Grupo 5
Integrantes del equipo: Eduardo Mussini, Mathías Pérez, Diego García

Clasificación de imágenes (detección de empleo de mascarilla facial)

22 de agosto 2025

1. Selección del Modelo Pre-entrenado.

Se seleccionó el modelo EfficientNetB0 preentrenado en ImageNet porque representa un buen balance entre rendimiento y eficiencia. Este modelo ya aprendió características generales de las imágenes, como bordes y texturas, lo que facilita la adaptación a la tarea de clasificación de rostros con y sin mascarilla. Además, al ser una arquitectura más moderna y liviana que otras opciones clásicas como VGG16, permite entrenar en menos tiempo y con menor riesgo de sobreajuste. Con respecto al conjunto de datos que emplea el modelo, este contiene 1,28 millones de imágenes de entrenamiento, 50,000 imágenes de validación y 1000 categorías diferentes. En este sentido, se consideró adecuado para el tamaño y la complejidad del conjunto de datos utilizado.

2. Preparación del dataset y ajuste del modelo.

Las imágenes se redimensionaron a 224×224 píxeles, ya que ese es el tamaño de entrada requerido por el modelo preentrenado EfficientNetB0. Además, se aplicó un split estratificado de los datos, reservando un 20% para validación y prueba, y el resto para entrenamiento.

Tal como fue mencionado anteriormente, se empleó EfficientNetB0 con pesos preentrenados en ImageNet y con la capa superior desactivada (include_top=False), de manera que sirviera como extractor de características. Posteriormente, se añadieron capas propias: una de Global Average Pooling para reducir la dimensionalidad de los mapas de activación, una capa de Dropout (0.3) para evitar sobreajuste, y finalmente una capa Dense con activación softmax adaptada a las tres clases del problema.

El ajuste se realizó en dos fases:

Fase 1 (entrenamiento de la cabeza): Se congelaron las capas convolucionales de EfficientNetB0, entrenando únicamente las capas añadidas. Se utilizó el optimizador Adam, función de pérdida categorical crossentropy y la métrica accuracy. El entrenamiento se configuró con un máximo de 30 épocas y early stopping sobre la pérdida de validación.

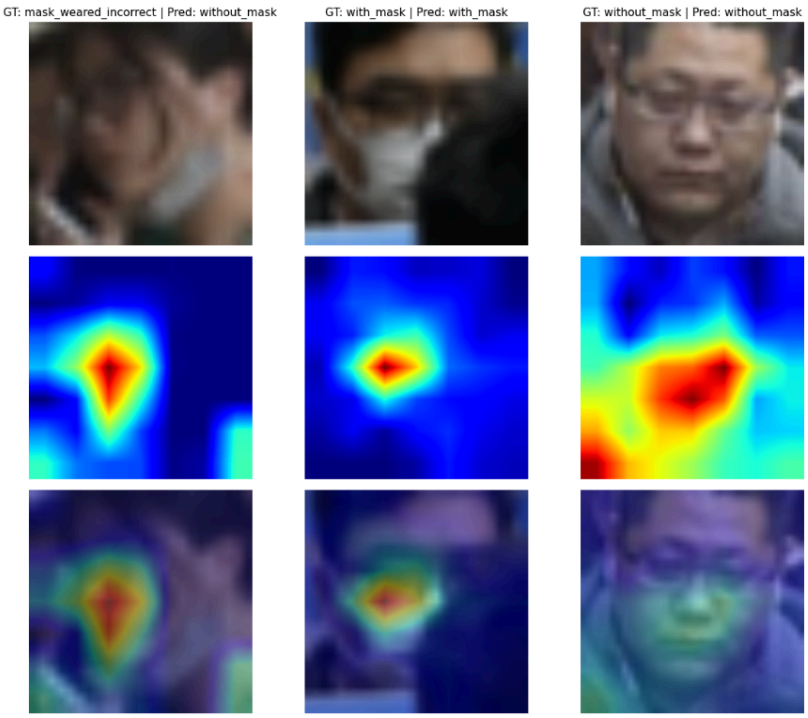
Fase 2 (fine-tuning): Se descongelaron selectivamente las capas superiores de EfficientNetB0 (bloques 6 y 7, excluyendo BatchNormalization) y se reentrenó el modelo con una tasa de aprendizaje muy baja (1e-5). Esta estrategia permitió refinar las representaciones del modelo manteniendo la estabilidad de los pesos preentrenados.

3. Selección y uso de una técnica de explicabilidad.

Para explicar las predicciones del modelo se utilizó Grad-CAM, una técnica que permite visualizar qué regiones de la imagen influyen más en la decisión de la red neuronal. Esta elección se debe a que es un método estándar en clasificación de imágenes con CNNs, no requiere modificar la arquitectura del modelo y resulta computacionalmente ligero. El procedimiento consistió en tomar los mapas de características previos a la capa de Global Average Pooling y calcular el gradiente del logit de la clase predicha con respecto a dichos mapas. A partir de esos gradientes se obtuvieron pesos de importancia por canal, que luego se combinaron con los mapas para generar un heatmap. Finalmente, este mapa se normalizó y se superpuso a la imagen original, mostrando de forma visual qué zonas fueron más determinantes en la predicción.

La aplicación en el problema de detección de mascarillas evidenció que el modelo suele concentrarse en la zona de la nariz, la boca y el contorno de la mascarilla, lo que confirma que está utilizando información relevante.

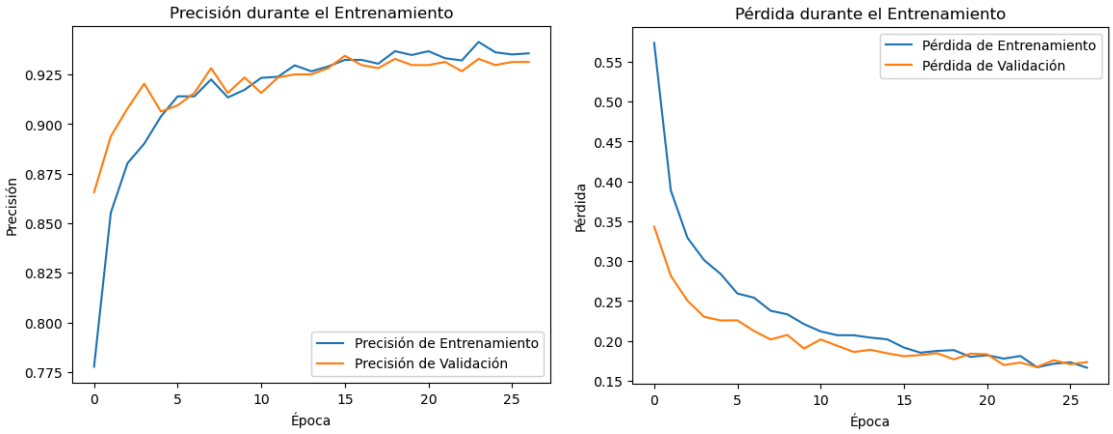
Gráfica de explicabilidad del modelo mediante Grad-CAM: Validación de predicciones de uso de mascarilla"



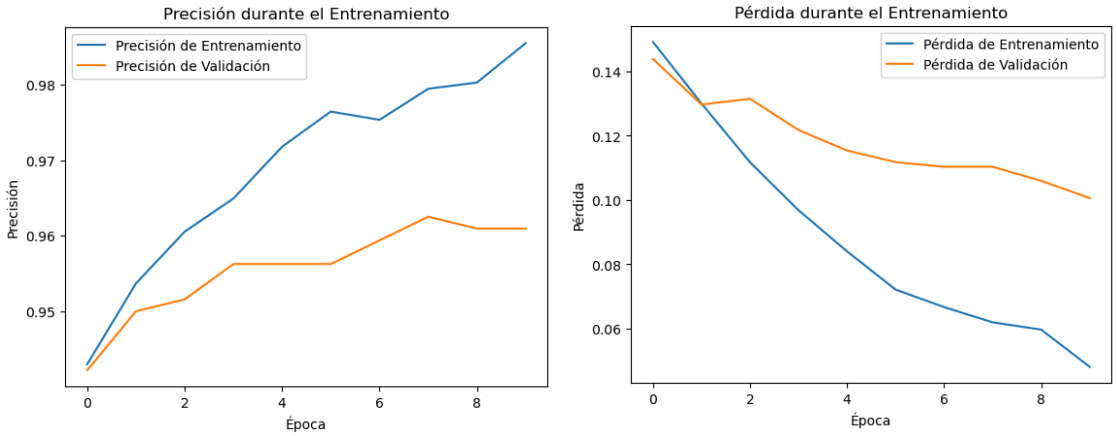
4. Evaluación y comparación del Modelo.

El desempeño del modelo se evaluó a partir de las métricas de entrenamiento, validación y prueba.

a) **Curvas de entrenamiento**: En la primera fase, el modelo alcanzó rápidamente una precisión de validación cercana al 93%, con curvas de pérdida que muestran una clara convergencia y sin señales de sobreajuste severo.

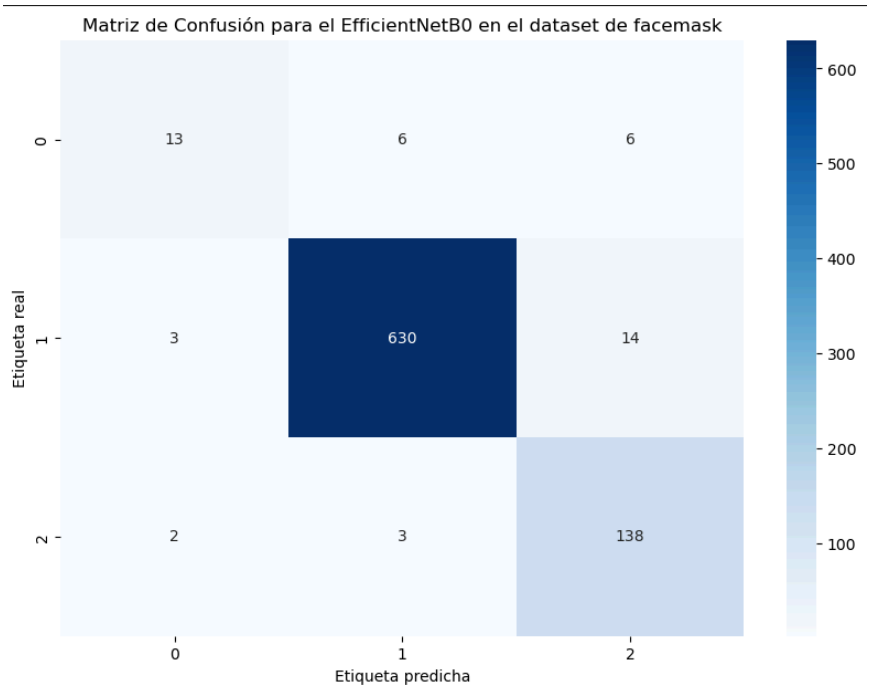


Tras aplicar el fine-tuning (segunda fase), la precisión mejoró hasta valores cercanos al 96% en validación, mientras que la pérdida continuó descendiendo de manera estable. Esto evidencia que liberar parcialmente las capas convolucionales permitió al modelo especializarse mejor en el dominio de mascarillas, manteniendo un buen equilibrio entre entrenamiento y validación.



b) **Matriz de confusión**: la clase with_mask es la mejor reconocida, con 630 aciertos frente a pocos errores. En cambio, la clase mask_weared_incorrect presenta mayor dificultad, con varios ejemplos confundidos tanto como con with_mask y without_mask. Esto indica

que los casos de uso incorrecto de mascarilla comparten características visuales con las demás clases, generando confusión en el modelo.



c) **Métricas de evaluación: comparación del desempeño con los modelos previos**

En conjunto, el transfer learning demuestra un mayor accuracy y la mayor robustez, combinando alta precisión y mejor recall, logrando de esta forma una mejor cobertura en la clase minoritaria, logrando el rendimiento más balanceado y consistente.

Además la explicabilidad seleccionada reafirma la confianza en este modelo, ya que además de clasificar mejor, podemos observar los aspectos de los rostros en los cuales el modelo se basa para clasificar, de este modo reforzamos la idea que sus predicciones no son al azar y que no se distrae en otros patrones de la imagen que no son influyentes para determinar las diferentes clases.

Reporte de Clasificación para EfficientNetB0 con Transfer Learning de facemask:

	precision	recall	f1-score	support
0	0.72	0.52	0.60	25
1	0.99	0.97	0.98	647
2	0.87	0.97	0.92	143
accuracy			0.96	815
macro avg	0.86	0.82	0.83	815
weighted avg	0.96	0.96	0.96	815

Resumen de rendimiento con los modelos previos:

Métrica \ Modelo	MLP	CNN	TL
Accuracy	0,93	0,94	0,96
Precisión (macro avg)	0,75	0,88	0,86
Recall (macro avg)	0,67	0,70	0,82
F1 (macro avg)	0,69	0,72	0,83

Análisis del rendimiento de los modelos (MLP, CNN, TL) para clasificar las diferentes clases.

El modelo logra un F1-score macro de 0.83, lo que indica un buen equilibrio global entre precisión y recall. La clase with_mask es la mejor representada, con F1=0.98, mientras que without_mask mantiene un desempeño sólido con F1=0.92 En contraste, mask_weared_incorrect muestra una un F1=0.60, reflejando la dificultad del modelo en esa categoría. En conjunto, el sistema alcanza una accuracy del 96%, pero la clase minoritaria, a pesar de desempeñarse mejor que en los otros modelos, sigue limitando el rendimiento balanceado. Comparando el modelo de transfer learning (TL) con el MLP y la CNN que realizamos antes, se observan diferencias claras en el desempeño por clase y en el F1-score global.

Resumen del rendimiento de los modelos mediante F1-Score por clase:

Clases \ Modelo	MLP	CNN	TL
mask_weared_incorrect (0)	0,24	0,32	0,60
with_mask (1)	0,96	0,97	0,98
without_mask (2)	0,87	0,88	0,92

En términos de rendimiento balanceado, el modelo de TL logra un F1 macro de 0.83, superando tanto al MLP (0.69) como a la CNN (0.72). Esto indica que maneja mejor el equilibrio entre clases, reduciendo la brecha entre la mayoritaria y las minoritarias. La clase mayoritaria (with_mask) alcanza un F1 de 0.98, ligeramente superior al obtenido con CNN (0.97) y al MLP (0.96). Para without_mask, el TL logra 0.92, desempeño superior a los modelos anteriores CNN (0.88) y MLP (0.87). La mayor diferencia aparece en mask_weared_incorrect: el TL consigue un F1=0.60, más alto que el CNN (0.32) y el MLP (0.24), lo que refleja una capacidad mucho mejor para capturar la clase minoritaria.