

Diagnóstico de Enfermedades en Hojas de Vid mediante Aprendizaje Profundo: Un Estudio Comparativo de Modelos

Diagnosis of Grape Leaf Diseases Using Deep Learning: A Comparative Study of Models

Misael Hernández Sandoval¹, Alfonso Ramírez Pedraza², Geminiano Martínez Ponce³, Luis Valentín Coronado¹,
Enoc Tapia Mendez¹, Sebastián Salazar Colores¹

¹ Centro de Investigaciones en Óptica A.C., León, Guanajuato, CP. 37150, México.

mihernandez@cio.mx, luismvc@cio.mx, enoc.tapia@uaq.mx, sebastian.salazar@cio.mx

² Universidad Autónoma de Querétaro, Campus San Juan del Río, Querétaro, CP. 76807, México.

aramirez_ixm@ipn.mx

³ CICATA Qro., Instituto Politécnico Nacional, Mexico City, CP. 76090, México.

geminis@cio.mx

Resumen

La clasificación temprana y precisa de enfermedades foliares en cultivos de vid (*Vitis vinifera* L.) es un desafío crítico en la agricultura de precisión, ya que estas enfermedades pueden afectar significativamente la fotosíntesis y el rendimiento general. Este estudio evalúa y compara el rendimiento de varios modelos de clasificación basados en aprendizaje profundo, utilizando técnicas de transferencia de aprendizaje para clasificar enfermedades foliares y deficiencias nutricionales en hojas de vid mediante análisis de imágenes. Se utilizó un conjunto de datos etiquetados, compuesto por cuatro categorías: sanas, mancha foliar, sarampión negro y podredumbre negra. Se analizaron métricas clave de evaluación—como función de pérdida, exactitud, precisión, recuperación, puntuación F1, matrices de confusión y curvas ROC—junto con aspectos computacionales como el tiempo de entrenamiento. Las arquitecturas livianas, particularmente MobileNetV2, demostraron un rendimiento superior, alcanzando más del 99% de precisión, con una fuerte capacidad de generalización y un bajo costo computacional, lo que las hace adecuadas para su implementación en tiempo real en dispositivos móviles. En contraste, modelos más complejos como ResNet50 y EfficientNetB0 presentaron problemas como sobreajuste o mala adaptación al conjunto de datos. En comparación con los métodos tradicionales de diagnóstico basados en inspección visual por expertos o análisis químicos—que pueden ser subjetivos, costosos y lentos—la visión por computadora ofrece una alternativa más rápida, objetiva y rentable. Estos resultados confirman el potencial de las técnicas de aprendizaje profundo como herramientas prácticas para la clasificación temprana de enfermedades, apoyando a los agricultores en la mejora del manejo y la productividad de los cultivos. Finalmente, se discuten recomendaciones prácticas y futuras direcciones de investigación para facilitar la adopción de estas tecnologías en entornos reales de viticultura.

Palabras clave: Agricultura de precisión; aprendizaje profundo; clasificación de imágenes; enfermedades de la vid; redes neuronales convolucionales.

Introducción

La vid (*Vitis vinifera* L.) es uno de los cultivos frutales más relevantes a nivel mundial, tanto por su valor económico como por su importancia histórica y cultural. De acuerdo con datos de la Organización Internacional de la Viña y el Vino (OIV), en el año 2022 se produjeron aproximadamente 80.1 millones de toneladas de uva, cultivadas en una superficie cercana a los 7.3 millones de hectáreas distribuidas globalmente (OIV, 2023). Este cultivo se destina a diversos usos, entre los cuales destacan el consumo en fresco y la elaboración de productos derivados como vinos, jugos y pasas.

No obstante, la producción vitivinícola enfrenta importantes desafíos fitosanitarios, especialmente aquellos derivados de enfermedades foliares causadas por hongos, bacterias, virus y plagas. Estas enfermedades impactan negativamente en la calidad del fruto, disminuyen el rendimiento del cultivo y pueden generar pérdidas económicas significativas. Entre las patologías más relevantes se encuentran la mancha foliar



(Isariopsis Leaf Spot), el sarampión negro (Esca o Black Measles) y la podredumbre negra (Black Rot), las cuales se manifiestan principalmente en las hojas mediante síntomas como manchas necróticas, decoloraciones y patrones irregulares. Estas alteraciones pueden conducir a una defoliación prematura y a una reducción considerable de la actividad fotosintética (Jones, 2016; Tatineni & Hein, 2023; Schoelz & Stewart, 2018; Fuchs, 2023; Fuchs, 2024; Surico *et al.*, 2008; Islam, 2020; Ji *et al.*, 2020).

La clasificación temprana de estas enfermedades es fundamental para limitar su propagación y aplicar tratamientos eficaces de manera oportuna. Sin embargo, los métodos tradicionales de diagnóstico, basados en inspección visual realizada por expertos, presentan limitaciones inherentes, tales como la subjetividad del observador, el alto consumo de tiempo y la necesidad de conocimientos especializados (Islam, 2020). En este contexto, las tecnologías de visión por computadora, y en particular aquellas fundamentadas en técnicas de aprendizaje profundo, se presentan como una alternativa eficaz para automatizar la clasificación de enfermedades a partir de imágenes digitales.

La principal contribución de este trabajo consiste en abrir una nueva perspectiva para la incorporación de tecnologías avanzadas en el ámbito de la agricultura de precisión, específicamente dirigidas a la clasificación temprana y precisa de enfermedades foliares en el cultivo de la vid. A diferencia de los métodos convencionales —como la inspección visual, sujeta a criterios subjetivos, o los análisis químicos, que resultan costosos, invasivos, poco accesibles y de respuesta lenta—, la aplicación de técnicas de visión por computadora ofrece una solución objetiva, rápida, no invasiva y económicamente viable. Además, su potencial implementación en dispositivos móviles permite democratizar el acceso a herramientas avanzadas de monitoreo fitosanitario, haciéndolas accesibles a productores de diversa escala.

En particular, el enfoque de transferencia de aprendizaje ha demostrado una alta efectividad en contextos agrícolas donde la disponibilidad de datos etiquetados es limitada. Esta técnica permite reutilizar modelos de redes neuronales convolucionales (CNN) previamente entrenados en grandes conjuntos de datos genéricos, como ImageNet, y adaptarlos a tareas específicas, como la clasificación de enfermedades en hojas de vid (J *et al.*, 2020; Lauguico *et al.*, 2020; Rajpal, 2020). Esto permite mejorar la precisión del diagnóstico y reducir los requerimientos computacionales para su implementación práctica en campo.

El objetivo de este estudio es implementar y comparar el desempeño de distintas arquitecturas de CNN preentrenadas, ampliamente reconocidas por su eficacia en tareas de clasificación de imágenes. Las arquitecturas evaluadas incluyen: InceptionV3, EfficientNetB0, VGG16, VGG19, ResNet50, DenseNet121, MobileNetV2, NASNetMobile, ConvNeXtSmall y Xception. La selección de estos modelos se basa en su rendimiento demostrado, eficiencia computacional y facilidad de adaptación mediante técnicas de transferencia de conocimiento (Chollet, 2017; He *et al.*, 2016; Howard *et al.*, 2017; Huang *et al.*, 2017; Liu *et al.*, 2022; Simonyan & Zisserman, 2014; Szegedy *et al.*, 2016; Tan & Le, 2019; Zoph *et al.*, 2018).

A través de una evaluación exhaustiva, se busca determinar la capacidad de dichas arquitecturas para identificar enfermedades foliares en hojas de vid a partir de imágenes digitales. Para ello, se utilizarán métricas estándar en el ámbito de la visión por computadora, tales como exactitud, pérdida, puntuación F1, recuperación, precisión, matriz de confusión y curvas ROC. Esta evaluación integral permitirá establecer recomendaciones fundamentadas sobre los modelos más adecuados para su eventual integración en sistemas automatizados de monitoreo fitosanitario.

Metodología Propuesta

Contexto del estudio

El presente estudio se enfoca en la clasificación automatizada del estado fitosanitario de hojas de vid (*Vitis vinifera* L.), mediante la aplicación de técnicas de visión por computadora basadas en aprendizaje profundo.

Para ello, se implementaron y compararon diez modelos de clasificación de imágenes basados en redes neuronales convolucionales (CNN) preentrenadas, utilizando el enfoque de transferencia de aprendizaje. Estas arquitecturas están diseñadas para detectar patrones visuales asociados a enfermedades, como necrosis, decoloraciones y deformaciones morfológicas en las hojas.

La metodología general adoptada en el estudio se resume en la Figura 1, e incluye las siguientes etapas: recolección de imágenes desde repositorios públicos, preprocesamiento de datos, entrenamiento y evaluación de los modelos, así como el análisis del rendimiento mediante métricas e inferencias visuales.





Figura 1. Esquema general de la metodología empleada para la detección automática de enfermedades foliares en hojas de vid, específicamente mancha foliar, sarampión negro y podredumbre negra, mediante técnicas de aprendizaje profundo con transferencia de conocimiento.
Fuente: Elaboración propia.

Conjunto de datos

Se utilizó un conjunto de datos público disponible en la plataforma Kaggle (Kesharwani, 2025), compuesto por imágenes de hojas de vid clasificadas en cuatro categorías según su condición fitosanitaria: mancha foliar, sarampión negro, podredumbre negra y hojas sanas. Un ejemplo representativo de cada categoría se muestra en la Figura 2.

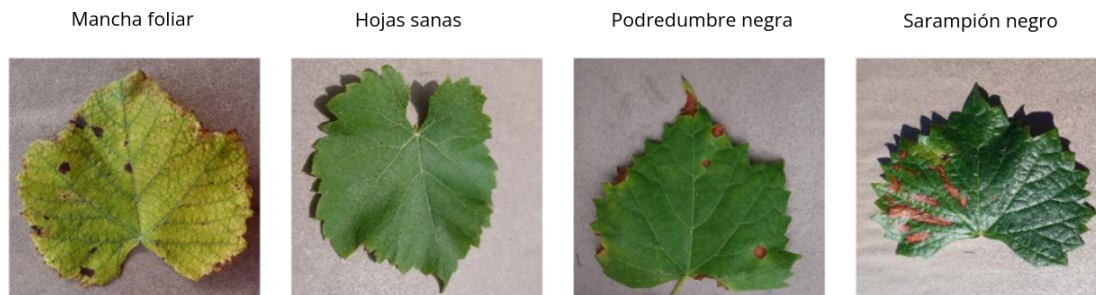


Figura 2. Ejemplos de imágenes del conjunto de datos utilizado, correspondientes a las cuatro clases consideradas: mancha foliar, hojas sanas, podredumbre negra y sarampión negro. Estas imágenes fueron empleadas como insumo para el entrenamiento y evaluación de los modelos de clasificación.
Fuente: Elaboración propia.

El conjunto se dividió en tres partes: entrenamiento, validación y prueba. La distribución de las imágenes por clase fue balanceada en cada subconjunto, con el fin de asegurar una representación equitativa y permitir una evaluación objetiva del rendimiento de los modelos. La Tabla 1 presenta un resumen de dicha distribución.

Tabla 1: Distribución del conjunto de datos por clase y partición.

Categoría	Entrenamiento (Número de imágenes)	Validación (Número de imágenes)	Prueba (Número de imágenes)
Mancha foliar	1722	344	86
Sarampión negro	1920	384	96
Podredumbre negra	1888	378	94
Hojas sanas	1692	339	84

Las imágenes presentan variabilidad en condiciones de iluminación, orientación y fondo, lo cual incrementa la robustez de los modelos y su capacidad de generalización ante escenarios reales en contextos agrícolas diversos.

Preprocesamiento de datos

El preprocesamiento constituye una etapa crítica para mejorar la calidad de los datos de entrada y optimizar el rendimiento de los modelos. En este estudio se empleó la clase ImageDataGenerator de la biblioteca Keras para aplicar transformaciones exclusivamente sobre el conjunto de entrenamiento.

El proceso incluyó normalización y técnicas de aumento de datos, con el fin de simular una mayor diversidad visual. Las transformaciones aplicadas fueron:

- Normalización `rescale=1./255`: Escala los valores de los píxeles al rango $[0, 1]$, facilitando la convergencia del modelo.
- Rotación aleatoria `rotation_range=40`: Simula diferentes orientaciones con rotaciones de hasta ± 40 .
- Desplazamientos horizontales y verticales `width_shift_range=0.2`, `height_shift_range=0.2`: Introduce variaciones posicionales del objeto.
- Cizalladura `shear_range=0.2`: Simula distorsiones diagonales.
- Zoom aleatorio `zoom_range=0.2`: Favorece la detección de patrones a distintas escalas.
- Volteo horizontal `horizontal_flip=True`: Refuerza la simetría visual mediante inversiones aleatorias.
- Relleno `fill_mode='nearest'`: Completa las áreas vacías usando el valor del píxel más cercano.

Estas técnicas contribuyen a reducir el sobreajuste y aumentar la capacidad del modelo para enfrentarse a condiciones visuales variadas en el entorno agrícola.

Modelos de clasificación implementados

Se implementaron y evaluaron diez arquitecturas de redes neuronales convolucionales (CNN) reconocidas por su efectividad en tareas de clasificación de imágenes. La selección se basó en su desempeño reportado, eficiencia computacional y facilidad de adaptación mediante transferencia de aprendizaje.

Las arquitecturas consideradas fueron:

- InceptionV3 (Szegedy, 2016): arquitectura híbrida que emplea convoluciones multiescala paralelas para maximizar la eficiencia.
- EfficientNetB0 (Tan, 2019): modelo escalable derivado de búsqueda automática de arquitectura, con escalado compuesto.



- VGG16 y VGG19 (Simonyan, 2014): arquitecturas secuenciales con estructuras repetitivas y capas densas.
- ResNet50 (He *et al.*, 2016): incorpora conexiones residuales para facilitar el entrenamiento de redes profundas.
- DenseNet121 (Huang, 2017): utiliza conexiones densas entre capas para fomentar la reutilización de características.
- MobileNetV2 (Howard *et al.*, 2017): diseñada para dispositivos móviles, con convoluciones separables en profundidad.
- NASNetMobile (Zoph *et al.*, 2018): arquitectura optimizada mediante búsqueda neuronal automatizada.
- ConvNeXtSmall (Liu *et al.*, 2022): arquitectura moderna que integra conceptos de transformers en una estructura CNN.
- Xception (Chollet, 2017): reemplaza las convoluciones estándar por convoluciones separables en profundidad, mejorando la eficiencia.

La diversidad de modelos considerados abarca desde arquitecturas livianas hasta configuraciones profundas y modernas, lo cual permite una evaluación integral según criterios de precisión, eficiencia y aplicabilidad.

Comparativa técnica de modelos

La Tabla 2 resume las principales características técnicas de las arquitecturas evaluadas, incluyendo el número de parámetros entrenables, el tipo de arquitectura, la técnica distintiva, la resolución de entrada y el tamaño aproximado del modelo en disco.

Tabla 2 : Comparativa técnica de modelos CNN utilizados en el estudio.

Modelo	Parámetros (M)	Tipo de arquitectura	Técnica distintiva	Entrada (px)	Tamaño (MB)
InceptionV3	23.8	Mixta	Convoluciones multiescala paralelas	299×299	92
EfficientNetB0	5.3	Escalable	Escalado compuesto (profundidad, ancho, resolución)	224×224	20
VGG16	138	Secuencial	Arquitectura clásica con capas densas	224×224	528
VGG19	144	Secuencial	Mayor profundidad estructural	224×224	548
ResNet50	25.6	Residual	Conexiones residuales (skip connections)	224×224	98
DenseNet121	8.0	Densa	Conexiones densas entre capas	224×224	33
MobileNetV2	3.4	Ligera	Convoluciones separables en profundidad	224×224	14
NASNetMobile	5.3	NAS	Arquitectura optimizada por búsqueda neuronal	224×224	21
ConvNeXtSmal	50	Moderna	Arquitectura CNN con diseño de transformer	224×224	197
Xception	22.9	Profunda	Convoluciones separables en profundidad	299×299	88



Esta comparativa permite identificar arquitecturas más adecuadas según el contexto de aplicación, considerando limitaciones de hardware, requisitos de precisión y escalabilidad.

Configuración de entrenamiento

Para asegurar una comparación justa entre las distintas arquitecturas, se definió un conjunto homogéneo de condiciones experimentales en términos de hardware, hiperparámetros y procedimientos. La Tabla 3 detalla la configuración empleada durante el entrenamiento de todos los modelos.

Tabla 3: Configuración de hiperparámetros utilizada durante el entrenamiento de los modelos.

Hiperparámetro	Valor
Tamaño de imagen de entrada	Según el modelo (ver Tabla 2)
Tamaño de lote (batch size)	32
Número de épocas	100
Tasa de aprendizaje inicial	0.0001
Función de pérdida	categorical_crossentropy
Optimizador	Adam
Tasa de abandono (dropout)	0.5
Capas entrenables	3 (bloques finales del backbone)
Número de clases de salida	4
Pasos por época	N.º de muestras / tamaño de lote

El entrenamiento se ejecutó en una estación de trabajo equipada con procesador AMD Ryzen 5 5600G (12 hilos), 48 GB de RAM, GPU NVIDIA RTX A4000 con 16 GB de VRAM y unidad SSD de 1 TB. Esta infraestructura permitió realizar múltiples experimentos de forma eficiente, reduciendo el tiempo de cómputo y facilitando la exploración de ajustes.

Resultados

Para evaluar el rendimiento de cada modelo se utilizaron métricas cuantitativas ampliamente reconocidas en el ámbito del aprendizaje profundo, tales como la función de pérdida y la exactitud global. Asimismo, se emplearon métricas específicas para tareas de clasificación multiclase como la precisión, la sensibilidad o recuperación y la puntuación F1. Estas métricas permiten no solo valorar la capacidad predictiva general del modelo, sino también su efectividad para discriminar correctamente entre las distintas clases, aspecto fundamental en escenarios con desbalance de clases.

El análisis comparativo de estas métricas busca identificar las arquitecturas que logren un equilibrio óptimo entre rendimiento, eficiencia computacional y capacidad de generalización, con vistas a su aplicación en sistemas de monitoreo fitosanitario automatizados dentro del contexto de la agricultura de precisión.



Métricas de evaluación

La validación rigurosa del desempeño de los modelos es esencial para garantizar su capacidad de generalización en entornos reales. Para este propósito, se recurrió a métricas estandarizadas en tareas de clasificación supervisada, incluyendo la función de pérdida, la precisión, la matriz de confusión, la puntuación F1-score y la curva Característica Operativa del Receptor (ROC) junto con su área bajo la curva (AUC).

Función de pérdida y precisión: La función de pérdida representa la discrepancia entre las predicciones del modelo y las etiquetas reales; valores más bajos indican una mejor aproximación del modelo a la distribución real de los datos. La exactitud o accuracy evalúa la proporción de predicciones correctas sobre el total de muestras evaluadas. Su cálculo se expresa mediante:

$$\text{exactitud} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

donde TP, TN, FP y FN representan verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente.

Matriz de confusión: La matriz de confusión proporciona una representación detallada del rendimiento del modelo, desglosando aciertos y errores por clase. Resulta particularmente útil para analizar patrones de confusión entre categorías similares y diagnosticar posibles sesgos de clasificación.

Puntuación F1: La puntuación F1 combina armónicamente la precisión y la sensibilidad, siendo especialmente útil en contextos donde las clases están desbalanceadas:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = 2 \frac{(\text{Precision})(\text{Recall})}{(\text{Precision})+(\text{Recall})} \quad (4)$$

Este valor oscila entre 0 y 1, donde puntuaciones más cercanas a 1 indican un mejor equilibrio entre precisión y sensibilidad.

Curva ROC y AUC: La curva ROC gráfica la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) a lo largo de distintos umbrales de decisión. El área bajo esta curva sirve como indicador global del poder discriminativo del modelo:

$$TPR = \frac{TP}{TP+FN} \quad (5)$$

$$FPR = \frac{FP}{FP+FN} \quad (6)$$

Un AUC cercano a 1 denota un alto rendimiento, mientras que valores próximos a 0.5 indican una capacidad de clasificación similar al azar.

Importancia del conjunto de validación: Durante el entrenamiento, las métricas fueron monitoreadas tanto en el conjunto de entrenamiento como en el de validación. Mientras que el primero permite optimizar los



parámetros internos del modelo, el segundo ofrece una estimación de su capacidad de generalización sobre datos no vistos. Un modelo robusto debería exhibir alta precisión y baja pérdida en ambos conjuntos, evitando fenómenos como el sobreajuste (overfitting) o el subentrenamiento (underfitting).

Dinámica del entrenamiento y rendimiento de clasificación

Las curvas de evolución de la pérdida y la precisión a lo largo de las épocas de entrenamiento permiten evaluar la dinámica de aprendizaje de los modelos. Un entrenamiento efectivo se refleja en una disminución progresiva y sostenida de la pérdida, junto con un incremento estable de la precisión.

La Figura 3 ilustra la evolución de la precisión durante 100 épocas para cada arquitectura evaluada. Se observa que múltiples modelos alcanzan altos niveles de precisión en las primeras 20 a 30 épocas, lo cual evidencia una rápida convergencia. En particular, los modelos MobileNetV2, VGG19 e InceptionV3 estabilizan su precisión en valores cercanos al 100%.

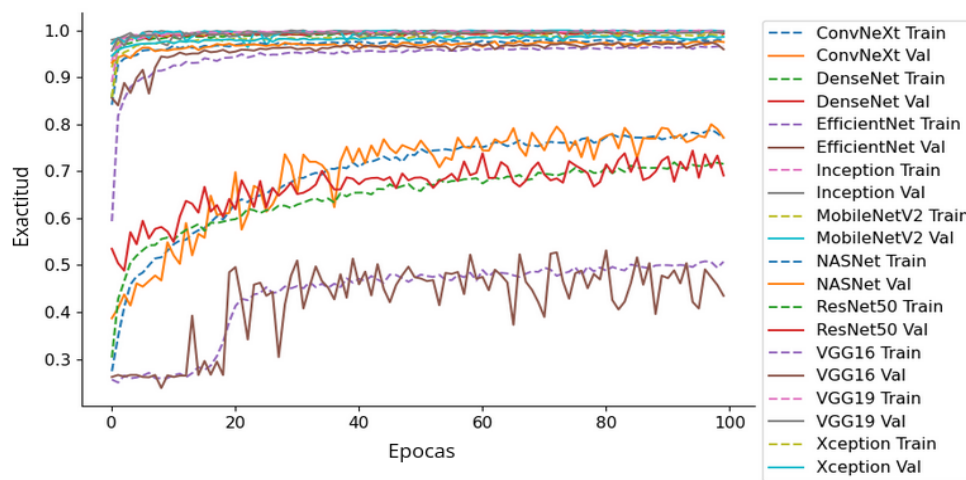


Figura 3. Evolución de la exactitud durante el entrenamiento para los modelos evaluados.
Fuente: Elaboración propia.

La Figura 4, por su parte, muestra la evolución de la función de pérdida. Una reducción sostenida a lo largo de las épocas indica una correcta optimización. MobileNetV2 evidenció una pérdida mínima desde las primeras fases del entrenamiento, mientras que EfficientNetB0 presentó valores persistentemente altos, lo cual sugiere dificultades de convergencia o un ajuste inadecuado al conjunto de datos.

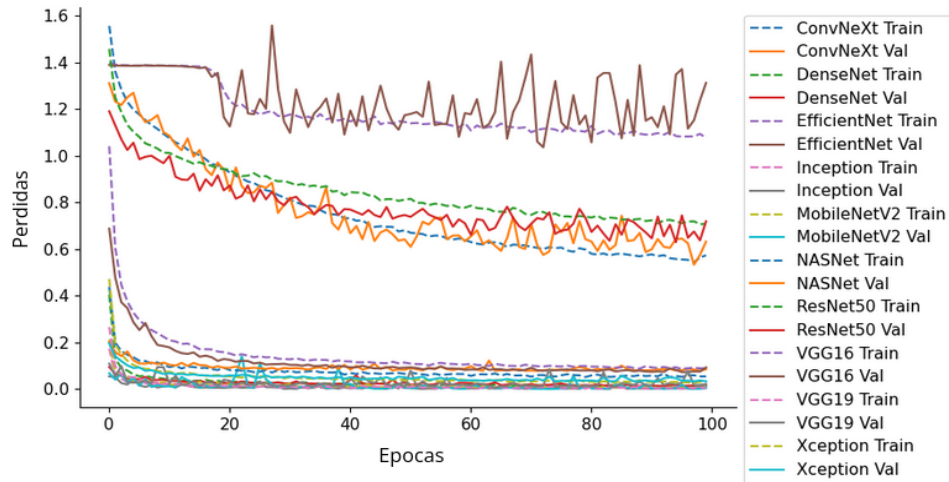


Figura 4. Evolución de la función de pérdida durante el entrenamiento para los modelos evaluados.
 Fuente: Elaboración propia.

La Tabla 4 resume los valores finales de pérdida y precisión en la época 100, tanto para los conjuntos de entrenamiento como de validación. Además, se reporta el tiempo total de entrenamiento de cada arquitectura, una métrica crucial al considerar implementaciones prácticas en tiempo real o en dispositivos de capacidad limitada.

Tabla 4. Pérdida y precisión final al término del entrenamiento (época 100) y tiempo total de entrenamiento por modelo.

Modelo	Pérdida Entren	Exactitud Entren	Pérdida Val	Exactitud Val	Tiempo de Entren (min)
Xception	0.0322	0.9887	0.0333	0.9861	122.62
VGG16	0.0888	0.9666	0.0925	0.9597	69.36
ConvNeXtSmall	0.5728	0.7702	0.6316	0.7715	78.10
NASNetMobile	0.0540	0.9791	0.0846	0.9750	67.92
NASNetMobile	0.0110	0.9955	0.0160	0.9931	67.98
DenseNet121	0.0093	0.9974	0.0047	0.9986	66.29
DenseNet121	0.7073	0.7156	0.7182	0.6910	68.29
MobileNetV2	0.0031	0.9992	0.0214	0.9965	70.24
ResNet50	1.0757	0.5074	1.3114	0.4347	67.10
VGG19	0.0037	0.9986	0.0146	0.9972	121.18
EfficientNetB0	-	-	-	-	-
InceptionV3	-	-	-	0.9853	-
MobileNet + clasificador propio (Nagi, 2021)	-	-	-	-	-
SVM + GLCM features (Javidan, 2023)	-	-	-	0.9871	-
Deep Learning (CNN) (Javidan, 2023)	-	-	-	0.8682	-
Deep Learning (CNN) (Javidan, 2023)	-	-	-	0.9405	-
Deep Learning	-	-	-	0.9303	-



Modelo	Pérdida Entren	Exactitud Entren	Pérdida Val	Exactitud Val	Tiempo de Entren (min)
(GoogLeNet)	-	-	-	0.8000	-
(Javidan, 2023)	-	-	-	0.5600	-
Random F (Uniyal et al., 2024)	-	-	-	-	-
Gradient Boos (Uniyal et al., 2024)	-	-	-	-	-
Naive Bayes (Uniyal et al., 2024)	-	-	-	-	-

Entre los modelos evaluados, MobileNetV2 sobresale por su rendimiento excepcional, alcanzando una precisión de validación de 99.86% y una pérdida de 0.0047, lo cual es significativamente superior a otros modelos reportados en la literatura. Por ejemplo, el modelo MobileNet con clasificador propio reportado por Nagi & Tripathy (2021) alcanzó una precisión de 98.53%, mientras que SVM + GLCM features de Javidan *et al.* (2023) obtuvo una precisión de 98.71%. Estos valores son considerablemente inferiores a los logrados por MobileNetV2, lo que destaca su superioridad en términos de precisión.

Adicionalmente, InceptionV3 y VGG19 también mostraron excelentes resultados, con precisiones de 99.72% y 99.65%, respectivamente, aunque con tiempos de entrenamiento más largos (121.18 y 70.24 minutos). Por otro lado, EfficientNetB0 presentó el peor rendimiento, con una precisión de validación de 43.47% y una pérdida de 1.3114, valores que indican una falta de adaptación a este dominio específico, y muy por debajo de otros enfoques de aprendizaje profundo reportados en la literatura.

En términos generales, modelos livianos como MobileNetV2, NASNetMobile y DenseNet121 lograron una combinación excepcional de rendimiento predictivo y tiempos de entrenamiento reducidos, lo que los posiciona como opciones óptimas para aplicaciones en dispositivos móviles o sistemas embebidos dentro de la agricultura de precisión. Estos resultados superan de manera consistente los reportados por otros modelos, destacando la efectividad de estas arquitecturas para la clasificación en tiempo real en entornos prácticos.

Capacidad de discriminación: Matriz de confusión y curvas ROC

Con el objetivo de complementar el análisis cuantitativo, se evaluó la capacidad discriminativa de los modelos mediante el uso de matrices de confusión y curvas ROC. La Figura 5 presenta una comparación visual entre el mejor modelo (MobileNetV2) y el de peor desempeño (EfficientNetB0), evidenciando las diferencias en su habilidad para clasificar correctamente las distintas clases.



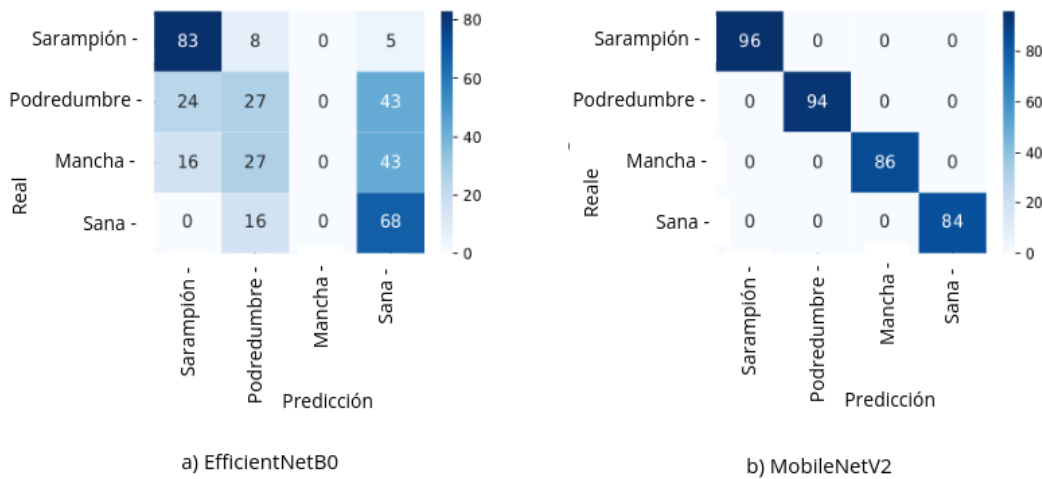


Figura 5. Comparación de matrices de confusión para los modelos: a) peor modelo y b) mejor modelo.
 Fuente: Elaboración propia.

El modelo EfficientNetB0 evidenció una capacidad de clasificación deficiente, con una alta tasa de errores en todas las clases. Por ejemplo, de las 100 muestras reales correspondientes a la clase podredumbre negra, solo 43 fueron clasificadas correctamente, mientras que 60 fueron erróneamente etiquetadas como mancha foliar y 5 como sarampión negro. Asimismo, de las 100 hojas saludables, únicamente 27 fueron reconocidas correctamente, mientras que 68 fueron clasificadas como podredumbre negra y 5 como sarampión negro, sin aciertos en la clase mancha foliar. Este patrón errático sugiere una limitada capacidad del modelo para discriminar entre enfermedades con características visuales similares, lo que afecta negativamente métricas clave como exactitud, recuperación y puntuación F1.

En marcado contraste, la matriz de confusión correspondiente al modelo MobileNetV2 muestra una clasificación perfecta: las 100 imágenes de cada clase fueron correctamente identificadas, generando una matriz perfectamente diagonal. Este desempeño refleja no solo una capacidad sobresaliente de generalización, sino también una alta robustez frente a la variabilidad visual de los datos.

En términos cuantitativos, mientras EfficientNetB0 alcanzó precisiones por debajo del 60% en varias clases, MobileNetV2 logró una precisión total superior al 99.8%, lo que subraya su potencial para aplicaciones en campo, especialmente en sistemas de diagnóstico embebidos con restricciones de hardware.

Las curvas ROC presentadas en la Figura 6 permiten analizar la relación entre la sensibilidad y la especificidad por clase. MobileNetV2 presenta curvas cercanas al óptimo (AUC = 1), mientras que EfficientNetB0 muestra un rendimiento significativamente inferior, con áreas bajo la curva sustancialmente más bajas.

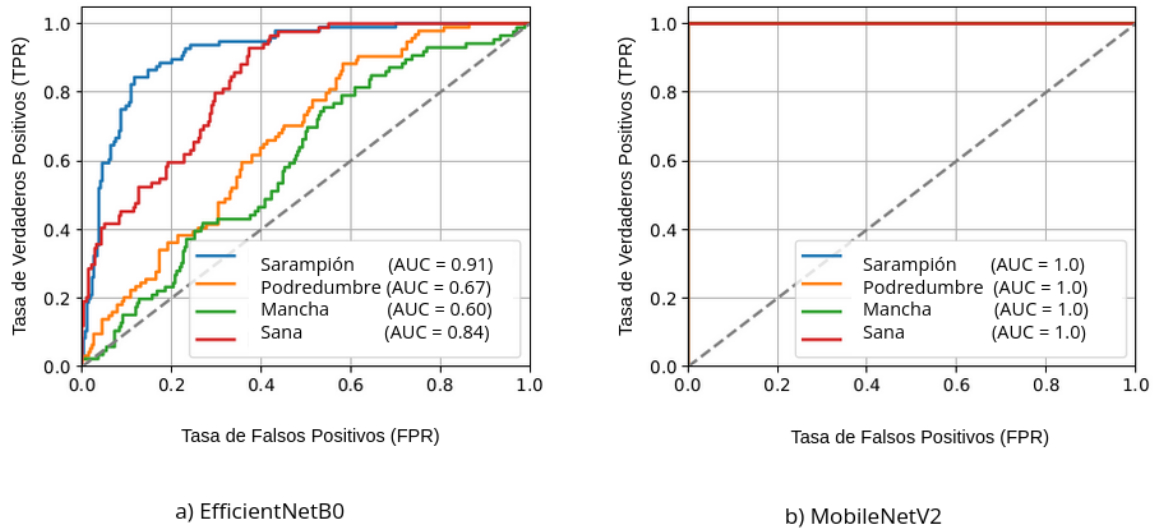


Figura 6. Curvas ROC por clase para los modelos: a) EfficientNetB0 y b) MobileNetV2.
Fuente: Elaboración propia.

El modelo MobileNetV2 alcanzó un desempeño ideal, con un AUC de 1.00 en todas las clases, lo que indica una discriminación perfecta sin generación de falsos positivos ni negativos.

En contraste, EfficientNetB0 evidenció un comportamiento inconsistente, con las siguientes áreas bajo la curva:

- Sarampión negro: AUC = 0.91
- Podredumbre negra: AUC = 0.67
- Mancha foliar: AUC = 0.60
- Sana: AUC = 0.84

Estos valores revelan una capacidad limitada para discriminar clases como mancha foliar y podredumbre negra, cuyas curvas se alejan considerablemente del ideal. Tal desempeño deficiente compromete su aplicabilidad en escenarios reales donde la fiabilidad diagnóstica es crítica.

En síntesis, la comparación entre las curvas ROC refuerza la superioridad de MobileNetV2, tanto en su capacidad para clasificar correctamente como en su poder de discriminación por clase, posicionándolo como una alternativa idónea para sistemas automatizados de monitoreo fitosanitario.

Evaluación cualitativa: Inferencias visuales

La Figura 7 muestra ejemplos de inferencias realizadas por el modelo MobileNetV2, ilustrando imágenes reales de hojas de vid con sus respectivas predicciones. La concordancia entre la categoría real y la predicha en cada imagen valida la efectividad del modelo incluso en condiciones visuales complejas, lo cual es fundamental para aplicaciones prácticas en campo.

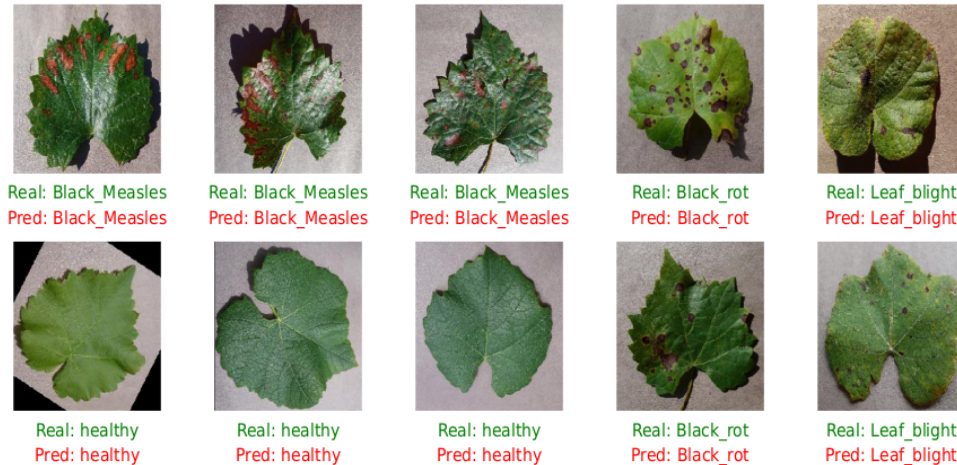


Figura 7. Ejemplos de inferencias visuales realizadas por el modelo MobileNetV2 en la clasificación de enfermedades y deficiencias nutricionales en hojas de vid.
 Fuente: Elaboración propia.

Métricas de clasificación (Precisión, Recuperación, Puntuación F1) por modelo

Para una evaluación integral del desempeño de los modelos, se calcularon métricas adicionales de clasificación: precisión, recuperación y puntuación F1, promediadas de forma ponderada. La Tabla 5 resume estos valores y permite apreciar el equilibrio entre aciertos y errores, particularmente relevante en presencia de clases desbalanceadas.

Tabla 5. Métricas de clasificación (promedio ponderado) para cada arquitectura evaluada.

Modelo	Precisión	Recuperación	Puntuación F1
Xception	0.9864	0.9861	0.9861
VGG16	0.9704	0.9666	0.9665
ConvNeXtSmall	0.7912	0.7702	0.7666
NASNetMobile	0.9784	0.9791	0.9778
DenseNet121	0.9946	0.9955	0.9944
MobileNetV2	1.0000	1.0000	1.0000
ResNet50	0.0544	0.2333	0.0883
VGG19	0.9973	0.9972	0.9972
EfficientNetB0	0.3701	0.4944	0.4147
InceptionV3	0.9973	0.9972	0.9972

Los resultados confirman a MobileNetV2 como la arquitectura más robusta y eficiente, alcanzando valores perfectos en todas las métricas de desempeño. Le siguen modelos como InceptionV3, VGG19 y DenseNet121, con resultados sobresalientes. En cambio, EfficientNetB0 y ResNet50 mostraron limitaciones notables en capacidad predictiva.

La Figura 8 complementa el análisis al comparar visualmente el rendimiento de los modelos según la métrica puntuación F1. Esta representación facilita una interpretación rápida del comportamiento relativo entre arquitecturas.

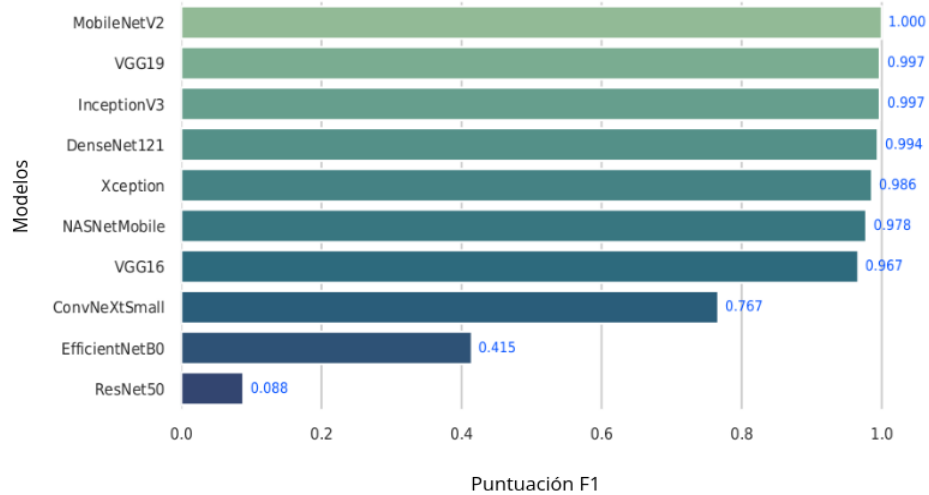


Figura 8. Comparación del rendimiento de modelos en términos de puntuación F1.
Fuente: Elaboración propia.

La métrica puntuación F1 es especialmente adecuada para problemas de clasificación con clases desbalanceadas, ya que equilibra la precisión (proporción de verdaderos positivos entre las predicciones positivas) y el recall (proporción de verdaderos positivos entre los casos positivos reales). Entre sus ventajas destacan:

- Proporciona una visión más realista del desempeño que la precisión simple cuando las clases están desbalanceadas.
- Facilita comparaciones objetivas entre modelos mediante un único valor representativo.
- Es clave en aplicaciones donde tanto falsos positivos como falsos negativos tienen consecuencias críticas.

Por estas razones, la puntuación F1 se posiciona como una métrica robusta y confiable para la evaluación del desempeño en tareas de clasificación automatizada, como el diagnóstico fitosanitario basado en imágenes.

Resultados y conclusiones

El presente estudio ha demostrado la efectividad del uso de arquitecturas de aprendizaje profundo, a través de técnicas de transferencia de aprendizaje, para la clasificación precisa de enfermedades foliares en hojas de vid mediante imágenes. Entre los modelos evaluados, MobileNetV2 se destacó por su sobresaliente rendimiento, alcanzando una precisión superior al 99.8%, con un bajo margen de error y una notable capacidad de generalización.

Este modelo superó a arquitecturas más complejas, las cuales presentaron problemas de sobreajuste o dificultades para adaptarse adecuadamente al conjunto de datos empleado.

La principal contribución de este trabajo radica en evidenciar el potencial de las tecnologías de visión por computadora aplicadas a la agricultura de precisión, especialmente en el contexto del cultivo de uva. A diferencia de los métodos tradicionales —como la inspección visual, sujeta a criterios subjetivos del personal técnico, o el análisis químico, que puede ser costoso, invasivo y de acceso limitado—, la implementación de modelos basados en inteligencia artificial representa una alternativa eficiente, no invasiva y económica para el diagnóstico temprano de enfermedades.

Asimismo, la posibilidad de implementar modelos ligeros en dispositivos móviles o sistemas embebidos permite democratizar el acceso a herramientas avanzadas de diagnóstico, beneficiando tanto a pequeños productores como a grandes viticultores, y optimizando la toma de decisiones agronómicas.

Recomendaciones

- **Implementación práctica:** Se recomienda la adopción de arquitecturas ligeras como *MobileNetV2* para su integración en dispositivos móviles o plataformas embebidas, ya que ofrecen una excelente relación entre precisión y eficiencia computacional.
- **Ampliación del conjunto de datos:** Futuras investigaciones deberían considerar la incorporación de bases de datos más amplias y diversas, que contemplen variabilidad en condiciones climáticas, etapas fenológicas del cultivo y distintos contextos geográficos, con el fin de mejorar la robustez y generalización de los modelos.
- **Modelos multimodales:** Se sugiere explorar la integración de enfoques multimodales que combinen imágenes con otros tipos de datos, como espectroscopia, sensores ambientales o análisis del suelo, lo cual podría incrementar la precisión diagnóstica y contextual del sistema.
- **Diseño centrado en el usuario:** Es fundamental el desarrollo de interfaces intuitivas y accesibles, orientadas a facilitar el uso de estas herramientas por parte de agricultores y técnicos sin formación especializada en inteligencia artificial.
- **Validación en campo:** Finalmente, se recomienda validar el desempeño de los modelos en condiciones reales de campo, evaluando tanto su eficacia operativa como la aceptabilidad por parte de los usuarios finales en entornos productivos reales.

En síntesis, la aplicación de técnicas de aprendizaje profundo en el diagnóstico automatizado de enfermedades foliares en la vid representa una vía prometedora hacia la consolidación de la agricultura de precisión. Este enfoque no solo contribuye a mejorar la productividad y sostenibilidad del sector vitivinícola, sino que también fomenta la adopción de soluciones tecnológicas accesibles, escalables y adaptadas a los desafíos del contexto agrícola actual.

Referencias

- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. CVPR.
- Fuchs, M. (2023). Grapevine virology highlights: 2018–2023. In Proc. ICVG, Thessaloniki, Greece, 18–26.
- Fuchs, M. (2024). Grapevine viruses: Did you say more than a hundred? J. Plant Pathol.
- Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., & Welinder, P. (2017). MobileNets. arXiv:1704.04861.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. CVPR.
- Islam, M. T. (2020). Plant disease detection using CNN model and image processing. *Int. J. Eng. Res. Technol.*, 9(10), 291–297.
- Javidan, S. M., Banakar, A., Asefpour Vakilian, K., & Ampatzidis, Y. (2023). Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. *Smart Agricultural Technology*, 3, 100081. <https://doi.org/10.1016/j.sagtech.2023.100081>
- Ji, M., Zhang, L., & Wu, Q. (2020). Automatic grape leaf diseases identification via UnitedModel. *Information Processing in Agriculture*, 7(3), 418–426.
- Jones, A. L., *et al.* (2016). Grapevine Leafroll: A complex viral disease affecting a high-value fruit crop. *Plant Disease*, 100(10), 2185–2204.
- Kesharwani, Y. (2025). Grapes Disease Dataset. Kaggle. Retrieved October 12, 2025, from <https://www.kaggle.com/datasets/ykesharwani/grapes-disease-dataset>.
- Lauguico, S., *et al.* (2020). Grape leaf multi-disease detection using transfer learning. In TENCON, 767–772.



- Liu, Z., *et al.* (2022). A ConvNet for the 2020s. CVPR.
- Nagi, R., & Tripathy, S. S. (2021). Grapevine leaf disease identification using transfer learning. Proceedings of the 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), 1-5. IEEE. <https://doi.org/10.1109/WIECON-ECE54711.2021.9829703>
- Organización Internacional de la Viña y el Vino (OIV). (2023). Evaluación anual del sector vitivinícola mundial – 2023.
- Rajpal, N. (2020). Black rot disease detection in the grape plant using color segmentation and ML. In ICACCCN, 976–979.
- Schoelz, J. E., & Stewart, L. R. (2018). The role of viruses in the phytobiome. *Annual Review of Virology*, 5, 93–111.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks. arXiv:1409.1556.
- Surico, G., Mugnai, L., & Marchi, G. (2008). The Esca disease complex. In *Integrated Management of Diseases Caused by Fungi, Phytoplasma and Bacteria* (pp. 119–136). Springer.
- Szegedy, C., *et al.* (2016). Rethinking the Inception Architecture. CVPR.
- Tatineni, S., & Hein, G. L. (2023). Plant viruses of agricultural importance. *Phytopathology*, 113(2), 117–141.
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling. ICML.
- Uniyal, S., Dhoundiyal, P., Sharma, V., & Vats, S. (2024). An intelligent approach to grape leaf disease diagnosis through machine learning. Proceedings of the 2nd International Conference on Disruptive Technologies (ICDT), 1-5. IEEE. <https://doi.org/10.1109/ICDT61202.2024.10489767>
- Zoph, B., *et al.* (2018). Learning transferable architectures. CVPR.

