

# CLASIFICACIÓN DE TUMORES DE MAMA BENIGNOS Y MALIGNOS A PARTIR DE CARACTERÍSTICAS MORFOLÓGICAS MEDIANTE ALGORITMOS DE MINERÍA DE DATOS

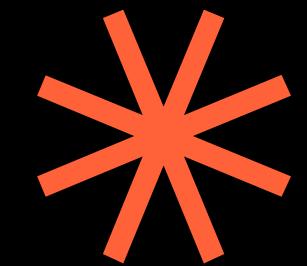
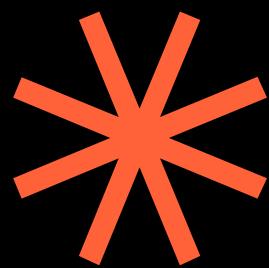


DIPLOMADO DE MINERÍA DE DATOS Y TÉCNICAS ESTADÍSTICAS G38  
Martinez Briones Angel Jesus





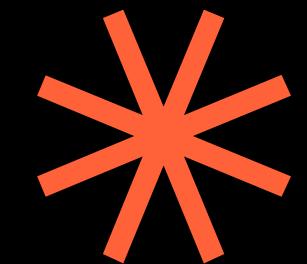
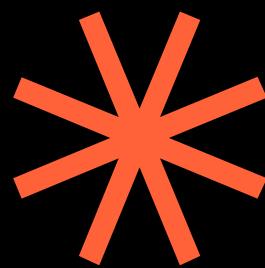
LAS ESTRATEGIAS IMPULSADAS POR LA OMS  
PARA ENFRENTAR ESTA ENFERMEDAD SE CENTRAN  
PRINCIPALMENTE EN LA EDUCACIÓN PARA LA  
DETECCIÓN TEMPRANA Y EL ACCESO OPORTUNO AL  
TRATAMIENTO.



Se estima que en el 2022  
**fallecieron 670,000** personas por cáncer de  
mama en todo el mundo.



## WISCONSIN BREAST CANCER DATASET (DIAGNOSTIC), CREADO ORIGINALMENTE POR EL DR. WILLIAM H. WOLBERG Y COLABORADORES EN LA UNIVERSIDAD DE WISCONSIN-MADISON.



¿qué tan efectivos son los modelos de **regresión logística y máquinas de soporte vectorial** para clasificar tumores mamarios benignos y malignos a partir de características morfológicas?



# ANALISIS DE LOS MODELOS

Modelo de regresión logística y un modelo SVM como métodos. Para evaluar el desempeño del modelo, el conjunto de datos fue dividido en subconjuntos de **entrenamiento 70% y prueba 30%**, y se emplearon métodos de evaluación como la curva ROC, AUC y matriz de confusión.

Cuadro I: Modelo SVM (Kernel Lineal)

Predicción / Referencia	B	M
Predicción (B)	99	5
Predicción (M)	0	57

Accuracy 96.8 %  
AUC 0.992  
Sensibilidad 91.9 %  
Especificidad 100 %  
Balanced Accuracy 95.97 %

0 falsos positivos = estrés, pruebas innecesarias  
5 falsos negativos = riesgo clínico



Cuadro II: Modelo de Regresión Logística

Predicción / Referencia	B	M
Predicción (B)	91	6
Predicción (M)	9	56

Accuracy 91.3 %  
AUC 0.956  
Sensibilidad 90.3 %  
Especificidad 91.9 %  
Balanced Accuracy 91.1 %

6 falsos positivos = estrés, pruebas innecesarias  
9 falsos negativos = riesgo clínico



# ANALISIS DE LOS MODELOS

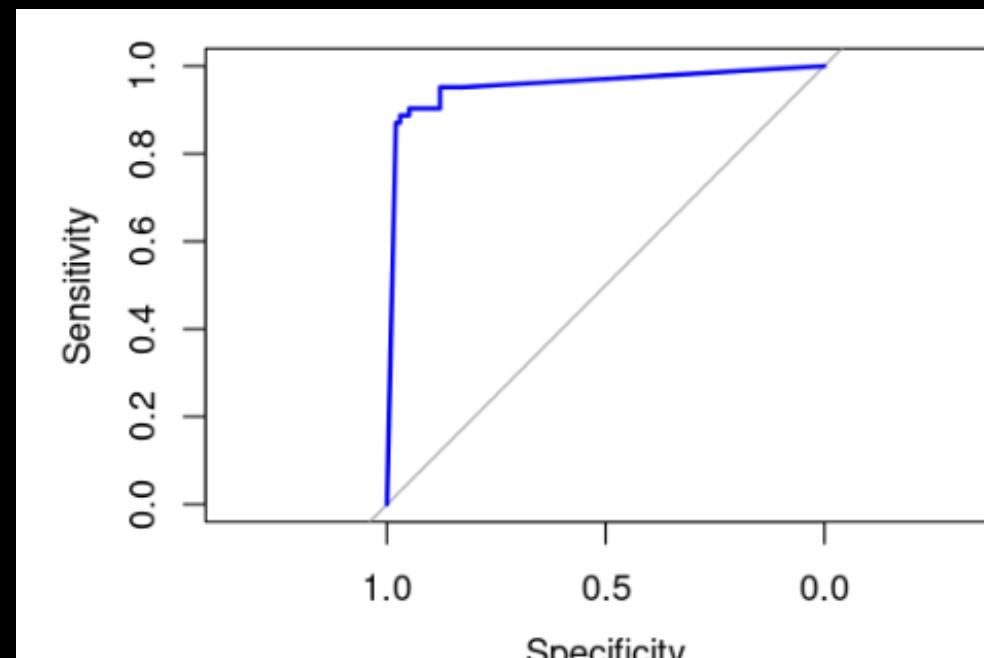


Figura 1: Curva ROC del Modelo de regresión logistica.



Cuadro III: Comparación de métricas

Métrica	Regresión Logística	SVM (Lineal)	Ganador
Accuracy	91.3 %	96.8 %	SVM
AUC	0.956	0.992	SVM
Sensibilidad (M)	90.3 %	91.9 %	SVM
Especificidad (B)	91.9 %	100 %	SVM
Balanced Accuracy	91.1 %	96.0 %	SVM

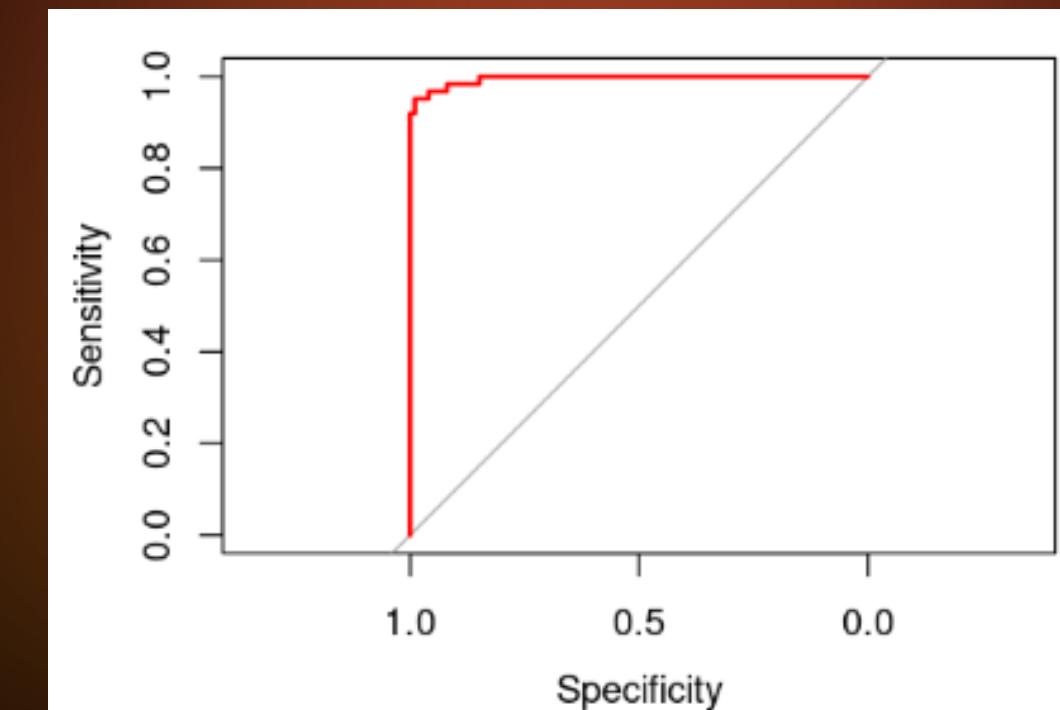


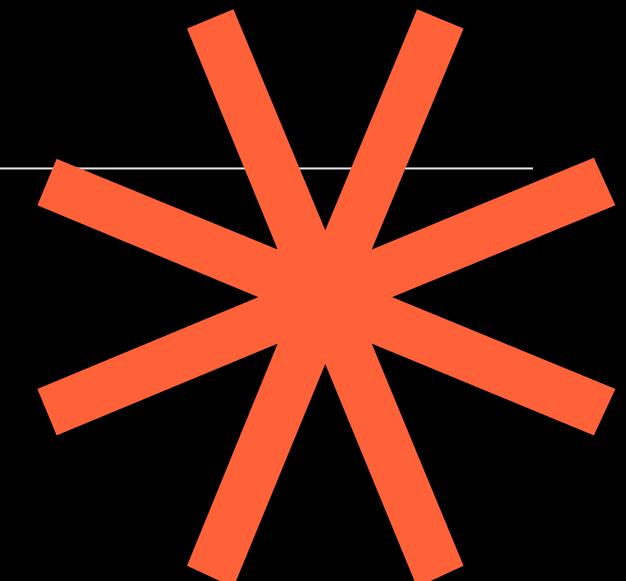
Figura 2: Curva ROC del Modelo SVM.



El uso de características morfológicas procesadas mediante minería de datos se valida como una **alternativa de apoyo diagnóstico viable** para reducir la brecha de mortalidad en países de renta baja.

En el contexto de países con infraestructura sanitaria limitada descrita por la OMS, un falso negativo es el escenario de mayor riesgo, ya que posterga un tratamiento. Por tanto, aunque el SVM es el modelo ganador, su implementación debe ser considerada como una **herramienta de apoyo y no como un sustituto definitivo** del juicio clínico.

Smith, J., & Jones, L. (2024). Avances en la detección morfológica del cáncer de mama mediante algoritmos de soporte vectorial. Journal of Medical AI, 12(2), 45-60. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12635515/>





GRACIAS  
POR LA  
ATENCION

---



-

amtzb16@gmail.com