



UNIVERSIDAD CATÓLICA SAN PABLO

Laboratorio: Clasificación binaria

Computer Science — Tópicos en Inteligencia
Artificial

Harold Alejandro Villanueva Borda

1. Introducción

En este informe se presentan los resultados de una serie de experimentos de clasificación de imágenes para detectar melanoma utilizando diferentes modelos de redes neuronales. Los modelos fueron entrenados, validados y probados utilizando un dataset de imágenes descargado desde Kaggle (<https://www.kaggle.com/drscarlat/melanoma>).

2. Descripción del Dataset

2.1. Adquisición de Datos

El dataset utilizado para estos experimentos fue descargado desde Kaggle, con un tamaño aproximado de 5.26 GB. Se dividió en tres conjuntos:

- Conjunto de entrenamiento: 10,682 imágenes
- Conjunto de validación: 3,562 imágenes
- Conjunto de prueba: 3,561 imágenes

2.2. Distribución de Clases

El dataset está compuesto por dos clases:

- Melanoma (Clase 0)
- No Melanoma (Clase 1)

3. Metodología

Para evaluar el rendimiento de los modelos, se implementaron y entrenaron tres arquitecturas distintas de redes neuronales. A continuación, se describen los detalles de cada modelo, incluyendo su arquitectura, tamaño del modelo, y resultados obtenidos.

4. Modelos Implementados

4.1. Modelo 1: Model1_Dense_ReLU

Este modelo consiste en una arquitectura de red completamente conectada utilizando la función de activación ReLU. La arquitectura del modelo se presenta en la Tabla 1.

Cuadro 1: Arquitectura de Model1_Dense_ReLU

Capa (Tipo)	Forma de Salida	Parámetros
Flatten	(None, 150528)	0
Dense (128)	(None, 128)	19,267,712
Batch Normalization	(None, 128)	512
Dropout	(None, 128)	0
Dense (64)	(None, 64)	8,256
Batch Normalization	(None, 64)	256
Dense (1)	(None, 1)	65
Total		19,276,801

4.1.1. Resultados de Model1_Dense_ReLU

El reporte de clasificación para este modelo es el siguiente:

Precisión: 0.85

Recall: 0.85

F1-score promedio: 0.84

Soporte: 3561

4.2. Modelo 2: Model2_CNN_LeakyReLU

Este modelo emplea una arquitectura de red convolucional con la función de activación Leaky-ReLU. La arquitectura de este modelo se presenta en la Tabla 2.

4.2.1. Resultados de Model2_CNN_LeakyReLU

El reporte de clasificación para este modelo es el siguiente:

Cuadro 2: Arquitectura de Model2_CNN_LeakyReLU

Capa (Tipo)	Forma de Salida	Parámetros
Conv2D (32 filtros)	(None, 222, 222, 32)	896
MaxPooling2D	(None, 111, 111, 32)	0
Conv2D (64 filtros)	(None, 109, 109, 64)	18,496
MaxPooling2D	(None, 54, 54, 64)	0
Conv2D (64 filtros)	(None, 52, 52, 64)	36,928
Flatten	(None, 173056)	0
Dense (64)	(None, 64)	11,075,648
Dropout	(None, 64)	0
Dense (1)	(None, 1)	65
Total		11,132,033

Precisión: 0.92

Recall: 0.92

F1-score promedio: 0.92

Soporte: 3561

4.3. Modelo 3: Model3_DeepCNN_ReLU

Este modelo consiste en una red convolucional profunda utilizando la función de activación ReLU y múltiples capas de normalización. La arquitectura de este modelo se presenta en la Tabla 3.

Cuadro 3: Arquitectura de Model3_DeepCNN_ReLU

Capa (Tipo)	Forma de Salida	Parámetros
Conv2D (32 filtros)	(None, 222, 222, 32)	896
Batch Normalization	(None, 222, 222, 32)	128
MaxPooling2D	(None, 111, 111, 32)	0
Conv2D (64 filtros)	(None, 109, 109, 64)	18,496
Batch Normalization	(None, 109, 109, 64)	256
MaxPooling2D	(None, 54, 54, 64)	0
Conv2D (128 filtros)	(None, 52, 52, 128)	73,856
Batch Normalization	(None, 52, 52, 128)	512
MaxPooling2D	(None, 26, 26, 128)	0
Flatten	(None, 86528)	0
Dense (128)	(None, 128)	11,075,712
Batch Normalization	(None, 128)	512
Dropout	(None, 128)	0
Dense (64)	(None, 64)	8,256
Batch Normalization	(None, 64)	256
Dropout	(None, 64)	0
Dense (1)	(None, 1)	65
Total		11,178,945

4.3.1. Resultados de Model3_DeepCNN_ReLU

El reporte de clasificación para este modelo es el siguiente:

Precisión: 0.93

Recall: 0.93

F1-score promedio: 0.93

Soporte: 3561

5. Gráficos de los modelos

A continuación se muestran los gráficos generados por cada modelo:

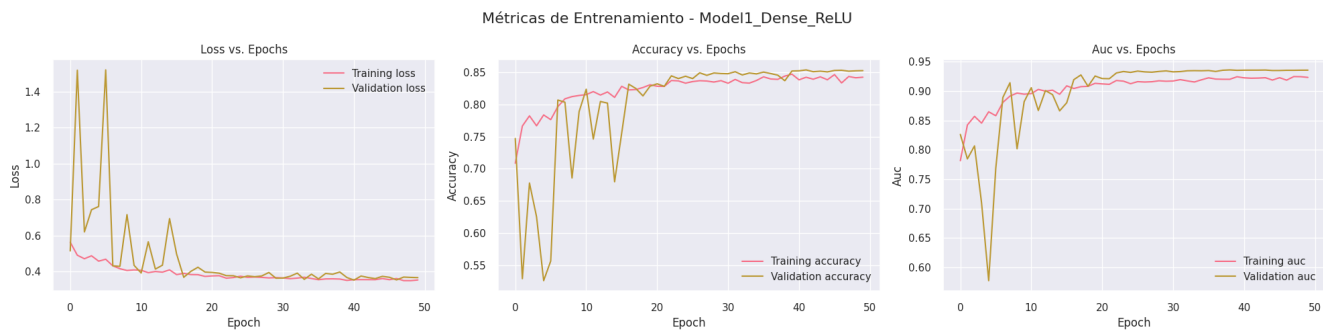


Figura 1: Métricas de entrenamiento para el modelo 1

- **Loss vs. Epochs:** Observamos una disminución en la pérdida tanto en el conjunto de entrenamiento como en el de validación. Sin embargo, en ciertos puntos de la curva, particularmente en las primeras épocas, se pueden ver fluctuaciones que indican inestabilidad en el proceso de optimización. Esto podría señalar un aprendizaje irregular en las primeras etapas del entrenamiento.
- **Accuracy vs. Epochs:** La precisión muestra un aumento progresivo, alcanzando valores cercanos al 85 % al final del entrenamiento. Las curvas de entrenamiento y validación están alineadas, lo cual sugiere que el modelo no sufre de un sobreajuste significativo.
- **AUC vs. Epochs:** La curva AUC mantiene valores relativamente altos, alcanzando un AUC cercano a 0.93. Esto indica que el modelo tiene un buen rendimiento en la clasificación de las clases, con una capacidad aceptable para distinguir entre ellas.

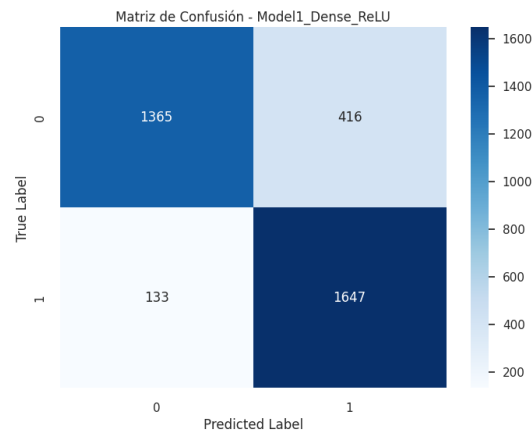


Figura 2: Métricas de entrenamiento para el modelo 1

- El modelo predice correctamente 1365 instancias de la clase negativa y 1647 de la clase positiva, mientras que se observan 416 falsos positivos y 133 falsos negativos. Esto sugiere una ligera tendencia a clasificar incorrectamente algunos ejemplos de la clase negativa como positivos.

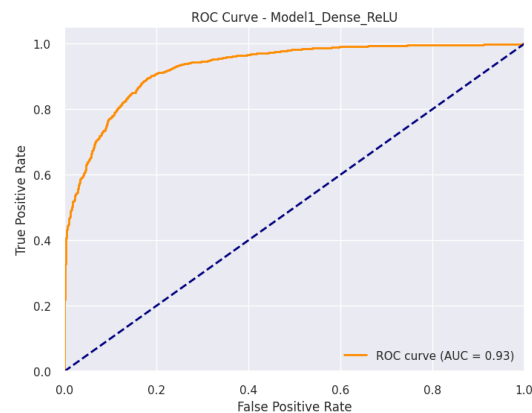


Figura 3: Métricas de entrenamiento para el modelo 1

- La curva ROC muestra un área bajo la curva (AUC) de 0.93, lo que indica una buena capacidad de discriminación del modelo. Sin embargo, existen áreas de mejora, ya que un AUC más cercano a 1 sería ideal.

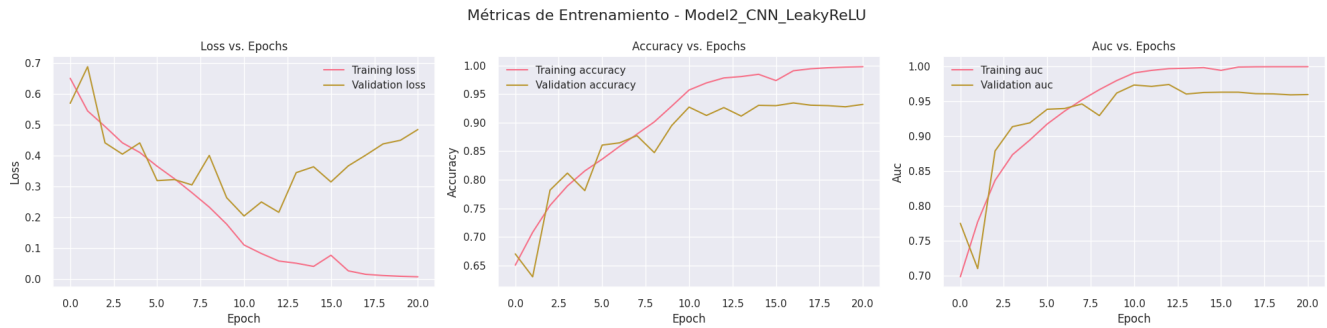


Figura 4: Métricas de entrenamiento para el modelo 1

- **Loss vs. Epochs:** La pérdida disminuye de manera consistente tanto para el conjunto de entrenamiento como para el de validación, sin las fluctuaciones observadas en el modelo anterior. Esto sugiere un aprendizaje más estable.
- **Accuracy vs. Epochs:** La precisión aumenta de manera progresiva, alcanzando valores superiores al 95 %. La alineación entre la precisión de entrenamiento y la de validación indica que el modelo generaliza bien sin signos de sobreajuste.
- **AUC vs. Epochs:** El modelo alcanza un AUC de 0.97 al final del entrenamiento, lo cual es un excelente indicador de su capacidad para clasificar correctamente entre las clases.

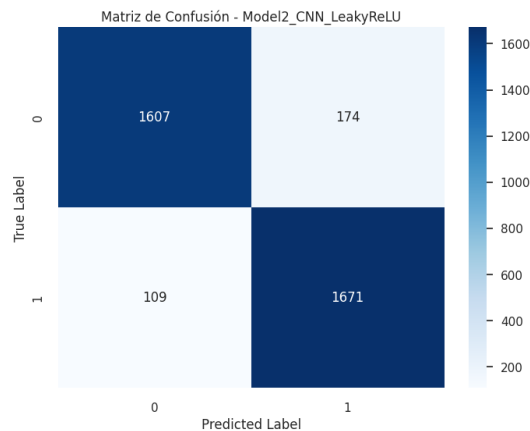


Figura 5: Métricas de entrenamiento para el modelo 1

- El modelo logra una alta precisión en ambas clases, con 1607 verdaderos negativos y 1671 verdaderos positivos. Los errores de clasificación son menores en comparación con el Modelo 1, con solo 174 falsos positivos y 109 falsos negativos.

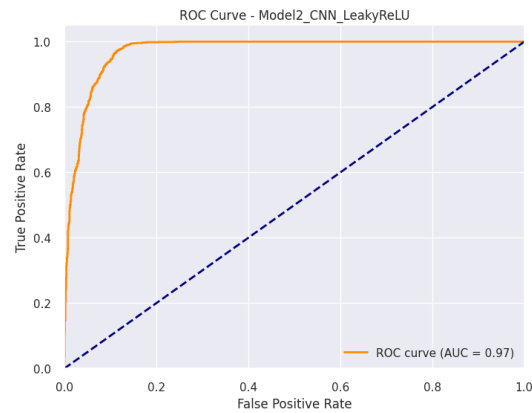


Figura 6: Métricas de entrenamiento para el modelo 1

- La curva ROC muestra un AUC de 0.97, indicando una capacidad muy alta de discriminación. Este modelo es capaz de distinguir entre clases de manera eficiente, con una tendencia mucho menor a clasificar incorrectamente.

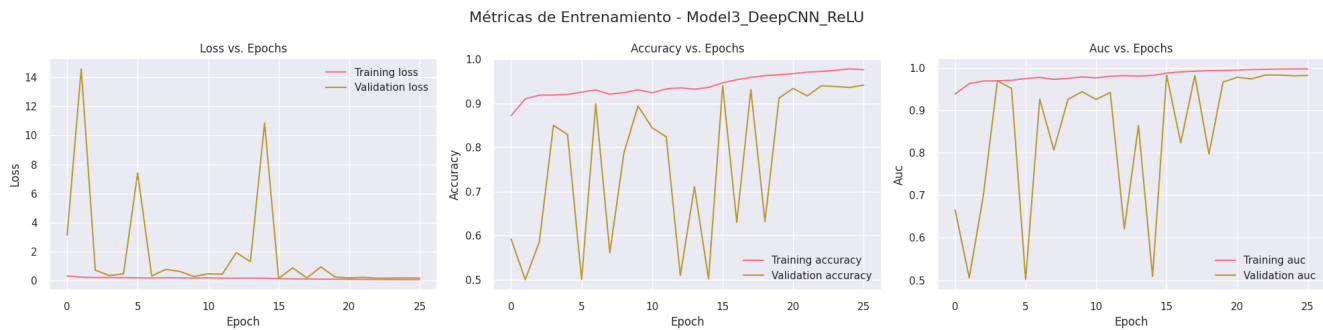


Figura 7: Métricas de entrenamiento para el modelo 1

- Loss vs. Epochs: La pérdida de entrenamiento y validación disminuye rápidamente, aunque muestra algunas fluctuaciones al principio del entrenamiento. A medida que avanza el entrenamiento, el modelo logra estabilizarse y alcanzar una baja pérdida.
- Accuracy vs. Epochs: La precisión es alta, superando el 95 %, y se mantiene consistente entre el entrenamiento y la validación. Esto indica que el modelo generaliza adecuadamente y se adapta bien a los datos.
- AUC vs. Epochs: La curva AUC también alcanza valores cercanos a 0.98, sugiriendo una excelente capacidad para diferenciar entre clases.

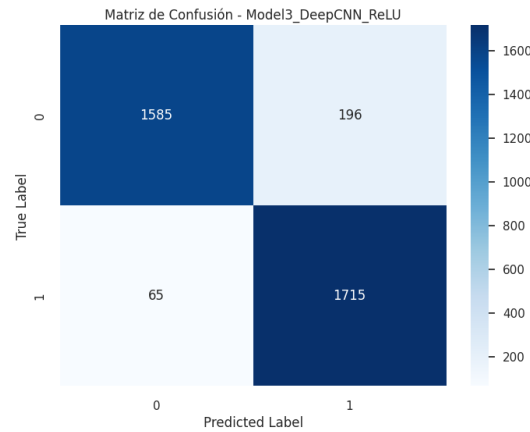


Figura 8: Métricas de entrenamiento para el modelo 1

- En términos de predicciones, el modelo identifica correctamente 1585 instancias de la clase negativa y 1715 de la clase positiva. Tiene un bajo número de errores con 196 falsos positivos y 65 falsos negativos, lo que representa el menor error entre los tres modelos.

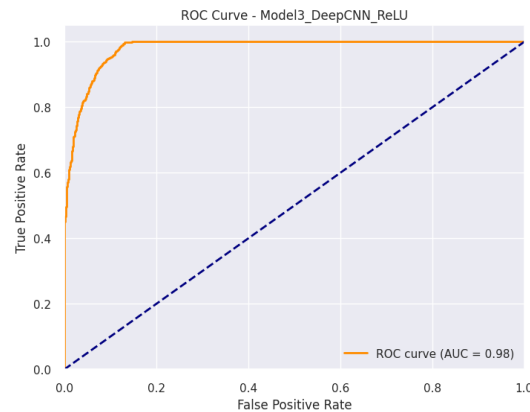


Figura 9: Métricas de entrenamiento para el modelo 1

- La curva ROC muestra un AUC de 0.98, lo cual es excelente y confirma la alta capacidad de discriminación del modelo. Este resultado sugiere que el modelo tiene el menor compromiso entre falsos positivos y falsos negativos.

6. Comparación de Modelos y Conclusión

El mejor desempeño fue logrado por el modelo `Model3_DeepCNN_ReLU`, con una precisión de 93 % en el conjunto de prueba y un AUC de 0.9796, lo cual lo convierte en el modelo más adecuado para la clasificación de melanoma en este experimento.

- El Modelo 3: Deep CNN con activación ReLU es el que mejor rendimiento muestra en términos de estabilidad en el entrenamiento, precisión en la predicción y capacidad de discriminación, representado por el alto valor de AUC y la baja tasa de falsos positivos y falsos negativos. Este modelo es el más adecuado para la tarea de clasificación de melanoma en este conjunto de datos.
- El Modelo 2: CNN con Leaky ReLU también muestra un desempeño notable, aunque con un AUC ligeramente menor que el Modelo 3. Es una buena alternativa en caso de que se desee una arquitectura menos compleja que aún ofrezca un excelente rendimiento.
- El Modelo 1: Dense con ReLU tiene un rendimiento aceptable, pero presenta mayor inestabilidad durante el entrenamiento y una mayor tasa de errores en la clasificación. Si bien es una opción válida, los modelos de convolución (Modelos 2 y 3) demuestran ser más efectivos para esta tarea, probablemente debido a su capacidad para extraer características espaciales en los datos de imagen.

7. Referencias

- Dataset de Kaggle: <https://www.kaggle.com/drscarlat/melanoma>

8. Repositorio

La implementación se encuentra en el siguiente repositorio: [GitHub](#).