



universidad
de león



Escuela de Ingenierías Industrial, Informática y Aeroespacial

GRADO EN INGENIERÍA INFORMÁTICA

Trabajo de Fin de Grado

Aplicación de técnicas de inteligencia artificial para
la predicción de la supervivencia y el grado del
glioma

Use of artificial intelligence techniques to predict the
survival and the grade of the glioma

Autor: Santiago Valbuena Rubio
Tutor: María Teresa García Ordás
Cotutor: Óscar García-Olalla Olivera

(Junio, 2022)



UNIVERSIDAD DE LEÓN
Escuela de Ingenierías Industrial, Informática y
Aeroespacial

GRADO EN INGENIERÍA INFORMÁTICA
Trabajo de Fin de Grado

ALUMNO: Nombre y apellidos del alumno

TUTOR: Nombre y apellidos del alumno

TÍTULO: Título del Trabajo

TITLE: Title of the work

CONVOCATORIA: Mes, Año

RESUMEN:

El resumen reflejará las ideas principales de cada una de las partes del trabajo, pudiendo incluir un avance de los resultados obtenidos. Constará de un único párrafo y se recomienda una longitud no superior a 300 palabras. En cualquier caso, no deberá superar una página de longitud.

ABSTRACT:

Palabras clave: Lorem, ipsum, dolor, sit, amet.

Firma del alumno:

VºBº Tutor/es:



Índice de contenidos

1	Introducción.....	8
1.1	GLIOMA Y GLIOBLASTOMA	8
1.2	INTELIGENCIA ARTIFICIAL	9
1.3	APRENDIZAJE PROFUNDO	9
1.4	MÉTRICAS PRINCIPALES	10
1.5	TRANSFERENCIA DE APRENDIZAJE.....	12
1.5.1	REDES PRE-ENTRENADAS	12
2	Estado del arte	13
3	Planificación	15
3.1	Diagrama Gantt.....	15
4	Núcleo del trabajo.....	16
4.1	CONJUNTO DE DATOS.....	16
4.1.1	COMPOSICIÓN.....	16
4.1.2	IMÁGENES	16
4.1.2.1	FORMATO DE LOS ARCHIVOS.....	16
4.1.2.2	TIPOS DE IMÁGENES	18
4.1.3	EL PROBLEMA DE LA FALTA DE DATOS	20
4.2	PRE-PROCESADO DE LOS DATOS	20
4.2.1	TRATAMIENTO DE DATOS FALTANTES.....	21
4.2.2	TRANSFORMACIÓN DE LOS DATOS	21
4.2.3	TRATAMIENTO DE LAS IMÁGENES.....	22
4.2.3.1	RE-ORIENTACIÓN DE LAS IMAGENES	22
4.2.3.2	NORMALIZACIÓN DE LAS IMÁGENES	22
4.2.3.3	SEGMENTACIÓN DE LAS IMÁGENES	23
4.3	ARQUITECTURA DE LA RED.....	24
4.4	PRUEBAS DE MODELOS PRE-ENTRENADOS	25
4.4.1	ResNet101	26
4.4.2	ResNet50	27
4.4.3	EfficientNetB4.....	29
4.4.4	VGG16	30



4.4.5	InceptionV3	32
4.4.6	InceptionResNetV2	33
4.4.7	DenseNet121	34
4.4.8	DenseNet201	36
4.4.9	ELECCIÓN DEL MEJOR MODELO	38
4.5	OPTIMIZACIÓN DEL MEJOR MODELO.....	38
4.5.1	BATCH NORMALIZATION	39
4.5.1.1	DOS CAPAS.....	40
4.5.1.2	UNA CAPA	41
4.5.1.3	SIN CAPAS	43
4.5.2	NEURONAS	44
4.5.2.1	32-64 NEURONAS	44
4.5.2.2	64-128 NEURONAS	44
4.5.2.3	128-256 NEURONAS	46
4.5.2.4	256-256 NEURONAS	47
4.5.3	DROPOUT	48
4.5.3.1	DROPOUT 0.2.....	48
4.5.3.2	DROPOUT 0.3.....	49
4.5.3.3	DROPOUT 0.4.....	49
4.5.3.4	DROPOUT 0.5.....	50
4.5.4	FUNCIÓN DE ACTIVACIÓN	51
4.5.4.1	RELU	52
4.5.4.2	TANH	52
4.5.5	TASA DE APRENDIZAJE.....	53
4.5.5.1	TASA DE APRENDIZAJE 0.0005	53
4.5.5.2	TASA DE APRENDIZAJE 0.001	54
4.5.5.3	TASA DE APRENDIZAJE 0.002	55
4.5.6	RESTO DE PARÁMETROS	56
5	Resultados.....	56
6	Conclusiones.....	57
7	Lista de referencias bibliográficas	58
8	Anexos	60



Índice de figuras

Figura 3.1 Organización Trello. Fuente [4]	¡Error! Marcador no definido.
Figura 3.2 Diagrama de Gantt de la planificación. Fuente propia.....	15
Figura 5.1 Matriz de confusión. Fuente propia.	¡Error! Marcador no definido.
Figura 5.2 Imágenes ejemplo dataset 1. Fuente propia.	¡Error! Marcador no definido.
Figura 5.3 Arquitectura red convolucional. Fuente propia..	¡Error! Marcador no definido.
Figura 5.4 Recall del modelo. Fuente propia.	¡Error! Marcador no definido.
Figura 5.5 Imágenes ejemplo dataset 2. Fuente propia.	¡Error! Marcador no definido.
Figura 5.6 Numero de imágenes clases dataset 2. Fuente propia.	¡Error! Marcador no definido.
Figura 5.7 Arquitectura red convolucional dataset 2. Fuente propia.	¡Error! Marcador no definido.



Índice de cuadros y tablas

Tabla 5.1 Ratio de valores nulos en los datos. Fuente propia;**Error!** **Marcador** **no**
definido.

Tabla 5.2 Tabla de datos dataset 1. Fuente propia.....;**Error!** **Marcador no definido.**



Glosario de signos

coeficientes de atenuación[23]

1 Introducción

1.1 GLIOMA Y GLIOBLASTOMA

El cáncer es una de las principales causas de muerte en el mundo con más de 18 millones de casos y 9,5 millones de muertes en 2018, pero se estima que estas cifras empeorarán aún más hasta los 29,5 millones de casos y 16,4 millones de muertes en el año 2040 [1]. Los gliomas malignos son los tumores cerebrales más comunes, con distintos grados de agresividad y distintas regiones donde pueden aparecer [2]. La clasificación establecida por la Organización Mundial de la Salud (OMS) [3], divide los gliomas en cuatro tipos, Astrocytomas, Oligodendrogliomas, Ependymomas y Oligo-astrocytomas. Cada uno de estos tipos es dividido en fases, las cuales tienen en cuenta características como la dispersión del tumor por el resto de los órganos o nódulos linfáticos, el tamaño del tumor o el nivel de penetración.

El sistema de estadificación del tumor más común es el TNM [4], T se refiere al tumor original, N indica que el cáncer se ha propagado a los nódulos linfáticos y M indica que el cáncer se ha propagado y se ha producido metástasis.

Dependiendo de la malignidad los tumores de tipo Astrocytomas se subdividen en [5]:

- Pilocytic, grado I
- Diffuse – grado II
- Anaplastic – grado III
- Glioblastoma multiforme – grado IV

Este último, el Glioblastoma multiforme es el tipo de glioma más común, afectando a un 60-70% de los casos, seguido por el Anaplastic Astrocytoma con un 10-15% de los casos [5]. El término de multiforme hace referencia a la heterogeneidad de este tumor el cual puede adoptar distintas formas y encontrarse en distintas regiones del cerebro. El glioblastoma multiforme puede surgir desde un tumor de menor grado o puede aparecer sin ningún tumor previo. Tiene una tasa de supervivencia a 5 años del 22% en personas de entre 20-44 años, del 9% en personas de entre 45-54 años y en tan solo un 6% en personas

de 55-64 años [6], siendo el tumor cerebral con menos porcentaje de supervivencia de todos los gliomas.

Sin tratamiento la supervivencia al glioblastoma multiforme es de unos 3 a 6 meses, por lo que como todos los tumores, pero en especial este tan maligno, un pronto diagnóstico puede aumentar las posibilidades de supervivencia. Los tratamientos incluyen quimioterapia y radioterapia, pero el que ha demostrado un mayor aumento en la esperanza de supervivencia es la resección del tumor, la cual se clasifica en dos [7]:

- GTR (Gross Total Resection): Eliminar el tumor completamente, aunque dependiendo de la localización y estado del glioma no siempre es posible.
- STR (Sub Total Resection): Eliminación parcial del tumor.

Ambas técnicas son importantes en nuestro estudio ya que el conjunto de datos utilizado para entrenar los distintos modelos de aprendizaje cuenta con este dato, en el que se informa si el paciente se realizó una cirugía parcial, total o no se realizó cirugía.

El glioblastoma multiforme (GBM) es clasificado como High Grade Glioma (HGG), mientras que el resto de los tumores de menor grado son clasificados como Low Grade Glioma (LGG) [8]. Esta es la clasificación utilizada para entrenar los modelos descritos más adelante.

1.2 INTELIGENCIA ARTIFICIAL

La inteligencia artificial tiene como objetivo el desarrollo de programas capaces de realizar tareas que realiza el ser humano intentando incluso superar la capacidad humana para realizar dicha tarea.

La inteligencia artificial es un término muy amplio que engloba numerosas áreas y técnicas. Una de ellas es el aprendizaje automático el cual consiste en el desarrollo y estudio de técnicas que pueden aprender automáticamente a partir de unos datos, mediante el ajuste de una ecuación matemática con el objetivo de optimizar unas métricas. Dentro del aprendizaje automático se encuentra el aprendizaje profundo, el cual nace a mediados del siglo XX con el estudio de los algoritmos que dan a las computadoras la capacidad de aprender de los ejemplos y la experiencia en vez de estar explícitamente programados.

1.3 APRENDIZAJE PROFUNDO

El aprendizaje profundo es una rama del aprendizaje automático el cual se centra en aprender representaciones de datos como capas sucesivas de representaciones con un significado cada vez mayor. Estas redes están formadas por capas de neuronas las cuales se encuentran interconectadas entre sí y mediante el desarrollo de filtros en cada una de las capas se consiguen realizar filtros más complejos conforme más profundizas en la red, permitiendo extraer características en el caso de las redes convolucionales cada vez más complejas.

Estas redes son capaces de aprender con la experiencia mediante el entrenamiento a partir de los datos iniciales y conseguir resultados superiores en la mayoría de los casos que las técnicas de aprendizaje automático convencionales gracias a su capacidad de aprendizaje, a su no linealidad y a su arquitectura. Una de las principales ventajas es su capacidad para extraer las características de las imágenes sin necesidad de extraerlas manualmente como se hacía tradicionalmente con los descriptores.

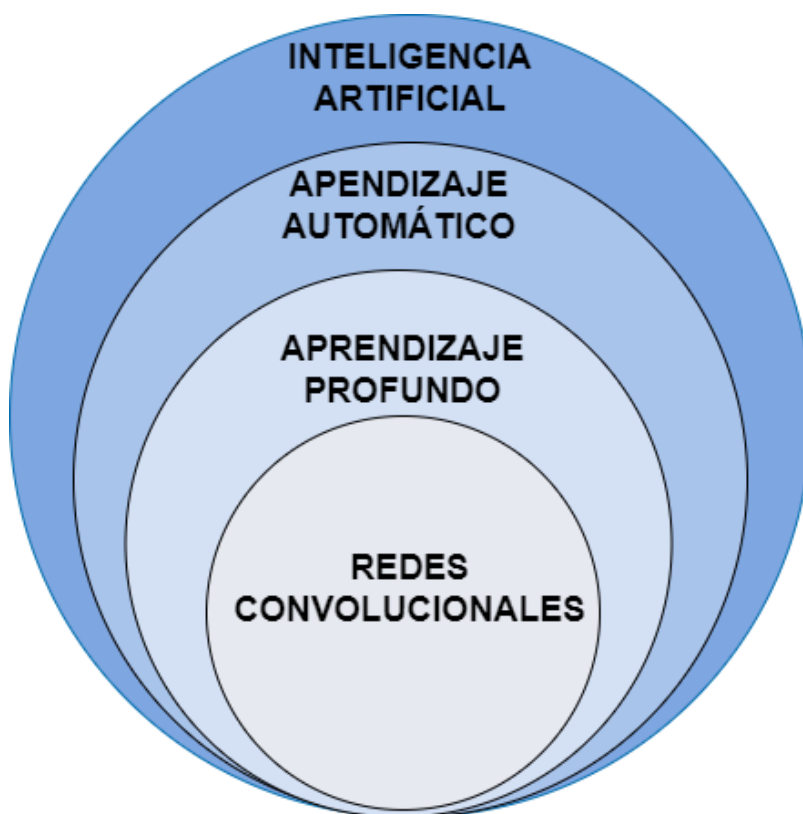


FIGURA X

1.4 MÉTRICAS PRINCIPALES

Dependiendo del tipo de problema las métricas utilizadas serán distintas. Como el problema que vamos a abordar es un problema de clasificación, las métricas que usaremos serán obtenidas a partir de la matriz de confusión. La matriz de confusión representa como se han predicho los datos respecto a su valor real. La matriz estará formada por 4 valores, en el caso de que tengamos dos clases, como vemos en la **figura X**:

- TP (True Positive): Son los datos que han sido predichos como positivos y realmente eran positivos.
- FP (False Positive): Son los datos que han sido predichos como positivos, pero realmente eran negativos.
- TN (True Negative): Son los datos que han sido predichos como negativos y realmente eran negativos.
- FN (False Negative): Son los datos que han sido predichos como negativos y realmente eran positivos.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Fuente <https://www.r-bloggers.com/2020/12/weighting-confusion-matrices-by-outcomes-and-observations-2/>

A partir de esta matriz obtendremos las métricas que usaremos para la clasificación, las cuales serán las siguientes:

- Accuracy: Mide el porcentaje total de elementos clasificados correctamente.
- Recall: Mide el porcentaje total de elementos clasificados como positivos sobre el total de positivos.

- Precision: Mide el porcentaje total de elementos clasificados como positivos sobre el total de elementos clasificados como positivos.
- F1-Score: Es la media armónica de la precisión y el recall.

1.5 TRANSFERENCIA DE APRENDIZAJE

La transferencia de aprendizaje o en inglés “transfer learning”, es el conjunto de técnicas que permiten transferir conocimientos adquiridos en la resolución de unos problemas para resolver otros problemas diferentes. Estas técnicas son utilizadas para reentrenar redes convolucionales que han sido entrenadas en un conjunto de datos diferente pero que han aprendido muy bien a extraer características de las imágenes y pueden ser utilizadas para otro problema diferente. La principal ventaja es el ahorro de recursos ya que el modelo ya se encuentra entrenado y se necesitarán muchas menos iteraciones y muchos menos datos para conseguir buenos resultados en comparación con las redes entrenadas desde cero. Gracias a la transferencia de aprendizaje podemos abordar nuestro problema, ya que solamente contamos con unos cientos de imágenes para entrenar los modelos y no sería posible sin ella, ya que se necesitan muchos más datos para entrenar una red desde cero obteniendo tan buenos resultados.

1.5.1 REDES PRE-ENTRENADAS

Resnet101, 50, EfficientNetB4, InceptionV3, InceptionResnetv2, VGG16, DenseNet121, 201

Las redes convolucionales se centraron inicialmente en la clasificación de imágenes, así como en la detección de objetos. Todas las redes que se describen a continuación han sido entrenadas con ImageNet [14], la cual es una base de datos que cuenta con más de 14 millones de imágenes utilizadas para el entrenamiento y la validación de las redes.

Se han seleccionado distintos modelos pre-entrenados que han conseguido muy buenos resultados en la clasificación de las imágenes de ImageNet [14] y que están integrados con la librería Keras para poder ser utilizados fácilmente. Las redes que utilizaremos serán: ResNet50, ResNet101, EfficientNetB4, InceptionV3, InceptionResNetV2, VGG16, DenseNet121, DenseNet201.

2 Estado del arte

El 2012 comenzó a realizarse una competición conocida como BraTS [9], la cual consiste en obtener la mejor puntuación en distintas pruebas como la segmentación del tumor o la clasificación del tiempo de supervivencia del paciente. Esta competición ofrece anualmente un conjunto de imágenes y otros datos, recopilados a través de distintas entidades médicas colaboradoras con la competición. Estos datos son ofrecidos para la competición y para el resto de investigadores. Estos datos son una fuente valiosísima para el entrenamiento de los distintos modelos de aprendizaje automático y son utilizados por gran parte de los trabajos previos de los que hablaremos a continuación, al igual que en este trabajo.

En estudios previos en los que se ha intentado predecir la supervivencia de los pacientes con glioblastoma se han utilizado combinaciones de técnicas de aprendizaje profundo con técnicas de aprendizaje clásicas, como en el caso del trabajo realizado por Lina y Shahram [10]. En su trabajo utilizaron distintos métodos para extraer las características de las imágenes y una vez extraídas fueron clasificadas en dos o tres clases diferenciando entre supervivientes de corto, medio o largo plazo, utilizando distintas técnicas de aprendizaje automático. En el caso de tres clases los mejores resultados fueron obtenidos utilizando “Complex and median tree” con un 62,5% de precisión. Y en el caso de la clasificación de dos clases entre superviviente de corto y largo plazo, los mejores resultados fueron obtenidos con regresión logística obteniendo un 68,8% de precisión.

En el trabajo [11] obtuvieron una precisión del 51,5% en la predicción de la supervivencia de los pacientes utilizando redes convolucionales, pero una vez más como en el caso anterior los mejores resultados fueron obtenidos con técnicas clásicas, concretamente usando un SVC (Support Vector Classifier) obteniendo un 72,2% de precisión en el conjunto de entrenamiento, un 57,1% en el conjunto de validación y un 42,9% en el conjunto de prueba.

Por otra parte, los trabajos orientados a la clasificación del grado del glioblastoma han obtenido resultados prometedores con unas mejores precisiones al ser una tarea más

sencilla que la de determinar la supervivencia del paciente a la que afectan muchos más factores.

En el trabajo [12] se realizó la extracción de 180 características y se obtuvo una precisión de un 89,81% utilizando técnicas de regresión logística.

Por otro lado, en el estudio [13], se realizaron ambas predicciones junto con la segmentación del tumor. En un primer lugar realizan la segmentación del tumor y sobre esta utilizan una red convolucional 3D para clasificar el tumor entre las distintas clases. Finalmente realizan una técnica híbrida al igual que los estudios anteriores entre aprendizaje profundo y aprendizaje tradicional para poder realizar la predicción de la supervivencia del paciente. En este estudio se obtuvo en el conjunto de prueba un 48,40% de precisión y en el de validación un 58,6% de precisión en la predicción de la supervivencia utilizando redes convolucionales para extraer características de las imágenes, y junto con la edad utilizando regresión lineal para obtener las predicciones.

Tras analizar el estado del arte podemos observar que la aproximación que ha obtenido unos resultados más prometedores es la utilización de técnicas híbridas entre el aprendizaje profundo y las técnicas clásicas y que existe un gran potencial de mejora de los modelos realizados hasta la fecha ya que las precisiones obtenidas son inferiores al 69% cuando se intenta realizar una clasificación del tiempo de supervivencia en dos clases, inferiores al 62,5% en el caso de tres clases de supervivientes, e inferiores al 59% cuando se intenta dar una predicción del tiempo estimado de supervivencia. Mientras que en la clasificación del tumor si se han obtenido unos mejores resultados, aunque aún se encuentran por debajo del 90%.

3 Planificación

3.1 Diagrama Gantt

Durante el mes de diciembre de 2021 se decidió cual sería la temática del TFG desarrollado y no fue hasta el 21 de febrero cuando se comenzó a desarrollar el trabajo. Lo principal al inicio fue la recopilación de trabajos previos y la obtención de ideas para desarrollar el trabajo. A continuación, los esfuerzos se centraron en tratar los datos ya que estos se encontraban sucios, en formatos médicos y sin tratar. Estos fueron limpiados y transformados a los formatos necesarios para el entrenamiento de los modelos. Las siguientes semanas fueron dedicadas al desarrollo de los modelos. En un primer momento fueron diseñados distintos modelos que usaban redes neuronales 3D, pero debido a la falta de datos las predicciones no fueron las esperadas, por lo que se decidió cambiar el enfoque y utilizar “transfer learning” para realizar las predicciones. Se probaron disantos modelos y se ajustaron los parámetros del modelo que mejor rendimiento mostró para obtener los mejores resultados.

Tras esto se redacto el documento y se realizaron las pruebas finales.

TAREAS/FECHAS	27/02/2022	06/03/2022	13/03/2022	20/03/2022	27/03/2022	03/04/2022	10/04/2022	17/04/2022	24/04/2022	01/05/2022
ESTADO DEL ARTE Y RECOPIACIÓN DE IDEAS E INFORMACIÓN										
TRATAMIENTO DE LOS DATOS										
DESARROLLO DEL MODELO										
EVALUACIÓN										
ELABORACIÓN DE LA MEMORIA										

Figura 3.1 Diagrama de Gantt de la planificación. Fuente propia.

4 Núcleo del trabajo

4.1 CONJUNTO DE DATOS

4.1.1 COMPOSICIÓN

El conjunto de datos utilizado es obtenido de las BraTS 2020 [15, 16, 17] la cual es una competición para la segmentación, clasificación del grado y clasificación de la supervivencia al glioma. El conjunto de datos está formado por 31GB con imágenes y datos de 369 pacientes. Para cada uno de estos pacientes podemos encontrar su edad, supervivencia en días y si se han realizado una resección de tipo GTR, STR o no se han realizado ninguna resección. Además, podemos encontrar 4 tipos de escáneres cerebrales diferentes: T1, T1ce, T2 y T2-Flair. Junto con la segmentación del tumor que no es utilizada en este estudio. No todos los pacientes cuentan con todos los datos como edad o supervivencia, por lo que es necesario tratarlos como explicaremos en posteriores secciones.

4.1.2 IMÁGENES

4.1.2.1 FORMATO DE LOS ARCHIVOS

Las imágenes se encuentran en archivos NifTI (.nii.gz) este es un formato para imágenes médicas en la que podemos encontrar la imagen junto con más información sobre la misma.

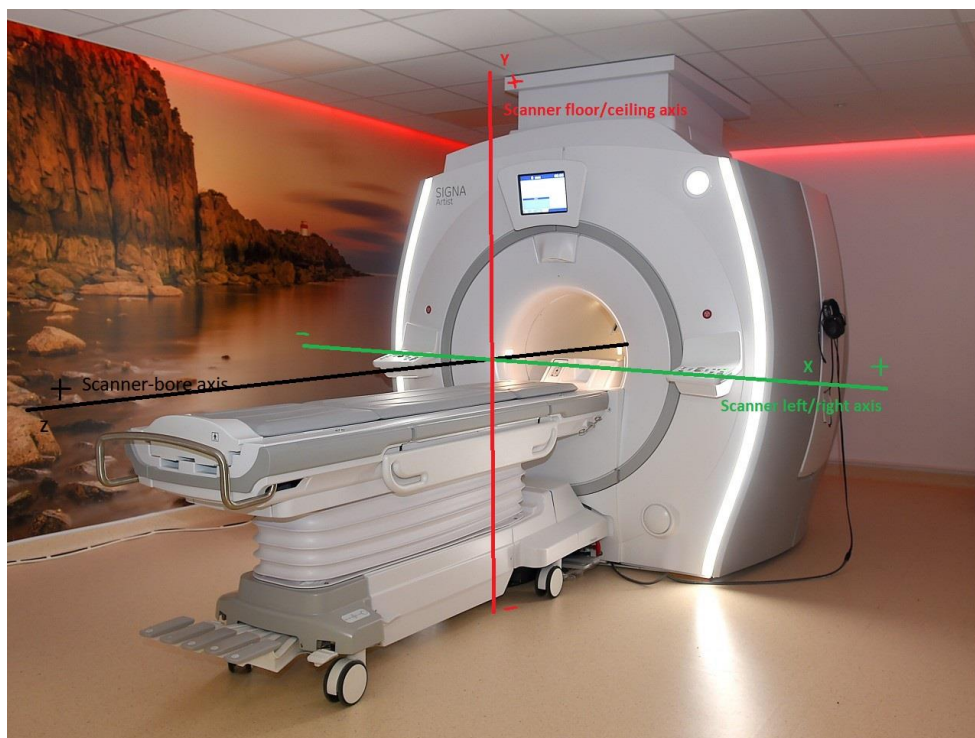
Cada imagen NifTI está formada por tres componentes.

- Una matriz N-D que contiene los datos de la imagen. En nuestro caso se trata de una matriz de 3-Dimensiones que contiene un mapeado de los cerebros de los pacientes. Gracias a esto podemos obtener cualquier región o corte del cerebro del paciente.
- Una matriz afín 4x4 con información sobre la posición y orientación de la imagen en un espacio determinado.
- Una cabecera con metadatos e información sobre la imagen.

Las imágenes en dos dimensiones están formadas por píxeles, las cuales son las unidades más pequeñas en las que podemos dividir una imagen. Cuando utilizamos imágenes en tres dimensiones existe un concepto similar a los píxeles, pero reciben el nombre de voxeles. Los voxeles al igual que los píxeles en dos dimensiones, constituyen la unidad mínima de un objeto tridimensional. Para crear una imagen tridimensional los voxeles deben de tener una opacidad diferente ya que sino solamente se observaría la capa más externa de la imagen.

La matriz afín nos permite conocer la orientación de la imagen y poder fijar una misma orientación a todas nuestras imágenes. Pero ¿Para qué? Cuando tenemos varias imágenes de una misma persona, probablemente queramos unir las para tener un modelo más completo del paciente. Las coordenadas afines nos permiten esto ya que nos dan un espacio de referencia para ambas imágenes. Si no fuera por esto, no sería posible, ya que cada imagen se encontraría en una orientación diferente y unas coordenadas de sus voxeles diferentes. ¿Qué es el espacio de referencia? Un espacio de referencia es un espacio común que nos permite comprender las dimensiones y orientaciones de un objeto. Para definir un espacio necesitamos definir el origen de sus ejes, sus direcciones y sus unidades.

Existen distintos espacios de referencia [19], los basados en el escáner y los basados en el sujeto. Los basados en el escáner también son conocidos como “XYZ”. Este espacio cuenta con un isocentro (el punto 0,0,0) en el que se cruzan los tres ejes y las unidades de los ejes son milímetros por lo que el punto (10, -5, -3) significa que se encuentra 10mm a la derecha del isocentro del observador, 5mm debajo del isocentro y 3mm entre los pies y el isocentro. Los ejes de este modelo son fijados con la posición del escáner, podemos ver una representación de ellos en la [figura x](#).



(Tengo que mejorar la foto)

Pero el modelo XYZ tiene un problema y es que la cabeza del paciente no siempre se encuentra en la misma posición dentro del scanner ya que puede por ejemplo estar apoyado hacia un costado. Es por esto por lo que este modelo fijo ha evolucionado a un modelo en el que los ejes dependen de la ubicación de la cabeza del sujeto. Estos modelos siguen manteniendo los mismos ejes que el modelo XYZ centrado en el escáner, pero el orden y la dirección de los ejes provienen de la orientación del sujeto.

El modelo más común es el modelo RAS, (Right Anterior Superior) en el que los ejes son reordenados de tal forma que el primer eje (X) será el eje más próximo al eje izquierda-derecha del paciente. Es decir, si el paciente se encuentra con la cabeza apoyada hacia la derecha el eje X será el eje suelo-techo, el Y será el posterior anterior del paciente es decir el verde en la **figura x** y el Z será el bore. En cuanto a los signos, los positivos serán el derecho (Right), anterior (Anterior) y superior (Superior) del paciente, de hay su nombre.

Para diferenciar en la literatura y evitar confusiones con otros modelos, nipy le asigna el nombre de RAS+, siendo este el sistema de coordenadas que depende del paciente, con los ejes positivos RAS.

4.1.2.2 TIPOS DE IMÁGENES

Las imágenes en tres dimensiones tienen un tamaño de 240x240x155 y podemos encontrar cuatro tipos de imágenes diferentes en el conjunto de datos [20]:

- T1: Muestran la anatomía normal del tejido blando y grasa. Sirven por ejemplo para confirmar que una masa contiene grasa.
- T1ce: Son imágenes con contraste que permiten ver los vasos sanguíneos u otros tejidos blandos más claramente.
- T2: Muestran líquidos y alteraciones como tumores, inflamación o traumatismo.
- T2-Flair: Utiliza contraste para detectar un amplio rango de lesiones.

Todas ellas muestran información complementaria. Podemos ver un ejemplo de todas ellas en la [figura x](#).

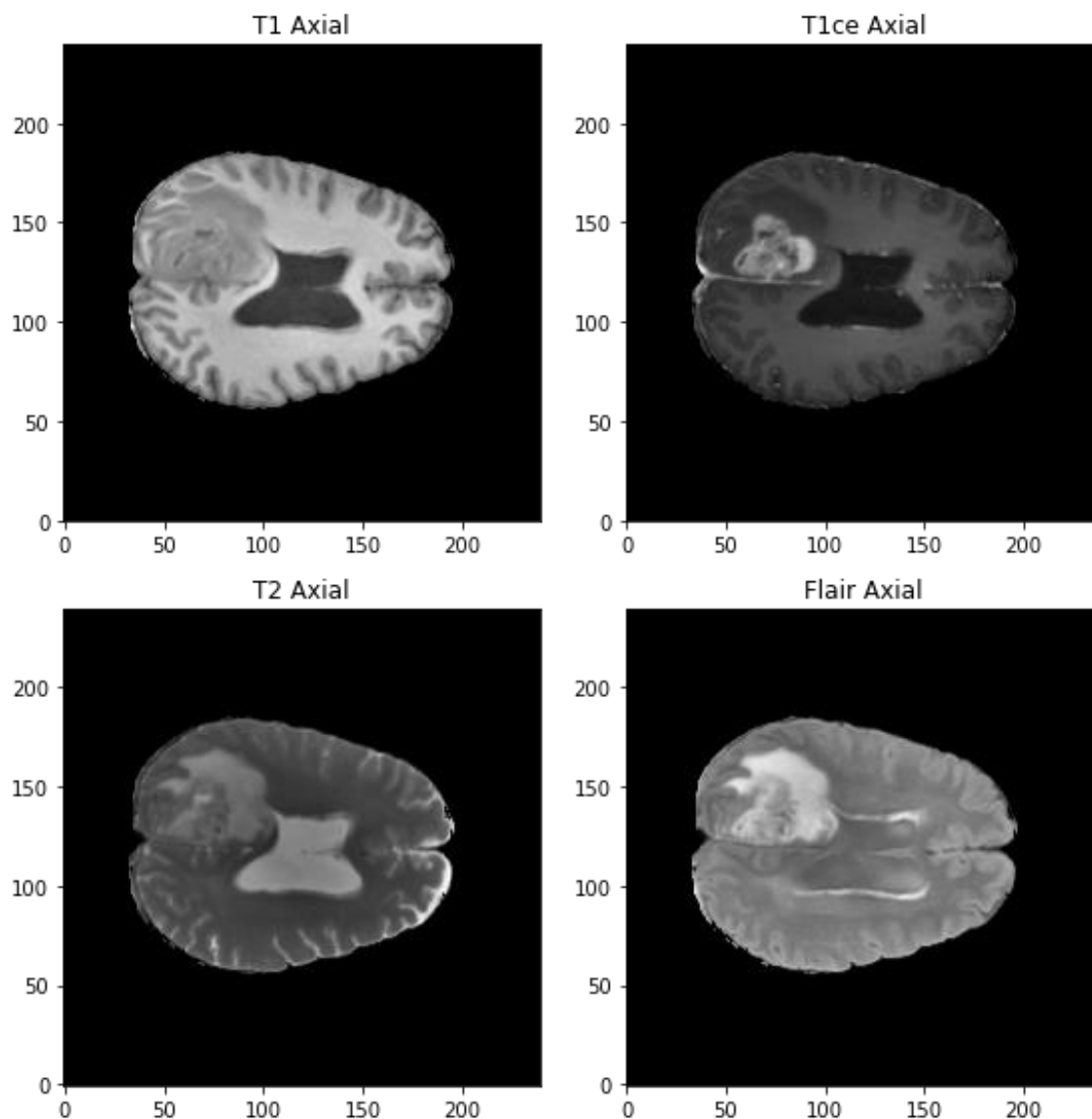


Figura X 90mm

4.1.3 EL PROBLEMA DE LA FALTA DE DATOS

Como hemos visto el conjunto de datos utilizado solamente cuenta con datos de 369 pacientes. Pero el número de datos de cada clase no está equilibrado, ya que 293 pacientes pertenecen a la clase HGG, mientras que solo 76 pertenecen a la clase LGG. Por lo que es necesario realizar un balanceo de los datos para evitar que el modelo asigne más peso a la clase dominante, ya que de otra forma si siempre predijera la clase dominante obtendría unos buenos resultados, cuando en la práctica sería inútil.

Existen dos posibilidades para abordar este problema. La primera de ellas es realizar sobremuestreo, la cual consiste en aumentar la clase minoritaria para equilibrar ambas. Y la segunda de ella es realizar submuestreo, la cual ha sido la utilizada y consiste en reducir la clase dominante para equilibrar ambas.

De esta manera se han fijado el número de elementos de ambas clases a 76 pacientes y para aumentar los datos para entrenar y validar los modelos, se ha tratado cada una de las cuatro imágenes de cada paciente como si fueran imágenes de pacientes diferentes. Por lo que el número de imágenes para el entrenamiento, validación y test es de 608 imágenes. De esta forma se obtienen dos cosas, por un lado, se consigue que la red sea capaz de clasificar el grado y la supervivencia del tumor en distintas imágenes y se consigue aumentar el número de imágenes para el entrenamiento, validación y prueba.

Aun con este número de imágenes los modelos entrenados desde cero tanto 3D como 2D, no darían buenos resultados ya que necesitan un volumen de datos mayor para poder realizar clasificaciones precisas, por lo que se utilizarán como comentamos anteriormente técnicas de transferencia de aprendizaje con distintos modelos pre-entrenados para realizar la clasificación.

4.2 PRE-PROCESADO DE LOS DATOS

En un primer lugar los datos con cargados en un DataFrame de la librería de Python Pandas, el cual contiene la identificación del paciente, el grado de su tumor, la supervivencia del paciente en días, su tipo de resección y las rutas a cada una de las imágenes del paciente. Las imágenes serán cargadas más adelante utilizando un generador ya que al ser tan pesadas no pueden ser cargadas todas a la vez en memoria.

4.2.1 TRATAMIENTO DE DATOS FALTANTES

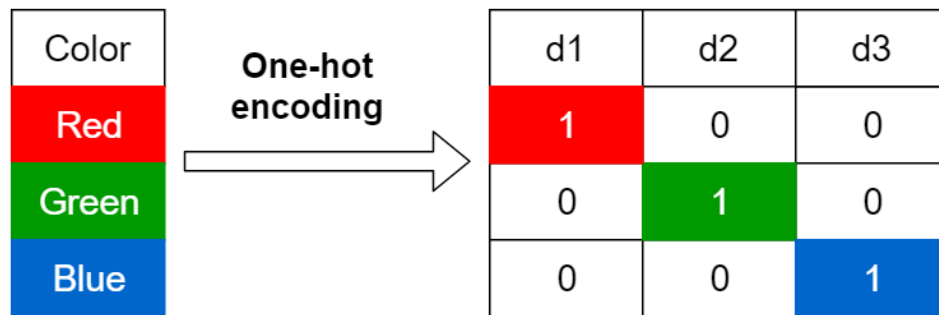
Analizando los datos podemos observar que todos los pacientes con un grado de tumor de tipo LGG, no tienen información sobre su edad, supervivencia ni tipo de resección. Esto se debe en gran medida a que estos pacientes tienen un pronóstico bastante favorable [21] y la mayoría no se realiza cirugía. Por ello asumiremos que ninguno de ellos se realizó la cirugía. Su edad será fijada de forma aleatoria teniendo en cuenta la media del resto de las edades y la desviación típica, de tal forma que las edades generadas serán como máximo la media más la desviación típica y como mínimo la media menos la desviación típica.

Para determinar el tiempo de supervivencia me he basado en el estudio [22], según el cual el 76% de las personas con una resección inferior al 90% del tumor LGG sobreviven más de 5 años. Como hemos determinado que estas personas no se han realizado ninguna resección, asumiremos que el 76% han sobrevivido más de 5 años y el 24% menos. Por lo que se generarán de manera aleatoria, de tal forma que hay un 24% de probabilidades de sobrevivir entre 4 y 5 años y un 76% de probabilidades de sobrevivir entre 5 y 7 años. Los números generados son un número aleatorio de días en una de esas dos clases.

4.2.2 TRANSFORMACIÓN DE LOS DATOS

Una vez comprobado que no hay ningún dato faltante hemos normalizado la edad de los pacientes entre el máximo y el mínimo de las edades y se han transformado los datos en formato texto a numérico para que pueda ser entrenado el modelo. Existen dos técnicas para realizar esto.

- Ordinal Encoding: Esta técnica sustituye cada valor único por un número desde el 0 hasta el infinito. El resultado es una única columna en la que cada clase es representada por un número distinto. Esta técnica es la utilizada para codificar los grados del tumor (0 LGG y 1 HGG) y para codificar la supervivencia del paciente (0 menos de 1 año, 1 entre 1-5 años y 2 para los supervivientes de más de 5 años).
- One Hot Encoding: Esta técnica a diferencia de la anterior que creaba una sola columna con un número para cada clase, crea una columna para cada valor único de tal manera que asigna un 1 a la columna a la que pertenece la clase y un 0 al resto. Podemos ver un ejemplo en la [figura x](#). En nuestro caso hemos aplicado esta técnica para codificar el tipo de cirugía en GTR, STR o Sin cirugía.



Fuente(<https://towardsdatascience.com/encoding-categorical-variables-one-hot-vs-dummy-encoding-6d5b9c46e2db>)

4.2.3 TRATAMIENTO DE LAS IMÁGENES

Las imágenes son tratadas al ser cargadas en memoria por el generador. Este generador carga un grupo de imágenes del tamaño del “batch” del modelo. Para ello se obtiene la ruta de la imagen guardada anteriormente en el DataFrame de Pandas y se realizan las siguientes transformaciones.

4.2.3.1 RE-ORIENTACIÓN DE LAS IMÁGENES

Las imágenes en tres dimensiones pueden tener distintas orientaciones dependiendo de la orientación del sujeto en el momento de la realización del escáner. Por lo que las imágenes son reorientadas a un espacio común de tal forma que todas las imágenes que se le pasen al modelo tendrán la misma orientación. Las imágenes son orientadas utilizando la librería nibabel [19] al eje RAS. Esto se consigue mediante el siguiente código de Python.

```
img = nib.load(filepath)
img = nib.as_closest_canonical(img)
```

4.2.3.2 NORMALIZACIÓN DE LAS IMÁGENES

Las imágenes son matrices de tres dimensiones. El contenido de estas matrices no son enteros del 0 al 255 como la mayoría de las imágenes, sino que son números decimales los cuales representan unidades Hounsfield (HU) [23]. Estas unidades son universalmente usadas en tomografías y escáneres de manera estandarizada. Son obtenidas por la

transformación lineal de los coeficientes de atenuación medidos. Esta basado en las densidades del agua pura que se corresponde con 0 HU y del aire que se corresponde con -1000 HU. Generalmente los valores de los escáneres oscilan entre -1000 (aire) y +2000 HU para los huesos más densos.

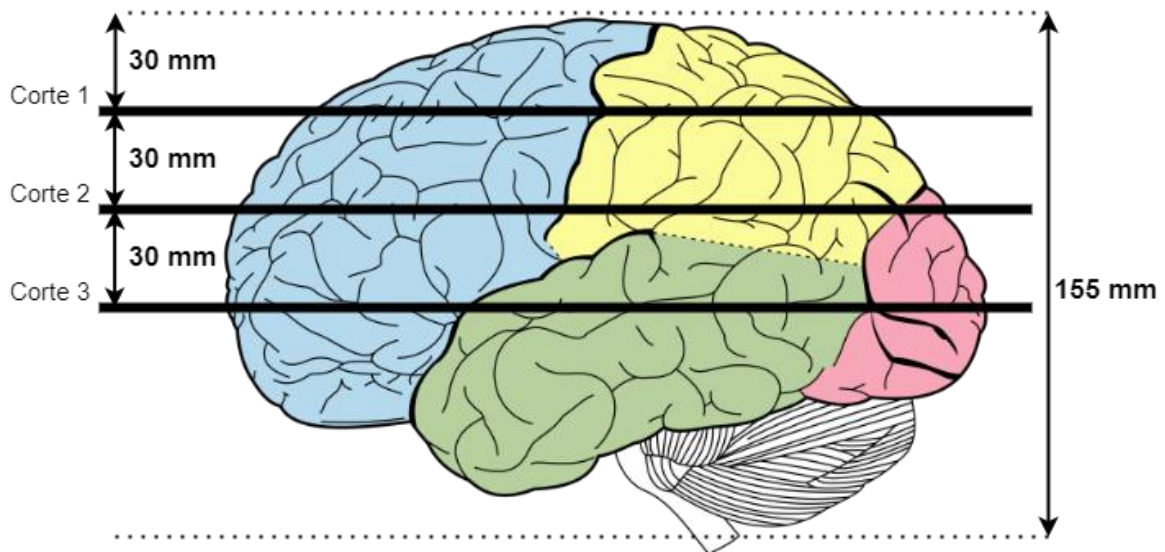
Para evitar que los huesos aparezcan en las imágenes y confundan a la red he acotado el rango de valores entre $[-1000, 800]$, de tal manera que se evitan los huesos que tienen una medida de unos 1000 HU [24].

Una vez delimitados los valores han sido normalizados entre este rango para mejorar los resultados del entrenamiento.

4.2.3.3 SEGMENTACIÓN DE LAS IMÁGENES

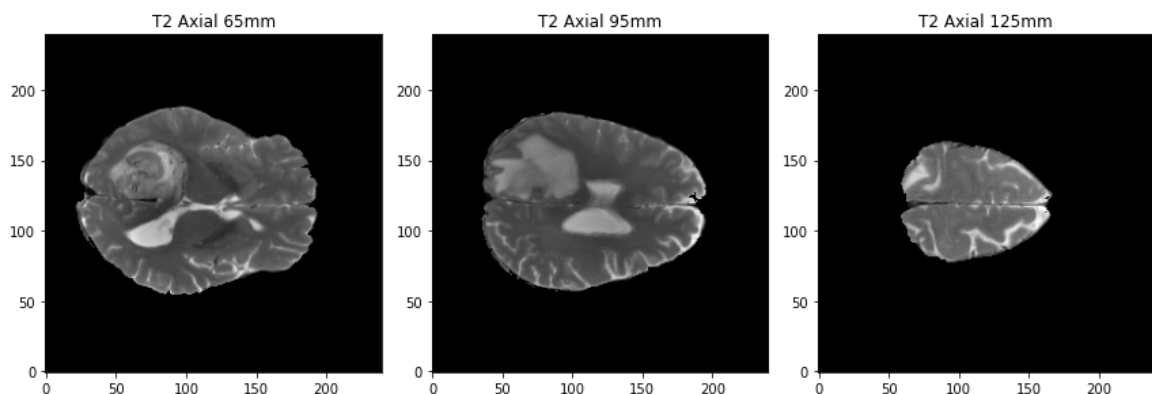
Los modelos pre-entrenados utilizados han sido entrenados con imágenes de tamaño $224 \times 224 \times 3$, aunque las dos primeras dimensiones pueden variar en un cierto margen. Es por eso por lo que necesitamos ajustar nuestras imágenes para introducirlas en estos modelos. Nuestras imágenes tienen un tamaño de $240 \times 240 \times 155$ por lo que nuestro tamaño objetivo será $240 \times 240 \times 3$. No es necesario modificar las dos primeras dimensiones, pero si la tercera. Como hemos visto las imágenes son modelos en tres dimensiones del cerebro por lo que para reducir la dimensionalidad tomaremos tres segmentos del cerebro. Estos cortes se han realizado por tres zonas distintas del cerebro separados por 30mm. Podemos apreciar de manera visual en la **figura x** como se han realizado estos cortes y en la **figura x** un ejemplo de cómo se verían en una imagen T2 estos tres segmentos con los que entrenaremos al modelo. Podemos diferenciar claramente en ellos distintos tamaños del tumor al tratarse de regiones distintas dentro del modelo 3D completo.

De esta forma ya podremos pasarle nuestra imagen segmentada, normalizada y con una orientación fija a nuestro modelo.



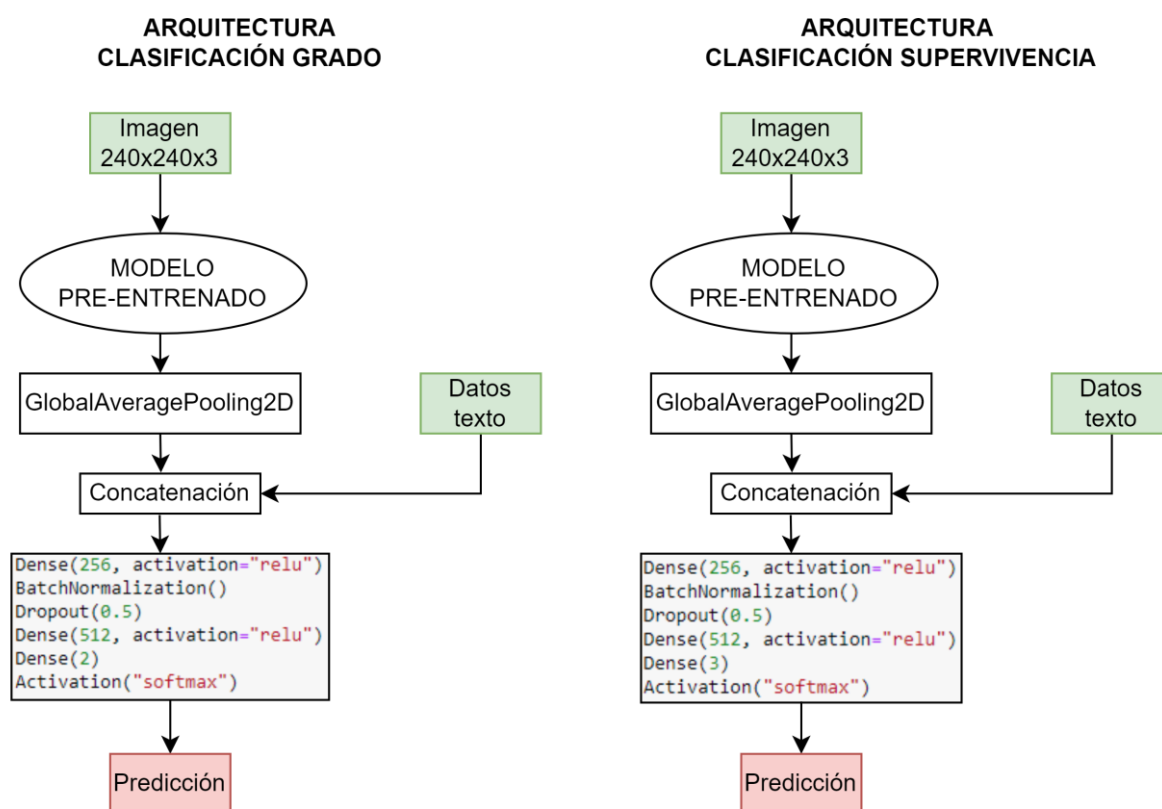
¿FUENTE PROPIA? El cerebro es de ahí pero las líneas las hice yo Cerebro->

https://es.m.wikipedia.org/wiki/Archivo:Lobes_of_the_brain_NL.svg



4.3 ARQUITECTURA DE LA RED

La red es alimentada con las imágenes en el formato que hemos comentado anteriormente, pasando en primer lugar por el modelo pre-entrenados, a continuación, la salida del modelo pasará por una capa GlobalAveragePooling2D. Tras esta capa es donde se fusionan los resultados de la convolución de las imágenes con los datos en formato de texto formados por edad y estado de la resección del tumor los cuales son utilizados junto con las imágenes realizar la clasificación. Una vez unidos pasan a la parte de clasificación que variará de unos experimentos a otros. Los primeros experimentos para realizar las pruebas con los distintos modelos pre-entrenados tendrán las mismas capas que en la **figura X**, en la que podemos ver un resumen de la arquitectura de las dos redes.



4.4 PRUEBAS DE MODELOS PRE-ENTRENADOS

La supervivencia y el grado del glioma han sido predichos utilizando dos redes distintas. Esto tiene una ventaja sobre utilizar una misma red para realizar ambas predicciones y es que son totalmente independientes. Ya que, si solamente una red se encargara de predecir ambas, habría parámetros compartidos entre ambas y no podrían ser optimizadas al máximo ya que existiría una cierta dependencia entre ambas y puede que lo que fuera mejor para una empeorara la predicción de la otra. Por eso se ha tomado la decisión de separar ambas redes.

Se han probado distintos modelos pre-entrenados para determinar cual de ellos tiene un mejor rendimiento en nuestro problema. Una vez determinado cual de ellos es el mejor se realizará un ajuste de los hiperparámetros del modelo con mejor rendimiento para tratar de optimizarlo.

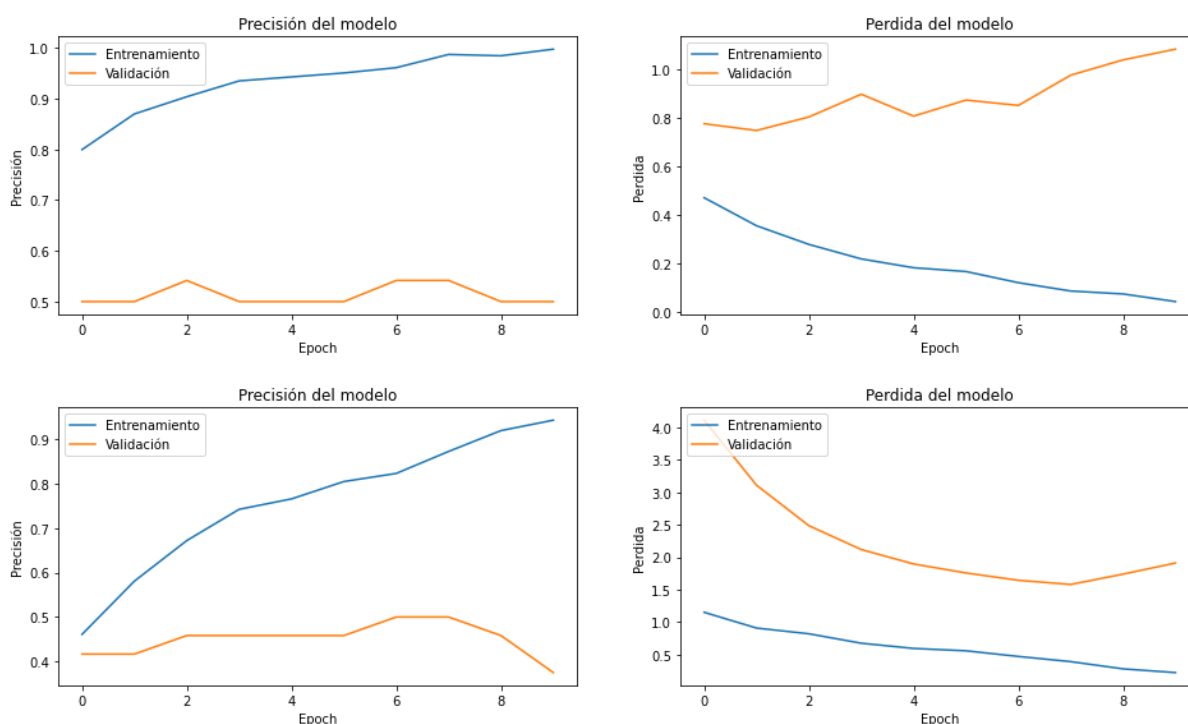
Todos los modelos que se describen a continuación han sido entrenados con los mismos datos. Además, se han utilizado los mismos parámetros para todos ellos, siendo los más importantes: Learning rate 0.0002, optimizador Adam, batch size 16, epochs 10.

Para la clasificación se ha utilizado la arquitectura comentada anteriormente, con 256-512 neuronas para la primera y la segunda capa densa respectivamente, BatchNormalization y una capa de dropout con una tasa de 0.5.

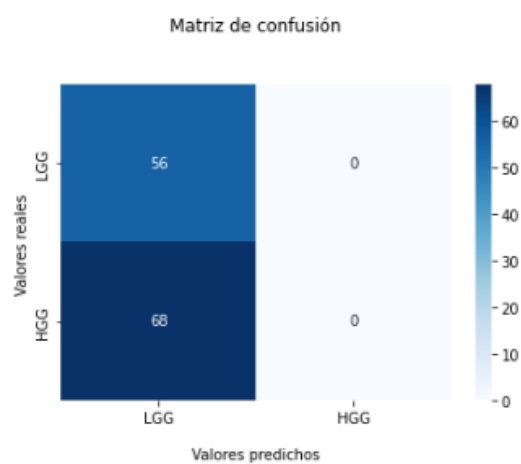
Posteriormente todas las redes han sido probadas con el mismo conjunto de prueba, el cual es un conjunto diferente al de entrenamiento y validación y no ha sido nunca visto por la red neuronal entrenada. Las distintas pruebas se describen a continuación.

4.4.1 ResNet101

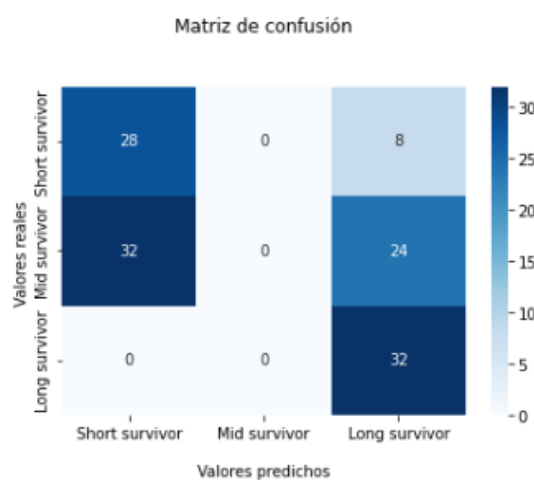
Utilizando esta arquitectura se ha obtenido una f1-score de 0.42 en la clasificación de la supervivencia y de un 0.31 en la clasificación del grado del tumor. Los resultados obtenidos tanto con ResNet101 como con ResNet50 han sido los peores de todos, ya que estas redes no han sido capaces de realizar una buena clasificación y simplemente han predicho todos los valores a una o varias clases.



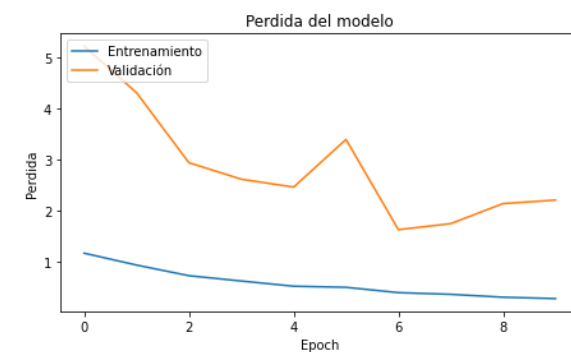
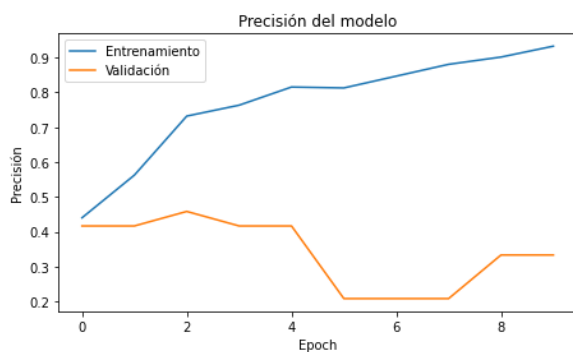
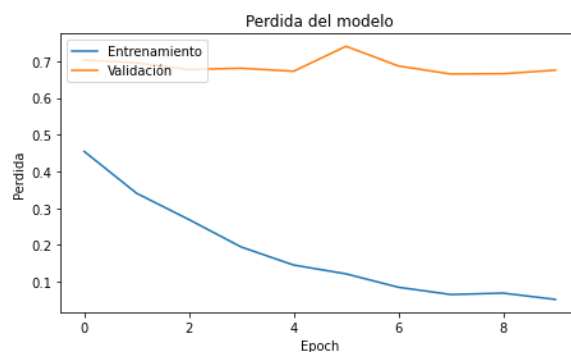
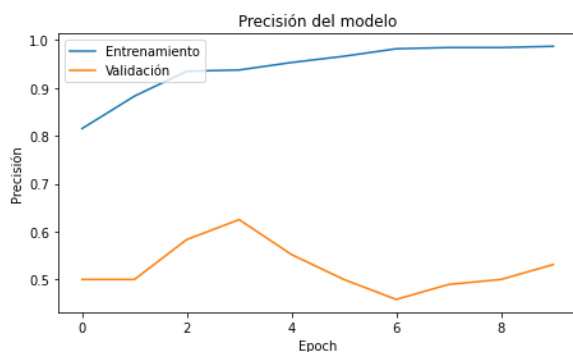
	precision	recall	f1-score
LGG	0.45	1.00	0.62
HGG	0.00	0.00	0.00
macro avg	0.23	0.50	0.31
accuracy	0.45		



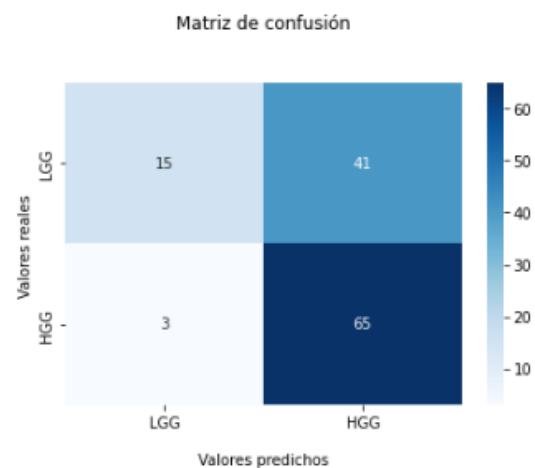
	precision	recall	f1-score
short	0.47	0.78	0.58
mid	0.00	0.00	0.00
long	0.50	1.00	0.67
macro avg	0.32	0.59	0.42
accuracy	0.48		



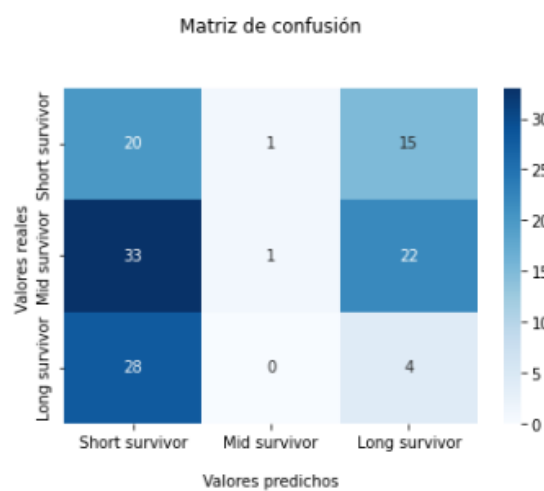
4.4.2 ResNet50



	precision	recall	f1-score
LGG	0.83	0.27	0.41
HGG	0.61	0.96	0.75
macro avg	0.72	0.61	0.58
accuracy	0.65		

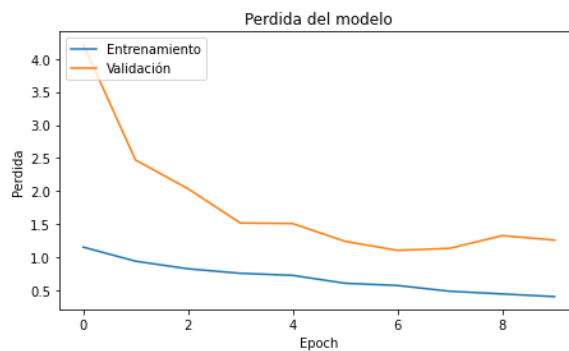
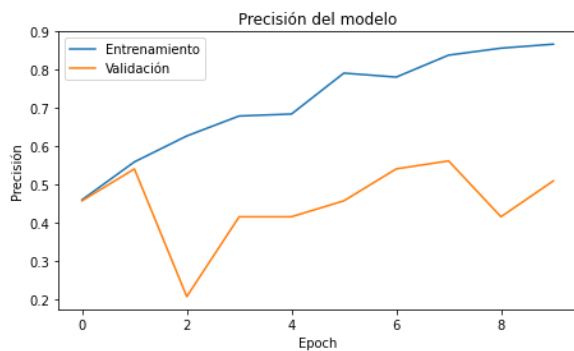
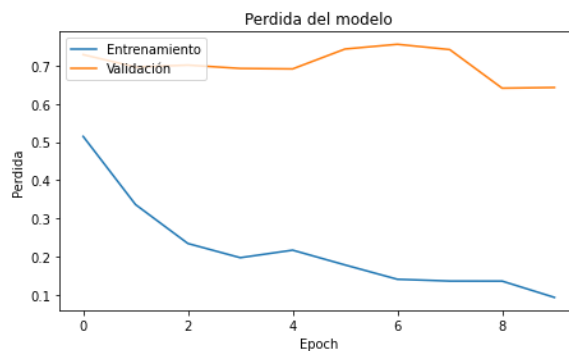
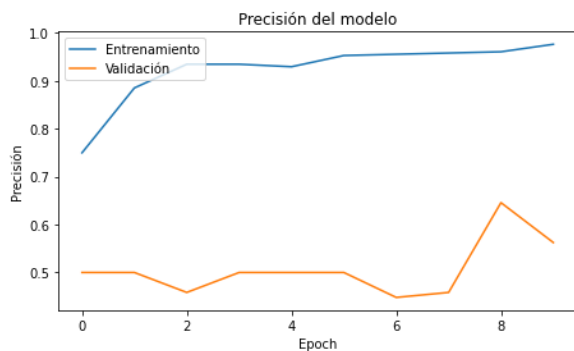


	precision	recall	f1-score
short	0.25	0.56	0.34
mid	0.50	0.02	0.03
long	0.10	0.12	0.11
macro avg	0.28	0.23	0.16
accuracy	0.20		

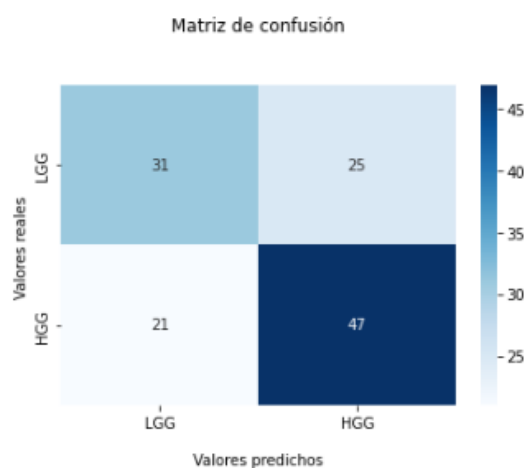


4.4.3 EfficientNetB4

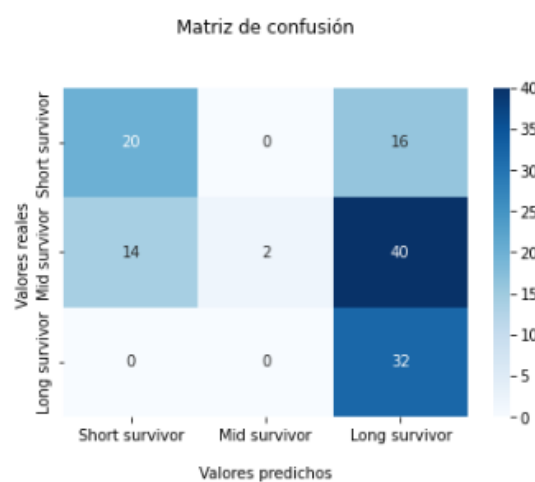
Los resultados de esta red han sido mejores que los de ResNet pero se quedan por detrás de las mejores con un 62% en la predicción del grado y un 39% en la predicción de la supervivencia.



	precision	recall	f1-score
LGG	0.60	0.55	0.57
HGG	0.65	0.69	0.67
macro avg	0.62	0.62	0.62
accuracy	0.63		

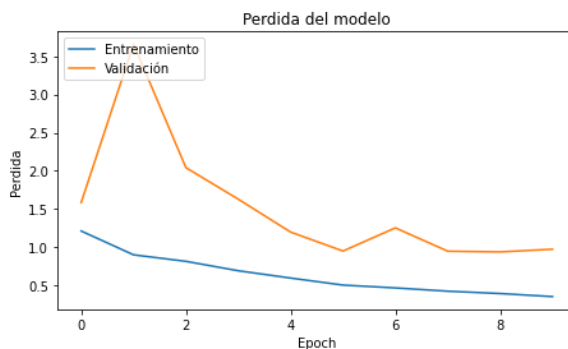
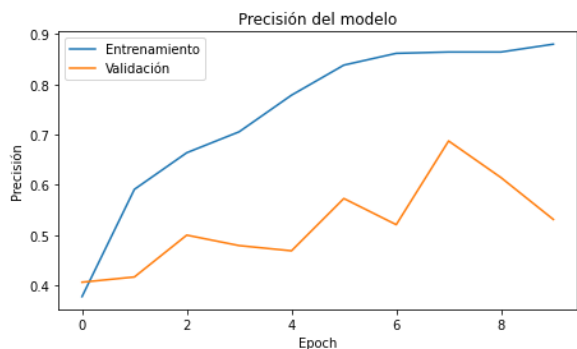
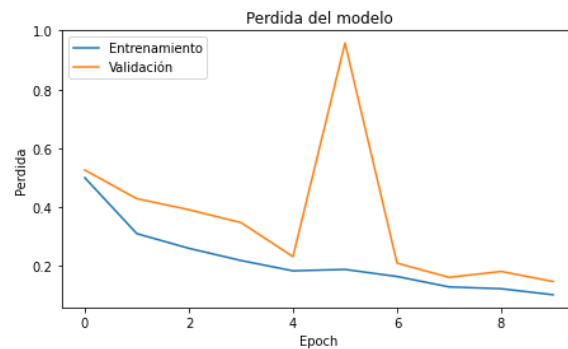
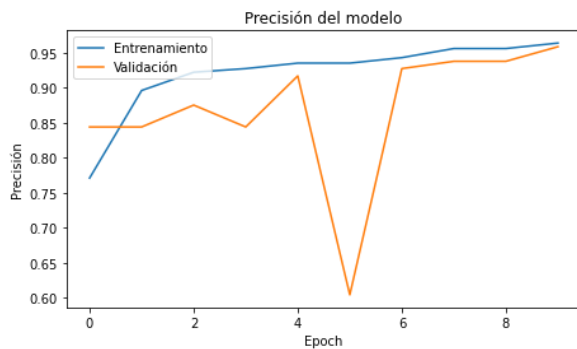


	precision	recall	f1-score
short	0.59	0.56	0.57
mid	1.00	0.04	0.07
long	0.36	1.00	0.53
macro avg	0.65	0.53	0.39
accuracy	0.44		

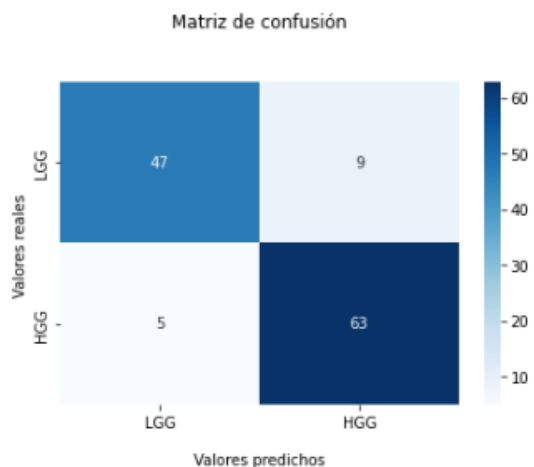


4.4.4 VGG16

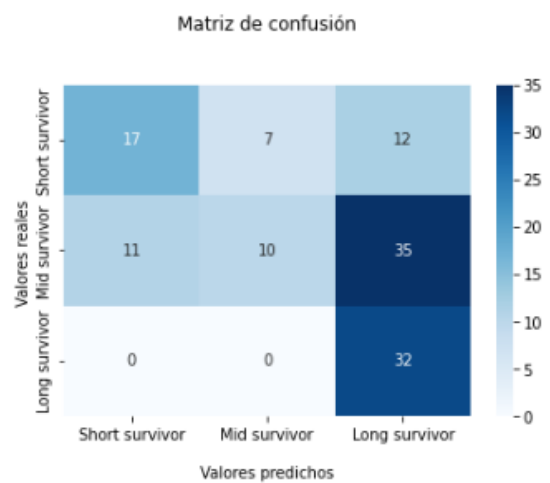
Esta red ha obtenido unos buenos resultados con un 89% en la clasificación del grado del tumor y un 46% en la clasificación de la supervivencia.



	precision	recall	f1-score
LGG	0.90	0.84	0.87
HGG	0.88	0.93	0.90
macro avg	0.89	0.88	0.89
accuracy	0.89		

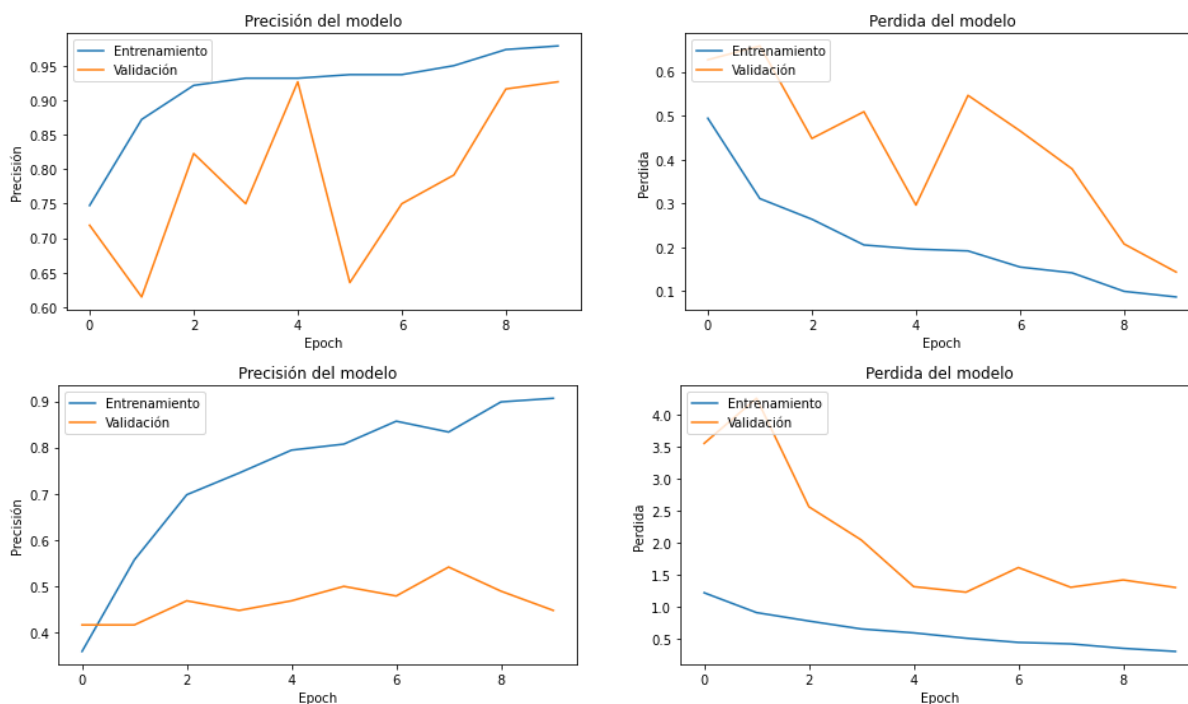


	precision	recall	f1-score
short	0.61	0.47	0.53
mid	0.59	0.18	0.27
long	0.41	1.00	0.58
macro avg	0.53	0.55	0.46
accuracy	0.48		



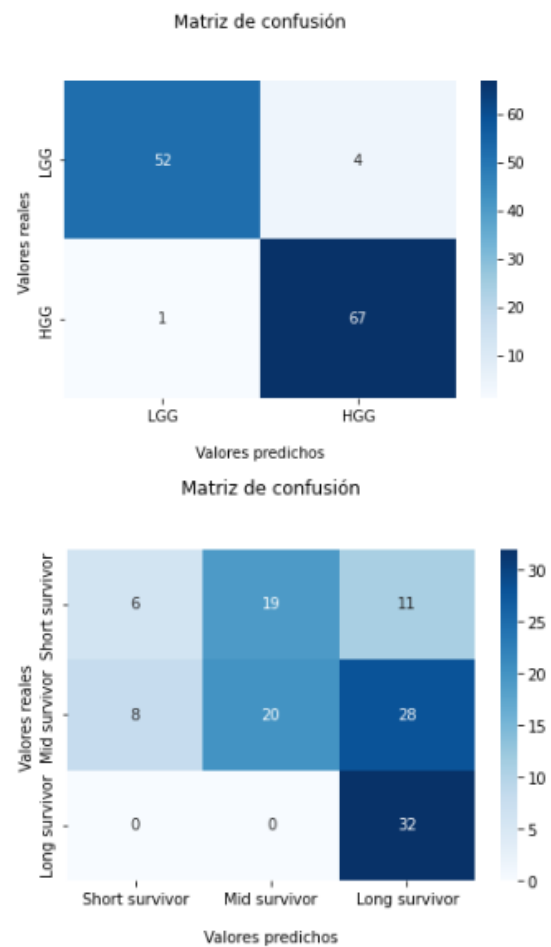
4.4.5 InceptionV3

Utilizando esta arquitectura se ha obtenido una f1-score de 0.43 en la clasificación de la supervivencia y de un 0.96 en la clasificación del tumor. Esta arquitectura consigue una de las mejores puntuaciones en la clasificación del grado, pero no son los mejores en la clasificación de la supervivencia.



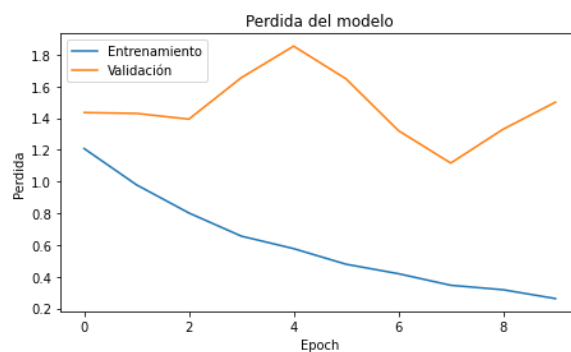
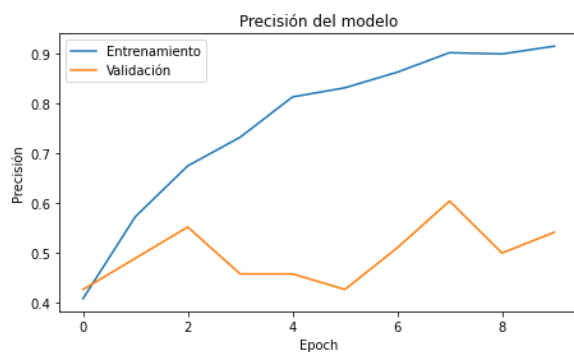
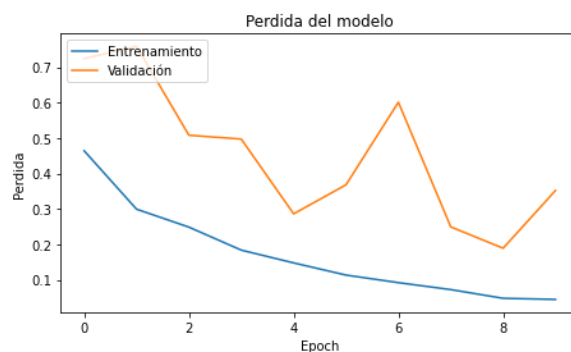
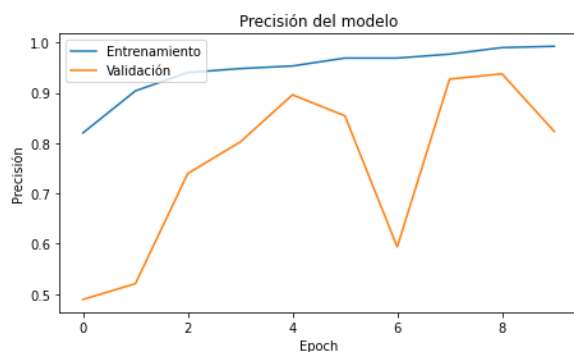
	precision	recall	f1-score
LGG	0.98	0.93	0.95
HGG	0.94	0.99	0.96
macro avg	0.96	0.96	0.96
accuracy	0.96		

	precision	recall	f1-score
short	0.43	0.17	0.24
mid	0.51	0.36	0.42
long	0.45	1.00	0.62
macro avg	0.46	0.51	0.43
accuracy	0.47		

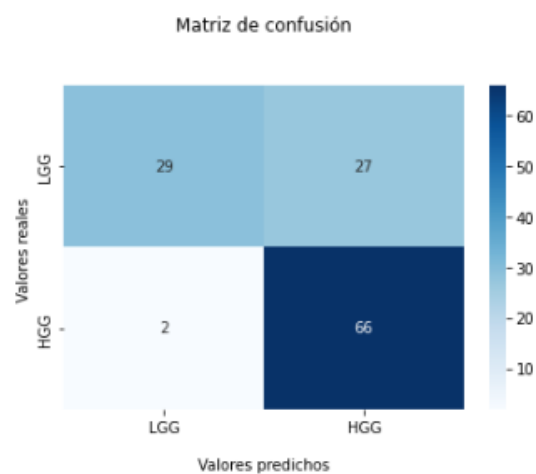


4.4.6 InceptionResNetV2

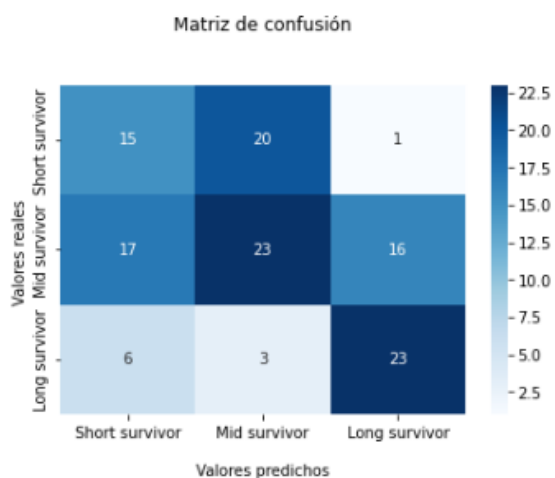
Utilizando esta arquitectura se ha obtenido una f1-score de 0.50 en la clasificación de la supervivencia y de un 0.74 en la clasificación del tumor.



	precision	recall	f1-score
LGG	0.94	0.52	0.67
HGG	0.71	0.97	0.82
macro avg	0.82	0.74	0.74
accuracy	0.77		

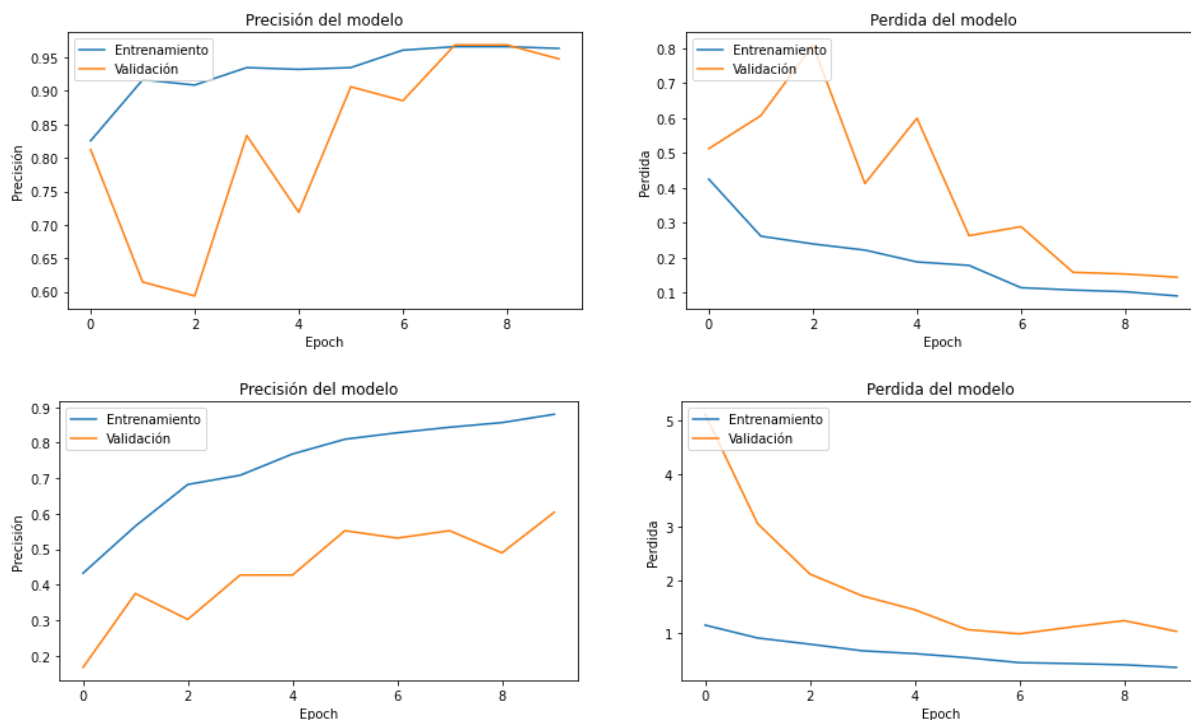


	precision	recall	f1-score
short	0.39	0.42	0.41
mid	0.50	0.41	0.45
long	0.57	0.72	0.64
macro avg	0.49	0.52	0.50
accuracy	0.49		

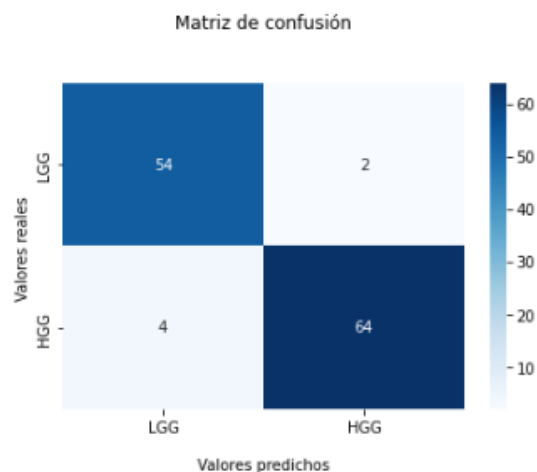


4.4.7 DenseNet121

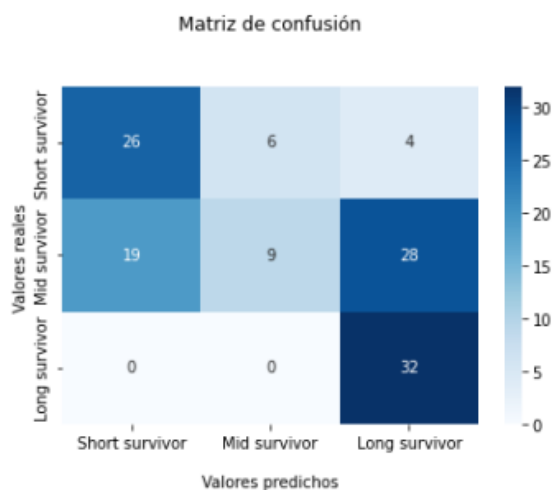
Utilizando esta arquitectura se ha obtenido una f1-score de 0.95 en la clasificación de la supervivencia y de un 0.52 en la clasificación del tumor. Esta es la arquitectura elegida al obtener los resultados más balanceados. Obtiene una precisión del 95% en la clasificación del tipo de tumor, mejorando el estado del arte actual, aún sin optimizar sus parámetros.



	precision	recall	f1-score
LGG	0.93	0.96	0.95
HGG	0.97	0.94	0.96
macro avg	0.95	0.95	0.95
accuracy	0.95		

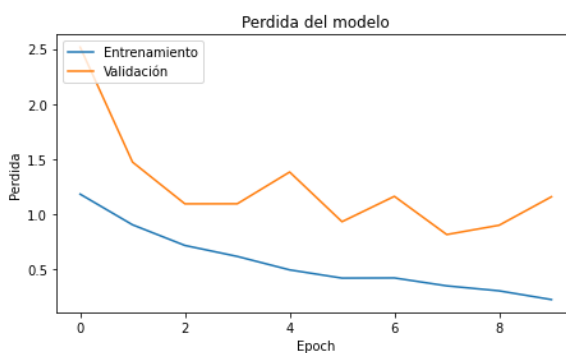
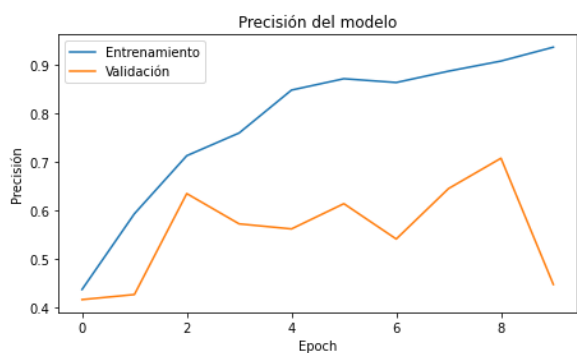
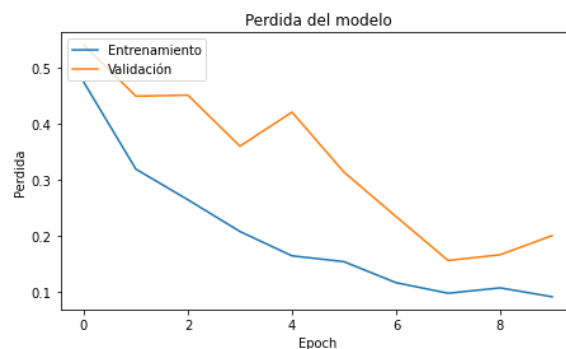
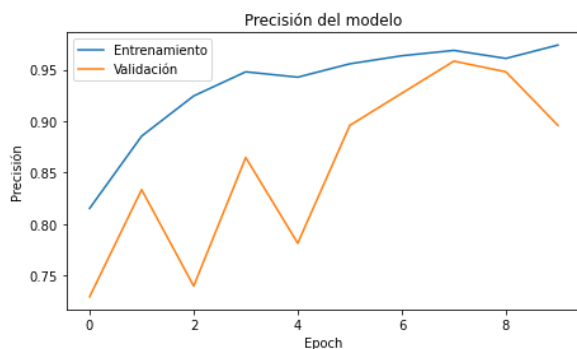


	precision	recall	f1-score
short	0.58	0.72	0.64
mid	0.60	0.16	0.25
long	0.50	1.00	0.67
macro avg	0.56	0.63	0.52
accuracy	0.54		

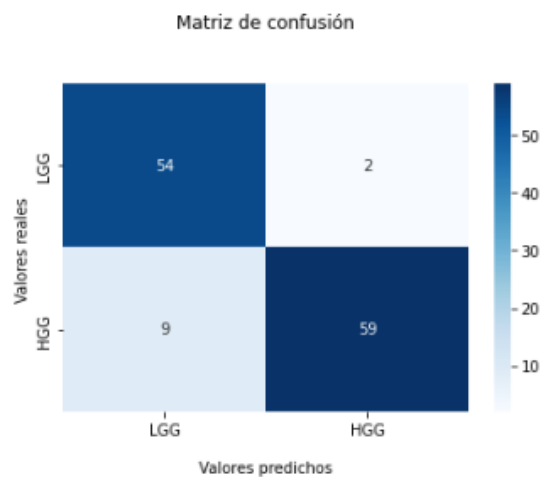


4.4.8 DenseNet201

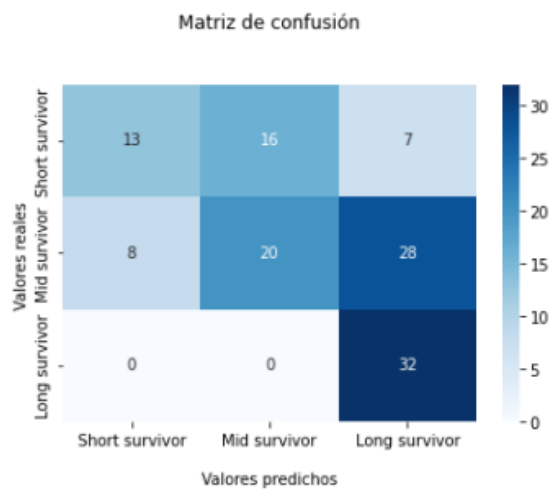
Utilizando esta arquitectura se ha obtenido una f1-score de 0.51 en la clasificación de la supervivencia y de un 0.91 en la clasificación del tumor. Los resultados son parecidos a los obtenidos con DenseNet121, pero ligeramente peores. Teniendo en cuenta que además su coste computacional es más elevado al tener una mayor profundidad, se ha elegido DenseNet121.



	precision	recall	f1-score
LGG	0.86	0.96	0.91
HGG	0.97	0.87	0.91
macro avg	0.91	0.92	0.91
accuracy	0.91		



	precision	recall	f1-score
short	0.62	0.36	0.46
mid	0.56	0.36	0.43
long	0.48	1.00	0.65
macro avg	0.55	0.57	0.51
accuracy	0.52		



4.4.9 ELECCIÓN DEL MEJOR MODELO

Podemos observar en la **tabla x** un resumen con los resultados de los diferentes modelos. Aunque los mejores resultados en la predicción del grado fueron obtenidos por la arquitectura InceptionV3, los resultados para la supervivencia no fueron tan positivos. Por lo que la red la cual optimizaremos para obtener los mejores resultados posibles será la DenseNet121 ya que ha obtenido los mejores resultados balanceados entre todas las redes.

Utilizando los mismos datos utilizados para entrenar los modelos, realizaremos distintas pruebas para encontrar los mejores hiperparámetros y arquitectura de clasificación para la red DenseNet121. Al ser dos redes independientes buscaremos los parámetros que más favorezcan a cada red por separado.

	F1-Macro Grado	F1-Macro Supervivencia
ResNet101	31	42
ResNet50	58	16
EfficientNetB4	62	39
VGG16	89	46
InceptionV3	96	43
InceptionResNetV2	74	50
DenseNet121	95	52
DenseNet201	91	51

4.5 OPTIMIZACIÓN DEL MEJOR MODELO

Para encontrar los parámetros con los que mejor funcione DenseNet101 para clasificar nuestro problema se han llevado a cabo distintos experimentos de manera manual. Estos experimentos podrían llevarse a cabo utilizando técnicas como GridSearchCV, pero por las características de nuestro problema no es posible su utilización debido a que nuestra red tiene varias entradas y esta función no es compatible con el entrenamiento con varias entradas desde un generador.

Inicialmente se han utilizado como parámetros 32-64 neuronas en sus capas densas respectivamente, 0.3 como tasa de dropout, 0.0005 como tasa de entrenamiento y relu

como función de activación. Conforme se han ido realizando los experimentos se ha seleccionado la mejor opción para cada modelo para realizar los siguientes experimentos. En las siguientes secciones se muestran las distintas pruebas realizadas, así como los resultados obtenidos para cada una de ellas. Además, se puede observar un resumen de los experimentos en la **tabla x** con los mejores resultados para cada opción resaltados.

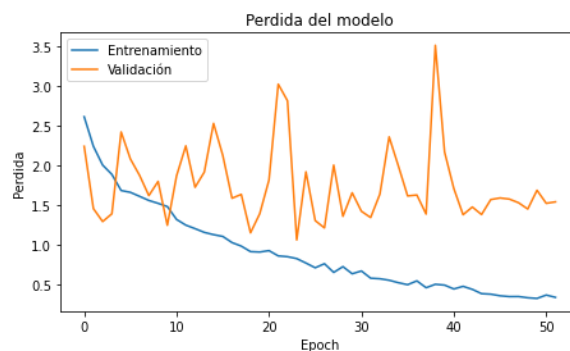
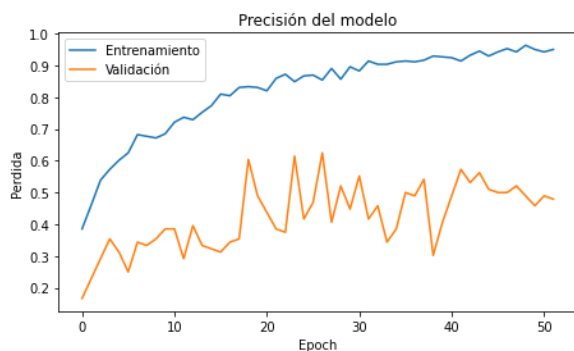
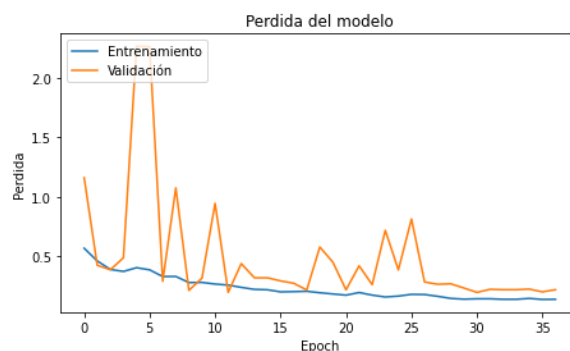
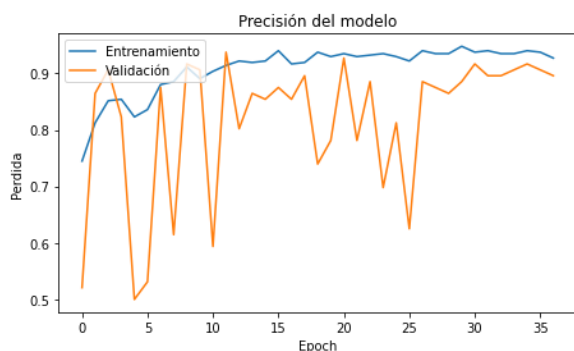
	Grado F1-Macro	Supervivencia F1-Macro
BATCH - BATCH	88	48
BATCH - NO BATCH	96	44
NO BATCH - NO BATCH	94	51
32, 64	96	51
64, 128	90	47
128, 256	89	29
256, 256	97	16
DROPOUT 0.2	93	60
DROPOUT 0.3	97	51
DROPOUT 0.4	93	52
DROPOUT 0.5	96	58
RELU	97	60
TANH	95	50
0.0005	97	60
0.001	93	48
0.002	92	34

4.5.1 BATCH NORMALIZATION

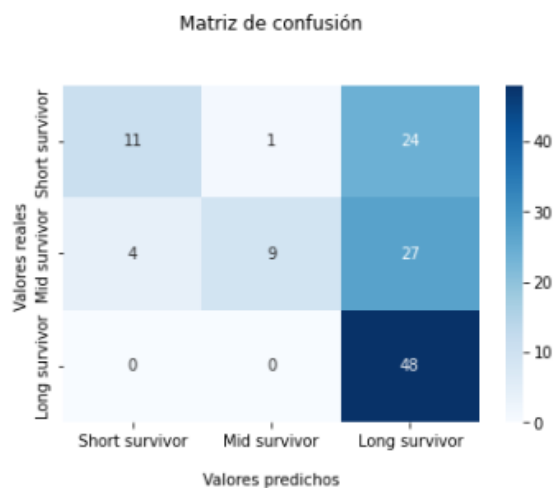
Esta capa normaliza su salida usando la media y desviación típica de cada conjunto de datos, manteniendo la media cercana a 0 y la desviación típica cercana a 1. Se han realizado pruebas utilizando solamente una capa de BatchNormalization después de la

primera capa densa, dos capas una detrás de cada una de las capas densas y sin ninguna capa de BatchNormalization. Los mejores resultados han sido obtenidos en el caso de la clasificación del grado con una sola capa de BatchNormalization y en el caso de la clasificación de la supervivencia sin ninguna capa.

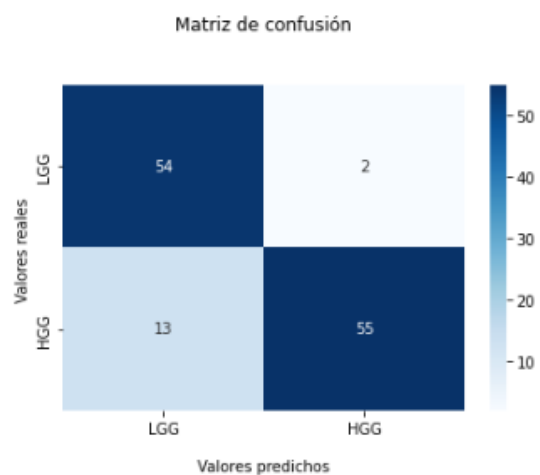
4.5.1.1 DOS CAPAS



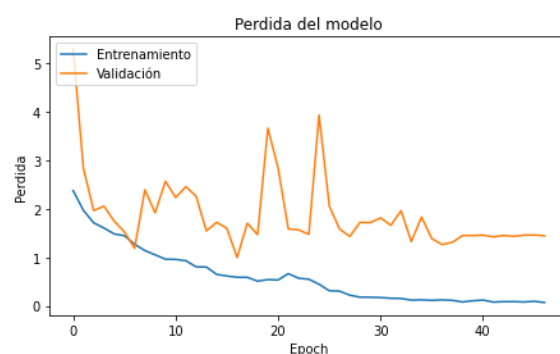
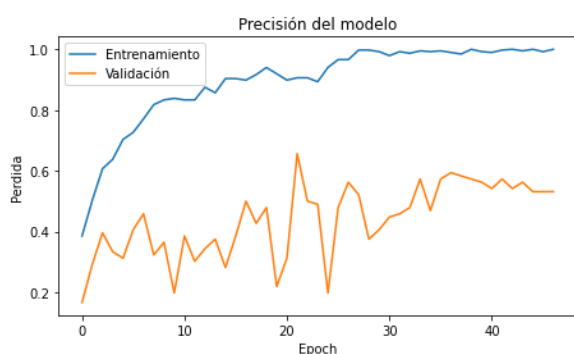
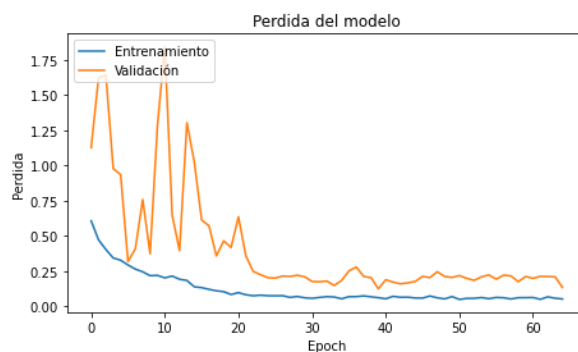
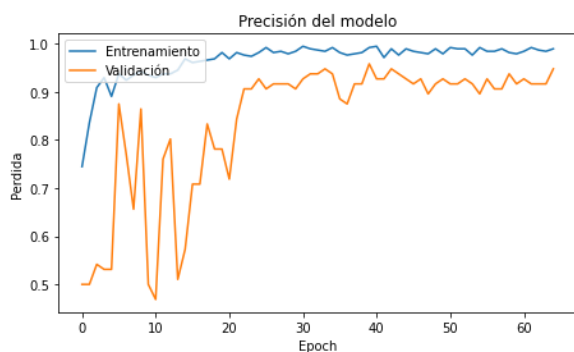
	precision	recall	f1-score
short	0.73	0.31	0.43
mid	0.90	0.23	0.36
long	0.48	1.00	0.65
macro avg	0.71	0.51	0.48
accuracy	0.55		



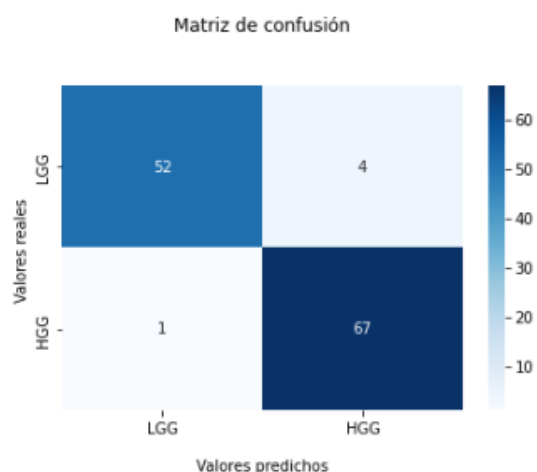
	precision	recall	f1-score
LGG	0.81	0.96	0.88
HGG	0.96	0.81	0.88
macro avg	0.89	0.89	0.88
accuracy	0.88		



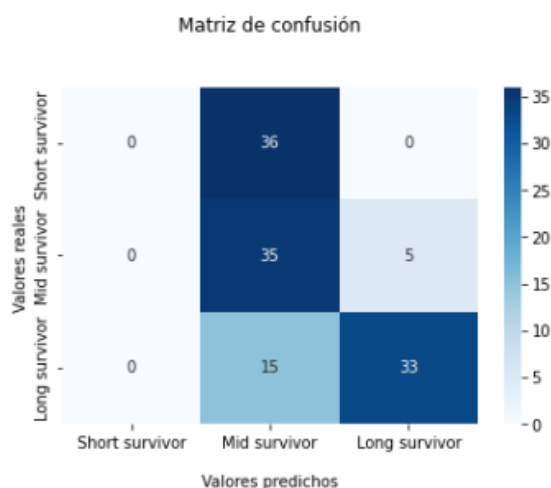
4.5.1.2 UNA CAPA



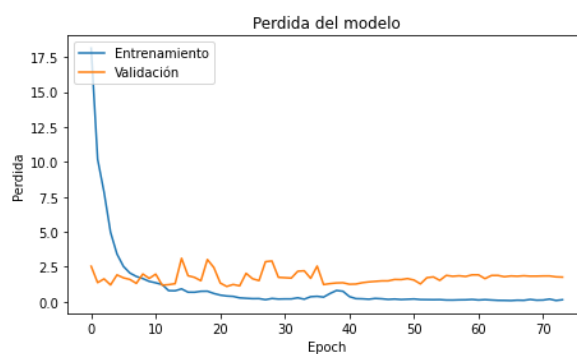
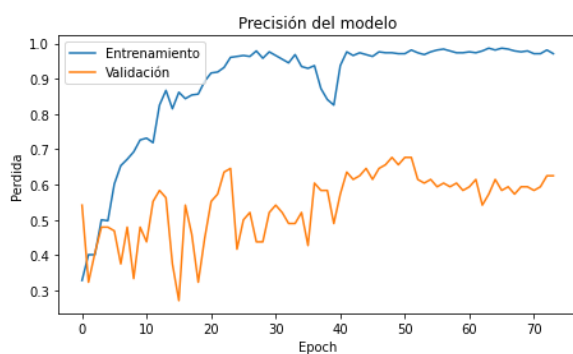
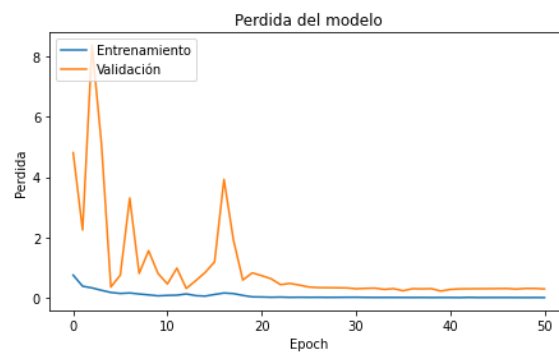
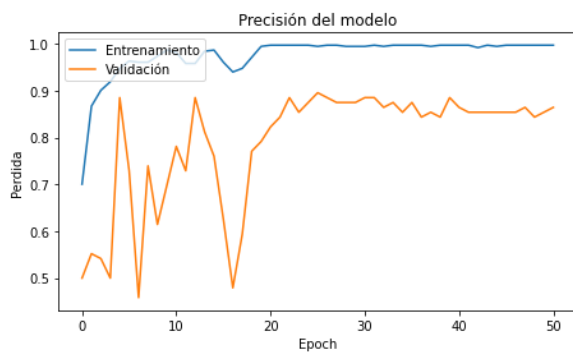
	precision	recall	f1-score
LGG	0.98	0.93	0.95
HGG	0.94	0.99	0.96
macro avg	0.96	0.96	0.96
accuracy	0.96		



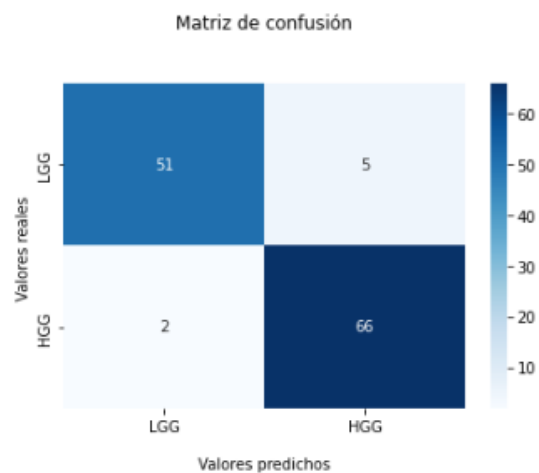
	precision	recall	f1-score
short	0.00	0.00	0.00
mid	0.41	0.88	0.56
long	0.87	0.69	0.77
macro avg	0.43	0.52	0.44
accuracy	0.55		



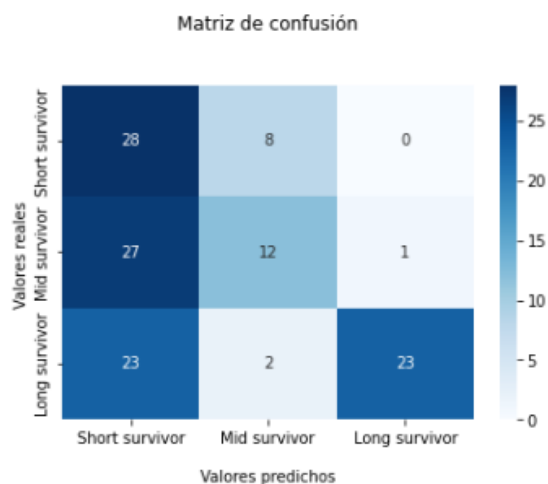
4.5.1.3 SIN CAPAS



	precision	recall	f1-score
LGG	0.96	0.91	0.94
HGG	0.93	0.97	0.95
macro avg	0.95	0.94	0.94
accuracy	0.94		



	precision	recall	f1-score
short	0.36	0.78	0.49
mid	0.55	0.30	0.39
long	0.96	0.48	0.64
macro avg	0.62	0.52	0.51
accuracy	0.51		

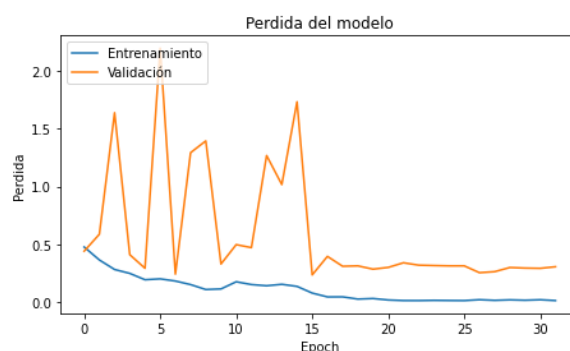
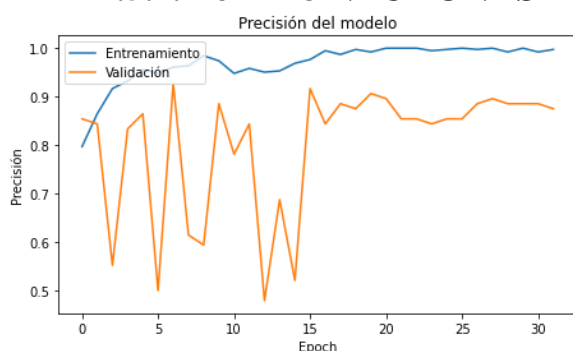


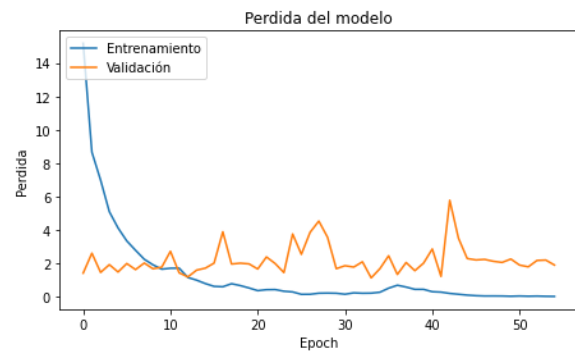
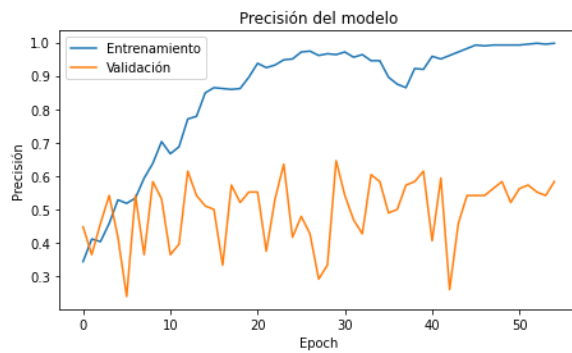
4.5.2 NEURONAS

Se han realizado pruebas con distintos números de neuronas para determinar cuales son los mejores para cada uno de los modelos. Se realizarán pruebas con los siguientes pares de neuronas, siendo el primer número el número de neuronas correspondiente a la primera capa densa y el segundo número el correspondiente a la segunda capa. Los pares probados fueron 32-64, 64-128, 128-256 y 256-256. Los mejores resultados fueron obtenidos con 32-64 neuronas para la clasificación de la supervivencia y con 256-256 para la predicción del grado del tumor.

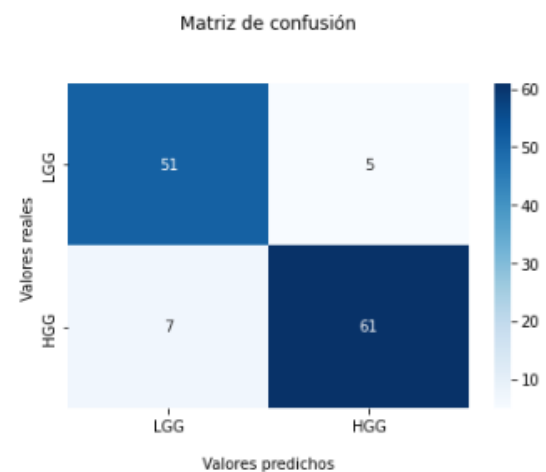
4.5.2.1 32-64 NEURONAS

4.5.2.2 64-128 NEURONAS

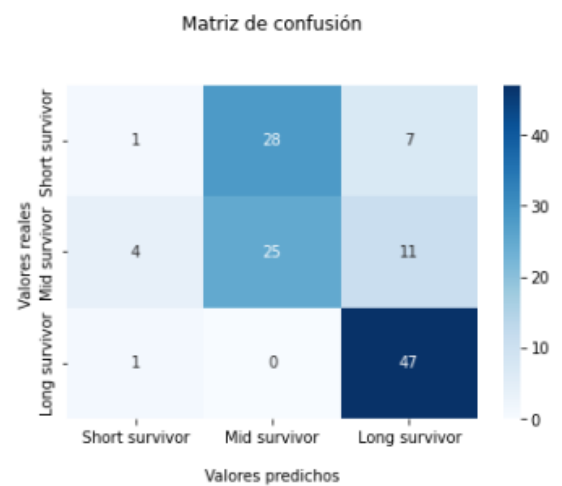




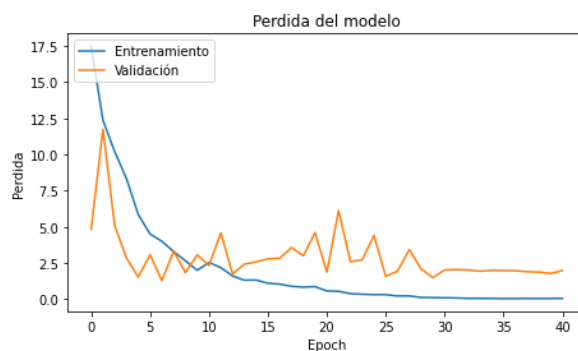
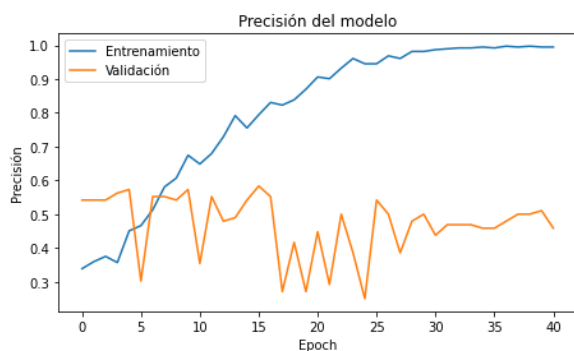
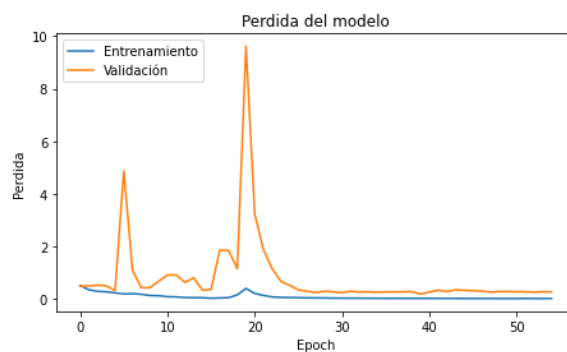
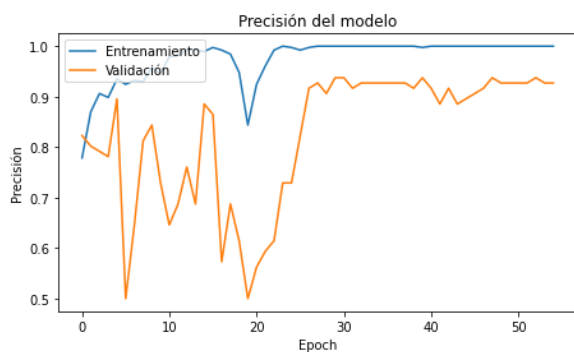
	precision	recall	f1-score
LGG	0.88	0.91	0.89
HGG	0.92	0.90	0.91
macro avg	0.90	0.90	0.90
accuracy	0.90		



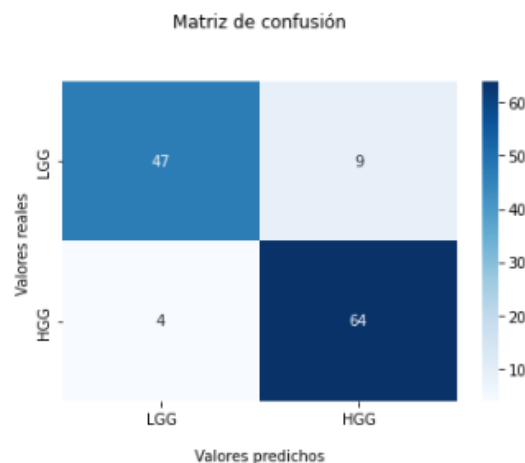
	precision	recall	f1-score
short	0.17	0.03	0.05
mid	0.47	0.62	0.54
long	0.72	0.98	0.83
macro avg	0.45	0.54	0.47
accuracy	0.59		



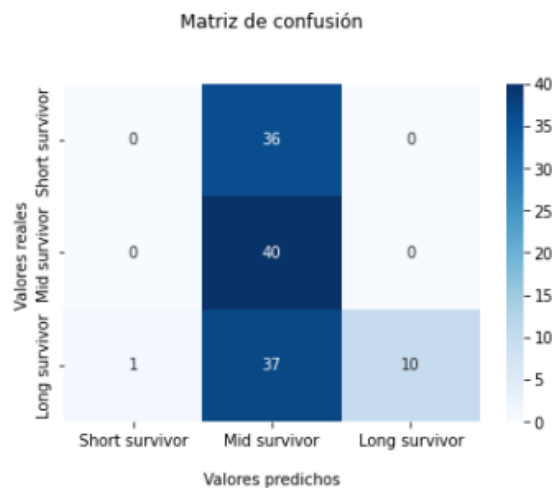
4.5.2.3 128-256 NEURONAS



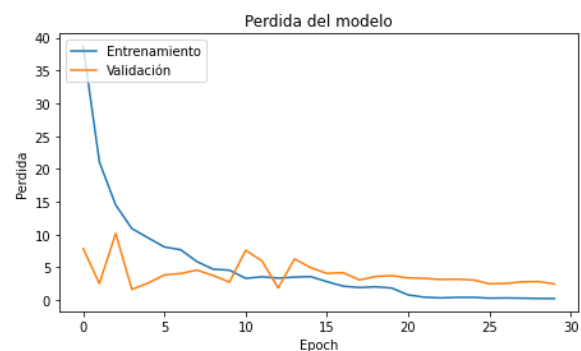
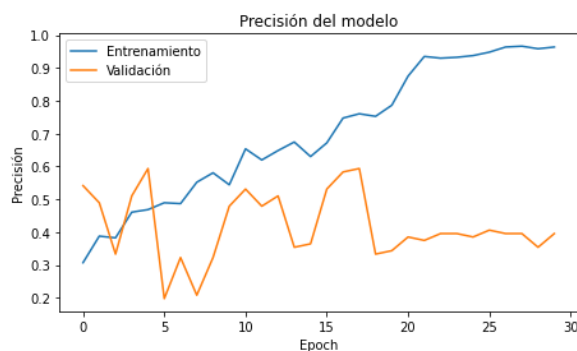
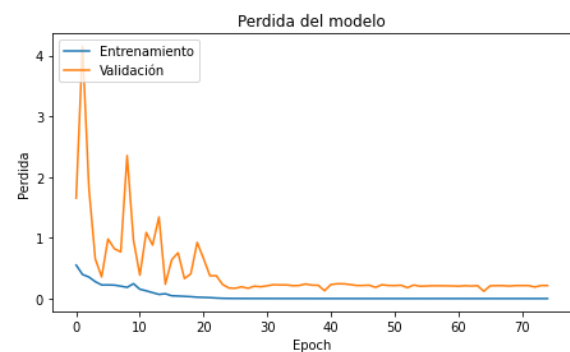
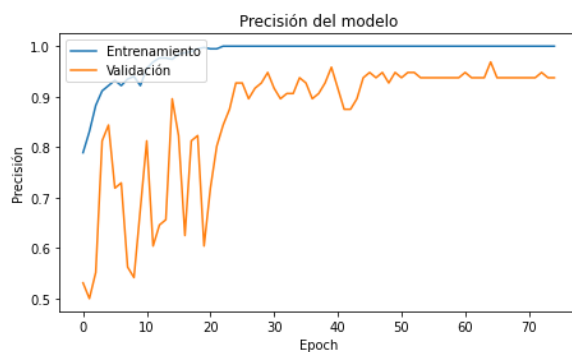
	precision	recall	f1-score
LGG	0.92	0.84	0.88
HGG	0.88	0.94	0.91
macro avg	0.90	0.89	0.89
accuracy	0.90		



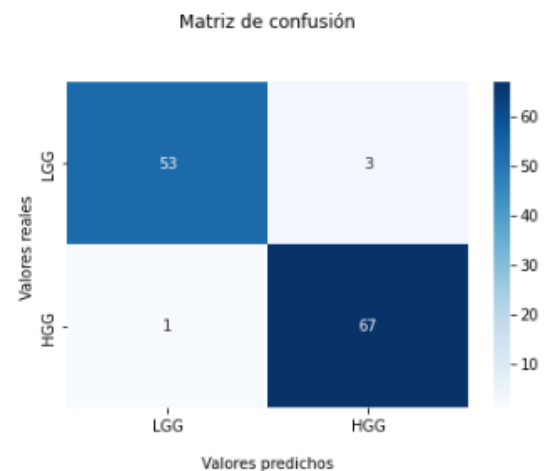
	precision	recall	f1-score
short	0.00	0.00	0.00
mid	0.35	1.00	0.52
long	1.00	0.21	0.34
macro avg	0.45	0.40	0.29
accuracy	0.49		



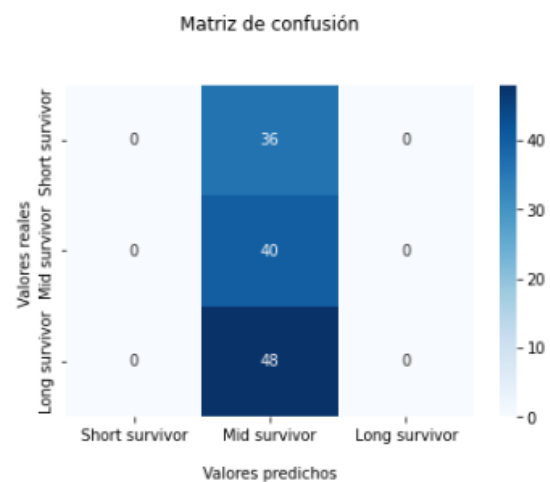
4.5.2.4 256-256 NEURONAS



	precision	recall	f1-score
LGG	0.98	0.95	0.96
HGG	0.96	0.99	0.97
macro avg	0.97	0.97	0.97
accuracy	0.97		



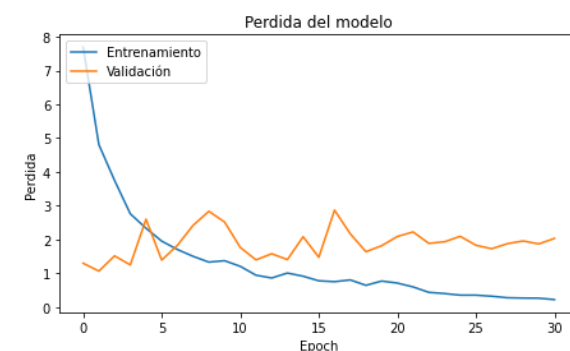
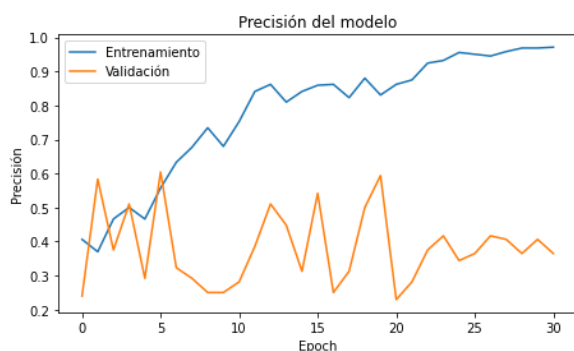
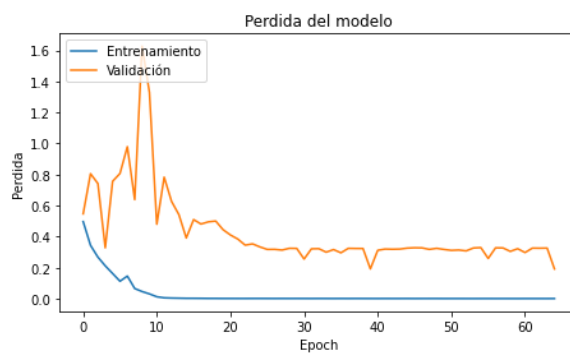
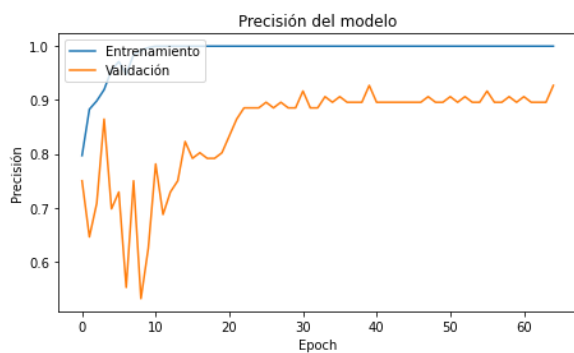
	precision	recall	f1-score
short	0.00	0.00	0.00
mid	0.32	1.00	0.49
long	0.00	0.00	0.00
macro avg	0.11	0.32	0.16
accuracy	0.32		



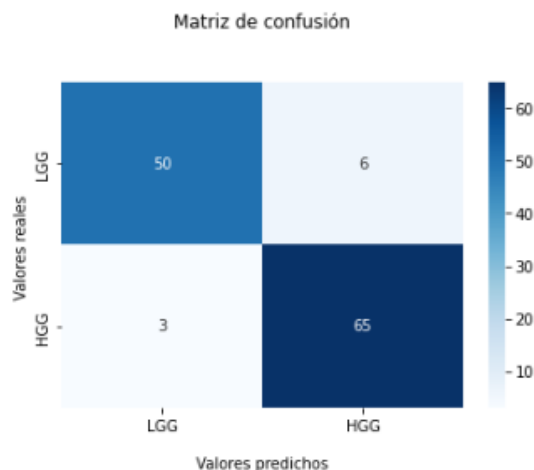
4.5.3 DROPOUT

La capa de Dropout pone un número de entradas aleatorias a 0, esto permite prevenir el sobre entrenamiento de la red ya que la obliga a no utilizar siempre las mismas neuronas para realizar la predicción. El número de entradas que son puestas a cero depende del porcentaje especificado en la capa. Este será el número que cambiaremos para probar el comportamiento de la red cambiando entre una tasa de 0.2, 0.3, 0.4 y 0.5. Los mejores resultados fueron obtenidos en el caso de la predicción de la supervivencia con una tasa de 0.2 y en el caso de la predicción del grado con 0.3.

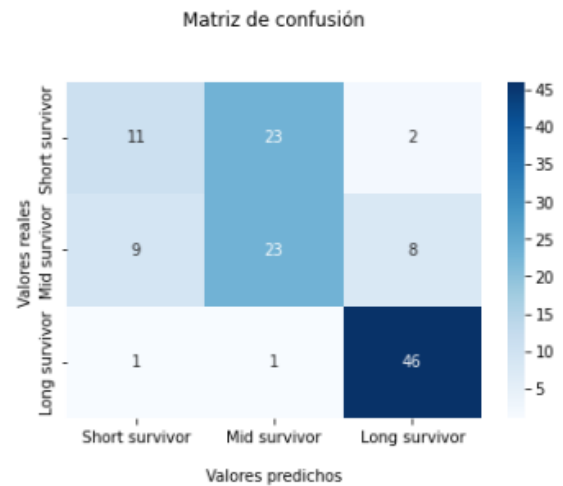
4.5.3.1 DROPOUT 0.2



	precision	recall	f1-score
LGG	0.94	0.89	0.92
HGG	0.92	0.96	0.94
macro avg	0.93	0.92	0.93
accuracy	0.97		

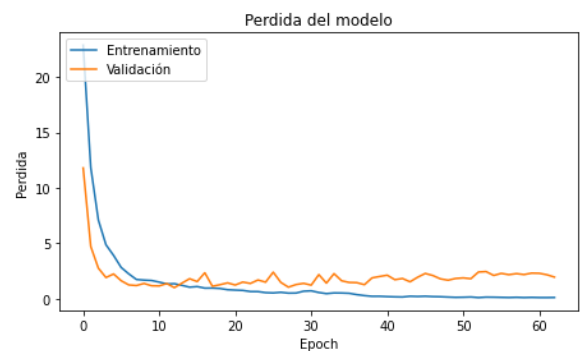
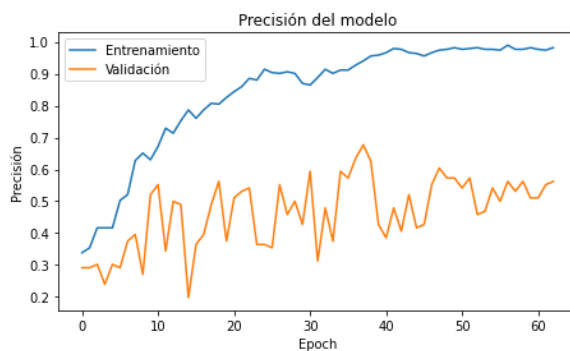
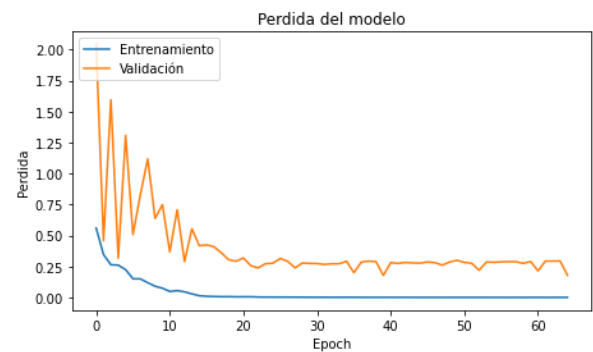
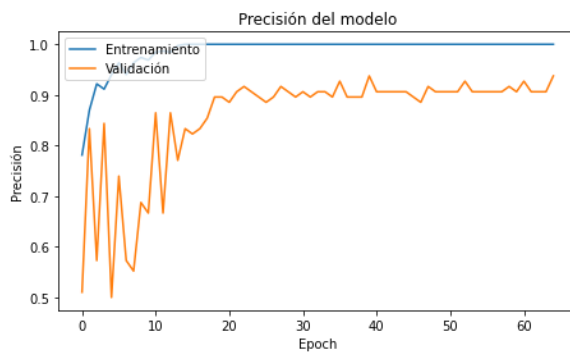


	precision	recall	f1-score
short	0.52	0.31	0.39
mid	0.49	0.57	0.53
long	0.82	0.96	0.88
macro avg	0.61	0.61	0.60
accuracy	0.65		

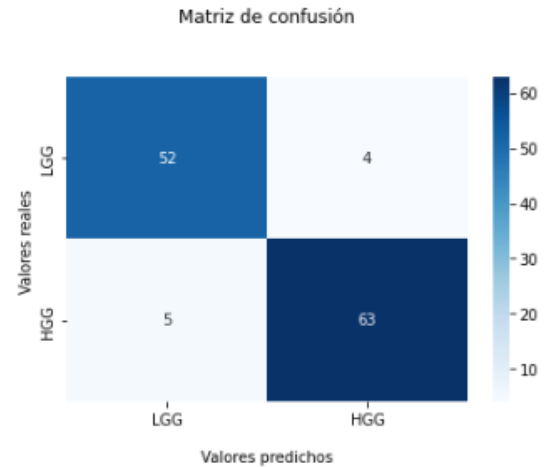


4.5.3.2 DROPOUT 0.3

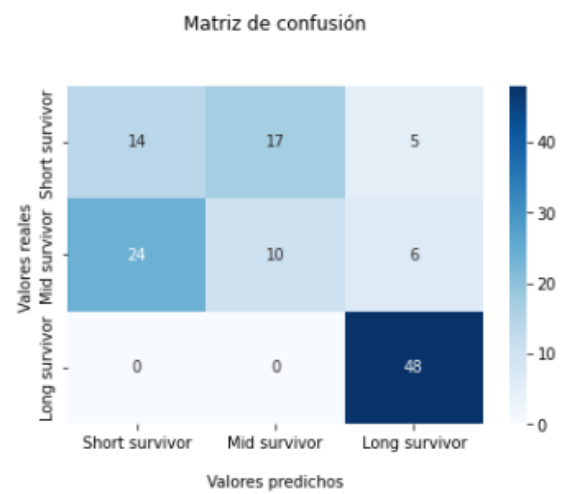
4.5.3.3 DROPOUT 0.4



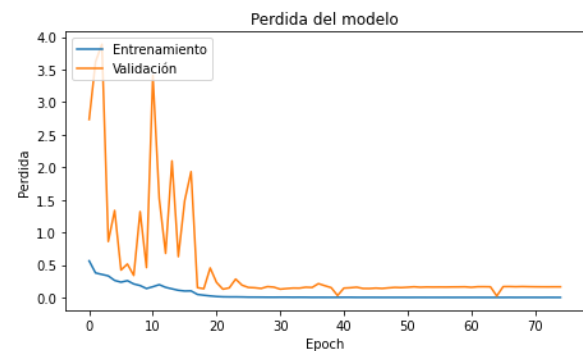
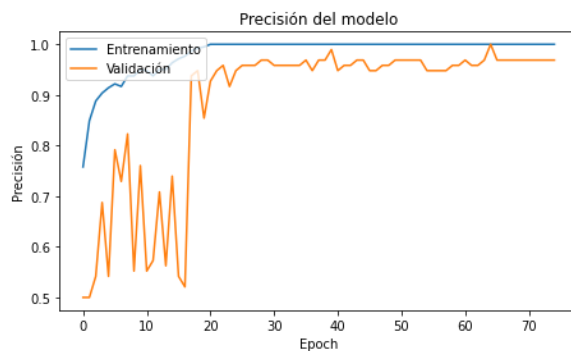
	precision	recall	f1-score
LGG	0.91	0.93	0.92
HGG	0.94	0.93	0.93
macro avg	0.93	0.93	0.93
accuracy	0.93		

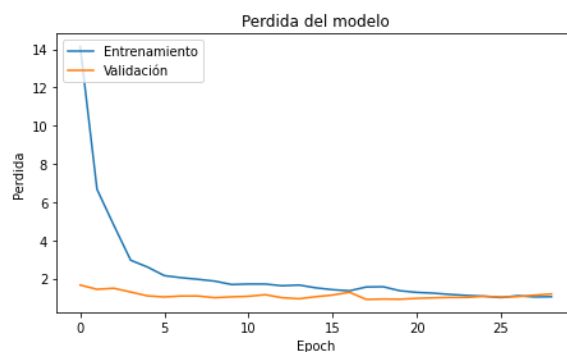
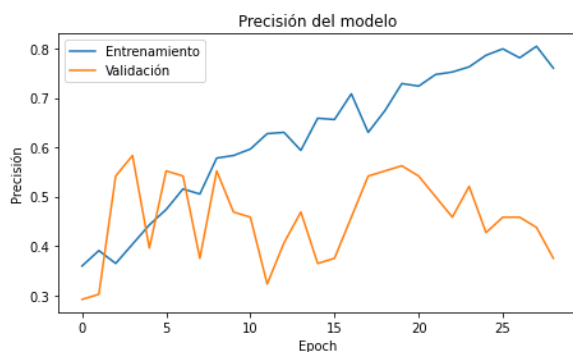


	precision	recall	f1-score
short	0.37	0.39	0.38
mid	0.37	0.25	0.30
long	0.81	1.00	0.90
macro avg	0.52	0.55	0.52
accuracy	0.58		

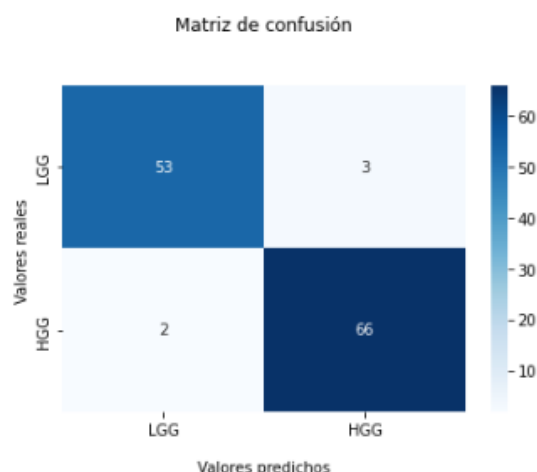


4.5.3.4 DROPOUT 0.5

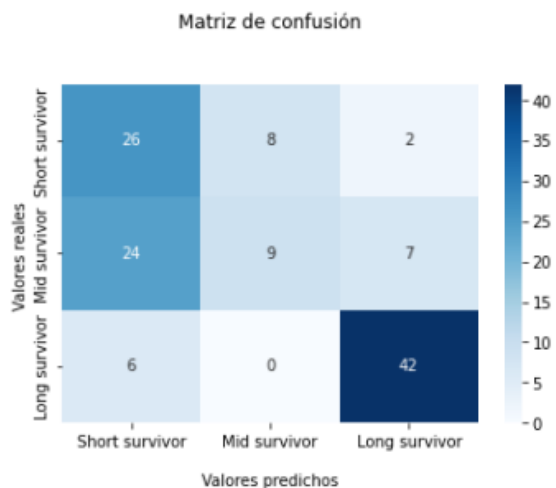




	precision	recall	f1-score
LGG	0.96	0.95	0.95
HGG	0.96	0.97	0.96
macro avg	0.96	0.96	0.96
accuracy	0.96		



	precision	recall	f1-score
short	0.46	0.72	0.57
mid	0.53	0.23	0.32
long	0.82	0.88	0.85
macro avg	0.61	0.61	0.58
accuracy	0.62		

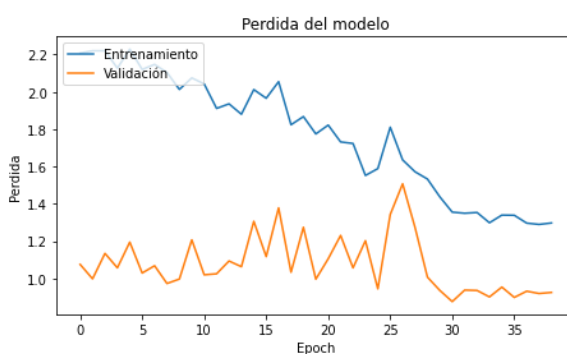
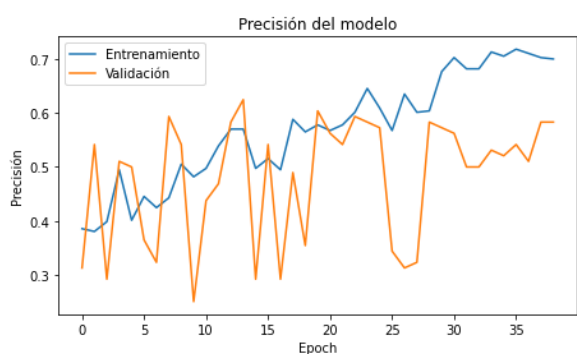
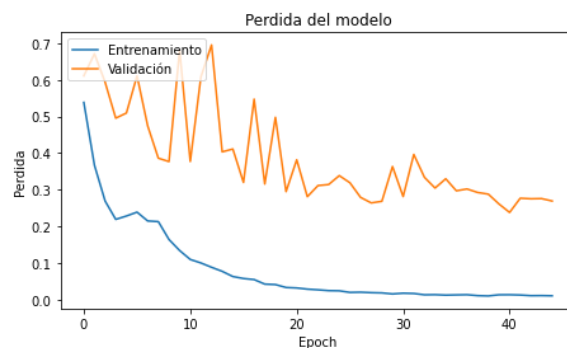
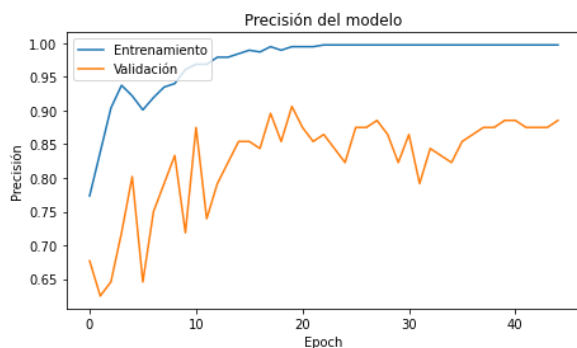


4.5.4 FUNCIÓN DE ACTIVACIÓN

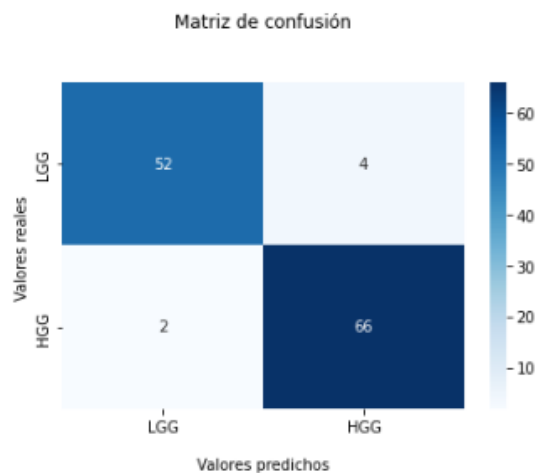
La función de activación transforma la salida dependiendo de su tipo de función. Se han probado las dos más utilizadas. La función relu transforma la salida manteniendo los valores positivos y convirtiendo en 0 los valores negativos, mientras que la función tanh aplica la función tangente a la entrada. Los mejores resultados fueron obtenidos con la función relu en ambos modelos.

4.5.4.1 RELU

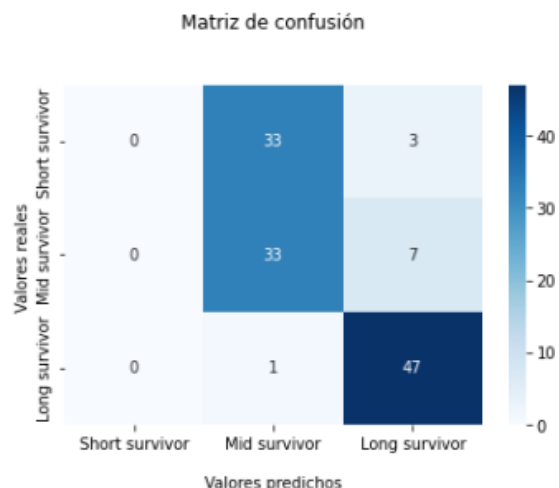
4.5.4.2 TANH



	precision	recall	f1-score
LGG	0.96	0.93	0.95
HGG	0.94	0.97	0.96
macro avg	0.95	0.95	0.95
accuracy	0.95		



	precision	recall	f1-score
short	0.00	0.00	0.00
mid	0.49	0.82	0.62
long	0.82	0.98	0.90
macro avg	0.44	0.60	0.50
accuracy	0.65		



4.5.5 TASA DE APRENDIZAJE

La tasa de aprendizaje indica al modelo el tamaño del paso que debe de dar para actualizar los pesos de las neuronas en base a la perdida del modelo. Este parámetro es fundamental ya que si la tasa de aprendizaje es demasiado pequeña el modelo tardará mucho en llegar a su mínimo, el cual puede ser un mínimo global o local. Pero si la tasa de aprendizaje es demasiado alta el modelo no terminará convergiendo al mínimo. Podemos ver de manera visual esta explicación en la [figura X](#).

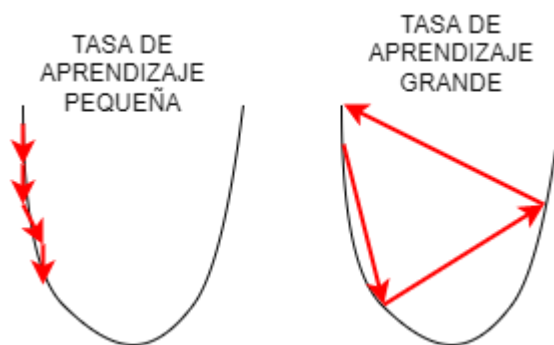
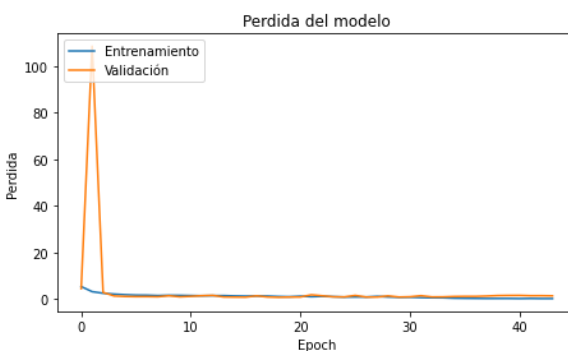
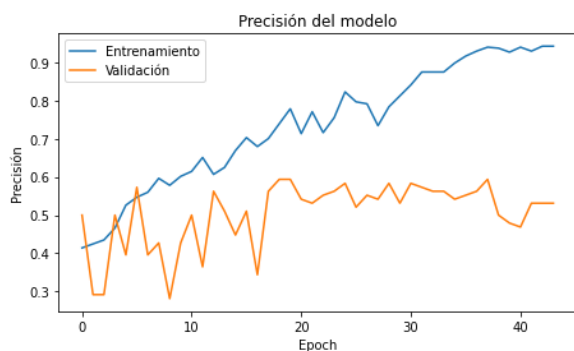
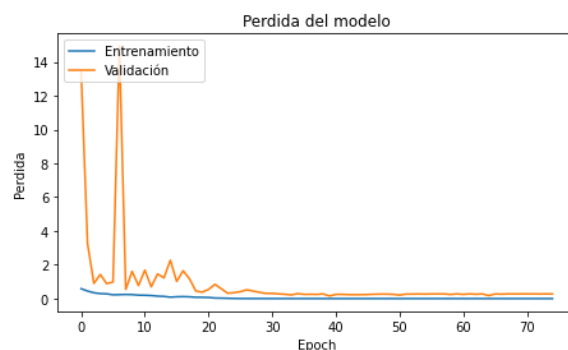
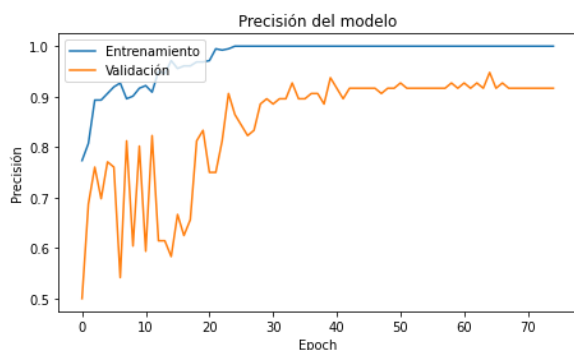


FIGURA X

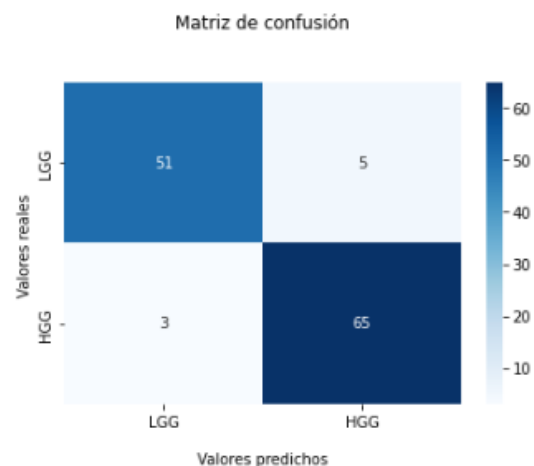
Se han realizado pruebas con una tasa de aprendizaje de 0.0005, 0.001 y 0.002, obteniéndose los mejores resultados en ambos modelos con la tasa de aprendizaje de 0.0005.

4.5.5.1 TASA DE APRENDIZAJE 0.0005

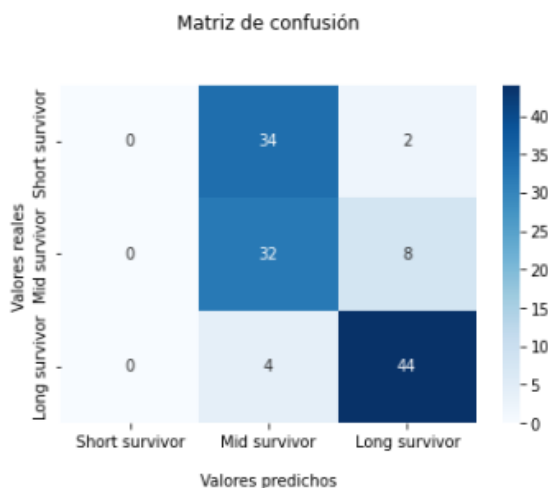
4.5.5.2 TASA DE APRENDIZAJE 0.001



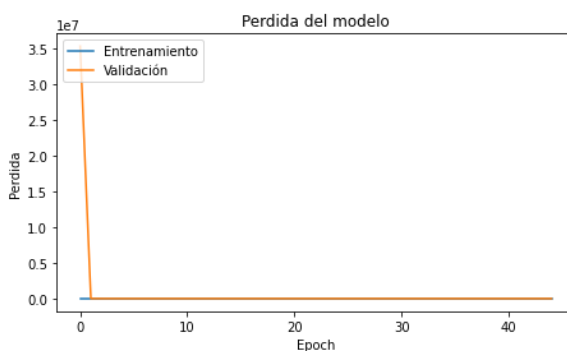
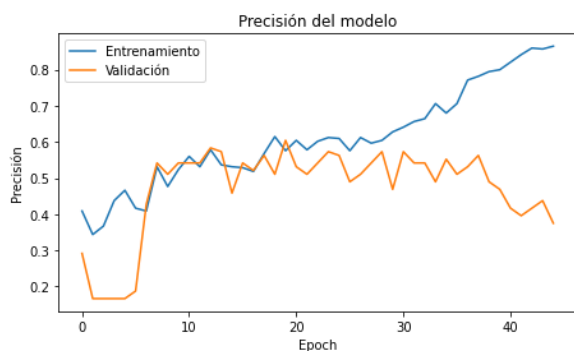
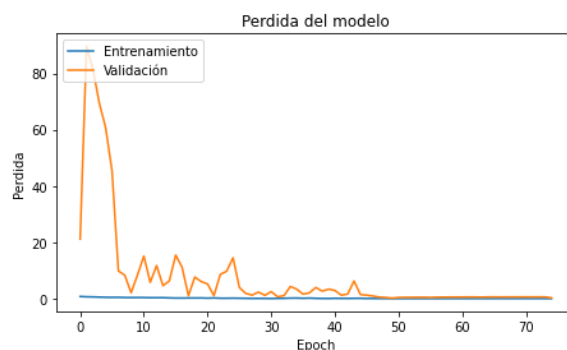
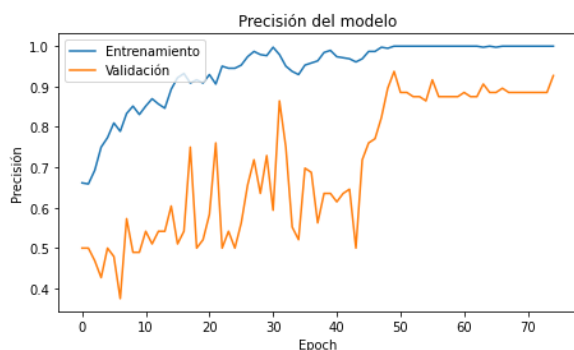
	precision	recall	f1-score
LGG	0.94	0.91	0.93
HGG	0.93	0.96	0.94
macro avg	0.94	0.93	0.93
accuracy	0.94		



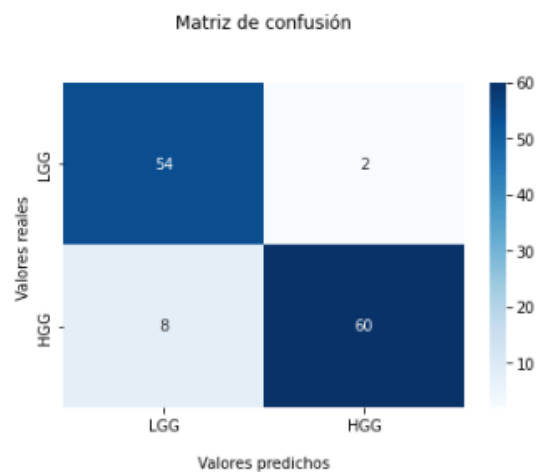
	precision	recall	f1-score
short	0.00	0.00	0.00
mid	0.46	0.80	0.58
long	0.81	0.92	0.86
macro avg	0.42	0.57	0.48
accuracy	0.61		



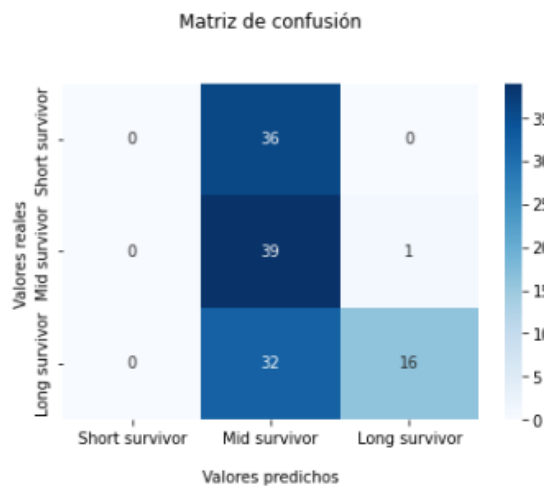
4.5.5.3 TASA DE APRENDIZAJE 0.002



	precision	recall	f1-score
LGG	0.87	0.96	0.92
HGG	0.97	0.88	0.92
macro avg	0.92	0.92	0.92
accuracy	0.92		



	precision	recall	f1-score
short	0.00	0.00	0.00
mid	0.36	0.97	0.53
long	0.94	0.33	0.49
macro avg	0.44	0.44	0.34
accuracy	0.44		



4.5.6 RESTO DE PARÁMETROS

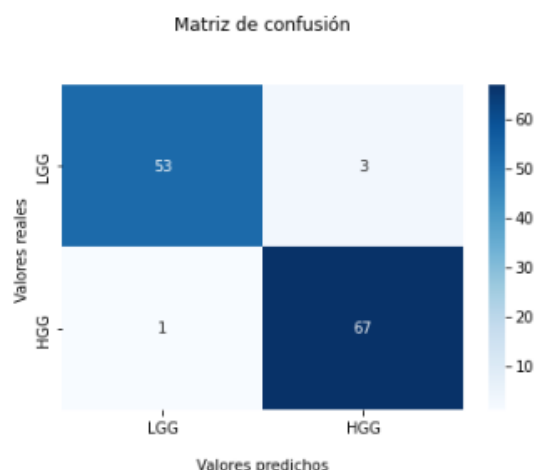
El número de iteraciones utilizados para entrenar los modelos han sido 75, el tamaño del lote utilizado es 16, el optimizador utilizado ha sido Adam, se ha balanceado el peso para cada una de las clases para compensar la diferencia entre ellas, además se han utilizado dos funciones para prevenir el sobre entrenamiento del modelo.

- **EarlyStopping:** Esta función finaliza el entrenamiento si la mejora en la métrica seleccionada es inferior a un cierto umbral tras un cierto número de iteraciones. Se ha utilizado la precisión de validación para realizar el seguimiento, con un umbral de 0.01 y una paciencia de 25 iteraciones. Además, se restauran los mejores pesos obtenidos.
- **ReduceLROnPlateau:** Esta función reduce la tasa de aprendizaje si la mejora en la métrica seleccionada es inferior a un cierto umbral tras un número de iteraciones. Se ha utilizado la métrica de precisión de validación para realizar el seguimiento, con un umbral de 0.01 y una paciencia de 15 iteraciones. La tasa de aprendizaje es reducida en una tasa de 0.1. Es decir, la nueva tasa será igual a la tasa anterior multiplicada por 0.1.

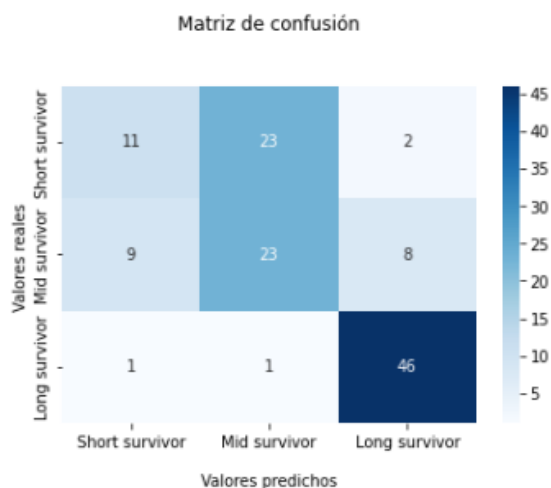
5 Resultados

Finalmente, el modelo ha sido entrenado con los mejores parámetros. En el caso de la clasificación del grado del tumor en HGG y LGG se ha obtenido una precisión del 97% en el conjunto de prueba, mientras que en el caso de la clasificación de la supervivencia en corta, media o larga se ha obtenido un 65% de precisión. Los resultados completos pueden apreciarse a continuación.

	precision	recall	f1-score
LGG	0.98	0.95	0.96
HGG	0.96	0.99	0.97
macro avg	0.97	0.97	0.97
accuracy	0.97		



	precision	recall	f1-score
short	0.52	0.31	0.39
mid	0.49	0.57	0.53
long	0.82	0.96	0.88
macro avg	0.61	0.61	0.60
accuracy	0.65		



6 Conclusiones

Con un 65% de precisión en la clasificación de la supervivencia se mejora el estado del arte actual en un 2.5% el cual había sido obtenido por el estudio [10], mientras que en la clasificación del grado del tumor se mejoró el estado del arte actual [12] en un 7% alcanzando una precisión del 97% en el conjunto de prueba. Lo que demuestra que puede aplicarse únicamente técnicas de aprendizaje profundo para realizar las clasificaciones, mejorando a las técnicas híbridas que usan técnicas convencionales junto con redes convolucionales. Y, además se demuestra el poder de la transferencia de aprendizaje y su utilidad para su aplicación en problemas complejos con una cantidad de datos reducida.



7 Lista de referencias bibliográficas

- [1] <https://www.cancer.gov/espanol/cancer/naturaleza/estadisticas>
- [2] <https://www.nature.com/articles/s41598-020-74419-9#Sec2>
- [3] <https://link.springer.com/article/10.1007/s00401-007-0243-4>
- [4] <https://www.cancer.org/es/tratamiento/como-comprender-su-diagnostico/estadificaciondelcancer.html>
- [5] <https://www.spandidos-publications.com/mco/1/6/935#b3-mco-01-06-0935>
- [6] <https://www.cancer.org/es/cancer/tumores-de-encefalo-o-de-medula-espinal/deteccion-diagnostico-clasificacion-por-etapas/tasas-de-supervivencia.html>
- [7] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6438173/>
- [8] <https://www.med.upenn.edu/cbica/brats2020/data.html>
- [9] <https://www.braintumorsegmentation.org/>

[10]

https://ieeexplore.ieee.org/abstract/document/8251258?casa_token=fmI_gS3pW_IAAAA_A:OAPgshnMiBG1yNSeCTAdmQ4qGutBnht0CMP3WDiAKp7AX9T81VFHyO7wGjdeNI8shMYsGJGxjNA

[11] https://link.springer.com/chapter/10.1007/978-3-030-11726-9_38

[12]

https://ieeexplore.ieee.org/abstract/document/8037508?casa_token=PefqwdE8HjoAAAAA:zIe00ELyE8WzGMJ101aICQcXMkTJjcds43ZYaWOMUAOLpNKHkd3BWcSrt2-M9heonVvqwPF7Hs

[13] <https://www.nature.com/articles/s41598-020-74419-9>

[14] <https://image-net.org/index.php>

[15] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694

[16] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

[17] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge", arXiv preprint arXiv:1811.02629 (2018) [18]

[19] <https://nipy.org/nibabel/>

[20] <https://www.msmanuals.com/es-mx/professional/temas-especiales/principios-de-estudios-por-la-imagen-radiol%C3%B3gicas/resonancia-magn%C3%A9tica>

[21] <https://www.analesdepediatria.org/es-gliomas-bajo-grado-revision-10-articulo-S1695403314000873#:~:text=La%20supervivencia%20global%20fue%20del%2088%2C3%25%2C%20con%20una,fue%20del%2083%2C7%25>

[22]

<https://ascopubs.org/doi/full/10.1200/JOP.2016.018622#:~:text=A%20larger%20study%20of%20216,of%20resection%20was%20%3C%2090%25>



[23] <https://radiopaedia.org/articles/hounsfield-unit?lang=us>

[24] https://doi.org/10.1142/9789813235533_0031

8 Anexos

(<https://ieeexplore.ieee.org/document/9702609>) artículo por revisar