

<div>FORMATO DE PROPUESTA DE PROYECTO</div> <div>Clasificación de imágenes con redes neuronales</div> <div>Semestre: 2025-1</div> <div>Fecha de entrega: 18 de Octubre, 2024</div>
Nombre de los Integrantes:
Integrante 1: Partida Contreras Marían de los Ángeles
Integrante 2: Pérez Montes Leslie Guadalupe
Integrante 3: Téllez Soto José Emanuel
1. Título del Proyecto.
Diagnóstico automatizado de Cáncer de Mama con Redes Neuronales Convolucionales usando Imágenes de Resonancia Magnética o Tomografía Axial Computarizada.
2. Objetivo
Desarrollar y entrenar un modelo de clasificación de imágenes basado en redes neuronales convolucionales (CNN) para identificar y clasificar tumores en imágenes de resonancias magnéticas, con el propósito de apoyar el diagnóstico temprano y preciso del cáncer de mama mediante el uso de técnicas de aprendizaje profundo.
3. Herramientas para la implementación del proyecto
<ul style="list-style-type: none">● Python: Para el desarrollo de scripts de procesamiento de datos, simulaciones, y análisis estadísticos.● Matplotlib y Seaborn: Para la visualización de datos.● Pandas: Para la manipulación de datos en formato tabular.● NumPy y SciPy: Para cálculos numéricos avanzados.● Jupyter Notebooks: Para documentar y ejecutar análisis interactivos.● TensorFlow: Para el aprendizaje profundo. Proporciona una API flexible y modular que facilita el desarrollo de CNN y otros tipos de redes neuronales● Keras: que simplifica la creación y entrenamiento de modelos.● GitHub o GitLab: Para el control de versiones y colaboración en equipo.● FastAPI: Para crear APIs web rápidas, eficientes y fáciles de mantener, con validación automática de datos y documentación generada automáticamente.● Pydantic: Para validación de datos.● Docker: Para empaquetar, distribuir y ejecutar aplicaciones en contenedores.
4. Orígenes de datos.
Tomaremos imágenes de resonancias magnéticas de varios repositorios encontrados en la siguiente liga: https://www.cancerimagingarchive.net/browse-collections/ , nosotros usaremos la imagen de la resonancia magnetica de un ceno con y sin saturación para nuestro modelo. Estas imágenes ya están clasificadas, por lo que nos hará más fácil el entrenamiento de nuestra red neuronal.
5. Principales actividades a realizar.
Recolección y preparación de datos: <ul style="list-style-type: none">- Obtener y limpiar los datos necesarios para el análisis, posteriormente usar Pandas para cargar y manipular los datos, transformándolos en formatos adecuados para el análisis.- Desarrollo y entrega de un pipeline de preprocesamiento y manejo de datos para la clasificación de imágenes.- Utilizar herramientas de control de versiones como GitHub para mantener un historial del progreso y colaborar en equipo.

Diseño de la arquitectura de la CNN:

- Determinar las capas convolucionales, las capas de pooling y las capas densas que usaremos en la CNN.
- Determinar la función de activación.

Compilación del modelo:

- Definir la función de pérdida para medir el error.
- Definir el optimizador que usaremos para ajustar los pesos de la red.
- Definir las métricas para evaluar el desempeño durante el entrenamiento

Entrenamiento del modelo:

- En esta fase, se alimenta el modelo con los datos de entrenamiento para que aprenda los patrones subyacentes.

Validación del modelo:

- Durante el entrenamiento, usaremos técnicas como la **validación cruzada** para evaluar el desempeño del modelo en diferentes subconjuntos de datos y evitar sobreajuste (overfitting).

Ajuste y optimización del modelo:

- Dado que el modelo se va a implementar en un entorno clínico, es necesario optimizarlo para que sea eficiente en términos de velocidad de inferencia y uso de recursos.
- Esto puede implicar el ajuste de parámetros internos (como los coeficientes de una regresión) o de hiperparámetros.

Pruebas finales:

- Realizaremos pruebas adicionales para optimizar el rendimiento del modelo o la simulación.
- Ajustaremos parámetros y realizaremos iteraciones para mejorar la precisión de los resultados.
- Tras entrenar y evaluar el modelo, es probable que se necesite ajustes adicionales. Puede ser necesario volver a entrenar con diferentes configuraciones o probar modelos alternativos.

Implementación del modelo:

- Implementar el modelo por medio de una API para el consumo del modelo.
- Combinaremos **Bucker** y **FastAPI** para cargar grandes de datos, en nuestro caso imágenes, y usar el modelo.

6. Observaciones y comentarios generales del Alumno.

1. Calidad y disponibilidad de los datos

- **Disponibilidad de datos:** El factor clave para este proyecto es el acceso a un conjunto de imágenes de resonancia magnética de calidad.
- **Preprocesamiento de las imágenes:** Las imágenes de resonancia magnética a menudo contienen ruido o variaciones que serán eliminadas o corregidas antes de ser procesadas por nuestra red neuronal.

2. Complejidad del modelo y poder computacional

- **Profundidad del modelo:** A medida que las redes se hacen más profundas, el modelo puede ser más preciso, pero también más propenso a problemas como el sobreajuste (overfitting) y el aumento del tiempo de entrenamiento.

3. Evaluación del rendimiento del modelo

- **Métricas de evaluación:** Para el diagnóstico de cáncer, será crucial utilizar métricas de evaluación adecuadas.
- **Validación cruzada:** Si el conjunto de datos es limitado, aplicaremos técnicas como la validación cruzada para asegurar que el modelo generaliza bien y no está simplemente aprendiendo patrones de un subconjunto específico de datos.

4. Preocupaciones éticas y legales

- **Privacidad de los datos:** Dado que se está trabajando con datos médicos sensibles, cumpliremos con las regulaciones de privacidad. Las imágenes de resonancia magnética serán anonimizadas para proteger la identidad de los pacientes.

5. Validación

- **Validación clínica y la aprobación regulatoria:** Son necesarias antes de desplegar el modelo en entornos médicos reales.

7. Observaciones y resultado de la revisión del profesor.

