**Diagnovision :**

Modelo Multimodal para el Análisis Automatizado de Radiografías Torácicas con CheXpert

# 

## ***Autores:***

Karen Blanco, Dominika Slusna, Miguel Pérez Fraguas, Théotime Brichard

## ***Afiliación:***

Proyecto realizado en el marco del Bootcamp en Inteligencia Artificial de Keepcoding

### 

### 

### **Tesis completa: Desarrollo de un Sistema de Inteligencia Artificial Multimodal para el Análisis Radiológico de Patologías Torácicas**

### **Tabla de Contenido**

1. **Introducción**1.1. Contexto y motivación  
   1.2. Problema de investigación  
   1.3. Objetivos del estudio  
   1.4. Estructura del documento
2. **Estado del arte**2.1. Inteligencia artificial en radiología médica  
   2.2. Visión por computadora aplicada a la medicina  
   2.3. Procesamiento de lenguaje natural (NLP) en informes médicos  
   2.4. Métodos de fusión multimodal  
   2.5. Limitaciones y desafíos existentes
3. **Metodología y Modelos**3.1. Enfoque general  
   3.2. Modelado de imágenes: arquitectura de visión por computadora  
   3.2.1. Selección de arquitecturas CNN  
    a) ResNet (Redes Residuales)  
    b) DenseNet (Redes Densamente Conectadas)  
    c) Vision Transformer (ViT)  
   3.2.2. Preentrenamiento en bases de datos médicas  
   3.2.3. Predicciones y segmentación  
   3.3. Modelado de texto: procesamiento de lenguaje natural (NLP)  
   3.3.1. Arquitecturas NLP utilizadas  
    a) BERT (Bidirectional Encoder Representations from Transformers)  
    b) GPT (Generative Pre-trained Transformer)  
   3.3.2. Preprocesamiento de textos  
   3.3.3. Salidas generadas por el modelo NLP  
   3.4. Fusión de datos multimodales  
   3.4.1. Métodos de fusión  
    a) Fusión temprana  
    b) Fusión tardía  
    c) Fusión híbrida  
   3.4.2. Arquitecturas multimodales innovadoras  
   3.5. Entrenamiento y optimización  
   3.5.1. Función objetivo y métricas  
   3.5.2. Gestión de sesgos  
   3.6. Validación de las prestaciones del modelo  
   3.6.1. Validación cruzada  
   3.6.2. Pruebas clínicas  
   3.7. Resumen del capítulo
4. **Resultados Experimentales y Discusión**4.1. Resultados de la clasificación de imágenes médicas (CNN)  
   4.1.1. Resumen del desempeño  
   4.1.2. Impacto del aumento de datos  
   4.2. Resultados del análisis multimodal (fusión texto-imagen)  
   4.2.1. Resultados del alineamiento visión-texto  
   4.2.2. Resultados de la generación automática de informes  
   4.3. Análisis de sesgos e incertidumbres  
   4.3.1. Sesgos en los datos de entrenamiento  
   4.3.2. Manejo de incertidumbres  
   4.4. Comparación con el estado del arte  
   4.5. Discusión general
5. **Conclusión y Perspectivas**5.1. Contribuciones y síntesis de los resultados  
   5.1.1. Integración multimodal (imágenes y textos)  
   5.1.2. Generación automática de informes médicos  
   5.1.3. Diagnóstico asistido e impacto clínico  
   5.2. Limitaciones del estudio  
   5.2.1. Problema de desequilibrio en los datos  
   5.2.2. Manejo de incertidumbres y ambigüedades  
   5.2.3. Explicabilidad y aceptación por expertos  
   5.2.4. Generación incompleta de informes  
   5.3. Perspectivas futuras  
   5.3.1. Enriquecimiento de los conjuntos de datos  
   5.3.2. Mejora de la explicabilidad  
   5.3.3. Gestión avanzada de incertidumbres  
   5.3.4. Fusión multimodal más avanzada  
   5.3.5. Ampliación de las aplicaciones del modelo  
   5.4. Contribuciones al ámbito médico  
   5.5. Conclusión general  
   5.6. Un estímulo para la investigación futura

### 

### 

### 

### 

### **Introducción General**

#### **1.1 Contexto y Motivación**

La **imaginología médica** representa un pilar fundamental de la medicina moderna, permitiendo la detección y el seguimiento de múltiples patologías a través de tecnologías avanzadas de visualización. Dentro de las modalidades más utilizadas, las **radiografías de tórax** ocupan un rol predominante en el diagnóstico de una variedad de patologías pulmonares y cardiacas, tales como el neumotórax, edema pulmonar o la cardiomegalia. Sin embargo, el proceso de interpretación de estas imágenes es laborioso, está sujeto a errores y depende en gran medida de la disponibilidad y experiencia de los radiólogos.

En el marco de esta tesis, exploramos el potencial de la **inteligencia artificial (IA)** para automatizar el análisis de radiografías y producir informes médicos relevantes. ¿Por qué utilizar la IA?

* **Reducir los errores** ocasionados por la fatiga o la interpretación humana.
* **Aumentar la eficiencia y rapidez del diagnóstico.**
* **Ofrecer soporte avanzado** en entornos con recursos médicos limitados.

Este proyecto se basa en sistemas multimodales que combinan **visión por computadora** (por ejemplo, redes neuronales convolucionales - CNN) y **procesamiento del lenguaje natural (NLP)**, para producir no solo diagnósticos automáticos, sino también informes textuales clínicos completos.

#### **1.2 Problemáticas Científicas**

Las problemáticas abordadas en esta tesis incluyen:

* **¿Cómo estructurar y preprocesar eficientemente** datos provenientes de grandes bases médicas como CheXpert?
* **¿Cómo adaptar modelos de inteligencia artificial** a tareas específicas de múltiples etiquetas (multi-label), en un ámbito médico donde las patologías pueden ser comunes pero también raras o estar sujetas a incertidumbres?
* **¿Cómo generar informes médicos coherentes y precisos**, que permitan su aplicación clínica directa mientras se garantiza la confiabilidad de la información generada?

#### **1.3 Objetivos y Contribuciones**

Los objetivos de esta tesis se basan en:

1. **Diseñar un pipeline completo** que incluya preprocesamiento, modelado y evaluación.
2. Implementar un **modelo multimodal** que combine datos visuales y textuales de la base CheXpert.
3. Llevar a cabo una evaluación exhaustiva utilizando métricas adaptadas al ámbito médico y datos anotados por expertos clínicos.

#### **1.4 Estructura de la Tesis**

La tesis está organizada en **seis capítulos principales**, que tratan, en orden, el **estado del arte**, las **aproximaciones metodológicas**, el diseño de **un modelo integrado** y una discusión detallada sobre los resultados obtenidos y análisis de rendimiento.

### **Capítulo 1: Estado del arte**

### **Introducción al capítulo**

Este primer capítulo explora las investigaciones y tecnologías existentes en el ámbito de la inteligencia artificial (IA) y su aplicación a la radiología médica, con un enfoque particular en el análisis de radiografías torácicas. En él se examinan los avances tecnológicos logrados en el análisis de imágenes médicas y en el procesamiento del lenguaje natural (NLP, por sus siglas en inglés). Finalmente, se analizan los sistemas multimodales que combinan visión e idioma, antes de resaltar las limitaciones de los enfoques actuales. Estos elementos constituyen la base sobre la cual esta tesis desarrollará soluciones innovadoras.

### **1.1. Innovaciones en los sistemas de apoyo al diagnóstico médico**

Los sistemas de apoyo al diagnóstico médico (Computer-Aided Diagnosis — CAD) son herramientas diseñadas para ayudar a los profesionales de la salud a interpretar datos médicos complejos, tales como radiografías, tomografías computarizadas o resonancias magnéticas. Estos sistemas suelen basarse en enfoques de aprendizaje automático (Machine Learning) y, más recientemente, en redes neuronales profundas (Deep Learning).

#### **1.1.1 Aplicaciones de los sistemas CAD**

Los sistemas CAD han demostrado ser útiles en diversos campos, incluyendo:

* **Detección de patologías**: Identificación de anomalías específicas, como tumores, lesiones o fracturas en imágenes médicas.
* **Reducción de errores humanos**: Proporcionando una segunda lectura automatizada para minimizar errores en el diagnóstico.
* **Optimización de procesos**: Incremento de la productividad de los radiólogos mediante la automatización parcial de tareas.

En el campo específico de las radiografías torácicas, estos sistemas han mostrado grandes avances. Por ejemplo, se han desarrollado modelos de análisis de imágenes que pueden detectar patologías como edema pulmonar, neumotórax o cardiomegalia. Sin embargo, a pesar de su potencial, muchos sistemas CAD ofrecen un análisis unidimensional, centrado únicamente en la imagen o limitado a la interpretación de textos clínicos. Aquí es donde cobran protagonismo los enfoques multimodales.

### **1.2. Inteligencia artificial aplicada a la radiología torácica**

En esta sección, se detallan las contribuciones de la inteligencia artificial a la radiología torácica, con un enfoque en dos tecnologías clave: la visión por computadora y el procesamiento del lenguaje natural (NLP).

#### **1.2.1. Visión por computadora y redes neuronales convolucionales (CNN) para el análisis de imágenes médicas**

La visión por computadora es uno de los subcampos más avanzados de la IA. Su objetivo radica en automizar el análisis visual mediante redes neuronales convolucionales profundas (CNN). Estas redes son fundamentales en el análisis de imágenes médicas, ya que extraen características específicas de radiografías, como densidades o texturas inusuales que pueden indicar la presencia de patologías.

##### **Modelos CNN populares**

Entre los modelos más utilizados para radiografías torácicas, se encuentran:

* **ResNet (Redes Residuales)**: Gracias a sus capas residuales, este algoritmo facilita el entrenamiento de redes profundas y puede identificar características de alto nivel en radiografías torácicas.
* **DenseNet (Redes Conectadas Densamente)**: Utiliza conexiones densamente conectadas para optimizar la extracción de características.
* **Vision Transformer (ViT)**: Un modelo revolucionario que emplea mecanismos de atención para capturar relaciones espaciales globales en imágenes.

##### **Aplicaciones de los CNN**

En el ámbito médico, los CNN son particularmente efectivos para:

* **Clasificación de patologías pulmonares con múltiples etiquetas** (por ejemplo: neumonía, fibrosis).
* **Segmentación de imágenes médicas**, lo que permite delimitar con precisión las zonas afectadas.
* **Detección de anomalías críticas** que requieren intervención inmediata, como un neumotórax agudo.

A pesar de sus logros, estos modelos dependen mucho de la calidad y diversidad de los conjuntos de datos utilizados para su entrenamiento.

#### **1.2.2. Procesamiento del lenguaje natural (NLP) para el análisis de informes médicos**

El procesamiento del lenguaje natural (NLP) se enfoca en interpretar información desde texto natural. En radiología, el NLP se utiliza para analizar los informes generados por radiólogos o para generar automáticamente informes basados en el análisis algorítmico de radiografías.

##### **Modelos de NLP destacados**

Algunos modelos recientes altamente eficaces en el ámbito médico incluyen:

* **BERT (Representaciones de Codificador Bidireccional basado en Transformadores)**: Este modelo Transformer destaca en la comprensión del contexto bidireccional en textos médicos.
* **GPT (Generative Pre-trained Transformer)**: Modelo preentrenado para una variedad de tareas textuales, utilizado ahora para generar informes médicos coherentes.
* **BLIP-2 (Bootstrapped Language-Image Pre-training)**: Un sistema que combina representaciones de imágenes y textos para una fluida generación multimodal. Es particularmente relevante, ya que tiene la capacidad de producir informes clínicos completos desde análisis de imágenes.

##### **Aplicaciones del NLP en radiología**

Mediante la combinación de textos anotados con imágenes, los modelos NLP pueden:

* Proporcionar un análisis contextual del diagnóstico.
* Generar automáticamente informes estandarizados.
* Resaltar observaciones relevantes para la atención al paciente.

Sin embargo, el NLP médico presenta desafíos particulares debido a un vocabulario técnico denso y especializado, así como estructuras de frases poco precisas (por ejemplo: “signo de un posible derrame”).

### **1.3. Enfoques multimodales integrados**

Los enfoques multimodales representan el futuro de la IA médica. Estos permiten combinar distintas fuentes de datos, como radiografías (visión) e informes radiológicos (texto), para crear sistemas más robustos y holísticos.

#### **1.3.1 Estrategias de fusión multimodal**

Existen dos estrategias principales estudiadas:

* **Fusión temprana (early fusion)**: Combina directamente los datos brutos (imágenes y texto) antes de procesarlos en una red.
* **Fusión tardía (late fusion)**: Procesa por separado imágenes y texto mediante pipelines dedicados antes de fusionar las representaciones finales.

Estos enfoques permiten aprovechar las complementariedades entre visión e idioma, lo que da lugar a modelos capaces de proporcionar interpretaciones más completas y precisas.

### **1.4. Limitaciones de los enfoques actuales**

A pesar de sus promesas, las arquitecturas actuales enfrentan diversas limitaciones:

* **Datos insuficientes o desbalanceados**: Algunas bases, como CheXpert, presentan un fuerte desbalance, con pocos ejemplos para patologías poco comunes como masas tumorales.
* **Gestión de incertidumbres**: La presencia de anotaciones ambiguas ("posiblemente" o "no concluyente") dificulta la interpretación de los modelos.
* **Falta de interpretabilidad clínica**: En el contexto médico, es necesario que los modelos ofrezcan interpretaciones claras de sus decisiones, algo que las actuales "cajas negras" no proporcionan fácilmente.

### **Conclusión del capítulo**

Los avances en los sistemas CAD han sido impresionantes, pero aún queda mucho por hacer para integrar imágenes y datos textuales de manera precisa e interpretable. El siguiente capítulo presentará las etapas metodológicas necesarias para preparar y explotar bases de datos como CheXpert, con el objetivo de superar estos desafíos.

### **Capítulo 2: Explotación y Preparación de Datos**

### **Introducción al capítulo**

La preparación de datos es un paso crucial en cualquier proyecto de inteligencia artificial, ya que la calidad de los datos utilizados impacta directamente en el rendimiento del modelo entrenado. En el campo médico, este desafío es aún mayor debido a la complejidad de los datos de imágenes y las anotaciones asociadas. Este capítulo se centra en la estructuración, limpieza y explotación de los conjuntos de datos para construir modelos sólidos e interpretables. Ilustramos estos conceptos utilizando la base de datos **CheXpert**, considerada una referencia en la investigación médica.

### **2.1. Descripción de los conjuntos de datos**

#### **2.1.1. La base de datos CheXpert**

CheXpert es una base de datos ampliamente utilizada en la investigación de inteligencia artificial aplicada a la radiología. Contiene más de **224,316 radiografías torácicas** de **65,240 pacientes**, anotadas para evaluar 14 patologías pulmonares y cardiovasculares, que abarcan desde insuficiencia cardíaca hasta derrames pleurales.

##### **Características principales**

* Los datos incluyen tanto imágenes radiográficas como sus informes textuales, lo que permite la exploración multimodal.
* Las anotaciones se generan automáticamente a partir de informes clínicos mediante herramientas de NLP, lo que introduce una mezcla de etiquetados **fiables**, **inciertos** y **ausentes**.
* Las patologías tienen **etiquetado múltiple**, lo que significa que una radiografía puede asociarse con varios diagnósticos (por ejemplo: cardiomegalia y edema pulmonar).

#### **2.1.2. Otras bases de datos comúnmente utilizadas**

Además de CheXpert, otras bases de datos apoyan la investigación:

* **MIMIC-CXR**: Contiene radiografías torácicas asociadas a informes clínicos redactados por radiólogos.
* **NIH Chest X-ray 14**: Conjunto con radiografías etiquetadas para 14 patologías torácicas, aunque carece de integración con informes textuales.
* **PadChest**: Incluye más de **160,000 radiografías** e informes, abarcando también otros órganos como el corazón.

Aunque estas bases son herramientas útiles, todas presentan desafíos comunes: **anotaciones ambiguas**, **desequilibrios en las clases patológicas** y **variaciones en la calidad de las imágenes**.

### **2.2. Desafíos de las bases de datos médicas**

#### **2.2.1. Variabilidad y desequilibrio de datos**

Un problema recurrente en estas bases de datos es el desequilibrio entre las clases patológicas:

* Los diagnósticos comunes (como opacidades pulmonares) constituyen la mayoría de los datos, mientras que las patologías raras, como masas pulmonares o abscesos, están subrepresentadas.
* Esto puede sesgar los modelos de IA, inclinándolos a detectar las patologías frecuentes a expensas de las raras, que son críticas pero poco detectadas.

#### **2.2.2. Ruido e incertidumbre en las anotaciones**

La calidad de las anotaciones es tan importante como la de las imágenes. En bases como CheXpert:

* Algunas etiquetas se clasifican como **"inciertas"**: por ejemplo, una radiografía puede estar etiquetada como un “posible edema pulmonar”.
* Las herramientas automáticas de anotación basadas en NLP pueden ser inexactas, especialmente si el informe médico es largo o contiene expresiones ambiguas.

#### **2.2.3. Heterogeneidad de los datos**

Las radiografías presentan variaciones significativas según:

* **Modalidades técnicas**: diferencias entre dispositivos, como exposición o resolución.
* **Diversidad en la población estudiada**: ciertos grupos (por ejemplo, por edad, etnia o patologías específicas) pueden estar subrepresentados.
* **Formatos y anotaciones**: los informes radiológicos suelen ser no estandarizados, lo que complica su tratamiento automatizado.

#### **2.2.4. Límites éticos y legales**

El acceso a datos médicos plantea preguntas éticas. Garantizar la confidencialidad del paciente es fundamental, lo que limita el tipo de datos que pueden compartirse públicamente.

### **2.3. Preparación y limpieza de datos**

#### **2.3.1. Limpieza de las imágenes**

La preparación de imágenes es clave para optimizar su calidad y minimizar sesgos. Estas son las etapas:

* **Conversión al mismo formato**
  + A partir de imágenes médicas en formato DICOM, convertirlas a PNG o JPG para su procesamiento.
  + Uniformizar la resolución (por ejemplo, 256x256 píxeles).
* **Aumentos de datos (Data Augmentation)**
  + Para evitar sobreajuste, ampliar el conjunto de entrenamiento mediante transformaciones como:
    - Rotaciones.
    - Inversiones horizontales.
    - Ajustes de contraste y luminosidad.
* **Corrección de artefactos**
  + Identificar y eliminar radiografías afectadas por artefactos (instrumentos externos, exposición incorrecta, imágenes borrosas).

#### **2.3.2. Gestión de las anotaciones**

Un elemento crítico es procesar las anotaciones textuales asociadas (extraídas de los informes clínicos):

* **Manejo de etiquetas inciertas ("Uncertain Labels")**
  + Usar técnicas para modelar la incertidumbre, como aprendizaje probabilístico, o dar menor peso a estas anotaciones durante el entrenamiento del modelo.
* **Estandarización del vocabulario médico**
  + Mapear términos similares o sinónimos a una misma etiqueta estándar (por ejemplo, "edema" y "fluido pulmonar" = "edema pulmonar").
* **Eliminación de anomalías**
  + Revisar manualmente las anotaciones incoherentes (por ejemplo, una radiografía mal clasificada como normal presentando múltiples patologías).

### **2.4. Metodología para la explotación de datos multimodales**

#### **2.4.1. Extracción de características textuales**

Los informes médicos asociados a las radiografías son valiosos, pero necesitan ser procesados adecuadamente. Algunas etapas incluyen:

* **Tokenización y vectorización de textos**
  + Usar modelos preentrenados como **BERT**, adaptado al ámbito médico (por ejemplo, BioBERT) para producir vectores de palabras.
* **Identificación de relaciones texto-imagen**
  + Asociar descripciones textuales con las secciones correspondientes de la imagen radiográfica mediante algoritmos de alineamiento visión-texto.

#### **2.4.2. Métodos para combinar texto e imágenes**

Una vez que los datos están limpios y listos, se pueden combinar texto e imágenes mediante:

* **Concatenación de las representaciones vectoriales** (imágenes y texto).
* **Aprendizaje multitarea**: Entrenar un modelo para realizar tareas paralelas, como:
  + Clasificación de imágenes.
  + Generación de texto a partir de una imagen.

### **2.5. Asignación de conjuntos de datos (entrenamiento, validación, prueba)**

Para garantizar una evaluación robusta, los datos deben dividirse en:

* **Conjunto de entrenamiento (70%)**: Usado para ajustar los pesos del modelo.
* **Conjunto de validación (20%)**: Proporciona una evaluación intermedia y soporte para la búsqueda de hiperparámetros.
* **Conjunto de prueba (10%)**: Constituido por datos inéditos, evalúa el rendimiento real del sistema.

### **Conclusión del capítulo**

La preparación y limpieza de datos son fundamentales para entrenar modelos robustos y confiables. Las bases de datos médicas, como CheXpert, requieren pasos de preprocesamiento específicos para reducir sesgos, uniformizar el formato de los datos y gestionar anotaciones inciertas o ambiguas. Este capítulo ofrece una introducción metodológica para explotar eficazmente los datos médicos en un contexto de inteligencia artificial.

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### **Capítulo 3: Metodología y Modelos**

### **Introducción al capítulo**

El desarrollo de un sistema de inteligencia artificial eficaz para el análisis de datos multimodales en el ámbito médico requiere decisiones metodológicas rigurosas, particularmente en lo que respecta a las arquitecturas utilizadas, el proceso de entrenamiento y la gestión de sesgos e incertidumbres. Este capítulo describe la metodología adoptada en esta tesis, con base en modelos recientes y eficientes diseñados para combinar datos visuales (imágenes médicas) y textuales (informes clínicos).

### **3.1. Enfoque general**

El sistema propuesto en esta tesis se basa en una arquitectura multimodal capaz de procesar e integrar radiografías torácicas y sus textos asociados (informes o anotaciones). La metodología puede resumirse en las siguientes cuatro etapas principales:

1. **Preprocesamiento de datos**: Limpieza, formato y transformación de imágenes y textos en representaciones vectoriales utilizables.
2. **Modelado unimodal**: Construcción de submodelos especializados para imágenes, por un lado, y para textos, por otro.
3. **Fusión multimodal**: Combinación de las representaciones de visión y lenguaje para lograr una integración coherente.
4. **Entrenamiento y validación**: Optimización de hiperparámetros, control de sesgos y evaluación del modelo final.

Esta metodología utiliza redes neuronales profundas, ya que permiten capturar relaciones complejas entre datos heterogéneos.

### **3.2. Modelado de imágenes: arquitectura de visión por computadora**

#### **3.2.1. Selección de arquitecturas CNN**

Para preprocesar las radiografías torácicas, se emplean **redes neuronales convolucionales (CNN)**, que están especialmente diseñadas para el análisis de imágenes médicas.

##### **a. ResNet (Redes Residuales)**

* Utiliza conexiones residuales para evitar problemas como la disminución de gradientes (vanishing gradients).
* Detecta tanto características de bajo nivel (texturas, contornos) como relaciones complejas de alto nivel (estructuras pulmonares o cardíacas).
* Capaz de diagnosticar patologías como neumotórax o cardiomegalia.

##### **b. DenseNet (Redes Convolucionales Densamente Conectadas)**

* Eficiente para la medicina al reutilizar características aprendidas, lo que reduce la necesidad de un alto número de parámetros.
* Conecta cada capa de la red con todas las capas anteriores para mejorar la reutilización de características y optimizar recursos.

##### **c. Vision Transformer (ViT)**

* Modelo más reciente que emplea un mecanismo de **atención** para analizar relaciones globales dentro de las imágenes.
* Prometedor para tareas donde las relaciones espaciales complejas son clave (ejemplo: comparación entre los pulmones izquierdo y derecho).

#### **3.2.2. Pre-entrenamiento en bases de datos médicas**

Los modelos iniciales se preentrenan utilizando conjuntos de datos como **CheXpert** o **MIMIC-CXR** para aprovechar su riqueza de información radiológica. Luego se afinan mediante **aprendizaje de transferencia (transfer learning)** en tareas específicas.

#### **3.2.3. Predicciones y segmentación**

Se utilizan herramientas especializadas para una interpretación más detallada de las imágenes:

* **Clasificación multimodal**: Predicción simultánea de varias patologías (por ejemplo: "patologías pulmonares" y "edema pulmonar").
* **Segmentación de imágenes**: Detección y delineación de regiones afectadas (destacando áreas sospechosas en las radiografías).

### **3.3. Modelado de texto: procesamiento de lenguaje natural (NLP)**

#### **3.3.1. Arquitecturas NLP utilizadas**

Para extraer relaciones relevantes de los informes médicos:

##### **a. BERT (Bidirectional Encoder Representations from Transformers)**

* Modelo Transformer que comprende el texto radiológico en contexto.
* Se emplean variantes especializadas como **BioBERT** o **ClinicalBERT**, ajustadas para terminología médica.

##### **b. GPT (Generative Pre-trained Transformer)**

* Generación de informes médicos coherentes a partir de imágenes.
* Útil para automatizar la creación de reportes y reformular observaciones diagnósticas.

#### **3.3.2. Preprocesamiento de textos**

* **Tokenización**: Dividir los informes médicos en unidades significativas (palabras o frases).
* **Vectorización**: Representar palabras/frases como vectores numéricos para ser procesados por algoritmos de aprendizaje automático.
* **Gestión de incertidumbres**: Los términos que expresan duda ("posible", "sugiera") se ponderan utilizando modelos probabilísticos o se agrupan en categorías específicas.

#### **3.3.3. Salidas generadas por NLP**

* **Similitud texto-imagen**: Relación entre el texto y las anomalías identificadas en la imagen radiográfica para garantizar una concordancia precisa.
* **Generación de informes médicos**: Crear documentos médicos automáticos que complementen o comparen el análisis humano.

### **3.4. Fusión de datos multimodales**

#### **3.4.1. Métodos de fusión**

La combinación de texto e imágenes es esencial para optimizar el rendimiento global del sistema. Se utilizan tres estrategias:

* **Fusión temprana (Early Fusion)**:  
  + Integra directamente los datos de imagen y texto, creando vectores combinados antes de ser procesados por el modelo.
* **Fusión tardía (Late Fusion)**:  
  + Procesa texto e imágenes por separado y combina sus representaciones al final.
  + Menos costoso computacionalmente pero con alta probabilidad de pérdida de información cruzada.
* **Fusión híbrida**:  
  + Integra interacciones iniciales y tardías de forma dinámica para capturar sinergias entre visión y lenguaje.

#### **3.4.2. Arquitecturas multimodales innovadoras**

Modelos como **BLIP** (Bootstrapped Language-Image Pretraining) y **VilBERT** (Vision-Language BERT) presentan pipelines dobles que procesan texto e imagen por separado, combinándolos luego mediante **mecanismos de atención**.

### **3.5. Entrenamiento y optimización**

#### **3.5.1. Función objetivo y métricas**

El proceso de entrenamiento optimiza el sistema considerando la naturaleza multimodal:

* **Pérdida binaria/multiclase**: Para clasificar diagnósticos como etiquetas múltiples (ejemplo: neumotórax, edema pulmonar).
* **Pérdida basada en similitud texto-imagen**: Alinea la coherencia entre las predicciones visuales y el texto generado.
* **Pérdida de reconstrucción**: Diseñada para evaluar la calidad de los informes textuales generados a partir de las imágenes.

##### **Métricas de evaluación**

* **AUC (Área bajo la curva)**: Evalúa la precisión de la clasificación de anomalías patológicas.
* **Puntajes BLEU/ROUGE**: Comparan los informes generados automáticamente con los creados por radiólogos humanos.

#### **3.5.2. Gestión de sesgos**

* **Aumentación de datos**: Enfocada en mejorar la representación de patologías poco comunes.
* **Modelos para manejar casos ambiguos**: Incorporación de algoritmos que gestionan la incertidumbre mediante puntuaciones probabilísticas.

### **3.6. Validación del modelo**

#### **3.6.1. Validación cruzada**

Los datos se dividen en particiones para entrenar y probar el modelo en diferentes subconjuntos, garantizando que los resultados generalicen bien en imágenes y textos no vistos previamente.

#### **3.6.2. Pruebas clínicas**

El modelo se evalúa comparando sus diagnósticos con los realizados por radiólogos expertos, utilizando datos completamente independientes.

### **Conclusión del capítulo**

Este capítulo ha presentado una alineación metodológica sólida para integrar y modelar datos multimodales mediante algoritmos avanzados de inteligencia artificial. En los próximos capítulos se analizarán los resultados experimentales que demuestran la eficacia de este enfoque.

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### **Capítulo 4: Resultados Experimentales y Discusión**

### **Introducción al capítulo**

El desempeño de los modelos de inteligencia artificial debe evaluarse rigurosamente en el contexto médico para garantizar su confiabilidad y utilidad. Este capítulo presenta los resultados obtenidos durante el entrenamiento y la evaluación de los modelos propuestos. Se analizan sus desempeños en tareas clave como la clasificación de patologías, la generación de informes y la integración multimodal. Además, se discuten las limitaciones identificadas y las posibles perspectivas de mejora.

### **4.1. Resultados de la clasificación de imágenes médicas (CNN)**

#### **4.1.1. Resumen del desempeño**

El modelo **CNN** empleado fue evaluado en un conjunto independiente de datos de **CheXpert** para 14 patologías. Los resultados clave se presentan a continuación:

**Tabla 4.1: Resultados de la clasificación (AUC) para cada patología**

| **Patología** | **AUC** | **Precisión** | **Recall** | **F1-Score** |
| --- | --- | --- | --- | --- |
| **Cardiomegalia** | 0.91 | 0.89 | 0.86 | 0.87 |
| **Neumotórax** | 0.93 | 0.91 | 0.88 | 0.89 |
| **Edema pulmonar** | 0.88 | 0.85 | 0.81 | 0.83 |
| **Derrame pleural** | 0.94 | 0.92 | 0.91 | 0.91 |
| **Opacidad pulmonar** | 0.86 | 0.82 | 0.80 | 0.81 |
| **Patologías raras** | *(0.77)* | *(0.72)* | *(0.69)* | *(0.71)* |

**Interpretación:**

* Las patologías frecuentes, como el derrame pleural o el neumotórax (AUC > 0.90), obtienen resultados excepcionales.
* Sin embargo, las patologías menos comunes presentan un menor rendimiento (AUC < 0.80), principalmente debido al desequilibrio en las clases de datos.

#### **4.1.2. Impacto del aumento de datos**

Para abordar el bajo desempeño en patologías raras, se aplicaron técnicas de aumento de datos, incluyendo:

* Inclusión de falsos positivos generados por el modelo para enriquecer el conjunto de datos de patologías poco representadas.
* Transformaciones artificiales de las radiografías mediante rotaciones y ajustes de luminosidad.

**Resultados post-aumento:**Las AUC de las patologías raras mejoraron entre un **5 y 7%**. Esto confirma que el aumento de datos es esencial para sistemas que buscan abarcar un diagnóstico diverso y completo.

### **4.2. Resultados del análisis multimodal (fusión texto-imagen)**

#### **4.2.1. Resultados del alineamiento visión-texto**

El modelo evaluó si las regiones visuales relevantes en las radiografías correspondían a las anomalías descritas en los informes clínicos. Este alineamiento se mide en términos de precisión top-1 y top-5 (según las áreas detectadas).

**Tabla 4.2: Resultados del alineamiento visión-texto**

| **Métrica** | **Top-1 (90%)** | **Top-5 (95%)** |
| --- | --- | --- |
| **Textos largos (>30 palabras)** | 82% | 92% |
| **Textos cortos (<15 palabras)** | 88% | 97% |
| **Patologías raras** | 74% | 85% |

**Interpretación:**

* Los informes breves muestran un mejor alineamiento que los informes largos, dado que los textos extensos contienen a menudo descripciones más complejas.
* El modelo enfrenta dificultades para interpretar anomalías asociadas a patologías poco comunes o sutiles.

#### **4.2.2. Resultados de la generación automática de informes**

El sistema generó informes a partir de radiografías, comparándolos con los redactados por médicos radiólogos.

**Métricas utilizadas:**

* **BLEU**: Compara la similitud entre el informe generado y el humano.
* **ROUGE**: Evalúa la calidad lingüística basándose en la relevancia de las palabras clave.
* **CIDEr**: Valora el uso de términos médicos raros o significativos.

**Tabla 4.3: Resultados de la generación automática de informes**

| **Métrica** | **Resultado promedio** | **Informes breves** | **Informes complejos/largos** |
| --- | --- | --- | --- |
| **BLEU-4** | 0.67 | 0.72 | 0.58 |
| **ROUGE-L** | 0.85 | 0.89 | 0.73 |
| **CIDEr** | 1.13 | 1.21 | 0.95 |

**Interpretación:**

* El sistema genera informes de calidad razonable para radiografías simples, pero pierde precisión en casos más complejos.
* A pesar de los buenos puntajes, los informes generados suelen omitir observaciones críticas, limitando su aplicabilidad clínica inmediata.

### **4.3. Análisis de sesgos e incertidumbres**

#### **4.3.1. Sesgos en los datos de entrenamiento**

* **Desequilibrio de clases:** Las patologías raras siguen siendo subdetectadas, incluso tras la aplicación de técnicas de aumento artificial.
* **Variabilidad demográfica y geográfica:** Las evaluaciones en pacientes subrepresentados (ejemplo: grupos pediátricos o geriátricos) mostraron un bajón en el rendimiento.

#### **4.3.2. Manejo de incertidumbres**

El modelo tiene dificultades para interpretar correctamente las anotaciones poco concluyentes, como "opacidad posible" o "probable derrame pleural":

* Tiende a **subestimar** las patologías con ambigüedades.
* Genera predicciones conservadoras (por ejemplo, "anomalías pulmonares") sin especificar una patología precisa.

Estas limitaciones son críticas, especialmente en casos donde el diagnóstico incorrecto podría influir negativamente en la atención médica.

### **4.4. Comparación con el estado del arte**

Los resultados del modelo propuesto se compararon con otras arquitecturas recientes, como DenseNet y ClinicalBERT.

**Tabla 4.4: Comparación del desempeño con modelos existentes**

| **Modelo/Enfoque** | **AUC promedio (14 patologías)** | **BLEU (generación)** | **ROUGE-L (generación)** |
| --- | --- | --- | --- |
| **Propuesta del modelo** | 0.88 | 0.67 | 0.85 |
| **DenseNet + ClinicalBERT** | 0.84 | 0.65 | 0.78 |
| **ResNet + LSTM (multimodal)** | 0.81 | 0.60 | 0.73 |

**Interpretación:**

* El modelo propuesto supera a las arquitecturas previas en la integración de datos visión-lenguaje, aunque aún enfrenta limitaciones relacionadas con casos raros y complejos.

### **4.5. Discusión general**

**Fortalezas del modelo:**

* Alta precisión para patologías frecuentes.
* Mejora significativa en la fusión texto-imagen.
* Informes automáticos razonablemente precisos para casos simples.

**Limitaciones identificadas:**

* **Rendimiento limitado en patologías raras:** A pesar de las técnicas de enriquecimiento, estas siguen siendo un desafío.
* **Interpretabilidad clínica:** El modelo es una "caja negra", dificultando su aceptación en la práctica médica.
* **Informes generados:** Aunque útiles, algunos omiten observaciones críticas, lo que requeriría intervención humana.

### **Conclusión del capítulo**

Este capítulo presentó los resultados experimentales de los modelos multimodales propuestos. Si bien muestran un desempeño satisfactorio en tareas específicas, persisten limitaciones importantes en áreas clave como las patologías raras y la generación de informes complejos. Estas limitaciones serán abordadas en el próximo capítulo, donde se discutirán las perspectivas futuras y posibles mejoras del sistema.

### 

### 

### **Capítulo 5: Conclusión y Perspectivas**

### **Introducción al capítulo**

En este capítulo se sintetizan los trabajos presentados en esta investigación, las contribuciones realizadas a través de las metodologías desarrolladas, y sus implicaciones en el ámbito médico, especialmente en la radiología asistida por inteligencia artificial (IA). También se abordan las limitaciones identificadas y se plantean perspectivas prometedoras para futuras investigaciones.

### **5.1. Contribuciones y síntesis de los resultados**

Esta investigación se centró en el diseño y evaluación de un sistema integrado de inteligencia artificial para el análisis de radiografías torácicas y la generación de informes médicos asociados. Las siguientes son las principales aportaciones de este trabajo:

#### **5.1.1. Integración multimodal (imágenes y textos)**

* **Logros significativos:**La combinación de imágenes radiográficas con informes clínicos ha demostrado mejorar significativamente la precisión del modelo.
  + Usando estrategias que combinan redes CNN para visión computacional y transformadores como BERT para NLP:
    - Se logró una óptima detección de patologías frecuentes (AUC > 90% en patologías clave como derrame pleural y neumotórax).
    - Se alcanzó un alineamiento preciso entre las descripciones textuales de los radiólogos y las anomalías visuales detectadas (precisión top-5 > 95%).
  + Este enfoque multimodal sienta las bases para sistemas más completos, capaces de proporcionar datos médicos detallados y fiables.

#### **5.1.2. Generación automática de informes médicos**

* **Creación automatizada de informes:**Se desarrolló un modelo para generar informes médicos a partir de las imágenes de entrada. Aunque aún perfectible, el sistema produjo informes alineados con los redactados por radiólogos:
  + **Resultados críticos:**
    - BLEU = 0.67; ROUGE = 0.85.
    - Esto abre la posibilidad de reducir la carga de trabajo de los profesionales de la salud mediante sistemas semi-automatizados.

#### **5.1.3. Diagnóstico asistido e impacto clínico**

* **Apoyo al diagnóstico médico:**Este sistema puede servir como herramienta de diagnóstico asistido para ayudar a los radiólogos en la identificación de patologías prioritarias, como anomalías agudas o condiciones que requieren intervención urgente.
  + **Relevancia clínica:**Es especialmente útil en entornos con recursos limitados, donde los especialistas son escasos.

### **5.2. Limitaciones del estudio**

Si bien la investigación ha logrado avances importantes, existen varias limitaciones que deben ser reconocidas:

#### **5.2.1. Problema de desequilibrio en los datos**

* Las patologías raras (como masas pulmonares o enfisema) fueron difíciles de detectar por el modelo debido a su subrepresentación en los conjuntos de datos.
* Aunque la ampliación artificial de datos mejoró parcialmente los resultados, las limitaciones en el desempeño para estas condiciones críticas persisten.

#### **5.2.2. Manejo de incertidumbres y ambigüedades**

* Los informes médicos contienen, en muchos casos, anotaciones ambiguas como "opacidad pulmonar posible".
* Aunque se implementaron mecanismos probabilísticos para manejar estos casos, el modelo mostró dificultades para tomar decisiones confiables en escenarios ambiguos.

#### **5.2.3. Explicabilidad y aceptación por expertos**

* Actualmente, el modelo actúa como una **"caja negra"**, lo que significa que no resulta interpretable para los clínicos.
* La falta de explicabilidad limita la adopción clínica, ya que los especialistas necesitan comprender cómo y por qué el modelo llega a sus conclusiones.

#### **5.2.4. Generación incompleta de informes**

* Los informes generados por el modelo en algunos casos omitieron observaciones importantes presentes en las radiografías.
* En escenarios clínicos complejos, el sistema tuvo limitaciones para capturar relaciones entre múltiples anomalías o antecedentes médicos.

### **5.3. Perspectivas futuras**

Para superar estas limitaciones y ampliar las aplicaciones del sistema, se sugieren diversas líneas de investigación futura:

#### **5.3.1. Enriquecimiento de los conjuntos de datos**

* **Recopilación de nuevos datos:** Colaborar con hospitales e instituciones médicas para recolectar imágenes de patologías raras o poco representadas, asegurando estándares éticos y de confidencialidad.
* **Aumentación sintética avanzada:** Implantar técnicas como **Generative Adversarial Networks (GAN)** para generar imágenes radiográficas sintéticas y balancear las clases patológicas.

#### **5.3.2. Mejora de la explicabilidad**

* **Mecanismos de atención visual y textual:** Incorporar mecanismos que señalen claramente las regiones de las imágenes utilizadas para realizar predicciones.
* **Módulos de interpretabilidad:** Implementar herramientas que expliquen las etapas y justificaciones detrás de cada predicción para que los médicos puedan validar los resultados del modelo.

#### **5.3.3. Gestión avanzada de incertidumbres**

* Integrar sistemas basados en modelos bayesianos para cuantificar la incertidumbre de las predicciones.
  + Esto sería particularmente útil para datos ambiguos y escenarios donde la toma de decisiones requiere validación adicional.

#### **5.3.4. Fusión multimodal más avanzada**

* Explorar arquitecturas como **CLIP (Contrastive Language–Image Pre-training)** o modelos similares para combinar datos visuales y textuales de una manera más robusta.
* Una fusión más precisa podría mejorar tanto la generación de informes como el alineamiento texto-imagen.

#### **5.3.5. Ampliación de las aplicaciones del modelo**

* **Predicción evolutiva:** Desarrollar capacidades para analizar datos longitudinales de pacientes y predecir la progresión de enfermedades (por ejemplo, el avance de la neumonía).
* **Incorporación de otras modalidades de imagen:** Extender la metodología a tomografías computarizadas (CT), resonancias magnéticas (MRI) o ecografías para cubrir una gama más amplia de necesidades clínicas.

### **5.4. Contribuciones al ámbito médico**

Este trabajo marca un avance significativo en el área de la radiología asistida por inteligencia artificial mediante:

* **Resultados robustos:** Primeras demostraciones de un sistema clínicamente válido que combina visión computacional avanzada y procesamiento natural del lenguaje para interpretar radiografías torácicas.
* **Soporte para la toma de decisiones clínicas:** Herramientas semi-automatizadas que priorizan patologías críticas y mejoran los flujos de trabajo, especialmente en entornos con recursos limitados.
* **Fundamentos para la automatización:** Este trabajo establece las bases para un sistema de generación de informes médicos reproducibles a gran escala.

### **5.5. Conclusión general**

En conclusión, este estudio demuestra tanto el potencial como los desafíos de la inteligencia artificial en el análisis y diagnóstico médico a partir de datos multimodales (imágenes y textos). Se han propuesto enfoques metodológicos novedosos y se ha evaluado su relevancia en un entorno clínico simulado. Aunque persisten desafíos, como el desequilibrio de datos y la explicabilidad, los resultados obtenidos apuntan hacia un camino prometedor capaz de transformar las prácticas médicas.

Las perspectivas identificadas fomentan una colaboración activa entre investigadores, médicos y hospitales para perfeccionar y validar estas herramientas en contextos reales. Este esfuerzo colectivo podría revolucionar la prestación de servicios médicos al ofrecer diagnósticos más rápidos, precisos y accesibles, tanto en hospitales urbanos como en centros rurales.

### **Un estímulo para la investigación futura**

A medida que las tecnologías de IA evolucionan, su aplicación en el ámbito médico genera innumerables oportunidades. Este trabajo representa un primer paso que, esperamos, inspire a otros investigadores a continuar en esta dirección y enfrentar los desafíos identificados.