

Diagnóstico automático de afecciones dermatológicas con Transfer Learning

Juan Francisco Terán

December 2024

1 Introducción

El problema de detectar y clasificar afecciones dermatológicas a partir de imágenes es una tarea crucial en el ámbito de la salud, ya que puede ayudar en el diagnóstico temprano y tratamiento de diversas enfermedades de la piel.

En este proyecto, abordaré este problema aplicando una arquitectura de Convolutional Neural Network (CNN) utilizando Transfer Learning con el modelo pre-entrenado Xception. Usaré el conjunto de datos "Human Against Machine with 10000 training images (HAM10000)", que contiene 10,015 imágenes de lesiones cutáneas previamente etiquetadas en siete categorías: queratosis actínica y carcinoma intraepitelial, carcinoma basocelular, lesiones benignas similares a queratosis, dermatofibroma, melanoma, nevos melanocíticos y lesiones vasculares.

Mi objetivo es construir un modelo que pueda clasificar correctamente las imágenes de lesiones cutáneas en una de estas siete categorías.

Mostraré cómo mi modelo logra una puntuación de Accuracy del 80% en los datos de prueba, lo cual es comparable a algunos trabajos anteriores de otros autores. También discutiré acerca de las limitaciones y desafíos de mi trabajo y sugeriré posibles direcciones para trabajos futuros relacionados con la detección y clasificación de afecciones dermatológicas usando Deep Learning.

2 Trabajo Relacionado

Muchos autores en el mundo han escrito artículos científicos de sus trabajos sobre el conjunto de datos "HAM10000", implementando variedad de modelos pre-entrenados.

- Jain [1] compara el rendimiento de varios modelos pre-entrenados: VGG19, ResNet50, InceptionResnetV2, InceptionV3, Xception y MobileNet. Su clasificador Xception obtuvo el mejor rendimiento con un Train Accuracy del 90% y otro 90% de Test Accuracy. Para ello, Jain realizó un Fine Tuning completo, reentrenando todas las capas de cada modelo anterior.

- Experimentando con el modelo pre-entrenado DINOv2, Yu [2] obtiene un modelo con un Test Accuracy del 83% sobre el Test Data, según él, un Accuracy mayor en comparación al que proporcionan los clasificadores ResNet y ResNet+Rein. Además, muestra una ventaja de Accuracy en los modelos que utilizan el metodo CUFIT para implementar Fine Tuning en presencia de imagenes con etiquetas ruidosas.

3 Definición del Problema

3.1 Tarea

La tarea de predicción consiste en determinar la clase de una afección dermatológica a partir de imágenes de lesiones en la piel. Utilizando técnicas de Transfer Learning con el modelo pre-entrenado Xception, la idea es clasificar las imágenes en una de las siguientes categorías:

- akiec: Queratosis actínica y carcinoma intraepitelial / enfermedad de Bowen
- bcc: Carcinoma basocelular
- bkl: Lesiones benignas similares a queratosis
- df: Dermatofibroma
- mel: Melanoma
- nv: Nevos melanocíticos
- vasc: Lesiones vasculares

3.2 Algoritmo

Construí un modelo de clasificación supervisado utilizando Transfer Learning con el modelo pre-entrenado Xception para diagnosticar automáticamente afecciones dermatológicas a partir de imágenes de lesiones en la piel.

Transfer Learning es una técnica de Machine Learning donde un modelo pre-entrenado en una gran cantidad de datos se reutiliza y ajusta para una tarea específica. En este caso, se utiliza el modelo Xception, que ha sido pre-entrenado en el conjunto de datos ImageNet, y se ajusta para clasificar imágenes de lesiones dermatológicas.

4 Evaluación Experimental

4.1 Datos

El conjunto de datos utilizado es el "Human Against Machine with 10000 training images (HAM10000)", que contiene imágenes dermatológicas

etiquetadas para la detección y clasificación de afecciones dermatológicas. Este Dataset se compone de 10,015 imágenes previamente etiquetadas en siete categorías diferentes ("akiec", "bcc", "bkl", "mel", "nv" y "vasc"). Para preparar los datos para el reentrenamiento del modelo, realicé los siguientes pasos:

1. Carga y Organización de Imágenes: Organicé las imagenes en directorios según su categoría utilizando los metadatos proporcionados en los archivos CSV HAM10000_metadata.csv y ISIC2018_Task3_Test_GroundTruth.csv.

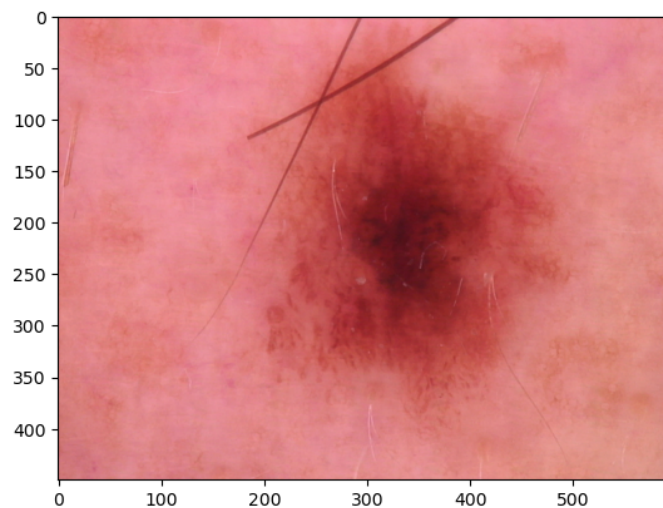


Figure 1: Imagen de muestra sin preprocesar

2. Creación de Directorios: Creé directorios separados para el conjunto de entrenamiento y el conjunto de prueba, y dentro de estos, subdirectorios para cada una de las categorías.
3. Copia de Imágenes: Copié las imagenes a los directorios correspondientes según su categoría.
4. Data Augmentation: Apliqué técnicas de aumento de datos para aumentar la variabilidad de las imágenes de entrenamiento. Las técnicas incluyeron rotación, desplazamiento, zoom, y volteo horizontal y vertical.
5. Preprocesamiento de Imágenes: Preprocesé las imagenes para normalizar los valores de los píxeles.

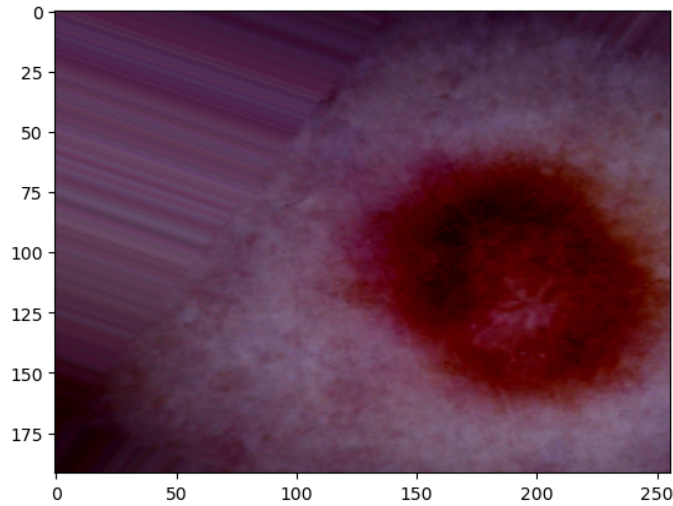


Figure 2: Imagen de muestra preprocesada para entrada de Xception

Conjuntos de datos:

- Entrenamiento: Compuesto por 10,015 imágenes distribuidas en las siete categorías mencionadas. Este conjunto lo utilicé para reentrenar el modelo de Deep Learning.
- Prueba: Compuesto por 1,512 imágenes etiquetadas, utilizadas para evaluar el rendimiento del modelo reentrenado.
- "HAM10000_metadata.csv": Metadata del Dataset de entrenamiento. Se compone de 10,015 observaciones y 8 atributos, incluyendo el identificador de cada imagen, su clase y variedad de características de los pacientes que registraron cada lesión (tipo de diagnóstico, edad, sexo, ubicación de la lesión).
- "ISIC2018_Task3_Test_GroundTruth.csv": Metadata del Dataset de prueba. Se compone de 1,512 registros y 8 variables, tal y como la metadata del conjunto de entrenamiento.

4.2 Metodología

Utilicé la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), que consta de seis pasos: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. A continuación, describo los pasos realizados:

- Paso 1 y 2 (comprensión del negocio y de los datos): El objetivo del proyecto es diagnosticar automáticamente afecciones dermatológicas uti-

lizando Transfer Learning. Para ello, usé el Dataset HAM10000, que contiene 10,015 imágenes etiquetadas de diferentes afecciones dermatológicas. Realicé un análisis exploratorio de los datos para entender la distribución y características de las variables.

- Descripción del dataset: Definí los tipos de atributos y realicé estadísticas descriptivas.
- Visualización de datos: Utilicé gráficos de barras para los atributos.

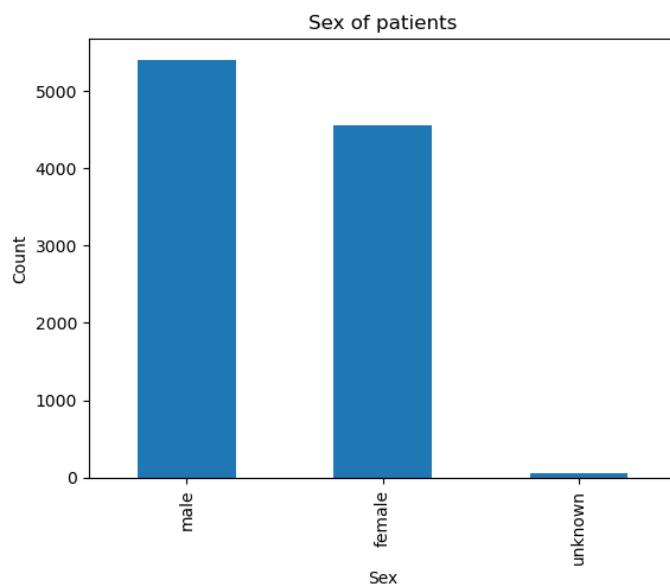


Figure 3: Grafico de barras para "sex"

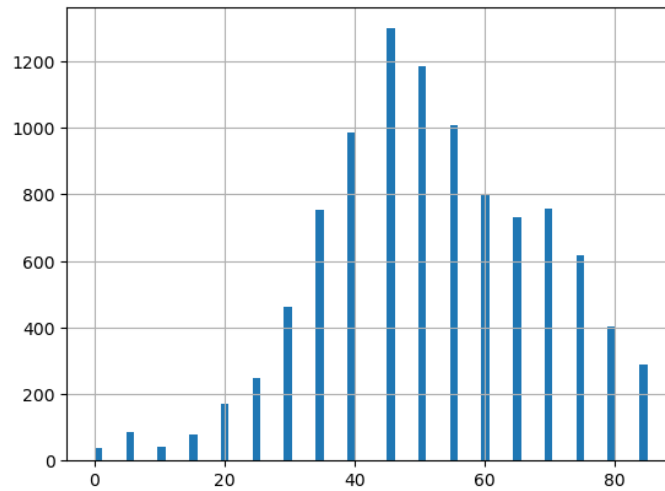


Figure 4: Histograma para "age"

- Paso 3 (preparación de los datos): Creé directorios para las imágenes de entrenamiento y prueba, y distribuí las imágenes en sus respectivas categorías.
- Paso 4 (modelado): Usé Transfer Learning con el modelo pre-entrenado Xception para clasificar las imágenes:
 - Data Augmentation: Apliqué técnicas de aumento de datos para mejorar la generalización del modelo.

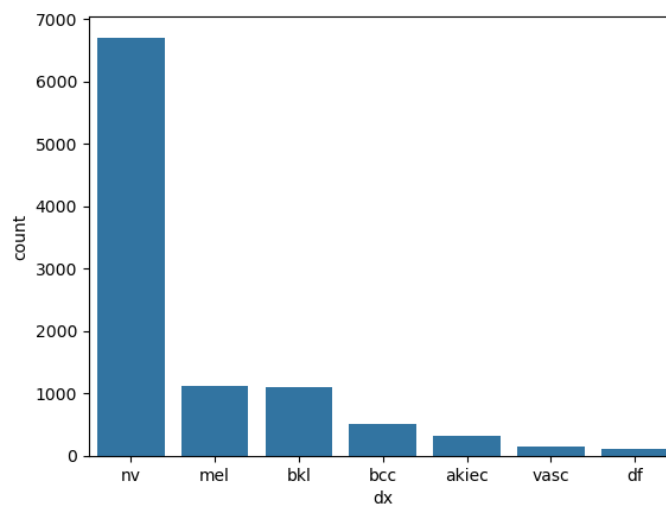


Figure 5: Distribución de la clase "dx"

- División del dataset: Dividí el Dataset en conjuntos de entrenamiento, validación y prueba.
- Definición del modelo: Cargué el modelo Xception pre-entrenado y añadí capas adicionales para la clasificación.
- Entrenamiento del modelo: Entrené el modelo utilizando el conjunto de datos de entrenamiento y validación.
- Paso 5 (evaluación): Evalué el rendimiento del modelo utilizando métricas y la matriz de confusión:
 - Matriz de confusión: Construí la matriz de confusión para evaluar el rendimiento del modelo en el conjunto de prueba.
 - Reporte de clasificación: Obtuve el reporte de clasificación para obtener métricas detalladas de Accuracy, Precision, Recall y F1-Score.
- Paso 6 (Despliegue): Llevar el análisis a producción.
 - El análisis y el modelo se implementaron en un entorno de Jupyter Notebook para facilitar el entendimiento del trabajo computacional realizado.
 - El modelo reentrenado se desplegó en un Space de Hugging Face utilizando Streamlit para facilitar su acceso y uso en producción con una aplicación simple. Ver en Hugging Face.

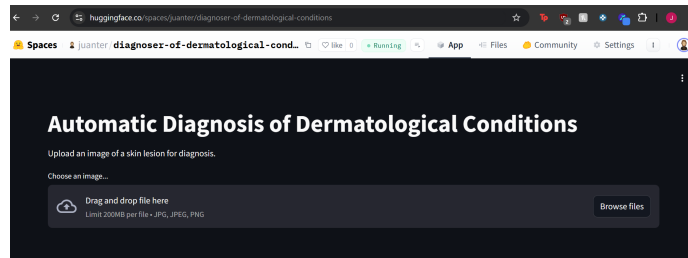


Figure 6: Aplicación desplegada en Hugging Face

4.3 Resultados

Presento los resultados que obtuve:

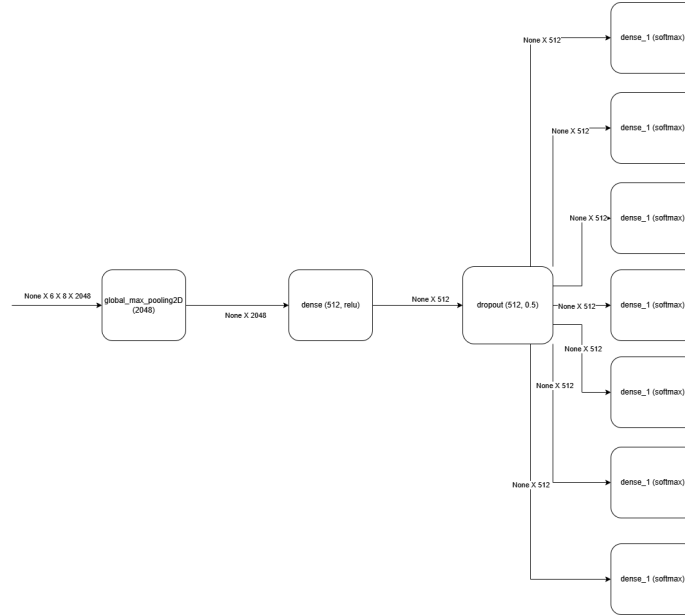


Figure 7: Arquitectura definida para las capas finales de la red

Nombre	epochs	Fine Tune	validation_steps
Xception	30	116 (desde "block13_sepconv1_act")	validation_data.samples//32

Table 1: Ajustes de Hiperparámetros para Xception

1. Accuracy del Modelo: El Accuracy del modelo durante el entrenamiento y la validación se muestra en el siguiente gráfico:

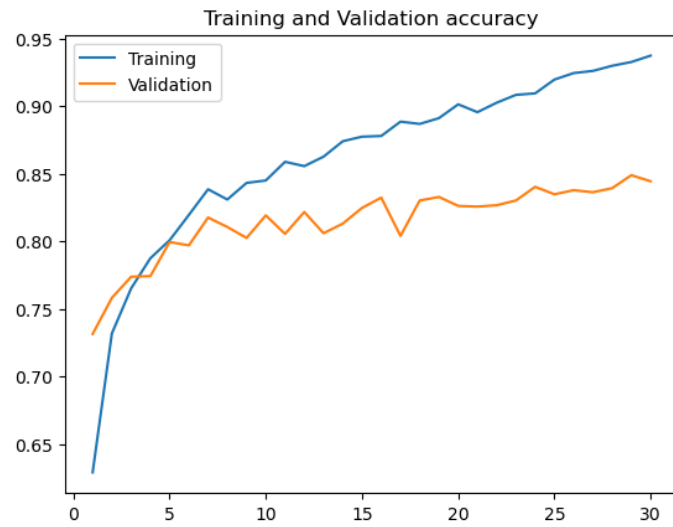


Figure 8: Accuracy para entrenamiento y validación

Nombre	Train Accuracy	Validation Accuracy
Xception	94%	85%

Table 2: Accuracy para entrenamiento y validación

2. Loss del Modelo: El Loss del modelo durante el entrenamiento y la validación se muestra en el siguiente gráfico:

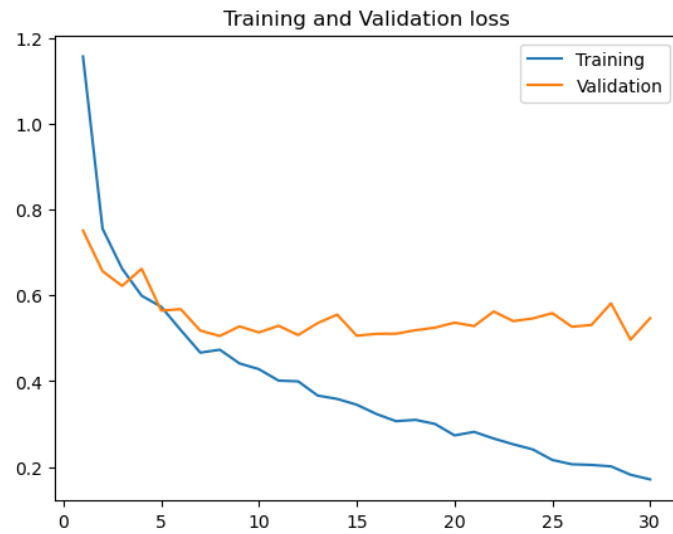


Figure 9: Loss para entrenamiento y validación

- Matriz de Confusion: La matriz de confusión para las predicciones del modelo en el conjunto de datos de prueba se muestra a continuación:

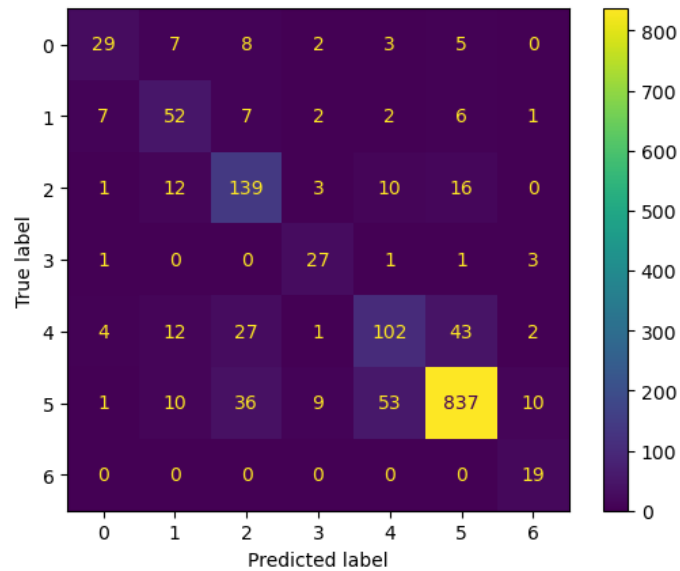


Figure 10: Matriz de Confusión para Xception

- Reporte de Clasificación: El reporte de clasificación que incluye Accuracy,

Precisión, Recall y F1-Score para cada clase se muestra a continuación:

	precision	recall	f1-score	support
0	0.54	0.67	0.60	43
1	0.68	0.56	0.61	93
2	0.77	0.64	0.70	217
3	0.82	0.61	0.70	44
4	0.53	0.60	0.56	171
5	0.88	0.92	0.90	908
6	1.00	0.54	0.70	35
accuracy			0.80	1511
macro avg	0.74	0.65	0.68	1511
weighted avg	0.80	0.80	0.80	1511

Figure 11: Reporte de Clasificación

Nombre	Test Accuracy
Xception	80%

Table 3: Accuracy para Test

5. Accuracy por Clase: El Accuracy por clase se muestra en la siguiente tabla:

Clase	Accuracy
akiec	67%
bcc	56%
bkl	64%
df	61%
mel	60%
nv	92%
vasc	54%

Table 4: Accuracy por Clase

Análisis de Resultados:

- Desbalance de Clases: Las etiquetas de clase del Dataset están desbalanceadas, con una notable diferencia entre la cantidad de registros pertenecientes a "nv" y las demás clases.
- Rendimiento del Modelo: El modelo Xception mostró un buen rendimiento en términos de Accuracy y Loss durante el entrenamiento y la validación.

Sin embargo, el desbalance de clases puede haber afectado el rendimiento en algunas clases específicas para el conjunto de prueba.

- Mejoras Potenciales: Experimentar con diferentes modelos de clasificación y parámetros podría mejorar los resultados.

4.4 Discusión

El modelo de Transfer Learning usando Xception alcanzó un rendimiento de 80% en la clasificación de afecciones dermatológicas, lo cual es un resultado prometedor. No obstante, no logró superar el rendimiento mencionado por Jain [1], del 90%, ni el rendimiento del autor Yu [2], quien alcanzó un 83%.

Quiero resaltar dos aspectos relevantes:

1. Preprocesamiento de Datos: El uso de técnicas de Data Augmentation (aumento de datos) y la normalización de las imágenes fueron cruciales para mejorar el rendimiento del modelo. Estas técnicas ayudaron a aumentar la diversidad del conjunto de datos de entrenamiento, lo cual es esencial para evitar el Overfitting y mejorar la generalización del modelo. El modelo funciona de manera efectiva para predecir correctamente la clase "nv", no obstante no tiene un gran desempeño al clasificar las otras etiquetas.
2. Hiperparámetros del Modelo: La combinación de hiperparámetros utilizada en el modelo Xception fue determinada experimentalmente. Se realizaron múltiples pruebas variando hiperparámetros como el número de épocas, la definición de las capas finales de la red, las capas para hacer Fine Tuning, entre otros. La mejor combinación encontrada fue la que permitió alcanzar un

$$\text{Accuracy} = 80\%.$$

El modelo Xception demostró ser una técnica útil para la clasificación de imágenes dermatológicas, gracias a su capacidad para extraer características relevantes de las imágenes. Sin embargo, es posible que explorando otros modelos de Transfer Learning, ajustando más los hiperparámetros, o combinando diferentes técnicas de preprocesamiento, se puedan obtener mejores resultados.

5 Conclusiones

Resumo las conclusiones más importantes:

1. Rendimiento del Modelo: El modelo entrenado logró un Validation Accuracy del 85% y un Test Accuracy del 80%. Estos resultados son prometedores y demuestran la efectividad del uso de Transfer Learning en la clasificación de imágenes dermatológicas.

2. Desbalance de Clases: Se observó un desbalance significativo en las clases del Dataset, con una predominancia de la clase "nv". A pesar de esto, el modelo logró manejar razonablemente bien este desbalance, aunque futuras mejoras podrían enfocarse en técnicas de balanceo de clases.
3. Importancia del Preprocesamiento: El preprocesamiento de las imágenes y la generación de datos aumentados fueron cruciales para mejorar el rendimiento del modelo. El uso de Data Augmentation ayudó a mejorar la generalización del modelo pero no fue suficiente por el desbalance de clases.

Para un trabajo futuro, aplicar técnicas más avanzadas de preprocesamiento de imágenes podría mejorar la calidad de los datos y, por ende, los resultados del análisis. Además, la implementación de un Ensemble Model para combinar el reconocimiento por imagen de las lesiones con las características personales de los pacientes, es un área que vale la pena explorar a fin de obtener el desempeño más exacto posible.

Como propuesta, aumentar la variedad de datos mediante la recopilación de más imágenes de diferentes fuentes podría mejorar la robustez del modelo, dado que una de las principales deficiencias encontradas fue la limitación en la variedad de datos.

6 Bibliografía

- [1] Jain, S., Singhanian, U., Tripathy, B., Nasr, E. A., Aboudaif, M. K., & Kamrani, A. K. (2021). Deep learning-based transfer learning for classification of skin cancer. *Sensors*, 21(23), 8142.
- [2] Yu, Y., Ko, M., Shin, S., Kim, K., & Lee, K. (2024). Curriculum Fine-tuning of Vision Foundation Model for Medical Image Classification Under Label Noise. *arXiv preprint arXiv:2412.00150*.

7 Anexos

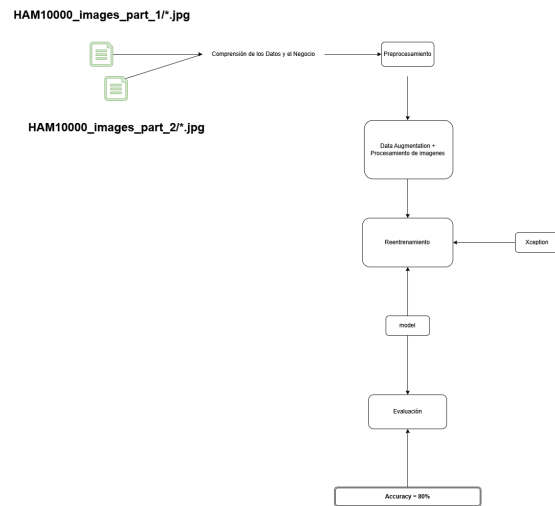


Figure 12: Flujo de trabajo general del proyecto