

# Detección de fracturas de las vértebras cervicales en radiografías

---

## Integrantes:

- **Sebastián Aristondo Pérez** - 20880
- **José Daniel Gonzalez Carrillo** - 20293

## Situación problemática

---

La espina cervical es una estructura dinámica que se encarga de la protección de la inervación nerviosa, que es la unión de fibras que permiten la transmisión de información entre nervios del sistema periférico. También, da movilidad al cuello y a la cabeza, además de proteger la médula espinal. La espina cervical se divide en tres grupos en términos anatómicos y funcionales. Estos son C1, C2, que son el axis y el atlas, y C3-C7, que son las vértebras cervicales inferiores. (McMordie et al., 2021)

Según (RSNA, 2022), ocurren más de 1.5 millones de fracturas de las vértebras al año en Estados Unidos. Siendo las vértebras cervicales las más afectadas, principalmente hombres. La mayoría de los accidentes se dan por choques de vehículos, bicicletas, buceo e incidentes con armas de fuego. Al quebrarse una vértebra cervical, se compromete el bombeo de sangre al cerebro, la movilidad de la persona e incluso puede causar la muerte. Este último factor tiene de 5-10% de incidencia. (Naguib et al., 2023)

En general, el proceso de diagnóstico de una cervical rota es complejo. Se pueden usar diferentes métodos, entre los que se encuentran radiografías o rayos X, los cuales pueden no ser totalmente certeros debido a fracturas pequeñas. Los dos métodos más utilizados para la detección de fracturas son la tomografía computarizada y la resonancia magnética. (Martín-Ferrer, 2006)

## Problema científico

---

El área cervical está compuesta por estructuras óseas y blandas las cuales cumplen roles biológicos críticos para la supervivencia del cuerpo. Dentro de estos funcionamientos se encuentra el soporte de la cabeza, el suministro de sangre al cerebro y la protección de la médula espinal. Debido a sus responsabilidades biológicas cualquier enfermedad o lesión en el área puede resultar en la muerte. (Hermanns, 2002)

Por esta razón es necesario el diagnóstico temprano y certero de lesiones en la columna cervical. Cuando se sospecha que un paciente tiene politraumatismo es necesario realizar una cribado de imágenes y ultrasonidos. Tras realizar las radiografías y rayos X el médico o radiólogo será el encargado de interpretarlas para poder dar un diagnóstico. Lastimosamente las estadísticas muestran que el 80% de los errores en los diagnósticos en las unidades de emergencias se deben a interpretaciones equivocadas de médicos novatos o doctores que llevan muchas horas en turno. (Hermanns, 2002)

Con el objetivo de mitigar los errores en los diagnósticos, se buscará construir un sistema de diagnóstico basado en inteligencia artificial para brindar un apoyo y segunda opinión a los doctores al momento de interpretar las imágenes.

## Objetivos

---

El **objetivo principal** del estudio es:

- Determinar la probabilidad de que exista una fractura cervical, dada una tomografía computarizada del paciente, por medio de algoritmos de visión artificial.

Para cumplir esto se definieron los siguientes **objetivos específicos**:

- Efectuar un preprocesamiento en las imágenes de las tomografías computarizadas para estandarizarlas.
- Desarrollar un modelo de aprendizaje profundo que sea capaz de identificar fracturas en las cervicales.

## Descripción de tareas de limpieza

---

- Buscar NA's en los datos
- Buscar datos duplicados
- Buscar datos atípicos
- Evaluar desequilibrio en los datos
- Condiciones de datos sensibles o privados

## Descripción de datos

---

Descripción de datos

A continuación se hace una descripción de los archivos que se tienen, debido a que solo describir el archivo csv que se tiene es muy poco para la dimensión del proyecto.

- Segmentations: Esta es una carpeta que tiene archivos NIFTI, los cuales son archivos anotados por una red neuronal y validados por radiólogos, que determinan que vértebra se observa en una imagen de tomografía computarizada.
- Train images: Se tiene un conjunto de 711601 imágenes, las cuales están organizadas en carpetas por ID de observación. Cada carpeta tiene en promedio 352 imágenes, que son tomografías computarizadas.
- Train.csv: Este archivo contiene información sobre los ID de observación (posiblemente cada paciente) y 8 columnas más. La columna `patient_overall`, tiene la probabilidad de que un paciente tenga alguna cervical rota. Luego tenemos 7 columnas más, que son `C1-C7`. El valor de estas es la probabilidad de que cada vértebra esté rota. Todas las variables, a excepción de ID, son numéricas.

# Análisis exploratorio

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import os
```

```
root = '/kaggle/input/rsna-2022-cervical-spine-fracture-detection/'
data = pd.read_csv(root + 'train.csv')
```

```
data.head()
```

	StudyInstanceUID	patient_overall	C1	C2	C3	C4	C5	C6	C7
0	1.2.826.0.1.3680043.6200	1	1	1	0	0	0	0	0
1	1.2.826.0.1.3680043.27262	1	0	1	0	0	0	0	0
2	1.2.826.0.1.3680043.21561	1	0	1	0	0	0	0	0
3	1.2.826.0.1.3680043.12351	0	0	0	0	0	0	0	0
4	1.2.826.0.1.3680043.1363	1	0	0	0	0	1	0	0

```
data.describe()
```

	patient_overall	C1	C2	C3	C4	C5	C6
count	2019.000000	2019.000000	2019.000000	2019.000000	2019.000000	2019.000000	2019.000000
mean	0.475978	0.072313	0.141159	0.036157	0.053492	0.080238	0.137197
std	0.499546	0.259070	0.348272	0.186726	0.225068	0.271728	0.344140
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000



```
data.dtypes
```

```
StudyInstanceUID    object
patient_overall      int64
C1                   int64
C2                   int64
C3                   int64
C4                   int64
C5                   int64
C6                   int64
C7                   int64
dtype: object
```

```
print("La cantidad de personas que formamron parte del estudio es de: ", data.shape[0])
```

La cantidad de personas que formamron parte del estudio es de: 2019

```
print("La cantidad de datos falta es de: ", data.isnull().sum().sum())
```

La cantidad de datos falta es de: 0

```
print("La cantidad de id's repetidos es de : ", data.duplicated().sum())
```

La cantidad de id's repetidos es de : 0

```
columnas_cervicales = ["C1", "C2", "C3", "C4", "C5", "C6", "C7"]
data_cervicales = data[columnas_cervicales]
```

## Posibles cervicales dañadas

Primero haremos un análisis sobre aquellas radiografías donde se cree que hay una posible fractura en las cervicales. Para esto, se hará un análisis de las radiografías que tienen un valor de distinto a en la columna de cada `cervical`.

```
incidencia_de_fractura = data_cervicales[data_cervicales !=0].count()
print(incidencia_de_fractura)
```

```
C1    146
C2    285
C3     73
C4    108
C5    162
C6    277
C7    393
dtype: int64
```

## Cervicales dañadas

Ahora veremos aquellas cervicales que en definitiva están dañadas. Para esto, se hará un análisis de las radiografías que tienen un valor de 1 a en la columna de cada `cervical`.

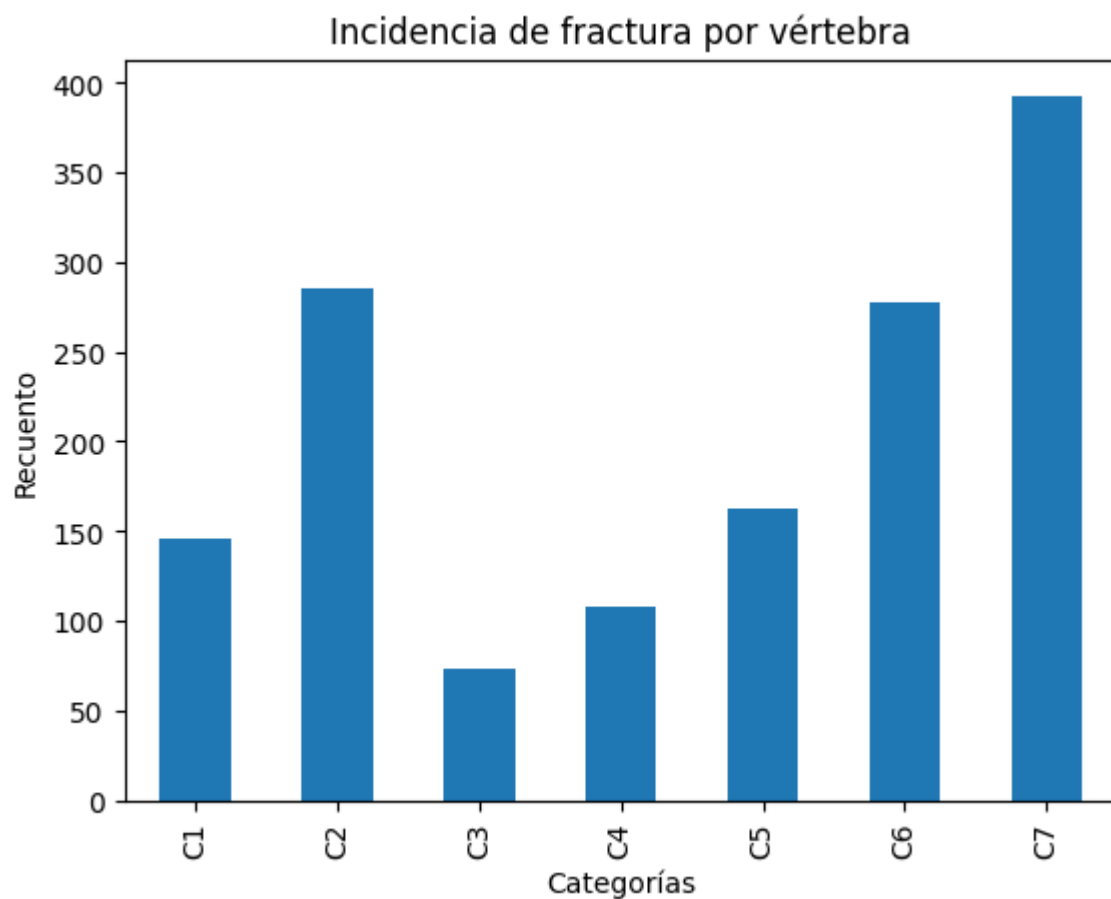
```
incidencia_de_fractura = data_cervicales[data_cervicales == 1].count()  
print(incidencia_de_fractura)
```

```
C1    146  
C2    285  
C3     73  
C4    108  
C5    162  
C6    277  
C7    393  
dtype: int64
```

A pesar de que la descripción de los datos nos indica que algunas cervicales pueden tener un valor entre 0 y 1 indicando que hay una `probabilidad` de fractura, podemos ver que el dataset unicamente cuenta con valores de 1 o 0 por lo que no hay fracturas sin diagnosticar.

## Frecuencia de fractura por cervical

```
incidencia_de_fractura.plot(kind='bar')  
  
# Agregar etiquetas y título  
plt.xlabel('Categorías')  
plt.ylabel('Recuento')  
plt.title('Incidencia de fractura por vértebra')  
  
# Mostrar el gráfico  
plt.show()
```



Se puede observar que la mayor cantidad de pacientes tienen una fractura en la séptima vértebra, con más de 400 ocurrencias. La menor cantidad se da en la tercera vértebra, con menos de 100 ocurrencias.

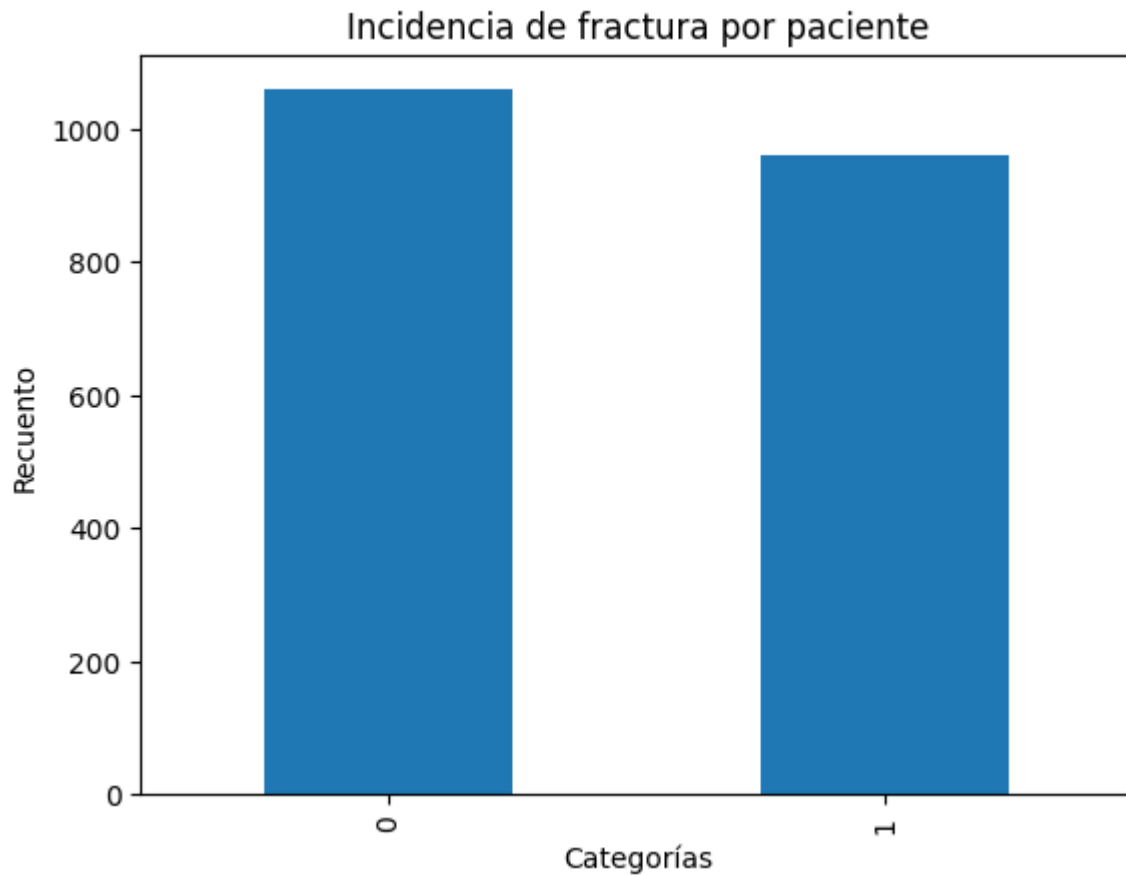
## Observaciones con fracturas

Ahora veremos cuántas observaciones tienen fractura y cuáles no.

```
fractura = data['patient_overall'].value_counts()

fractura.plot(kind='bar')
plt.xlabel('Categorías')
plt.ylabel('Recuento')
plt.title('Incidencia de fractura por paciente')

# Mostrar el gráfico
plt.show()
```



Se puede observar que las observaciones con fractura y las que no tienen fractura están balanceadas.

Ya vimos que hay algunas cervicales que son más propensas a sufrir fracturas, pero ahora veremos si hay cierta combinación de cervicales que se fracturan más que otras.

```
dicc_combinacion = {}  
df = data.copy()  
  
def combinatoria(fila):  
    acu = ""  
    for c in columnas_cervicales:  
        if df.loc[fila, c] == 1:  
            acu += str(c) + "-"  
  
    # Elimina el guion adicional al final de la cadena  
    acu = acu.rstrip('-')  
  
    # Guarda la cadena en el diccionario con la fila como clave  
  
    if acu != "":  
        if acu not in dicc_combinacion:  
            dicc_combinacion[acu] = 1  
        else:  
            dicc_combinacion[acu] += 1  
  
for i in range(len(df)):
```

```
combinatoria(i)
```

```
# Imprime el diccionario con las combinaciones para la fila 0
lista_de_tuplas = [(llave, valor) for llave, valor in dicc_combinacion.items()]
lista_de_tuplas_ordenadas = sorted(lista_de_tuplas, key=lambda x: x[1])
diccionario_ordenado = dict(lista_de_tuplas_ordenadas)

for key in diccionario_ordenado:
    print("Combinación: ", key, " - Cantidad: ", diccionario_ordenado[key])
```

```
Combinación: C2-C3-C4 - Cantidad: 1
Combinación: C3-C4-C5 - Cantidad: 1
Combinación: C1-C2-C3 - Cantidad: 1
Combinación: C3-C5-C6-C7 - Cantidad: 1
Combinación: C1-C2-C3-C5 - Cantidad: 1
Combinación: C1-C2-C4 - Cantidad: 1
Combinación: C2-C5-C6-C7 - Cantidad: 1
Combinación: C3-C4-C6 - Cantidad: 1
Combinación: C2-C4 - Cantidad: 1
Combinación: C1-C2-C5 - Cantidad: 1
Combinación: C1-C4 - Cantidad: 1
Combinación: C1-C3-C5-C6 - Cantidad: 1
Combinación: C2-C3-C4-C7 - Cantidad: 1
Combinación: C1-C3-C5 - Cantidad: 1
Combinación: C2-C3-C4-C5 - Cantidad: 1
Combinación: C4-C5-C7 - Cantidad: 1
Combinación: C2-C4-C5-C6 - Cantidad: 1
Combinación: C3-C6-C7 - Cantidad: 1
Combinación: C1-C6 - Cantidad: 1
Combinación: C1-C3-C5-C7 - Cantidad: 1
Combinación: C2-C3-C5 - Cantidad: 1
Combinación: C2-C4-C5 - Cantidad: 1
Combinación: C1-C2-C4-C7 - Cantidad: 1
Combinación: C2-C4-C5-C7 - Cantidad: 1
Combinación: C4-C6 - Cantidad: 1
Combinación: C3-C6 - Cantidad: 1
Combinación: C1-C3-C4-C5-C6 - Cantidad: 1
Combinación: C3-C7 - Cantidad: 1
Combinación: C2-C4-C6-C7 - Cantidad: 1
Combinación: C1-C2-C6-C7 - Cantidad: 1
Combinación: C3-C4-C5-C6 - Cantidad: 2
Combinación: C1-C5-C6-C7 - Cantidad: 2
Combinación: C2-C3-C6 - Cantidad: 2
Combinación: C1-C2-C3-C7 - Cantidad: 2
Combinación: C1-C5-C6 - Cantidad: 2
Combinación: C4-C6-C7 - Cantidad: 2
Combinación: C1-C4-C5-C6 - Cantidad: 2
Combinación: C2-C5 - Cantidad: 2
Combinación: C2-C3-C7 - Cantidad: 2
Combinación: C1-C2-C6 - Cantidad: 2
```



Combinación: C2-C3-C4-C5-C6-C7 - Cantidad: 2  
Combinación: C2-C3-C4-C5-C6 - Cantidad: 3  
Combinación: C3-C4-C5-C6-C7 - Cantidad: 3  
Combinación: C2-C6 - Cantidad: 3  
Combinación: C1-C6-C7 - Cantidad: 3  
Combinación: C3-C5 - Cantidad: 3  
Combinación: C1-C2-C7 - Cantidad: 4  
Combinación: C1-C5 - Cantidad: 4  
Combinación: C4-C7 - Cantidad: 5  
Combinación: C5-C7 - Cantidad: 5  
Combinación: C4-C5-C6-C7 - Cantidad: 6  
Combinación: C4-C5-C6 - Cantidad: 6  
Combinación: C2-C3 - Cantidad: 7  
Combinación: C1-C7 - Cantidad: 7  
Combinación: C2-C7 - Cantidad: 9  
Combinación: C2-C6-C7 - Cantidad: 9  
Combinación: C3-C4 - Cantidad: 13  
Combinación: C3 - Cantidad: 18  
Combinación: C4-C5 - Cantidad: 20  
Combinación: C5-C6-C7 - Cantidad: 21  
Combinación: C5 - Cantidad: 28  
Combinación: C4 - Cantidad: 28  
Combinación: C5-C6 - Cantidad: 36  
Combinación: C1-C2 - Cantidad: 42  
Combinación: C1 - Cantidad: 64  
Combinación: C6-C7 - Cantidad: 77  
Combinación: C6 - Cantidad: 82  
Combinación: C2 - Cantidad: 180  
Combinación: C7 - Cantidad: 223

El dataset cuenta con 2019 personas, por lo tanto tomaremos como fracturas frecuentes aquellas que sean 1% de la población. Bajo este criterio podemos ver que las combinaciones más comunes de fractura son:

Cervicales: C4-C5 - Cantidad: 20

Cervicales: C5-C6-C7 - Cantidad: 21

Cervicales: C5 - Cantidad: 28

Cervicales: C4 - Cantidad: 28

Cervicales: C5-C6 - Cantidad: 36

Cervicales: C1-C2 - Cantidad: 42

Cervicales: C1 - Cantidad: 64

Cervicales: C6-C7 - Cantidad: 77

Cervicales: C6 - Cantidad: 82

Cervicales: C2 - Cantidad: 180

Cervicales: C7 - Cantidad: 223

```
# Ruta de la carpeta principal que contiene las subcarpetas con los elementos
ruta_principal = root + 'train_images'

# Inicializar un diccionario para llevar un registro de la cantidad de elementos por carpeta
elementos_por_carpeta = {}

# Recorre las carpetas y cuenta los elementos
for carpeta, _, archivos in os.walk(ruta_principal):
    cantidad_elementos = len(archivos)
    elementos_por_carpeta[carpeta] = cantidad_elementos

cantidad = 0
for elemento in elementos_por_carpeta:
    cantidad += elementos_por_carpeta[elemento]
print(f"Hay {cantidad} imágenes.")
print(f"En promedio las carpetas tienen {cantidad // len(elementos_por_carpeta)} imágenes.")
```

Hay 711601 imágenes.

En promedio las carpetas tienen 352 imágenes.

## Hallazgos

No todas las cervicales tienen la misma frecuencia de fractura, tras los análisis se pudo obtener que la cervical tiende a fracturarse. El orden es el siguiente C7 con 393 fracturas, C2 con 285, C6 con 277, C5 con 162, C1 con 146, C4 con 108, y C3 con 73. Muchos de los datos mostraban que los pacientes tenían más de la cervical lastimada, por ello se buscó si existía algún tipo de patrón o dependencias entre las cervicales dañadas. Se pudo concluir que algunas cervicales si tienden a fracturarse al mismo tiempo como lo son la C5-C6, C1-C2 y C6-C7, estas son las fracturas múltiples más frecuentes aunque cabe destacar que estas son superadas por las fracturas individuales de la C6, C2 Y C7

Se encontró que hay 711601 imágenes en total de la tomografía computarizada de las cervicales. Estas imágenes están distribuidas en los distintos ID de observaciones. En promedio se tienen 352 imágenes por cada ID. Todas las imágenes están en formato .dcm y todas tienen un tamaño de 512x512 píxeles. Además, se tienen 87 segmentaciones de vértebras en formato NIFTI, las cuales servirán para entrenar un modelo de segmentación que generalice a todas las imágenes del conjunto de datos. Por otra parte, también se identificó que las segmentaciones no solo tienen cervicales, si no que también vértebras torácicas.

Finalmente, se puede mencionar que en el conjunto de datos no existen fracturas potenciales en los estudios realizados, si no que siempre son garantizadas. Esto se evidencia en que siempre hay un 100% de probabilidad de que la vértebra esté fracturada, en los pacientes que la tienen.

## Conclusiones de pasos a seguir

A partir de este análisis exploratorio inicial, se pudo identificar que se tiene un reto bastante grande para realizar la clasificación de las cervicales rotas. Para poder llevar a cabo la predicción de probabilidad de estas y cumplir con los objetivos, se llevarán a cabo distintas acciones. Estas consisten básicamente en preprocesamiento de imágenes y entrenamiento de modelos de aprendizaje profundo. Estas acciones son las siguientes:

- Muestra de datos: Determinar el tamaño de muestra que se utilizará para entrenar el modelo.
- Estandarizar color: Convertir las imágenes a una escala de grises.
- Preprocesamiento de imágenes: Crear un modelo de aprendizaje profundo para el preprocesamiento de las imágenes. Este modelo creará máscaras, permitiendo eliminar el ruido de la tomografía.
- Clasificación de datos: Tras aplicar las máscaras a las imágenes es necesario asignarles labels para poder distinguir qué sección de la columna se está viendo.
- Limpieza de datos: Una vez las imágenes tienen sus respectivos labels es necesario eliminar las tomografías que no son de las cervicales.
- Determinar qué arquitectura de Deep Learning se utilizará para realizar la clasificación y si se usará un modelo propio o pre entrenado.

## Referencias

---

Hermanns, M. (2002). Parallel Programming in Fortran 95 using OpenMP. [https://www.openmp.org/wp-content/uploads/F95\\_OpenMPv1\\_v2.pdf](https://www.openmp.org/wp-content/uploads/F95_OpenMPv1_v2.pdf)

Martín-Ferrer, S. (2006). Traumatismos de la columna cervical alta: Clasificación tipológica, indicaciones terapéuticas y abordajes quirúrgicos (a propósito de 286 casos). *Neurocirugía*, 17(5), 391–419. [https://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1130-14732006000500001](https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1130-14732006000500001)

McMordie, J. H., Viswanathan, V. K., y Gillis, C. C. (2021). Cervical Spine Fractures Overview. PubMed; StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK448129/>

Naguib, S. M., Hamza, H. M., Hosny, K. M., Saleh, M. K., y Kassem, M. A. (2023). Classification of Cervical Spine Fracture and Dislocation Using Refined Pre-Trained Deep Model and Saliency Map. *Diagnostics*, 13(7), 1273–1273. <https://doi.org/10.3390/diagnostics13071273>

RSNA. (2022, Julio 28). RSNA 2022 Cervical Spine Fracture Detection. Kaggle.com. <https://www.kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection/overview>