

# EXAMEN PARCIAL

## CUADERNOS DE CÓDIGO PARA DOCENTES

Este examen está disponible desde el 28 de agosto hasta el 05 de setiembre.

Presentado por: Eginhardo Navarro Honda

---

### INDICACIONES

1. Cada pregunta se responde con un cuaderno de código independiente en formato ipynb, excepto la pregunta 1, que se presenta en PDF.
2. Cargue en su repositorio de Github sus archivos ipynb, PDF y CSV.
3. Adjunte en Blackboard la dirección web de su repositorio en Github.

	PREGUNTA	PUNTOS
<input type="checkbox"/>	Pregunta 1: Problemática y planteamiento de trabajo.	6
<input type="checkbox"/>	Pregunta 2: Análisis exploratorio de datos.	6
<input type="checkbox"/>	Pregunta 3: Revisión de artículos asistida por IA	8

### Pregunta 1: (6 puntos)

- (a) Explique la problemática para un tema de investigación de su interés.

Deep learning en imágenes médicas

Se plantea la pregunta

⊞ ¿Cuáles son las dificultades tecnológicas al usar Deep Learning en el procesamiento de imágenes médicas, y cómo se puede mejorar la seguridad en Deep Learning?

Recientemente se ha estado afianzándose Deep learning como un potente instrumento para un vasto dominio en imágenes.

Esto es, desde la generación, clasificación, reconstrucción, predicción, segmentación, diagnóstico e interpretación.

Deep learning usa redes neurales convolucionales que usa ingentes cantidades de datos etiquetados para simular el complejo aprendizaje del cerebro humano.

No obstante, estos aprendizajes no siempre representan a la población y conlleva sesgos en los modelos afectando la equidad y precisión en un diagnóstico.

Además, la administración de la data para el entrenamiento de los modelos de Deep Learning pueden presentar dificultades importantes en la privacidad y seguridad.

- (b) En base a su pregunta de investigación elabore un mapa mental (usando EDOTOR.net o XMIND) que relacione la observación, hipótesis, experimentación y conclusión con el formato IMRyD (Introducción, Metodología, Resultados y Discusión)

Identificación y entrenamiento de redes neuronales de aprendizaje profundo en datos relacionados a la biomedicina

⊞ Introducción

- La inteligencia artificial (IA) permite que las computadoras simulen la inteligencia humana.
- El aprendizaje automático alude al uso de modelos y reducción de dimensionalidad para facilitar que un procesador "aprenda" relaciones dentro de una base de datos.

▫ Objetivo:

Presentar a los nuevos usuarios la utilidad de utilizar adecuadamente las redes neuronales de aprendizaje profundo para los conjuntos de datos biomédicos.

▫ Hipótesis

Las redes neuronales de aprendizaje profundo son rápidas y pueden producir una precisión similar a la humana sin la intervención del usuario.

⊞ Metodología

▫ Se divide en cuatro submódulos constituido por un único cuaderno Jupyter basado en Python usando la biblioteca PyTorch.

- Clasificación de imágenes:

Proceso de determinar un resultado discreto basado en características continuas dentro de los datos de entrada.

- Incremento de datos:

Proceso de crear sintéticamente más datos de imágenes sobre la base de transformaciones como rotación, escala y recorte.

- Segmentación de imágenes:

Método idóneo con alta precisión espacial donde se identifica el lugar de una clase específica dentro de una imagen. Es decir, es la capacidad de generar un mapa de píxeles a partir de datos de entrada.

- Regresión de datos:

Método de entrenamiento para predecir salidas numéricas continuas, como no ocurren en las Convolutional Neural Networks (CNNs o ConvNets) donde la salida para los píxeles es discreta.

⊞ Resultados y Discusión

▫ Al adaptar una CNN preentrenada con arquitectura ResNet18, se obtiene una precisión relativamente alta ( $\geq 90\%$ ).

▫ Al transformar aleatoriamente datos de un conjunto de datos, la red neural posee una menor posibilidad de memorizar datos de entrenamiento, por lo que debe adaptarse a diversas situaciones.

▫ A medida que los datos de entrada fluyen a lo largo de una CNN tradicional, las características en los datos se codifican para predecir una clasificación única para la entrada.

▫ Dentro de una U-Net CNN que usa una ruta de codificación y decodificación, la salida final tiene el mismo tamaño que la entrada. En general, está demostrado que una red entrenada entrega precisiones altas ( $>85\%$ ).

▫ Al usar modelos de regresión lineal simple, árbol de decisiones y regresión de bosque aleatorio, el resultado esperado en el MAE y el RMSE posee una desviación estándar relativamente alta ( $>0,01$ ).

▫ El aprendizaje profundo es una técnica de IA que utiliza una red neuronal para aprender y predecir características simples y complejas en un conjunto de datos.

▫ Conclusión:

Este método es útil para el análisis de datos de imágenes biomédicas por su alta capacidad de identificar relaciones complejas dentro de grandes cantidades de datos por lo que se anima a los usuarios a explorar los tipos de redes neuronales y los métodos de aprendizaje profundo usados en la medicina (Woessner, et al., 2024).

**Referencia**

Woessner, A. E., Anjum, U., Salman, H., Lear, J., Turner, J. T., Campbell, R., . . . Gauch, S. (2024, July). Identifying and training deep learning neural networks on biomedical-related datasets. *Briefings in Bioinformatics*, 25(Supplement\_1). doi:<https://doi.org/10.1093/bib/bbae232>

FileEditInsertToolsViewHelp

Untitled 2 — Edited

Mind Map

Outliner

Topic

Subtopic

Relationship

Summary

Boundary

Insert

ZEN

Pitch

Panel

Identification y entrenamiento de redes neuronales de aprendizaje profundo en datos relacionados a la biomedicina

Introducción

- La inteligencia artificial (IA) permite que las computadoras simulen la inteligencia humana.
- El aprendizaje automático ayuda al uso de modelos y reducción de dimensionalidad para facilitar que un procesador "aprenda" relaciones dentro de una base de datos.

Metodología

- Dividido en cuatro submódulos constituido por un único cuaderno Jupyter basado en Python usando la biblioteca PyTorch.

Resultados y Discusión

- Al adaptar una CNN preentrenada con arquitectura ResNet18, se obtiene una precisión relativamente alta (≥90%).
- Al transformar aleatoriamente datos de un conjunto de datos, la red neural posee una menor posibilidad de memorizar datos de entrenamiento, por lo que debe adaptarse a diversas situaciones.
- A medida que los datos de entrada fluyen a lo largo de una CNN tradicional, las características en los datos se codifican para predecir una clasificación única para la entrada.
- Dentro de una U-Net CNN que usa una ruta de codificación y decodificación, la salida final tiene el mismo tamaño que la entrada. En general, está demostrado que una red entrenada entrega precisiones altas (≥85 %).
- Al usar modelos de regresión lineal simple, árbol de decisiones y regresión de bosque aleatorio, el resultado esperado en el MAE y el RMSE posee una desviación estándar relativamente alta (>0.0).
- El aprendizaje profundo es una técnica de IA que utiliza una red neuronal para aprender y predecir características simples y complejas en un conjunto de datos.
- Conclusión. Este método es útil para el análisis de datos de imágenes biomédicas por su alta capacidad de identificar relaciones complejas dentro de grandes cantidades de datos por lo que se anima a los usuarios a explorar los tipos de redes neuronales y los métodos de aprendizaje profundo usados en la medicina.

Referencia

- Woessner, A. E., Anjum, U., Salman, H., Lear, J., Turner, J. T., Campbell, R., ... Gauch, S. (2024, July). Identifying and training deep learning neural networks on biomedical-related datasets. *Briefings in Bioinformatics*, 25(Supplement\_1), doi: <https://doi.org/10.1093/bib/bbae232>

Objetivo

- Presentar a los nuevos usuarios la utilidad de utilizar adecuadamente las redes neuronales de aprendizaje profundo para los conjuntos de datos biomédicos.

Hipótesis

- Las redes neuronales de aprendizaje profundo son rápidas y pueden producir una precisión similar a la humana sin la intervención del usuario.

Clasificación de imágenes

- Proceso de determinar un resultado discreto basado en características continuas dentro de los datos de entrada.

Incremento de datos

- Proceso de crear sintéticamente más datos de imágenes sobre la base de transformaciones como rotación, escala y recorte.

Segmentación de imágenes

- Método idóneo con alta precisión espacial donde se identifica el lugar de una clase específica dentro de una imagen. Es decir, es la capacidad de generar un mapa de píxeles a partir de datos de entrada.

Regresión de datos

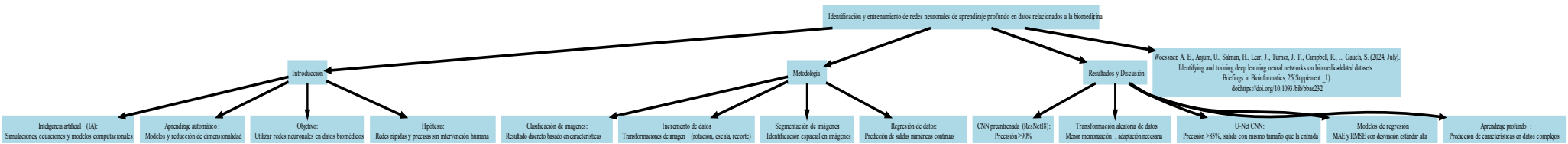
- Método de entrenamiento para predecir salidas numéricas continuas, como no ocurren en las Convolutional Neural Networks (CNNs o ConvNets) donde la salida para los píxeles es discreta.

Topic: 21

46%

ENG

10:45 PM



(c) Justifique las 4 palabras clave de su artículo.

1. Deep Learning

Deep learning como un instrumento de la IA ha revolucionado el análisis y la interpretación de datos médicos por su capacidad para el «aprendizaje».

2. Biomedical Images

Las imágenes biomédicas dotan una vista pormenorizada del interior del cuerpo, por ende, permite la detección temprana y el seguimiento de patologías. Además, con la ayuda de Deep learning y redes neurales convolucionales (Convolutional Neural Networks), sus algoritmos son «entrenados» para detectar automáticamente anomalías.

3. Neural Networks

Constituyen la base del deep learning, porque favorecen la creación de modelos con capacidad para «aprender» sobre la base de grandes cantidades de datos.

En particular, para imágenes biomédicas, las CNN (Convolutional Neural Networks), son eficientes para llevar a cabo tareas de clasificación, segmentación y detección de anomalías.

4. Algorithms

Son procedimientos que establecen cómo las redes neurales ajustan sus ponderaciones y aprenden de los datos.

**Pregunta 2: (6 puntos)**

(a) Usando SCOPUS (<https://www.scopus.com/>) obtenga un archivo CSV con al menos 10 artículos científicos para estas palabras clave. Asegúrese de disponer de la columna “Abstract”.

PALABRAS PARA LA BÚSQUEDA

```
TITLE-ABS-KEY(("deep") AND ("learning")) AND (("biomedical") AND ("images")) AND (("neural") AND ("networks")) AND ("algorithms")
```

(b) Usando matplotlib y/o seaborn grafique un diagrama de barras con los tipos de publicación versus la cantidad.

Material visto en clase:

[https://github.com/inefable12/consultas\\_scopus\\_api/blob/main/Consultas en SCOPUS API.ipynb](https://github.com/inefable12/consultas_scopus_api/blob/main/Consultas_en_SCOPUS_API.ipynb)

(c) Usando Missingno evalúe si existen datos faltantes.

Referencia: <https://github.com/ResidentMario/missingno>

URL para la pregunta 2:

1. [https://colab.research.google.com/github/EHN8829/DUPD\\_PARCIAL/blob/main/Pregunta2\\_Consultas en SCOPUS API.ipynb](https://colab.research.google.com/github/EHN8829/DUPD_PARCIAL/blob/main/Pregunta2_Consultas_en_SCOPUS_API.ipynb)

2. [https://github.com/EHN8829/DUPD\\_PARCIAL/blob/main/Pregunta2\\_Consultas en SCOPUS API.ipynb](https://github.com/EHN8829/DUPD_PARCIAL/blob/main/Pregunta2_Consultas_en_SCOPUS_API.ipynb)

**Pregunta 3: (8 puntos)**

(a) Elabore una lista de Python con 3 preguntas dirigidas al abstract.

a.1. ¿Cuáles son los principales temas abordados en los artículos sobre aprendizaje profundo y biomedicina?

- a.2. ¿Cuál es el propósito principal del módulo de aprendizaje descrito en el "NIGMS Sandbox for Cloud – based Learning"?
- a.3. ¿Qué desafíos enfrentan los investigadores en la interpretación de datos biomédicos y cómo el aprendizaje ayudaría a resolverlos desafíos?

- (b) Disponga de LLAMA y realice estas consultas almacenando las respuestas en archivos de texto o en el mismo DataFrame de pandas (según le convenga)
- (c) Asigne un indicador numérico (como por ejemplo un puntaje) para luego filtrar su dataset.
- (d) Guarde su DataFrame inicial y final en archivos en formato CSV y adjunte en su Github.

URL para la pregunta 3:

1. [https://colab.research.google.com/github/EHN8829/DUPD\\_PARCIAL/blob/main/Pregunta3\\_IA\\_Consultas\\_Publicaciones.ipynb](https://colab.research.google.com/github/EHN8829/DUPD_PARCIAL/blob/main/Pregunta3_IA_Consultas_Publicaciones.ipynb)
2. [https://github.com/EHN8829/DUPD\\_PARCIAL/blob/main/Pregunta3\\_IA\\_Consultas\\_Publicaciones.ipynb](https://github.com/EHN8829/DUPD_PARCIAL/blob/main/Pregunta3_IA_Consultas_Publicaciones.ipynb)