Hacia una IA en salud confiable y explicable: Innovaciones en análisis de datos e imagen médica

[Carlos Minutti-Martinez, Boris Escalante-Ramírez, Jimena Olveres]

Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación (INFOTEC)

Universidad Nacional Autónoma de México (UNAM)

2025

Tabla de Contenidos

Problema

1 Interpretabilidad con IA generativa

Desafíos en la comprensión de imágenes médicas

- Disponibilidad **limitada** de expertos **humanos**.
- Inconvenientes de **fatiga y estimación imprecisa** en el análisis manual.

Desafíos de las Redes Neuronales Convolucionales (CNN) en el análisis de imágenes médicas

- Los desafíos incluyen la **escasez de datos anotados**, **conjuntos de datos de imágenes médicas limitados**.
- Los modelos CNN grandes **carecen de explicabilidad**, una **característica crucial** para el análisis confiable de imágenes médicas.
- Grandes **recursos computacionales** para el entrenamiento y la inferencia.

Proyecto PumaMedNet

1 Interpretabilidad con IA generativa

- Tiene como objetivo diseñar una **arquitectura CNN para la clasificación de imágenes médicas** con **bajo costo computacional**.
- **Específico del dominio**: Entrenado con imágenes médicas para un Aprendizaje por Transferencia eficiente.
- Modelos **altamente explicables**, con capacidades de **mitigación de sesgos**.
- Autocodificador Variacional Denoising β-Variacional (**VAE**) como el esqueleto del modelo.
- Nuestra **versión inicial** se enfoca en imágenes de **Rayos X de tórax**, entrenada y validada en el conjunto de datos ChestX-ray14.
- Los nuevos modelos funcionan con imágenes de **Cerebro** e **imágenes de Mama**.

Arquitectura del modelo

1 Interpretabilidad con IA generativa

- Se eligió un **Autocodificador** como la estructura base para describir las **características visuales** de las imágenes (**Aprendizaje No Supervisado**).
- El Autocodificador permite generar un vector de variables latentes que **captura información esencial de la imagen**.
- La arquitectura VAE tiene una aproximación de **espacio latente continuo** a una distribución normal.
- Un β-VAE incorpora el hiperparámetro β, que apunta a una representación **desentrelazada** y controla las representaciones aprendidas en el espacio latente mediante una **penalización de la divergencia KL**.

Diagrama esquemático

1 Interpretabilidad con IA generativa

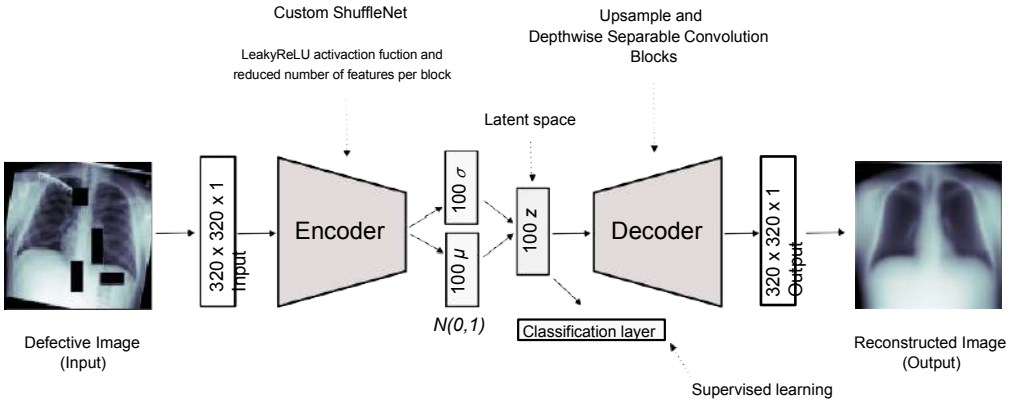
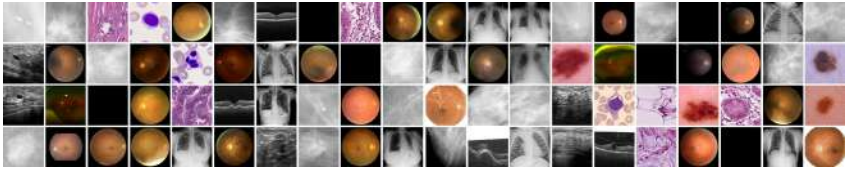


Figura: Diagrama esquemático de la arquitectura del modelo propuesto.

Pre-entrenamiento

1 Interpretabilidad con IA generativa

Utilizamos **pre-entrenamiento débilmente supervisado y no supervisado** para mejorar el rendimiento del reconocimiento de imágenes en un gran conjunto de metadatos de imágenes médicas, el conjunto de datos MiMeta. Compuesto por 17 conjuntos de datos públicos que abarcan 28 tareas y comprenden 372.895 imágenes, esta estrategia de pre-entrenamiento permite que el modelo capture características específicas del dominio y matices visuales inherentes a los datos de imágenes médicas.

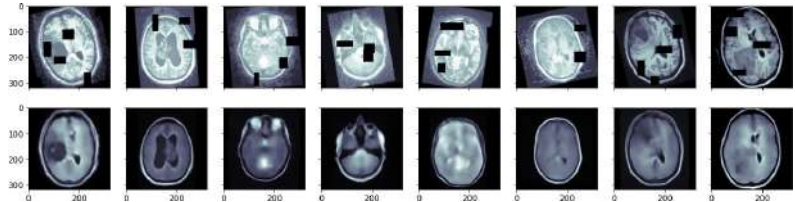


Modelado de contexto

1 Interpretabilidad con IA generativa

Utilizamos **principios de diseño inspirados en arquitecturas de transformadores** como contexto y atención.

Para capturar el contexto, utilizamos **aumento de datos extensivo**, incluyendo rotaciones aleatorias, volteos, desenfoque, transformaciones de perspectiva y **borrado aleatorio** (similar a las representaciones de palabras enmascaradas en los modelos de lenguaje), lo que permite que el decodificador **aprenda el contexto mediante la predicción de parches de imagen faltantes o defectuosos**.



Pesos de atención

1 Interpretabilidad con IA generativa

Pesos de atención para priorizar áreas de reconstrucción.

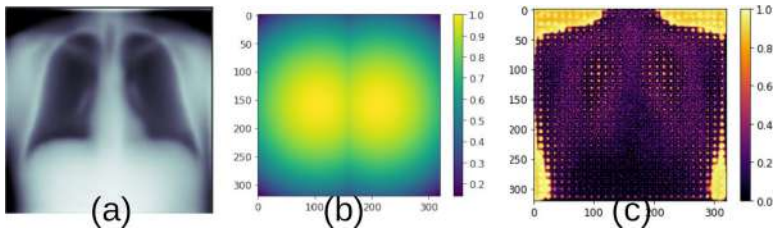
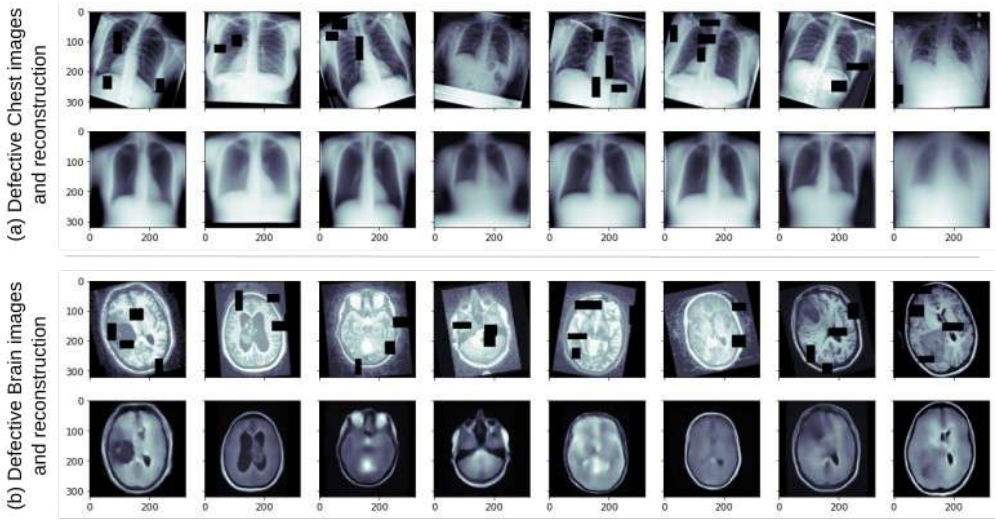


Figura: (a) Imagen de muestra. (b) Pesos de atención antes, (c) Pesos de atención ahora.

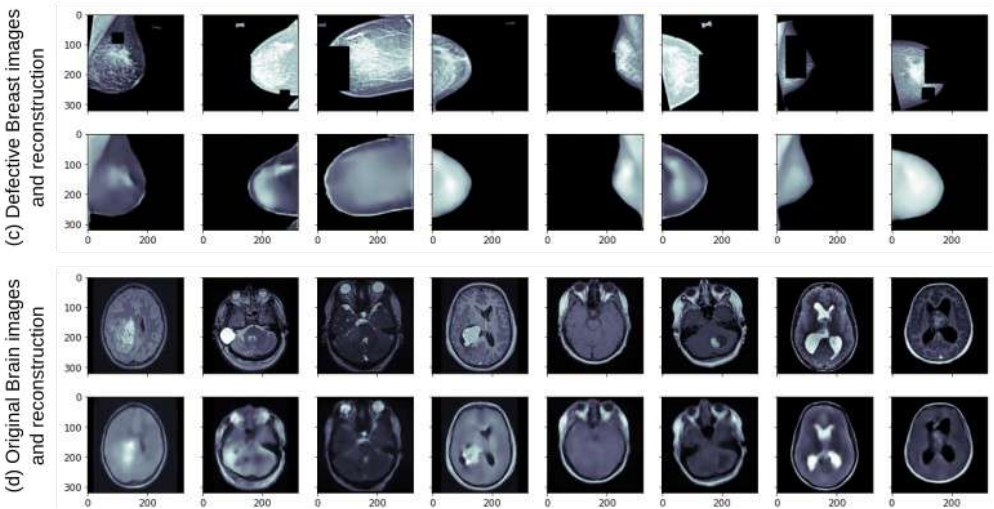
Reconstrucción

1 Interpretabilidad con IA generativa



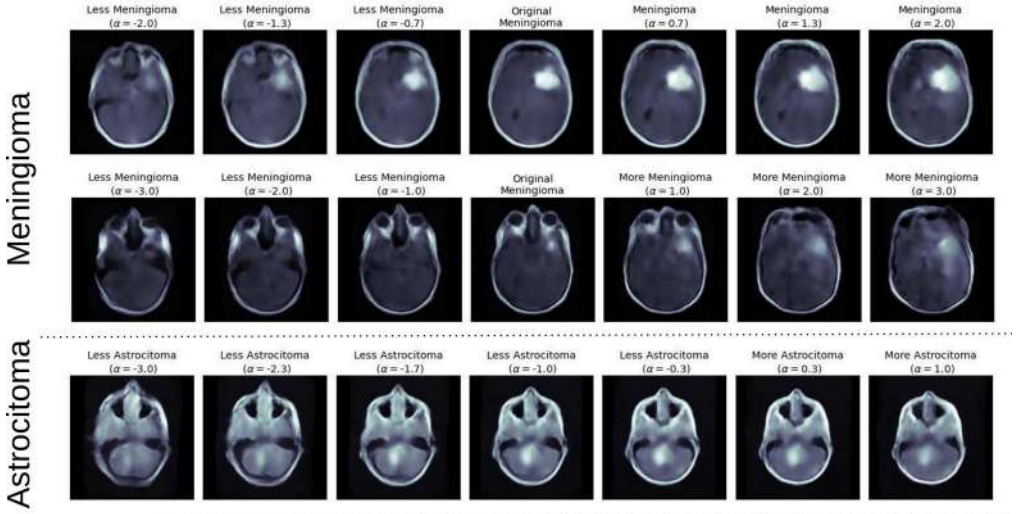
Reconstrucción

1 Interpretabilidad con IA generativa



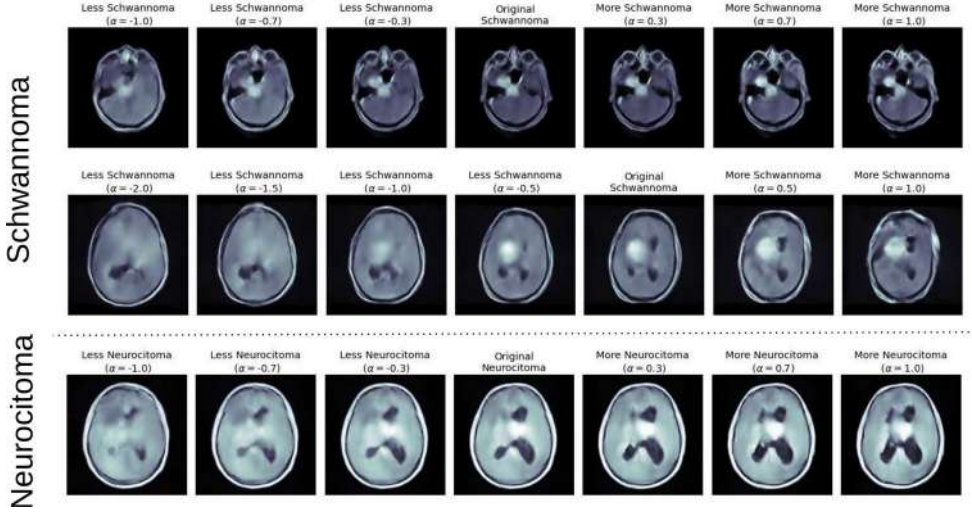
Explicabilidad

1 Interpretabilidad con IA generativa



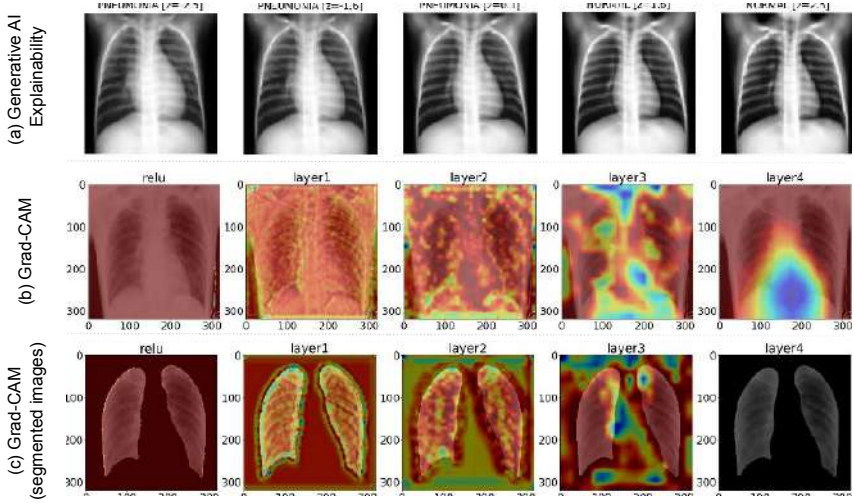
Explicabilidad

1 Interpretabilidad con IA generativa



Detección y mitigación de sesgos

1 Interpretabilidad con IA generativa



Resumen y Conclusiones

1 Interpretabilidad con IA generativa

Nuestro trabajo presenta un marco de IA generativa que aprovecha los autocodificadores para el análisis de imágenes médicas

- Características clave:
  - Mejora la interpretabilidad, el control de sesgos y la eficiencia de datos.
  - Permite la exploración intuitiva del proceso de toma de decisiones del modelo.
- Fortalezas:
  - Puede revelar sesgos que evaden la detección mediante técnicas convencionales como mapas de saliencia o visualizaciones basadas en gradientes.
  - Crucial para generar confianza y facilitar la adopción de herramientas clínicas impulsadas por IA.
- Objetivos futuros:
  - Escalar el marco para manejar diversas modalidades de imágenes médicas.
  - Tener diferentes tamaños de versiones del modelo.

Tabla de Contenidos

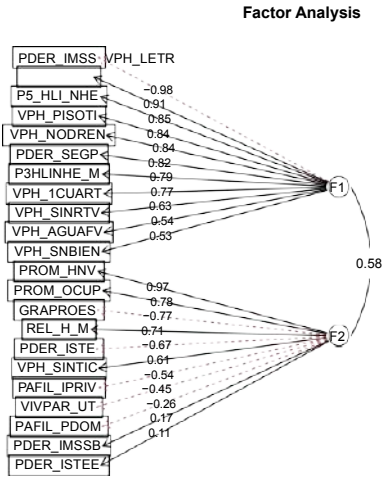
Problema

2 Análisis multifactorial con datos médicos

- Al estudiar enfermedades, existen múltiples factores que contribuyen a como se desarrolla la misma.
- Estudiar estos factores requiere de unificar múltiples fuentes de datos, por ejemplo, hospitalizaciones, censos de población, contaminantes atmosféricos, indices económicos, información geográfica.
- Es necesario el uso de modelos estadísticos para generar modelos explicables.
- También es necesario el uso de modelos de Machine-Learning para estudiar la compleja interacción entre variables y posibles efectos no lineales.

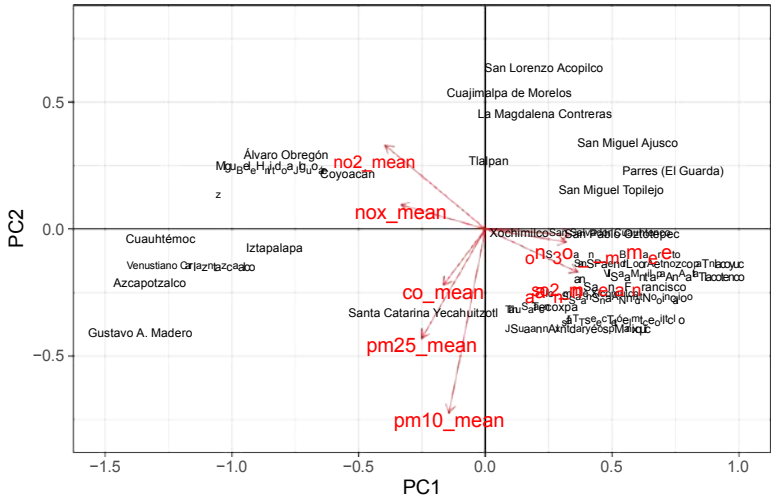
Indices socioeconómicos

2 Análisis multifactorial con datos médicos



Contaminantes atmosféricos

2 Análisis multifactorial con datos médicos



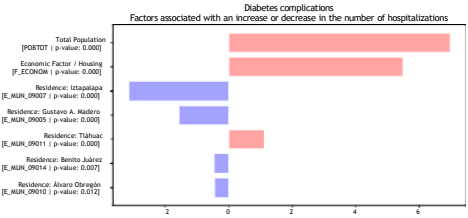
Diabetes mellitus

2 Análisis multifactorial con datos médicos

- Se mide el efecto de los factores socio-económicos y de contaminantes en el número y severidad de las hospitalizaciones.
- Se penaliza fuertemente el número de variables explicativas en el modelo para solo mantener las que muestran un efecto relevante y disminuir problemas de multicolinealidad.

Diabetes mellitus (Regresión)

2 Análisis multifactorial con datos médicos



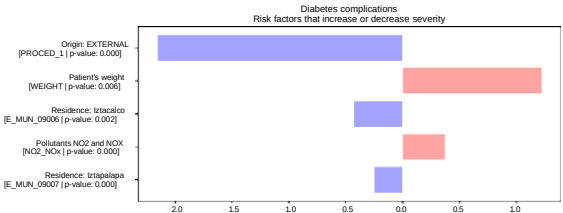


Figura: Factores relevantes en el número y gravedad de las hospitalizaciones por complicaciones de la Diabetes.

Diabetes mellitus (GBM)

2 Análisis multifactorial con datos médicos

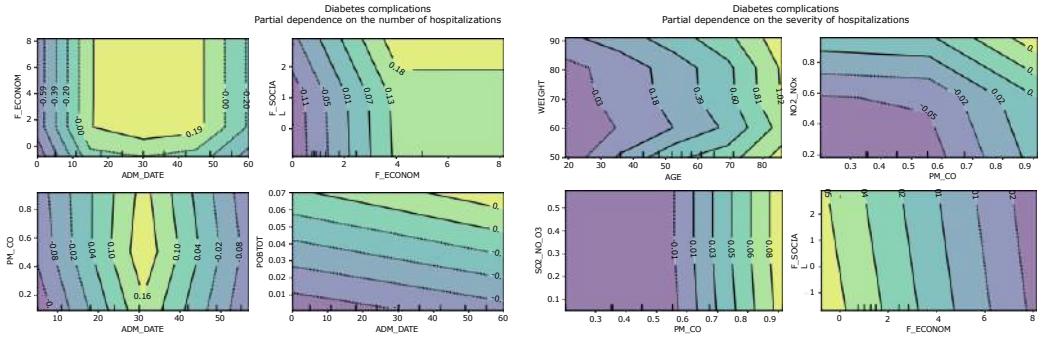


Figura: Dependencia parcial del número y gravedad de las hospitalizaciones por complicaciones de la Diabetes.

Q&A

Gracias por escucharnos.
Sus comentarios serán muy apreciados.