

Desarrolle un modelo para predecir el tiempo de supervivencia de cáncer de pulmón, basándose en imágenes dos dimensionales (tomografías computarizadas, TC) y datos clínicos.

Realice los siguientes pasos para la predicción:

1. Carga de datos y preprocesamiento
 - a. Cargar los dos tipos de datos (datos clínicos e imágenes), los datos vienen ya divididos en prueba y entrenamiento.
 - b. Note que las tomografías computarizadas almacenan la intensidad bruta de los vóxeles en unidades Hounsfield (UH). Esto se define como Aire = -1000 HU, Pulmón $\approx$ -500 HU, Agua = 0 HU, Tejidos blandos (y sangre) $\approx$ +50 HU, Hueso $\approx$ +1000 HU. Se puede aplicar una ventana de pulmón de -1200 a +800 para ver el tumor de pulmón y normalizar las tomografías. Para cada TC, hay una máscara de segmentación del mismo tamaño, compuesta por ceros en todos los lugares donde no hay tumor, y 1 en caso contrario. El objetivo es aumentar el tamaño de los datos de entrenamiento cortando cada TC 3D en el centro del tumor en cada eje, e utilizar una máscara para identificar la posición del tumor en la TC.
 - c. Note que algunos TC de entrenamiento no tienen ningún tumor delimitado (la máscara de segmentación está compuesta únicamente por ceros). Estas tomografías no se utilizan en las creaciones de las tomografías 2D.
 - d. Visualizar una de las imágenes
 - e. Variables de objetivo y censura
 - f. Organizar los datos del tren en el entrenamiento y la validación
2. Creación de una red neuronal de convolución para el análisis de la supervivencia
 - a. Calcular la máscara que representa el conjunto de riesgos de cada muestra.
 - b. Cálculo de la función de pérdida PH de Cox
 - c. Cálculo del índice de concordancia (sobre los datos de validación en cada época)
 - d. Entrenamiento de la red neuronal convolucional de supervivencia
 - e. Arquitectura de la red neuronal convolucional de supervivencia. La arquitectura de la CNN es similar a la que utiliza el TAC 3D, pero adaptada a las imágenes 2D.
 - i. El tamaño de los TAC (imagen en escala de grises) es de 92x92. Se normalizan por lotes antes de alimentar la red.
 - ii. Dos capas convolucionales que utilizan un tamaño de kernel de 3x3, stride 1, padding='valid', y una función de activación ReLu, seguida de una normalización por lotes y una capa de pooling máximo con un tamaño de pool de 2, dividiendo la dimensión espacial con un tamaño de 2.
 - iii. Red totalmente conectada, compuesta por dos capas densas ocultas y una capa densa de salida. Se aplica una normalización por lotes a la salida de las dos capas densas ocultas. La última capa densa emite las predicciones (puntuación de riesgo). Se añade una capa de abandono con una tasa de abandono del 20% para reducir el sobreajuste.
3. Predicción de riesgo

El entregable deberá ser un archivo .ipynb que contenga los pasos de arriba.

Extra:

- o ¿Como se pueden comparar los resultados usando 2 o 3 dimensiones?