

Proyecto SP1 IA - Dermatología

Byron Haldair García Guzmán - 2200336

Jose Ivan Barreno Bulux - 22003949

Luis Fernando Guzman Orozco - 21000542

Introducción

Las enfermedades dermatológicas representan un reto diagnóstico debido a la amplia variedad de síntomas y su similitud con otras enfermedades cutáneas. Para resolver esta incertidumbre se realizó un modelo de Inteligencia Artificial (IA) capaz de predecir enfermedades dermatológicas a partir de un conjunto de datos clínicos, empleando técnicas de aprendizaje automático.

Descripción del Dataset

- 366 registros de pacientes descritos.
- 34 características clínicas y de laboratorio.
- Una clase que representa una de las seis posibles enfermedades dermatológicas.

Las enfermedades pueden ser:

- Psoriasis
- Seborrheic dermatitis
- Lichen planus
- Pityriasis rosea
- Chronic dermatitis
- Pityriasis rubra pilaris

Algunos aspectos importantes Dataset:

- La mayoría de las variables son ordinales (0-3).
- Hay una variable con valores nulos (Age) que se debe de completar.
- Todas las variables son numéricas.

Metodología

1. Carga y pre-procesamiento de datos

- Lectura del Dataset.
- Reemplazo de valores faltantes NaN en la columna Age por una media.
- Separar variables en independientes y dependientes.
- Balanceo de las clases con Smote.
- Normalización de los datos con StandarScaler.
- One-Hot Encoding a la variable Clase.

2. División del conjunto de datos

- Conjunto de entrenamiento: 80% de los datos.
- Conjunto de prueba: 20% de los datos.
- Validación interna: se reservó el 10% del conjunto de entrenamiento para validación durante el entrenamiento.

3. Construcción del modelo de IA

Red neuronal desarrollada con TensorFlow/Keras.

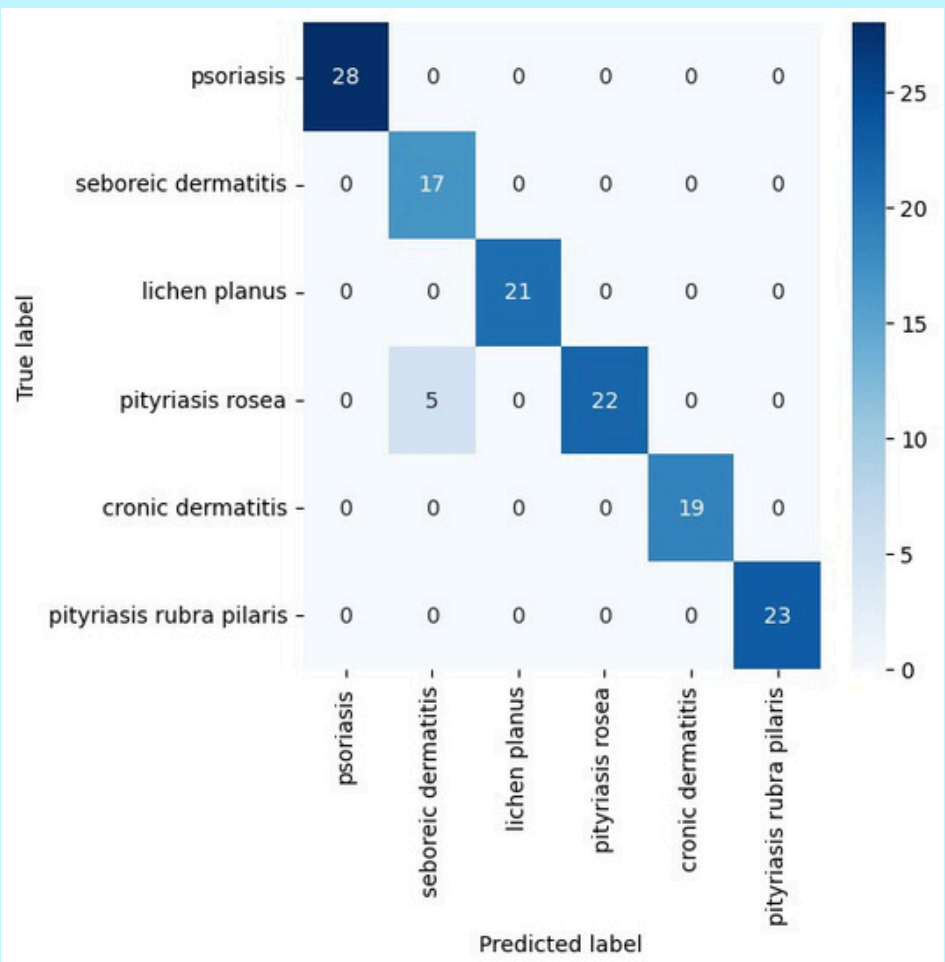
- Input Layer: 34 neuronas.
- Hide Layers:
 - 2 capas con activación ReLU, regularización L2 y normalización por lotes.
 - Capas de Dropout para evitar el Overfitting.
- Output Layer: capa con activación Softmax y 6 neuronas (6 clases).

4. Entrenamiento del modelo

- Función de pérdida: CategoricalCrossentropy.
- Optimizador: Adam.
- Métrica de evaluación: Accuracy.
- Técnica de regularización: EarlyStopping para detener el entrenamiento cuando el modelo deja de mejorar, ayudando a evitar el Overfitting.

RESULTADOS

	Accuracy (%)	Loss (%)
Modelo 1	97.0370	0.1852
Modelo 2	44.4444	6.4049
Modelo 3	16.6667	7.0377



CONCLUSIONES

- El modelo puede apoyar a dermatólogos en diagnósticos de enfermedades crónicas.
- El pre-procesar los datos es un paso crítico para el entrenamiento correcto de un modelo de IA y que este proporcione resultados exitosos.
- Las técnicas de regularización como Dropout, Batch Normalization, L2 y EarlyStopping mejora el rendimiento del modelo evitando el Overfitting.

MEJORAS A FUTURO

- Optimización de hiperparámetros
- Aumentar la cantidad de capas densas o utilizar otras arquitecturas más avanzadas como redes convolucionales o árboles.
- Implementar k-fold cross-validation (validación cruzada estratificada) para evaluar el rendimiento de manera más robusta y evitar sobreajuste a una sola partición.