Detección de Cáncer de Piel con Redes Neuronales Convolucionales: Comparación de Arquitecturas Usando el Dataset HAM10000

Dante David Pérez Pérez - A01709226
Tecnológico de Monterrey

ABSTRACT

Diagnosticar el cáncer de piel en etapas tempranas es crítico para la supervivencia del paciente . Este estudio implementa y compara dos arquitecturas de clasificación de lesiones cutáneas. Se utilizó un dataset personalizado de The International Skin Imaging Collaboration con 6 tipos de lesiones: melanoma, carcinoma basocelular, carcinoma escamocelular, queratosis actínica, queratosis seborreica y nevus. El modelo MobileNetV2 utiliza transfer learinig y el modelo 2 utiliza una arquitectura de 4 bloques entrenada desde cero. MobileNetV2 demostró superior rendimiento general, mientras que el modelo 2 tiene mejor precisión bias-varianza.

KEYWORDS

Cáncer de piel, CNN (Redes Neuronales Convolucionales), Dermatología, Diagnóstico médico,, MobileNetV2, Transfer learning

I. NOMENCLATURA

Nomenclatura	Término	Definición
AKC	Queratosis Actínica	*Actinic Keratosis*
BCC	Carcinoma Basocelular	*Basal Cell Carcinoma*
MEL	Melanoma	Melanoma
NEV	Nevus	Nevus
SCC	Carcinoma Escamocelular	*Squamous Cell Carcinoma*
SEK	Queratosis Seborreica	*Seborrheic Keratosis*

Tabla 1

II. INTRODUCCIÓN

El cáncer en la piel es una enfermedad que nos puede afectar a todos, el melanoma es una de las variantes más letales y su diagnóstico temprano en los pacientes en crítico para salvar la vida de alguien. Pacientes detectados en etapa temprana tienen un 99% de supervivencia a 5 años, mientras que solamente un 30% de probabilidades en etapas avanzadas [3]. Los problemas más grandes a los que se enfrentan es la falta de análisis de la célula e interpretación de un experto.

Las CNN han evolucionado bastante a tal punto que el análisis de imágenes médicas, muestran un desempeño comparable a dermatólogos expertos, esto no indica que es mejor realizar la prueba con la CNN, pero para un diagnóstico rápido es buena idea utilizarla. Este estudio aborda la implementación de dos arquitecturas CNN, y su análisis comparativo, el dataset que se utilizó fue sacado de The International Skin Imaging Collaboration [4], por lo que fue personalizado para 6 tipos de enfermedades en la piel.

III. TRABAJO RELACIONADO

Como punto de partida para desarrollar mi modelo tome varios artículos científicos para no empezar desde cero, estos dos artículos fueron los que más me ayudaron a desarrollar el modelo final y en los que más me base.

1. Mannava et al. [1]:

Mannava implementa dos arquitecturas, una personalizada con 3 bloques con capas de Conv2D y MaxPooling y una segunda arquitectura que es más eficiente con MobilNetV2 y la pruebo con 4 diferentes optimizadores como lo son Adam y SDG.

2. Rezaoana et al. [2]:

En este artículo usan más de 25,000 imágenes y se aumentan el doble. Igualmente, se implementan bloques convolucionales para mejorar la captura de características en paralelo. Pero una de las limitaciones es que se necesitan una gran cantidad de datos para su entrenamiento óptimo.

Los puntos en común entre ambos artículos son que en ambos se usa *Data Augmentation* para mejorar el entrenamiento y ambos hacen énfasis en hacer un dataset con gran volumen para mejores resultados.

IV. METODOLOGÍA

A. Dataset y Preprocesamiento

a) Dataset y preprocesamiento

El dataset utilizado fue personalizado, se buscaron imágenes de The International Skin Imaging Collaboration,

concretamente de 6 tipos de enfermedades en la piel que son Carcinoma Basocelular (BCC), Carcinoma Escamoso (SCC), Melanoma (MEL), Queratosis Actínica (ACK), Queratosis Seborreica (SEK) y Nevus (NEV).

El procesamiento que se realizó primero fue reorganizar las imágenes por tipo de enfermedad para tenerlas clasificadas correctamente, después la separación del dataset en 70% entrenamiento, 10% validación y 20% prueba.

Clase	Entrenamiento	Validación	Prueba
AKC	1814	259	519
BCC	1993	284	571
MEL	1779	254	509
NEV	1931	276	553
SCC	1201	171	345
SEK	1868	267	535

Tabla 2

Igualmente, se implementó *Data Augmentation* y escalamiento para la etapa de entrenamiento. Esta fue la configuración que se implementó:

Técnica	Descripción	Valor usado
Rescale	Normaliza los valores de los píxeles dividiendo entre 255	1./255
Horizontal Flip	Voltea horizontalmente las imágenes aleatoriamente	True
Rotation Range	Rota las imágenes aleatoriamente dentro de un rango	±15°
Width Shift Range	Desplaza horizontalmente una fracción del ancho de la imagen	10% (0.1)
Height Shift Range	Desplaza verticalmente una fracción de la altura de la imagen	10% (0.1)
Shear	Aplica transformaciones de corte	0.1
Range	(shearing)	(10%)
Zoom Range	Aplica zoom aleatorio dentro de un rango	±10% (0.1)
Fill Mode	Método usado para rellenar los píxeles generados tras una transformación	nearest

Tabla 3

B. Arquitecturas de Modelos

■ Modelo 1:MobileNetV2 Adaptada

Basado en el trabajo de Mannava se implementó una arquitectura MobileNetV2 con transferencia de aprendizaje adaptado al dataset. El uso de MobilNetV2 radica en su eficiencia en equipos de poco cómputo, además de separar en dos diferentes convoluciones, convolución en profundidad que procesa cada canal de entrada independiente para poder capturar mejor y convolución puntual que procesa mediante filtros 1x1 para combinar la información de los canales. Hacer esto reduce significativamente el procesamiento manteniendo las características, MobilNetV2 expande las características en muchas más canales y aplica convolución en profundidad en este espacio donde hay más información, después comprime de nuevo las dimensiones para preservar toda la información importante.

Así es como funciona MobilNetV2 primero usa Depthwise Convolution canal por canal, es decir en RGB.

Filtro separado para cada canal

Filtro_R: Filtro_G: Filtro_B:

$$\begin{bmatrix} 0.1 & 0.2 & 0.1 \\ 0.2 & 0.4 & 0.2 \\ 0.1 & 0.2 & 0.1 \end{bmatrix} \rightarrow \begin{bmatrix} 0.2 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.1 \\ 0.2 & 0.1 & 0.2 \end{bmatrix} \rightarrow \begin{bmatrix} 0.1 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 \\ 0.1 & 0.3 & 0.1 \end{bmatrix}$$

$\text{Canal\_R} = \text{procesado: } 18 \quad \text{Canal\_G} = \text{procesado: } 14 \quad \text{Canal\_B} = \text{procesado: } 10$

Figura 1

De esta forma detecta mejor cada canal y puede analizar mejor en el paso 2, que es PointWise Convolution, esto nos da como resultado el patrón especifico de la enfermedad para poder diferenciarla después.

Combina información entre canales

$\text{Combinación\_1×1:} \quad [18,\ 14,\ 10] \times [0.3,\ 0.4,\ 0.3] = 18 \times 0.3 + 14 \times 0.4 + 10 \times 0.3 = 5.4 + 5.6 + 3.0 = 14.4$

Resultado: 14.4 Detecta patrón específico de pigmentación melanoma

Figura 2

Se usó **Fine Tuning y Transfer Learning** donde se congelaron las primeras 131 capas de MobilNetV2 para preservar características como bordes y texturas y las 23 capas restantes se quedaron en espera para ser entrenadas de acuerdo a las enfermedades. El modelo preentrenado de MobilNetV2 con los pesos de ImageNet.

ImageNet es un conjunto de datos de imágenes clasificadas en miles de categorías

GlobalAveragePooling2D reduce una matriz a un solo número por capa, por lo que no importa en que posición esté, se reporta con el peso.

$$\text{melanoma\_detector} = \begin{bmatrix} 0.1 & 0.2 & 0.1 & 0.1 & 0.1 & 0.2 & 0.1 \\ 0.2 & \mathbf{0.9} & 0.2 & 0.1 & 0.1 & \mathbf{0.8} & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.1 & 0.1 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

$$\text{GAP} = \frac{0.9 + 0.8 + 47 \times 0.2}{49} = \frac{1.7 + 9.4}{49} = \frac{11.1}{49} \approx 0.2265$$

Figura 3

Después utilizamos Dense para tener un cabezal con 1024 neuronas para que aprenda patrones, estas mismas las borramos con DropOut y volvemos con otra capa de Dense ahora con 512 neuronas. La capa final es un, Dense con 6 neuronas para las 6 clases.

Como optimizador se utilizó SDG y este sirve para hacer ajustes a los pesos y minimizar el error, momentum es para acelerar el aprendizaje y hacer cambios más eficientes y rápidos.

Componente	**Tipo de Capa**	**Forma de Salida**	**Detalles**
Capa de Entrada	Input	(None, 224, 224, 3)	Imágenes dermatoscópicas RGB
Extractor de Características	MobileNetV2 Base	(None, 7, 7, 1280)	Preentrenado en ImageNet
Agrupamiento Global	GlobalAveragePooling2D	(None, 1280)	Reducción de dimensión espacial
Cabezal Clasificador	Dense + ReLU	(None, 1024)	Primera capa de clasificación
	Dropout	(None, 1024)	Regularización (p = 0.5)
	Dense + ReLU	(None, 512)	Segunda capa de clasificación
	Dropout	(None, 512)	Regularización (p = 0.3)
Capa de Salida	Dense + Softmax	(None, 6)	Predicción de 6 clases de lesiones cutáneas

Tabla 4

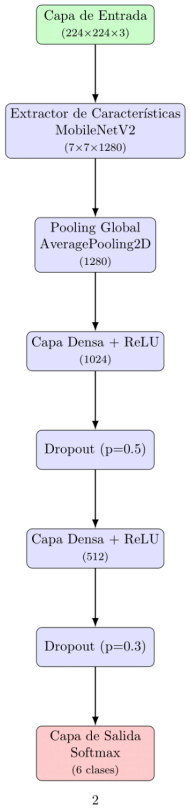

Figura 4

■ Modelo 2: Arquitecturas con capas densas

Esta implementación es basada al segundo artículo, pero se añadieron más capas para agarrar mejor las características. La primera capa es para agarrar características como bordes,

contornos de la lesión, texturas muy superficiales. La segunda capa es para los patrones de pigmentación como puntos o líneas y texturas un poco más complejas. La tercera capa es para buscar patrones reticulares y puntos dermatológicos. La última capa es para características más generales como la distribución de los colores y características globales. El patrón de usar dos Conv2D, BatchNorm MaxPooling2D y Dropout es para capturar las características complejas de cada fase, además de ser de uso clínico, los patrones deben de ser más específicos para determinar si es una enfermedad u otra.

Lo que hace Conv2D es aplicar kernels de 3x3 px por toda la imagen.

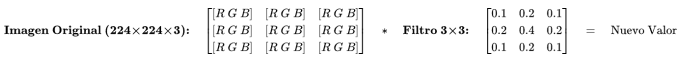

Figura 5

Después BatchNormlalization normaliza los datos en el ejemplo de abajo, primero son datos de iluminación de la imagen, algo que puede confundir al modelo porque varían mucho entre ambos ejemplos, por lo que se pasa la matriz a valores de -1 a 1 para tener un patrón más normalizado.

$$\text{Imagen1}[50,\ 60,\ 70] \rightarrow [-1,\ 0,\ 1]$$
$$\text{Imagen2}[150,\ 160,\ 170] \rightarrow [-1,\ 0,\ 1]$$

Figura 6

MaxPooling2D toma regiones 2x2 de la imagen y conserva solamente el valor máximo, en el ejemplo de una matriz 4x4 toma el valor máximo 2x2. Esto es importante porque una lesión puede estar en cualquier parte, por lo que usar esto no importa donde este, sino que toma el valor más alto para mantener las características esenciales

$$\begin{bmatrix} 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 \\ 8 & 2 & 0 & 0 \\ 1 & 0 & 12 & 3 \end{bmatrix} \xrightarrow{\text{MaxPooling } 2\times 2} \begin{bmatrix} 0 & 5 \\ 8 & 12 \end{bmatrix}$$

Figura 7

Y el DropOut elimina aleatoriamente N% de las neuronas, ya que sin DropOut el modelo memorizaría, por lo que así hacemos que busque patrones

$$\text{Antes del Dropout:} \quad [5,\ 8,\ 3,\ 12,\ 6,\ 9,\ 1]$$
$$\text{Durante Entrenamiento:} \quad [5,\ 0,\ 3,\ 0,\ 6,\ 9,\ 0]$$

Figura 8

Componente	Tipo de Capa	Forma de Salida	Detalles
Capa de Entrada	Input	(None, 224, 224, 3)	Imágenes dermatoscópicas RGB normalizadas
Bloque Convolucional 1	2×Conv2D + BatchNorm	(None, 224, 224, 32)	32 filtros 3×3, activación ReLU, padding same
	MaxPooling2D	(None, 112, 112, 32)	Pooling 2×2, reducción espacial
	Dropout	(None, 112, 112, 32)	Regularización (p = 0.25)
Bloque Convolucional 2	2×Conv2D + BatchNorm	(None, 112, 112, 64)	64 filtros 3×3, activación ReLU, padding same
	MaxPooling2D	(None, 56, 56, 64)	Pooling 2×2, reducción espacial
	Dropout	(None, 56, 56, 64)	Regularización (p = 0.25)
Bloque Convolucional 3	2×Conv2D + BatchNorm	(None, 56, 56, 128)	128 filtros 3×3, activación ReLU, padding same
	MaxPooling2D	(None, 28, 28, 128)	Pooling 2×2, reducción espacial
	Dropout	(None, 28, 28, 128)	Regularización (p = 0.25)
Bloque Convolucional 4	2×Conv2D + BatchNorm	(None, 28, 28, 256)	256 filtros 3×3, activación ReLU, padding same

	MaxPooling2D	(None, 14, 14, 256)	Pooling 2×2, reducción espacial
	Dropout	(None, 14, 14, 256)	Regularización (p = 0.3)
Agrupamiento Global	GlobalAveragePooling2D	(None, 256)	Reducción de dimensión espacial completa
Cabezal Clasificador	Dense + ReLU + BatchNorm	(None, 512)	Primera capa de clasificación densa
	Dropout	(None, 512)	Regularización agresiva (p = 0.5)
	Dense + ReLU + BatchNorm	(None, 256)	Segunda capa de clasificación densa
	Dropout	(None, 256)	Regularización agresiva (p = 0.5)
Capa de Salida	Dense + Softmax	(None, 6)	Predicción probabilística de 6 clases de lesiones cutáneas

Tabla 5

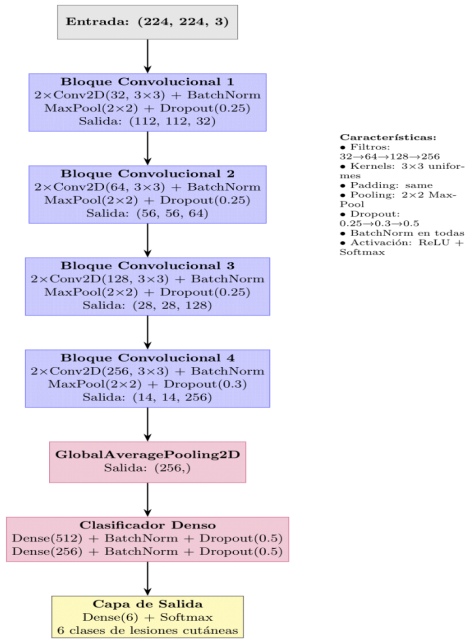

Figura 9

V. Entrenamiento y primeros resultados

A. MobilNetV2

La implementación de MobileNetV2 alcanzó una precisión final de 73.44% en el conjunto de prueba, con una mejor precisión de validación de 64.69% durante el entrenamiento. El modelo fue entrenado durante 71 épocas utilizando un total de 4,097,606 parámetros, de los cuales 81.5% fueron entrenables mediante la estrategia de fine-tuning selectivo implementado. Fueron 71 épocas, ya que se detuvo por early stopping, esto quiere decir que después de 10 épocas no hubo mejoras en el modelo, pero se configuró para que durara 100 épocas.

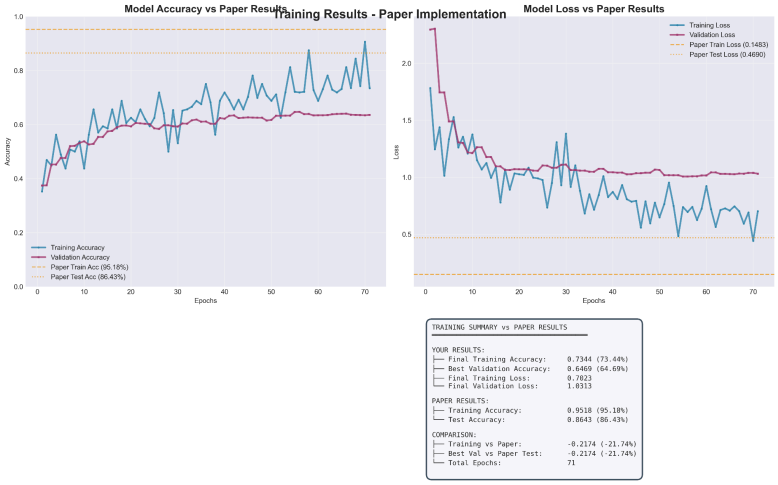

Figura 9

1) Convergencia del Modelo

En las gráficas se ve que se espera un accuracy de 95.18%, pero se logra alcanzar solamente 73.44%, en la Figura 1 se puede ver que iban estables desde la época 35, desde ahí no hubo tanto avance en validation

La diferencia entre precisión de entrenamiento (73.44%) y validación (64.69%) sugiere un ligero sobre ajuste, aunque el gap de 8.75% se mantiene dentro de rangos aceptables para aplicaciones médicas con datasets limitados. Por lo que todavía está dentro del rango para no ser considerado OverFitting.

Clase	Precisión	Recall	F1-Score	Interpretación Clínica
NEV (Nevus)	0.8539	0.8770	0.865	Excelente detección de lesiones benignas
MEL (Melanoma)	0.6362	0.6699	0.652	Rendimiento moderado para lesiones malignas
BCC (Carcinoma Basocelular)	0.5977	0.6480	0.621	Detección aceptable de cáncer no-melanoma
AKC (Queratosis Actínica)	0.6117	0.5857	0.598	Rendimiento moderado para lesiones precancerosas
SEK (Queratosis Seborreica)	0.5609	0.6280	0.592	Dificultad en diferenciación de lesiones benignas
SCC (Carcinoma Escamocelular)	0.5634	0.3478	0.430	Menor rendimiento, posiblemente por menor representación

Tabla 6

La matriz de confusión (Figura 2) revela patrones y se dividen en dos formas

- Fortalezas del Modelo:

NEV (Nevus): Es la clase con más precisión, lo que se le puede adjudicar a la gran cantidad de datos de entrenamiento que tuvo.
BCC: 370 verdaderos positivos de 571 casos, indicando buena detección de carcinomas basocelulares.
MEL: 341 verdaderos positivos de 509 casos, rendimiento aceptable para melanomas.

- Debilidades Identificadas:

SCC: Solo 120 verdaderos positivos de 345 casos (34.8% recall), esta tuvo mayor dificultad de detección debido a que fue la clase con menor datos, siendo la más débil.
Confusión BCC-AKC: 74 casos de AKC clasificados como BCC.
Confusión MEL-SEK: 67 casos de SEK clasificados como MEL, representando falsos positivos críticos para melanoma

Esta es una tabla comparativa para ver las diferencias entre mi implementación y la del paper. Hay una gran diferencia en precisión y eficacia de los datos. Pero en el paper no se menciona con cuántas épocas fue entramado su modelo para tener una mejor idea de los resultados.

Métrica	Implementación Actual	Paper Original	Diferencia
Precisión de Prueba	64.51%	86.43%	-21.92 %
Precisión de Entrenamiento	73.44%	95.18%	-21.74 %
Épocas de Entrenamiento	71	No especificado	-
Parámetros Totales	4.1M	~3.5M	+0.6 M

Tabla 7

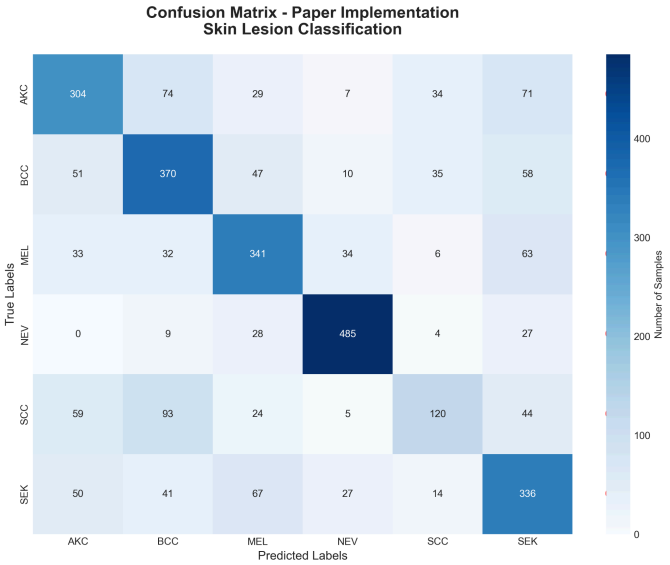

Figura 10

B. Arquitectura adaptada

La implementación de la CNN mejorada con la arquitectura de 4 bloques convolucionales alcanzó una precisión final de 58.11% en el conjunto de validación, con una precisión de entrenamiento de 59.38% al finalizar las 100 épocas.
El modelo muestra una diferencia mínima entre entrenamiento y validación (1.27%), sugiriendo un mejor balance bias-varianza comparado con MobileNetV2, aunque con precisión general inferior.

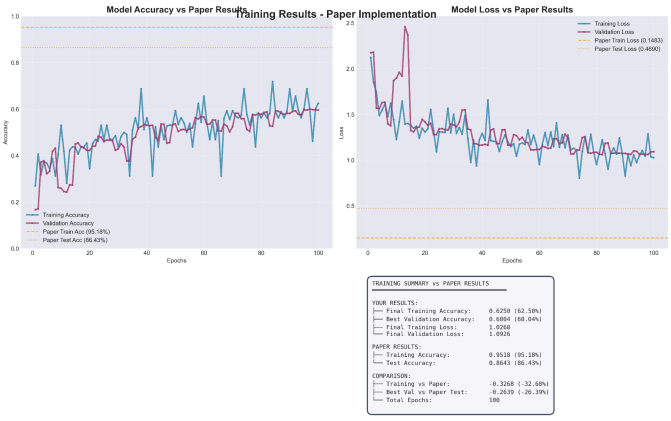

Figura 11

Analizando la tabla de confusión, NEV sigue siendo una clase muy fuerte debido a su gran volumen de datos, pero en este modelo hay clases que se confunden entre ellas, las cuales son:
SEK-AKC (138 casos): Mayor confusión del modelo, clasificando queratosis seborreicas como queratosis actínicas
AKC-BCC (112 casos): Dificultad persistente en distinguir lesiones queratinizadas
BCC-AKC (97 casos): Confusión entre carcinomas basocelulares y queratosis actínicas
MEL-SEK (63 casos): Falsos negativos críticos - melanomas clasificados como lesiones benignas

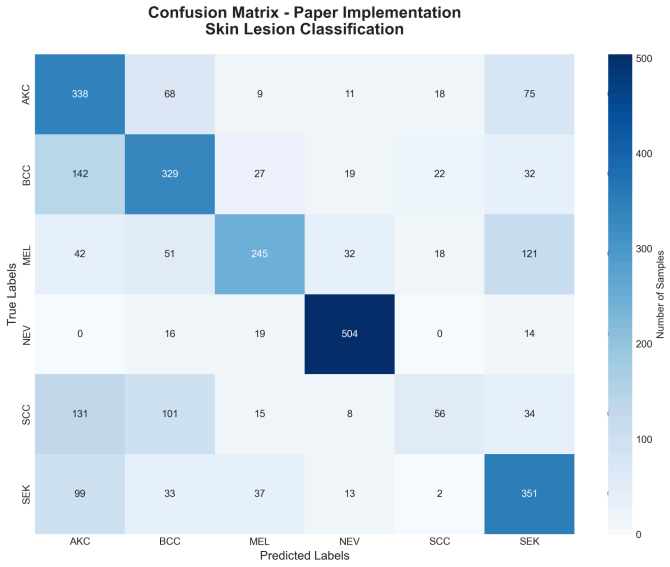

Figura 12

A continuación se muestra una tabla comparativa para los dos modelos, el modelo MobileNetV2 y Modelo con arquitectura Convolutiva. En general, el segundo modelo muestra picos pocos variables, mientras que MobilNetV2 varía mucho y hace que aparezcan picos abruptos.

Métrica	MobileNetV2	CNN Mejorada	Diferencia	Interpretación
Precisión General	64.51%	58.11%	-6.40%	MobileNetV2 superior
Training-Validation Gap	8.75%	1.27%	-7.48%	CNN menos sobreajuste

Épocas Entrenamiento	71	100	+29	CNN requiere más tiempo
Parámetros Totales	4.1M	1.4M	-2.7M	CNN más eficiente en memoria
Estrategia	Transfer Learning	Entrenamiento completo	--	--

Tabla 8

VI. Conclusiones

Después de haber comparado ambos y de haber investigado y practicado ambas formas de arquitecturas con diferentes enfoques. La clase NEV fue una clase muy fuerte por su gran representación en el dataset, pero las demás clases había confusiones, esto debido a las limitaciones del dataset que a comparación de los artículos contemplados usaron 4 veces más el dataset e igualmente las similitudes entre lesiones como principal fuente de error. El uso de transfer learning fue de gran uso para tener una precisión más alta, mientras que el segundo modelo demostró una mejor métrica en cuanto bias-varianza. Algunas mejoras que se pueden implementar es balancear las clases correctamente con pesos indicados para cada una. Aunque los resultados tengan una gran brecha entre los papers originales, se puede indicar que es viable poder aplicar estas tecnologías en situaciones en la vida real como consultorios o citas médicas.

Referencias

[1] M. C. Mannava, B. Tadjgadapa, D. Anil, A. S. Dev, and A. T, "CNN Comparative Analysis for Skin Cancer Classification," in *Proc. 13th Int. Conf. Commun., Netw. Comput. Eng. (ICCCNT)*, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984442.

[2] N. Rezaoana, M. S. Hossain, and K. Andersson, "Detection and Classification of Skin Cancer by Using a Parallel CNN Model," in *Proc. IEEE WIE Conf. Elect. Comput. Eng. (WIECON-ECE)*, 2020, pp. 380-387, doi: 10.1109/WIECON-ECE52138.2020.9398001.

[3] American Cancer Society, "Cancer Facts & Figures 2023," Atlanta: American Cancer Society, 2023. [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2023.html

[4] ISIC 2017 Challenge