

Recurrencia de cáncer de mama

Mariela Arias

LTIC

ENES-Morelia

Email: marielarojo11@gmail.com

Resumen—La recurrencia en el cáncer de mama es cuando después del tratamiento, el cáncer reaparece. La recurrencia puede ser local, que se refiere a que aparece de nuevo en el mismo seno o la cicatriz de la operación, o regional, que es cuando aparece en los ganglios linfáticos cercanos. Hay una gran diferencia entre la recurrencia y la progresión del cáncer. Cuando el cáncer reaparece a muy pocos meses del final del tratamiento, lo más probable es que sea una progresión y se puede deber a que en la cirugía no se removió por completo o que el cáncer ha vuelto resistente al tratamiento.[1]

La detección temprana del cáncer aumenta considerablemente la tasa de supervivencia, según la Sociedad Americana del Cáncer la tasa aumenta 5 años en 99 % de las detecciones tempranas. En la actualidad existen diversas fuentes de datos sobre la recurrencia y detección de cáncer, y estas utilizan diferentes tecnologías, de esta manera se puede utilizar el aprendizaje automático para su detección.

I. INTRODUCCIÓN

El cáncer de mama es actualmente el tipo más común de cáncer en el mundo, con al menos 276.000 nuevos casos ocurriendo solo en los EE. UU. en 2020.[2] En la búsqueda de nuevos métodos de bajo costo, menos invasivos y rápidos para poder detectar el cáncer de mamá, se ha encontrado la alternativa del uso de métodos de aprendizaje automático y de esta manera poder clasificar el cáncer con base en perfiles morfológicos.

En este proyecto ocupamos una fuente de datos donde podemos clasificar la recurrencia o no recurrencia de cáncer de mama a partir de la edad, tamaño del tumor, si el paciente había entrado en la menopausia o no, entre otros. Por otro lado, se ha demostrado que el ADN y ARN ayudan a la detección y clasificación del cáncer de mama. Los ARN extracelulares (exARN) se encuentran presentes en el suero humano, pero no está claro en qué medida estos exARN circulantes pueden reflejar estados fisiológicos y de enfermedades humanas. Para analizar eficazmente en exARN y la relevancia que tiene en estudios clínicos, se desarrolló una tecnología de secuenciación de exARN de baja entrada llama SILVER-seq.[3]

La tecnología SILVER-seq reveló una relación entre los niveles de exARN en suero con el sexo y la edad del paciente.

II. MATERIALES Y MÉTODOS

Para el entrenamiento del modelo usamos un conjunto de datos[4] de 286 pacientes, de los cuales 201 corresponden a la no recurrencia y 85 a la recurrencia. Los atributos que consideramos son los siguientes: la edad, menopausia, tamaño del tumor, nodos invadidos, tumor maligno, en que pecho

se presentó y si fue cuádriceps mamario. Entrenamos tres modelos diferentes: k-nearest neighbors, regresión logística y máquina de vectores de soporte (SVM). Para su entrenamiento dividimos los datos en 70 % training y 30 % test.

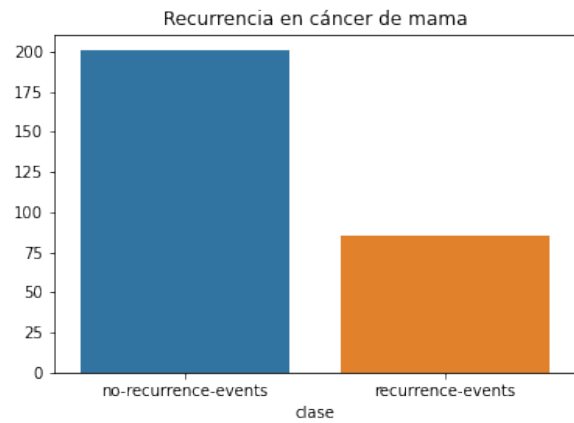


Figura 1: Manejamos 286 muestras de diferentes pacientes, las cuales 201 corresponden a la no recurrencia y 85 a la recurrencia.

II-A. Modelo kNN

Para comenzar a entrenar al modelo kNN, primero evaluamos los puntajes por cada valor de k, tomando un rango de 2 a 16, consideramos a k=9 como un buen parámetro para ahora poner en marcha los datos, hacer predicciones y obtener el reporte de precisión, recall y accuracy.

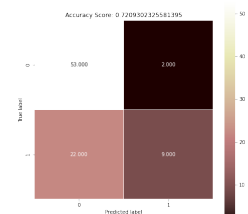


Figura 2: En nuestra matriz de confusión del modelo kNN, tenemos los siguientes resultados: 53 casos fueron verdaderos negativos, 9 fueron verdaderos positivos, 22 falsos negativos y 9 falsos positivos. Y accuracy = 72 %.

II-B. Modelo SVM

El modelo de maquinas de soporte vectorial lo entrenamos de manera lineal, se calcularon las predicciones y el reporte de precisión, recall y accuracy.

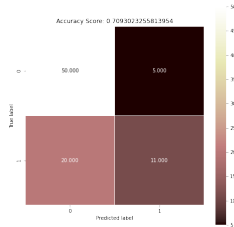


Figura 3: En nuestra matriz de confusión del modelo SVM, tenemos los siguientes resultados: 50 casos fueron verdaderos negativos, 11 fueron verdaderos positivos, 20 falsos negativos y 5 falsos positivos. Y accuracy = 70 %.

II-C. Modelo Regresión Logística

Después de entrenar el modelo de regresión logística obtuvimos las predicciones y reporte de precisión, recall y accuracy.

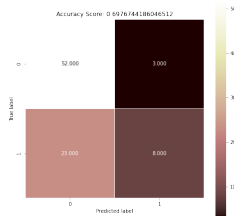
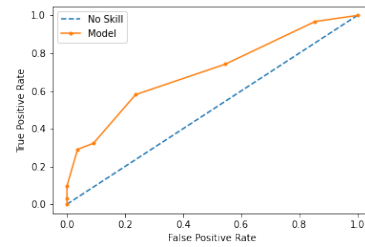


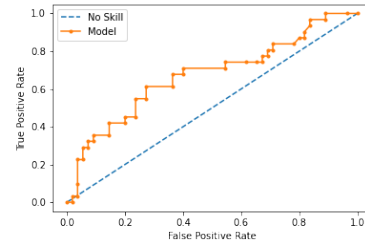
Figura 4: En nuestra matriz de confusión del modelo de regresión logística, tenemos los siguientes resultados: 52 casos fueron verdaderos negativos, 8 fueron verdaderos positivos, 23 falsos negativos y 3 falsos positivos. Y accuracy = 69 %.

III. EXPERIMENTOS Y RESULTADOS

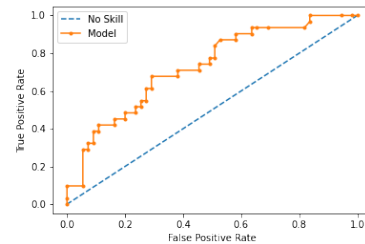
Los resultados de los modelos fueron favorables, ya que al analizarlos en la curva ROC[5], en los tres modelos el área de la curva se encuentra en el lado de verdaderos positivos. Pero hablando específicamente sobre el resultado de accuracy, no se obtuvieron valores realmente altos, ninguno superó el 75 %.



(a) Curva ROC del modelo kNN



(b) Curva ROC del modelo SVM



(c) Curva ROC del modelo de Regresión Logística

Figura 5: La curva ROC está en el área de verdaderos positivos

Después de entrenar por separado cada modelo, implementamos una validación cruzada con ayuda de k-fold y le asignamos $k = 10$. Tomamos en cuenta el valor de accuracy de cada modelo y también el valor de su desviación estándar. Obtuvimos resultados diferentes, pues al inicio el modelo con mayor puntaje fue kNN y ahora es el modelo SVM con un puntaje de 76 %.

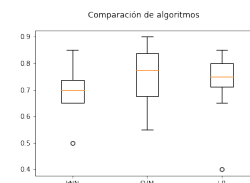


Figura 6: El modelo kNN tiene un puntaje de 69 % y su desviación estándar es 8.6, el modelo SVM tiene un puntaje de 76 % y desviación estándar de 10.9 y el modelo de Regresión logística tiene un puntaje de 73 % y desviación estándar de 12.4.

IV. CONCLUSIONES

La detección temprana del cáncer aumenta la tasa de supervivencia, por lo que es importante la constante revisión. Al tener diversas fuentes de datos respecto al cáncer de mama, a la recurrencia o detección por primera vez, podemos tener en cuenta los diferentes aspectos que se toman en la evaluación, como por ejemplo la edad o si es que ya se encuentra en la etapa de menopausia, y así tomar las medidas necesarias para cuidar la salud.

ACKNOWLEDGMENT

El código escrito e imagenes se encuentran en el repositorio de GitHub https://github.com/marielaAriass/recurrencia_cancerdemama

REFERENCIAS

- [1] Tratamiento del cáncer de seno recurrente. [Online]. Available: <https://www.cancer.org/es/cancer/cancer-de-seno/tratamiento/tratamiento-del-cancer-del-seno-segun-su-etapa/tratamiento-para-el-cancer-de-seno-recurrente.html>
- [2] Worldwide cancer data | world cancer research fund international. [Online]. Available: <https://www.wcrf.org/cancer-trends/worldwide-cancer-data/>
- [3] Z. Zhou, Q. Wu, Z. Yan, H. Zheng, C.-J. Chen, Y. Liu, Z. Qi, R. Calandrelli, Z. Chen, S. Chien, H. I. Su, and S. Zhong, "Extracellular RNA in a single droplet of human serum reflects physiologic and disease states," vol. 116, no. 38, pp. 19 200–19 208. [Online]. Available: <https://pnas.org/doi/full/10.1073/pnas.1908252116>
- [4] UCI machine learning repository: Breast cancer data set. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>
- [5] J. Cerda and L. Cifuentes, "Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos," vol. 29, no. 2, pp. 138–141. [Online]. Available: http://www.scielo.cl/scielo.php?script=sci_abstract&pid=S0716-10182012000200003&lng=es&nrm=iso&tlng=es