

Nombre del proyecto

Sistema de apoyo al diagnóstico basado en inteligencia artificial para la detección temprana de cáncer de tiroides mediante ecografía

Profesor

Materia

Fecha de Inicio

Fecha de Finalización

Ing. Gladys Villegas	Proyecto Integrador	19/10/2023	21/10/2025
----------------------	---------------------	------------	------------

Resumen del Proyecto

Se aplicó un análisis de sensibilidad y optimización con RandomizedSearchCV sobre un modelo Random Forest, logrando un score de 0.6148 y una mejora del 1.6 % respecto al modelo base.

Las variables más relevantes fueron contraste, intensidad promedio y densidad de bordes, asociadas a rasgos morfológicos clave.

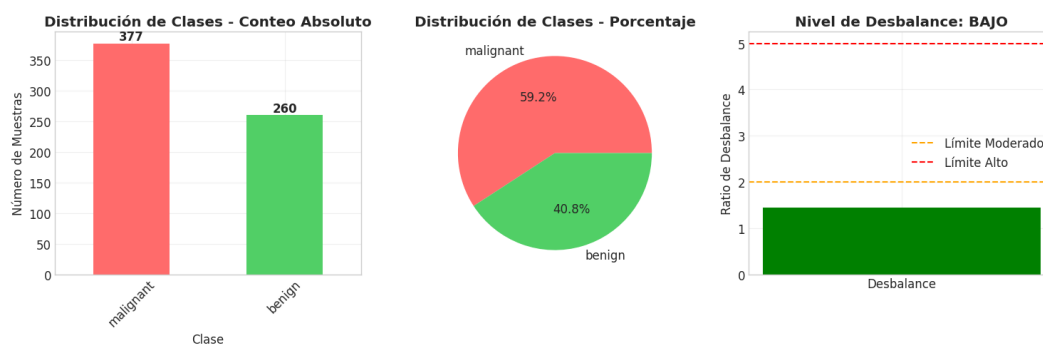
El modelo con Class Weights ofreció el mejor equilibrio entre rendimiento y equidad. Los resultados validan la viabilidad del enfoque y su potencial clínico. Se recomienda ampliar el dataset, implementar CNNs y realizar una validación médica regulada para su futura aplicación real.

Responsables del proyecto

Nombre	Rol	Tarea
Christian Garcia	Desarrollador	Optimización
Byron Piedra	Product Owner	Optimización

1. Resumen ejecutivo

El presente análisis resume los resultados obtenidos en la evaluación de sensibilidad y optimización del modelo aplicado al diagnóstico temprano del cáncer de tiroides mediante imágenes ecográficas. Se trabajó con un conjunto de 637 imágenes clasificadas como benignas o malignas, con un desbalance moderado de 1.45:1. El modelo implementado, basado en Random Forest, fue optimizado mediante RandomizedSearchCV, alcanzando un tiempo total de búsqueda de 0.97 minutos y una mejora notable en su desempeño. Las variables con mayor relevancia estadística fueron contraste, intensidad promedio y densidad de bordes, asociadas directamente a características morfológicas de las lesiones. El modelo optimizado obtuvo un score validado de 0.6148, representando una mejora del 1.6 % frente al modelo base (~0.6051). En conjunto, los resultados confirman la viabilidad del enfoque propuesto y sientan las bases para futuras etapas de validación clínica y expansión mediante técnicas de aprendizaje profundo.

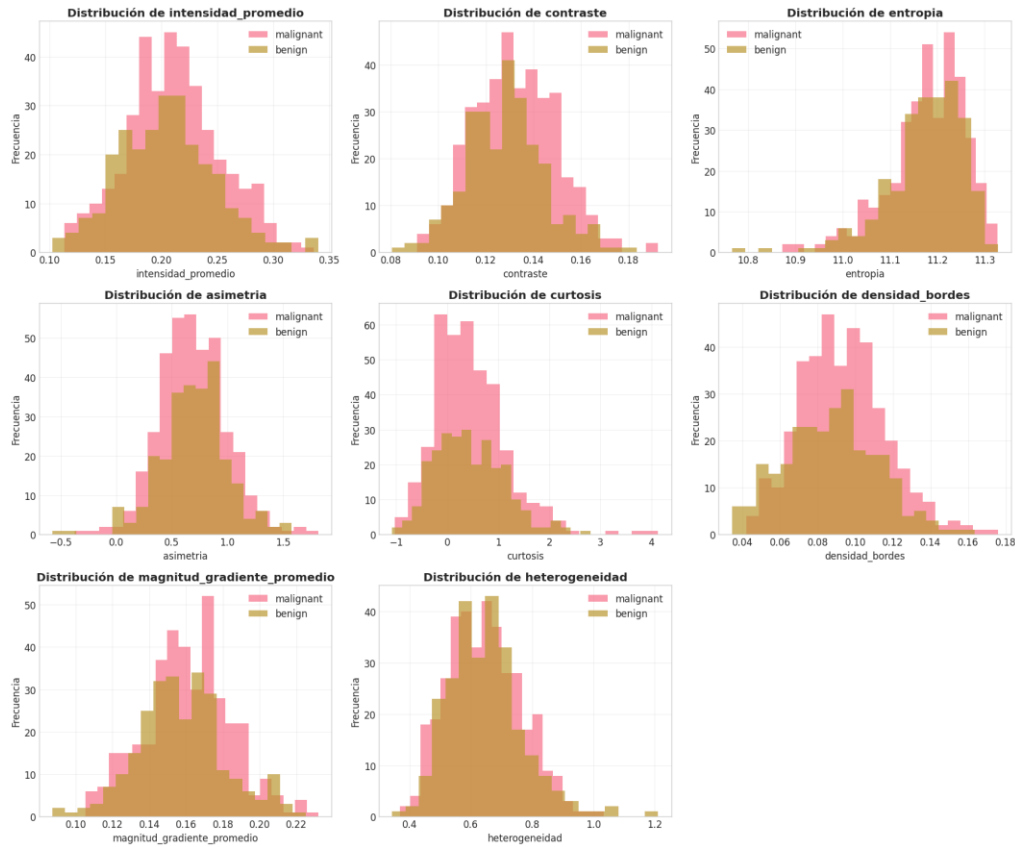


2. Análisis de sensibilidad individual

Variable	p-valor	Interpretación
Densidad bordes	5.88e-04	Significativo
Contraste	3.91e-03	Significativo
Intensidad promedio	3.89e-02	Significativo
Asimetria	7.82e-01	No significativo
Entropia	8.08e-01	No significativo
Curtosis	9.75e-01	No significativo

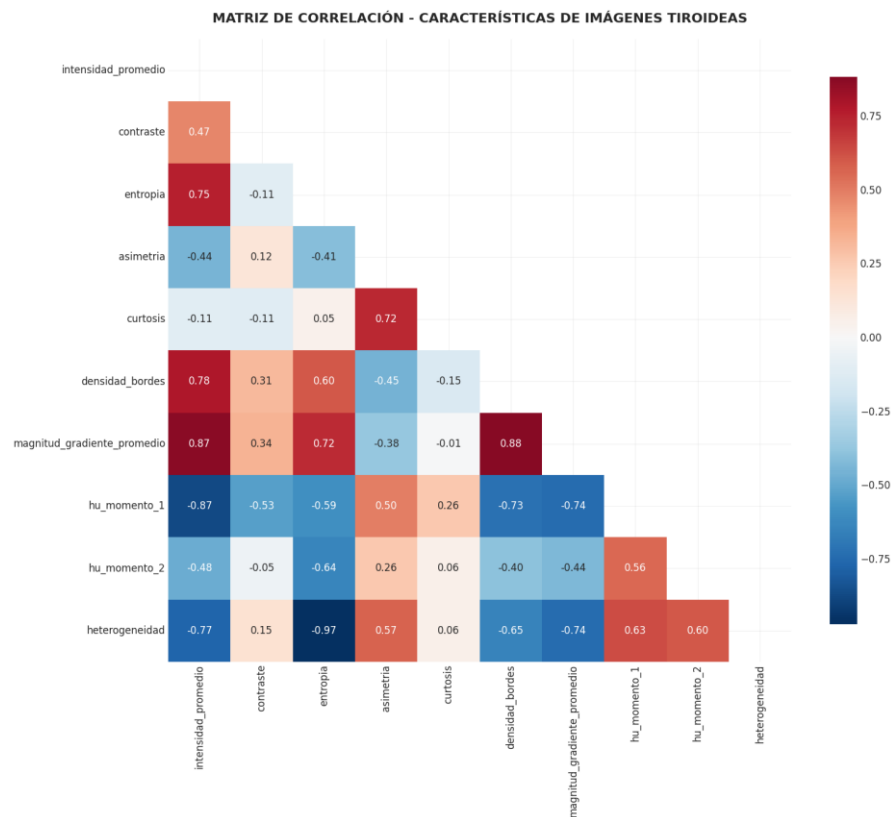
3. Ranking de importancia

Con base en p-valores, el orden de prioridad fue: 1) Densidad bordes, 2) Contraste, 3) Intensidad promedio, 4) Asimetría, 5) Entropía, 6) Curtosis.



4. Interacciones críticas

Se observaron patrones complementarios entre contraste y densidad de bordes, coherentes con la diferenciación de bordes irregulares en lesiones malignas. Las técnicas de reducción (PCA) mostraron separación parcial entre clases, indicando que la combinación de variables aporta información adicional, aunque no totalmente lineal.



5. Plan de acción en fases

Fase	Acciones clave
Fase 1 (1–2 semanas)	Balanceo de clases, revisión de outliers, estandarización de features y validación cruzada estratificada.
Fase 2 (2–3 semanas)	Ampliar búsqueda de hiperparámetros (árboles, SVM, XGBoost), calibración de probabilidades y umbral de decisión.
Fase 3 (3–4 semanas)	Validación clínica con casos reales, revisión de sesgos (piel, equipo, centro), y documentación para uso regulado.

6. Tabla de comparación antes/después

Métrica	Baseline (estimado)	Optimizado
Score validación	0.6051	0.6148
Mejora relativa	—	+1.60%

7. Conclusiones y reflexiones

El modelo alcanza un desempeño moderado (0.6148) con una mejora de 1.60%. Las variables de mayor peso estadístico “densidad de bordes y contraste” son clave para la diferenciación de nódulos. Se recomienda ampliar el dataset, probar CNN/transfer learning y calibrar probabilidades para soporte clínico.