

# Clasificador de Radiografías de Tórax mediante Técnicas de Minería de Datos e Inteligencia Artificial

Juan David Mateus  
Facultad de Ingeniería  
Universidad Cooperativa de Colombia  
Villavicencio, Colombia  
juan.mateuss@campusucc.edu.co

Jhonnatan Mendez  
Facultad de Ingeniería  
Universidad Cooperativa de Colombia  
Villavicencio, Colombia  
jhonnatan.mendez@campusucc.edu.co

**Abstract**—La exploración y análisis de grandes volúmenes de datos médicos, específicamente radiografías de tórax, representa un desafío significativo en el campo de la minería de datos médicos. Este proyecto implementa técnicas avanzadas de minería de datos e inteligencia artificial para procesar y analizar radiografías de tórax, con el objetivo de detectar patrones y características asociadas a diversas enfermedades pulmonares. Utilizando técnicas de pre-procesamiento de imágenes, extracción de características y algoritmos de aprendizaje profundo, se desarrolla un sistema capaz de clasificar automáticamente diferentes patologías pulmonares, mejorando así la eficiencia y precisión del diagnóstico médico.

**Index Terms**—Minería de datos médicos, Procesamiento de imágenes, Deep Learning, Radiografías de tórax, Clasificación automática

## I. INTRODUCTION

En la era actual del Big Data médico, la minería de datos se ha convertido en una herramienta fundamental para el análisis y procesamiento de información médica. Las radiografías de tórax, como fuente rica en datos visuales, presentan un campo ideal para la aplicación de técnicas de minería de datos e inteligencia artificial. El proceso de extracción de conocimiento a partir de estas imágenes médicas involucra múltiples etapas, desde la recopilación y preprocesamiento de datos hasta la aplicación de algoritmos de aprendizaje profundo para la identificación de patrones patológicos.

Este proyecto implementa una metodología integral de minería de datos (KDD - Knowledge Discovery in Databases) aplicada al análisis de radiografías torácicas, combinando técnicas de procesamiento de imágenes, extracción de características y algoritmos de clasificación avanzados para desarrollar un sistema robusto de detección de enfermedades pulmonares.

## II. PLANTEAMIENTO DEL PROBLEMA

### A. Descripción del problema

El análisis de grandes volúmenes de radiografías de tórax presenta desafíos significativos en el contexto de la minería de datos médicos:

- Necesidad de procesar y analizar eficientemente grandes conjuntos de datos de imágenes médicas
- Complejidad en la extracción de características relevantes para el diagnóstico
- Variabilidad en la calidad y formato de las imágenes
- Requerimiento de técnicas avanzadas de preprocesamiento y normalización de datos
- Necesidad de algoritmos robustos para la clasificación automática de patologías

### B. Preguntas del problema

- 1) ¿Qué técnicas de minería de datos son más efectivas para el procesamiento y análisis de radiografías de tórax?
- 2) ¿Cómo se puede optimizar el proceso de extracción de características relevantes en las imágenes médicas?
- 3) ¿Qué algoritmos de clasificación proporcionan mejores resultados en la detección de patologías pulmonares?
- 4) ¿Cómo se puede validar la efectividad del modelo de minería de datos en el contexto médico?

### C. Objetivos

1) *Objetivo general:* Desarrollar un sistema de clasificación automática de radiografías de tórax mediante técnicas de minería de datos y aprendizaje profundo, implementando una metodología KDD para la detección efectiva de enfermedades pulmonares.

#### 2) *Objetivos específicos:*

- Implementar técnicas de preprocesamiento y normalización de imágenes médicas para optimizar la calidad de los datos
- Desarrollar algoritmos de extracción de características relevantes utilizando técnicas de minería de datos
- Diseñar e implementar un modelo de clasificación basado en redes neuronales convolucionales
- Evaluar y validar el rendimiento del sistema utilizando métricas estándar de minería de datos

### D. Justificación

La implementación de técnicas de minería de datos en el análisis de radiografías de tórax representa una solución

innovadora a los desafíos actuales en el diagnóstico médico. Este enfoque permite:

- Procesamiento eficiente de grandes volúmenes de datos médicos
- Extracción automática de características relevantes
- Mejora en la precisión del diagnóstico mediante algoritmos avanzados
- Reducción del tiempo de análisis y diagnóstico
- Apoyo objetivo en la toma de decisiones médicas

### III. MARCO DE REFERENCIA

#### A. Marco Teórico

El proyecto se fundamenta en los siguientes conceptos y técnicas de minería de datos:

- **Proceso KDD (Knowledge Discovery in Databases)**
  - Selección y preprocesamiento de datos
  - Transformación de datos
  - Minería de datos
  - Evaluación e interpretación de patrones
- **Técnicas de Preprocesamiento de Imágenes**
  - Normalización y estandarización
  - Reducción de ruido
  - Mejora de contraste
  - Segmentación de regiones de interés
- **Algoritmos de Aprendizaje Profundo**
  - Redes neuronales convolucionales (CNN)
  - Técnicas de transferencia de aprendizaje
  - Optimización de hiperparámetros

### IV. METODOLOGIA

El proyecto se desarrolló siguiendo la metodología ágil **Scrum**, permitiendo una gestión eficiente y flexible del proceso de desarrollo. Scrum se caracteriza por dividir el proyecto en ciclos cortos denominados *sprints*, facilitando la adaptación a cambios y la mejora continua.

#### A. Metodología Scrum

- **Roles**
  - Product Owner: Responsable de definir los requisitos del sistema
  - Scrum Master: Facilitador del proceso
  - Equipo de desarrollo: Implementación técnica
- **Eventos**
  - Sprint Planning: Planificación de tareas cada 2 semanas
  - Daily Scrum: Reuniones diarias de seguimiento
  - Sprint Review: Revisión de resultados
  - Sprint Retrospective: Mejora continua del proceso

#### B. Fases del Proyecto

##### 1) Fase de Preparación de Datos

- Recopilación de radiografías
- Preprocesamiento y normalización
- Etiquetado de datos

##### 2) Fase de Desarrollo del Modelo

- Implementación de algoritmos de extracción de características
- Desarrollo del modelo de clasificación
- Optimización y ajuste de parámetros

##### 3) Fase de Evaluación

- Validación del modelo
- Análisis de métricas de rendimiento
- Ajustes y mejoras iterativas

##### 4) Fase de Implementación

- Desarrollo de la interfaz web
- Integración del modelo
- Pruebas de usuario