque tradicional para las tareas de clasificación de imágenes se basaba en características diseñadas manualmente como SIFT (Lowe, 2004), HoG (Dalal & Triggs, 2005), SURF (Bay et al., 2008), etc, para luego usar algún algoritmo de aprendizaje supervisado como máquinas de vectores de soporte (SVM), kvecinos más cercanos (KNN) y redes bayesianas. Sin embargo, estos enfoques dependían en gran medida de las características predefinidas. La ingeniería de características en sí misma es un proceso complejo y laborioso que debe revisarse cada vez que el problema en cuestión o el conjunto de datos asociado cambia considerablemente. Este problema se da en todos los intentos tradicionales de detectar enfermedades de las plantas , ya que se basaban en gran medida en características diseñadas manualmente, técnicas de mejora de imágenes y otras metodologías complejas.

Además, los enfoques tradicionales para la clasificación de enfermedades mediante *machine learning* suelen centrarse en un número reducido de clases, normalmente dentro de un mismo cultivo. Por ejemplo, se han usado imágenes térmicas y estéreo para detectar el moho polvoriento (*powdery mildew*) en hojas de tomate (Raza et al., 2015), imágenes RGBD para detección del tizón del manzano (*apple scab*) (Chéné et al., 2012) y sensores basados en aeronaves para la detección del virus del rizado amarillo del tomate (*tomato yellow leaf curl*) mediante el uso de un conjunto de pasos clásicos de extracción de características (Garcia-Ruiz et al., 2013).

Posteriormente han aparecido estudios que usan entrenamiento supervisado de extremo a extremo utilizando una arquitectura de red neuronal convolucional como (Krizhevsky et al., 2012) que demostraron que este tipo de arquitecturas tienen un buen desempeño incluidos en problemas de clasificación de imágenes con un número de clases, superando con creces a los enfoques tradicionales. El hecho de no depender de la laboriosa fase de la ingeniería de características, además de la generalización que alcanza convierte este enfoque en un candidato prometedor para la detección computacional de enfermedades en plantas. (Mohanty et al., 2016)

Para aplicaciones de diagnóstico de enfermedades en plantas es habitual seleccionar arquitecturas que equilibren precisión y coste computacional. Arquitecturas ligeras como *MobileNetV2* y *NASNetMobile* han demostrado ser adecuadas para sistemas móviles por su bajo número de parámetros y alta velocidad de inferencia, manteniendo precisiones comparables a modelos más pesados tras *fine tuning* (Lu et al., 2023). Por su parte, *EfficientNetB0*, diseñado mediante *compound scaling*, ofrece una eficiencia de parámetros superior y suele vencer a tradicionales en relación precisión y tamaño (Atila et al., 2021). Estas características hacen que *MobileNetV2*, *EfficientNetB0* y *NASNetMobile* sean opciones sólidas para aplicaciones prácticas de diagnóstico en tomate, especialmente cuando el despliegue en dispositivos con recursos limitados.

El uso de modelos preentrenados en bases de datos extensas como *ImageNet* constituye una práctica consolidada en tareas de clasificación de enfermedades foliares, dado que permiten aprovechar representaciones visuales previamente aprendidas (bordes, texturas, patrones) para iniciar el entrenamiento en un dominio específico con menor cantidad de datos. Estudios recientes demuestran que el uso de *transfer learning* a partir de arquitecturas entrenadas sobre *ImageNet* acelera la convergencia, reduciendo el número de épocas (*epochs*), cantidad de datos y riesgo de sobreajuste (*overfitting*) (Al Sahili & Awad, 2022). No obstante, también se advierte que la diferencia de dominio entre *Imagenet* y las imágenes reales de hojas puede afectar a la generalización, lo cual resalta la importancia de una augmentación adecuada, ajustes de hiperparámetros y validación con imágenes de campo (Dong et al., 2023).

En este trabajo se emplean modelos basados en las arquitecturas *MobileNetV2*, *EfficientNetB0* y *NASNetMobile* preentrenados con *ImageNet* como punto de partida, lo que permite centrar el esfuerzo en la adaptación al problema de enfermedades de tomate mediante fine-tuning y validación rigurosa.

En la literatura sobre clasificación automática de enfermedades foliares mediante redes neuronales convolucionales, la evaluación del rendimiento del modelo rara vez se limita a la precisión global. Se emplean métricas complementarias como precisión, sensibilidad (*recall*) y *F1-Score* para reflejar la capacidad del modelo tanto de minimizar falsos positivos como falsos negativos (Salka et al., 2025), lo cual es crítico en un contexto agronómico donde tanto detectar tarde una enfermedad como etiquetar erróneamente una hoja sana pueden acarrear consecuencias importantes. De la misma manera, es interesante analizar la función de pérdidas de validación durante el entrenamiento, ya que permite evaluar la convergencia y ayudar a la detección de sobreajuste (Saleem et al., 2020).

Por otra parte, no siempre es suficiente con observar métricas sobre el modelo en su conjunto, sino que a veces también es relevante analizar métricas como las mencionadas (precisión, sensibilidad y *F1-Score*) por clases, para determinar si existe algún sesgo hacia alguna clase (Chimate et al., 2025). En esta línea para apoyarse en alguna métrica más visual también es interesante el uso de la matriz de confusión que permite ver

cómo se comporta el modelo entre diferentes clases pudiendo visualizar el número de imágenes por clases, las predicciones sobre las mismas con sus aciertos y fallos de un vistazo (Chimate et al., 2025).

Más allá de entrenar una única red convolucional sobre imágenes de hojas, los estudios recientes incorporan un enfoque adicional que consiste en crear *ensembles* o las combinaciones de diferentes arquitecturas y luego combinan sus predicciones por votación (*hard-voting*, *soft-voting*) o mediante promedios ponderados. En esta línea nos encontramos estudios como el realizado en el artículo que utilizaron tres arquitecturas, *DenseNet201*, *InceptionV3* y *Xception* y un método de *weighted soft-voting* para lograr un precisión superior al de cada modelo individual (Octacian et al., 2024).

## 4. Implementación y desarrollo

En este capítulo se presenta tanto el *hardware* como el *software* usados en este proyecto. Además se explica la procedencia y estructura del conjunto de datos que serán usados para el estudio. Finalmente, se desarrolla el preprocesamiento que se realiza a este conjunto de datos junto con los modelos entrenados para conseguir un clasificador.

### 4.1. Herramientas usadas

Para llevar a cabo este proyecto, se ha usado Google Colab (abreviatura de Google Colaboratory) que se accedía desde el ordenador portátil del autor del documento. Este ordenador es un ASUS TUF Gaming FX505GT que cuenta con las siguientes características:

- 16 GB de RAM con formato DDR4.
- Almacenamiento compuesto por un disco duro con tecnología SSD de 512GB.
- Procesador Intel Core i7-9750H CPU a 2.60 GHz, con 6 procesadores principales y 6 procesadores lógicos.
- Tarjeta gráfica NVIDIA GeForce GTX 1650 con 4GB de RAM.

Google Colab es un servicio gratuito de Google que permite escribir y ejecutar código en la nube sin necesidad de instalar nada en tu equipo. Los recursos que ofrece de forma gratuita varían con el tiempo, pero las características que suele ofrecer son las siguientes:

- GPU NVIDIA Tesla T4 con 16 GB de VRAM ó CPU Intel Xeon con alrededor de 13 GB de RAM.
- Almacenamiento temporal se corresponde con unos 100 GB de espacio en disco.
- La duración de la sesión puede ser de hasta 12 horas, aunque en la práctica podrían terminarse antes según uso y carga del sistema.

Por otra parte el lenguaje de programación usado ha sido Python, un lenguaje que es ampliamente utilizado por científicos de datos. En las últimas décadas Python se ha

enriquecido con numerosas librerías relacionadas con técnicas de ML que facilitan el uso de las mismas. En concreto para este proyecto se ha utilizado la versión 3.12.11 de Python.

En cuanto a las librerías de Python usadas para la implementación, se presentan a continuación:

- NumPy: es una librería que ofrece la posibilidad de crear matrices y vectores multidimensionales y provee además un gran número de operaciones matemáticas de alto nivel.
- Pandas: es una librería que ofrece la estructura de datos llamada DataFrame que facilita la manipulación y el análisis de datos. Es una extensión de la librería NumPy. Ha sido usada para tratar y transformar los datos.
- Plantcv: es una librería de Python de código abierto diseñada específicamente para el análisis de imágenes de plantas. Se ha usado para lectura de las imágenes.
- Tensorflow: es una librería de software de código abierto creada por Google para desarrollar y entrenar modelos de machine learning (ML) y deep learning (DL). Recibe este nombre porque trabaja con tensores, estructuras de datos multidimensionales, como matrices o vectores que fluyen a través de un grafo computacional de operaciones. Se ha usado para crear y entrenar los modelos descritos en este proyecto.
- Seaborn: esta librería permite la visualización de los datos a través de distintos tipos de gráficas. Ha sido usada para realizar los distintos gráficos como las matrices de confusión para evaluar los modelos.
- Sklearn: es una librería que contiene un gran número funciones relacionadas con modelos de machine learning (ML) y deep learning (DL). Se ha usado para extraer las principales métricas de los modelos tras su entrenamiento.

## 4.2. Procedencia y descripción de los datos

Los datos provienen de la plataforma online de ciencia de datos de Google, Kaggle, que funciona como una mezcla de red social, repositorio de datasets y espacio de competición. En concreto, el conjunto de datos usado es el llamado "Tomato Leaves Dataset". Según su descripción en la misma plataforma se trata de un conjunto de datos de más de 20.000 imágenes de hojas de tomate con 11 clases, 10 enfermedades y una clase sana. Estas imágenes se han recopilado tanto en entornos de laboratorio como en entornos naturales. (Motwani & Khan, 2025)

En concreto se pueden extraer dos directorios que servirán como conjunto de datos

para el entrenamiento y conjunto de datos para validación, ambos cuentan con 11 directorios con imágenes dentro. Cada uno de estos subdirectorios representa una de las clases que serán brevemente expuestas a continuación:

Cuadro 4.1.: Clases del conjunto de datos

<b>Conjunto de datos</b>	
Clase	Traducción Clase
Healthy	Saludable
Bacterial_spot	Manchas bacterianas
Early_blight	Tizón precoz
Late_blight	Tizón tardío
Leaf_Mold	Hojas con moho
Powdery_mildew	Moho polvoriento
Septoria_leaf_spot	Hojas manchadas de septoriosis
Spidermite_Two-spotted_spider_mite	Picadura de araña roja de dos manchas
Target_Spot	Punto blanco
Tomato_mosaic_virus	Virus mosaico
Tomato_Yellow_Leaf_Curl_Virus	Virus de la hoja amarilla

### 4.3. Preprocesado de los datos

Al realizar la carga de datos se tiene un directorio con dos subdirectorios, cada uno de ellos representará un conjunto de datos, uno de datos de entrenamiento (*train*) y otro de validación (*valid*). Cada uno de estos directorios contienen a su vez 11 directorios, representando cada una de las clases.

A continuación, se llevan a cabo las primeras tareas de exploración de las imágenes.

En primer lugar, seleccionando el directorio que contiene los datos de entrenamiento se realiza una función para obtener el número de imágenes existentes por cada tipo de tamaño. Gracias a esta función se sabe que se cuentan con imágenes de variables tamaños, concretamente existen 760 tipos de tamaños distintos. En la siguiente tabla se exponen los 5 tipos de tamaño que más se repiten:

Cuadro 4.2.: Los 5 tamaños de imágenes más comunes

<b>Top 5 tamaños de imágenes</b>	
Tamaño (píxeles)	Nº imágenes
256x256	18942
227x227	4120
640x640	1207
533x800	151
800x600	10

De cara a construir un modelo todas las imágenes tienen que tener el mismo tamaño y debida a esta primera toma de contacto se toma la decisión de transformar todas las imágenes a tamaño de 256x256 píxeles. Sin embargo, debido a los modelos usados se termina convirtiendo a imágenes de tamaño 224x224, ya que es un tamaño estándar que entiende cualquier modelo y es el más cercano a los dos tipos de tamaños de imágenes más usuales en nuestros datos.

En segundo lugar, se realiza una muestra aleatoria de una imagen por clase, para visualizar el tipo de imágenes que se van a tratar en la Figura 4.1.

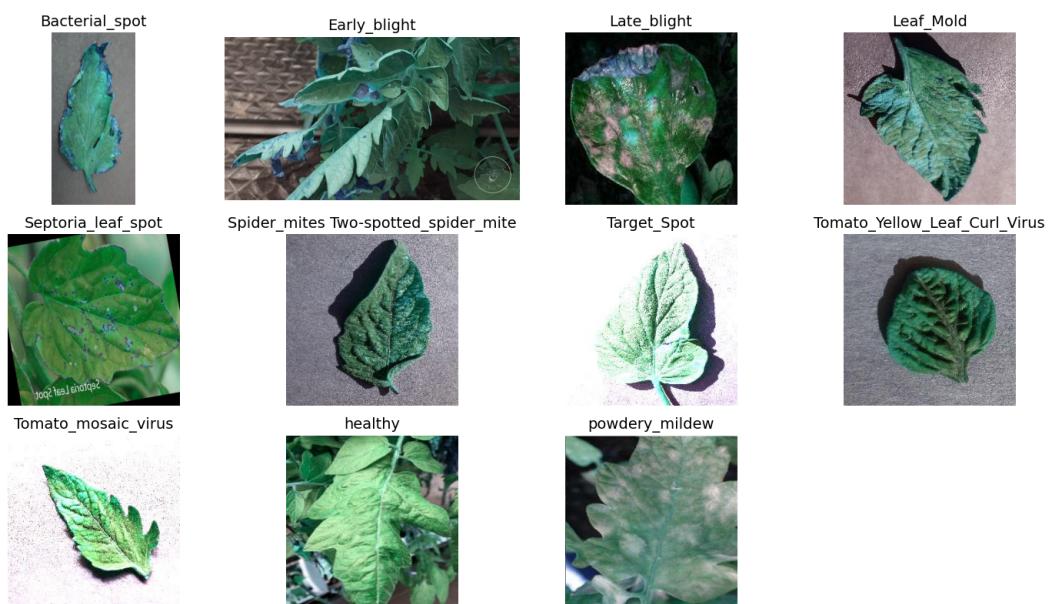


Figura 4.1.: Imagen aleatoria por clase

En la Figura 4.1 se puede observar la gran variedad de imágenes que existen, con distintos brillos, distintos fondos o incluso giradas:

- En cuanto al brillo se puede observar que la imagen de Target\_Spot tiene mucho más brillo que la de Tomato\_Yellow\_Leaf\_Curl\_Virus.
- Con respecto al fondo podemos ver distinciones entre la imagen Bacterial\_spot con un fondo grisáceo plano, la imagen Late\_blight con fondo completamente negro y la foto Early\_blight en la que se ve el resto de la planta de tomate, no solo se ve una hoja.
- También se puede destacar la diferencia entre las posiciones de las hojas, algunas como Bacterial\_spot tienen el tallo abajo, otras como Leaf\_Mold tienen el tallo arriba y otras como Septoria\_leaf\_spot tienen el tallo horizontal.
- Además se tiene un ejemplo de imagen rotada, concretamente la Septoria\_leaf\_spot

Debido a esta gran variedad de imágenes y de enfermedades se llegó a la conclusión de que no tiene mucho sentido usar funciones especiales de la librería PlantCV, ya que esta está muy orientada al fenotipado clásico (área, forma, color, índices), y esas características a veces no capturan la complejidad de patrones de enfermedades, que suelen ser más sutiles y no lineales. La estrategia que se ha seguido es pasar a un pipeline de *deep learning* con imágenes preprocesadas de forma estándar. De esta manera el modelo aprenderá por si mismo las características discriminantes en lugar de imponer un conjunto de *features* manuales.

A continuación se dispuso a formar los conjuntos de datos que usará el modelo. Hasta ahora se tienen datos para el entrenamiento del modelo y para la validación del mismo, sin embargo, no se tienen datos para realizar pruebas sobre el modelo resultante. Por lo tanto, se formará un nuevo conjunto de datos de pruebas a partir del conjunto de entrenamiento, concretamente, seleccionando un 20 % de sus datos.

Para realizar esta tarea se recorre el directorio de datos de entrenamiento y aleatoriamente se seleccionan imágenes de cada subdirectorio (clase) y se añaden a un nuevo directorio que será el de datos de prueba.

Otra tarea importante en cuanto al procesado de los datos es normalizar los mismos. Para ello, se hace uso de *ImageDatagenerator* de Tensorflow. Se usa esta función porque también sirve para aplicar la técnica de aumento de datos (*data augmentation*) (Dong et al., 2023) en los datos de entrenamiento. Esta técnica consiste en generar datos adicionales a partir de los ya existentes aplicando transformaciones que mantienen la esencia de la información original, pero la presentan de manera distinta. Su objetivo principal es enriquecer el conjunto de entrenamiento para mejorar la capacidad de generalización del modelo resultante.

Es decir, a los tres conjuntos de datos se le aplaza normalización que consiste en transformar los pixeles de cada imagen de valores que van de 0 a 255 (escala RGB) a valores entre 0 y 1. Además a las imágenes del conjunto de entrenamiento se le aplica la técnica de *data augmentation* que se ha mencionado anteriormente, concretamente se le aplican las siguientes transformaciones:

- Rota aleatoriamente la imagen hasta más menos 20 grados.
- Aplica un zoom aleatorio entre 80 % y 120 % del tamaño original.
- Desplaza la imagen horizontalmente hasta un 20 % de su ancho.
- Desplaza la imagen verticalmente hasta un 20 % de su alto.
- Aplica una transformación de cizallamiento (*shear*), como si se deformara la imagen en diagonal.
- Gira horizontalmente las imágenes aleatoriamente.

- Cambia el brillo aleatoriamente entre 80 % y 120 %.
- Cuando una transformación (como una rotación o desplazamiento) deja espacios vacíos en la imagen, estos se rellenan con el valor del píxel más cercano.

Estas transformaciones se aplicarán aleatoriamente cada vez que el generador entrega un lote al modelo durante el entrenamiento. De esta manera, el modelo el modelo nunca verá dos veces exactamente la misma versión de la imagen.

Como ya se ha mencionado se ha usado *ImageDataagenerator* para formar los conjuntos de datos. Concretamente lo que permite es crear generadores, que no son más que objetos que producen datos de forma incremental, lote a lote, en lugar de cargar todo el *dataset* en memoria de golpe. El generador se conecta a un directorio con datos y los va leyendo en lotes de un tamaño determinado, en este caso se ha usado un tamaño de 32, es decir, va formando lotes de 32 imágenes y a estas se les aplica normalización y también las transformaciones de *data augmentation* mencionadas si son datos de entrenamiento.

De esta manera, quedan los siguientes conjuntos de datos:

- Entrenamiento: Tiene 20.686 imágenes utilizadas para entrenar el modelo. Se trata de la mayor parte de los datos, ya que el modelo necesita muchos ejemplos para aprender patrones.
- Validación: Tiene 6.683 imágenes. Estos datos se utilizan para evaluar el rendimiento del modelo durante el entrenamiento, sin afectar a los parámetros del modelo.
- Prueba: Hay 5.165 imágenes para la prueba final. Este conjunto de datos se utiliza una vez finalizado el entrenamiento para medir objetivamente el rendimiento del modelo con datos nuevos que nunca se han visto.

Otro dato importante del conjunto de entrenamiento es el número de imágenes que se tiene por clase, ya que si hubiera muchas más imágenes de una clase que de otras, el modelo podría incluir un sesgo no deseado. El número de imágenes por clase del conjunto de entrenamiento es el siguiente:

La mayoría de clases tienen entre 2000 y 2400 imágenes, lo cual es aceptable. Sin embargo, existen clases claramente minoritarias, como *Spidermite\_Two-spotted\_spider-mite* y *Target\_Spot* con menos de 1500 imágenes, o *powdery\_mildew* con menos de 1000 imágenes. Esto puede causar que el modelo aprenda mejor las clases con más ejemplos y tienda a confundirse en las minoritarias porque tiene menos exposición a ellas.

Para evitar el posible problema de sesgo en el modelo se aplicará una técnica de balanceo al conjunto de entrenamiento. Específicamente se usará ponderación de clases con un diccionario *class\_weight* que se puede añadir al modelo en forma de parámetro. Esto dará más peso a los errores en clases minoritarias, para que el modelo no las

Cuadro 4.3.: Número de imágenes por clases en el conjunto de entrenamiento

Número de imágenes por clases	
Clase	Nº imágenes
Bacterial_spot	2261
Early_blight	1964
Late_blight	2491
Leaf_Mold	2204
Septoria_leaf_spot	2306
Spidermite_Two-spotted_spider_mite	1398
Target_Spot	1462
Tomato_Yellow_Leaf_Curl_Virus	1632
Tomato_mosaic_virus	1723
healthy	2441
powdery_mildew	804

ignore.

Los pesos de las clases se han calculado usando la función `compute_class_weight` que hace que las clases con menor muestras tengan mayor peso y las que tengan más muestras tengan menor peso. Haciendo, en promedio, que todas las clases tengan la misma importancia. Los pesos aplicados son los que se muestran a continuación:

Cuadro 4.4.: Pesos asignados a cada clase en el entrenamiento del modelo

Pesos por clases		
Clase	Nº imágenes	Peso
Bacterial_spot	2261	0.8317
Early_blight	1964	0.9575
Late_blight	2491	0.7549
Leaf_Mold	2204	0.8532
Septoria_leaf_spot	2306	0.8155
Spidermite_Two-spotted_spider_mite	1398	1.3451
Target_Spot	1462	1.2862
Tomato_Yellow_Leaf_Curl_Virus	1632	1.1522
Tomato_mosaic_virus	1723	1.0914
healthy	2441	0.7703
powdery_mildew	804	2.3389

## 4.4. Modelado

En esta sección se describe la estrategia de modelado empleada, común a todos los experimentos realizados posteriormente. El enfoque adoptado se basa en el uso de redes neuronales convolucionales (CNN), dado que constituyen la arquitectura de

referencia en tareas de clasificación de imágenes al ser capaces de extraer de manera automática y jerárquica características relevantes de los datos visuales.

Considerando las limitaciones de recursos computacionales disponibles, se optó por la técnica de *transfer learning*. Esta metodología permite aprovechar modelos previamente entrenados sobre grandes bases de datos, de modo que las capas iniciales ya contienen representaciones generales de las imágenes. Posteriormente, dichas representaciones se ajustan a la tarea específica de clasificación de enfermedades en hojas de tomate.

Concretamente se han usado los modelos *MobileNetV2*, *NASNetMobile* y *EfficientNetB0* preentrenados con el dataset de *ImageNet*. De esta manera se aprovechan los pesos previamente aprendidos, que contienen representaciones visuales generales como bordes, texturas y formas, para inicializar la red. Posteriormente, esta red se adapta a la clasificación de enfermedades en hojas de tomate.

Además, cabe mencionar que para algunos modelos se ha usado la técnica de *fine tuning* que va un poco más allá que *transfer learning*, ya que tras entrenar un modelo con una gran base de datos se descongelan algunas de sus capas para que el modelo final se adapte más a la tarea específica que en este caso es la clasificación de enfermedades en hojas de tomate.

Al haber usado la técnica de *fine tuning* tiene sentido añadir las características principales de las arquitecturas de redes neuronales usadas. Esta tarea se lleva a cabo en la sección 4.5. Con el objetivo de gestionar mejor los recursos disponibles, se implementaron distintos callbacks:

- *HistorySaver*: permite guardar el historial de métricas en un archivo externo. Esto resulta especialmente útil en entornos con recursos limitados o sesiones interrumpibles (como Google Colab), ya que garantiza que la información del entrenamiento no se pierda. Concretamente se ha usado para guardar las métricas de precisión y pérdida de entrenamiento y validación.
- *ModelCheckpoint*: encargado de almacenar en disco el modelo con mejor desempeño en validación, según la métrica de precisión (`val_accuracy`). De este modo, se asegura la conservación de la mejor versión del modelo entrenado, evitando depender únicamente de los pesos finales.

Estos callbacks se han usado en todos los entrenamientos. Sin embargo, en algunos entrenamientos se han usado además otros dos callbacks:

- *EarlyStopping*: detiene el entrenamiento de forma anticipada si la pérdida de validación no mejora durante un número determinado de épocas consecutivas. Esto evita sobreentrenamiento y reduce el tiempo de cómputo innecesario.
- *ReduceLROnPlateau*: ajusta de manera dinámica la tasa de aprendizaje cuando la

pérdida de validación alcanza una meseta. Gracias a esta reducción progresiva, el modelo puede seguir afinando sus parámetros con pasos cada vez más pequeños, lo que mejora la convergencia.

En conjunto, estos callbacks permitieron no solo optimizar el uso de los recursos computacionales disponibles, sino también obtener modelos más robustos y con mejor capacidad de generalización.

Por último cabe destacar que para que los modelos puedan ser reconstruidos por cualquiera con acceso a los mismos datos de los que se parte es necesario fijar a un valor fijo disintas semillas aleatorias:

- Para la inicialización de los pesos en el modelo se usa `tf.random.set_seed(42)`.
- Para la generación de los conjuntos de datos con `ImageDataagenerator` se usa el parámetro `seed` con valor 42.
- Para el barajado de los datos del conjunto de entrenamiento `random.seed(42)` y `np.random.seed(42)`.

#### 4.4.1. Modelo *MobileNetV2 1*

Este modelo usa de capa base *MobileNetV2* preentrenada con los pesos de *ImageNet*. Posteriormente se congelan las capas del modelo base para evitar que los pesos se modifiquen durante el entrenamiento y se use correctamente la técnica de *transfer learning*. A continuación se añade una capa de *GlobalAveragePooling* y una capa densa con 11 neuronas y función de activación *softmax*, ya que nos enfrentamos a una clasificación de 11 clases. Esta información es la que se muestra en la Figura 4.2.

Layer (type)	output Shape	Param #
mobilenetv2_1.00_224 (Functional)	(None, 7, 7, 1280)	2,257,984
global_average_pooling2d_2 (GlobalAveragePooling2D)	(None, 1280)	0
dense_6 (Dense)	(None, 11)	14,091
<b>Total params: 2,272,075 (8.67 MB)</b> <b>Trainable params: 14,091 (55.04 KB)</b> <b>Non-trainable params: 2,257,984 (8.61 MB)</b>		

Figura 4.2.: Resumen del modelo *MobileNetV2 1*

Finalmente para la compilación se usa el optimizador Adam, la función de pérdida de entropía cruzada categórica y se escoge la métrica de precisión para evaluar el rendimiento.

Se usa la función de pérdida de entropía cruzada categórica porque al crear los conjuntos de datos con `ImageDatagenerator` como se indica en la sesión 4.3 se usa el parámetro `class mode` con valor `categorical` que convierte las etiquetas a vectores `one-hot encoded`.

#### 4.4.2. Modelo *MobileNetV2* 2

Este modelo usa de capa base *MobileNetV2* preentrenada con los pesos de *ImageNet*. Posteriormente se congelan las capas del modelo base para evitar que los pesos se modifiquen durante el entrenamiento, pero en esta ocasión se descongelan algunas capas, usando la técnica de *fine tuning*. Concretamente, se descongela a partir de la capa 100 con el objetivo de que el modelo se adapte mejor a nuestro caso concreto como se explica en la Subsección 4.5.1. A continuación se añade una capa de *GlobalAveragePooling* y una capa densa con 11 neuronas y función de activación *softmax*, ya que nos enfrentamos a una clasificación de 11 clases. Esta información es la que se muestra en la Figura 4.3.

Layer (type)	Output Shape	Param #
mobilenetv2_1.00_224 (Functional)	(None, 7, 7, 1280)	2,257,984
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 1280)	0
dense_1 (Dense)	(None, 11)	14,091

Total params: 2,272,075 (8.67 MB)  
Trainable params: 1,875,531 (7.15 MB)  
Non-trainable params: 396,544 (1.51 MB)

Figura 4.3.: [Resumen del modelo \*MobileNetV2\* 2](#)

Finalmente para la compilación se usa el optimizador Adam, la función de pérdida de entropía cruzada categórica y se escoge la métrica de precisión para evaluar el rendimiento. Sin embargo, no se usa un *learning rate* por defecto ( $1e-3$ ), se usa un *learning rate* menor ( $1e-5$ ).

Se usa la función de pérdida de entropía cruzada categórica porque al crear los conjuntos de datos con `ImageDatagenerator` como se indica en la Sección 4.3 se usa el parámetro `class mode` con valor `categorical` que convierte las etiquetas a vectores `one-hot encoded`.

#### 4.4.3. Modelo *MobileNetV2* 3

Este modelo usa de capa base *MobileNetV2* preentrenada con los pesos de *ImageNet*. Posteriormente se congelan las capas del modelo base para evitar que los pesos se modifiquen durante el entrenamiento, pero en esta ocasión se descongelan algunas capas, usando la técnica de *fine tuning*. Concretamente, se descongela a partir de la capa 80 con el objetivo de que el modelo se adapte mejor a nuestro caso concreto como se explica en la Subsección 4.5.1. A continuación se añade una capa de *GlobalAveragePooling* y una capa densa con 11 neuronas y función de activación *softmax*, ya que nos enfrentamos a una clasificación de 11 clases. Esta información es la que se muestra en la Figura 4.4.

Layer (type)	Output Shape	Param #
mobilenetv2_1.00_224 (Functional)	(None, 7, 7, 1280)	2,257,984
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 1280)	0
dense_1 (Dense)	(None, 11)	14,091

Total params: 2,272,075 (8.67 MB)  
 Trainable params: 1,875,531 (7.15 MB)  
 Non-trainable params: 396,544 (1.51 MB)

Figura 4.4.: [Resumen del modelo \*MobileNetV2\* 3](#)

Finalmente para la compilación se usa el optimizador Adam, la función de pérdida de entropía cruzada categórica y se escoge la métrica de precisión para evaluar el rendimiento. Sin embargo, no se usa un *learning rate* por defecto ( $1e-3$ ), se usa un *learning rate* menor ( $5e-6$ ).

Se usa la función de pérdida de entropía cruzada categórica porque al crear los conjuntos de datos con *ImageDataGenerator* como se indica en la Sección 4.3 se usa el parámetro *class mode* con valor *categorical* que convierte las etiquetas a vectores *one-hot encoded*.

#### 4.4.4. Modelo *EfficientNetB0* 1

Este modelo usa de capa base *EfficientNetB0* preentrenada con los pesos de *ImageNet*. Posteriormente se congelan las capas del modelo base para evitar que los pesos se modifiquen durante el entrenamiento y se use correctamente la técnica de *transfer learning*. A continuación se añade una capa de *GlobalAveragePooling* y una capa densa con 11 neuronas y función de activación *softmax*, ya que nos enfrentamos a una clasificación de 11 clases. Esta información es la que se muestra en la Figura 4.5.

Layer (type)	Output Shape	Param #
efficientnetb0 (Functional)	(None, 7, 7, 1280)	4,049,571
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1280)	0
dense (Dense)	(None, 11)	14,091

Total params: 4,063,662 (15.50 MB)  
Trainable params: 14,091 (55.04 KB)  
Non-trainable params: 4,049,571 (15.45 MB)

Figura 4.5.: Resumen del modelo *EfficientNetB0* 1

Finalmente para la compilación se usa el optimizador Adam, la función de pérdida de entropía cruzada categórica y se escoge la métrica de precisión para evaluar el rendimiento.

Se usa la función de pérdida de entropía cruzada categórica porque al crear los conjuntos de datos con ImageDataGenerator como se indica en la Sección 4.3 se usa el parámetro *class mode* con valor categorical que convierte las etiquetas a vectores *one-hot encoded*.

#### 4.4.5. Modelo *EfficientNetB0* 2

Este modelo usa de capa base *EfficientNetB0* preentrenada con los pesos de *ImageNet*. Posteriormente se congelan las capas del modelo base para evitar que los pesos se modifiquen durante el entrenamiento, pero en esta ocasión se descongelan algunas capas, usando la técnica de *fine tuning*. Concretamente, se descongela a partir de la capa 100 con el objetivo de que el modelo se adapte mejor a nuestro caso concreto como se explica en la Subsección 4.5.2. A continuación se añade una capa de *GlobalAveragePooling* y una capa densa con 11 neuronas y función de activación *softmax*, ya que nos enfrentamos a una clasificación de 11 clases. Esta información es la que se muestra en la Figura 4.6.

Layer (type)	output Shape	Param #
efficientnetb0 (Functional)	(None, 7, 7, 1280)	4,049,571
global_average_pooling2d_3 (GlobalAveragePooling2D)	(None, 1280)	0
dense_3 (Dense)	(None, 11)	14,091

Total params: 4,063,662 (15.50 MB)  
Trainable params: 3,854,439 (14.70 MB)  
Non-trainable params: 209,223 (817.28 KB)

Figura 4.6.: Resumen del modelo *EfficientNetB0* 2

Finalmente para la compilación se usa el optimizador Adam, la función de pérdida de entropía cruzada categórica y se escoge la métrica de precisión para evaluar el rendimiento. Sin embargo, no se usa un *learning rate* por defecto ( $1e-3$ ), se usa un *learning rate* menor ( $1e-5$ ).

Se usa la función de pérdida de entropía cruzada categórica porque al crear los conjuntos de datos con `ImageDatagenerator` como se indica en la Sección 4.3 se usa el parámetro *class mode* con valor *categorical* que convierte las etiquetas a vectores *one-hot encoded*.

#### 4.4.6. Modelo *EfficientNetB0* 3

Este modelo usa de capa base *EfficientNetB0* preentrenada con los pesos de *ImageNet*. Posteriormente se congelan las capas del modelo base para evitar que los pesos se modifiquen durante el entrenamiento, pero en esta ocasión se descongelan algunas capas, usando la técnica de *fine tuning*. Concretamente, se descongela a partir de la capa 100 con el objetivo de que el modelo se adapte mejor a nuestro caso concreto como se explica en la Subsección 4.5.2. A continuación se añade una capa de *GlobalAveragePooling* y una capa densa con 11 neuronas y función de activación *softmax*, ya que nos enfrentamos a una clasificación de 11 clases. Esta información es la que se muestra en la Figura 4.6.

Layer (type)	output Shape	Param #
efficientnetb0 (Functional)	(None, 7, 7, 1280)	4,049,571
global_average_pooling2d_3 (GlobalAveragePooling2D)	(None, 1280)	0
dense_3 (Dense)	(None, 11)	14,091

Total params: 4,063,662 (15.50 MB)  
 Trainable params: 3,854,439 (14.70 MB)  
 Non-trainable params: 209,223 (817.28 KB)

Figura 4.7.: Resumen del modelo *EfficientNetB0* 3

Finalmente para la compilación se usa el optimizador Adam, la función de pérdida de entropía cruzada categórica y se escoge la métrica de precisión para evaluar el rendimiento. Sin embargo, no se usa un *learning rate* por defecto ( $1e-3$ ), se usa un *learning rate* menor ( $3e-6$ ).

Se usa la función de pérdida de entropía cruzada categórica porque al crear los conjuntos de datos con `ImageDatagenerator` como se indica en la Sección 4.3 se usa el parámetro *class mode* con valor *categorical* que convierte las etiquetas a vectores *one-hot encoded*.

#### 4.4.7. Modelo *NASNetMobile* 1

Este modelo usa de capa base *NASNetMobile* preentrenada con los pesos de *ImageNet*. Posteriormente se congelan las capas del modelo base para evitar que los pesos se modifiquen durante el entrenamiento y se use correctamente la técnica de *transfer learning*. A continuación se añade una capa de *GlobalAveragePooling* y una capa densa con 11 neuronas y función de activación *softmax*, ya que nos enfrentamos a una clasificación de 11 clases. Esta información es la que se muestra en la Figura 4.8.

Layer (type)	output shape	Param #
nasnet_mobile (Functional)	(None, 7, 7, 1056)	4,269,716
global_average_pooling2d_2 (GlobalAveragePooling2D)	(None, 1056)	0
dense_2 (Dense)	(None, 11)	11,627

Total params: 4,281,343 (16.33 MB)  
Trainable params: 11,627 (45.42 KB)  
Non-trainable params: 4,269,716 (16.29 MB)

Figura 4.8.: [Resumen del modelo \*NASNetMobile\* 1](#)

Finalmente para la compilación se usa el optimizador Adam, la función de pérdida de entropía cruzada categórica y se escoge la métrica de precisión para evaluar el rendimiento.

#### 4.4.8. Modelo *NASNetMobile* 2

Este modelo usa de capa base *NASNetMobile* preentrenada con los pesos de *ImageNet*. Posteriormente se congelan las capas del modelo base para evitar que los pesos se modifiquen durante el entrenamiento, pero en esta ocasión se descongelan algunas capas, usando la técnica de *fine tuning*. Concretamente, se descongela a partir de la capa 60 con el objetivo de que el modelo se adapte mejor a nuestro caso concreto como se explica en la Subsección 4.5.3. A continuación se añade una capa de *GlobalAveragePooling* y una capa densa con 11 neuronas y función de activación *softmax*, ya que nos enfrentamos a una clasificación de 11 clases. Esta información es la que se muestra en la Figura 4.9.

Finalmente para la compilación se usa el optimizador Adam, la función de pérdida de entropía cruzada categórica y se escoge la métrica de precisión para evaluar el rendimiento. Sin embargo, no se usa un *learning rate* por defecto (1e-3), se usa un *learning rate* menor (5e-6).

Se usa la función de pérdida de entropía cruzada categórica porque al crear los conjuntos de datos con *ImageDataGenerator* como se indica en la Sección 4.3 se usa el

Layer (type)	Output Shape	Param #
nasnet_mobile (Functional)	(None, 7, 7, 1056)	4,269,716
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1056)	0
dense (Dense)	(None, 11)	11,627

Total params: 4,281,343 (16.33 MB)  
Trainable params: 4,233,471 (16.15 MB)  
Non-trainable params: 47,872 (187.00 KB)

Figura 4.9.: [Resumen del modelo NASNetMobile 2](#)

parámetro *class mode* con valor categorical que convierte las etiquetas a vectores *one-hot encoded*.

## 4.5. Arquitecturas

En esta sección se explicarán las principales características de las arquitecturas empleadas para construir los modelos de la sección 4.4

### 4.5.1. MobileNetV2

MobileNetV2 es una arquitectura de CNN muy eficiente diseñada para aplicaciones de visión integrada en móviles. Ha sido desarrollada por Google para mejorar a su antecesor MobileNetV1 proporcionando una mejor precisión y reduciendo la capacidad de cómputo. Sus características claves son las siguientes (GeeksforGeeks contributors, 2025b):

1. Bloques residuales invertidos (*Inverted Residuals*): Primero se expande el número de canales mediante una convolución 1x1 (capa de expansión). Luego se aplica una convolución "depthwise" (convolución separada por canal) para realizar el filtrado espacial. Finalmente, se reduce de nuevo el número de canales a la dimensión deseada mediante otra 1x1 (capa de proyección). Esta estructura permite conservar dimensiones de entrada/salida mientras se reduce el coste computacional.
2. Convoluciones separables en profundidad (*Depthwise Separable Convolutions*): Al igual que su antecesor, utiliza convoluciones separables, es decir, se descompone una convolución estándar en dos operaciones; primero una convolución "depthwise" por cada canal, luego una "pointwise 1x1" que mezcla los canales. Esto reduce considerablemente el número de parámetros y de cálculos, lo que mejora la eficiencia.

3. Embotellamientos lineales (*Linear Bottlenecks*): Entre capas se insertan *bottlenecks* lineales, que ayudan a que la "variedad"(*manifold*) de los datos de entrada no se comprima excesivamente. Esta práctica contribuye a conservar mayor información, lo cual mejora la precisión del modelo.
4. Función de activación ReLU6: emplea la función de activación ReLU6, una versión modificada de la ReLU clásica que limita los valores de activación al rango [0, 6]. Esto favorece la cuantización (es decir, el uso eficiente de formatos de menor precisión) en dispositivos móviles manteniendo un buen equilibrio entre eficiencia y precisión.

A continuación se realiza una división funcional de la red para entender qué aprenden las capas (Sandler et al., 2019):

- Zona temprana: Comprende los primeros bloques, aproximadamente hasta la capa 20. En esta etapa, la red aprende características muy básicas, como bordes, texturas simples, colores, orientación y gradientes. Se trata de filtros universales que casi cualquier red visual aprende.
- Zona intermedia: Incluye los bloques medios, aproximadamente desde la capa 20 hasta la 80. Aquí se capturan características más complejas, como combinaciones de formas, patrones locales, partes de objetos y texturas específicas (por ejemplo, de hojas o plantas). En este punto, la red comienza a adaptarse a dominios más específicos.
- Zona tardía o cabeza del modelo: corresponde a los últimos bloques, desde la capa 80 hasta la final. Esta zona extrae características de muy alto nivel, relacionadas con la semántica del dominio, las relaciones globales dentro de la imagen y la adaptación al problema concreto que se desea resolver.

#### 4.5.2. *EfficientNetB0*

*EfficientNet* es una familia de redes neuronales convolucionales (CNN) cuyo objetivo es lograr un alto rendimiento con menos recursos computacionales en comparación con arquitecturas anteriores. Sus características claves son las siguientes (GeeksforGeeks contributors, 2025a):

1. Tallo (*Stem*): Comienza con una etapa de stem que consiste en una convolución estándar seguida de normalización por lotes (*Batch Normalization*) y activación ReLU6.
2. Cuerpo (*Body*): se compone de una serie de bloques MBConv con diferentes configuraciones (expansión de canales, tamaños de kernel, stride, mecanismo "*squeeze-and-excitation*").

3. Head (*Cabeza*): Incluye un bloque convolutional final para clasificación.
4. Bloques internos: Al igual que *MobileNetV2* usa convoluciones separables en profundidad (*Depthwise Separable Convolutions*) y bloques residuales invertidos (*Inverted Residuals*). Además añade bloques *Squeeze-and-Excitation* para calibrar la importancia de los canales.

A continuación se realiza una división funcional de la red para entender qué aprenden las capas (P., 2023):

- Zona temprana: Desde la entrada (*stem*) hasta los primeros bloques *MBConv*, aproximadamente hasta la capa 20. En esta etapa, la red aprende características muy básicas, como bordes, texturas simples, colores, orientación y gradientes. Se trata de filtros universales que casi cualquier red visual aprende.
- Zona intermedia: Bloques medios de *MBConv* con mayor expansión y mayor profundidad, aproximadamente de la capa 20 a la 100. Aquí se capturan características más complejas, como combinaciones de formas, patrones locales, partes de objetos y texturas específicas (por ejemplo, de hojas o plantas). En este punto, la red comienza a adaptarse a dominios más específicos.
- Zona tardía: corresponde a los últimos bloques, desde la capa 100 hasta la final. Esta zona extrae características de muy alto nivel, relacionadas con la semántica del dominio, las relaciones globales dentro de la imagen y la adaptación al problema concreto que se desea resolver.

#### 4.5.3. *NASNetMobile*

*NASNetMobile* es una versión optimizada para dispositivos móviles de la arquitectura *NASNet* (*Neural Architecture Search Network*) desarrollada por *Google Brain*. Su diseño se basa en la búsqueda automática de arquitectura *NAS* para encontrar estructuras de red eficientes para tareas de clasificación de imágenes de gran escala. Sus características claves son las siguientes (Zoph et al., 2018):

1. La entrada típica es de 224x224 píxeles en el caso móvil.
2. Su diseño emplea dos tipos de células (*cells*): las llamadas células normales (*normal cells*) y células de reducción (*reduction cells*).
3. Las *normal cells* producen mapas de características que conservan la resolución espacial (altura x anchura).
4. Las *reduction cells* reducen la resolución espacial (por ejemplo, con factor 2 en altura y anchura) para avanzar hacia características más globales.

A continuación se realiza una división funcional de la red para entender qué aprenden las capas (TensorFlow developers, 2024):

- Zona temprana: Comprende los primeros bloques, aproximadamente hasta la capa 50. En esta etapa, la red aprende características muy básicas, como bordes, texturas simples, colores, orientación y gradientes. Se trata de filtros universales que casi cualquier red visual aprende.
- Zona intermedia: Incluye los bloques medios, aproximadamente desde la capa 50 hasta la 150. Aquí se capturan características más complejas, como combinaciones de formas, patrones locales, partes de objetos y texturas específicas (por ejemplo, de hojas o plantas). En este punto, la red comienza a adaptarse a dominios más específicos.
- Zona tardía o cabeza del modelo: corresponde a los últimos bloques, desde la capa 150 hasta la final. Esta zona extrae características de muy alto nivel, relacionadas con la semántica del dominio, las relaciones globales dentro de la imagen y la adaptación al problema concreto que se desea resolver.

## 5. Evaluación y resultados

En este capítulo se presentan los resultados obtenidos al usar los datos preprocesados explicado en la sección 4.3 a los modelos descritos en la sección 4.4. Para comenzar, se considera necesario realizar una breve explicación sobre las métricas que ayudan a realizar esta tarea (5.1). Finalmente se realizará una comparación entre los modelos con mejor rendimiento en la Sección 5.10.

### 5.1. Métricas

El rendimiento de los modelos propuestos se evalúan utilizando diferentes métricas como, la precisión, la sensibilidad, *F1-score* y soporte. Las métricas de evaluación se establecen utilizando fórmulas que usan TP, TN, FP y FN, que denotan verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente. De esta manera se definen las siguientes métricas:

- Precisión (*Precision*): Mide la proporción de verdaderos positivos sobre el total de predicciones positivas.

$$\text{Precisión} = \frac{TP}{TP + FP}$$

De cara a la interpretación de lo que nos informa es de todas las imágenes que el modelo predijo como una clase, cuántas realmente pertenecen a esa clase.

- Sensibilidad (*Recall*): mide la proporción de verdaderos positivos detectados.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Nos informa de todas las imágenes que son realmente de una clase, cuántas fueron detectadas correctamente.

- *F1-Score*: Es el promedio armónico entre precisión y sensibilidad.

$$\text{F1-Score} = \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}}$$

Otra métrica útil para la evaluación de los modelos usada es la matriz de confusión. Se trata de una tabla que compara las predicciones del modelo con las etiquetas reales, mostrando en su diagonal principal el número de aciertos (instancias correctamente clasificadas) y fuera de la diagonal los errores (instancias que el modelo confundió con otra clase). Cada fila representa las predicciones reales y cada columna las predicciones del modelo, lo que permite identificar con precisión qué clases son más propensas a ser confundidas entre sí.

## 5.2. Modelo *MobileNetV2* 1

En esta sección se van a mostrar los resultados del entrenamiento del modelo descrito en la subsección 4.4.1. A este modelo se han aplicado varios entrenamientos distintos explicados en las subsecciones 5.2.1, 5.2.2 y 5.2.3.

### 5.2.1. Entrenamiento 1

Este entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Solo se ha sometido a 10 épocas y se han usado los callbacks de *ModelCheckpoint* e *HistorySaver*, ya que el objetivo de este entrenamiento es observar la tendencia inicial para poderlo comparar con el entrenamiento de la sección 5.2.2 y poder determinar qué tanto puede influir los pesos en las clases.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.1.

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.1 en el entrenamiento. A continuación se describe lo que se puede interpretar:

- Precisión de entrenamiento: Comienza alrededor de 0.6 y sube rápidamente a un valor cercano a 0.8, mostrando que el modelo aprende bien de los datos de entrenamiento.
- Pérdida de entrenamiento: Disminuye de un valor cercano a 1 hasta cerca de 0.6, lo que confirma que el modelo está optimizando correctamente con la función de pérdida elegida para los datos de entrenamiento.

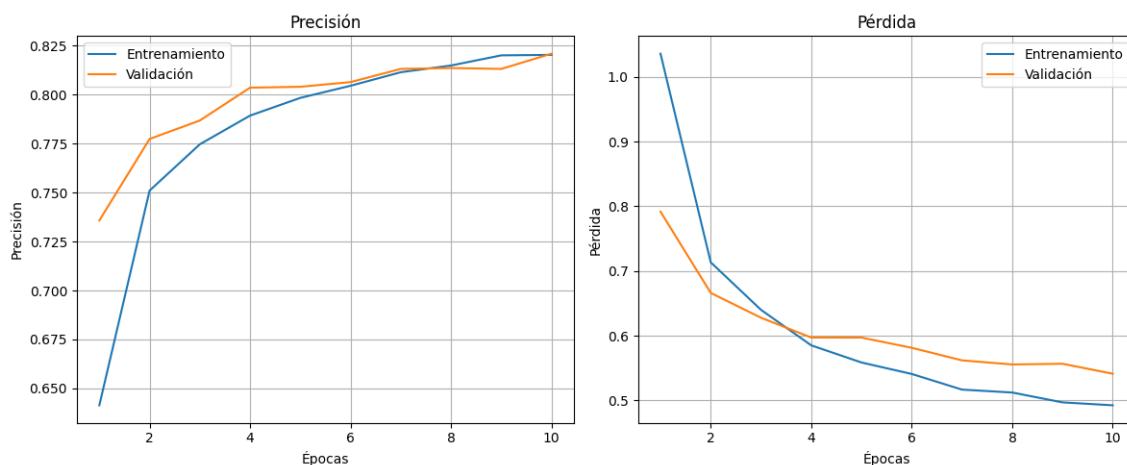


Figura 5.1.: Histórico de métricas del entrenamiento 1 del modelo *MobileNetV2* 1

- Precisión de validación: Aumenta de un valor cercano a 0.7 a un valor cercano a 0.8, lo que muestra que el modelo aprende bien de datos distintos a los de entrenamiento.
- Pérdida de validación: Disminuye de un valor cercano a 0.8 a un valor cercano a 0.5, confirmando que la función de pérdida optimiza bien el modelo.

En resumen, ambas métricas de precisión tienden a subir acercándose a 1 y las métricas de pérdidas a bajar acercándose a 0. Ambas tendencias se realizan de manera progresiva sin señales de *overfitting*.

La precisión de entrenamiento se estabiliza alrededor de 0.8, con incrementos pequeños al final. Esto puede indicar que el modelo se está acercando a su límite de capacidad con la configuración actual. Para exprimir aún más este modelo con intención de mejorarlo se podría aumentar el número de épocas con el callback *EarlyStopping* para que pare el entrenamiento en caso de que no mejore y el callback de *ReduceLROnPlateau* para que reduzca el *learning rate* en el entrenamiento y haga que no se estanquen sus valores.

Para poder comparar este entrenamiento con el de la subsección 5.2.2 es necesario realizar un análisis de métricas por clases. Para ello se van a mostrar las métricas del método *classification\_report* de la librería *Sklearn* en la Figura 5.2 y la matriz de confusión en la Figura 5.3.

Se puede decir que el modelo de clasificación multiclase alcanzó una precisión del 81.4 % y un F1-score medio de 0.81, lo que indica un buen equilibrio entre precisión y capacidad de detección en las once clases evaluadas. Las clases *Tomato\_**Yellow\_Leaf - \_Curl\_Virus*, *Tomato\_mosaic\_virus* y *healthy* destacan por sus altos valores de F1-score (superiores a 0,89), mostrando que el modelo identifica correctamente los patrones

	precision	recall	f1-score	support	
Bacterial_spot	0.800	0.791	0.795	565	
Early_blight	0.780	0.692	0.734	491	
Late_blight	0.818	0.768	0.793	622	
Leaf_Mold	0.801	0.842	0.821	550	
Septoria_leaf_spot	0.757	0.705	0.730	576	
Spider_mites Two-spotted_spider_mite	0.849	0.837	0.843	349	
	Target_Spot	0.647	0.868	0.742	365
Tomato_Yellow_Leaf_Curl_Virus	0.939	0.948	0.944	407	
	Tomato_mosaic_virus	0.948	0.840	0.890	430
	healthy	0.949	0.862	0.904	610
powdery_mildew	0.640	0.935	0.760	200	
accuracy			0.814	5165	
macro avg	0.812	0.826	0.814	5165	
weighted avg	0.823	0.814	0.815	5165	

Figura 5.2.: [Resumen de métricas del modelo MobileNetV2 1 en el entrenamiento 1](#)

más distintivos de estas categorías. Asimismo, el desempeño equilibrado entre las métricas macro y ponderadas (*weighted*) sugiere que no existe un sesgo notable hacia clases con mayor cantidad de muestras.

Sin embargo, algunas clases como Target\_Spot, powdery\_mildew, Early\_blight y Septoria\_leaf\_spot presentan valores más bajos, reflejando cierta confusión con otras enfermedades de síntomas visualmente similares.

De esta manera la matriz de confusión de la Figura 5.3 refuerza las conclusiones obtenidas a partir de las métricas de evaluación, mostrando que el modelo clasifica correctamente la mayoría de las muestras en casi todas las clases. Se observa una fuerte concentración de valores en la diagonal principal, especialmente en las clases Tomato\_Yellow\_Leaf\_Curl\_Virus, Tomato\_mosaic\_virus y healthy. Estas clases presentan patrones visuales distintivos que el modelo ha aprendido de forma eficaz.

En contraste, se aprecian errores recurrentes entre clases con síntomas similares, como Early\_blight, Late\_blight y Septoria\_leaf\_spot donde las confusiones cruzadas son más frecuentes. También se observan predicciones erróneas entre Target\_Spot y otras enfermedades foliares, así como entre powdery\_mildew y Leaf\_Mold, lo que explica los valores más bajos de precisión en estas clases. En conjunto, la matriz de confusión visualiza de manera clara la capacidad del modelo para distinguir la mayoría de las enfermedades, evidenciando a la vez los retos en la diferenciación de patologías con síntomas visualmente parecidos.

Este modelo tiene un comportamiento sólido y equilibrado, generaliza bien y consigue que incluso las clases con menos muestras tengan buenos niveles de recall.

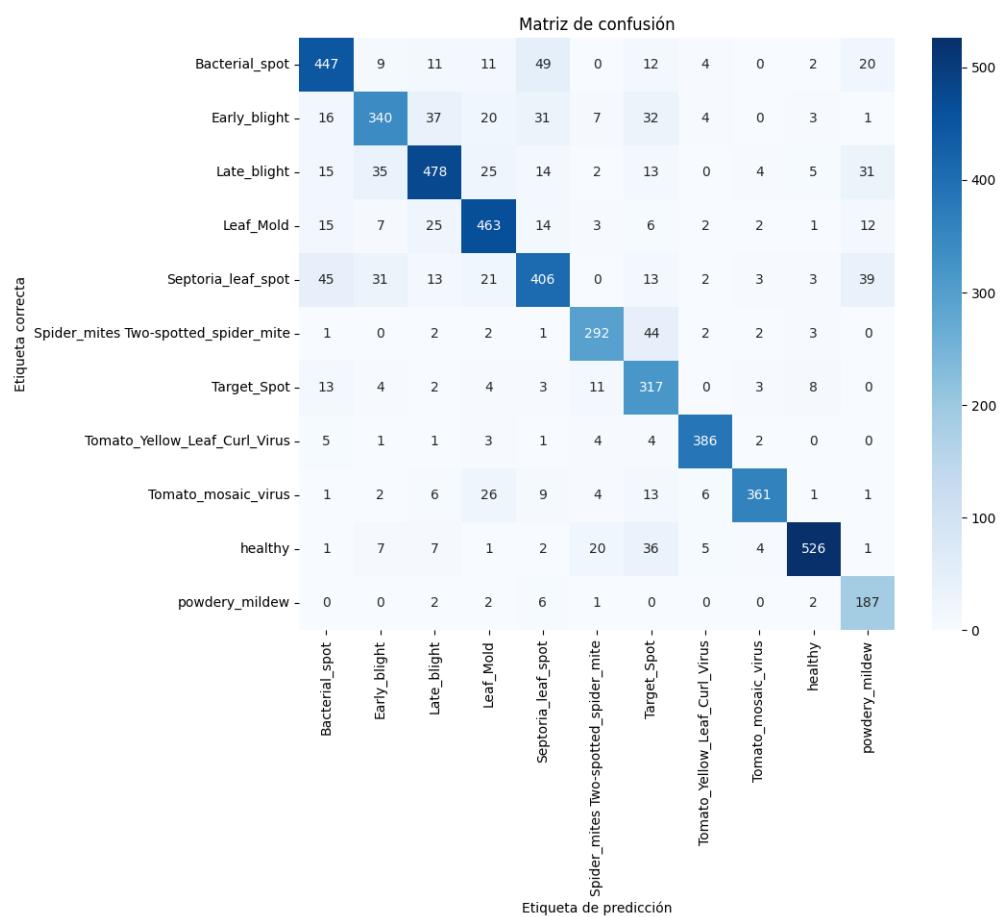


Figura 5.3.: [Matriz de confusión del modelo MobileNetV2 1 en el entrenamiento 1](#)

### 5.2.2. Entrenamiento 2

Este entrenamiento es igual que el anterior, pero sin tener en cuenta los pesos de las clases.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.4.

En la Figura 5.4 podemos observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.1 en el entrenamiento. A continuación se describe lo que se puede interpretar:

- Precisión de entrenamiento: Comienza alrededor de 0.6 y sube rápidamente a un valor cercano a 0.8, mostrando que el modelo aprende bien de los datos de entrenamiento.

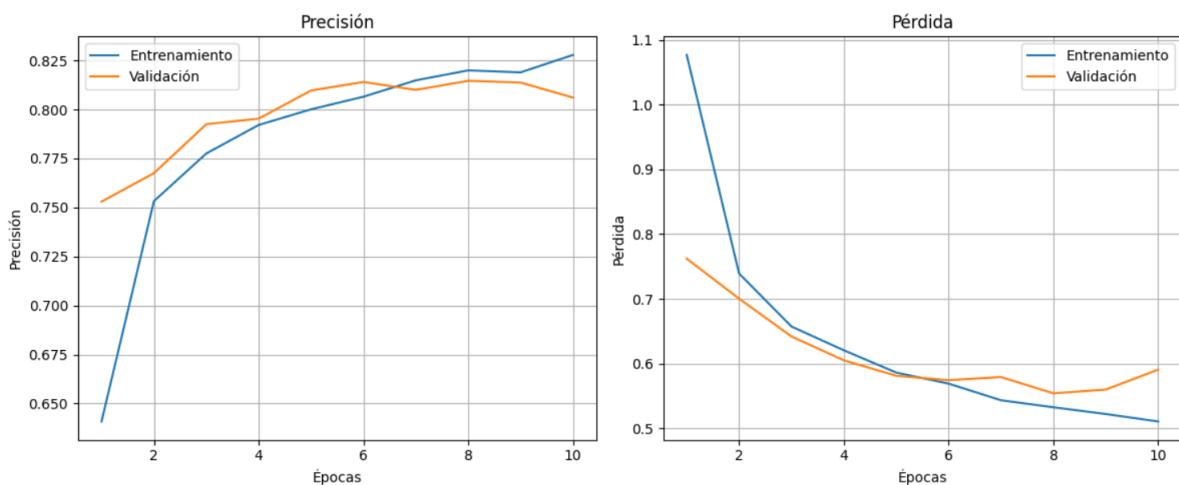


Figura 5.4.: Histórico de métricas del entrenamiento 2 del modelo MobileNetV2 1

- Pérdida de entrenamiento: Disminuye de un valor cercano a 1 hasta cerca de 0.6, lo que confirma que el modelo está optimizando correctamente con la función de pérdida elegida para los datos de entrenamiento.
- Precisión de validación: Aumenta de un valor cercano a 0.7 a un valor cercano a 0.8, lo que muestra que el modelo aprende bien de datos distintos a los de entrenamiento.
- Pérdida de validación: Disminuye de un valor cercano a 0.8 a un valor cercano a 0.5, confirmando que la función de pérdida optimiza bien el modelo.

En resumen, ambas métricas de precisión tienden a subir acercándose a 1 y las métricas de pérdidas a bajar acercándose a 0. Ambas tendencias se realizan de manera progresiva sin señales de *overfitting*.

El rendimiento global es prácticamente el mismo que el conseguido con el entrenamiento de la subsección 5.2.1, por lo que para poder compararlos más detalladamente es necesario realizar un análisis de métricas por clases. Para ello se van a mostrar las métricas del método `classification_report` de la librería `Sklearn` en la Figura 5.5 y la matriz de confusión en la Figura 5.6.

El modelo ha alcanzado una precisión del 81 % y un F1-score medio de 0.81, lo que indica un buen equilibrio entre precisión y capacidad de detección en las once clases evaluadas. Las clases `Tomato_Leaf_Curl_Virus`, `Tomato_mosaic_virus` y `healthy` siguen destacando por sus altos valores de F1-score (superiores a 0.89), mostrando que el modelo identifica correctamente los patrones más distintivos de estas clases. Asimismo, el desempeño equilibrado entre las métricas macro y ponderadas (*weighted*) sugiere que no existe un sesgo notable hacia clases con mayor cantidad de muestras.

	precision	recall	f1-score	support
Bacterial_spot	0.742	0.851	0.793	565
Early_blight	0.757	0.684	0.719	491
Late_blight	0.763	0.793	0.778	622
Leaf_Mold	0.717	0.878	0.789	550
Septoria_leaf_spot	0.824	0.611	0.702	576
Spider_mites Two-spotted_spider_mite	0.928	0.705	0.801	349
Target_Spot	0.712	0.811	0.758	365
Tomato_Yellow_Leaf_Curl_Virus	0.997	0.912	0.953	407
Tomato_mosaic_virus	0.905	0.905	0.905	430
healthy	0.875	0.928	0.901	610
powdery_mildew	0.821	0.805	0.813	200
accuracy			0.808	5165
macro avg	0.822	0.807	0.810	5165
weighted avg	0.815	0.808	0.807	5165

Figura 5.5.: [Resumen de métricas del modelo MobileNetV2 1 en el entrenamiento 2](#)

Sin embargo, algunas clases como Target\_Spot, powdery\_mildew, Early\_blight y Septoria\_leaf\_spot presentan valores más bajos, reflejando cierta confusión con otras enfermedades de síntomas visualmente similares.

La matriz de confusión de La Figura 5.6 refuerza las conclusiones obtenidas a partir de las métricas de evaluación, mostrando que el modelo clasifica correctamente la mayoría de las muestras en casi todas las clases. Se observa una fuerte concentración de valores en la diagonal principal, especialmente en las clases Tomato\_Yellow\_Leaf\_Curl\_Virus, Tomato\_mosaic\_virus y healthy. Estas clases presentan patrones visuales distintivos que el modelo ha aprendido de forma eficaz.

En contraste, se aprecian errores recurrentes entre clases con síntomas similares, como Early\_blight, Late\_blight y Septoria\_leaf\_spot donde las confusiones cruzadas son más frecuentes. También se observan predicciones erróneas entre Target\_Spot y otras enfermedades foliares, así como entre powdery\_mildew y Leaf\_Mold, lo que explica los valores más bajos de precisión en estas clases. En conjunto, la matriz de confusión visualiza de manera clara la capacidad del modelo para distinguir la mayoría de las enfermedades, evidenciando a la vez los retos en la diferenciación de patologías con síntomas visualmente parecidos.

En general se obtiene un buen rendimiento, existiendo un buen balance entre precisión y sensibilidad en prácticamente todas las clases. Aún así, se pueden observar diferencias en algunas clases con respecto al modelo resultante de la subsección 5.2.1. En concreto se puede determinar una disminución en la métrica de sensibilidad en en clases menos representadas en el *dataset* como las clases powdery\_mildew (de 0.93 a 0.80), Septoria\_leaf\_spot (de 0.71 a 0.61) y Spidermite\_Two-spotted\_spider\_mite (de 0.84 a 0.70). Esto demuestra la función de los pesos en las clases, es decir el balanceo de las mismas.

El modelo obtenido en la subsección 5.2.1 equilibra mejor los resultados sin sacrificar

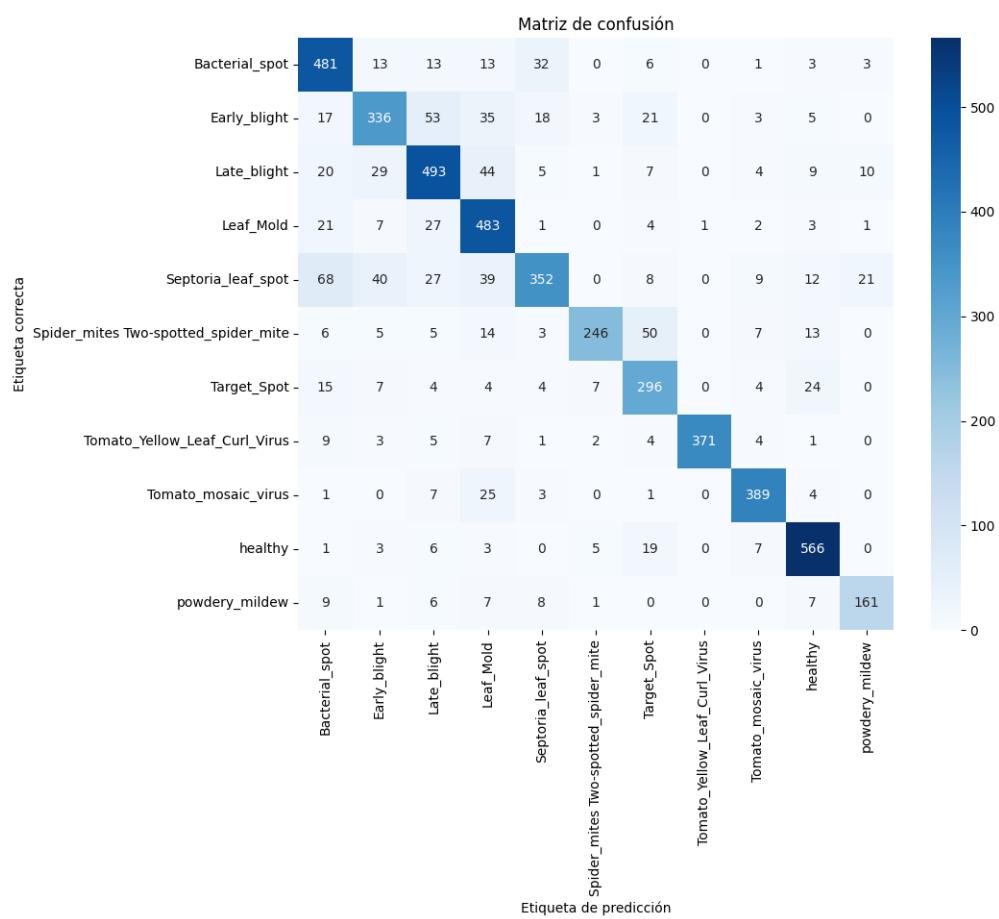


Figura 5.6.: [Matriz de confusión del modelo MobileNetV2 1 en el entrenamiento 2](#)

precisión general, por lo que es la opción más sólida para un sistema de diagnóstico de enfermedades, donde todas las clases tienen importancia clínica. Por tanto el modelo que se seguirá mejorando en la suubsección subsección 5.2.3

### 5.2.3. Entrenamiento 3

Este entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Se ha sometido a 20 épocas y además de los callbacks de *ModelCheckpoint* e *HistorySaver*, se han usado también los callbacks de *ReduceLROnPlateau* y *EarlyStopping*.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.7.

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.1 en el entrenamiento. A continuación se

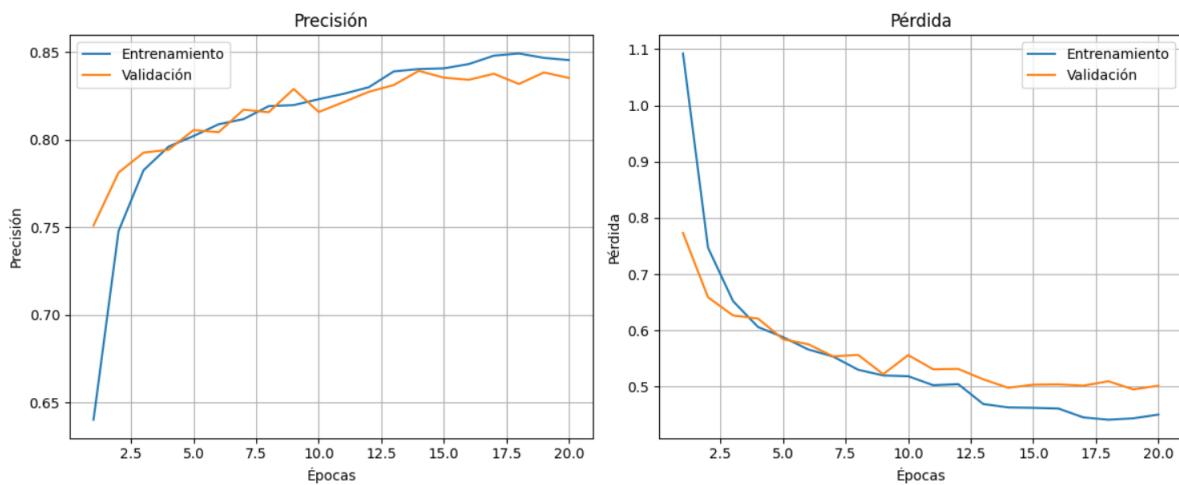


Figura 5.7.: Histórico de métricas del entrenamiento 3 del modelo MobileNetV2 1

describe lo que se puede interpretar:

- Precisión de entrenamiento: Comienza alrededor de 0.65 y sube rápidamente a un valor cercano a 0.85, mostrando que el modelo aprende bien de los datos de entrenamiento.
- Pérdida de entrenamiento: Disminuye de un valor cercano a 1.1 hasta cerca de 0.5, lo que confirma que el modelo está optimizando correctamente con la función de pérdida elegida para los datos de entrenamiento.
- Precisión de validación: Aumenta de un valor cercano a 0.75 a un valor cercano a 0.85, lo que muestra que el modelo aprende bien de datos distintos a los de entrenamiento.
- Pérdida de validación: Disminuye de un valor cercano a 0.8 a un valor cercano a 0.5, confirmando que la función de pérdida optimiza bien el modelo.

En resumen, ambas métricas de precisión tienden a subir acercándose a 1 y las métricas de pérdidas a bajar acercándose a 0. Ambas tendencias se realizan de manera progresiva sin señales de *overfitting*. Cabe destacar que el callback de ReduceLROnPlateau se ha activado en la época 13 reduciendo el learning rate de 1e-3 a 5e-4 y en la época 18 reduciendo el learning rate de 5e-4 a 2.5e-4, explicando la estabilización que existe.

Para poder comparar este entrenamiento con el resto es necesario realizar un análisis de métricas por clases. Para ello se van a mostrar las métricas del método `classification_report` de la librería *Sklearn* en la Figura 5.8 y la matriz de confusión en la Figura 5.9.

El modelo ha alcanzado una precisión del 83.1 % y un F1-score medio de 0.83, lo que

	precision	recall	f1-score	support
Bacterial_spot	0.803	0.809	0.806	565
Early_blight	0.827	0.711	0.765	491
Late_blight	0.813	0.817	0.815	622
Leaf_Mold	0.800	0.844	0.821	550
Septoria_leaf_spot	0.751	0.724	0.737	576
Spider_mites Two-spotted_spider_mite	0.903	0.822	0.861	349
Target_Spot	0.702	0.838	0.764	365
Tomato_Yellow_Leaf_Curl_Virus	0.992	0.929	0.959	407
Tomato_mosaic_virus	0.929	0.907	0.918	430
healthy	0.864	0.936	0.899	610
powdery_mildew	0.838	0.830	0.834	200
accuracy			0.831	5165
macro avg	0.838	0.833	0.834	5165
weighted avg	0.834	0.831	0.831	5165

Figura 5.8.: [Resumen de métricas del modelo MobileNetV2 1 en el entrenamiento 3](#)

indica un buen equilibrio entre precisión y capacidad de detección en las once clases evaluadas. Las clases Tomato\_Yellow\_Leaf\_Curl\_Virus, Tomato\_mosaic\_virus y healthy siguen destacando por sus altos valores de F1-score (superiores a 0,90), mostrando que el modelo identifica correctamente los patrones más distintivos de estas clases. Asimismo, el desempeño equilibrado entre las métricas macro y ponderadas (*weighted*) sugiere que no existe un sesgo notable hacia clases con mayor cantidad de muestras.

Sin embargo, algunas clases como Target\_Spot, Early\_blight y Septoria\_leaf-spot presentan valores más bajos, reflejando cierta confusión con otras enfermedades de síntomas visualmente similares.

La matriz de confusión de la Figura 5.9 refuerza las conclusiones obtenidas a partir de las métricas de evaluación, mostrando que el modelo clasifica correctamente la mayoría de las muestras en casi todas las clases. Se observa una fuerte concentración de valores en la diagonal principal, especialmente en las clases Tomato\_Yellow\_Leaf\_Curl\_Virus, Tomato\_mosaic\_virus y healthy. Estas clases presentan patrones visuales distintivos que el modelo ha aprendido de forma eficaz.

En contraste, se aprecian errores recurrentes entre clases con síntomas similares, como Early\_blight, Late\_blight, Septoria\_leaf\_spot y Target\_Spot donde las confusiones cruzadas son más frecuentes. En conjunto, la matriz de confusión visualiza de manera clara la capacidad del modelo para distinguir la mayoría de las enfermedades, evidenciando a la vez los retos en la diferenciación de patologías con síntomas visualmente parecidos.

Este modelo tiene un comportamiento sólido y equilibrado, generaliza bien y consigue que incluso las clases con menos muestras tengan buenos niveles de precisión y sensibilidad. Para seguir refinando este modelo se descongelarán algunas capas del modelo base (*fine tuning*) y se reducirá el *learning rate*, estos cambios se aplican en el modelo de la Subsección 4.4.2.

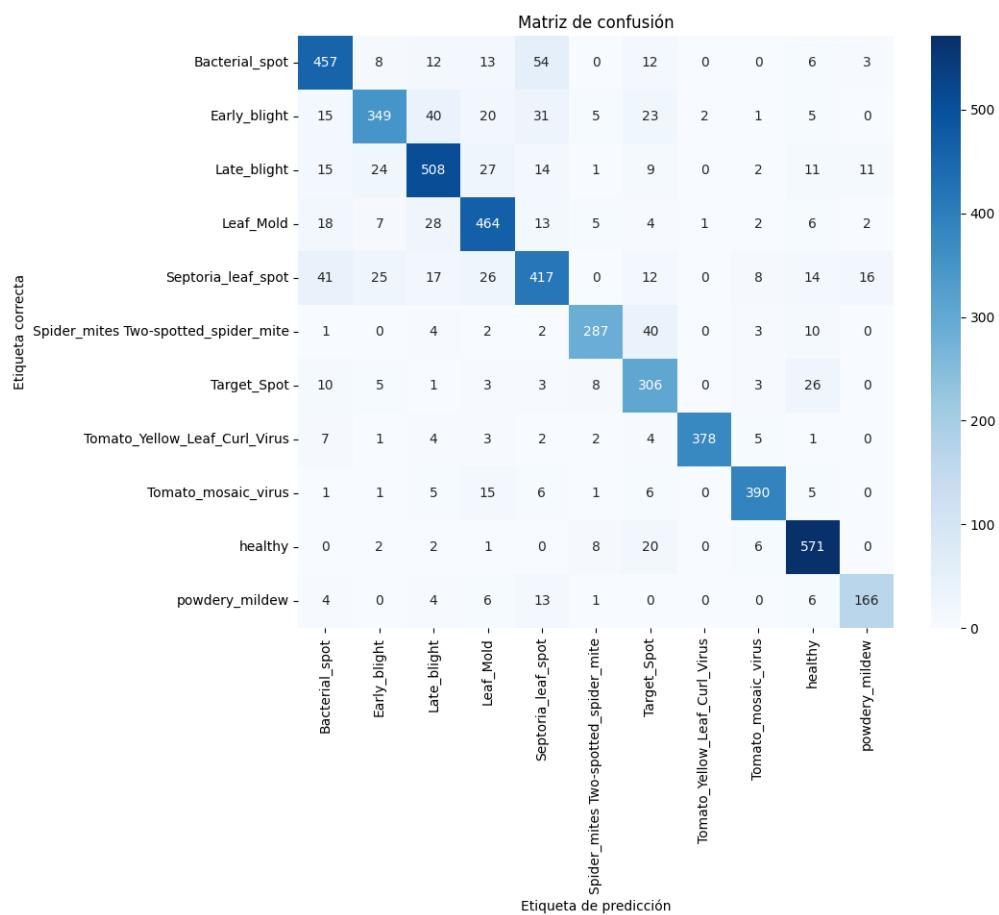


Figura 5.9.: [Matriz de confusión del modelo MobileNetV2 1 en el entrenamiento 3](#)

### 5.3. Modelo **MobileNetV2 2**

En esta sección se explicarán los resultados obtenidos al entrenar el modelo explicado en la Subsección 4.4.2. El entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Se ha sometido a 20 épocas y además de los callbacks de *ModelCheckpoint* e *HistorySaver*, se han usado también los callbacks de *ReduceLROnPlateau* y *EarlyStopping*.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.10.

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.2 en el entrenamiento. A continuación se describe lo que se puede interpretar:

- Precisión de entrenamiento: Comienza alrededor de 0.5 y sube rápidamente a un

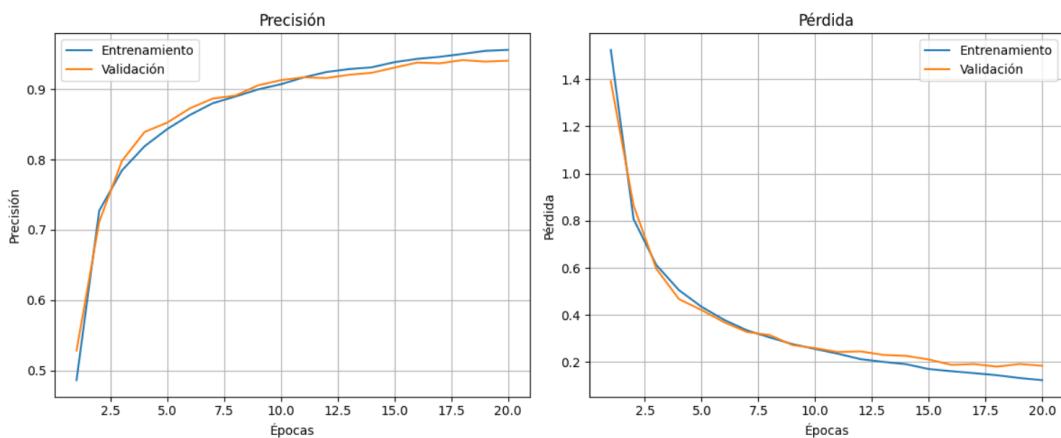


Figura 5.10.: [Histórico de métricas del entrenamiento del modelo MobileNetV2 2](#)

valor cercano a 0.9, mostrando que el modelo aprende bien de los datos de entrenamiento.

- Pérdida de entrenamiento: Disminuye de un valor cercano a 1.4 hasta cerca de 0.2, lo que confirma que el modelo está optimizando correctamente con la función de pérdida elegida para los datos de entrenamiento.
- Precisión de validación: Aumenta de un valor próximo a 0.5 a un valor cercano a 0.9, lo que muestra que el modelo aprende bien de datos distintos a los de entrenamiento.
- Pérdida de validación: Disminuye de un valor cercano a 1.4 a un valor próximo a 0.2, confirmando que la función de pérdida optimiza bien el modelo.

En resumen, ambas métricas de precisión tienden a subir acercándose a 1 y las métricas de pérdidas a bajar acercándose a 0. Ambas tendencias se realizan de manera progresiva sin señales de *overfitting*.

En términos generales se tiene un buen modelo, pero para poder comparar con el resto es necesario realizar un análisis de métricas por clases. Para ello se van a mostrar las métricas del método `classification_report` de la librería *Sklearn* en la Figura 5.11 y la matriz de confusión en la Figura 5.12.

El modelo ha alcanzado una precisión del 94 % y un F1-score medio de 0.93, lo que indica un buen equilibrio entre precisión y capacidad de detección en las once clases evaluadas. Las clases `Tomato_Yellow_Leaf_Curl_Virus`, `Tomato_mosaic_virus` y `healthy` siguen destacando por sus altos valores de F1-score (superiores a 0.95), mostrando que el modelo identifica correctamente los patrones más distintivos de estas clases. Asimismo, el desempeño equilibrado entre las métricas macro y ponderadas (*weighted*) sugiere que no existe un sesgo notable hacia clases con mayor cantidad de muestras.

	precision	recall	f1-score	support
Bacterial_spot	0.981	0.908	0.943	565
Early_blight	0.932	0.835	0.881	491
Late_blight	0.909	0.949	0.928	622
Leaf_Mold	0.966	0.967	0.966	550
Septoria_leaf_spot	0.918	0.913	0.916	576
Spider_mites Two-spotted_spider_mite	0.981	0.880	0.927	349
Target_Spot	0.808	0.967	0.880	365
Tomato_Yellow_Leaf_Curl_Virus	0.997	0.973	0.985	407
Tomato_mosaic_virus	0.974	0.974	0.974	430
healthy	0.948	0.989	0.968	610
powdery_mildew	0.921	0.995	0.957	200
accuracy			0.939	5165
macro avg	0.940	0.941	0.939	5165
weighted avg	0.941	0.939	0.939	5165

Figura 5.11.: Resumen de métricas del modelo *MobileNetV2 2*

Sin embargo, algunas clases como *Target\_Spot* y *Early\_blight* presentan valores más bajos, reflejando cierta confusión con otras enfermedades de síntomas visualmente similares.

La matriz de confusión de la Figura 5.12 refuerza las conclusiones obtenidas a partir de las métricas de evaluación, mostrando que el modelo clasifica correctamente la mayoría de las muestras en casi todas las clases. Se observa una fuerte concentración de valores en la diagonal principal, especialmente en las clases *Tomato\_Yellow\_Leaf\_Curl\_Virus*, *Tomato\_mosaic\_virus* y *healthy*. Estas clases presentan patrones visuales distintivos que el modelo ha aprendido de forma eficaz.

En contraste, se aprecian errores recurrentes entre clases con síntomas similares, como *Early\_blight*, *Septoria\_leaf\_spot* y *Target\_Spot* donde las confusiones cruzadas son más frecuentes. En conjunto, la matriz de confusión visualiza de manera clara la capacidad del modelo para distinguir la mayoría de las enfermedades, evidenciando a la vez los retos en la diferenciación de patologías con síntomas visualmente parecidos.

Este modelo tiene un comportamiento sólido y equilibrado, generaliza bien y consigue que incluso las clases con menos muestras tengan buenos niveles de precisión y sensibilidad. Para seguir refinando este modelo se descongelarán más capas del modelo base y se reducirá el *learning rate*, estos cambios se aplican en el modelo de la Subsección 4.4.3.

## 5.4. Modelo *MobileNetV2 3*

En esta sección se explicarán los resultados obtenidos al entrenar el modelo explicado en la Subsección 4.4.3. El entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Se ha sometido a 20 épocas y además de los callbacks de

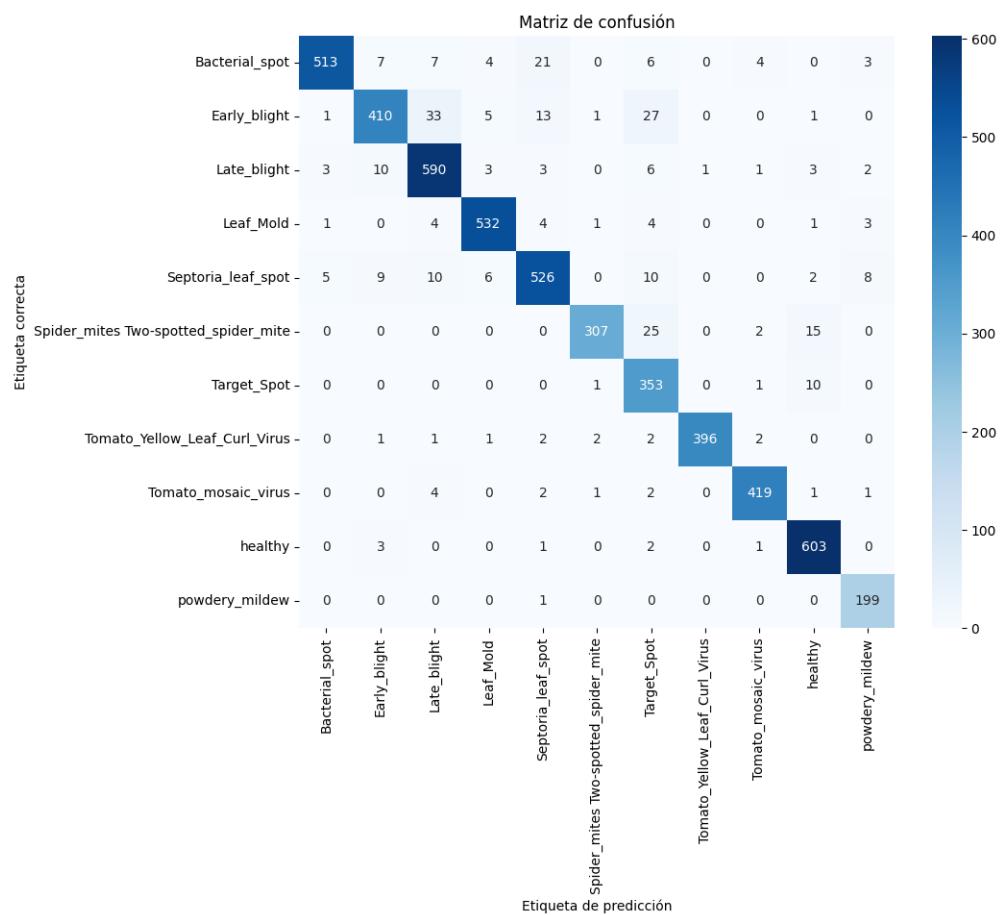


Figura 5.12.: [Matriz de confusión del modelo MobileNetV2 2](#)

*ModelCheckpoint* e *HistorySaver*, se han usado también los callbacks de *ReduceLROnPlateau* y *EarlyStopping*.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.10.

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.2 en el entrenamiento. A continuación se describe lo que se puede interpretar:

- Precisión de entrenamiento: Comienza alrededor de 0.5 y sube rápidamente a un valor cercano a 0.9, mostrando que el modelo aprende bien de los datos de entrenamiento.
- Pérdida de entrenamiento: Disminuye de un valor cercano a 1.75 hasta cerca de 0.25, lo que confirma que el modelo está optimizando correctamente con la función de pérdida elegida para los datos de entrenamiento.

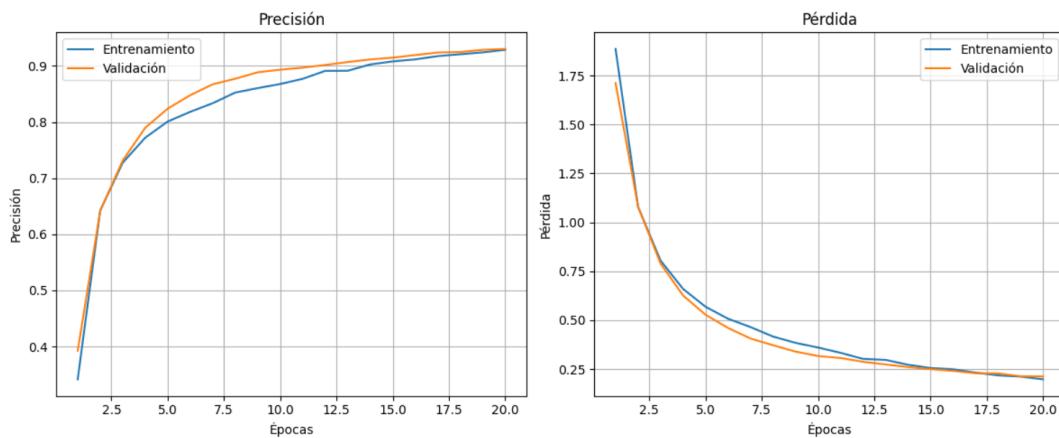


Figura 5.13.: Histórico de métricas del entrenamiento del modelo MobileNetV2 3

- Precisión de validación: Aumenta de un valor próximo a 0.5 a un valor cercano a 0.9, lo que muestra que el modelo aprende bien de datos distintos a los de entrenamiento.
- Pérdida de validación: Disminuye de un valor cercano a 1.75 a un valor próximo a 0.25, confirmando que la función de pérdida optimiza bien el modelo.

En resumen, ambas métricas de precisión tienden a subir acercándose a 1 y las métricas de pérdidas a bajar acercándose a 0. Ambas tendencias se realizan de manera progresiva sin señales de *overfitting*.

En términos generales se tiene un buen modelo, pero para poder comparar con el resto es necesario realizar un análisis de métricas por clases. Para ello se van a mostrar las métricas del método `classification_report` de la librería *Sklearn* en la Figura 5.14 y la matriz de confusión en la Figura 5.15.

	precision	recall	f1-score	support
Bacterial_spot	0.956	0.878	0.915	565
Early_blight	0.917	0.900	0.909	491
Late_blight	0.887	0.929	0.907	622
Leaf_Mold	0.933	0.938	0.936	550
Septoria_leaf_spot	0.884	0.837	0.860	576
Spider_mites Two-spotted_spider_mite	0.980	0.854	0.913	349
	0.795	0.975	0.876	365
Target_Spot	0.997	0.973	0.985	407
Tomato_Yellow_Leaf_Curl_Virus	0.979	0.977	0.978	430
Tomato_mosaic_virus	0.963	0.974	0.968	610
healthy	0.881	0.965	0.921	200
powdery_mildew				
accuracy			0.924	5165
macro avg	0.925	0.927	0.924	5165
weighted avg	0.927	0.924	0.924	5165

Figura 5.14.: Resumen de métricas del modelo MobileNetV2 3

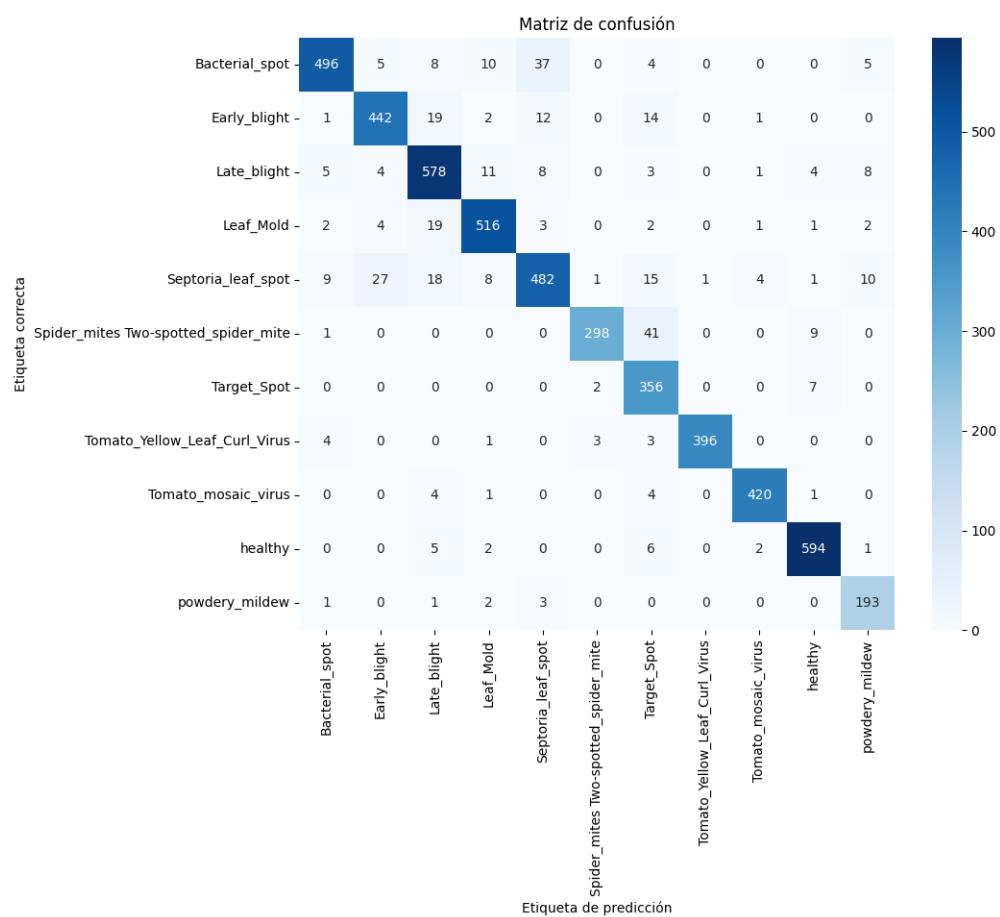


Figura 5.15.: [Matriz de confusión del modelo MobileNetV2 3](#)

El modelo ha alcanzado una precisión del 92 % y un F1-score medio de 0.92, lo que indica un buen equilibrio entre precisión y capacidad de detección en las once clases evaluadas. Las clases Tomato\_Yellow\_Leaf\_Curl\_Virus, Tomato\_mosaic\_virus y healthy siguen destacando por sus altos valores de F1-score (superiores a 0,95), mostrando que el modelo identifica correctamente los patrones más distintivos de estas clases. Asimismo, el desempeño equilibrado entre las métricas macro y ponderadas (*weighted*) sugiere que no existe un sesgo notable hacia clases con mayor cantidad de muestras.

Sin embargo, algunas clases como Target\_Spot y Septoria\_leaf\_spot presentan valores más bajos, reflejando cierta confusión con otras enfermedades de síntomas visualmente similares.

La matriz de confusión de la Figura 5.15 refuerza las conclusiones obtenidas a partir de las métricas de evaluación, mostrando que el modelo clasifica correctamente la mayoría de las muestras en casi todas las clases. Se observa una fuerte concentración de valores en la diagonal principal, especialmente en las clases Tomato\_Yellow\_Leaf\_Curl\_Virus, Tomato\_mosaic\_virus y healthy. Estas clases presentan patrones visuales distintivos

que el modelo ha aprendido de forma eficaz.

En contraste, se aprecian errores recurrentes entre clases con síntomas similares, como *Septoria\_leaf\_spot* y *Target\_Spot* donde las confusiones cruzadas son más frecuentes. En conjunto, la matriz de confusión visualiza de manera clara la capacidad del modelo para distinguir la mayoría de las enfermedades, evidenciando a la vez los retos en la diferenciación de patologías con síntomas visualmente parecidos.

Este modelo tiene un comportamiento sólido y equilibrado, generaliza bien y consigue que incluso las clases con menos muestras tengan buenos niveles de precisión y sensibilidad.

En este modelo se han desbloqueado más capas del modelo base dando más libertad para aprender características específicas (como texturas finas de enfermedades). Pero también rompe parcialmente la estructura previa de ImageNet, reduciendo la estabilidad de las características más generales. Eso demuestra que se observen mejoras en alguna clase, pero pérdida de rendimiento general.

## 5.5. Modelo *EfficientNetB0* 1

En esta sección se van a mostrar los resultados del entrenamiento del modelo descrito en la subsección 4.4.4. Este entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Solo se ha sometido a 10 épocas y se han usado los callbacks de *ModelCheckpoint* e *HistorySaver*, ya que el objetivo de este entrenamiento es observar la tendencia inicial.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.16.

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.4 en el entrenamiento. A continuación se describe lo que se puede interpretar:

- Precisión de entrenamiento: Queda constante en aproximadamente 0.09, un valor muy lejano a 1, lo que demuestra que el modelo no aprende de los datos de entrenamiento.
- Pérdida de entrenamiento: Queda prácticamente constante en un valor cercano a 2.4, bastante lejano a 0.
- Precisión de validación: Oscila entre valores cercanos a 0.1, por lo que el modelo tampoco generaliza.

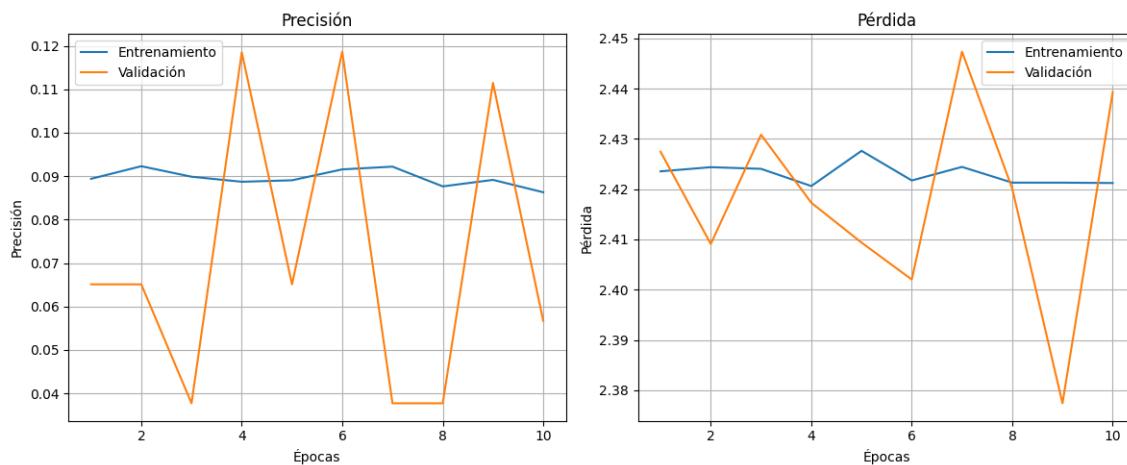


Figura 5.16.: [Histórico de métricas del del modelo EfficientNetB0 1](#)

- Pérdida de validación: Oscila entre valores cercanos a 2.4, bastante lejano a 0.

El modelo no está aprendiendo absolutamente nada más allá del azar. Para mejorar esta situación se descongelarán algunas capas superiores del modelo base *EfficientNetB0*, haciendo uso de la técnica de *fine tuning*.

## 5.6. Modelo *EfficientNetB0 2*

En esta sección se van a mostrar los resultados del entrenamiento del modelo descrito en la subsección 4.4.5. Este entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Solo se ha sometido a 10 épocas y se han usado los callbacks de *ModelCheckpoint* e *HistorySaver*, ya que el objetivo de este entrenamiento es observar la tendencia inicial.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.17.

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.5 en el entrenamiento. A continuación se describe lo que se puede interpretar:

- Precisión de entrenamiento: Aumenta desde un valor aproximado a 0.2 a un valor aproximado a 0.5, lo que demuestra que el modelo está aprendiendo de los datos de entrenamiento.
- Pérdida de entrenamiento: Disminuye de un valor cercano a 2.2 a un valor aproxi-

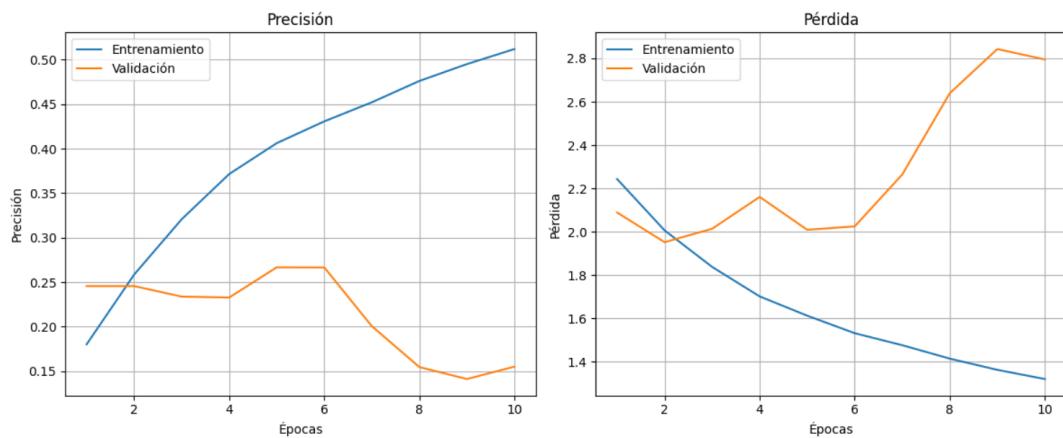


Figura 5.17.: Histórico de métricas del entrenamiento del modelo *EfficientNetB0 2*

mado a 1.4, mostrando una tendencia decreciente.

- Precisión de validación: Decrece de un valor cercano a 0.25 a un valor cercano a 0.15, por lo que el modelo no generaliza.
- Pérdida de validación: Aumenta desde un valor aproximado a 2.2 a un valor aproximado a 2.8, bastante lejano a 0.

En definitiva, se puede decir que el modelo está aprendiendo a memorizar el conjunto de entrenamiento, pero pierde capacidad de generalización, es decir tiene sobreajuste. Para mejorar esta situación, se va a disminuir el *learning rate* en el modelo de la sección 4.4.6

## 5.7. Modelo *EfficientNetB0 3*

En esta sección se van a mostrar los resultados del entrenamiento del modelo descrito en la subsección 4.4.6. Este entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Solo se ha sometido a 10 épocas y se han usado los callbacks de *ModelCheckpoint* e *HistorySaver*, ya que el objetivo de este entrenamiento es observar la tendencia inicial.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.18.

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.6 en el entrenamiento. A continuación se describe lo que se puede interpretar:

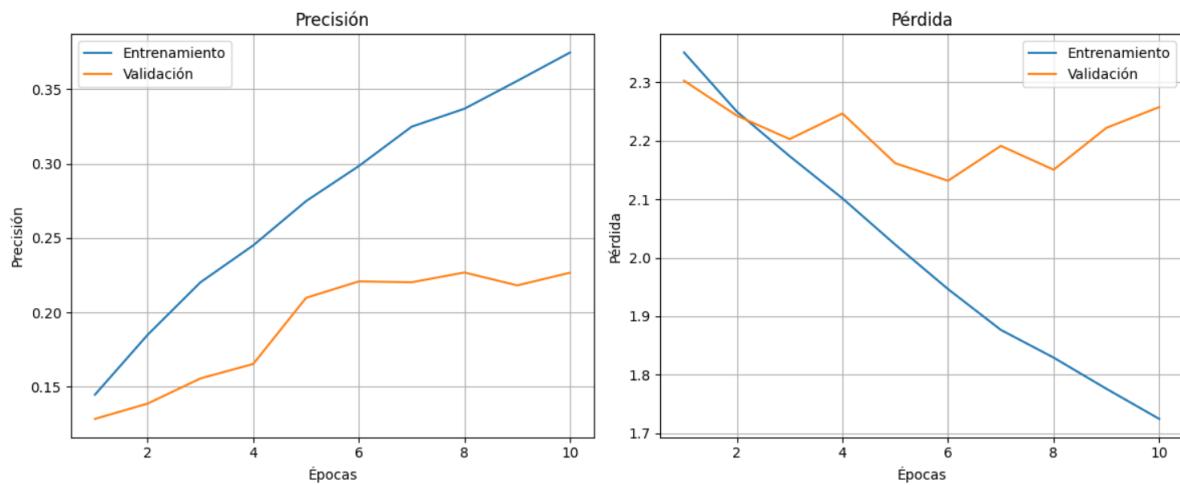


Figura 5.18.: Histórico de métricas del entrenamiento del modelo *EfficientNetB0 1*

- Precisión de entrenamiento: Aumenta desde un valor aproximado a 0.15 a un valor aproximado a 0.35, lo que demuestra que el modelo está aprendiendo de los datos de entrenamiento.
- Pérdida de entrenamiento: Disminuye de un valor cercano a 2.3 a un valor aproximado a 1.8, mostrando una tendencia descendente.
- Precisión de validación: Aumenta de un valor cercano a 0.15 a un valor cercano a 0.25, por lo que el está empezando a generalizar.
- Pérdida de validación: Oscila en un valor cercano a 2.2, por lo que no existe una generalización clara.

En definitiva, se puede decir que se ha disminuido el sobreajuste, el modelo ha comenzado a generalizar, pero no lo suficiente y con la suficiente estabilidad para considerarlo un buen resultado. Para afinar más este modelo se podría incluir una capa de *dropout* antes de la capa densa y entrenar más épocas con los callbacks de *ReduceLROnPlateau* y *EarlyStopping*. Aún así, no se esperan excelentes resultados, simplemente una mejora en la regularización.

## 5.8. Modelo *NASNetMobile 1*

En esta sección se van a mostrar los resultados del entrenamiento del modelo descrito en la subsección 4.4.7. Este entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Solo se ha sometido a 15 épocas y se han usado los callbacks de *ModelCheckpoint* e *HistorySaver*, ya que el objetivo de este entrenamiento es

observar la tendencia inicial.

Para mostrar los resultados de este entrenamiento se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.19.

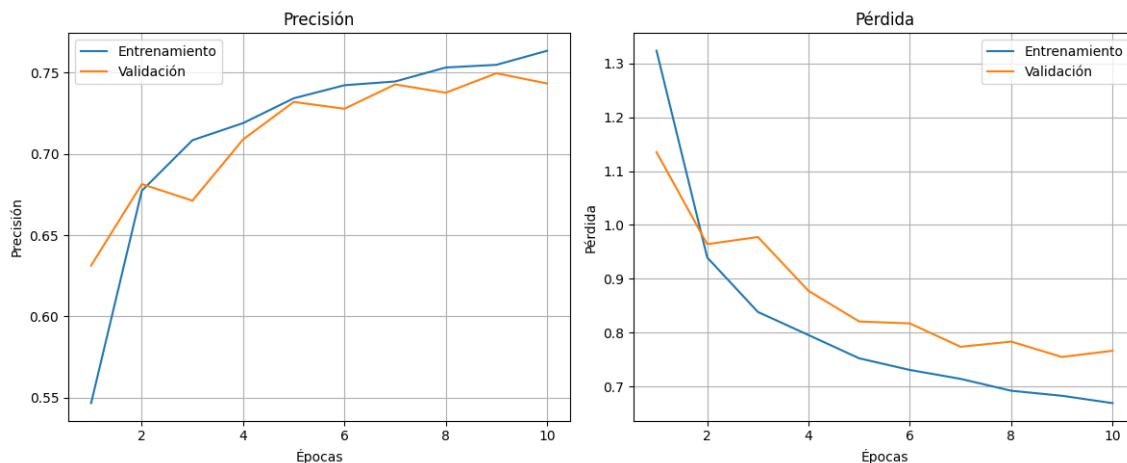


Figura 5.19.: [Histórico de métricas del entrenamiento del modelo NASNetMobile 1](#)

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.1 en el entrenamiento. A continuación se describe lo que se puede interpretar:

- Precisión de entrenamiento: Comienza alrededor de 0.5 y sube rápidamente a un valor cercano a 0.7, mostrando que el modelo aprende bien de los datos de entrenamiento.
- Pérdida de entrenamiento: Disminuye de un valor cercano a 1.3 hasta cerca de 0.7, lo que confirma que el modelo está optimizando correctamente con la función de pérdida elegida para los datos de entrenamiento.
- Precisión de validación: Aumenta de un valor cercano a 0.65 a un valor cercano a 0.75, lo que muestra que el modelo aprende bien de datos distintos a los de entrenamiento.
- Pérdida de validación: Disminuye de un valor cercano a 1.1 a un valor cercano a 0.8, confirmando que la función de pérdida optimiza bien el modelo.

El modelo mejora rápidamente durante las primeras 5 épocas. A partir de ese punto, tanto la precisión de entrenamiento como la de validación se estabilizan alrededor del 75 %, lo cual indica que el modelo ha aprendido bien las características generales, pero está alcanzando su límite.

Para mejorar este resultado se podría entrenar más épocas añadiendo los callbacks *EarlyStopping* para que pare la ejecución si el modelo no mejora y con *ReduceLROnPlateau* para que disminuya el *learning rate* en el entrenamiento y haga que no se estanquen sus valores.

## 5.9. Modelo *NASNetMobile* 2

En esta sección se van a mostrar los resultados del entrenamiento del modelo descrito en la subsección 4.4.8. Este entrenamiento se ha realizado usando los pesos de clases descritos en el Cuadro 4.4. Se ha sometido a 10 épocas y se han usado los callbacks de *ModelCheckpoint* e *HistorySaver*.

Para mostrar los resultados del mismo se va a comenzar mostrando las métricas recogidas en el fichero generado por el callback *HistorySaver* en la Figura 5.20.

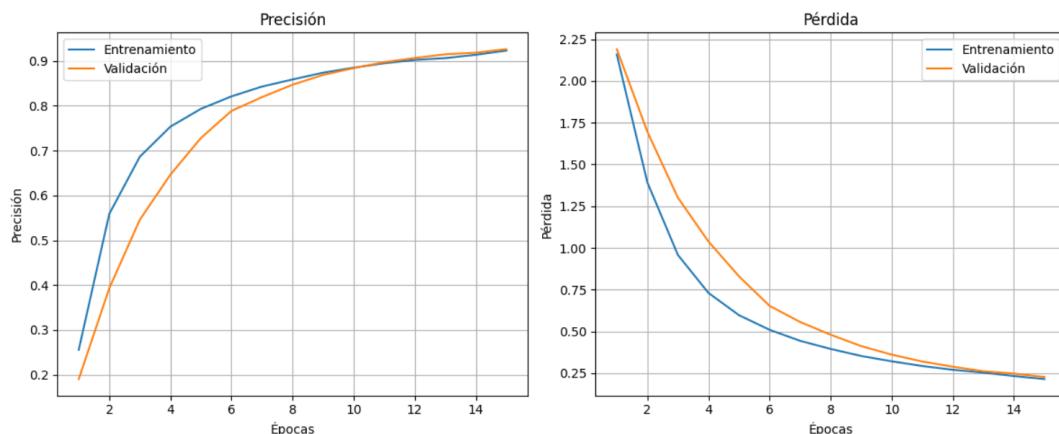


Figura 5.20.: [Histórico de métricas del entrenamiento del modelo \*NASNetMobile\* 2](#)

En esta figura se puede observar el comportamiento de las métricas de precisión y pérdida de entrenamiento y validación a lo largo de las épocas a las que se ha sometido el modelo descrito en la subsección 4.4.8 en el entrenamiento. A continuación se describe lo que se puede interpretar:

- Precisión de entrenamiento: Comienza alrededor de 0.2 y sube rápidamente a un valor cercano a 0.9, mostrando que el modelo aprende bien de los datos de entrenamiento.
- Pérdida de entrenamiento: Disminuye de un valor cercano a 2 hasta cerca de 0.25, lo que confirma que el modelo está optimizando correctamente con la función de pérdida elegida para los datos de entrenamiento.
- Precisión de validación: Aumenta de un valor próximo a 0.2 a un valor cercano a

0.9, lo que muestra que el modelo aprende bien de datos distintos a los de entrenamiento.

- Pérdida de validación: Disminuye de un valor cercano a 2 a un valor próximo a 0.25, confirmando que la función de pérdida optimiza bien el modelo.

En resumen, ambas métricas de precisión tienden a subir acercándose a 1 y las métricas de pérdidas a bajar acercándose a 0. Ambas tendencias se realizan de manera progresiva sin señales de *overfitting*.

En términos generales se tiene un buen modelo, pero para poder comparar con el resto es necesario realizar un análisis de métricas por clases. Para ello se van a mostrar las métricas del método `classification_report` de la librería *Sklearn* en la Figura 5.21 y la matriz de confusión en la Figura 5.22.

	precision	recall	f1-score	support
Bacterial_spot	0.950	0.812	0.876	565
Early_blight	0.908	0.866	0.886	491
Late_blight	0.835	0.949	0.888	622
Leaf_Mold	0.940	0.915	0.927	550
Septoria_leaf_spot	0.832	0.866	0.849	576
Spider_mites Two-spotted_spider_mite	0.945	0.980	0.962	349
Target_Spot	0.915	0.948	0.931	365
Tomato_Yellow_Leaf_Curl_Virus	0.988	0.971	0.979	407
Tomato_mosaic_virus	0.981	0.958	0.969	430
healthy	0.982	0.967	0.974	610
powdery_mildew	0.905	0.955	0.929	200
accuracy			0.920	5165
macro avg	0.925	0.926	0.925	5165
weighted avg	0.923	0.920	0.920	5165

Figura 5.21.: [Resumen de métricas del modelo \*NASNetMobile 2\*](#)

El modelo ha alcanzado una precisión del 92% y un F1-score medio de 0.92, lo que indica un buen equilibrio entre precisión y capacidad de detección en las once clases evaluadas. Las clases `Tomato_Yellow_Leaf_Curl_Virus`, `Tomato_mosaic_virus` y `healthy` siguen destacando por sus altos valores de F1-score (superiores a 0.95), mostrando que el modelo identifica correctamente los patrones más distintivos de estas clases. Asimismo, el desempeño equilibrado entre las métricas macro y ponderadas (*weighted*) sugiere que no existe un sesgo notable hacia clases con mayor cantidad de muestras.

Sin embargo, algunas clases como `Septoria_leaf_spot`, `Early_blight` y `Late_blight` presentan valores más bajos, reflejando cierta confusión con otras enfermedades de síntomas visualmente similares.

La matriz de confusión de la Figura 5.22 refuerza las conclusiones obtenidas a partir de las métricas de evaluación, mostrando que el modelo clasifica correctamente la mayoría de las muestras en casi todas las clases. Se observa una fuerte concentración de valores en la diagonal principal, especialmente en las clases `Tomato_Yellow_Leaf_Curl_Virus`,

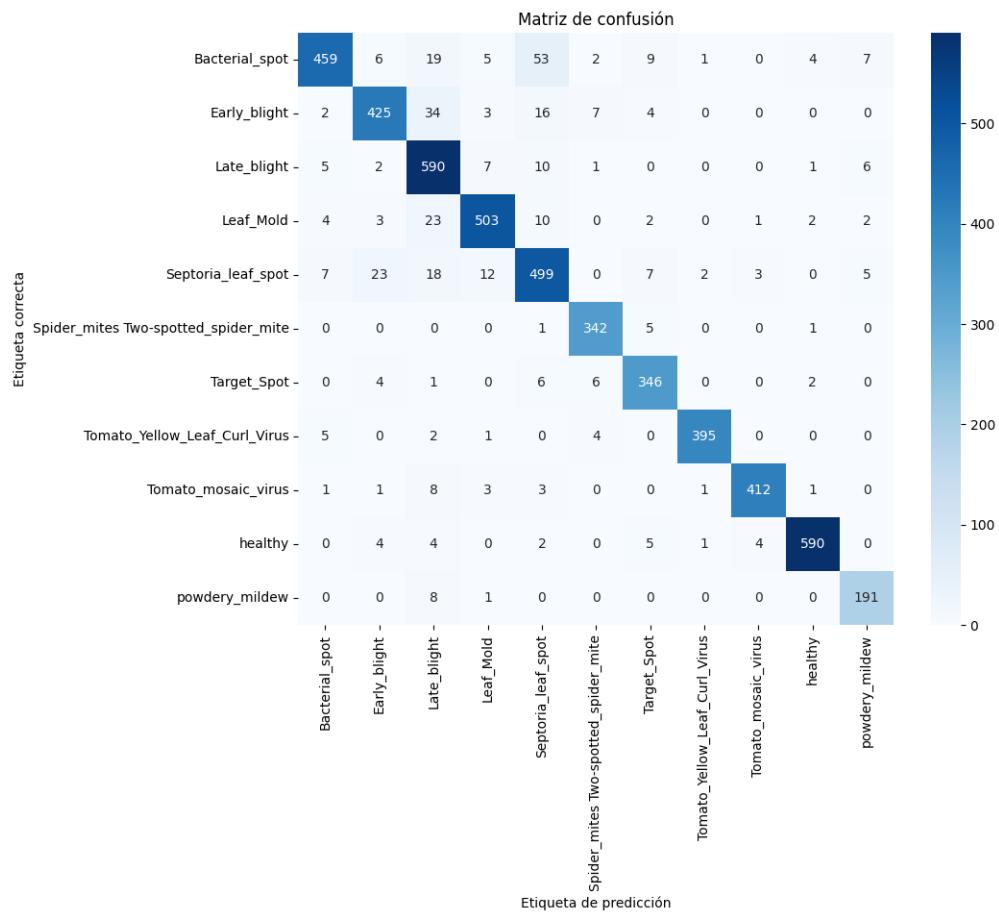


Figura 5.22.: Matriz de confusión del modelo *NASNetMobile 2*

Tomato\_mosaic\_virus y healthy. Estas clases presentan patrones visuales distintivos que el modelo ha aprendido de forma eficaz.

En contraste, se aprecian errores recurrentes entre clases con síntomas similares, como Septoria\_leaf\_spot, Early\_blight y Late\_blight donde las confusiones cruzadas son más frecuentes. En conjunto, la matriz de confusión visualiza de manera clara la capacidad del modelo para distinguir la mayoría de las enfermedades, evidenciando a la vez los retos en la diferenciación de patologías con síntomas visualmente parecidos.

Este modelo tiene un comportamiento sólido y equilibrado, generaliza bien y consigue que incluso las clases con menos muestras tengan buenos niveles de precisión y sensibilidad.

Descongelar desde la capa 60 hacia adelante ha sido una excelente elección. Se ha permitido que los bloques finales aprendan patrones específicos de las hojas y texturas de las enfermedades y el *learning rate* bajo fue clave para no destruir los pesos de *ImageNet*.

## 5.10. Comparativa de modelos

En esta sección se van a comparar los mejores modelos que se han hallado en este estudio, es decir, se van a comparar los modelos de las subsecciones 4.4.2, 4.4.3 y 4.4.8, a partir de los resultados obtenidos en las secciones 5.2, 5.3 y 5.8 respectivamente. Para ello, en el Cuadro 5.1 se presentan las características generales de los modelos junto con un número asociado en la columna "Número", que será usado como identificador del modelo en el Cuadro 5.2, donde se tienen las principales métricas de cada modelo.

Cuadro 5.1.: [Resúmenes de modelos](#)

Resumen de modelos				
Modelo	Número	Fine tuning	Learning rate	Épocas
MobileNetV2 2	1	100 capas	1e-5	20
MobileNetV2 3	2	80 capas	5e-6	20
NASNetMobile 2	3	60 capas	5e-6	15

Cuadro 5.2.: [Comparativa de modelos con métricas globales](#)

Resumen de métricas						
Modelo	Precisión macro	Precisión ponderada	Sensibilidad macro	Sensibilidad ponderada	F1-Score macro	F1-Score ponderada
1	0.940	0.941	0.941	0.939	0.939	0.939
2	0.925	0.927	0.927	0.924	0.924	0.924
3	0.925	0.923	0.926	0.920	0.925	0.920

Los tres modelos presentan rendimientos muy próximos, con ambas precisiones superiores al 92 %, lo que indica una buena capacidad de generalización en la clasificación de las once clases. También cabe destacar que los tres modelos en las tres métricas tienen un valor muy parecido en macro y weighted, lo que sugiere que el conjunto de datos está relativamente equilibrado entre clases, y que los modelos no presentan sesgos fuertes hacia ninguna clase.

El modelo MobileNetV2 2 obtiene las mejores métricas globales, alcanzando una precisión del 94 % y un f1-score de 93.9 %. Este modelo destaca por su rendimiento equilibrado en la mayoría de las clases, con especial solidez en Tomato\_Yellow\_Leaf\_Curl\_Virus ( $f1 = 0.985$ ), Tomato\_mosaic\_virus ( $f1 = 0.974$ ) y healthy ( $f1 = 0.968$ ).

En comparación, el modelo MobileNetV2 3 muestra un rendimiento ligeramente inferior con una precisión de 92.4 %, aunque mantiene una buena estabilidad entre clases. Las diferencias más notables respecto al modelo anterior se observan en Septoria\_leaf-spot y Spidermite\_Two-spotted\_spider\_mite, donde presenta valores de sensibilidad algo más bajos.

Por su parte, NASNetMobile 2 alcanza un rendimiento global similar con una precisión de 92.0 %, con un comportamiento especialmente bueno en las clases Spidermite\_Two

- `spotted_spider_mite` ( $f1 = 0.962$ ) y `Target_spot` ( $f1 = 0.931$ ). Sin embargo, en otras clases como `Bacterial_spot` o `Septoria_leaf_spot` su precisión se reduce notablemente, lo que sugiere una menor estabilidad interclase.

En conjunto, los resultados confirman que las arquitecturas basadas en MobileNetV2 ofrecen un mejor resultado. Las diferencias entre los modelos son moderadas, pero la configuración con fine tuning más profundo (100 capas) muestra una ligera ventaja consistente en las métricas promedio y una mayor estabilidad entre clases.

Estos resultados serán considerados en el apartado 6 para la selección del modelo final.

## 6. Conclusiones

Las conclusiones obtenidas tras el desarrollo completo de este proyecto son las siguientes:

1. Se realizó una revisión de la literatura sobre estudios relacionados con la detección de enfermedades en plantas, lo que permitió seleccionar arquitecturas preentrenadas ligeras y eficientes, como *MobileNetV2* y *NASNetMobile*.
2. Se analizaron los datos disponibles y se realizaron distintas tareas como adaptación de número de píxeles por imagen, normalización, creación de conjunto de pruebas, *data augmentation* y balanceo de clases con pesos predefinidos (*class weight*), para asegurar su homogeneidad, además de una estructuración equilibrada entre clases que facilitó un entrenamiento estable.
3. Se desarrollaron ocho modelos basados en arquitecturas (MobileNetV2, EfficientNetB0, NASNetMobile) preentrenadas con ImageNet, ajustando parámetros como el *learning rate* y el número de capas finamente ajustadas (*fine tuning*).
4. Se realizaron once entrenamientos, de los cuales tres modelos destacaron por su rendimiento. La comparativa detallada mostró que MobileNetV2 con *fine tuning* de 100 capas y learning rate de  $1e-5$  alcanzó los mejores resultados globales, con una accuracy del 93.9 % y un f1-score macro de 0.939, superando ligeramente al resto de configuraciones y mostrando una mayor estabilidad interclase.

En consecuencia, se selecciona este modelo (4.4.2) como el modelo final propuesto, al ofrecer el mejor equilibrio entre métricas globales e interclase, que garantizan una mayor precisión y capacidad de generalización. Este modelo cumple el objetivo de desarrollar un sistema preciso y adaptable para la clasificación automática de enfermedades en hojas de tomate. Los resultados obtenidos permiten afirmar que dicho objetivo se ha alcanzado satisfactoriamente, logrando un modelo con un rendimiento competitivo y una adecuada robustez frente a la variabilidad del conjunto de datos.

Como líneas futuras de trabajo, se plantea incorporar técnicas *ensembles* que permita combinar las diferentes arquitecturas para lograr mejorar la detección de las clases más conflictivas (Octacian et al., 2024).

## A. Anexo: Código fuente

El código fuente desarrollado para este Trabajo de Fin de Máster se encuentra disponible en el siguiente repositorio: <https://github.com/Pr0t0n9/TFM-tomatoes-diseases>. A continuación se va a explicar la estructura de directorios y ficheros del mismo:

### ■ Modelos

- modelo1: red convolucional base para clasificación multiclase.
- modelo2: versión optimizada con aumento de datos.

### ■ Datos

- train/: imágenes de entrenamiento organizadas por clase.
- valid/: imágenes de validación organizadas por clase.

### ■ Scripts

- preprocesamiento.py: normalización y segmentación de hojas.
- entrenamiento.py: definición y entrenamiento del modelo.
- evaluacion.py: métricas y generación de informes.

# Bibliografía

- Al Sahili, Z., & Awad, M. (2022). The Power of Transfer Learning in Agricultural Applications: AgriNet. *arXiv*. <https://doi.org/10.1016/j.ecoinf.2020.101182>
- Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, 61, 101-182. <https://doi.org/10.1016/j.ecoinf.2020.101182>
- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3), 346-359. <https://doi.org/10.1016/j.cviu.2007.09.014>
- Baysal-Gurel, F., & Miller, S. (2010). *Early Blight Management for Organic Tomato Production*. eOrganic. Consultado el 21 de octubre de 2025, desde <https://eorganic.org/node/4961>
- Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., Belin, É., & Chapeau-Blondeau, F. (2012). On the use of depth camera for 3D phenotyping of entire plants. *Computers and Electronics in Agriculture*, 82, 122-127. <https://doi.org/10.1016/j.compag.2011.12.007>
- Chimate, Y., Patil, S., Prathapan, K., Patil, J., & khot, J. (2025). Optimized sequential model for superior classification of plant disease. *Scientific reports*, 15(1), 3700. <https://doi.org/10.1038/s41598-025-86427-8>
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886-893. <https://doi.org/10.1109/CVPR.2005.177>
- Davis, R. M., Miyao, G., Subbarao, K. V., Stapleton, J. J., & Aegeerter, B. J. (s.f.). *Tomato (Solanum lycopersicum)-Bacterial Spot*. University of California. Consultado el 23 de octubre de 2025, desde <https://ipm.ucanr.edu/agriculture/tomato/bacterial-spot/#gsc.tab=0>
- Díaz-Pendón, J. A., Cañizares, M. C., Moriones, E., Bejarano, E. R., Czosnek, H., & Navas-Castillo, J. (2010). Tomato yellow leaf curl viruses: ménage à trois between the virus complex, the plant and the whitefly vector. *Molecular Plant Pathology*, 11(4), 441-450. <https://doi.org/10.1111/j.1364-3703.2010.00618.x>
- Dong, X., Wang, K., Huang, Q., Ge, Q., Zhao, K., Wu, X., Wu, X., Liang, L., & Hao, G. (2023). PD3D-PreTrain: A Series of Commonly Used Pre-Trained Models Support Image-Based Plant Disease Diagnosis. *Plant Phenomics*, 5, 0054. <https://doi.org/10.34133/plantphenomics.0054>

- Editors of Encyclopedia Britannica. (2024, agosto). *Late blight*. Encyclopedia Britannica. Consultado el 22 de octubre de 2025, desde <https://www.britannica.com/science/late-blight>
- Garcia-Ruiz, F., Sankaran, S., Maja, J. M., Lee, W. S., Rasmussen, J., & Ehsani, R. (2013). Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Computers and Electronics in Agriculture*, 91, 106-115. <https://doi.org/10.1016/j.compag.2012.12.002>
- GeeksforGeeks contributors. (2025a). *Efficientnet Architecture*. GeeksforGeeks. Consultado el 28 de octubre de 2025, desde <https://www.geeksforgeeks.org/computer-vision/efficientnet-architecture/>
- GeeksforGeeks contributors. (2025b). *Mobilenet V2 Architecture in Computer Vision*. GeeksforGeeks. Consultado el 28 de octubre de 2025, desde <https://www.geeksforgeeks.org/computer-vision/mobilenet-v2-architecture-in-computer-vision/#key-features-of-mobilenet-v2>
- Gevens, A., & Wilbur, J. (2014). *Tomato Late Blight*. University of Wisconsin-Madison. Consultado el 22 de octubre de 2025, desde <https://vegpath.plantpath.wisc.edu/diseases/tomato-late-blight/>
- Greenlife editors. (2025, enero). *Tomato yellow leaf curl virus*. Greenlife. Consultado el 23 de octubre de 2025, desde <https://www.greenlife.co.ke/tomato-yellow-leaf-curl-virus/>
- Hazzard, R. (2022). *Two-spotted Spider Mite*. University of Massachusetts Amherst. Consultado el 23 de octubre de 2025, desde <https://www.umass.edu/agriculture-food-environment/vegetable/fact-sheets/two-spotted-spider-mite>
- Inspired Pencil editors. (2021). *Red Spider Mite On Tomato Plants*. Inspired Pencil. Consultado el 23 de octubre de 2025, desde <https://ar.inspiredpencil.com/pictures-2023/red-spider-mites-on-tomato-plants>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. En F. Pereira, C. J. C. Burges, L. Bottou & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (pp. 1097-1105). Curran Associates, Inc.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Lu, J., Liu, X., Ma, X., Tong, J., & Peng, J. (2023). Improved MobileNetV2 crop disease identification model for intelligent agriculture (M. Murugappan, Ed.). *PeerJ Computer Science*, 9, e0123262. <https://doi.org/10.7717/peerj-cs.1595>
- McGrath, M. (s.f.). *Early blight on tomatoes*. Cornell CALS. Consultado el 21 de octubre de 2025, desde <https://blogs.cornell.edu/livegpath/gallery/tomato/early-blight/>
- Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, 7. <https://doi.org/10.3389/fpls.2016.01419>

- Motwani, A., & Khan, Q. (2025, octubre). Tomato Leaves Dataset. *Kaggle*. Consultado el 4 de octubre de 2025, desde <https://www.kaggle.com/datasets/ashishmotwani/tomato/data>
- Octacian, O., Badruzzaman, A., Ridho, M. Y., & Trisedya, B. D. (2024). Enhancing Weighted Averaging for CNN Model Ensemble in Plant Diseases Image Classification. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 8(2), 272-279. <https://doi.org/10.29207/resti.v8i2.5669>
- P., P. (2023). *What is EfficientNet? The Ultimate Guide*. Roboflow. Consultado el 28 de octubre de 2025, desde <https://blog.roboflow.com/what-is-efficientnet/>
- P. Keinath, A. (2023, julio). *A New Bacterial Spot Strain Is a Headache for Tomato Growers*. Growing Produce. Consultado el 23 de octubre de 2025, desde <https://www.growingproduce.com/crop-protection/a-new-bacterial-spot-strain-is-a-headache-for-tomato-growers/>
- Pacific Northwest contributors. (2025a). *Tomato (Solanum lycopersicum)-Bacterial Spot*. Pacific Northwest Handbooks. Consultado el 23 de octubre de 2025, desde <https://pnwhandbooks.org/plantdisease/host-disease/tomato-solanum-lycopersicum-bacterial-spot>
- Pacific Northwest contributors. (2025b). *Tomato (Solanum lycopersicum)-Early Blight*. Pacific Northwest Handbooks. Consultado el 21 de octubre de 2025, desde <https://pnwhandbooks.org/plantdisease/host-disease/tomato-solanum-lycopersicum-early-blight>
- Raza, S.-e.-A., Prince, G., Clarkson, J. P., & Rajpoot, N. M. (2015). Automatic detection of diseased tomato plants using thermal and stereo visible light images. *PLoS ONE*, 10(4), e0123262. <https://doi.org/10.1371/journal.pone.0123262>
- Renaissance Farms Editors. (s.f.). *Early Blight In Tomatoes*. Renaissance Farms. Consultado el 21 de octubre de 2025, desde <https://renaissancefarms.org/blog/early-blight-in-tomatoes/>
- Robinson, K., Creswell, T., & Dan, E. (2013, agosto). *Late blight disease found on Indiana tomato samples*. Purdue University. Consultado el 22 de octubre de 2025, desde <https://www.purdue.edu/newsroom/archive/releases/2013/Q3/late-blight-disease-found-on-indiana-tomato-samples.html>
- Saleem, M. H., Potgieter, J., & Rahman, A. K. M. (2020). Plant Disease Classification: A Comparative Evaluation of Convolutional Neural Networks and Deep Learning Optimizers. *Plants (Basel, Switzerland)*, 9(10), 1319. <https://doi.org/10.3390/plants9101319>
- Salka, T. D., Hanafi, M. B., Rahman, S. M. S. A. A., Zulperi, D. B. M., & Omar, Z. (2025). Plant leaf disease detection and classification using convolution neural networks model: a review. *Artificial Intelligence Review*, 58. <https://doi.org/10.1007/s10462-025-11234-6>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2019). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv*. <https://doi.org/10.1109/CVPR.2018.00474>

- Silva, M. (s.f.). *Cultivo de tomate: Cómo se realiza, plagas e importancia*. Agrotendencia TV. Consultado el 28 de julio de 2025, desde [https://agrotendencia.tv/agricultura/cultivos/hortalizas/el-cultivo-de-tomate/#Historia\\_del\\_tomate\\_o\\_jitomate](https://agrotendencia.tv/agricultura/cultivos/hortalizas/el-cultivo-de-tomate/#Historia_del_tomate_o_jitomate)
- TensorFlow developers. (2024). *tf.keras.applications.NASNetMobile*. TensorFlow. Consultado el 28 de octubre de 2025, desde [https://www.tensorflow.org/api\\_docs/python/tf/keras/applications/NASNetMobile](https://www.tensorflow.org/api_docs/python/tf/keras/applications/NASNetMobile)
- Wikipedia contributors. (s.f.-a). *Producción mundial del tomate*. Wikipedia. Consultado el 28 de julio de 2025, desde [https://es.wikipedia.org/wiki/Anexo:Producci%C3%B3n\\_mundial\\_de\\_tomate](https://es.wikipedia.org/wiki/Anexo:Producci%C3%B3n_mundial_de_tomate)
- Wikipedia contributors. (s.f.-b). *Solanum lycopersicum*. Wikipedia. Consultado el 28 de julio de 2025, desde [https://es.wikipedia.org/wiki/Solanum\\_lycopersicum](https://es.wikipedia.org/wiki/Solanum_lycopersicum)
- Zoph, B., Vasudevan, V., Shlens, J., & V. Le, Q. (2018). Learning Transferable Architectures for Scalable Image Recognition. *arXiv*. <https://doi.org/10.1109/CVPR.2018.00907>