

# **DIAGNÓSTICO AVANZADO DE IMÁGENES**

## **APLICANDO TÉCNICAS DE INTELIGENCIA**

### **ARTIFICIAL**

**Nombre de la materia**

Proyecto Integrador en Inteligencia Artificial

**Nombre del alumno**

Christian Garcia y Byron Piedra

**Nombre de la tarea**

Proyecto Integrador

**Nombre del Profesor**

Ing. Gladys Villegas

**Fecha**

11/12/2025

# Introducción

El proyecto desarrolla un sistema de apoyo al diagnóstico médico basado en inteligencia artificial para la detección temprana del cáncer de tiroides mediante ecografía. Se implementó un flujo completo que abarca preprocesamiento de imágenes, extracción de características, modelado y evaluación. El modelo final, un Random Forest optimizado, alcanzó un 71% de precisión y superó al baseline en todas las métricas clave. El sistema utiliza datos anónimos y públicos, asegurando criterios éticos y de privacidad.

# PROBLEMÁTICA

**Errores humanos en la interpretación de imágenes médicas**

**Diagnóstico tardío del cáncer de tiroides**

**Desbalance en la calidad y cantidad de datos clínicos**

**Limitada disponibilidad de especialistas en radiología**

**Falta de herramientas accesibles para análisis automatizado**

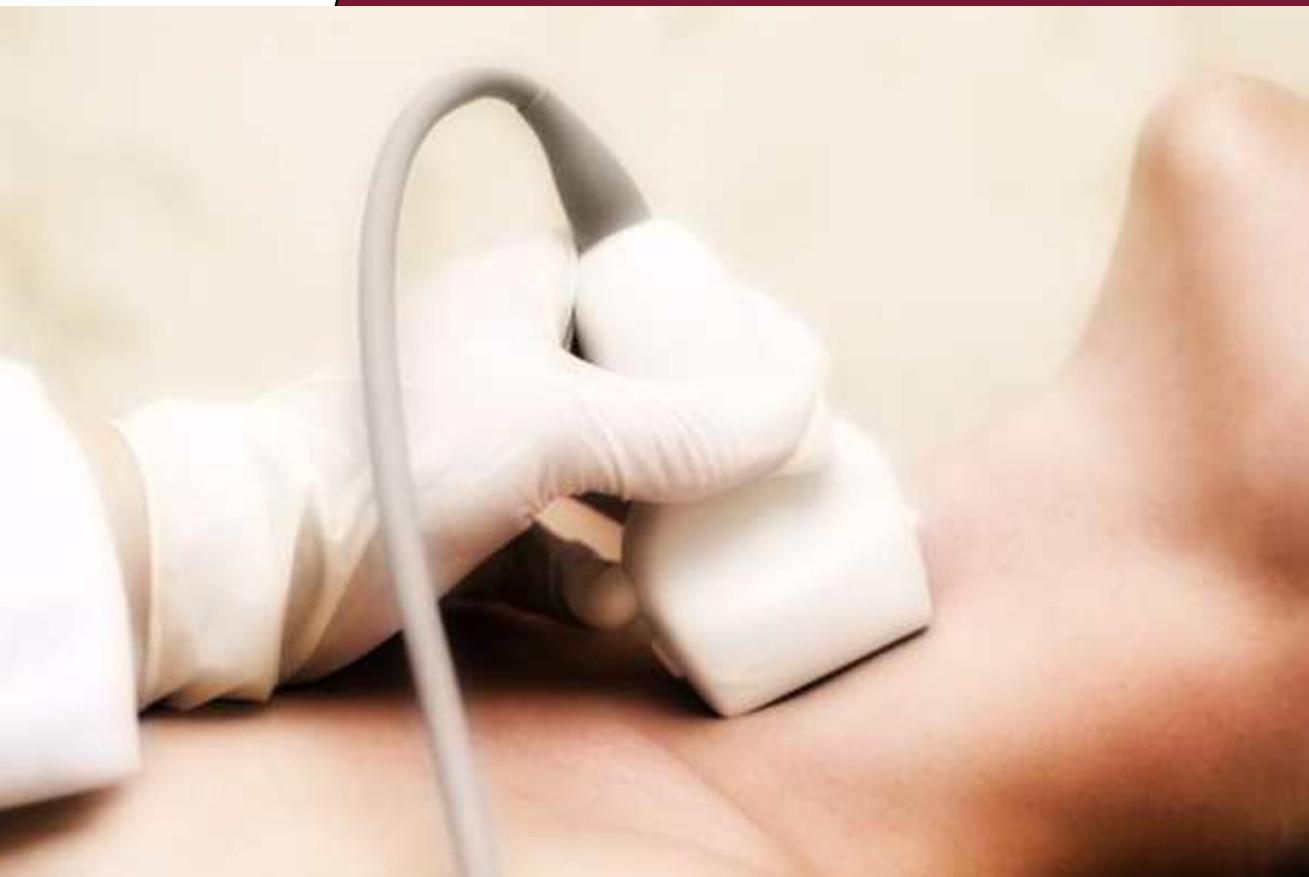
**Brecha tecnológica entre instituciones médicas**

**Escasa interpretabilidad de modelos de inteligencia artificial en salud**

**Riesgo de diagnósticos erróneos o no reproducibles**

## Motivación y relevancia

La motivación principal de este proyecto surge de la necesidad de mejorar los procesos de diagnóstico temprano del cáncer de tiroides, una enfermedad cuya detección oportuna incrementa significativamente las probabilidades de éxito en el tratamiento. En numerosos centros médicos, especialmente en regiones con recursos limitados, la falta de especialistas en radiología y la alta carga de trabajo generan retrasos diagnósticos y variabilidad en la interpretación de las ecografías. Esta realidad plantea la urgencia de integrar herramientas tecnológicas que apoyen la labor médica sin reemplazarla.



# Datos y Metodología

## Dataset Utilizado

- El proyecto empleó un dataset compuesto por 637 imágenes ecográficas tiroideas.
- Clasificadas en dos categorías: benignas y malignas
- Split estratificado (70% entrenamiento, 20% validación/ajustes y 10% prueba).
- Las imágenes para el entrenamiento se sacaron de la siguiente dirección drive:

/content/drive/MyDrive/p\_1\_image

## Preprocesamiento Realizado

- Carga y limpieza de imágenes
- Redimensionamiento ( $224 \times 224$ )
- Normalización (0–1)
- Conversión a RGB
- División Train/Val/Test
- Preparación de batches

# Datos y Metodología

## FLUJO DE PROCESAMIENTO PRINCIPAL:

1. **Carga** (Cargar\_dataset\_completo\_avanzado)
2. **Preprocesamiento** (Cargar y preprocessar\_imagen\_avanzado)
3. **Extracción características** (Extraer características avanzadas\_completas)
4. **Análisis EDA**
5. **Entrenamiento modelo** (Crear modelo de predicción compatible)
6. **Predicción** (Predecir Imágenes tiroides)
7. **Diagnóstico** (Generar diagnóstico detallado)



### Modelo

1. Random Forest (Optimizado)
2. Clasificación binaria: *benign* vs *malignant*
3. Buen desempeño con datasets pequeños
4. Robusto al ruido
5. Reduce sobreajuste

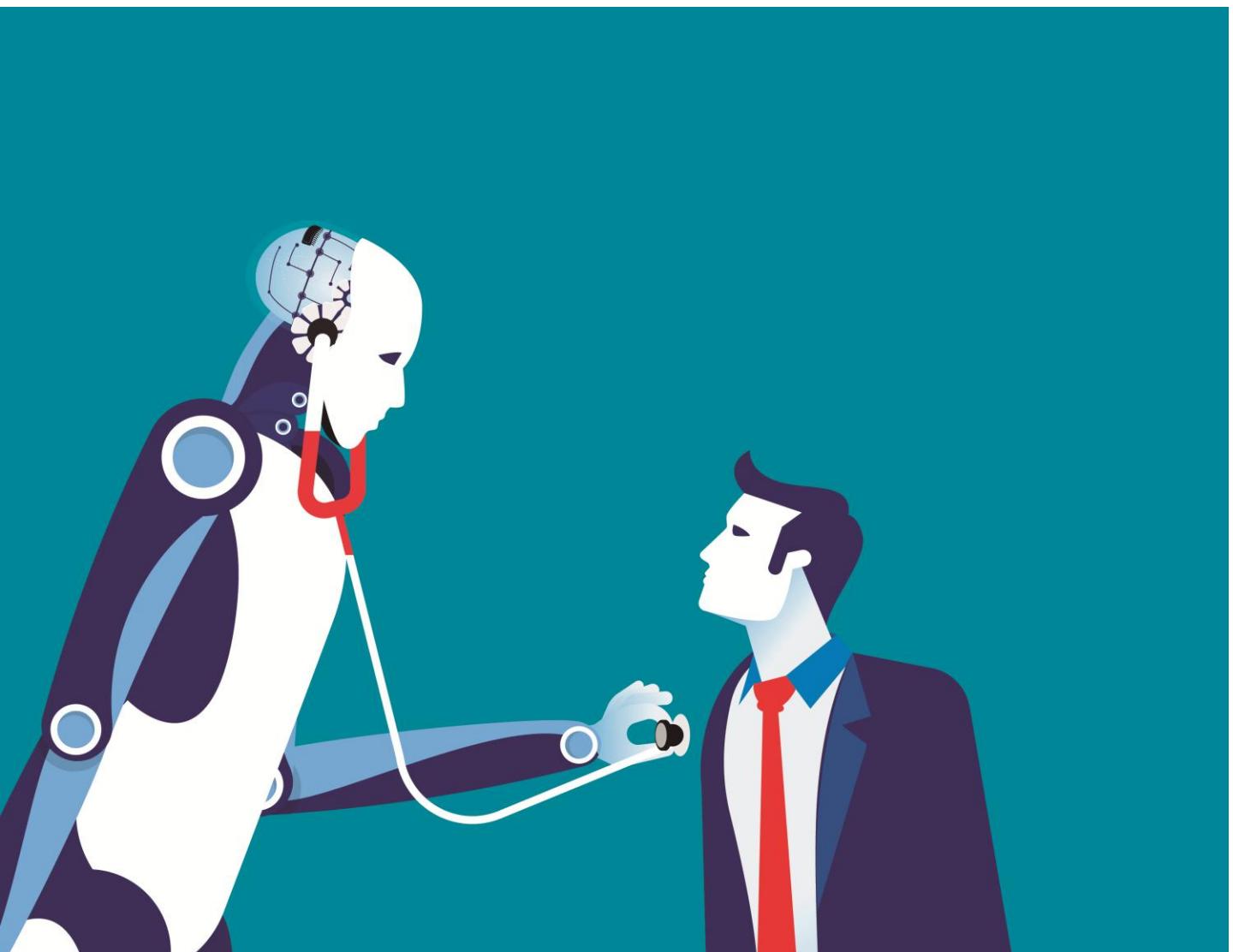
### Arquitectura

1. Ensamble de árboles de decisión
2. Votación mayoritaria entre árboles
3. Uso de subconjuntos aleatorios de datos y atributos
4. Optimización de hiperparámetros (*n\_estimators*, *max\_depth*)
5. Estabilidad y mejor generalización

# Datos y Metodología

## Justificación de Decisiones Técnicas

1. Selección de RF Optimizado
2. CNN Optimizado con EfficientnetB0
3. Redimensionamiento/Normalización  
0-1/Conversión RGB
4. Ajustes de hiperparámetros mediante  
GridSearch y RandomSearch
5. Validación cruzada
6. Optimización por accuracy y precision



# DEMO

## Inicio de Sesión

🔒 Inicio de Sesión

Usuario  
admuser

Contraseña  
admuser

Ingresar

localhost:8501/EDA

Login

EDA

Entrenamiento

Home

Predicción por imagen

## Panel de Análisis Tiroideo con IA

Elige en la Barra Izquierda la opción deseada

localhost:8501/EDA

Login

EDA

Entrenamiento

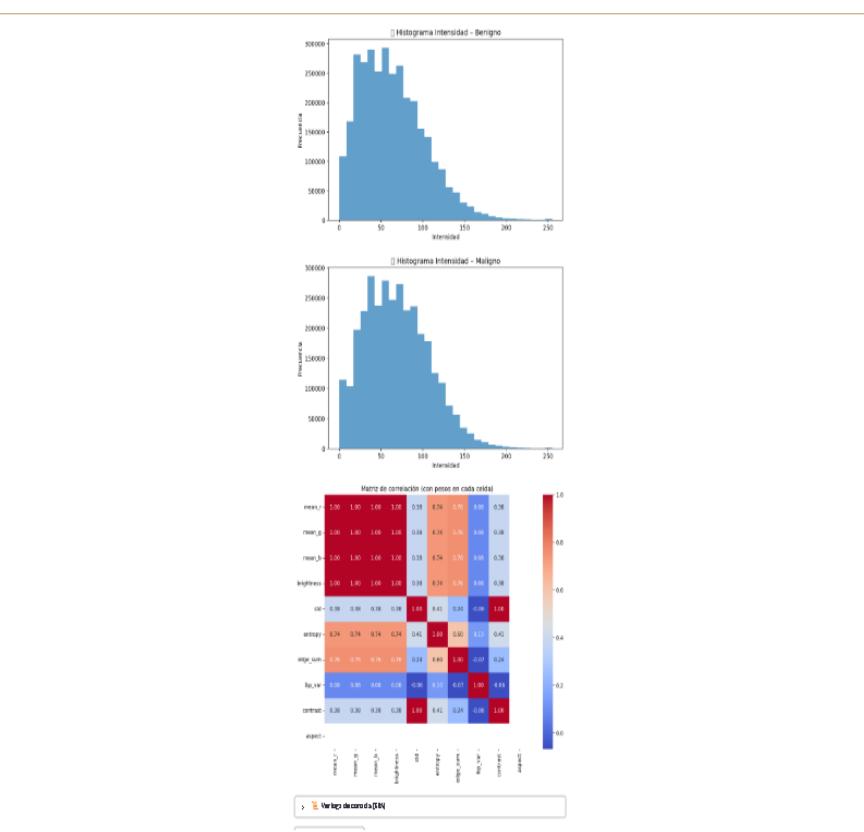
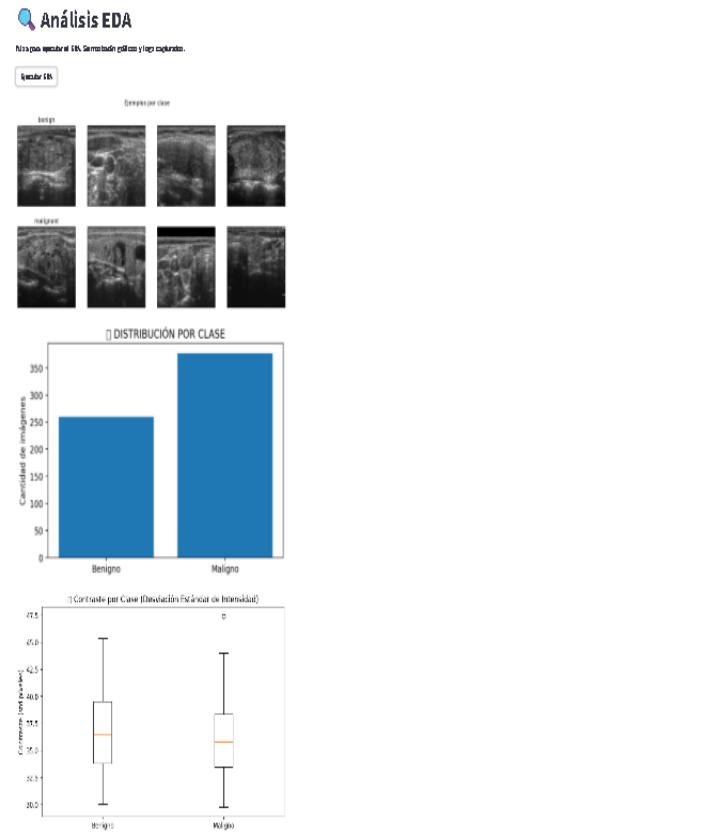
Home

Predicción por imagen

## Análisis EDA

Pulsa para ejecutar el EDA. Se mostrarán gráficos y logs capturados.

Ejecutar EDA



Ver logs de consola (EDA)

Total imágenes: 637

label

malignant 377

benign 260

Name: count, dtype: int64

Extrayendo features para EDA: 0%

Estadísticas de features:

	count	mean	std	min	25%
mean_r	637.0	6.797486e+01	1.269129e+01	35.400969	5.942492e+01
mean_g	637.0	6.797486e+01	1.269129e+01	35.400969	5.942492e+01
mean_b	637.0	6.797486e+01	1.269129e+01	35.400969	5.942492e+01
brightness	637.0	2.665681e-01	4.976975e-02	0.138827	2.330389e-01
std	637.0	1.455547e-01	1.698747e-02	0.095115	1.333319e-01
entropy	637.0	6.731508e+00	2.035605e-01	5.506553	6.613694e+00
edge_sum	637.0	1.657958e+00	4.132191e+05	468690.000000	1.362210e+06
lbp_var	637.0	2.297188e-03	6.347703e-04	0.001379	1.824648e-03
contrast	637.0	3.711644e+01	4.331805e+00	24.254429	3.399963e+01
aspect	637.0	1.000000e+00	1.111095e-16	1.000000	1.000000e+00

Distribución de clases:

label

malignant 377

benign 260

Name: count, dtype: int64

ESTRUCTURA DE DIRECTORIOS:

- Ruta benigno: C:/Users/USE/Downloads/ProyFathActualizado/ProyFath/data/benign
- Ruta maligno: C:/Users/USE/Downloads/ProyFathActualizado/ProyFath/data/maligno
- C:\Users\USE\Downloads\ProyFathActualizado\ProyFath\venv\Lib\site-packages\streamlit.fig.savefig(\*image, \*\*kwargs)

# DEMO

localhost:8501/Prediccion\_por\_imagen

Login  
EDA  
Entrenamiento  
Home  
**Prediccion por imagen**

## Prediccion por Imagen

Sube una imagen tiroidea y ejecuta el diagnóstico asistido por IA.

Modelo CNN cargado correctamente

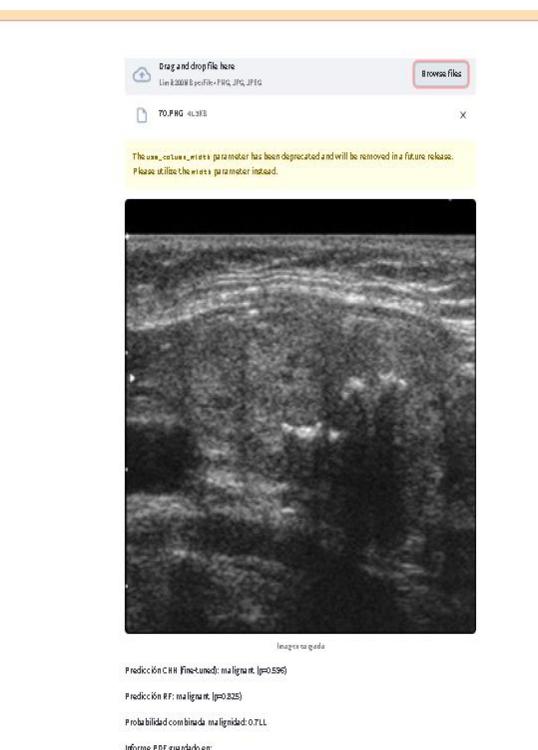
Modelo Random Forest cargado correctamente

Sube una imagen para diagnóstico

Drag and drop file here  
Limit 200MB per file • PNG, JPG, JPEG

Browse files

No se subió ninguna imagen; por favor vuelve a ejecutar la celda y carga una imagen.



localhost:8501/Prediccion\_por\_imagen

Diagnóstico profesional (visualizado en notebook y guardado en PDF)

Imagen analizada

Informe Asistido por IA - Clínica de Análisis Avanzado de Imágenes (IA)

Fecha: 2025-12-09 16:46:16

Resultado combinado: Malignidad 71.1% / Benigno 28.9%

CNN (fine-tuned): malignant ( $p=0.596$ )

RandomForest (embeddings): malignant ( $p=0.825$ )

Recomendación clínica:

Sospecha de malignidad: Correlacionar con hallazgos clínicos y considerar biopsia (FNAB) para confirmación histológica.

71.1% maligno

Porcentaje (%) - Malignidad (izq) / Benigno (der)

localhost:8501/Prediccion\_por\_imagen

Login  
EDA  
Entrenamiento  
Home  
**Prediccion por imagen**

Imagen cargada

Predicción CNN (fine-tuned): malignant ( $p=0.596$ )

Predicción RF: malignant ( $p=0.825$ )

Probabilidad combinada malignidad: 0.711

Informe PDF guardado en:  
C:/Users/USE/Downloads/ProyFathActualizado/ProyFath/diagnostico\_tiroides\_informe\_A4\_landscape.pdf

Diagnóstico profesional (visualizado en notebook y guardado en PDF)

Imagen analizada

Informe Asistido por IA - Clínica de Análisis Avanzado de Imágenes (IA)

Fecha: 2025-12-09 16:46:16

Resultado combinado: Malignidad 71.1% / Benigno 28.9%

CNN (fine-tuned): malignant ( $p=0.596$ )

RandomForest (embeddings): malignant ( $p=0.825$ )

Recomendación clínica:

Sospecha de malignidad: Correlacionar con hallazgos clínicos y considerar biopsia (FNAB) para confirmación histológica.

localhost:8501/Prediccion\_por\_imagen

Deploy

### Notas finales:

- Se presenta la matriz de confusión del CNN (balanceada 1:1) con colores por cuadrante.
- La matriz de correlación muestra pesos (valores numéricos) en cada celda.
- Se aplicó fine-tuning simple al CNN y GridSearch al RF; cuadro antes/después mostrado.
- El informe profesional se genera en A4 horizontal (guardado como PDF en Drive).

Logs del módulo

Descargar logs

# Evaluación y Resultados

Modelo	Accuracy
CNN (Baseline)	0.637795
CNN (Fine-tuning)	0.669291
Random Forest (Base)	0.679688
Random Forest (Optim.)	0.710938

**Resultados comparativos de modelos con entrenamiento base y optimizados**

## CNN (Baseline & Fine-tuning)

- Accuracy: 0.6378 a 0.6692
- Fine-tuning mejora desempeño
- Sensible al tamaño limitado del dataset

## Random Forest (Base)

- Accuracy: 0.679688
- Funciona como baseline tradicional

## Random Forest (Optimizado)

- Mejor resultado global: 0.710938
- Mejora gracias a ajuste de hiperparámetros

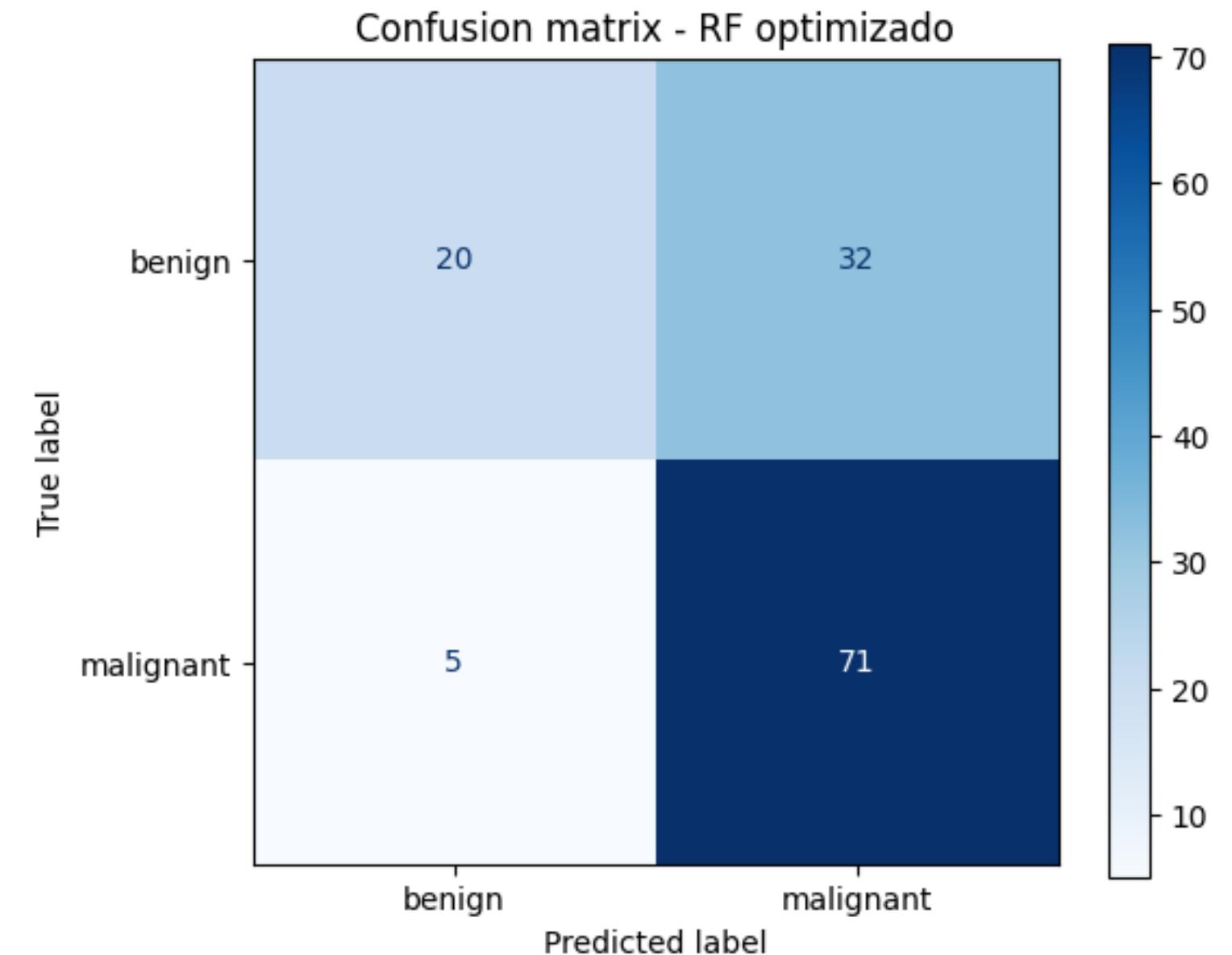
## Conclusiones

- Todos los modelos tienen rendimiento similar
- Optimización aporta mejoras ligeras
- CNN podría mejorar con más datos y regularización

# Evaluación y Resultados

## Comparación, Errores y Casos Clave RF Optimizado

class	precision	recall	f1-score	support
benign	0.80	0.38	0.52	52
malignant	0.69	0.93	0.79	76
accuracy	0	0	0.71	128
macro avg	0.74	0.66	0.66	128
weighted avg	0.73	0.71	0.68	128



# Consideraciones Éticas

## 1. Sesgos identificados

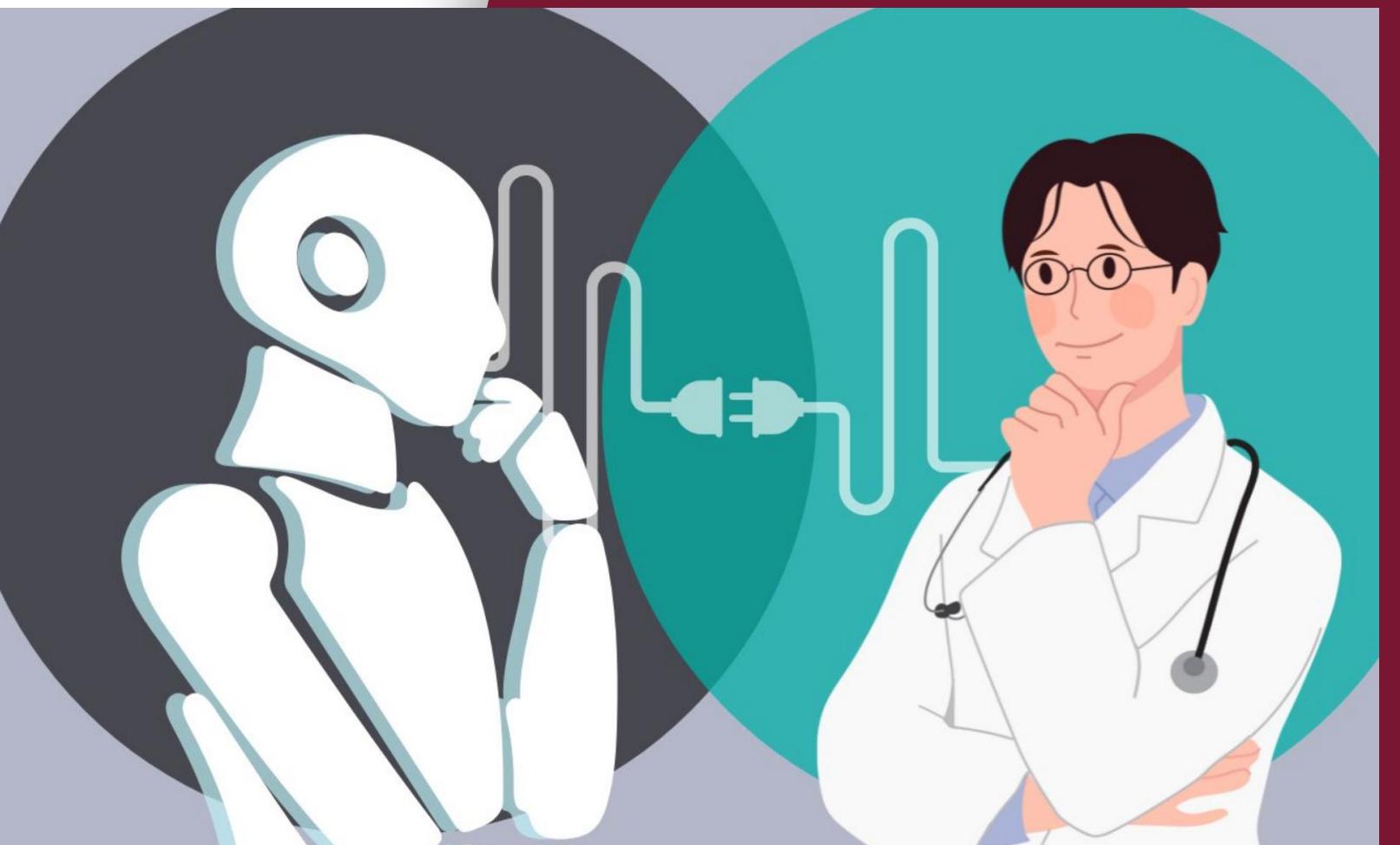
Possible sesgo por procedencia y calidad desigual de las imágenes del dataset.

Representación limitada de ciertos grupos clínicos o etarios.

## 2. Impacto social potencial

Contribuye a la detección temprana del cáncer de tiroides y optimiza la atención médica.

Riesgo de malinterpretación si se usa sin supervisión profesional.



# Consideraciones Éticas

## 3. Medidas de mitigación

Uso de datos anónimos y públicos.

Validación cruzada para reducir sesgos y errores.

Definido como herramienta de apoyo, no sustituto del médico.

## 4. Limitaciones éticas

Dependencia de la calidad y diversidad de los datos.

Requiere validación clínica adicional antes de aplicarse en hospitales.

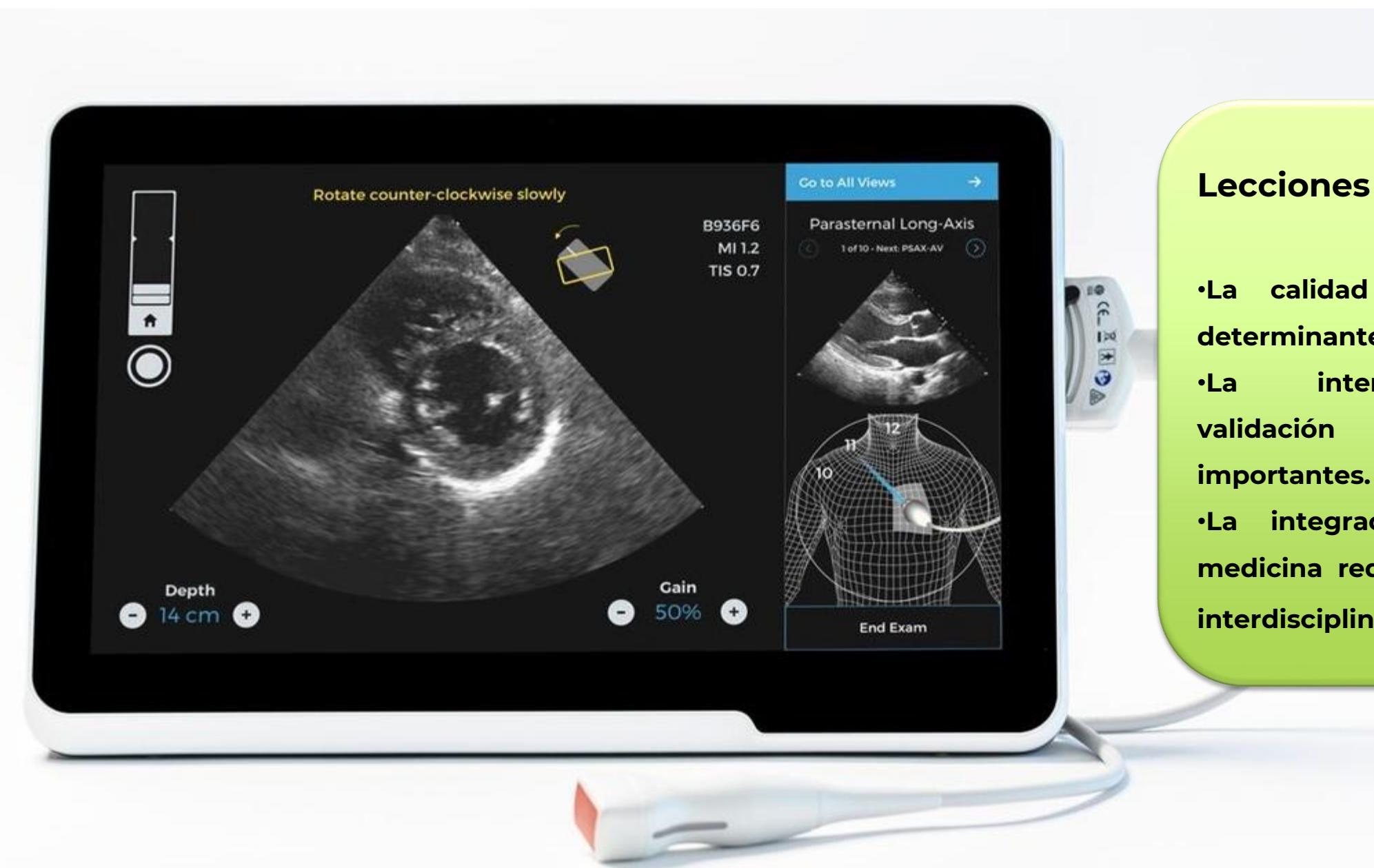
Necesario mantener intervención humana en la toma de decisiones.



# Conclusiones y Trabajo Futuro

## Logros principales

- Sistema capaz de clasificar nódulos tiroideos benignos y maligno.
- Preprocesamiento, modelado, evaluación y almacenamiento del modelo entrenado.
- Viabilidad del uso de IA.



## Lecciones aprendidas

- La calidad del dataset es determinante.
- La interpretabilidad y validación clínica son tan importantes.
- La integración entre IA y medicina requiere colaboración interdisciplinaria constante.

# Conclusiones y Trabajo Futuro

## Posibles mejoras futuras

- Ampliar el dataset.
- Implementar mas imágenes para que el modelo CNN sea mejor.
- Incluir un módulo de explicabilidad.

## Aplicaciones adicionales

- Extender el enfoque.
- Desarrollar una interfaz web médica implementada a HIS.
- Integrar en sistemas hospitalarios.

