

IA EN DERMATOLOGÍA

UTILIZANDO IA PARA EL DIAGNÓSTICO
DE ENFERMEDADES DERMATOLÓGICAS



ESCANEAME



NUESTRO EQUIPO



Kevin Hernandez
17001095

Edy Cocón
13001762

Javier Ferrer
13000522



Introducción

Utilizaremos inteligencia artificial para el diagnóstico de enfermedades dermatológicas. Para esto utilizamos un conjunto de 32 síntomas buscando realizar el diagnóstico de 6 enfermedades eritematoescamosas.

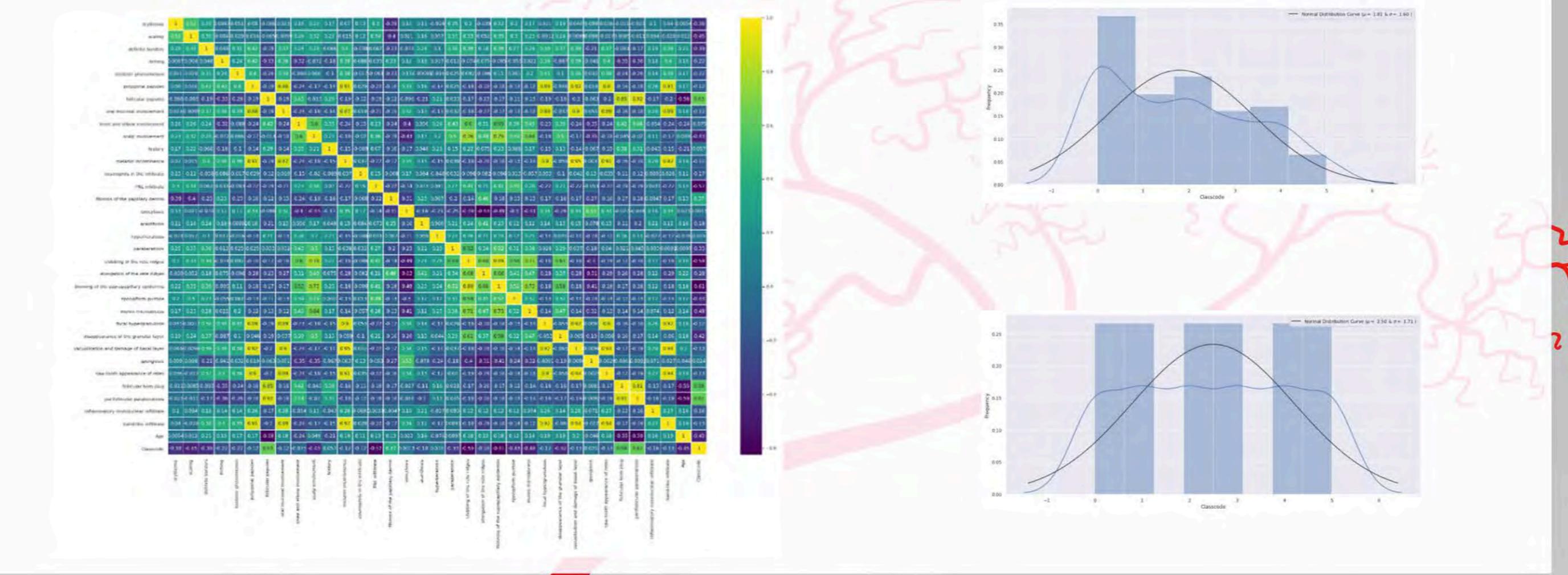
Existen más de 3000 enfermedades dermatológicas registradas, y el método de diagnóstico actual por parte de un médico es hacer un estudio clínico de los síntomas descritos por el paciente, realizar una observación de la zona afectada y realizar una comparación entre su experiencia y algún manual dermatológico, con esto compara los datos obtenidos con las fotografías y descripciones de las enfermedades descritas en el manual. Esto que es la obtención inicial se contrasta o confirma con un estudio histopatológico del tejido (Biopsia).

Es acá donde nace el interés de aplicar esta tecnología para realizar el diagnóstico. Se ha demostrado en varios casos que la inteligencia artificial es capaz de superar la eficiencia de los médicos al realizar este tipo de tareas, la comparación abstracción y predicción dentro de conjuntos extensos de datos.

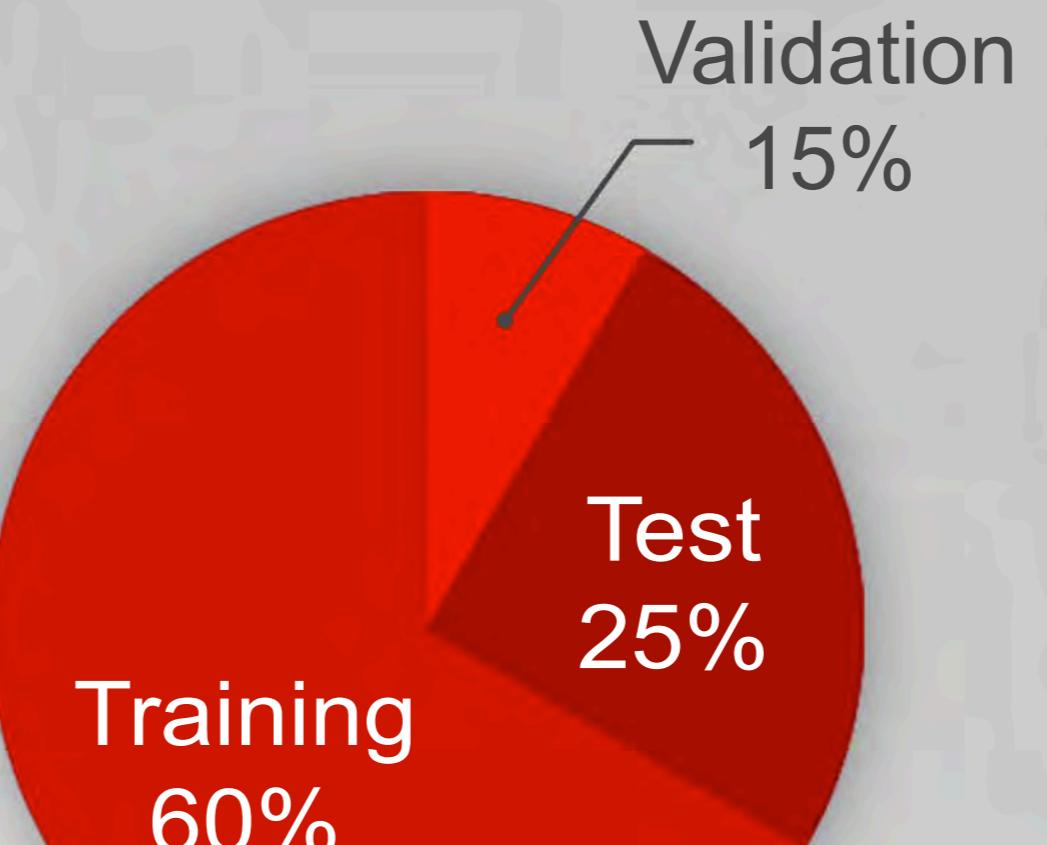
El Conjunto de Datos

Se cuentan con 366 casos de los cuales en la mayoría se describen 34 síntomas, 33 de manera numérica y 1 nominal. Cada síntoma se describe dentro de un rango de 0 a 3. Los síntomas se dividen en clínicos(12) e histopatológicos(22).

El primer paso, la exploración de los datos, se llevó a cabo para entender un poco más el problema y el contexto en el que vamos a trabajar. Esto nos llevó a realizar el la limpieza de los datos, el balanceo del conjunto de datos utilizando un algoritmo de OverSampling y un análisis de correlación.



6 Categorías
33 Features
366 Casos



La Metodología

El objetivo es encontrar la mejor arquitectura para solucionar el problema planteado, para ello elegimos utilizar una red neuronal densamente conectada. El lenguaje utilizado fue Python y todo el ecosistema de IA que este posee (Keras, SKLearn, Pandas, etc.).

Se crearon 4 aproximaciones y a cada una se le aplicó 4 arquitecturas diferentes, con lo cual trabajamos con 16 modelos

Data Set



Aproximaciones

- 1 - Todos los features
- 2 - Los features más relevantes
- 3 - Solo features clínicos
- 4 - Solo features histopatológicos

Arquitecturas

$$f(n) = 2^{(10-n)}$$

OVERSAMPLING

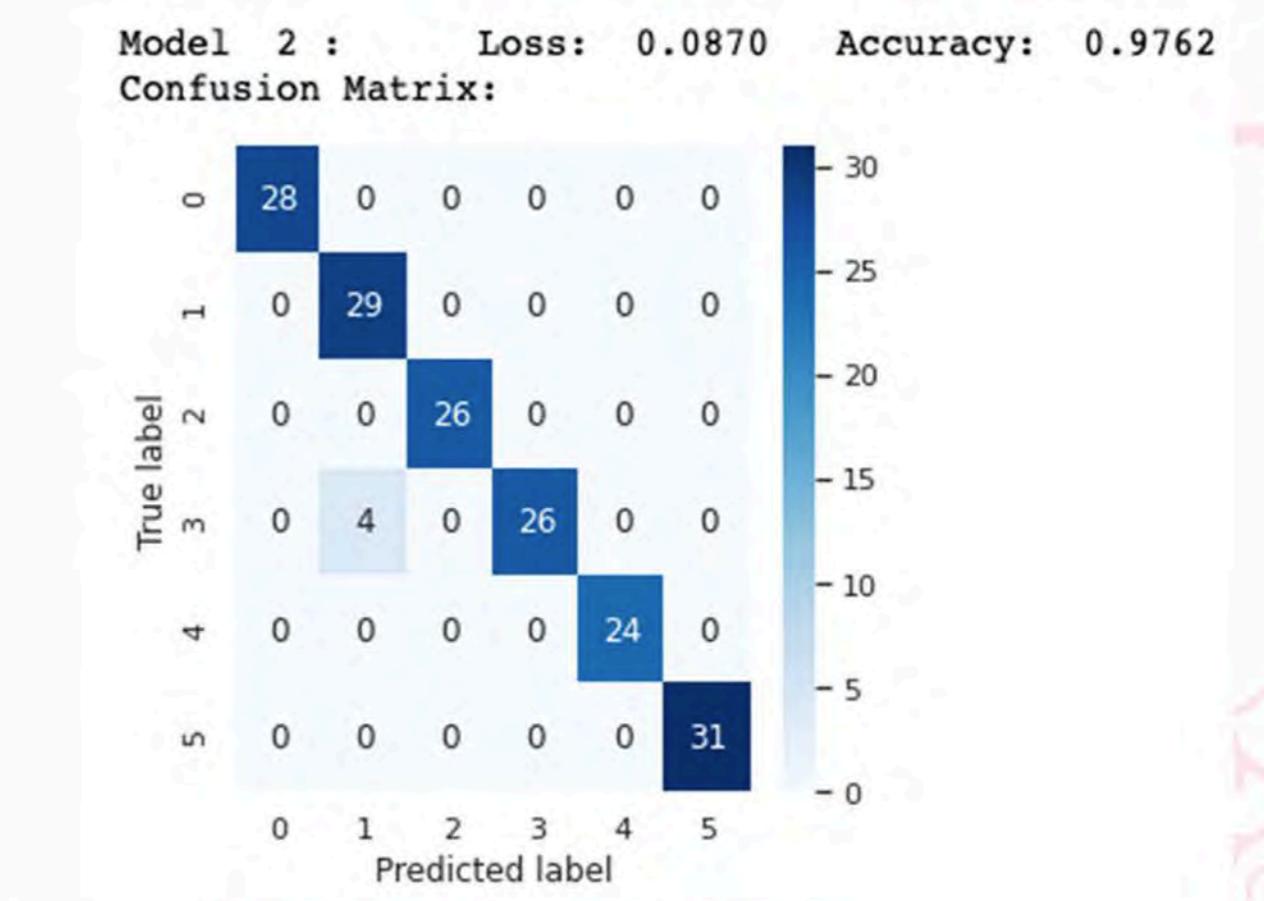
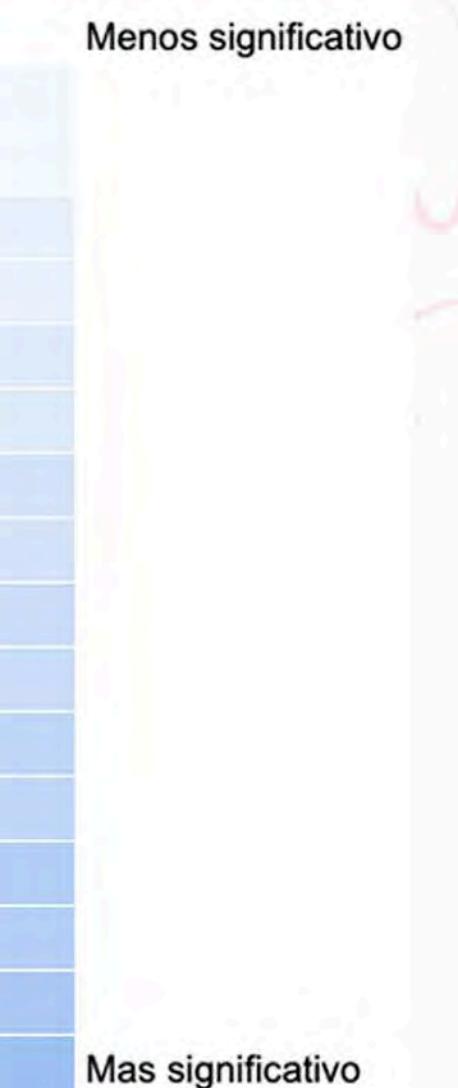
Es un método utilizado para balancear un conjunto de datos en el cual se genera nuevos ejemplos, ya sea duplicando o creando ejemplos sintéticos, para agregarlos a las clases cuyo número de ejemplos están por debajo de la cantidad de ejemplos que dispusimos

Los Resultados

Tras ejecutar los 12 modelos planteados nos topamos con los inconvenientes de tener un conjunto de datos pequeño.

- Se puede notar como mientras más capas agregamos al modelo las métricas se alejan del objetivo
- Los modelos más precisos son aquellos que toman en cuenta más features
- Se puede notar como mientras más capas agregamos al modelo menos las métricas se alejan del objetivo
- El número de Epochs es muy bajo para el proceso de entrenamiento debido a que el dataset de Training es muy pequeño. De esta forma, evitamos también que los modelos caigan en Overfitting

Aproximación	Hidden Layers	Loss	Accuracy
All Features	2	0.0884	0.9524
	3	0.087	0.9762
	4	0.155	0.9286
	5	0.1583	0.9702
	2	0.4604	0.8095
Clinical Features	3	0.5079	0.7381
	4	0.5824	0.75
	5	0.684	0.625
Histopathological Features	2	0.1681	0.9405
	3	0.1893	0.9167
	4	0.2451	0.8571
	5	0.2048	0.8988
Important Features	2	0.136	0.9286
	3	0.1318	0.9345
	4	0.1722	0.8988
	5	0.1581	0.9464



El mejor modelo

Aproximación:
Todos los features

Arquitectura:
3 Capas hidden

El número de Epochs es muy bajo para el proceso de entrenamiento debido a que el dataset de Training es muy pequeño. De esta forma, evitamos también que los modelos caigan en Overfitting.