

5.3.3 Análisis de los modos de fallo a partir de modelos interpretables

Los resultados del análisis anterior proporcionan una comprensión cualitativa de los errores de clasificación del modelo. Para cuantificar cómo se realizan las predicciones y, en consecuencia, identificar los casos en los que el modelo podría equivocarse, se ha seguido la metodología descrita en la sección 4.3.3 para entrenar dos modelos subrogados interpretables: un modelo de regresión lineal y un modelo de árbol de decisión. Estos modelos no están diseñados para diferenciar entre imágenes de control y de pacientes, sino para emular las predicciones del modelo original.

Así pues, utilizando como entrada las seis principales características de imagen identificadas previamente y como salida las predicciones del modelo original para la clase paciente se han entrenado ambos modelos subrogados. La curva ROC resultante para cada uno se muestra en la Figura 5.21. El umbral óptimo, determinado a partir de la mínima distancia al punto (0, 1) en la curva ROC, es de 0.48 para el modelo de regresión lineal y de 0.36 para el modelo de árbol de decisión.

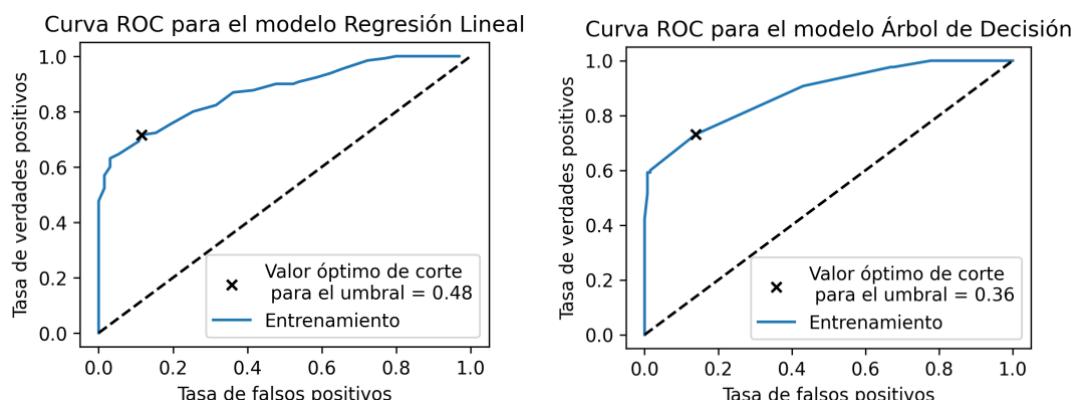


Figura 5.21. Curvas ROC obtenidas para el modelo de regresión lineal (izquierda) y para el modelo de regresión con árbol de decisión (derecha).

Los resultados presentados en la Tabla 5, demuestran que ambos modelos subrogados se entrenaron de manera satisfactoria, alcanzando una exactitud cercana al 80%. Además, la sensibilidad de ambos modelos es comparable, con una puntuación F1 de 0.782, lo que indica un buen rendimiento general en la emulación de las predicciones del modelo original.

Tabla 5. Resultados del entrenamiento de los modelos subrogados interpretables, modelo de regresión lineal y árbol de decisión.

Modelo	Datos de entrenamiento					
	Umbral seleccionado	Número de muestras	Exactitud	Precisión	Sensibilidad	Puntuación F1
Regresión lineal	0.48	260	0.800	0.861	0.861	0.782
Árbol de decisión	0.36	260	0.796	0.841	0.731	0.782

La evaluación de ambos modelos en los datos de prueba, cuyos resultados se muestran en Tabla 6, también es satisfactoria. El modelo de regresión lineal alcanzó una exactitud del 90% con una sensibilidad del 86%. Por su parte, el modelo de árbol de decisión, aunque presenta una exactitud ligeramente menor (85%), compensa con una mayor sensibilidad, alcanzando el 96%. Esto sugiere que el árbol de decisión, aunque menos preciso, es más eficaz en la identificación de casos positivos.

Tabla 6. Resultados de la prueba de los modelos subrogados interpretables, modelo de regresión lineal y árbol de decisión.

Modelo	Datos de Prueba					
	Umbral seleccionado	Número de muestras	Exactitud	Precisión	Sensibilidad	Puntuación F1
Regresión lineal	0.48	121	0.901	0.765	0.867	0.812
Árbol de decisión	0.36	121	0.851	0.630	0.967	0.763

Los modelos entrenados y su ecuación y forma interpretable se pueden encontrar en la Ecuación 5.1 para el modelo de regresión lineal y en la Figura 5.22 para el modelo de árbol de decisión.

Ecuación 5.1 Ecuación del modelo subrogado de regresión lineal entrenado para lograr la interpretabilidad del modelo original.

$$\begin{aligned}
 P(\text{Clase Paciente}) &= 0.395 * ((-0.0678 * \text{Número de agregados de colágeno}) + (0.596 \\
 &\quad * \text{Número de picos del VOG}) + (0.265 * \text{Intensidad media}) + (-0.449 \\
 &\quad * \text{Longitud de las fibras de colágeno}) + (-0.065 * \text{Porosidad}) + (0.122 \\
 &\quad * \text{Orientación de las fibras})) + 0.491
 \end{aligned}$$

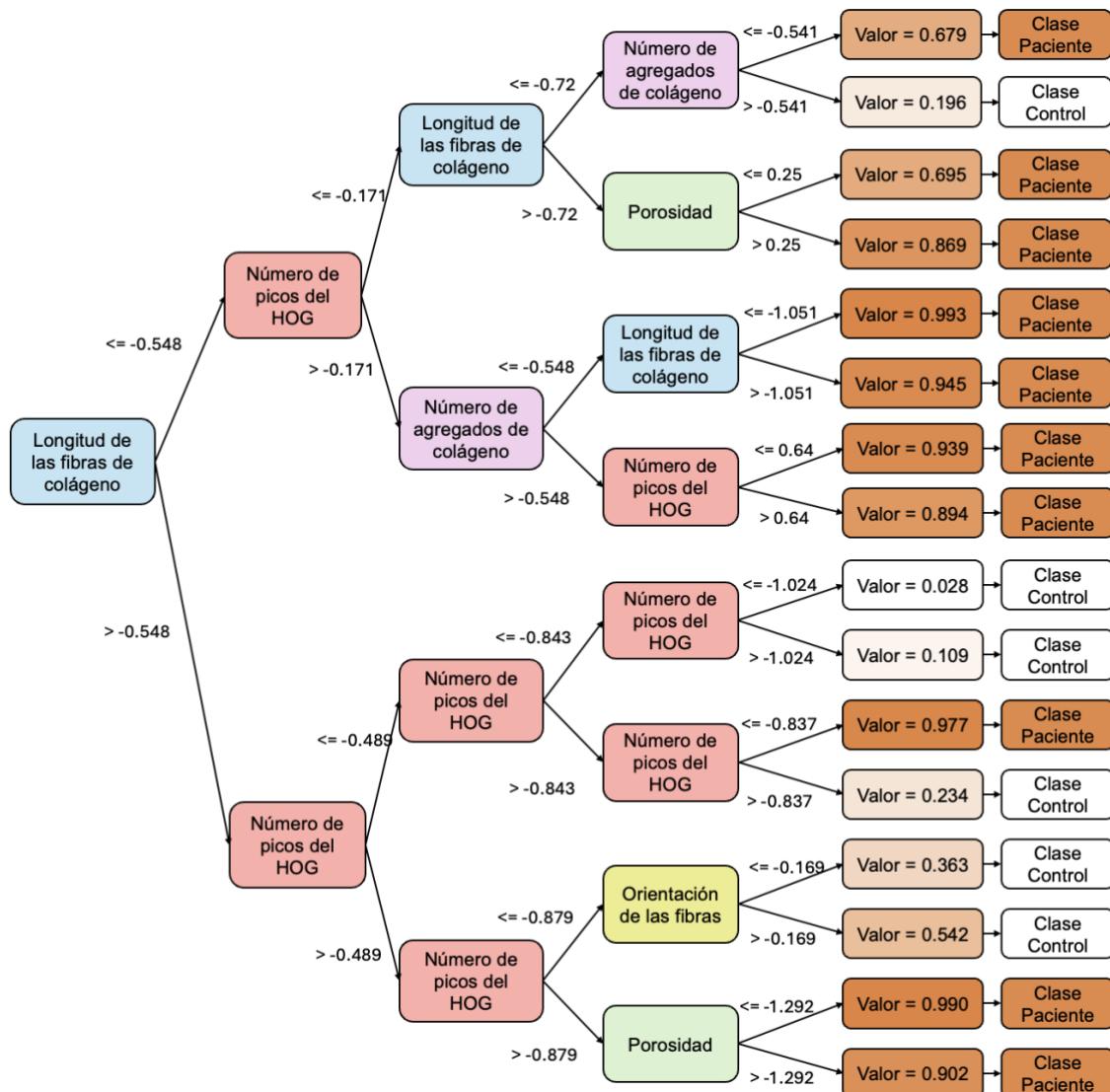


Figura 5.22. Representación del modelo subrogado de árbol de decisión entrenado para lograr la interpretabilidad del modelo original.

Dado que los resultados del entrenamiento de ambos modelos son satisfactorios, se puede concluir que estos modelos subrogados son útiles para interpretar cómo el modelo original realiza sus predicciones en diferentes escenarios. Aunque su capacidad predictiva no es total, estos modelos permiten obtener una aproximación razonable al razonamiento detrás de las decisiones del modelo. Esto significa que, aunque no es posible comprender con exactitud la causa de todas las predicciones del modelo original, los modelos subrogados ofrecen un valioso proxy para entender los patrones y características que influyen en las predicciones.

Es posible utilizar los modelos subrogados para ilustrar por qué el modelo original comete o no errores en sus predicciones. Al aplicar estos modelos, se puede analizar cuáles son las características que llevan a una clasificación y así entender qué factores influyen en los fallos del modelo. A continuación, se pueden tomar dos ejemplos específicos para ilustrar este análisis: