



INSTITUTO TECNOLÓGICO DE ESTUDIOS SUPERIORES DE MONTERREY
ESCUELA DE INGENIERÍA Y CIENCIAS
CONCENTRACIÓN INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA DE DATOS
MONTERREY, NUEVO LEÓN

TC3006C.103. INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA
DE DATOS I

**MÓDULO 2 ANÁLISIS Y REPORTE SOBRE EL DESEMPEÑO DEL
MODELO: REDES NEURONALES**

PORTAFOLIO ANÁLISIS

MARIO JAVIER SORIANO AGUILERA A01384282

10 DE SEPTIEMBRE DE 2023

Índice

1. Introducción	2
2. Dataset	2
3. Modelos	2
3.1. Modelo 1	2
3.2. Modelo 2	3
3.3. Modelo 3	5
3.4. Modelo 4	6
3.5. Modelo 5	7
3.6. Modelo 6	8
3.7. Modelo 7	9
3.8. Modelo 8	10
3.9. Modelo 9	11
3.10. Modelo 10	11
4. Análisis de Sesgo, Varianza y Ajuste del Modelo	12
5. Matriz de Confusión y Métricas de Evaluación	13
6. Conclusiones	13

1. Introducción

En el ámbito del aprendizaje automático y la inteligencia artificial, las redes neuronales son un enfoque poderoso para abordar problemas de clasificación binaria. En este informe, se presenta un análisis detallado de varios modelos de redes neuronales entrenados y evaluados para la clasificación binaria de datos. El objetivo principal de este informe es identificar el modelo con el mejor rendimiento y analizar el grado de sesgo, varianza y ajuste del modelo.

Para llevar a cabo esta evaluación, se han implementado varios modelos de redes neuronales, cada uno con configuraciones de hiperparámetros y técnicas de regularización específicas. Se examinarán los resultados de cada modelo, se analizarán los patrones de sesgo y varianza, y se determinará el nivel de ajuste de cada modelo.

2. Dataset

El conjunto de datos Cancer Data contiene información sobre 570 células cancerosas y 30 características que se utilizan para determinar si las células cancerosas en los datos son benignas o malignas. Estas características se utilizan para realizar diagnósticos médicos y son fundamentales en la detección temprana y el tratamiento del cáncer.

Las características incluyen medidas como el radio, la textura, el perímetro, el área, la suavidad, la compacidad, la concavidad y los puntos cónicos de células cancerosas. Estos valores se han recopilado y registrado para cada célula en el conjunto de datos.

El conjunto de datos se divide en dos tipos de cáncer: benigno (B) y maligno (M), lo que permite a los profesionales de la salud y a los científicos utilizar estas características para determinar la naturaleza del cáncer en las células analizadas. La capacidad de distinguir entre cáncer benigno y maligno es esencial para tomar decisiones médicas informadas y brindar el tratamiento adecuado a los pacientes.

Una red neuronal de clasificación binaria es una buena elección para el conjunto de datos "Cancer Data". Esto se debe a su capacidad para procesar las numerosas características de las células cancerosas y aprender relaciones complejas en los datos. Además, su escalabilidad y flexibilidad en la arquitectura permiten adaptarse a diferentes tamaños de conjunto de datos y experimentar con configuraciones de modelo.

3. Modelos

3.1. Modelo 1

Características del Modelo:

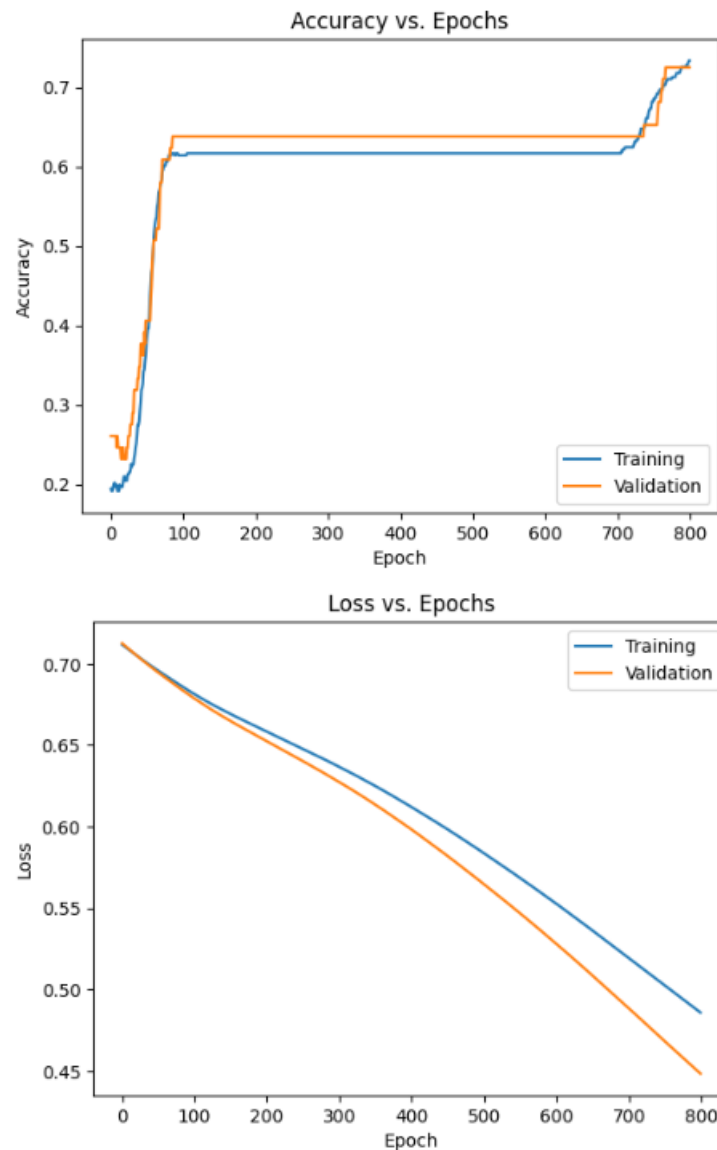


Figura 1: Modelo 1 gráfico accuracy y loss

- Arquitectura: Red neuronal con 9 capas ocultas y activación ReLU.
- Optimizador: SGD con learning rate de 0.0001.
- Precisión en el conjunto de prueba: 81.58 %

El modelo parece estar haciendo un progreso rapido en las primeras epocas pero luego el accuracy junto con loss tiene un estancamiento, depues empieza a mejorar un poco mas el modelo. Este tipo de optimizador SGD suele ser muy sensible en los hiperparámetros.

3.2. Modelo 2

Características del Modelo:

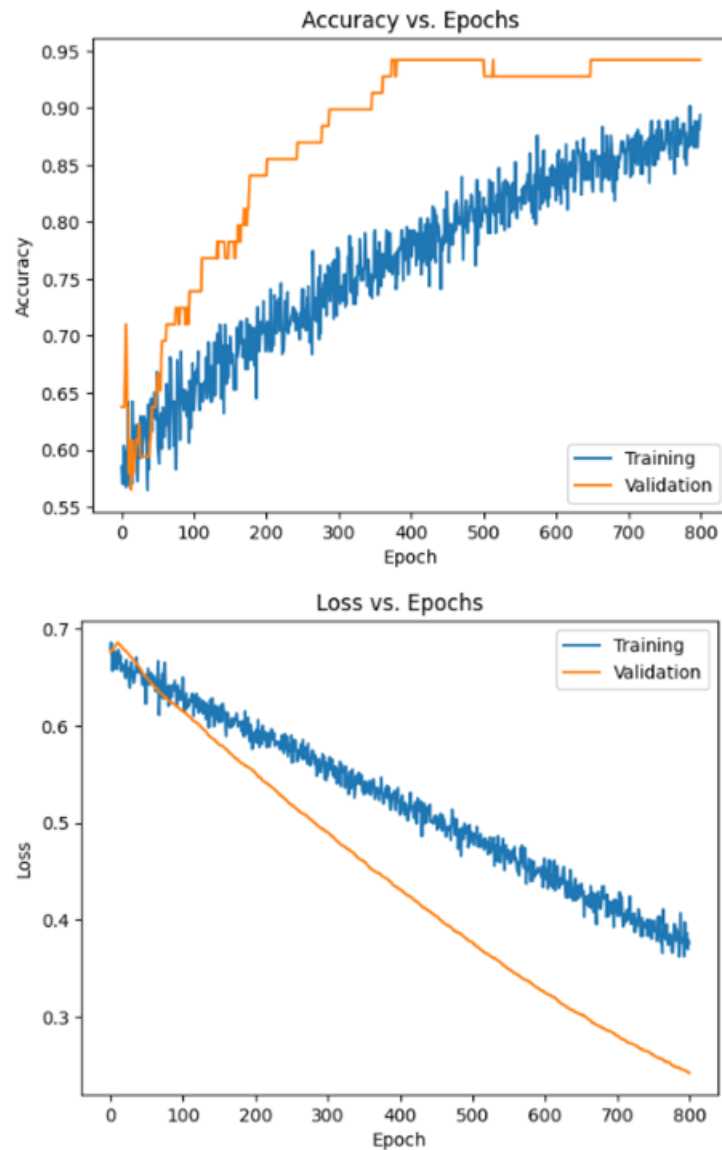


Figura 2: Modelo 2 gráfico accuracy y loss

- Arquitectura: Red neuronal similar al Modelo 1, pero con Dropout y Batch Normalization.
- Optimizador: SGD con learning rate de 0.0001.
- Precisión en el conjunto de prueba: 94.74 %

El modelo parece estar haciendo un mejor progreso que el anterior implementando Dropout y Batch Normalization, disminuyendo la perdida y aumentando la precisión tanto en el conjunto de entrenamiento como en el conjunto de validación a medida que avanza en las épocas, pero aun nuestro modelo no esta convergiendo que es lo que buscamos. Vemos como SGD necesita tener mas épocas o un learning rate mas alto para que tenga mejor ajuste nuestro modelo.

3.3. Modelo 3

Características del Modelo:

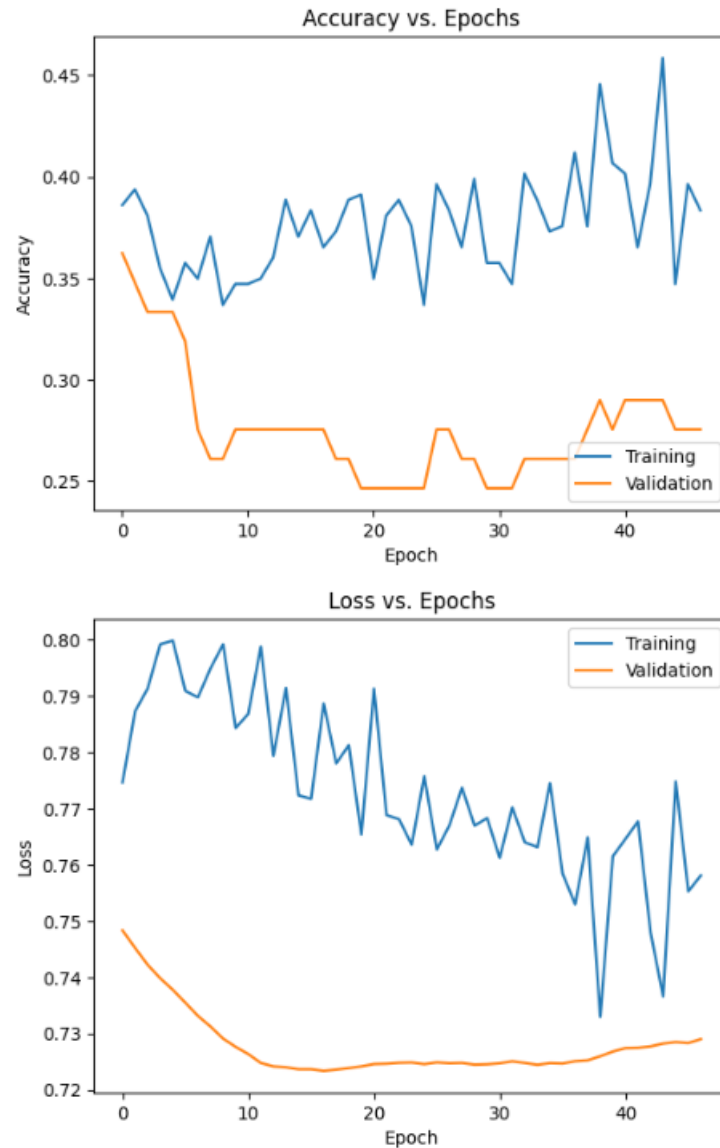


Figura 3: Modelo 3 gráfico accuracy y loss

- Arquitectura: Misma arquitectura que el Modelo 2.
- Optimizador: SGD con learning rate de 0.0001.
- Técnicas de Regularización: Early Stopping y Reducción del Learning Rate.
- Precisión en el conjunto de prueba: 94.74 %

Al implementar las tecnicas de earlystopping, learning rate reduction, el modelo al implementar earlys-topping y tener un learning rate muy bajo implementado con el optimizador SGD vemos que el ajuste del modelo se detiene pronto, siendo no muy bueno en los datos de entrenamiento y validación pero si en los

datos de prueba, se cree que esto se debe a que no se dieron proporciones adecuadas ya que se utilizaron en todas divisiones de entrenamiento 80 %, de prueba 20 % y de validacion 15 % a los datos de validación, entrenamiento y prueba. Como también que el modelo tuvo un estancamiento temprano en esos épocas como para que se estabilizara. Esto sera modificado en los últimos modelos.

3.4. Modelo 4

Características del Modelo:

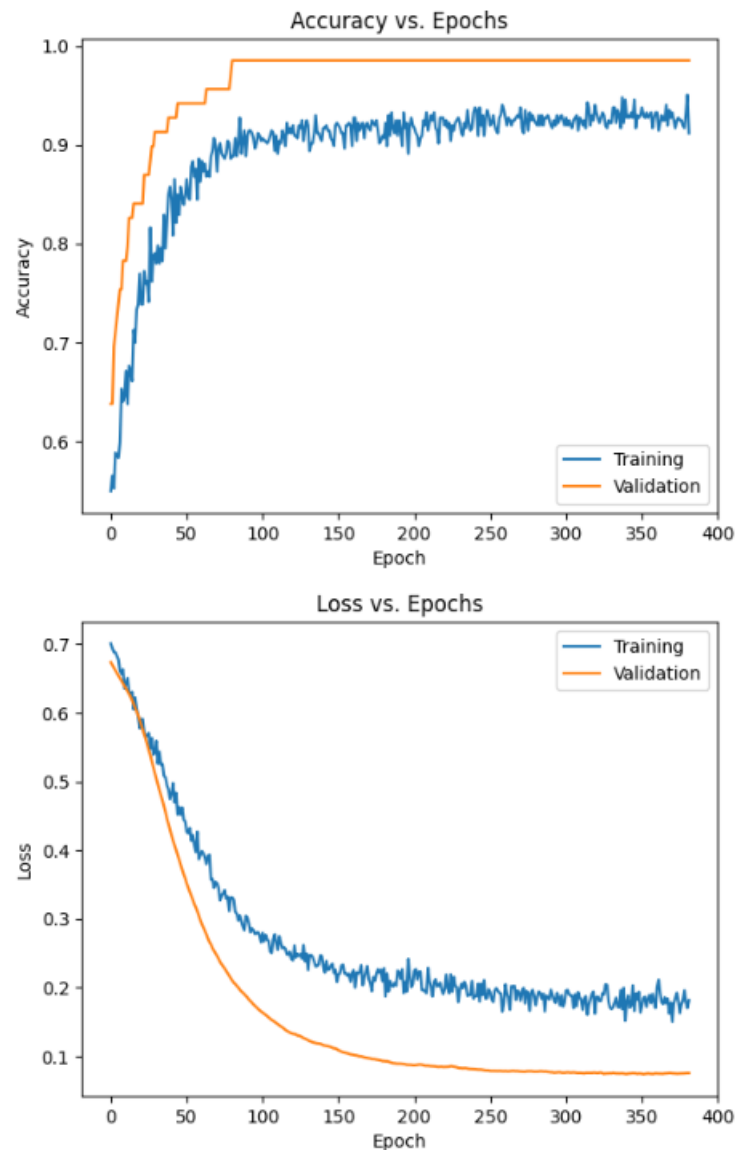


Figura 4: Modelo 4 gráfico accuracy y loss

- Arquitectura: Misma arquitectura que el Modelo 2.
- Optimizador: SGD con learning rate de 0.001.
- Técnicas de Regularización: Early Stopping y Reducción del Learning Rate.

- Precisión en el conjunto de prueba: 96.49 %

Se puede apreciar como ahora nuestro modelo si esta convergiendo en los graficos de accuracy y loss indicando que si esta identificando los patrones y aprendiendo adecuadamente nuestro modelo. Pero se esta ajustando muy lento y hay overfitting.

3.5. Modelo 5

Características del Modelo:

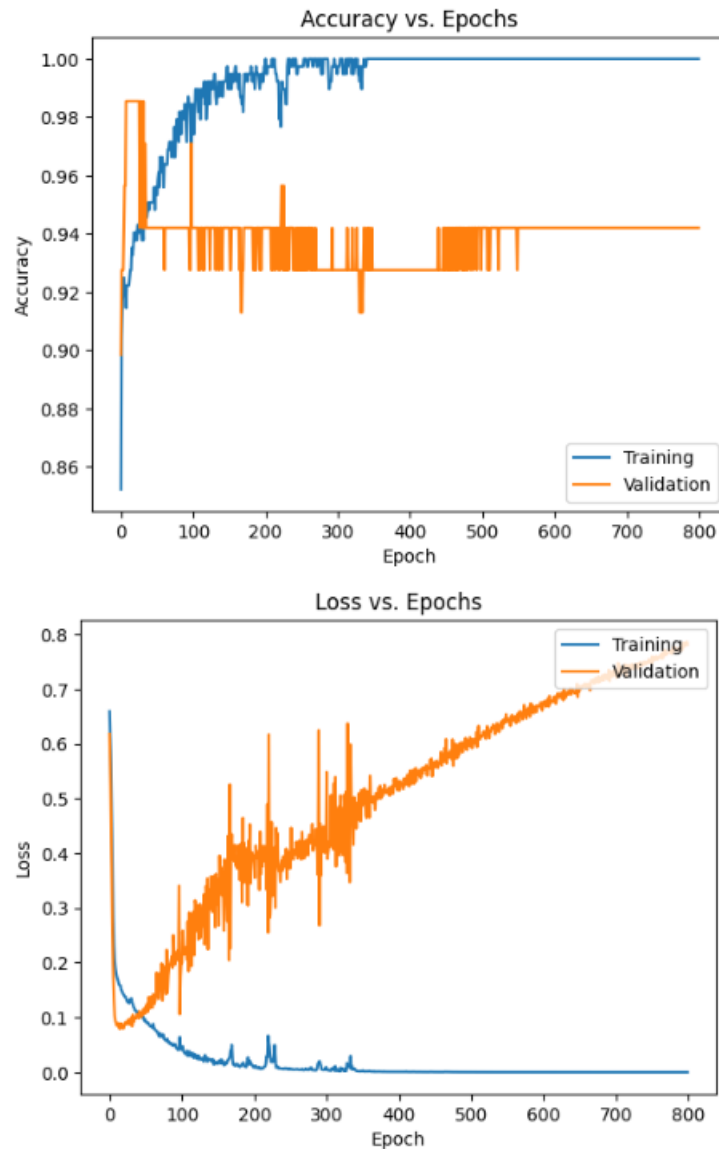


Figura 5: Modelo 5 gráfico accuracy y loss

- Arquitectura: Misma arquitectura que el Modelo 1.
- Optimizador: Adam con learning rate de 0.0001.
- Precisión en el conjunto de prueba: 92.98 %

Nuestro modelo muestra claramente overfitting.

3.6. Modelo 6

Características del Modelo:

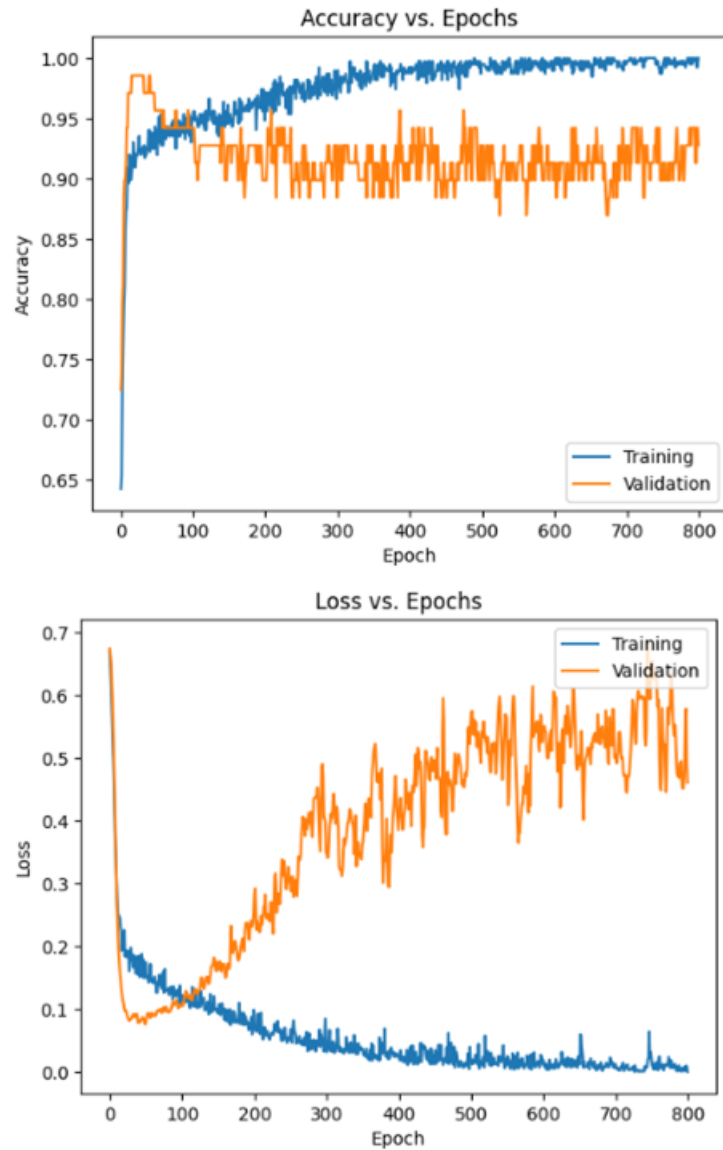


Figura 6: Modelo 6 gráfico accuracy y loss

- Arquitectura: Misma arquitectura que el Modelo 2.
- Optimizador: Adam con learning rate de 0.0001.
- Precisión en el conjunto de prueba: 96.49 %

Igual overfitting pero la convergencia parece extenderse.

3.7. Modelo 7

Características del Modelo:

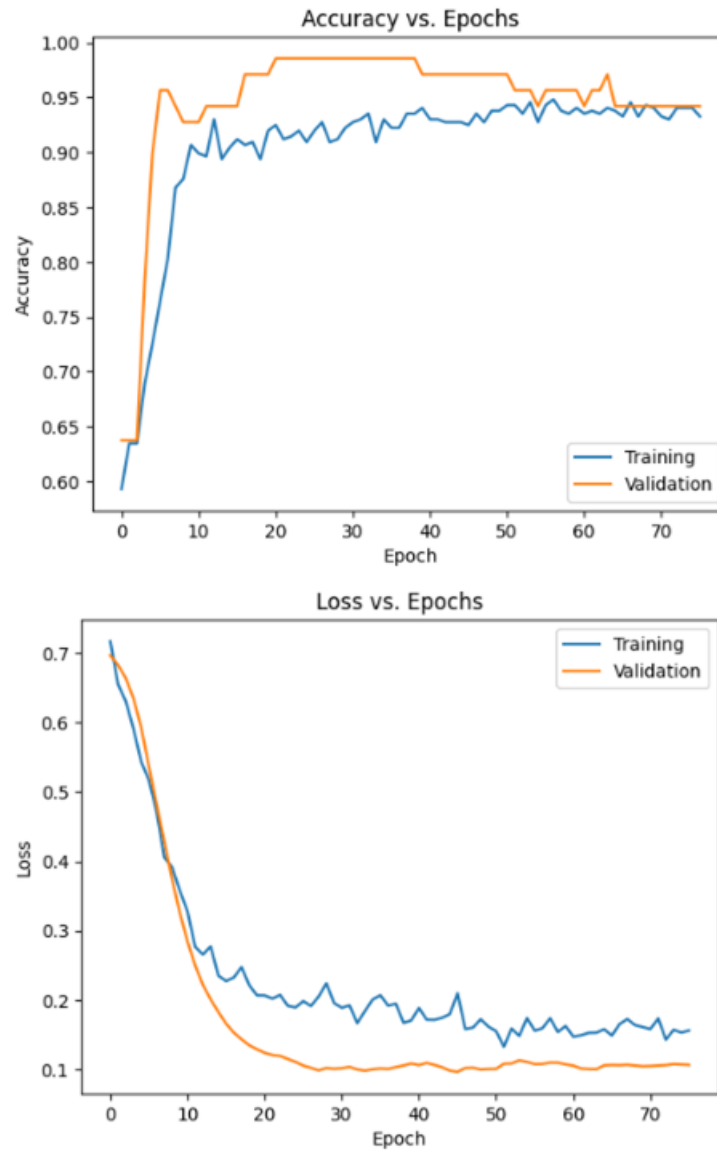


Figura 7: Modelo 7 gráfico accuracy y loss

- Arquitectura: Misma arquitectura que el Modelo 2.
- Optimizador: Adam con learning rate de 0.0001.
- Técnicas de Regularización: Early Stopping y Reducción del Learning Rate.
- Precisión en el conjunto de prueba: 95.61 %

El modelo se detiene pero se puede ver que tenemos algo de overfitting, para evitarlo se puede hacer que tenga menos paciencia en el numero de epocas y asi podriamos lograr un mejor modelo.

3.8. Modelo 8

Características del Modelo:

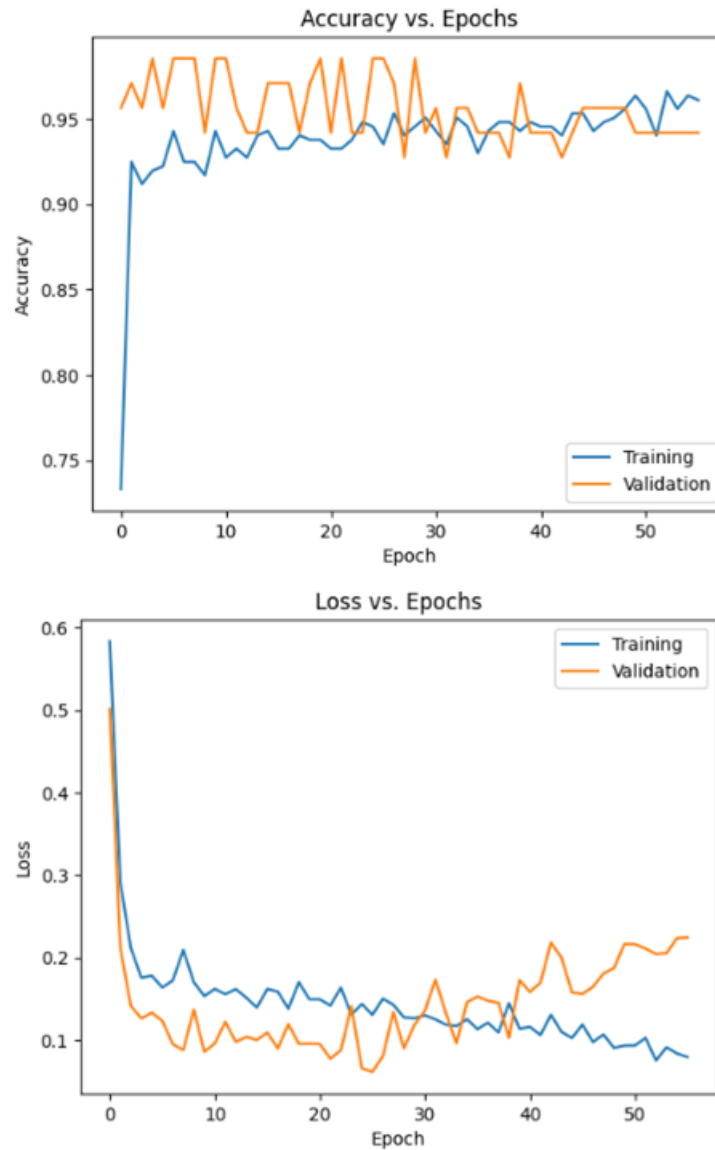


Figura 8: Modelo 8 gráfico accuracy y loss

- Arquitectura: Misma arquitectura que el Modelo 2.
- Optimizador: Adam con learning rate de 0.001.
- Técnicas de Regularización: Early Stopping y Reducción del Learning Rate.
- Precisión en el conjunto de prueba: 96.49 %

Es igual al anterior pero solo se modifico el learning rate, vemos como eso afecta al modelo mostrando un mejor accuracy y un menor loss pero se puede apreciar que tiene overfitting por como avanza la linea de validacion con la de entrenamiento

3.9. Modelo 9

Características del Modelo:

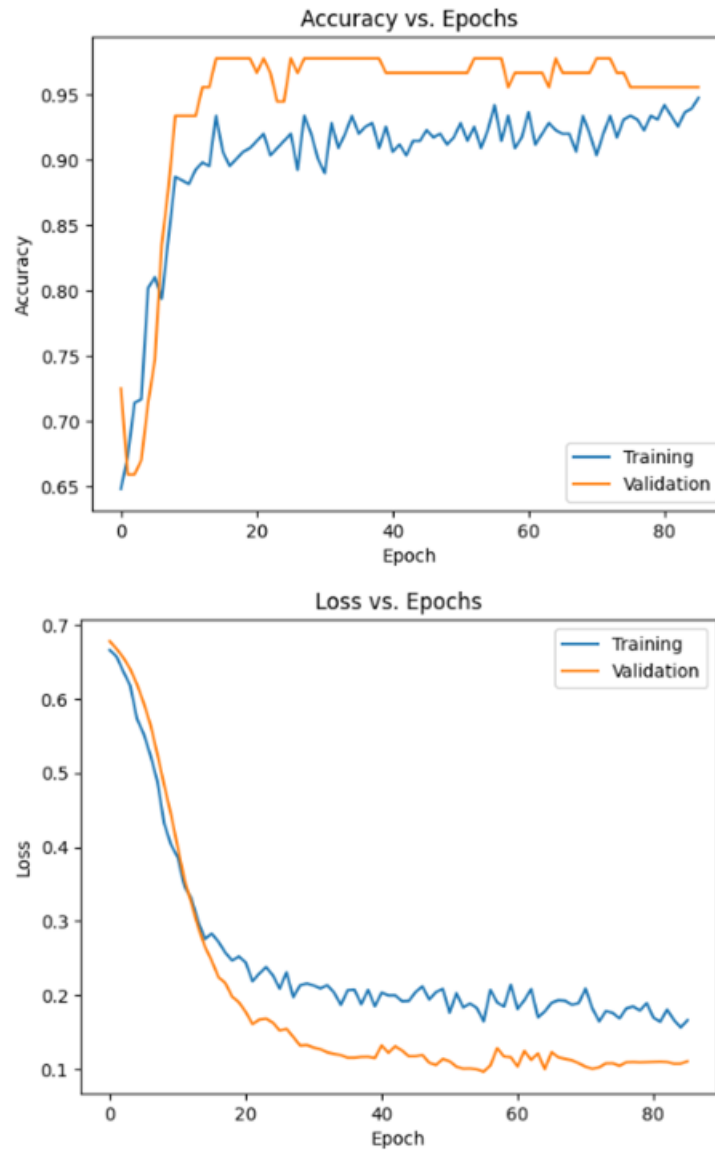


Figura 9: Modelo 9 gráfico accuracy y loss

- Arquitectura: Misma arquitectura que el Modelo 2.
- Optimizador: SGD con learning rate de 0.01.
- Técnicas de Regularización: Early Stopping y Reducción del Learning Rate.
- Precisión en el conjunto de prueba: 95.61 %

3.10. Modelo 10

Características del Modelo:

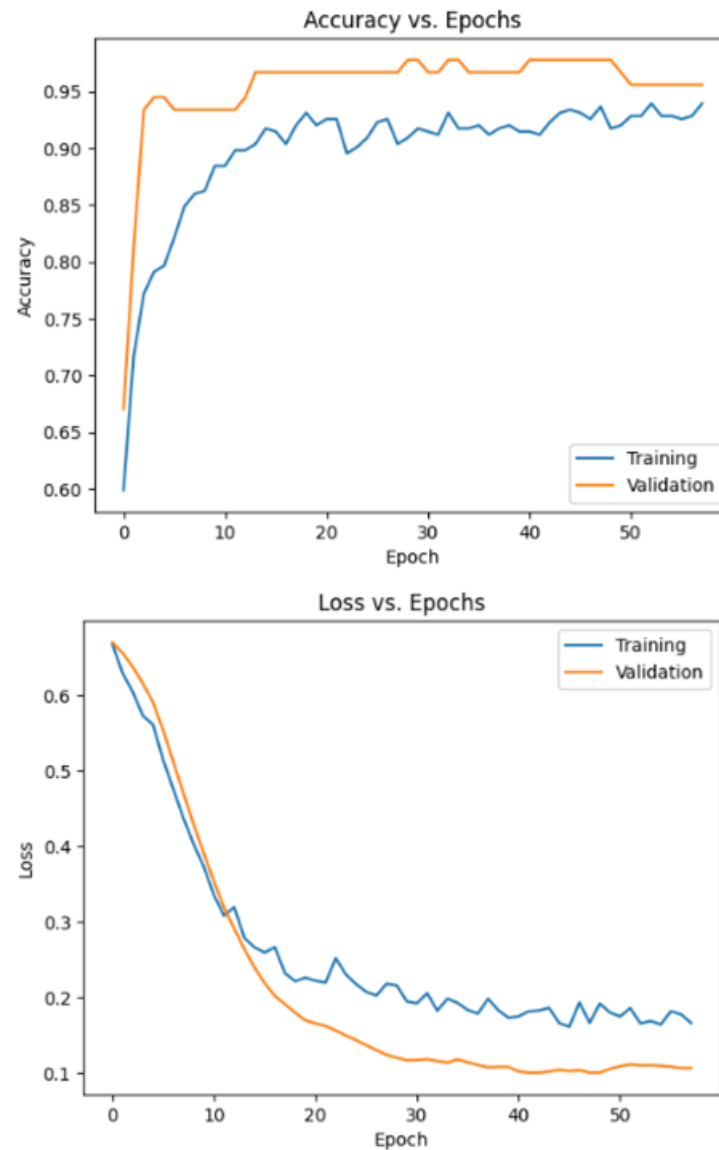


Figura 10: Modelo 10 gráfico accuracy y loss

- Arquitectura: Misma arquitectura que el Modelo 2.
- Optimizador: Adam con learning rate de 0.0001.
- Técnicas de Regularización: Early Stopping y Reducción del Learning Rate.
- Precisión en el conjunto de prueba: 95.61 %

4. Análisis de Sesgo, Varianza y Ajuste del Modelo

- El Modelo 1 muestra un sesgo medio y varianza media, pero lento aprendizaje, no converge, estancamiento en muchas épocas.
- El Modelo 2 mejora el sesgo y la varianza con Dropout y Batch Normalization, pero todavía no converge completamente.

- El Modelo 3 usa técnicas de regularización como Early Stopping y Reducción del Learning Rate, pero no converge adecuadamente debido al bajo learning rate y a la división de los datos (Train/Test/Validation).
- El Modelo 4 con learning rate más alto logra converger mejor, aun hay overfitting. Muestra un sesgo alto y baja varianza
- Los Modelos 5 y 6 muestran overfitting debido a un gran número de épocas. Muestra un sesgo bajo y alta varianza
- El Modelo 7 mejora con técnicas de regularización, pero aún tiene overfitting. Muestra un sesgo bajo y baja varianza.
- El Modelo 8 con learning rate más, muestra overfitting más evidente. Muestra un sesgo bajo y alta varianza.
- El Modelo 9 logra converger bien, muestra menos estancamiento que los otros optimizadores Sgd.
- El Modelo 10 también tiene un buen rendimiento, muestra menos estancamiento que los otros optimizadores Adam.

5. Matriz de Confusión y Métricas de Evaluación

Los Modelos 9 y 10 exhiben un rendimiento sólido en la tarea para la que fueron entrenados, con resultados muy similares. Ambos modelos logran una alta precisión del 94.87 % y 94.74 %, respectivamente, lo que indica que la mayoría de sus predicciones son correctas. También tienen una sólida sensibilidad (o recall) del 94.87 % y 92.31 %, lo que significa que identifican eficazmente los casos positivos. La especificidad, que mide la capacidad de identificar casos negativos, es del 94.87 % y 92.31 %, respectivamente. La exactitud general es alta, con un 96.49 % y 95.61 % para los Modelos 9 y 10, respectivamente. Además, el F1-Score, que combina precisión y sensibilidad, se mantiene alto en ambos modelos, con un 94.87 % y 93.51 %. Estos resultados sólidos sugieren que ambos modelos son efectivos en su tarea y proporcionan un rendimiento confiable.

6. Conclusiones

La aplicación de técnicas de regularización, como Early Stopping y la reducción del Learning Rate, se revela como una práctica esencial para prevenir el overfitting y mejorar significativamente el rendimiento del modelo. La elección adecuada del optimizador y la configuración del learning rate son determinantes cruciales en el proceso de entrenamiento de una red neuronal, influyendo en su capacidad para aprender y generalizar a partir de los datos. El cuidadoso ajuste en la partición de datos en conjuntos de entrenamiento, validación y prueba resulta fundamental para obtener resultados confiables y representativos. Los optimizadores SGD (Descenso de Gradiente Estocástico) y Adam (Adaptive Moment Estimation) son dos algoritmos esenciales en el aprendizaje automático. Una diferencia clave radica en la sensibilidad a la elección de la tasa de aprendizaje: SGD es más sensible a la tasa de aprendizaje, que es un hiperparámetro crítico, mientras que

Adam se adapta automáticamente, reduciendo la necesidad de ajustes precisos. Esta distinción hace que Adam sea más popular en muchas aplicaciones de aprendizaje automático.