



## IDENTIFICACIÓN DE PROBLEMAS FITOSANITARIOS EN EL CULTIVO DE FRESA POR MEDIO DE REDES NEURONALES CONVOLUCIONALES HÍBRIDAS

### IDENTIFICATION OF PLANT HEALTH PROBLEMS IN STRAWBERRY USING HYBRID CONVOLUTIONAL NEURAL NETWORKS

Giovanny Montes-Rodríguez y Juan Manuel González-Camacho\*

Colegio de Postgraduados, Campus Montecillo, Cómputo Aplicado, Montecillo, Texcoco, Estado de México, México.

\*Autor de correspondencia (jmgc@colpos.mx)

#### RESUMEN

La aplicación de nuevas tecnologías para diagnóstico oportuno de problemas fitosanitarios contribuye a mejorar el manejo de los cultivos agrícolas; en particular, los métodos de inteligencia artificial y visión por computadora facilitan el reconocimiento de enfermedades en cultivos de forma automática y confiable. En esta investigación se implementaron dos arquitecturas de redes neuronales convolucionales (CNN) híbridas para identificar 10 clases objetivo: tres enfermedades en frutos de fresa, antracnosis (*Colletotrichum* spp.), cenicilla (*Sphaerotheca macularis*) y moho gris (*Botrytis cinerea*); cuatro enfermedades en hojas de fresa, mancha angular (*Xanthomonas fragariae*), viruela (*Ramularia tulasnei*), cenicilla (*Sphaerotheca macularis*) y quemadura de la hoja (*Diplocarpon earlianum*); deficiencia nutricional de calcio en hojas de fresa, hojas sanas y frutos sanos. En la primera etapa se entrenó el modelo CNN con transferencia de aprendizaje MobileNetv2 (CNN-M) a partir de un conjunto de imágenes digitales RGB para extraer en forma automática un total de 62,720 características, a las cuales se aplicó un análisis de componentes principales. En la segunda etapa, las características transformadas en la última capa de neuronas del modelo CNN-M se utilizaron como entradas para optimizar los modelos de aprendizaje bosque aleatorio (RF) y máquina de soporte vectorial (SVM) y predecir las 10 clases objetivo. Los modelos CNN híbridos alcanzaron una precisión global de clasificación (ACC) superior a 97 %. CNN-M-SVM superó ligeramente a CNN-M-RF con un puntaje promedio F1macro de 97.6 % y ACC de 98.9 %. Los resultados muestran el alto potencial de los modelos convolucionales híbridos con transferencia de aprendizaje para desarrollar herramientas de identificación automática de enfermedades en cultivos de interés agrícola.

**Palabras clave:** *Fragaria* spp., aprendizaje profundo, detección de enfermedades, reconocimiento de patrones, redes neuronales híbridas, transferencia de aprendizaje.

#### SUMMARY

The application of new technologies for early diagnosis of plant health problems contributes to improved management of agricultural crops. In particular, artificial intelligence and computer vision methods facilitate automatic and reliable recognition of crop diseases. In this research, two hybrid convolutional neural network (CNN) architectures were implemented to identify 10 target classes: three strawberry fruit diseases: anthracnose (*Colletotrichum* spp.), powdery mildew (*Sphaerotheca macularis*) and gray mold (*Botrytis cinerea*); four strawberry leaf diseases: angular leaf spot (*Xanthomonas fragariae*), white spot (*Ramularia tulasnei*), powdery mildew

(*Sphaerotheca macularis*) and leaf scorch (*Diplocarpon earlianum*), as well as calcium nutritional deficiency in strawberry leaves, healthy leaves, and healthy fruits. In the first stage, the CNN model was trained using MobileNetv2 (CNN-M) learning transfer from a set of digital RGB images to automatically extract a total of 62,720 features, to which a principal component analysis was applied. In the second stage, the transformed features from the last neural layer of the CNN-M model were used as inputs to optimize the random forest (RF) and vector support machine (SVM) learning models and predict the 10 target classes. The hybrid CNN models achieved an overall classification accuracy (ACC) above 97 %. CNN-M-SVM slightly outperformed CNN-M-RF with an average F1macro score of 97.6 % and ACC of 98.9 %. Results demonstrated the high potential of hybrid convolutional models with learning transfer to develop tools for automatic disease identification in crops of agricultural interest.

**Index words:** *Fragaria* spp, deep learning, disease detection, hybrid neural networks, learning transfer, pattern recognition.

#### INTRODUCCIÓN

México es uno de los principales productores de fresa en el mundo; en 2023, el valor de las exportaciones representó 1220 mdd (SIAP, 2024). El manejo integrado de plagas y enfermedades es una estrategia esencial para mejorar la calidad y rendimiento de fresa; por ello, la incorporación de técnicas de inteligencia artificial y de visión por computadora que faciliten el reconocimiento y detección de problemas fitosanitarios es de gran interés (ur Rehman *et al.*, 2024). Los métodos tradicionales de identificación de enfermedades son laboriosos y propensos a errores; en contraste, las técnicas de visión por computadora facilitan su identificación temprana e intervención oportuna (Hu *et al.*, 2023).

Kim *et al.* (2021) aplicaron el modelo de red neuronal convolucional (CNN) con transferencia de aprendizaje ResNet152 para identificar en el cultivo de fresa las enfermedades pudrición de la corona, causada por hongos y bacterias, y tizón de la hoja (*Neopestalotiopsis* spp.), antracnosis (*Colletotrichum siamense*), cenicilla

(*Sphaerotheca macularis*), moho gris (*Botrytis cinerea*) con una precisión media promedio (mAP, *mean average precision*) de 96 %. Los modelos híbridos combinan las capacidades de extracción de características de CNN con la capacidad de clasificación de modelos de aprendizaje automático, como bosque aleatorio o las máquinas de soporte vectorial (Kumar *et al.*, 2024). Knauer *et al.* (2019) utilizaron un modelo CNN híbrido acoplado a un modelo de bosque aleatorio para identificar especies de árboles a partir de imágenes hiperespectrales, y obtuvieron una precisión global de clasificación (ACC) de 78 %.

Dong *et al.* (2020) utilizaron MobileNetv2 con una ACC de 89 % para clasificar coberturas de suelo y fue superior a los modelos MobileNet1, ResNet152, Xception e Inception-ResNetv2; asimismo, Huerta-Mora *et al.* (2021) compararon los modelos MobileNetv2 y Xception, este último fue superior con una ACC de 92 % con base en imágenes de la base de datos *plant village* (Hughes y Salathé, 2015). Los modelos CNN, vi-transformers, CNN híbridas y modelos CNN YOLO han mostrado alto desempeño para identificar enfermedades de interés agrícola (Chen *et al.*, 2023; Zhu *et al.*, 2024).

El objetivo de esta investigación fue implementar dos arquitecturas de redes neuronales convolucionales híbridas con base en el modelo CNN con transferencia de aprendizaje MobileNetv2 (CNN-M) acoplado a un modelo bosque aleatorio (RF, *random forest*) y a un modelo máquina de soporte vectorial (SVM, *support vector machine*) para clasificar 10 clases objetivo a partir de imágenes digitales RGB (*red, green, blue*).

## MATERIALES Y MÉTODOS

### Clases objetivo

Las 10 clases objetivo fueron: tres enfermedades en frutos de fresa, antracnosis (*Colletotrichum* spp.), cenicilla (*Sphaerotheca macularis*) y moho gris (*Botrytis cinerea*); cuatro enfermedades en hojas de fresa mancha angular (*Xanthomonas fragariae*), viruela (*Ramularia tulasnei*), cenicilla (*Sphaerotheca macularis*), y quemadura de la hoja (*Diplocarpon earlianum*), así como deficiencia nutricional de calcio en hojas, hojas sanas y frutos sanos.

### Conjunto de datos

En este estudio se realizaron ocho etapas desde la colecta de datos hasta la evaluación del desempeño en predicción de los modelos convolucionales híbridos CNN-M-SVM y CNN-M-RF (Figura 1). La extracción automática de características con CNN-M se realizó a partir de la capa de aplanado y de la última capa completamente conectada.

Las imágenes que se utilizaron en este estudio provienen de diferentes estudios reportados en la literatura. Se usaron imágenes con enfermedades en frutos y hojas de fresa (Afzaal *et al.*, 2021), hojas enfermas y sanas de la base de datos *plant village* (Hughes y Salathé, 2015), frutos frescos y podridos (Choi *et al.*, 2022), frutas (Shreya, 2023), frutos frescos y podridos (Sultana *et al.*, 2022) y deficiencia de calcio en hojas (Hariri y Avşar, 2022) (Cuadro 1).

Las imágenes se colectaron bajo diferentes condiciones de iluminación y se capturaron con diferente resolución, tamaño y fondo; por lo cual, éstas se homogeneizaron a un tamaño de 224 × 224 píxeles, resolución de 96 ppp y modelo de color RGB (*red, green, blue*); después, cada imagen se guardó en formato JPG de acuerdo con la entrada estándar del modelo MobileNetv2 (Figura 2).

### Hardware y software

Los algoritmos se ejecutaron en el lenguaje *Python* 3.10.11 con *tensorflow* 2.12 como *back end* y en la plataforma *Keras* 2.12 como *front end*; se utilizaron los siguientes módulos de *Python*: *Numpy* 1.22.4, *Pandas* 1.5.3, *OpenCV* 4.6.0, *Matplotlib* 3.7.1 y *Seaborn* 0.12.2. La ejecución de los códigos se realizó en la plataforma Web Google-Colab (Bisong, 2019).

En el entrenamiento de CNN-M y CNN-M-SVM se utilizó una GPU (*graphics processing unit*) Nvidia A100 con 40 GB de memoria de video HBM2G y 81 GB de memoria RAM con 256 GB de disco duro. Para el modelo CNN-M-RF se utilizó una capacidad RAM de 51 GB de memoria, 256 GB de disco duro y CPU Intel Xeon de 2.20 GHz; además, se utilizó un almacenamiento en Google Drive de 15 GB para guardar las imágenes, códigos y modelos finales.

### Arquitecturas de redes neuronales convolucionales

Una red neuronal convolucional (CNN) realiza tres tareas principales: 1) transforma una imagen de entrada a un tensor de dimensión  $a \times h \times ch$ , donde  $a$  es el ancho,  $h$  es la altura y  $ch$  es el número de canales RGB; realiza un producto escalar entre el tensor y el filtro, a este resultado se le aplica la función de activación ReLU y se realiza un agrupamiento con base en valores máximos (*maxpooling*) o valores promedio (*averagepooling*) para extraer las características más relevantes de la imagen; 2) transforma el tensor resultante de las capas de convolución a un vector unidimensional que contiene las características extraídas y éstas se introducen a un bloque de neuronas completamente conectadas, y 3) aplica una función de activación *softmax* a la última capa de salida para calcular la probabilidad de pertenencia para cada clase objetivo (Dong *et al.*, 2020).

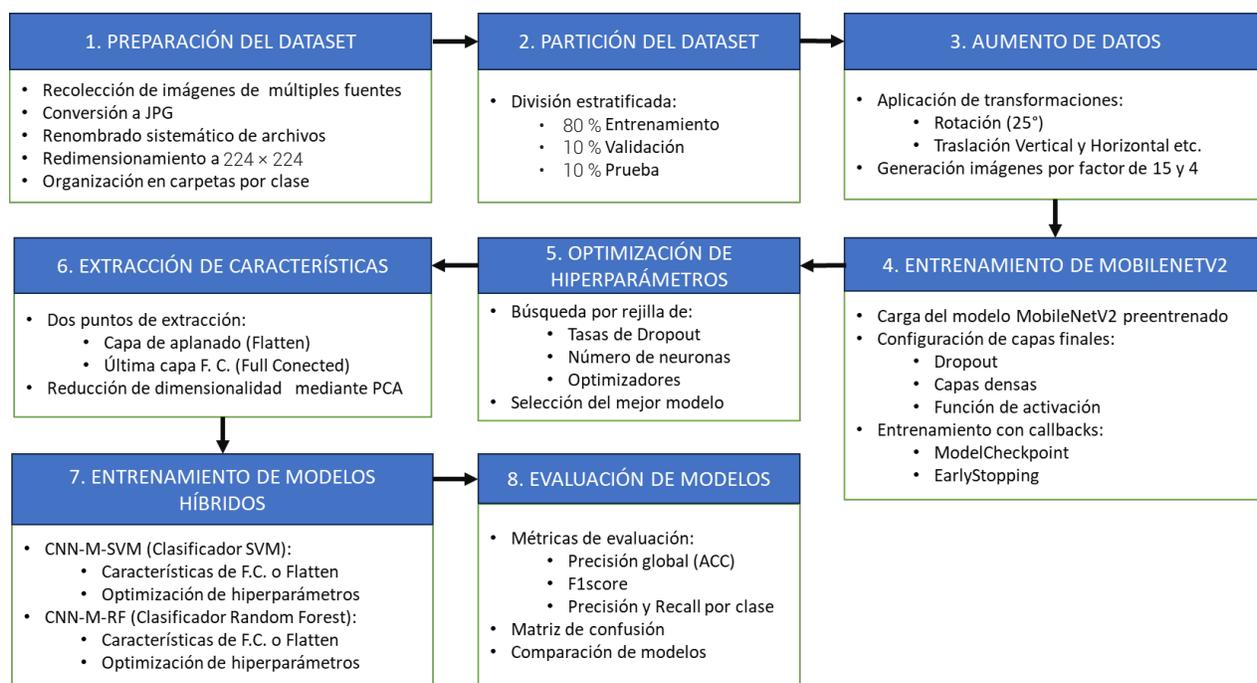


Figura 1. Esquema del procedimiento realizado en la presente investigación. Elaboración propia.

Cuadro 1. Descripción del conjunto de imágenes por clase objetivo considerada en este estudio.

| Clase objetivo | Nombre común          | Agente causal                 | Cantidad |
|----------------|-----------------------|-------------------------------|----------|
| <i>an</i>      | Antracnosis           | <i>Colletotrichum</i> spp.    | 97       |
| <i>cf</i>      | Cenicilla en fruto    | <i>Sphaerotheca macularis</i> | 135      |
| <i>mg</i>      | Moho gris del fruto   | <i>Botrytis cinerea</i>       | 477      |
| <i>ma</i>      | Mancha angular        | <i>Xanthomonas fragariae</i>  | 435      |
| <i>vi</i>      | Víruela               | <i>Ramularia tulasnei</i>     | 615      |
| <i>ch</i>      | Cenicilla en hoja     | <i>Sphaerotheca macularis</i> | 533      |
| <i>qh</i>      | Quemadura de la hoja  | <i>Diplocarpon earlianum</i>  | 1,109    |
| <i>dc</i>      | Deficiencia de calcio |                               | 805      |
| <i>hs</i>      | Hojas sanas           |                               | 875      |
| <i>fs</i>      | Frutos sanos          |                               | 465      |
|                |                       | Total                         | 5,546    |

El modelo CNN con transferencia de aprendizaje MobileNetv2 (CNN-M) está diseñado para dispositivos móviles, equipos con bajos recursos de cómputo y uso de imágenes con resolución 224 × 224 píxeles. CNN-M aplica filtros sobre el mapa de características en canales separados con filtro de 1 × 1, este método es más eficiente que una CNN estándar. Este modelo utiliza operaciones

lineales para reducir la dimensión de los datos antes de pasar a la capa de convolución, sin perder información en el proceso de extracción de características (Sandler et al., 2018). CNN-M es un modelo de clasificación pre-entrenado con imágenes de ImageNet (Deng et al., 2009).

En este estudio, los parámetros de la última capa de

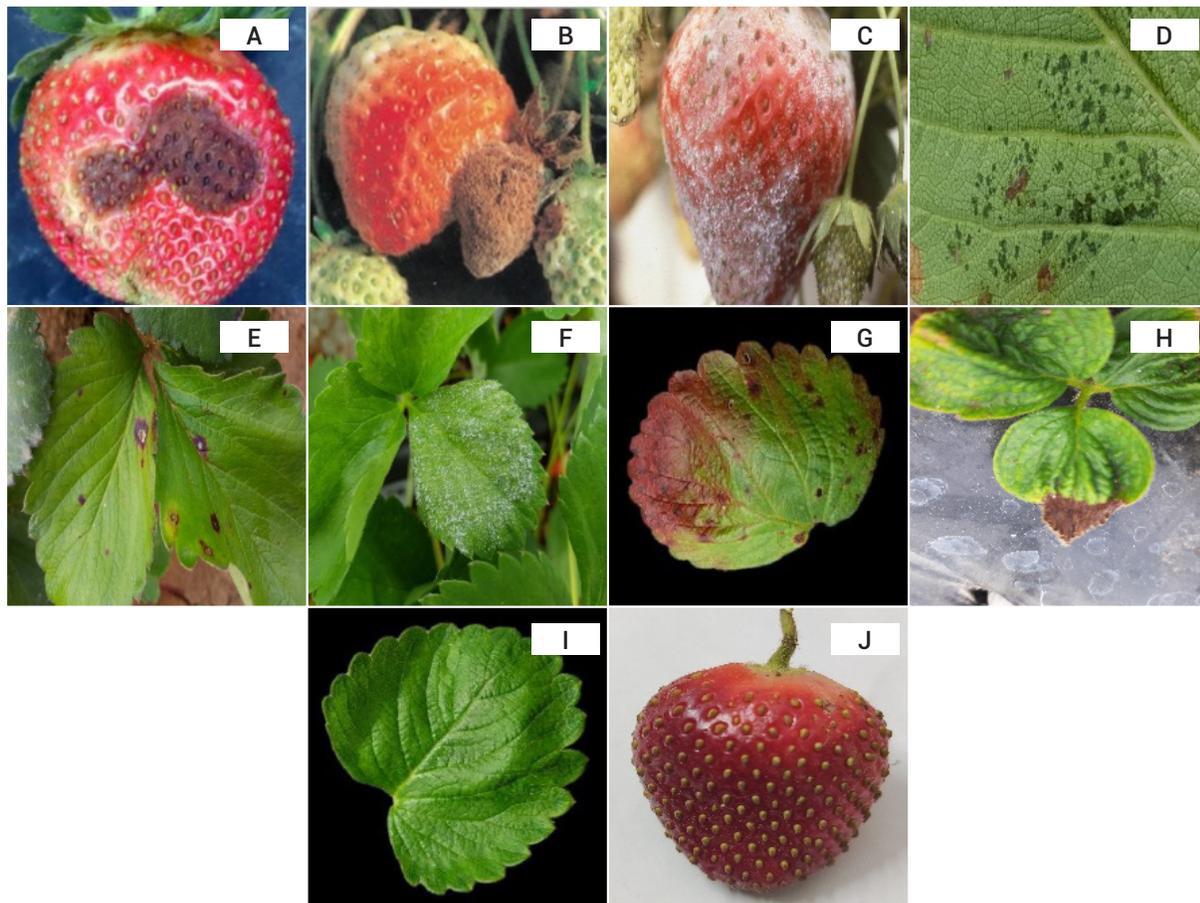


Figura 2. Ejemplos de las 10 clases objetivo: A) antracnosis (*Colletotrichum* spp.), B) cenicilla en fruto (*Sphaerotheca macularis*), C) moho gris del fruto (*Botrytis cinerea*), D) mancha angular (*Xanthomonas fragariae*), E) viruela (*Ramularia tulasnei*), F) cenicilla en hoja (*Sphaerotheca macularis*), G) quemadura de la hoja (*Diplocarpon earlianum*), H) deficiencia de calcio, I) hojas sanas, J) frutos sanos.

CNN-M (Figura 3-D) se definieron como entrenables y los hiperparámetros después de la capa de aplanado se establecieron como sigue: *dropout* (0.5, 0.6) desactiva de manera aleatoria neuronas durante el proceso de entrenamiento para prevenir sobreajuste (*overfitting*) del modelo. Se utilizaron dos capas de 256 neuronas completamente conectadas, la capa de salida utilizó una función de activación *softmax* para calcular la probabilidad de pertenencia de cada clase objetivo, la función de pérdida *categorical crossentropy* se utilizó como función objetivo durante el entrenamiento del modelo, la métrica precisión global de clasificación (ACC, *Accuracy*) se utilizó para monitorear el desempeño del modelo, y el optimizador *Adamax* se utilizó para determinar los parámetros de la red convolucional durante el entrenamiento (Géron, 2022).

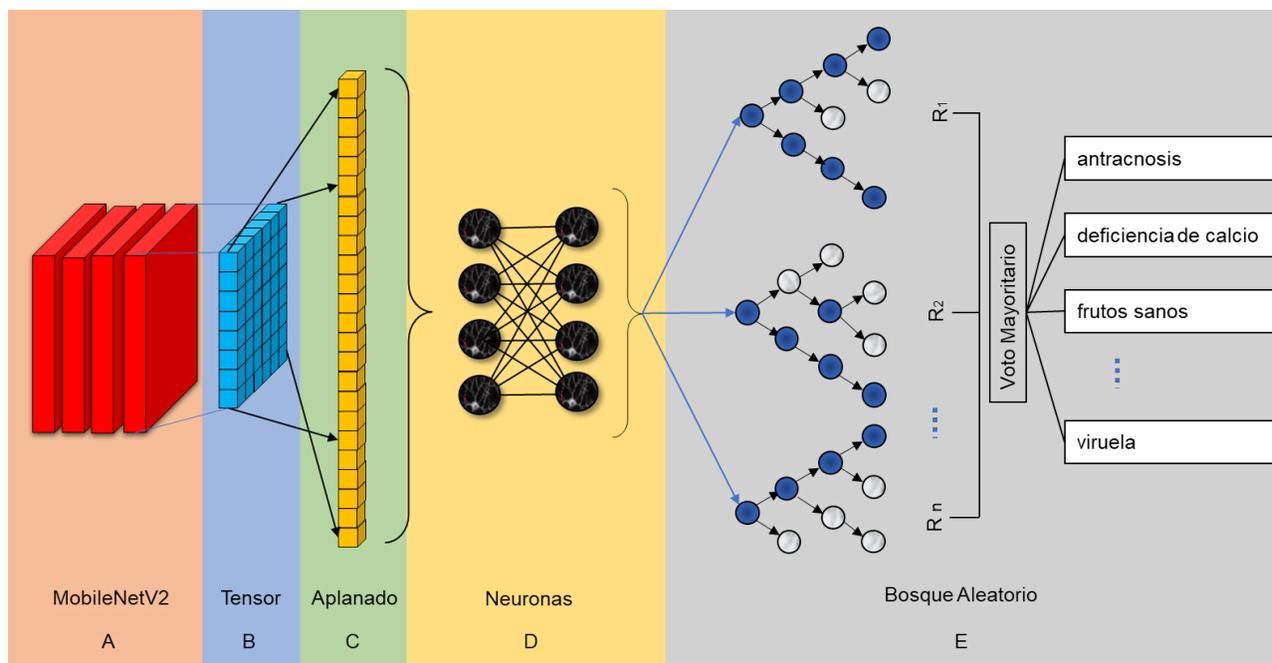
#### Modelos CNN-M híbridos

Los modelos CNN-M híbridos se construyen en dos

etapas, en la primera se entrena el modelo CNN-M y en la segunda se extraen las características de entrada de los modelos de aprendizaje. Estas entradas se obtienen en dos capas diferentes, en la capa de aplanado (F, *flatten*) y en la última capa de neuronas completamente conectadas (FC, *full connected*) del modelo CNN-M (Knauer *et al.*, 2019; Kumar *et al.*, 2024). En este estudio se utilizaron dos modelos híbridos CNN-M-SVM y CNN-RF (Figura 3).

CNN-M-SVM es un modelo CNN-M acoplado a un modelo de aprendizaje SVM que se entrena con las características extraídas de CNN-M para clasificar las clases objetivo. SVM maximiza el margen *dis* que separa las clases mediante un hiperplano, con base en puntos cercanos llamados vectores soporte (Raschka y Mirjalili, 2019).

$$dis = \max \frac{2}{\|w\|}$$



**Figura 3. Arquitectura de una red neuronal convolucional híbrida con bosque aleatorio (RF): A) MobileNetV2, B) tensor de salida, C) vector de características extraídas, D) capa de neuronas completamente conectadas, E) clasificador RF. Elaboración propia.**

donde *dis* es la distancia entre dos vectores y *w* es el vector de pesos del hiperplano.

La forma dual para expresar *dis* es buscar los vectores de soporte ( $x_i, x_j$ ) que maximicen la siguiente función:

$$W(\alpha) = \sum_{j=1}^l \alpha_j - \frac{1}{2} \sum_{ij=1}^l \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j)$$

donde  $\alpha_i$  y  $\alpha_j$  son multiplicadores de Lagrange,  $y_i$  es la clase objetivo de la muestra  $x_i$ ,  $y_j$  es la clase objetivo de la muestra  $x_j$  y  $K(x_i, y_j)$  es el kernel o núcleo del producto escalar, sujeta a:

$$\sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq \frac{C}{l}, i = 1, \dots, l$$

donde *C* es el parámetro de regularización y *l* es el número total de muestras (Raschka y Mirjalili, 2019).

El modelo CNN-M-RF híbrido utiliza como entrada del modelo RF las características extraídas con CNN-M (Figura 3). RF es un clasificador que consiste en un conjunto de árboles de decisión, los cuales se forman a partir de muestras aleatorias con reemplazo del conjunto de datos de entrenamiento. Los nodos de cada árbol se dividen con base en características extraídas de forma aleatoria. Cada árbol de RF realiza una predicción de clase;

en seguida, la predicción final de una clase objetivo se realiza por voto mayoritario (Khozeimeh *et al.*, 2022). El número de características *nc* en cada nodo se determinó por  $nc = \sqrt{m}$ , donde *m* es el número de características del conjunto de entrenamiento. En general RF es robusto y evita el sobreajuste (Raschka y Mirjalili, 2019).

**Métricas de desempeño**

La evaluación del desempeño de los modelos de predicción se realizó con base en ACC y F1macro; esta última es una métrica más conservadora que ACC cuando se tienen muestras de clases desbalanceadas. Las métricas se obtienen con las predicciones del conjunto de datos de prueba (*test*) y con la matriz de confusión que contiene diferentes conteos de valores: verdaderos positivos (TP, *true positive*) son valores verdaderos que se clasifican como verdaderos; verdaderos negativos (TN, *true negative*) son valores negativos que se clasifican como negativos; falsos positivos (FP) son valores negativos que se clasifican como positivos; falsos negativos (FN) son valores positivos que se clasifican como negativos (Raschka y Mirjalili, 2019).

ACC se define como:

$$ACC = (TP + TN) / (TP + FP + TN + FN)$$

Precisión ( $P$ ) se define como:

$$P = TP / (TP + FP)$$

Sensibilidad ( $S$ ) se calcula como:

$$S = TP / (TP + FN)$$

F1 se calcula como:

$$F1 = (2 \times P \times S) / (P + S)$$

F1 macro se calcula como:

$$F1_{macro} = \frac{1}{C} \sum_{i=1}^C F1_i$$

donde  $C$  es el número total de clases objetivo.

### Entrenamiento y prueba de los modelos convolucionales

Las imágenes del conjunto de datos se agruparon por clase objetivo en carpetas con el nombre de cada clase; posteriormente, las imágenes en cada carpeta se identificaron con el nombre de la clase y un número consecutivo ascendente. Las imágenes se redimensionaron a un tamaño  $224 \times 224$  píxeles, de acuerdo con la entrada del modelo MobileNetv2.

### Partición y aumento del conjunto de datos

El conjunto original de datos se particionó de forma aleatoria y estratificada por clase objetivo en tres

subconjuntos: entrenamiento (80 %), validación (10 %) y prueba (10 %) (Pham *et al.*, 2020). Los conjuntos de entrenamiento se aumentaron, para cada clase, por un factor de cuatro, excepto las clases antracnosis y cenicilla que se incrementaron por un factor de 15 por medio de transformaciones de cada imagen. Con las siguientes transformaciones: rotación ( $25^\circ$ ), volteado horizontal, volteado vertical, brillo (de 0.5 a 1.5), traslación horizontal (0.2), traslación vertical (0.2), distorsión (0.2) y zoom (0.3) el conjunto de entrenamiento se incrementó a 22,175 imágenes para mejorar la capacidad de aprendizaje de los modelos (Chollet, 2021).

### Selección de hiperparámetros óptimos

La selección de hiperparámetros óptimos del modelo CNN-M se realizó mediante una búsqueda por rejilla que maximicen ACC. Los intervalos de valores fueron para *dropout* = 0.3, 0.5, 0.7; *neuronas* = 300, 400, 600, y *optimizador* = adamax, rmsprop, adadelta. El entrenamiento de los modelos se limitó a 100 ciclos (*epochs*) con tamaño de lote (*batch*) de 50 imágenes en entrenamiento y 10 en validación, y se aplicó la técnica de regularización *earlystopping*, con una tolerancia de cinco ciclos para evitar el sobreajuste.

La selección de hiperparámetros óptimos de SVM y RF se realizó por medio de una búsqueda por rejilla y validación cruzada (CV, *cross validation*) con  $k = 3$  particiones. Los modelos SVM y RF utilizaron diferentes intervalos de valores de búsqueda para sus hiperparámetros (Cuadro 2).

**Cuadro 2. Intervalo de valores para la búsqueda de hiperparámetros óptimos de los modelos máquina de soporte vectorial (SVM) y bosque aleatorio (RF) con características de entrada extraídas del modelo de red neuronal convolucional con transferencia de aprendizaje MobileNetv2 (CNN-M).**

| Clasificador | Capa <sup>†</sup> | Hiperparámetros <sup>††</sup> | Intervalo de valores                     |
|--------------|-------------------|-------------------------------|--|
| CNN-M-RF     | F y FC            | $n$                           | 250, 300, 350, 400, 450                  |
|              |                   | $f$                           | gini, entropy, log loss                  |
|              |                   | $d$                           | 8, 12, 14, 16, 18, 20                    |
|              |                   | $mx$                          | $\sqrt{\text{log2}}$                     |
| CNN-M-SVM    | F                 | $C$                           | 0.001, 0.01, 0.1, 1                      |
|              |                   | $\gamma$                      | 0.001, 0.01, 0.1, 1                      |
|              | FC                | $C$                           | 0.001, 0.01, 0.1, 1, 10, 100, 1000, 2000 |
|              |                   | $\gamma$                      | 0.001, 0.01, 0.1, 1, 10, 100, 1000, 2000 |

<sup>†</sup>F: características extraídas de la capa de aplanado (*flatten*) de CNN-M, FC: características extraídas de la capa completamente conectada de CNN-M. <sup>††</sup>C: parámetro de regularización,  $\gamma$ : parámetro del *kernel* lineal,  $n$ : número de árboles,  $f$ : función objetivo,  $d$ : profundidad del árbol,  $mx$ : máximo de características para dividir el árbol.

El entrenamiento de los clasificadores SVM y RF, con base en las características extraídas en la capa de aplanado F, se realizó con un conjunto de características reducidas de 9,000 componentes principales que representaron 92 % de varianza total acumulada. La optimización de CNN-M se obtuvo con base en la curva de aprendizaje, *versus* ciclos de tiempo, para determinar el ciclo de entrenamiento con máximo desempeño (Kumar *et al.*, 2024). La evaluación en predicción de los modelos se realizó con el conjunto de datos de prueba (10 %), el cual consistió de una partición aleatoria estratificada por clase objetivo del conjunto original de imágenes.

## RESULTADOS Y DISCUSIÓN

### Hiperparámetros óptimos

Los hiperparámetros óptimos del modelo CNN-M fueron: *dropout* = 0.5, *neuronas* = 300 en las últimas dos capas y *optimizador* = *adamax* con ocho ciclos de entrenamiento (Figura 4). Los hiperparámetros del clasificador SVM acoplado a CNN-M con la capa F fueron (*C* = 0.001,  $\gamma$  = 0.001), y con la capa FC (*C* = 0.1,  $\gamma$  = 0.001) con un tiempo de entrenamiento de 6 h 45 m. Similarmente, los hiperparámetros del clasificador RF acoplado a CNN-M con la capa F fueron (*f* = *gini*, *d* = 20, *mx* = *log2*, *n* = 300) y con la capa FC fueron (*f* = *gini*, *d* = 16, *mx* = *sqrt*, *n* = 350) con un tiempo de entrenamiento de 2 h 55 m.

### Desempeño en predicción de los modelos híbridos

CNN-M-SVM fue ligeramente superior a CNN-M-RF,

con *F1* macro = 97.6 y *ACC* = 98.9. CNN-M-SVM alcanzó su máximo desempeño *F* = 9000 componentes y CNN-M-RF alcanzó *F1* macro = 97.2 y *ACC* = 98.7, con *FC* = 300 componentes (Cuadro 3).

El desempeño en predicción de los modelos híbridos a nivel de clases, con base en la matriz de confusión, CNN-M-SVM obtuvo su menor desempeño para clasificar antracnosis en frutos (*an*), mientras que frutos sanos (*fs*) y frutos con moho gris (*mg*) fueron confundidos por el modelo; las demás clases objetivo fueron clasificadas con valores de sensibilidad (*S*) superiores a 93 % (Figura 5).

En términos de *F1*, CNN-M-SVM obtuvo su menor valor para identificar la clase *an* con *F1* de 88.9 %, debido principalmente al número reducido de la clase objetivo *an* con (97 imágenes). El modelo alcanzó un alto desempeño para clasificar las nueve clases restantes con *F1* superiores a 96 % (Cuadro 4).

Los resultados obtenidos muestran que los modelos CNN híbridos mejoran la precisión global de clasificación (*ACC*) de los modelos CNN. En este estudio, el modelo de aprendizaje SVM fue ligeramente superior a RF; sin embargo, SVM utilizó 9000 características extraídas de CNN-M y RF utilizó 300 características de CNN-M, con desempeños ligeramente inferiores. En términos de tiempos computacionales CNN-M-SVM empleó aproximadamente dos veces más recursos que CNN-M-RF. Los resultados coinciden con lo reportado por Khozeimeh *et al.* (2022).

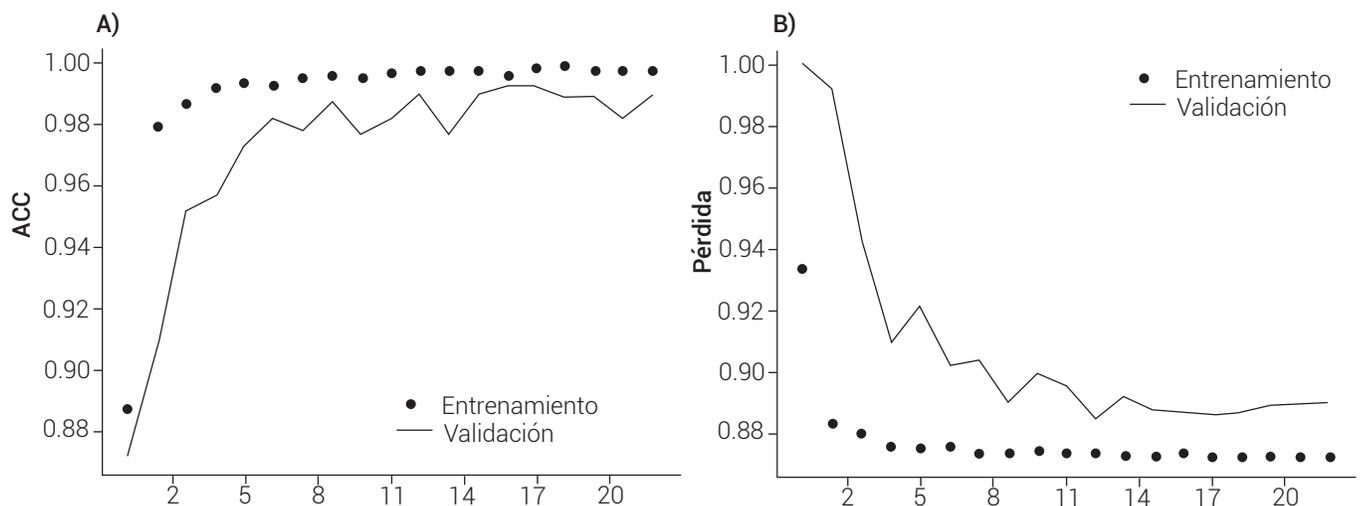
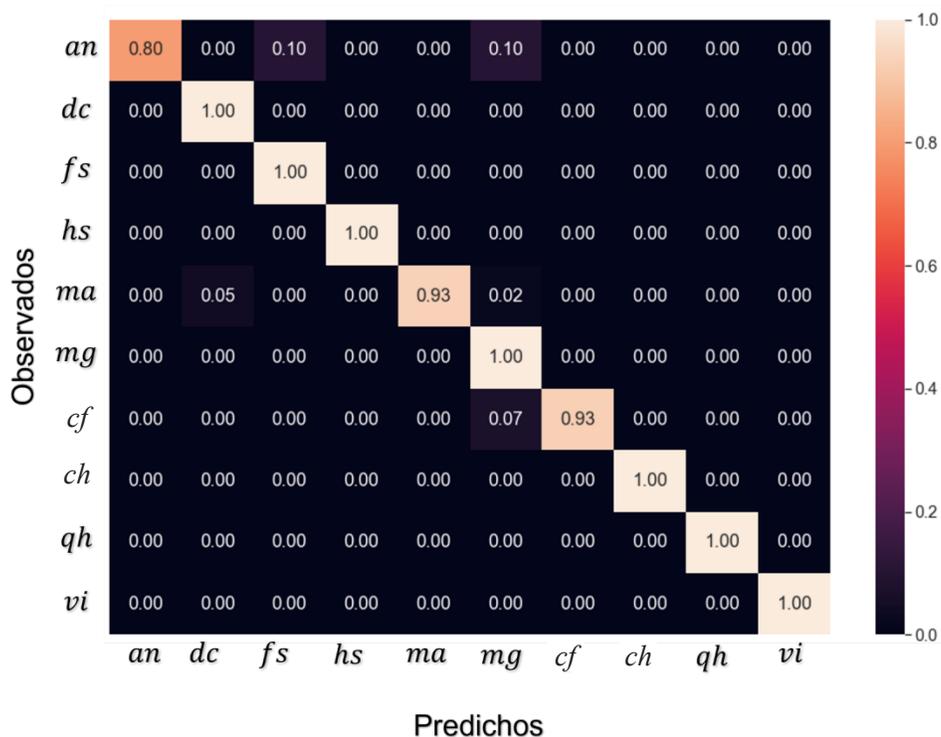


Figura 4. A) Curva de aprendizaje del modelo CNN-M en función de la precisión global de clasificación ACC, B) Curva de aprendizaje del modelo CNN-M en función de la pérdida o error.

**Cuadro 3. Evaluación del desempeño en predicción de los modelos convolucionales híbridos con base en el conjunto de prueba.**

| Modelo    | Capa <sup>†</sup> | NCE <sup>††</sup> | F1macro (%) | ACC (%) |
|-----------|-------------------|-------------------|-------------|---------|
| CNN-M-SVM | FC                | 300               | 96.8        | 98.6    |
|           | F                 | 9000              | 97.6        | 98.9    |
| CNN-M-RF  | FC                | 300               | 97.2        | 98.7    |
|           | F                 | 9000              | 96.8        | 98.6    |

<sup>†</sup>F: capa de aplanado, FC: completamente conectada, <sup>††</sup>NCE: componentes principales extraídos.



**Figura 5. Matriz de confusión para la clasificación de las 10 clases objetivo con el modelo CNN-M-SVM con el conjunto de datos de prueba (test): an: antracnosis, dc: deficiencia de calcio, fs: frutos sanos, hs: hojas sanas, ma: mancha angular, mg: moho gris en fruto, cf: cenicilla en fruto, ch: cenicilla en hoja, qh: quemadura de la hoja, vi: viruela.**

**CONCLUSIONES**

El modelos híbrido CNN-M-SVM fue ligeramente superior a CNN-N-RF; sin embargo, ambos modelos alcanzaron una precisión global de clasificación (ACC) superior a 98.5 %. A nivel de clases objetivo, CNN-M-SVM alcanzó un puntaje F1 superior a 96 % en nueve de las clases

objetivo, y un valor F1 de 88.9 % para identificar la clase antracnosis en frutos de fresa. Los resultados muestran el potencial de los modelos CNN híbridos para el desarrollo de plataformas web y aplicaciones en dispositivos móviles para el diagnóstico automático de enfermedades en cultivos a partir de imágenes.

**Cuadro 4. Métricas de Precisión (P), Sensibilidad (S) y F1 obtenidas de la matriz de confusión para el modelo CNN-M-SVM para cada clase objetivo.**

| Clase     | (%)  | (%)             | (%)  |
|-----------|------|-----------------|------|
| <i>an</i> | 80.0 | 100             | 88.9 |
| <i>dc</i> | 100  | 97.6            | 98.8 |
| <i>fs</i> | 100  | 97.9            | 98.9 |
| <i>hs</i> | 100  | 100             | 100  |
| <i>ma</i> | 93.2 | 100             | 96.5 |
| <i>mg</i> | 100  | 94.1            | 97.0 |
| <i>cf</i> | 92.9 | 100             | 96.3 |
| <i>ch</i> | 100  | 100             | 100  |
| <i>qh</i> | 100  | 100             | 100  |
| <i>vi</i> | 100  | 100             | 100  |
|           |      | <i>F1 macro</i> | 97.6 |

*an*: antracnosis, *dc*: deficiencia de calcio, *fs*: frutos sanos, *hs*: hojas sanas, *ma*: mancha angular, *mg*: moho gris en fruto, *cf*: cenicilla en fruto, *ch*: cenicilla en hoja, *qh*: quemadura de la hoja, *vi*: viruela.

**AGRADECIMIENTOS**

El autor principal expresa su agradecimiento al CONAHCYT (Actualmente SECIHTI) y al COMECYT por la beca otorgada para realizar esta investigación durante sus estudios de posgrado.

**BIBLIOGRAFÍA**

Afzaal U., B. Bhattarai, Y. R. Pandeya and J. Lee (2021) An instance segmentation model for strawberry diseases based on mask R-CNN. *Sensors* 21:6565, <https://doi.org/10.3390/s21196565>

Bisong E. (2019) Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. Apress. Berkeley, California, USA. 709 p, <https://doi.org/10.1007/978-1-4842-4470-8>

Chen S., Y. Liao, F. Lin and B. Huang (2023) An improved lightweight YOLOv5 algorithm for detecting strawberry diseases. *IEEE Access* 11:54080-54092, <https://doi.org/10.1109/ACCESS.2023.3282309>

Choi T., O. Would, A. Salazar-Gomez and G. Cielniak (2022) Self-supervised representation learning for reliable robotic monitoring of fruit anomalies. In: 2022 International Conference on Robotics and Automation. Institute of Electrical and Electronics Engineers. Philadelphia, Pennsylvania, USA. pp:2266-2272, <https://doi.org/10.48550/arXiv.2109.10135>

Chollet F. (2021) Deep Learning with Python. 2nd edition. Manning Publications. Shelter Island, New York, USA. 400 p.

Deng J., W. Dong, R. Socher, L. J. Li, K. Li and F. F. Li (2009) ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Institute of Electrical and Electronics Engineers. Miami, Florida, USA. pp:248-255, <https://doi.org/10.1109/CVPR.2009.5206848>

Dong K., C. Zhou, Y. Ruan and Y. Li (2020) MobileNetv2 model for image classification. 2020 2nd International Conference on Information Technology and Computer Application. Guangzhou, China. pp:476-480, <https://doi.org/10.1109/ITCA52113.2020.00106>

Géron A. (2022) Hands-On Machine Learning with Scikit-Learn, Keras & Tensorflow. Concepts, Tools, and Techniques to Build Intelligent Systems. 3rd edition. O'Reilly Media. Sebastopol, California, USA. 861 p.

Hariri M. and E. Avşar (2022) Tipburn disorder detection in strawberry leaves using convolutional neural networks and particle swarm optimization. *Multimedia Tools and Applications* 81:11795-11822, <https://doi.org/10.1007/s11042-022-12759-6>

Hu X., R. Wang, J. Du, Y. Hu, L. Jiao and T. Xu (2023) Class-attention-based lesion proposal convolutional neural network for strawberry diseases identification. *Frontiers in Plant Science* 14:1091600, <https://doi.org/10.3389/fpls.2023.1091600>

Huerta-Mora E. A., V. González-Huitrón, H. Rodríguez-Rangel y L. E. Amabilis-Sosa (2021) Detección de enfermedades foliares con arquitecturas de redes neuronales convolucionales. *RINDERESU, Revista Internacional de Desarrollo Regional Sustentable* 5:18-40.

Hughes D. P. and M. Salathé (2015) An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv:1511.08060*, <https://doi.org/10.48550/arXiv.1511.08060>

Khozeimeh F., D. Sharifrazi, N. H. Izadi, J. H. Joloudari, A. Shoeibi, R. Alizadehsani, ... and M. S. Islam (2022) RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance. *Scientific Reports* 12:11178, <https://doi.org/10.1038/s41598-022-15374-5>

Kim B., Y. K. Han, J. H. Park and J. Lee (2021) Improved vision-based detection of strawberry diseases using a deep neural network. *Frontiers in Plant Science* 11:559172, <https://doi.org/10.3389/fpls.2020.559172>

Knauer U., C. S. von Rekowsky, M. Stecklina, T. Krokotsch, T. P. Minh, V. Hauffe, ... and U. Seiffert (2019) Tree species classification based on hybrid ensembles of a convolutional neural network (CNN) and random forest classifiers. *Remote Sensing* 11:2788, <https://doi.org/10.3390/rs11232788>

Kumar R. R., R. Chauhan, S. A. Dhondiyal and A. Singh (2024) Deep learning-driven diagnosis a CNN-SVM hybrid approach for automated detection of strawberry leaf diseases. *IEEE 3rd World Conference on Applied Intelligence and Computing*. Gwalior, India. pp:1467-1470, <https://doi.org/10.1109/AIC61668.2024.10731051>

Pham B. T., C. Qi, L. S. Ho, T. Nguyen-Thoi, N. Al-Ansari, H. D. Nguyen, ... and I. Prakash (2020) A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil. *Sustainability* 12:2218, <https://doi.org/10.3390/su12062218>

Raschka S. and V. Mirjalili (2019) Python Machine Learning. Machine Learning and Deep Learning with Python, Scikit-Learn, and

- Tensorflow 2. 3rd edition. Packt Publishing. Birmingham, UK. 772 p.
- Sandler M., A. Howard, M. Zhu, A. Zhmoginov and L. C. Chen (2018)** MobileNetv2: inverted residuals and linear bottlenecks. *In*: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, Utah, USA. pp:4510-4520, <https://doi.org/10.1109/CVPR.2018.00474>
- Shreya M. (2023)** Fruits dataset (Images). Exploring 9 popular fruits through a comprehensive image dataset Kaggle. San Francisco, California, USA. <https://doi.org/10.34740/KAGGLE/DSV/5514079> (February 2025).
- SIAP, Servicio de Información Agroalimentaria y Pesquera (2024)** Panorama Agroalimentario. La Ruta de la Transformación Agroalimentaria 2018-2024. México. Servicio de Información Agroalimentaria y Pesquera. Ciudad de México, México. 207 p.
- Sultana N., M. Jahan and M. S. Uddin (2022)** Fresh and rotten fruits dataset for machine-based evaluation of fruit quality. Mendeley Data Repository. Amsterdam, The Netherlands. <https://doi.org/10.17632/bdd69gyhv8.1> (August 2025).
- ur Rehman M. M., J. Liu, A. Nijabat, M. Faheem, W. Wang and S. Zhao (2024)** Leveraging convolutional neural networks for disease detection in vegetables: a comprehensive review. *Agronomy* 14:2231, <https://doi.org/10.3390/agronomy14102231>
- Zhu X., Fan J., X. Yan, T. Mao, J. Huang, Y. Zhang and H. Nie (2024)** SDD: Strawberry disease detection framework based on computer vision neural networks. *In*: 2024 5th International Conference on Computer Vision, Image and Deep Learning. Zhuhai, China. pp:1077-1082, <https://doi.org/10.1109/CVIDL62147.2024.10603536>