

**1.** Indique la cantidad de datos muestrales, verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Interprete estos valores.

|  |  |
| --- | --- |
| Verdaderos Positivos (TP) | 50 |
| Verdaderos Negativos (TN) | 10 |
| Falsos Positivos (FP) | 17 |
| Falsos Negativos (FN) | 23 |
| Cantidad de datos muestrales | 100 |

Se puede observar que el sistema tiene una buena capacidad para identificar los casos positivos, pero comete muchos errores a la hora de clasificar los casos negativos, especialmente los falsos negativos, lo cual es preocupante ya que representan tumores malignos que el sistema no detectó.

**2.** Calcule las métricas para evaluar la calidad de los resultados. ¿Qué significa el resultado obtenido en cada métrica?

|  |  |  |
| --- | --- | --- |
| **Métrica** | **Valor** | **Interpretación** |
| Exactitud (ACC) | 0,6 | Solo el 60% de las predicciones fueron correctas. |
| Sensibilidad (SEN), Recall o TPR | 0,684931507 | El 68.5% de los tumores malignos fueron correctamente identificados. |
| Especificidad (SPE) o TNR | 0,37037037 | Solo el 37.03% de tumores no malignos fueron correctamente identificados. |
| Precisión o Valor Predictivo Positivo | 0,746268657 | El 74.6% de los tumores clasificados como malignos eran efectivamente malignos. |
| Valor Predictivo Negativo (NPV) | 0,303030303 | Solo el 30.3% de los tumores no malignos eran efectivamente no malignos. |
| Tasa de descubrimiento falso (FDR) | 0,253731343 | El 25.4% de los tumores clasificados como malignos eran no malignos. |
| Tasa de falsos negativos (FNR) | 0,315068493 | El 31.5% de los tumores clasificados como no malignos eran malignos. |
| Tasa de falsos positivos (FPR) | 0,62962963 | El 62.9% de los tumores no malignos se clasificaron como tumores malignos. |
| Índice de elevación (Lift) | 1,022285831 | El modelo es solo ligeramente mejor que el azar. |
| F1-score | 0,714285714 | Valor aceptable entre precisión y recall. Sin embargo, deja mucho que desear para estándares médicos |

**3.** Teniendo en cuenta las métricas obtenidas, describa la calidad de los resultados indicando si son confiables. ¿Qué se debe mejorar?

A través de las métricas, se puede observar como el modelo identifica aceptablemente los tumores malignos con una sensibilidad del 68.5% y una precisión del 74.6%. Sin embargo, falla significativamente al detectar los tumores no malignos tal y como indica la especificidad de 37.03% y el bajo valor predictivo negativo de 30.3%. Además, el modelo cuenta con una alta tasa de falsos positivos de un 62.9% y una tasa de falsos negativos de 31.5%, la cual resulta especialmente preocupante en contextos médicos. Consideraría que el modelo es parcialmente confiable a la hora de detectar tumores malignos, pero no es confiable al momento de identificar tumores no malignos, lo cual podría generar diagnósticos fallidos. Para mejorar el modelo, se debe reducir el número de falsos negativos, así como aumentar la especificidad, ya que hay una muy gran tasa de falsos positivos. Por último, creo que se debería de realizar un balance en los datos, ya que estaríamos hablando de un dataset de 73 tumores malignos (TP+FN) y 27 tumores no malignos (TN+FP) por lo que la alta exactitud del modelo a la hora de detectar tumores malignos puede deberse a que está favoreciendo a dicha clase.