

Implementación de un modelo de Machine Learning para la detección y clasificación de cáncer de piel.

Universidad Autónoma de Chihuahua.

Facultad de Ingeniería.

2025.

Asignatura: Data Science.

Profesor: Manuel Alberto Chávez Salcido.

Alumno: Rodrigo García Núñez.

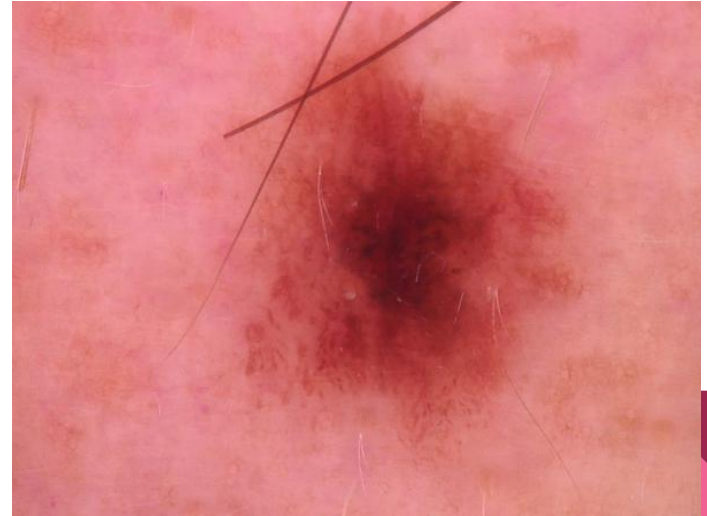
Objetivo

- Desarrollar un sistema de clasificación automática de lesiones de cáncer de piel utilizando técnicas de Machine Learning, capaz de distinguir entre lesiones benignas y malignas.
- Recopilar y preprocesar un conjunto de datos conformados por imágenes de lesiones de piel de alta calidad y datos clínicos relevantes.
- Implementar y entrenar modelos de aprendizaje profundo, como redes neuronales convolucionales (**CNN**), para la clasificación de lesiones de piel.
- Evaluar el rendimiento de los modelos utilizando métricas como precisión, sensibilidad y especificidad.



Descripción del problema

El cáncer de piel es una de las formas de cáncer más comunes a nivel mundial, y su detección temprana es crucial para aumentar las tasas de supervivencia. El diagnóstico tradicional de cáncer de piel se basa en la evaluación visual de lesiones cutáneas por parte de dermatólogos, seguida de biopsias en casos sospechosos.



Comprensión de Datos

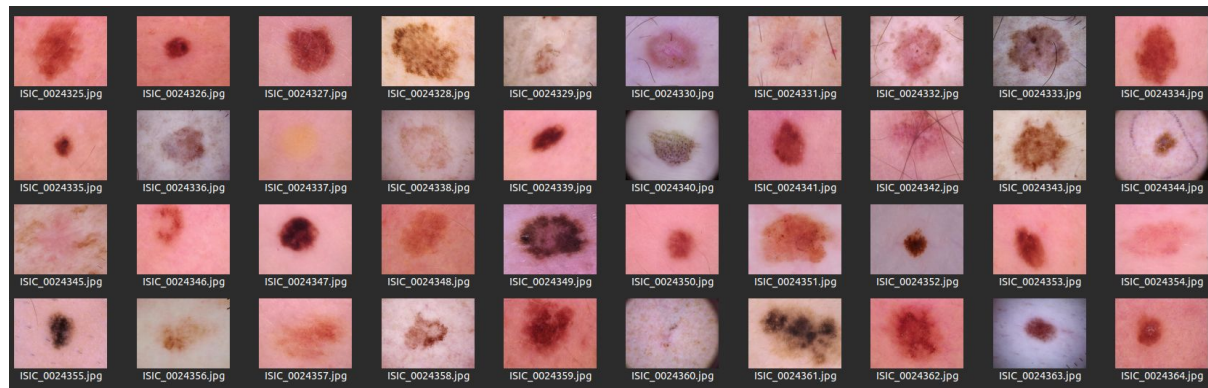
Para el desarrollo de este proyecto, se emplearán los conjuntos de datos HAM10000 de ISIC Archive, el cual representa una vasta colección de datos necesarios para la implementación del modelo de Machine Learning destinado a la clasificación de diferentes diagnósticos de cáncer de piel.



Comprensión de Datos

Se tiene una tabla compuesta de 17 atributos y 11, 720 registros; cada registro asociado mediante id a una imagen.

El DataSet es proporcionado por The International Skin Imaging Collaboration (ISIC), academia destinada al uso de imágenes digitales relacionadas al Cáncer de Piel, con el fin de apoyar el desarrollo de tecnologías dedicadas a reducir la mortalidad del Cáncer de Piel.



Procesamiento de Datos

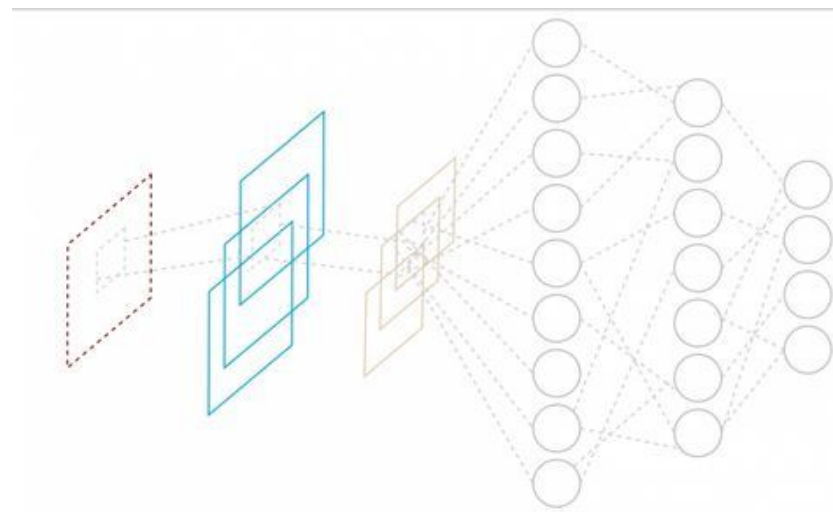
Con el fin de que el modelo procese datos con la mejor calidad posible, se implementaron estrategias de limpieza de datos, balanceo por undersampling ante clases mayoritarias, normalización, estandarización y codificación de datos.

Con estas estrategias, se logró obtener conjuntos de entrenamiento y validación vastos y balanceados, y suficientes registros de prueba. Con esto, nos aseguramos de que los datos que consume el modelo tengan la mayor calidad posible.



Implementación de Modelo.

Como se mencionó anteriormente, el modelo debe ser capaz de trabajar con imágenes y metadatos. Para cumplir con este requisito, se implementa una red neuronal convolucional que recibe a las imágenes, cuya salida será la entrada de una red neuronal densa. A esta entrada también se le integran los metadatos, de modo que el resultado final será resultado tanto de características de las imágenes, como de metadatos.

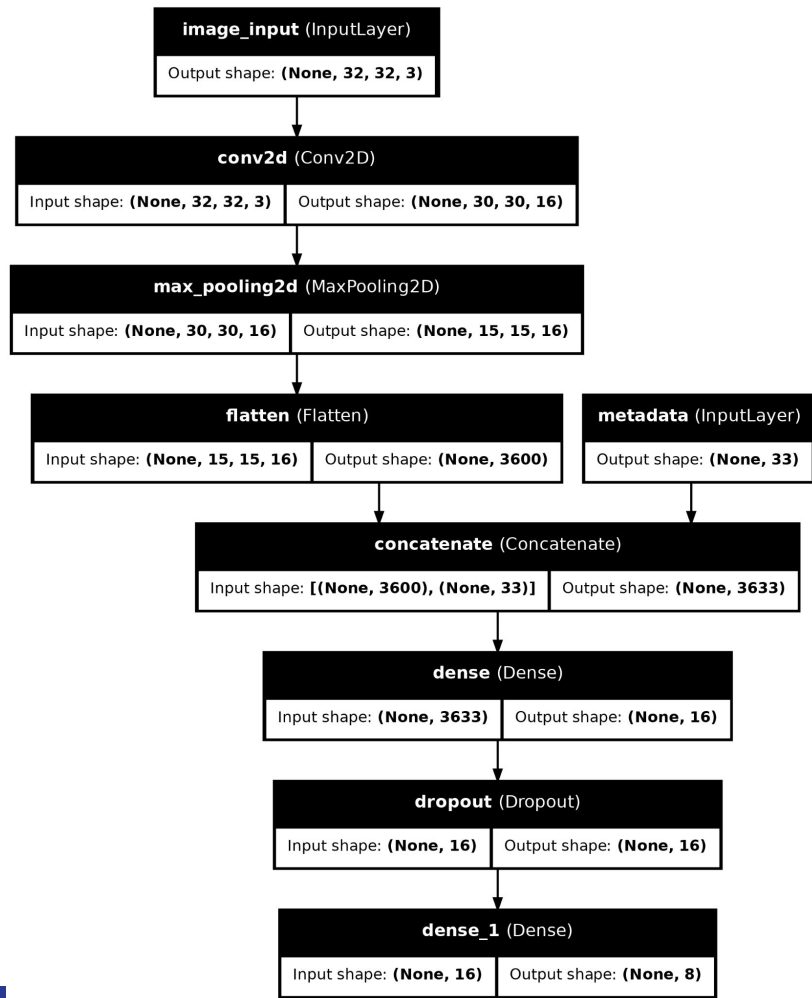


Implementación de Modelo.

Para implementar la red neuronal convolucional que procesa metadatos e imágenes, se utilizó la api funcional de tensorflow-keras, la cual permite implementar modelos con características especiales, como es el caso de nuestro modelo.



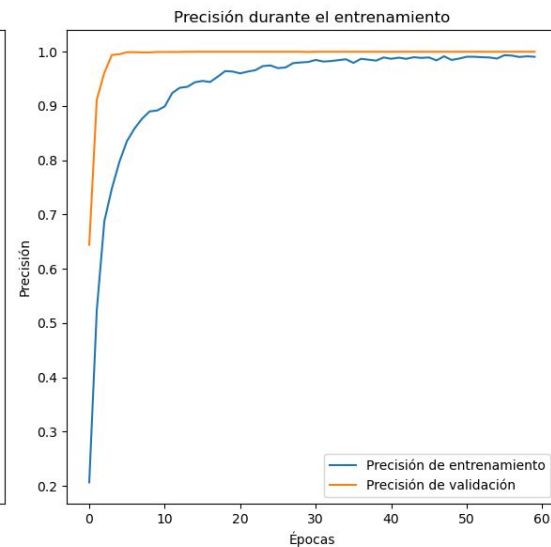
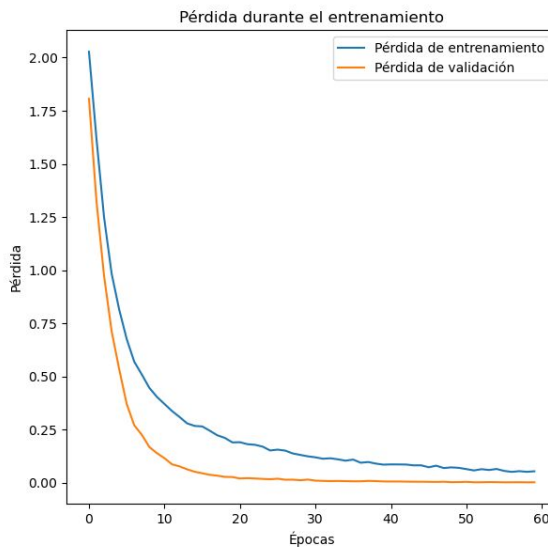
Estructura del Modelo



Desempeño del Modelo - Entrenamiento

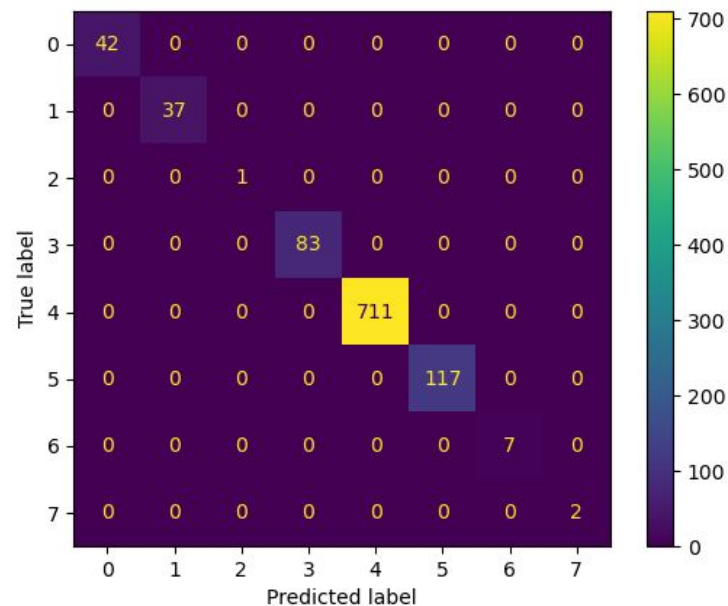
Al terminar el entrenamiento, el modelo obtuvo las siguientes métricas:

- categorical_accuracy: 0.9857
- loss: 0.0545
- val_categorical_accuracy: 1.0000
- val_loss: 0.0024



Desempeño del Modelo - Matriz de confusión con datos de prueba.

De la matriz de confusión, podemos notar que, aunque el dataset de validación no fue balanceado, el modelo clasificó bien todos los diagnósticos de las clases presentes en el dataset de prueba.



Métricas del Desempeño

- Tiempo total de predicción de 1000 registros: 0.25922679901123047
- Tiempo por predicción única: 0.0002592267990112305
- Precisión: 1.000
- Sensibilidad: 1.000
- Especificidad por clase:
 - Especificidad clase 'actinic keratosis' (0): 1.0000
 - Especificidad clase 'basal cell carcinoma' (1): 1.0000
 - Especificidad clase 'dermatofibroma' (2): 1.0000
 - Especificidad clase 'melanoma' (3): 1.0000
 - Especificidad clase 'nevus' (4): 1.0000
 - Especificidad clase 'pigmented benign keratosis' (5): 1.0000
 - Especificidad clase 'squamous cell carcinoma' (6): 1.0000
 - Especificidad clase 'vascular lesion' (7): 1.0000



Implementación en Google Cloud Platform

Este modelo fue importado a una aplicación alojada en un contenedor de google cloud run.

Dicha aplicación fue realizada utilizando el framework “flask”.



Skin Cancer Classifier

Selecciona tu edad aproximada:

5

Localización de la lesión:

Extremidad inferior

Biopsia de Concomitante:

Positivo

Primer diagnostico:

Benigno

Segundo diagnostico:

Proliferaciones benignas de melanocitos

Tercer diagnostico:

Nevo

Tipo de Confirmación de diagnóstico:

Histopatología

Lesión con melanocitos:

Positivo

Sexo:

Masculino

Link de la aplicación:

- <https://skincancerapp-935771581787.northamerica-south1.run.app>

Link al repositorio GitHub:

- https://github.com/RodrigoGarciaNunez/Skin_cancer_classifier/

