

INTELIGENCIA ARTIFICIAL: DETECCIÓN DE ECOCARDIOGRAMA

ALESSANDRO MORALES, GABRIELA SUBUYÚ



La ecocardiografía es una herramienta diagnóstica importante en la detección y monitoreo de enfermedades cardíacas. Dado un dataset con distintos datos de ecocardiografías, se pretende determinar la supervivencia de algunas pacientes bajo ciertas condiciones. Para esto se construyeron 3 modelos de inteligencia artificial que fueron utilizados para resolver el problema especificado.

DESCRIPCIÓN DEL DATASET

El dataset presenta información de pacientes que sufrieron un paro cardíaco en algún punto en su vida, algunos están vivos y otros no. El objetivo es determinar si el paciente sobrevivirá por lo menos un año después. Este se determina por la supervivencia y el tiempo que ha sobrevivido,

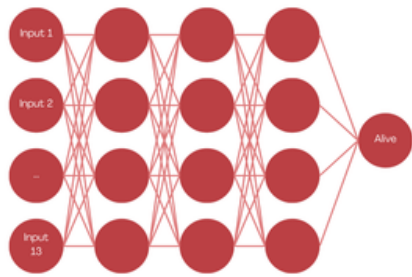
FEATURES:

- Age at heart attack
- Pericardial effusion
- Fractional shortening
- E-point septal separation
- Left ventricular end diastolic dimension
- Wall motion index

Target: Alive

METODOLOGÍA

La resolución del problema empezó por entender como las variables: survival, still alive y still alive at 1 estaban relacionadas. Luego de encontrar el target, se utilizó el balanceo de clases para determinar si era necesario realizar data augmentation, en este caso si era necesario.



Los 3 modelos realizados tienen la siguiente arquitectura, la cual permite solucionar un problema de clasificación. Los únicos resultados posibles son “Vive” y “No vive”.

Cada uno de los modelos realizados utilizaron callbacks: earlystopping, tensorboard y checkpoint.

RESULTADOS

Para la resolución del problema se trabajaron 3 modelos, con distintos enfoques: modelo realizado con preprocessing únicamente, modelo utilizando undersampling y modelo utilizando oversampling. Antes de realizar cada modelo se preprocesaron los datos

Modelo 1:

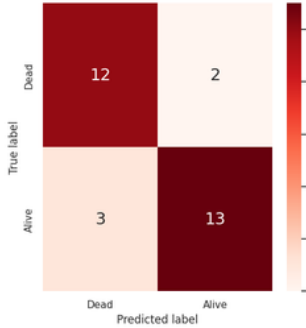
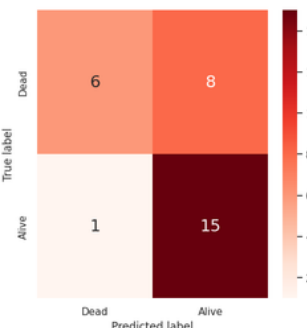
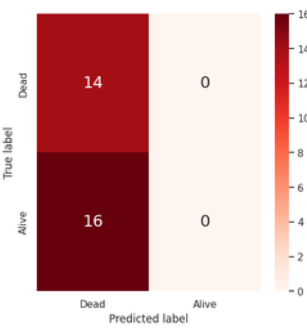
En este modelo utilizamos los datos preprocesados para crear el modelo; sin embargo, el balanceo de clases no fue muy favorable.

Modelo 2:

En este modelo, se buscó eliminar datos para igualar la cantidad de datos que hay de cada clase del target.

Modelo 3:

En este modelo, al igual también se buscó igualar la cantidad de datos, pero por medio de la creación de algunos datos



Accuracy: 0.7727
Loss: 0.6401

Accuracy: 0.7143
Loss: 0.6842

Accuracy: 0.8333
Loss: 0.6615

CONCLUSIONES

Los 3 modelos consiguieron un accuracy de arriba de 70. Sin embargo, es importante remarcar que cada uno de los modelos tuvo sus fallas.

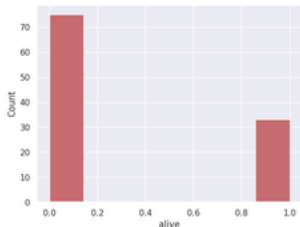
El primero, a pesar de tener buen accuracy, tiene errores a la hora de clasificar, ya que se le dieron muy pocos datos en los que una persona sigue viva. Por lo que se puede notar en la matriz de confusión que está sesgada.

El segundo modelo al tener muy pocos datos no es tan preciso para clasificar cada uno de los datos

El tercer modelo es el que está mejor, ya que al haberle realizado data augmentation contamos con más datos para que el modelo aprenda mejor.

MEJORAS A FUTURO

Lo más importante a mencionar es la obtención de más datos para la mejora de los datos de entrenamiento y validación. Como podemos ver en la gráfica, hay aproximadamente la mitad de datos que



corresponden a personas que sobrevivieron. Lo que causa que el modelo esté sesgado y, por lo tanto, no obtengamos los mejores resultados