UNIVERSITAT POLITÈCNICA DE VALÈNCIA

_ TELECOM ESCUELA TÉCNICA VLC SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería de Telecomunicación

DISEÑO Y DESARROLLO DE UN SISTEMA DE PREDICCIÓN DE SARCOMA DE EWING MEDIANTE TÉCNICAS DE APRENDIZAJE PROFUNDO SOBRE IMÁGENES HISTOLÓGICAS

Trabajo Fin de Grado

Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación

AUTOR/A: Rubio Gil, Ana

Tutor/a: Naranjo Ornedo, Valeriana

Cotutor/a externo: MESEGUER ESBRI, PABLO

CURSO ACADÉMICO: 2022/2023

Resumen

El sarcoma de Ewing es un tipo de cáncer que se desarrolla en los huesos y en el tejido blando que los rodea. Aunque la incidencia es baja a nivel global, en niños y adolescentes es uno de los tumores más frecuentes y agresivos, lo que genera un gran impacto social y representa un desafío para encontrar un diagnóstico y tratamiento óptimo, ya que la tasa de supervivencia está altamente relacionada con la cantidad de propagación del tumor. En la actualidad, se diagnostica mediante biopsias y su análisis histológico, aunque solamente con esto es difícil diferenciar entre sarcomas de Ewing y otro tipo de tumores de células redondas y pequeñas. Esto es un problema debido a que el tratamiento y pronóstico para cada uno de los diferentes tumores es distinto. Para la implementación de sistemas de ayuda basados en computador que asistan a los médicos en la toma de decisiones, se emplean escáneres de patología digital para la digitalización de las biopsias.

En este proyecto se pretende desarrollar un sistema de clasificación basado en aprendizaje profundo sobre imágenes de histología digitalizadas. Estos tipos de algoritmos son capaces de analizar grandes cantidades de datos y aprender patrones para realizar tareas que pueden llegar a imitar las capacidades humanas. Los métodos explorados se basan en redes neuronales convolucionales que están compuestas por distintas capas interconectadas entre sí para extraer características de las imágenes. El sistema diseñado será capaz de clasificar imágenes de sarcoma de Ewing y otros tipos de tumor de células redondas y pequeñas. Exploraremos distintos algoritmos de aprendizaje que se adaptan a la naturaleza de los datos para ver cuál ofrece mejores resultados, como son el supervisado y el débilmente supervisado, concretamente el de múltiples instancias.

Palabras clave: Sarcoma de Ewing, tumores de células redondas y pequeñas, aprendizaje profundo, clasificación de imágenes, aprendizaje supervisado, aprendizaje débilmente supervisado, aprendizaje de múltiples instancias.

Resum

El sarcoma de Ewing és un tipus de càncer que es desenvolupa en els ossos i en el teixit tou que els envolta. Encara que la incidència és baixa a nivell global, en xiquets i adolescents és un dels tumors més freqüents i agressius, la qual cosa genera un gran impacte social i representa un desafiament per a trobar un diagnòstic i tractament òptim, ja que la taxa de supervivència està estretament relacionada amb la quantitat de propagació del tumor. En l'actualitat, es diagnostica mitjançant biòpsies i l'anàlisi histològica, encara que solament amb això és difícil diferenciar entre sarcomes de Ewing i un altre tipus de tumors de cèl·lules rodones i petites. Això és un problema a causa de que el tractament i el pronòstic per a cadascun dels diferents tumors és diferent. Per a la implementació de sistemes d'ajuda basats en ordinador que assisteixen els metges en la presa de decisions, s'empleen escàners de patologia digital per a la digitalització de les biòpsies.

En aquest projecte es pretén desenvolupar un sistema de classificació basat en aprenentatge profund sobre imatges d'histologia digitalitzades. Aquests tipus d'algoritmes són capaços d'analitzar grans quantitats de dades i aprendre patrons per a realitzar tasques que poden arribar a imitar les capacitats humanes. Els mètodes explorats es basen en xarxes neuronals convolucionals que estan compostes per diferents capes interconnectades entre si per a extreure característiques de les imatges. El sistema dissenyat serà capaç de classificar imatges de sarcoma de Ewing i altres tipus de tumors de cèl·lules rodones i petites. Explorarem diferents algoritmes d'aprenentatge que s'adapten a la naturalesa de les dades per a veure quin ofereix millors resultats, com ara l'aprenentatge supervisat i el dèbilment supervisat, concretament el d'instàncies múltiples.

Paraules clau: Sarcoma d'Ewing, tumors de cèl·lules rodones i xicotetes, aprenentatge profund, classificació d'imatges, aprenentatge supervisat, aprenentatge feblement supervisat, aprenentatge de múltiples instàncies.

Abstract

Ewing sarcoma is a type of cancer that develops in the bones and the soft tissue surrounding them. Although the global incidence is low, in children and adolescents, it is one of the most frequent and aggressive tumors, which generates a significant social impact and represents a challenge to find an optimal diagnosis and treatment, as the survival rate is highly related to the tumor's extent of spread. Currently, it is diagnosed through biopsies and histological analysis, although it is difficult to differentiate between Ewing sarcoma and other types of small round cell tumors with only these methods. This is a problem because the treatment and prognosis differ for each tumor type. For the implementation of computer-aided diagnosis systems that assist doctors in decision-making, digital pathology scanners are used for the digitization of biopsies.

This project aims to develop a classification system based on deep learning on digitized histology images. These types of algorithms can analyze large amounts of data and learning patterns to perform tasks that can mimic human capabilities. The explored methods are based on convolutional neural networks, which consist of interconnected layers to extract features from the images. The designed system will be able to classify images of Ewing sarcoma and other types of small round cell tumors. We will explore different learning algorithms that adapt to the nature of the data to determine which one offers better results, such as supervised learning and weakly supervised learning, specifically multiple-instance learning.

Keywords: Ewing sarcoma, small round cell tumors, deep learning, image classification, supervised learning, weakly supervised learning, multiple instance learning.

A mis padres por su constante apoyo y por estar a mi lado en cada decisión que he tomado a lo largo de mis años universitarios, a Pablo por su tiempo y ayuda incalculable durante todo el desarrollo de este trabajo y a Valery por brindarme la oportunidad de llevar a cabo este proyecto y todos los aprendizajes que ha conllevado.

Índice general

Índice de figuras

Índice de tablas

Parte I

Memoria

Capítulo 1

Introducción

En esta primera parte se hará una breve explicación tanto del contexto médico como del aprendizaje profundo, ya que estos son los dos pilares fundamentales de este proyecto. Además, se presentará el estado del arte y se definirá un objetivo principal y los objetivos secundarios del proyecto.

1.1. Contexto médico

Los tumores son crecimientos anormales de células en el cuerpo que se producen cuando hay un desequilibrio entre la división y el crecimiento celular. Normalmente, el cuerpo regula la cantidad de células producidas, eliminando las que están dañadas o que ya no son necesarias y reemplazándolas por células nuevas y saludables. Sin embargo, si se altera este proceso, las células pueden comenzar a dividirse y multiplicarse excesivamente, formando tumores.

Los tumores pueden aparecer en cualquier parte del cuerpo y pueden ser clasificados en benignos y malignos. Los benignos son aquellos que solo crecen en una zona concreta y no pueden aparecer en ni invadir otras. Por otro lado, los malignos, también llamados cancerosos, son los que sí pueden crecer y diseminarse a otras partes del cuerpo [1]. Dentro de los tumores malignos o cancerosos existen los sarcomas, que son un tipo de cáncer que empieza en los huesos o en los tejidos blandos del cuerpo, como los cartílagos, los músculos o los vasos sanguíneos. Los tipos de sarcomas son clasificados según el lugar donde se forman [2]. Algunos ejemplos de ellos son los angiosarcomas, que se forman en las células que revisten los vasos sanguíneos o los vasos linfáticos, los rabdomiosarcomas, que se suelen formar en los músculos que se unen a huesos y que ayudan a mover el cuerpo y los condrosarcomas, que por lo general comienza en los huesos.

El Sarcoma de Ewing (SE) es un tipo de tumor óseo maligno poco común que se caracteriza por ser altamente agresivo y por poder propagarse rápidamente a otras partes del cuerpo [3]. Generalmente, comienza en los huesos largos, como el fémur o el hueso de la pelvis, pero también puede originarse en otros huesos, tejidos blandos o en la médula ósea. Se caracteriza por la presencia de una anomalía genética llamada translocación cromosómica, en la cual fragmentos de dos cromosomas diferentes se fusionan, generalmente el cromosoma 11 y el 22. Esta fusión da lugar a una proteína anormal llamada proteína de fusión EWSR1-FLI1 [4], la cual juega un papel importante en su desarrollo.

Este tipo de tumor es más común en adolescentes y adultos jóvenes, pero puede aparecer en cualquier edad. Según la American Society of Clinical Oncology [5], cada año, aproximadamente a 200

niños y adolescentes en los Estados Unidos se les diagnostica un sarcoma de Ewing. Los tumores de Ewing representan el 1 % de todos los cánceres infantiles diagnosticados en niños y adolescentes menores de 15 años y el 2 % de todos los cánceres diagnosticados en adolescentes de 15 a 19 años. Aproximadamente la mitad de todos los diagnósticos de sarcoma de Ewing se dan en personas de entre 10 y 20 años. Los tumores de Ewing también pueden afectar niños más pequeños y adultos jóvenes de entre 20 y 30 años.

La tasa de supervivencia a 5 años es un indicador que muestra el porcentaje de pacientes que han sido diagnosticados con una enfermedad y que siguen vivos al menos durante 5 años después de ese diagnóstico. En el caso del sarcoma de Ewing, la tasa general de supervivencia a 5 años para personas con un tumor de Ewing es 62 %, pero hay distintos factores que afectaran al valor de este indicador. Teniendo en cuenta el factor de la edad, en niños menores de 15 años la tasa de supervivencia a 5 años es del 75 %, mientras que para los adolescentes de 15 a 19 años es del 58 %. Por otro lado, considerando la extensión de la propagación del tumor, si el tumor se encuentra únicamente en su lugar de origen, es decir, está localizado, la tasa de supervivencia es 82 %. Si el tumor se ha diseminado a una región cercana, se considera regional, y la tasa de supervivencia es del 67 %. Sin embargo, si el tumor se ha propagado a una región distante en el momento del diagnóstico, se denomina metástasis, y la tasa de supervivencia disminuye al 39 %. Se puede apreciar que a mayor propagación hay una menor esperanza de supervivencia [5].

Propagación del tumor	**Tasa supervivencia a 5 años vista**
Localizado	82 %
Regional	67 %
Metástasis	39 %
General	62 %

Tabla 1.1: Tasa de supervivencia a 5 años según propagación del SE [5]

En los últimos 40 años, la tasa de supervivencia del sarcoma de Ewing ha experimentado un notable incremento. Este avance se debe principalmente al uso de la quimioterapia y al tratamiento multidisciplinar. A pesar de estos avances, el pronóstico de la enfermedad sigue siendo desfavorable, especialmente en casos de metástasis, que es considerado el principal factor pronóstico adverso [6]. Aunque el sarcoma de Ewing es una enfermedad poco común, en España es el tumor óseo maligno primario más frecuente en la infancia, superando incluso al osteosarcoma. Dado el mal pronóstico asociado a la diseminación avanzada del tumor y su frecuencia, es crucial realizar una detección temprana que permita iniciar un tratamiento precoz con atención médica especializada. Estos aspectos son fundamentales para mejorar el pronóstico y aumentar las posibilidades de recuperación, ya que como se evidencia en la Tabla 1.1 la tasa de supervivencia está altamente relacionada con la fase de propagación del tumor cuando es detectado.

Otro tipo de sarcoma de tejido blando es el de células redondeadas y pequeñas (TCRP). Es altamente maligno y suele aparecer en la infancia y adolescencia. Su nombre se debe a su apariencia altamente celular y a que a menudo carece de marcadores morfológicos específicos para su identificación [7]. Su diferenciación con el SE es difícil de apreciar, pero es vital contar con técnicas que la puedan llevar a cabo. Esto es porque las terapias empleadas para tratar cada uno de ellos es distinta y, por lo tanto, se necesita un diagnóstico diferencial para posibilitar la recuperación del paciente.

Diagnóstico

Una vez establecida la sospecha de que un paciente padece algún tipo de cáncer óseo, se procede al diagnóstico de este. Hay una variedad de pruebas que se pueden hacer para determinar si hay presencia de tumor y de que tipo es [8], y aunque no todas puedan usarse individualmente como diagnóstico definitivo, cada una aporta una información distinta. Una de ellas es la radiología simple, que consiste en la obtención de imágenes de una lesión osteolítica. Aunque este tipo de pruebas pueden no ser lo suficientemente sensibles para detectar tumores pequeños o lesiones óseas tempranas, en el caso de ser detectado puede proporcionar información importante sobre la presencia y las características del tumor. También existen otras técnicas de imagen como la Tomografía Axial Computerizada (TAC) y la Resonancia Magnética Nuclear (RMN). Estas permiten determinar la extensión del tumor, tanto en el hueso como en los tejidos blandos vecinos.

La técnica definitiva de diagnóstico es la que se realiza mediante la extracción de una biopsia y es la que emplearemos para obtener las imágenes que utilizaremos para nuestro modelo de aprendizaje. Se puede realizar con aguja gruesa guiada por TAC o con cirugía [9], como se puede apreciar en la Figura 1.1. Una vez se obtiene el tejido, se prepara mediante técnicas histológicas para su estudio histopatológico. Pueden ser necesarios otros estudios como la inmunohistoquímica y marcadores tumorales para realizar el diagnóstico diferencial con otros tumores.

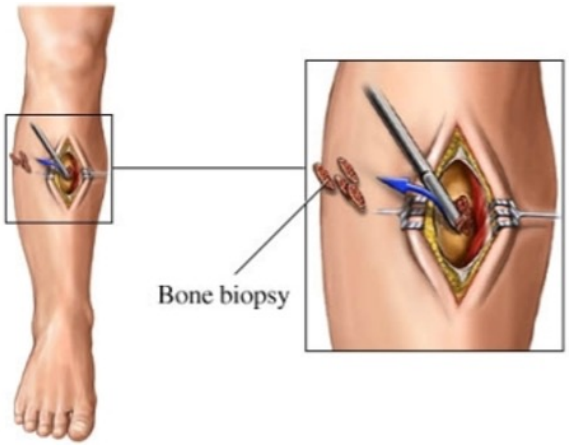

Figura 1.1: Biopsia de hueso mediante cirugía (obtenido de [10])

Para la creación de la base de datos de este proyecto se utilizarán imágenes conseguidas a partir de un tratamiento digital de las biopsias. Hay que llevar a cabo ciertos procesos para llegar a tener las imágenes de la base de datos que será utilizada para el sistema de aprendizaje. El primer paso es aplicar la técnica histológica a la biopsia. Esta consiste en hacer cortes delgados a la muestra para poder observarlos bajo el microscopio. A continuación, se lleva a cabo la tinción de las muestras con Hematoxilina Eosina (HE). De estos dos colorantes, la hematoxilina tiñe de violeta azulado intenso los ribosomas, la cromatina (material genético) dentro del núcleo y otras estructuras y la eosina tiñe de rosa anaranjado o rosado el citoplasma, el colágeno, el tejido conjuntivo y otras estructuras que rodean y sostienen la célula [11]. Una vez se tienen las muestras preparadas y teñidas, se lleva a cabo la digitalización mediante escáneres de patología digital, produciendo micromatrices de tejido digitalizado (TMA, del inglés *tissue microarrays*) por el proceso de digitalización llamado *whole slide imaging*. Este nombre se le da al proceso de digitalización, pero su resultado puede ser tanto los TMAs como las *whole slide images* (WSI). La diferencia entre estas dos es que

las WSI captan secciones histológicas completas de tejido para ser analizado, por lo que su tamaño es mucho mayor. Por otro lado, los TMAs contienen más de un núcleo de tejido, pero cada uno de estos núcleos individualmente es de mucho menor tamaño. En la Figura 1.2 se puede ver a la izquierda un ejemplo de TMA y a la derecha uno de WSI para poder apreciar sus diferencias.

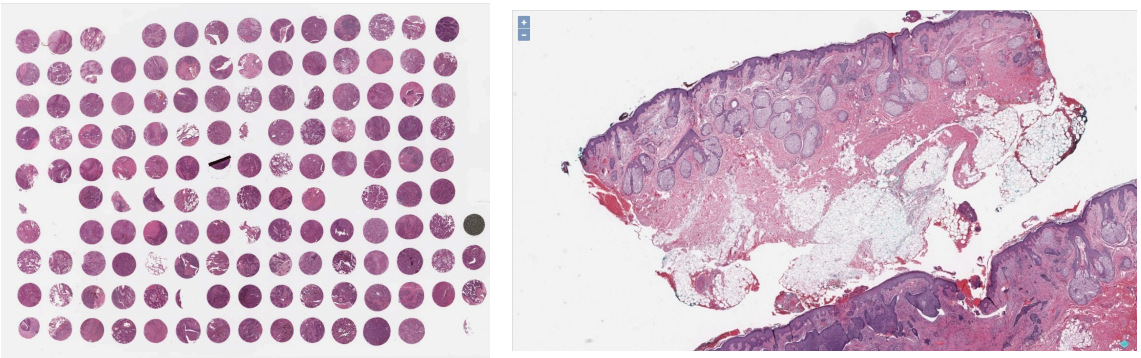

Figura 1.2: Ejemplo de TMA y WSI (obtenidas de [12] y [13]

1.2. Inteligencia artificial

Según la definición ofrecida por la página web de noticias del Parlamento Europeo [14]:

«La inteligencia artificial es la habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear.»

La inteligencia artificial (IA) posibilita que los sistemas tecnológicos tengan la capacidad de comprender su entorno, interactuar con él, encontrar soluciones a problemas y llevar a cabo acciones específicas. La máquina recibe información en forma de datos, los cuales pueden estar previamente preparados o recopilados a través de sensores incorporados, como una cámara. Luego, procesa estos datos y genera respuestas en función de ellos. Los sistemas de IA tienen la capacidad de ajustar su comportamiento en cierta medida, analizar las consecuencias de acciones anteriores y funcionar de manera independiente.

Dentro del ámbito de la inteligencia artificial, existe una rama conocida como aprendizaje automático o Machine Learning (ML). Su objetivo principal es desarrollar algoritmos y técnicas que permitan a las máquinas aprender de forma autónoma a partir de datos, sin necesidad de ser programadas de manera explícita. En el ML, se emplean modelos matemáticos y estadísticos para analizar conjuntos extensos de datos y descubrir patrones y regularidades en ellos. Estos modelos se entrenan utilizando datos de entrada, como imágenes, texto o sonido, junto con su correspondiente salida, ya sea etiquetas o clasificaciones o sin ella. De esta manera, los modelos pueden hacer predicciones sobre nuevas instancias de datos que nunca han sido vistas anteriormente.

1.2.1. Deep learning

El aprendizaje profundo (DL, del inglés *deep learning*), es una rama dentro del campo de la inteligencia artificial.

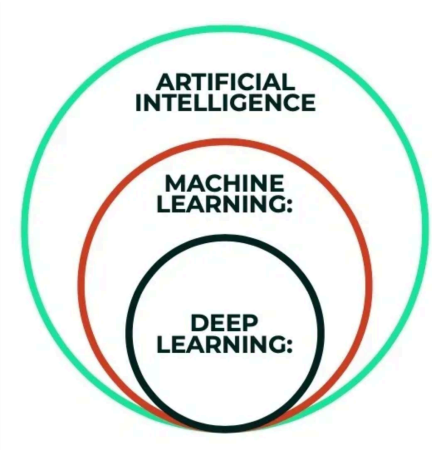

Figura 1.3: Artificial Intelligence - Machine Learning and Deep Learning (obtenida de [15])

Este tipo de aprendizaje se centra en el desarrollo de algoritmos y modelos de aprendizaje automático inspirados en el funcionamiento del cerebro humano. A diferencia de otros enfoques de aprendizaje automático, el DL se basa en redes neuronales artificiales (ANNs, del inglés *Artificial Neural Networks*) que están compuestas por múltiples capas de neuronas interconectadas. Puede haber decenas o cientos de capas que se encargan de procesar datos y extraer características complejas y abstractas para poder predecir la salida de un set de datos.

Una ANN es un modelo computacional que se comporta de manera similar a la de una neurona del sistema nervioso humano. Este tipo de redes están compuestas por múltiples capas, conectadas de manera que la salida de una de las capas es la entrada de la siguiente. La primera capa es denominada la capa de entrada, la última la capa de salida y las intermedias son las capas ocultas.

1.2.2. Algoritmos de aprendizaje

Los algoritmos de aprendizaje se dividen en diferentes tipos según qué etiquetas acompañan a los datos de entrada. Estos tipos incluyen el aprendizaje supervisado, no supervisado, semi-supervisado, por refuerzo y débilmente supervisado. La elección del algoritmo se decide en base a la aplicación que se va a dar al modelo y los datos que se disponen para su realización.

Aprendizaje supervisado

En el aprendizaje supervisado, se entrena un modelo utilizando datos que están etiquetados, es decir, cada dato de entrada tiene una etiqueta que indica la respuesta correcta. El objetivo del modelo es encontrar una función que pueda asignar correctamente las etiquetas a nuevos datos de entrada para que estas coincidan con el *ground truth* de los datos, que es su valor verdadero. Durante el entrenamiento, el algoritmo ajusta los parámetros del modelo basándose en los resultados de épocas anteriores para obtener mejores resultados en cada iteración. Para obtener buenos resultados en este tipo de aprendizaje, es necesario contar con una cantidad suficiente de datos de entrenamiento que estén correctamente etiquetados y no contengan ruido en las anotaciones.

Aprendizaje no supervisado

En el aprendizaje no supervisado contamos con datos sin etiquetar, es por eso que el sistema tiene que intentar entenderlos por sí mismo. En este tipo de aprendizaje se realiza un análisis para

observar patrones ocultos y así hacer agrupaciones según los datos que tengan similitudes.

Aprendizaje semisupervisado

El aprendizaje semisupervisado es un enfoque intermedio entre el aprendizaje supervisado y el aprendizaje no supervisado, ya que se utilizan conjuntos de datos que tienen tanto ejemplos etiquetados como ejemplos sin etiquetar.

Aprendizaje por refuerzo

El aprendizaje por refuerzo es una técnica de aprendizaje automático en la que un agente aprende a través de interacciones prueba-error con un entorno. El objetivo del agente es tomar acciones en el entorno que le lleven a obtener la máxima recompensa a lo largo del tiempo.

Aprendizaje debilmente supervisado

En el aprendizaje débilmente supervisado se tienen datos de entrada que están parcialmente etiquetados. Dentro de estos algoritmos hay un tipo llamado aprendizaje de múltiples instancias (MIL, del inglés *multiple instance learning*). Este consiste en la agrupación de los datos de entrenamiento en bolsas y etiquetado de estas en vez de etiquetado por cada instancia. Es decir, se tiene una bolsa que es acompañada por una etiqueta y contiene una determinada cantidad de datos. Una bolsa será considerada positiva con tal de contener una instancia que pertenezca a la clase positiva.

1.2.3. Redes neuronales artificiales

Perceptrón multicapa

El perceptrón multicapa (MLP, del inglés *multilayer perceptron*) es un tipo específico de ANN que se caracteriza por tener múltiples capas ocultas con neuronas completamente conectadas. En cada neurona se recibe una o más entradas, se realizan unas operaciones matemáticas y su resultado produce una salida. El procesamiento consiste que las entradas son ponderadas por unos pesos y sumadas y su resultado es pasado por una función de activación, para devolver una única salida, que será la entrada de la siguiente capa. Los pesos de los nodos de cada neurona determinan cuanta importancia tiene esa entrada en el resultado a la salida. Es por eso por lo que su actualización es necesaria para el proceso de aprendizaje del modelo.

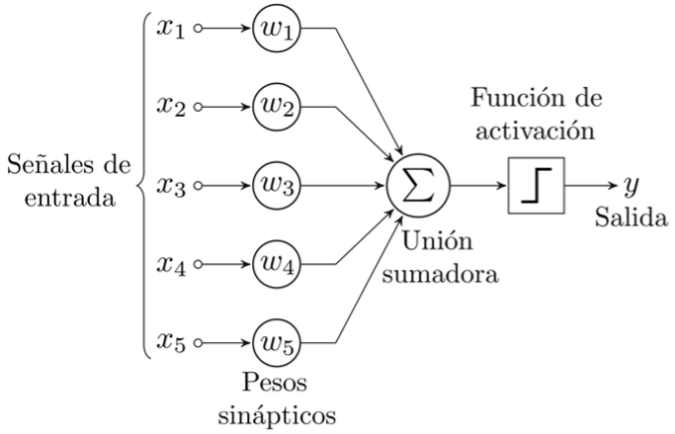

Figura 1.4: Perceptrón multicapa (obtenida de [16])

Funciones de activación

Las funciones de activación son una parte que destacar de las ANNs, ya que tienen un papel crucial en el proceso de cálculo de la salida de cada neurona, que será la entrada de la siguiente, y son capaces de introducir no linealidad, lo que permite modelar relaciones no lineales en los datos. Algunas de las funciones de activación más usadas son la función unidad linear rectificada (ReLU), sigmoide, softmax y tangente hiperbólica (tanh). En la Figura 1.5 se muestra un ejemplo de cada una de las funciones mencionadas en esta sección.

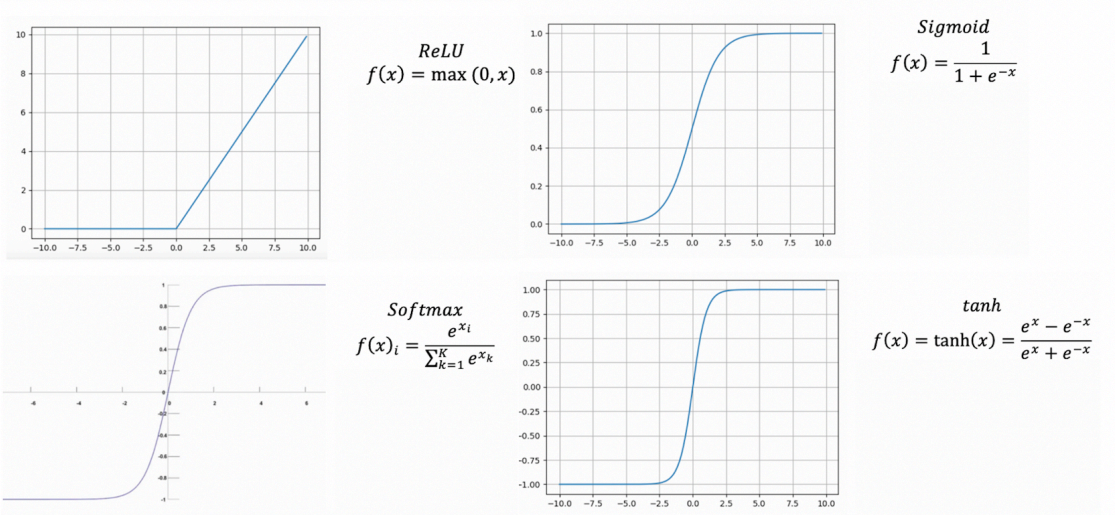

Figura 1.5: Funciones de activación (obtenida de [17])

Proceso de entrenamiento

Para entrenar una ANN, las dos etapas que se deben llevar a cabo son la propagación hacia delante, en inglés *forward propagation*, y la retropropagación, en inglés *back propagation*. Estas se repiten iterativamente durante un número de veces marcado, llamado épocas, hasta que el modelo diseñado minimiza la función de pérdidas y alcanza la convergencia.

En la *forward propagation*, los datos de entrada se propagan por la red neuronal hasta la salida. En el proceso de propagación, se les aplican las funciones de activación para obtener una predicción de los datos de entrada a la salida. Con las predicciones se calcula la discrepancia entre la salida predicha y el valor real de la etiqueta del dato que se está observando. Este error se puede calcular con las llamadas funciones de pérdidas. Entre ellas están la entropía cruzada y el error cuadrático medio (MSE, del inglés *mean square error*).

En la retropropagación, al contrario que en la forward propagation, lo que se propaga es el error y lo hace al sentido contrario, de la salida a la entrada. Esto se hace para que las neuronas de la red ajusten sus pesos para así minimizar el error en la predicción a la salida y, por lo tanto, obtener un mayor rendimiento del modelo. La magnitud de la actualización de los pesos se determina con una métrica llamada la tasa de aprendizaje, que se especifica antes del entrenamiento, y determina si los pasos para llegar al mínimo local deben ser más grandes o más pequeños.

Redes neuronales convolucionales

Una Red Neuronal Convolucional (CNN, del inglés *Convolutional Neural Network*), es un tipo de arquitectura de red neuronal que evoluciona del MLP para adaptarse al procesamiento de datos como imágenes, videos y audios. La principal aportación de las CNNs es que, gracias a sus capas convolucionales, son capaces de analizar espacialmente las entradas de la red para obtener una salida, a diferencia de los MLPs, que realizan una combinación lineal de sus entradas. Su nombre se debe a la aplicación de operaciones de convolución entre capas, que son operaciones matemáticas entre matrices, para extraer características de los datos. Están formadas por tres capas, que son las convolucionales, las de reducción, en inglés *pooling* y las densas de neuronas completamente conectadas, en inglés *fully connected*.

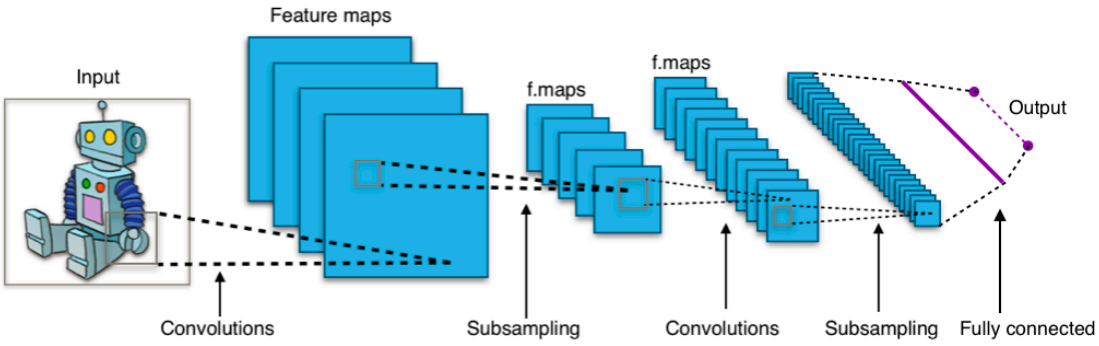

Figura 1.6: Estructura de una Red Neuronal Convolucional (obtenida de [18])

- Capas *fully connected*: Estas capas también son llamadas capas densas y su objetivo es llevar a cabo la clasificación, asignando una probabilidad a cada clase. A la entrada se tiene un vector de características de la imagen y a la salida tiene el mismo número de salidas como clases se está clasificando. En la Figura 1.7 se puede ver un ejemplo de este tipo de capas y como todas sus entradas y salidas están conectadas.

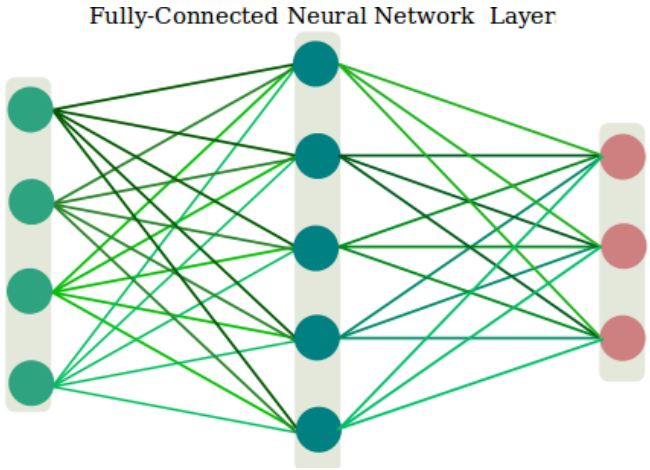

Figura 1.7: Ejemplo de una capa de *fully connected* (obtenida de [19])

- Capas de pooling: Estas capas ofrecen la posibilidad de reducir la dimensionalidad de los mapas de activación para quedarse solo con los valores más significativos. El tipo de capa de *pooling* más utilizado es *maxpooling* y en ella se seleccionan los valores máximos de unos subconjuntos de la matriz de características. En la Figura 1.8 se puede ver un esquema que ejemplifica la implementación de esta capa en la que se seleccionan los valores máximos de subconjuntos de matrices 2x2 para así reducir el tamaño de la matriz a la mitad.

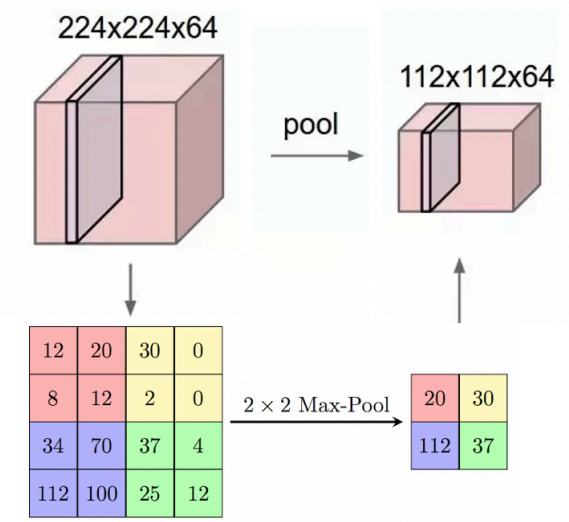

Figura 1.8: Ejemplo de una capa de *maxpooling* (obtenida de [20])

- Capas convolucionales: Estas capas se encargan de la extracción de características de las imágenes mediante una serie de operaciones que son las convoluciones y las funciones de activación. La convolución consiste en el producto elemento a elemento de una pequeña matriz llamada *kernel* sobre la matriz en la que se representa el valor de cada pixel de la imagen. Este proceso se repite deslizando el *kernel* por toda la imagen.

1.2.4. Aprendizaje de transferencia

El aprendizaje de transferencia, en inglés *transfer learning*, es un concepto utilizado en el campo del aprendizaje automático que se refiere a la aplicación y transferencia de conocimientos o habilidades aprendidos en una tarea o dominio a otro diferente pero relacionado. En lugar de comenzar el entrenamiento desde cero en un nuevo problema, como se haría en el caso del *machine learning* tradicional, el *transfer learning* busca aprovechar el conocimiento previo adquirido para mejorar el rendimiento en la tarea actual. La oportunidad que ofrece el aprendizaje de transferencia es la posibilidad de resolver problemas que requieren grandes cantidades de datos cuando no se dispone de ellos. En la Figura 1.9 se puede ver a la izquierda una representación del *machine learning* tradicional y a la derecha del *transfer learning*. Se puede observar como a la derecha los datasets son de tamaños distintos, siendo el Dataset 1 significantemente mayor que el 2, es por eso que se aplica lo aprendido del primero para que el aprendizaje del segundo ofrezca mejores resultados.

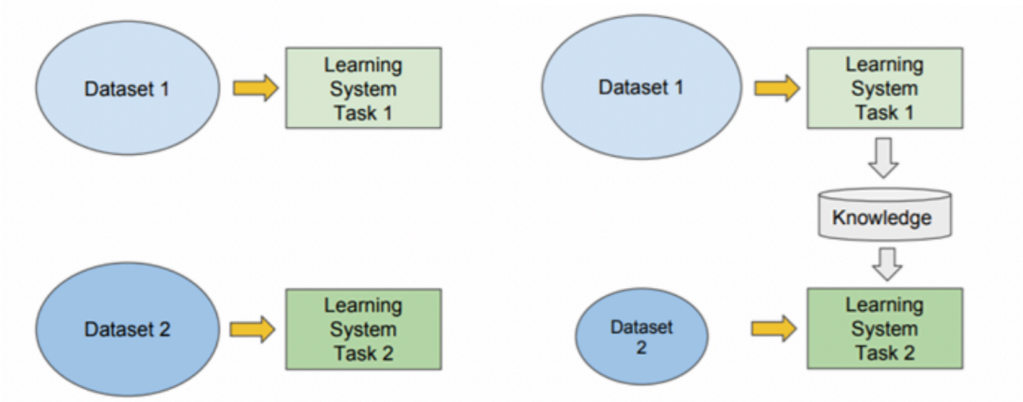

Figura 1.9: Diferencia entre aprendizaje de transferencia y aprendizaje tradicional (obtenido de [21])

Existen distintas formas de implementar el aprendizaje de transferencia, siendo las principales el *fine-tuning* y la extracción de características. La extracción de características consiste en utilizar todas las capas de un modelo preentrenado en una base de datos de gran tamaño, excepto la última *fully connected*, como extractor de características para una nueva tarea. Las capas que se utilizan del modelo preentrenado se congelan, lo cual quiere decir que no se actualizan sus pesos durante el entrenamiento. El *fine-tuning* también utiliza un modelo preentrenado, pero gracias a la estructura de las CNNs en bloques convolucionales, se pueden elegir reentrenar los bloques deseados. En este caso no se congela toda la red, sino que se puede congelar hasta un cierto número de bloques convolucionales y reentrenar el resto con los nuevos datos. Así, los primeros bloques obtienen características más generales, ya que se basan en conocimientos de otras tareas, y los siguientes buscan características más específicas a la tarea que se está realizando. Además de estos dos enfoques, también se puede implementar *transfer learning* reentrenando todas las capas del modelo. Aunque pueda parecer que los resultados serán iguales que si se hubiese entrenado la red desde cero, esto no es así, ya que al estar los pesos inicializados a valores que han sido útiles para fines similares, la red los tendrá que modificar menos y por eso el entrenamiento será más rápido.

ImageNet es una base de datos que consiste en millones de imágenes etiquetadas con una amplia variedad de objetos. Inicialmente, se empleó para un challenge llamado ImageNet Large Scale Visual Recognition Challenge (ILSVRC) y actualmente es ampliamente utilizada en el campo del aprendizaje automático. En este proyecto se han llevado a cabo entrenamientos utilizando la técnica de *transfer learning* con modelos preentrenados en el conjunto de imágenes de ImageNet, ya que en investigaciones previas se ha visto que ofrece muy buenos resultados.

1.3. Estado del arte

Aunque la detección de Sarcoma de Ewing a través de imágenes histológicas no ha sido ampliamente investigada, se han llevado a cabo investigaciones similares para otras patologías. Hay tres aproximaciones que son las que se considera que son de mayor interés para el estudio del estado del arte de este proyecto. Estas son el uso del aprendizaje profundo para clasificar imágenes histológicas, el uso de la técnica de aprendizaje débilmente supervisado, en concreto aprendizaje de múltiples instancias, para clasificación de imágenes histológicas y el uso de aprendizaje profundo

para detección de sarcoma de Ewing con radiografías.

La primera de ellas, que es el uso del aprendizaje profundo para clasificar imágenes histológicas, se implementa mediante el uso de redes neuronales convolucionales para detectar y clasificar los distintos tejidos de las imágenes. Liao y otros [22] llevaron a cabo el estudio de clasificación basada en aprendizaje profundo y predicción de mutaciones a partir de imágenes histopatológicas de carcinoma hepatocelular. Para el método propuesto se utilizaron WSIs de una base de datos y se excluyeron las que no tenían una buena resolución o etiqueta. La clasificación de las imágenes se realizó mediante una CNN con la estructura de aprendizaje residual profundo para superar el problema de degradación. Está CNN está compuesta por tres bloques que constan de dos filtros convolucionales *kernel* de 3×3, cada uno de los cuales va seguido de una capa BatchNormalization y luego una capa ReLU. La salida de cada bloque consiste en residuos del bloque previo y su propia salida. Este método ofreció resultados muy prometedores para la clasificación de tejidos con imágenes histológicas.

La siguiente aproximación que se va a exponer es la del uso de la técnica de aprendizaje débilmente supervisado, en concreto aprendizaje de múltiples instancias, para clasificación de imágenes histológicas. Spanhol y otros [23] llevaron a cabo el estudio de aprendizaje de múltiples instancias para la clasificación de imágenes histopatológicas de cáncer de mama. Estudiaron dos enfoques distintos, el primero asumieron que cada bolsa era una imagen y dentro de esa bolsa estaban los parches de esa imagen y el segundo asumieron que cada bolsa era un paciente y dentro de esa bolsa estaban las muestras completas que habían sido tomadas de ese paciente. Es decir, en el primer enfoque el contenido de cada bolsa son parches de una misma imagen y en el segundo cada bolsa contiene más de una imagen, todas asociadas a muestras tomadas del mismo paciente. En la Figura 1.10 se puede apreciar la diferencia entre la clasificación de instancias individuales y la clasificación de múltiples instancias, que es el tipo de clasificación que se utilizó en este método.

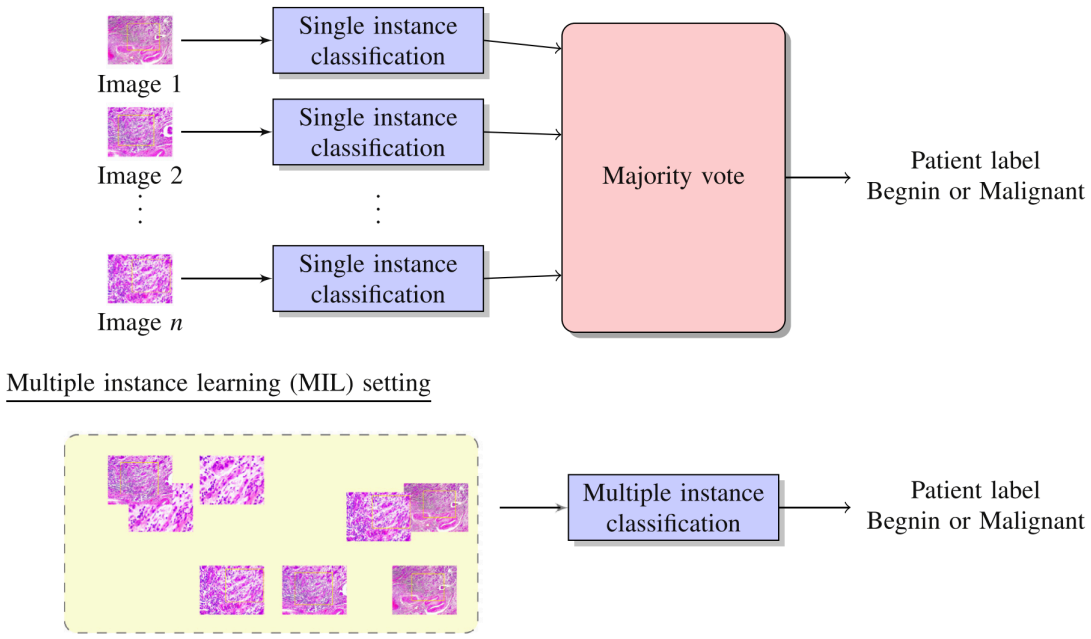

Figura 1.10: Diferencia entre clasificación de instancias individuales y clasificación de múltiples instancias (obtenido de [23])

La última aproximación que se ha considerado de interés ha sido el uso de aprendizaje profundo para detección de sarcoma de Ewing con radiografías. Esto es muy parecido a lo que se pretende implementar a este proyecto, ya que se busca conseguir lo mismo pero mediante imágenes histológicas. Consalvo y otros [24] llevaron a cabo un algoritmo de aprendizaje profundo de dos fases para detección y diferenciación del sarcoma de Ewing y osteomielitis aguda en radiografías pediátricas. En la primera de las dos fases realizaron la detección de casos patológicos y la segunda la diferenciación entre cuáles de estos so sarcoma de Ewing y cuáles son osteomielitis. Para ambos utilizaron una ResNet-50 como CNN, estando preentrenada en imágenes radiográficas relacionadas con sarcomas, para así poder realizar aprendizaje de transferencia. Además, utilizaron *data augmentation* para compensar el desbalanceo de clases.

1.4. Objetivo del proyecto

Como se ha mencionado en la introducción, uno de los factores determinantes a la hora de asegurar la supervivencia del paciente con Sarcoma de Ewing es el diagnóstico temprano. Debido a que el aumento de casos de cáncer ha incrementado la demanda de servicios de patología, actualmente hay una sobrecarga de trabajo de los patólogos encargados del análisis y diagnóstico de las muestras. Esto se traduce en un tiempo menor para analizar las pruebas de cada paciente y mayor agotamiento de los patólogos expertos. Además, la heterogeneidad celular y presencia de patrones vasculares similares dificulta la diferenciación entre tejido que contiene sarcoma de Ewing y el que contiene otros tumores de células redondas pequeñas. Teniendo en cuenta que diagnosticar correctamente entre cada uno de estos es importante, ya que el tratamiento es distinto, y sabiendo que esto causa discrepancias entre observadores, se puede apreciar la clara necesidad de un sistema de diagnóstico más fiable.

Es por esto por lo que el objetivo principal de este proyecto será desarrollar un sistema automatizado capaz de diferenciar tumores de Ewing frente al rabdomiosarcoma, que es un tipo de tumor de células redondas y pequeñas, mediante aprendizaje profundo. Debido a la naturaleza de nuestros datos, que consisten en imágenes histológicas de secciones cilíndricas de tejido, y el problema específico al que nos enfrentamos, vamos a llevar a cabo pruebas tanto de aprendizaje supervisado tradicional como de MIL para entrenar nuestro clasificador y comparar sus rendimientos. La posibilidad de realizar ambos tipos de entrenamiento se debe a que las imágenes tienen un tamaño grande pero aceptable para ser utilizadas completas en el aprendizaje supervisado, y también son lo suficientemente grandes como para ser recortadas y utilizadas en MIL. Sin embargo, en el caso de tener imágenes histológicas de tejidos completos, no sería factible utilizarlas sin recortar, ya que podrían tener dimensiones de hasta 100000x100000 píxeles, lo cual no es viable para el procesamiento en las GPUs. En este escenario, la única opción sería utilizar el enfoque de MIL.

Adémas del objetivo principal del proyecto, también hay unos objetivos secundarios que se busca cumplir:

1. Recolectar y preparar la base de datos: extraer cada muestra de biopsia cilíndrica, a la que llamaremos core, que contienen las micro matrices de tejido digitalizado.

2. Revisar el estado del arte: empezar revisando las técnicas de uso de deep learning para imágenes histológicas. Después, continuar revisando el uso del deep learning para cáncer de hueso y en concreto para diferenciar sarcoma de Ewing de otro tipo de cáncer.

3. Procesar las imágenes: filtrar los cores que han sido recolectados para la base de datos de manera que no quede ninguno incompleto o que no tenga tejido tumoral. Se realizarán tanto filtrados automáticos como manuales para asegurar que las imágenes que se utilizan para el entrenamiento son de alta calidad.

4. Explorar CNNs: dado el acceso a una extensa colección de datos etiquetados y el objetivo de realizar predicciones, se llevarán a cabo experimentos de aprendizaje supervisado con distintas arquitecturas de CNN con el objetivo de determinar cuáles ofrecen los mejores resultados en el contexto específico. La elección del tipo de aprendizaje a implementar se basa en que el tamaño de nuestras imágenes lo permite y en la evidencia que demuestra que, en general, el aprendizaje supervisado brinda el mejor rendimiento en la tarea de predicción, especialmente cuando se cuenta con datos adecuados, lo cual no siempre es fácil de obtener. Las arquitecturas que se explorarán son las que en el estudio del estado del arte se ha visto que han sido utilizadas para propósitos parecidos y se presentarán resultados de todas para poder comparar sus rendimientos.

5. Explorar agregaciones MIL: como se ha mencionado, además de utilizar el aprendizaje supervisado, se realizarán experimentos en los que el aprendizaje será débilmente supervisado, en concreto MIL. En él, las imágenes de cada uno de los datasets son parcheadas a un tamaño determinado y el entrenamiento se realiza con los datos dispuestos en bolsas que contienen todos los parches de una imagen. Teniendo en cuenta que un modelo de MIL consta de tres partes: extracción de características, agregación y clasificación, para la primera parte de extracción de características se utilizará la arquitectura que ha proporcionado los mejores resultados en el entrenamiento supervisado. Además, se explorarán distintas agregaciones de características previas a la clasificación, estas serán agregación de máximos, medias y *MILAttention*.

6. Análisis de resultados: una vez llevados a cabo todos los experimentos se analiza que resultados son los mejores con las métricas que se obtienen de los entrenamientos. En esta sección compararemos el rendimiento de las distintas arquitecturas y de los distintos tipos de aprendizaje para así poder ver si el aprendizaje de múltiples instancias ofrece mejores resultados que el supervisado.

Capítulo 2

Materiales y métodos

2.1. Materiales

2.1.1. Adquisición

Las imágenes que hemos utilizado para el desarrollo del modelo de aprendizaje profundo han sido obtenidas gracias a una colaboración con el Instituto Valenciano de Oncología (IVO), en especial con el patólogo Isidro Machado Puerto. Provienen de los proyectos asociados a sus investigaciones [25] y su tesis [26] y consisten en TMAs. Cada uno de ellos contiene múltiples cores individuales, todos pertenecientes a pacientes contenidos en la misma clase, SE o TCRP. Un paciente puede tener entre uno y cinco cores que le pertenecen y cada uno de ellos puede contener porciones variables de tumor. Detectar visualmente la presencia de SE o TCRP es fácil, ya que ambos tiñen de un color más intenso que el resto del tejido. Lo que supone un reto y una oportunidad para ser resuelto con IA es la diferenciación entre estos dos tejidos, que es lo que ayudaría a mejorar el pronóstico de esta patología. Las muestras que tenemos de tumor de células redondas y pequeñas son de rabdomiosarcoma, que es un tipo de sarcoma que suele comenzar en los músculos que se unen a huesos y que ayudan a mover el cuerpo. El proceso seguido para llevar a cabo la adquisición de la base de datos se resume en la Figura 2.1 para poder seguir de manera más cómoda su explicación.

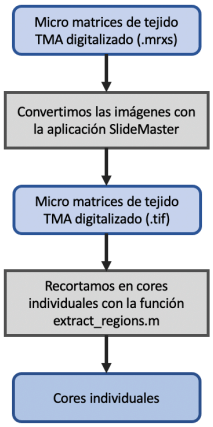

Figura 2.1: Proceso de adquisición

El primer paso a seguir es la conversión de formato .mrxs a .tif. El formato de archivo de imágenes con etiquetas (TIFF, del inglés Tag Image File Format), es un archivo informático que se emplea para almacenar información de imágenes y gráficos rasterizados. Se utiliza este formato para evitar las pérdidas de información que suponen la compresión a formato JPG. La visualización y conversión de las imágenes se realiza con SlideViewer y SlideMaster respectivamente.

El siguiente paso es recortar las imágenes que contienen múltiples muestras de tejido en imágenes con muestras individuales, llamados cores. Esto es realizado con una función que detecta cada uno de los cores con el método de las *bounding boxes*, que son los rectángulos mínimos que contienen a los objetos que hay presentes en una imagen. Para poder llevar a cabo este método es necesario tener la imagen en formato binario y así poder sacar una máscara. Para la binarización, después de probar que canal ofrece un mejor funcionamiento para la detección de cores, se ha pasado de formato RGB a CMYK y elegido un solo canal, que ha sido el de Magenta. Con el resultado de la conversión de modelo de color de los píxeles se puede utilizar el método de umbralización, llamado el método de Otsu [27], para binarizar la imagen. Este método será explicado más adelante en la parte de preprocesado de las imágenes.

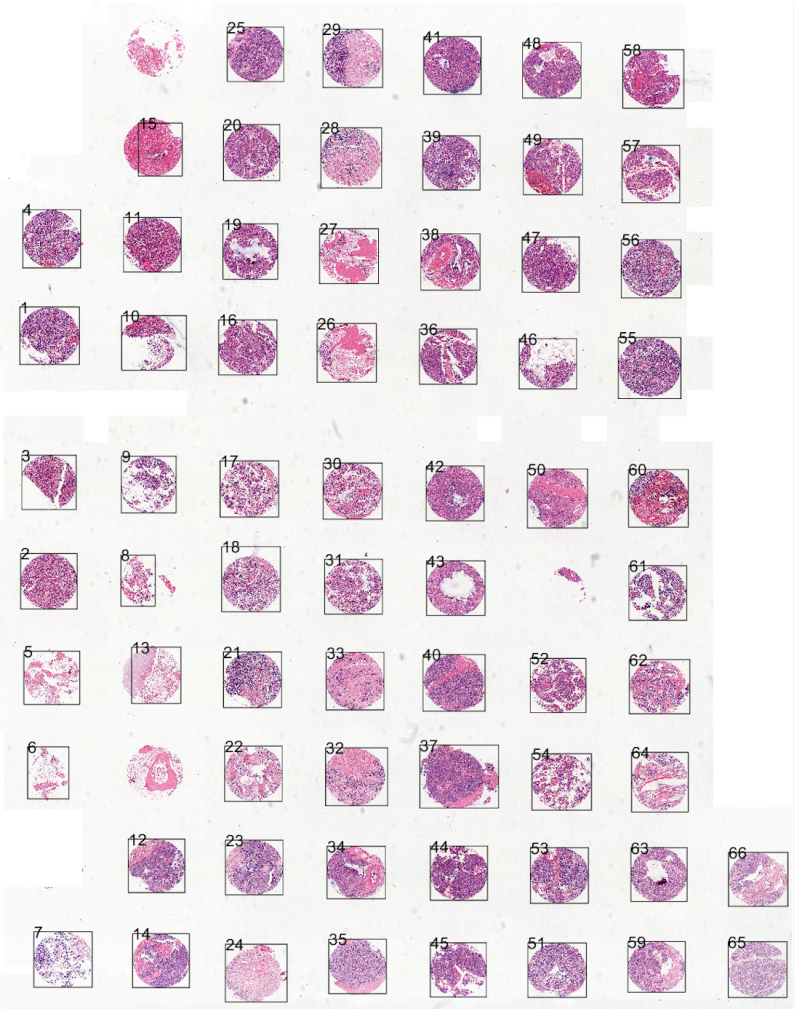

Figura 2.2: Plantillas de detección de *bounding boxes*

El método de las *bounding boxes* que se ha utilizado para la detección de los cores aporta información acerca de su tamaño y localización y así se puede recortar la imagen por las zonas deseadas. Las imágenes resultantes de este proceso tienen un tamaño de alrededor de 3000x3000 píxeles. Este tamaño de imagen es mucho menor que el que tendría una WSI, que es una imagen histológica de una muestra de tejido completa y puede tener un tamaño de 100000x100000 píxeles. Para este tipo de imágenes solo es posible implementar MIL, ya que su gran tamaño no permite utilizar las muestras completas para el entrenamiento. Debido a que los cores tienen un tamaño menor, ya que únicamente contienen una muestra cilíndrica de tejido, es posible implementar tanto MIL como aprendizaje supervisado tradicional, pues las GPUs soportan estos tamaños de imagen para el entrenamiento de modelos de aprendizaje profundo. Además de los recortes de los cores, la función mencionada también nos devuelve una imagen, como la que se puede ver en la Figura 2.2, en la que nos muestra cada uno de los *bounding boxes* que ha detectado a cada uno de los cores. Esto resulta útil para poder identificar a qué paciente pertenece cada uno de los recortes. Se muestra la figura rotada para hacer coincidir su formato con el de la plantilla de la Figura 2.3.

Además de los TMAs mencionados al principio de esta sección, desde el laboratorio de patología se prepararon plantillas con la misma estructura en las que se indica a qué paciente pertenece cada uno de los cores y qué tipo de patología padecen los pacientes de esa diapositiva. Se puede ver un ejemplo de estas platillas en la Figura 2.3. La estructura que siguen estas plantillas es que un cuadrado, en los que se indica el ID del paciente, corresponde a cada uno de los cores. Como ejemplo, PT418 es el ID de un paciente que tiene dos cores asociados, ya que hay dos cuadrados con ese mismo ID. Los cuadrados en los que se indica CONTROL son los asociados a los cores de control que no están asociados a ningún paciente y, por lo tanto, no se utilizan para el entrenamiento. Con esto y con la ayuda de las plantillas de *bounding boxes* generadas (Figura 2.2) se puede llevar a cabo el paso final de la adquisición de la base de datos, que es etiquetar las imágenes correctamente. Esto se realiza rellenando las platillas con el número de detección que corresponde a cada paciente. Así, se puede etiquetar cada imagen con el identificador de paciente y el número de muestra de ese paciente para que tengan la estructura '(nombre de paciente)_(número de muestra de ese paciente)'. Por ejemplo, PT418_1 sería la primera muestra de ese paciente y la siguiente PT418_2. Tener las etiquetas de esta manera permite asegurar que a la hora de hacer las particiones de datos para entrenamiento, validación y test exista la posibilidad de evitar que haya muestras del mismo paciente en distintos conjuntos. Es deseable evitar esta situación, ya que esto podría llevar a una evaluación sesgada y poco realista del rendimiento del modelo.

PT418	**PT418**	**PT419**	**PT419**	**PT420**	**PT420**	**PT421**	**PT421**		
PT413	**PT413**	**PT414**	**PT414**	**PT415**	**PT415**	**PT416**	**PT416**	**PT417**	**PT417**
PT408	**PT408**	**PT409**	**PT409**	**PT410**	**PT410**	**PT411**	**PT411**	**PT412**	**PT412**
PT403	**PT403**	**PT404**	**PT404**	**PT405**	**PT405**	**PT406**	**PT406**	**PT407**	**PT407**
PT398	**PT398**	**PT399**	**PT399**	**PT400**	**PT400**	**PT401**	**PT401**	**PT402**	**PT402**
PT392	**PT392**	**PT393**	**PT393**	**PT395**	**PT395**	**PT396**	**PT396**	**PT397**	**PT397**
PT386	**PT386**	**PT387**	**PT387**	**PT388**	**PT388**	**PT389**	**PT389**	**PT391**	**PT391**
CONTROL	**CONTROL**								

Figura 2.3: Plantillas de correspondencia entre paciente y muestra

Una vez finalizada la adquisición de datos, se puede realizar un recuento para determinar la magnitud de la base de datos disponible para el entrenamiento. Los resultados de este recuento se presentan en la Figura 2.1.

Cantidad	Ewing	Rabdomiosarcoma
TMAs	26	6
Cores	203	257
Parches por WSI (media $\pm$ desv. típica)	92 ± 15	101 ± 24

Tabla 2.1: Recuento de muestras de la base de datos después del filtrado

2.1.2. Software

El software utilizado se refiere al conjunto de programas, instrucciones y datos que se utilizan para controlar y operar un sistema informático. En el caso de este proyecto, ha sido el entorno de programación MATLAB® en su versión R2019a y el entorno de desarrollo Visual Studio Code, utilizado para programar en Python.

MATLAB®, el lenguaje de cálculo técnico desarrollado por MathWorks, es un entorno de programación para el desarrollo de algoritmos, análisis de datos, visualización y cálculo numérico [28]. Es conocido por su capacidad de trabajar con matrices y su amplia gama de herramientas y bibliotecas.

Visual Studio Code (VS Code) es un editor de código fuente desarrollado por Microsoft disponible para Windows, GNU/Linux y macOS. VS Code tiene una buena integración con Git, cuenta con soporte para depuración de código, y dispone de un sinnúmero de extensiones, que básicamente te da la posibilidad de escribir y ejecutar código en cualquier lenguaje de programación [29].

Para la programación en Python se han utilizado principalmente TensorFlow, que es una biblioteca de aprendizaje automático amplia y flexible, y Keras, que es una biblioteca de alto nivel que facilita el diseño y la construcción de modelos de redes neuronales utilizando TensorFlow como backend.

2.1.3. Hardware

Para la realización del proyecto se han utilizado un ordenador portátil y una tarjeta gráfica o GPU. El ordenador portátil empleado ha sido un MacBook Pro (13-inch, 2017, Two Thunderbolt 3 ports) que dispone de un procesador 2,3 GHz Intel Core i5 con tarjeta gráfica Intel Iris Plus Graphics 640 1536 MB y 8 GB de memoria RAM. Debido a que llevar a cabo el entrenamiento de las redes neuronales convolucionales de este proyecto requiere una alta capacidad computacional, se ha necesitado hacer uso de procesadores más potentes. Es por eso que se han empleado los servidores pertenecientes al equipo de investigación CVBLab, que constan de un procesador Intel i7 @4.20 GHz, tarjeta gráfica NVIDIA Titan V y una capacidad de memoria de 32 GB de RAM.

2.2. Métodos

2.2.1. Preprocesado

Como se ha mencionado anteriormente, para la binarización de las imágenes se ha realizado una umbralización con el método de Otsu. Este se encuentra dentro de las técnicas de umbralización más utilizadas en la literatura. Selecciona el máximo valor umbral de la varianza entre clases del histograma de una imagen [27]. Con el valor calculado por el método de Otsu, los píxeles que tengan un valor por encima de ese umbral serán convertidos al valor máximo y los que tengan un valor inferior al mínimo. Esta binarización permite detectar y recortar los cores con el método de las *bounding boxes*. En la Figura 2.4 se puede ver una sección de la imagen resultante de aplicar el método de Otsu a un TMA de la base de datos del proyecto para detectar las posiciones de los cores.

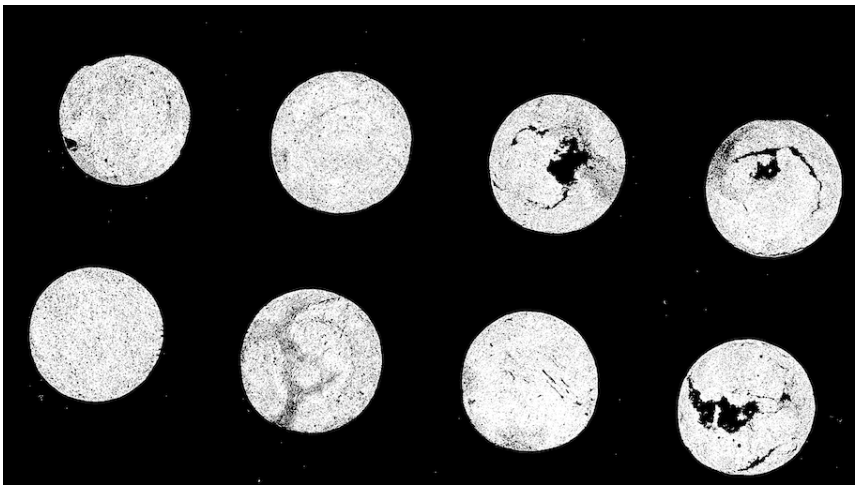

Figura 2.4: Resultado de aplicar el método de Otsu a un TMA

De los cores adquiridos mediante los procesos explicados anteriormente, no todos resultan útiles para llevar a cabo el entrenamiento del modelo de aprendizaje profundo. Esto se debe a que, como inicialmente se sacaron para análisis bajo el microscopio, y en este caso no es grave que hayan cores incompletos, dentro de los TMAs hay muestras de cores que no están totalmente completos, tienen formas muy variadas o incluyen artefactos. En el caso de ser analizado bajo el microscopio no habría problema en contar con formas distintas o cantidades de tejido muy variantes, ya que a simple vista es fácil detectar cuando una muestra no es útil y fijarse en la siguiente o ver que una muestra está incompleta y solo fijarse en la parte de tejido que se puede apreciar. A diferencia de cuando el análisis es bajo microscopio, en el caso de ser de manera automatizada se necesita tener una base de datos en la que todas las instancias tengan una estructura muy parecida. Esto es porque al aprender características de las imágenes para poder clasificarlas, las diferencias en cantidad de tejido o forma de la imagen podrían inducir a confusión si no se indica que esas imágenes no son útiles para el entrenamiento.

En el código mencionado anteriormente, en el que se extraen los distintos cores, hay ciertos fallos que se cometen en la detección. Uno de los errores que se presentan son que cuando un core tiene un tamaño más grande o no totalmente cuadrado, el código puede llegar a detectarlo como dos

partes y eso resulta en cores de por ejemplo 3000x1500 píxeles. El otro error es que cuando un core está incompleto se detecta como si fuese un core válido y esto genera una imagen en la que no hay casi tejido. En la Figura 2.5 se pueden ver ejemplos de ambas situaciones.

Figura 2.5: Ejemplos de los errores de detección

Sabiendo que es necesario tener una base de datos homogénea y viendo qué es lo que causa que haya diferencia entre imágenes, se ha creado un código que filtre las imágenes de manera que todas tengan una forma cuadrada y una cantidad mínima de tejido. La tarea de comprobar que las imágenes tengan una forma cuadrada se ha realizado analizando cada una de ellas y comprobando que su tamaño de anchura y altura sea muy parecida, descartando las que su cuentan con más de 10 % de diferencia entre estas dos medidas. Para comprobar que la cantidad de tejido súpera un mínimo se ha convertido la imagen a blanco y negro, convirtiendo los píxeles a binarios, y se ha comprobado que la cantidad de píxeles negros supera el 20 % del total de la imagen. Así, se ha asegurado que hay al menos una determinada cantidad de píxeles con una cantidad suficiente de tinción.

Además de utilizar el código para el filtrado automatizado de las imágenes, se ha considerado necesario hacer también un revisado manual con la ayuda del patólogo que nos proporcionó los TMAs. Es por eso que se tuvo una reunión con él en la que además de aprender a analizar una biopsia digitalizada y entender qué tejidos son de interés por ser tumorales y que tejidos no, se realizó una revisión de las imágenes una a una para que con su ayuda experta se pudiese decidir que imágenes eliminar por no contener suficiente tejido tumoral o contener muchos tejidos que no son de interés como es el colágeno o fibras musculares.

Para entender y poder apreciar la diferencia entre los tejidos mencionados se van a mostrar unos ejemplos en los que se ven con claridad cada uno de ellos. En la Figura 2.6 se puede ver como hay dos tejidos que se pueden diferenciar claramente de colores morado y rosa, siendo la parte rosa el colágeno y el resto células tumorales.

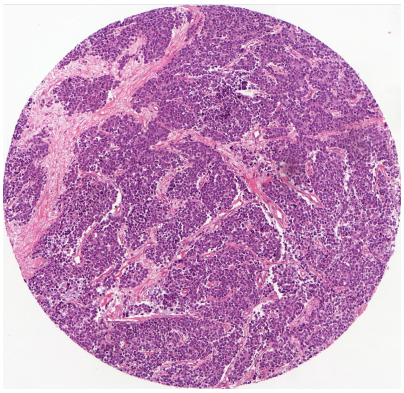

Figura 2.6: Core que contiene tejidos de colágeno

En la Figura 2.7 se muestra un ejemplo de cores de SE (izquierda) y de rabdomiosarcoma (derecha) en la que se pueden apreciar diferencias en sus células que permiten diferenciar entre estas dos patologías. En el core de SE se puede ver que la forma de las células moradas es muy redondeada y su núcleo no es muy oscuro. Por otro lado, en el core de rabdomiosarcoma se puede ver que los núcleos de las células son mucho más oscuros y no tan redondeados, tienen una forma más *spindle-like*. Las células con esta forma se caracterizan por ser alargadas y fusiformes y son características de los rabdomiosarcomas.

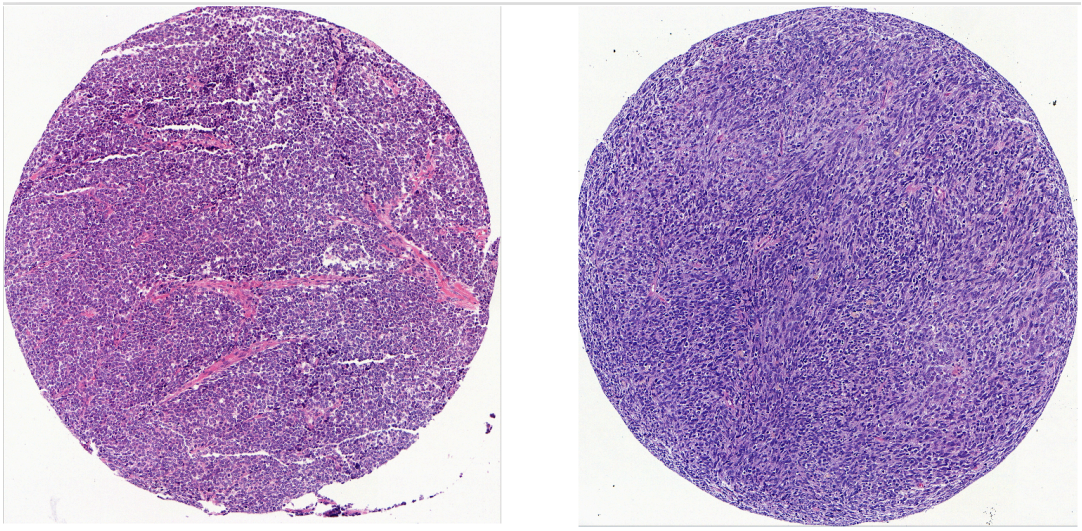

Figura 2.7: Core de SE (izquierda) y de rabdomiosarcoma (derecha)

Es importante recalcar que, aunque en el ejemplo mostrado se evidencia claramente la diferencia, esto no siempre es el caso. En muchas ocasiones, las células de rabdomiosacroma no tienen una forma tan pronunciada de *spindle-like* y se asemejan más a las de SE, siendo de forma redonda. Debido a esta similitud, se considera que la detección de estas células requiere un análisis profundo y detallado de las características celulares. Es por eso que se plantea que la inteligencia artificial podría ser una herramienta prometedora para facilitar su detección.

2.2.2. Aprendizaje supervisado

El aprendizaje supervisado consiste en el entrenamiento de algoritmos que clasifican datos o prevén resultados con precisión en un conjunto de datos etiquetados. El entrenamiento y optimización de parámetros se realiza utilizando las etiquetas. En este proyecto se han explorado distintas arquitecturas de red neuronal convolucional para poder comparar sus rendimientos y analizar cuál ofrece mejores resultados. Entre las arquitecturas que se han estudiado están la VGG16, ResNet-50, AlexNet e Inceptionv3. A continuación se hará una explicación de cada una de estas y como se han utilizado.

VGG16

La red neuronal VGG16 es un modelo de red convolucional profunda que consta de 16 capas, incluyendo capas convolucionales, capas de *pooling* y capas *fully connected* [30]. Se caracteriza por tener un tamaño de *kernel* de 3x3 en las capas convolucionales y 2x2 en las capas de *pooling*,

en concreto *maxpooling*, lo que supone una mejora respecto a modelos anteriores al utilizar núcleos de convolución más pequeños. Como se puede ver en la Figura 2.8, el tamaño de entrada de las imágenes es de 224x224x3 y las capas de la red están agrupadas en 5 bloques, a los que se les conectan capas densas al final para realizar la clasificación entre 1000 clases. A pesar de que el tamaño de entrada para el que está inicialmente diseñada es para el mencionado anteriormente, la red está preparada para soportar distintos tamaños, como en nuestro caso que han sido imágenes de 750x750x3.

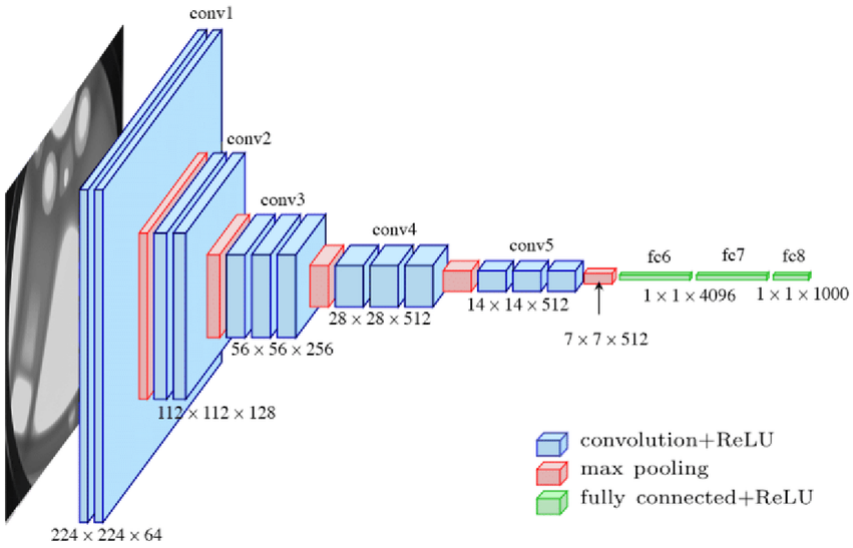

Figura 2.8: Esquema de la arquitectura de la VGG16 (obtenido de [31])

En este proyecto se utilizará esta arquitectura de distintas maneras. Se harán pruebas utilizándola tanto como extractor de características como haciendo *fine-tuning* y reentrenándola todo. Básicamente, se irá probando a reentrenar todas las capas, reentrenar solo los últimos dos bloques convolucionales y finalmente congelar todas las capas, utilizando los pesos que la red ha aprendido para la base de datos de ImageNet.

ResNet-50

La ResNet-50 es un modelo de red neuronal que consta de 50 capas de profundidad y destaca por su arquitectura basada en bloques residuales. Los bloques residuales permiten que las capas de la red aprendan las diferencias entre las representaciones originales y las representaciones deseadas, facilitando así el entrenamiento de redes más profundas [32]. Para ello, como se puede ver en la Figura 2.9, ResNet-50 conecta las salidas de diferentes bloques de convolución, permitiendo así que la información de gradiente pase a través de las mismas y que se realice el entrenamiento de las últimas capas de los modelos. Utiliza convoluciones de tamaño 3x3 y 1x1, para extraer características, junto con capas de *maxpooling*, *average pooling* y *fully connected*, para realizar la clasificación de las imágenes.

Para esta arquitectura de red se seguirá el mismo proceso de entrenamiento que se ha explicado para la VGG16. En primer lugar, se reentrena toda la red, a continuación se reentrenan solo los últimos dos bloques y como prueba final se utilizará toda la red con los pesos calculados entrenándola con la base de datos de ImageNet.

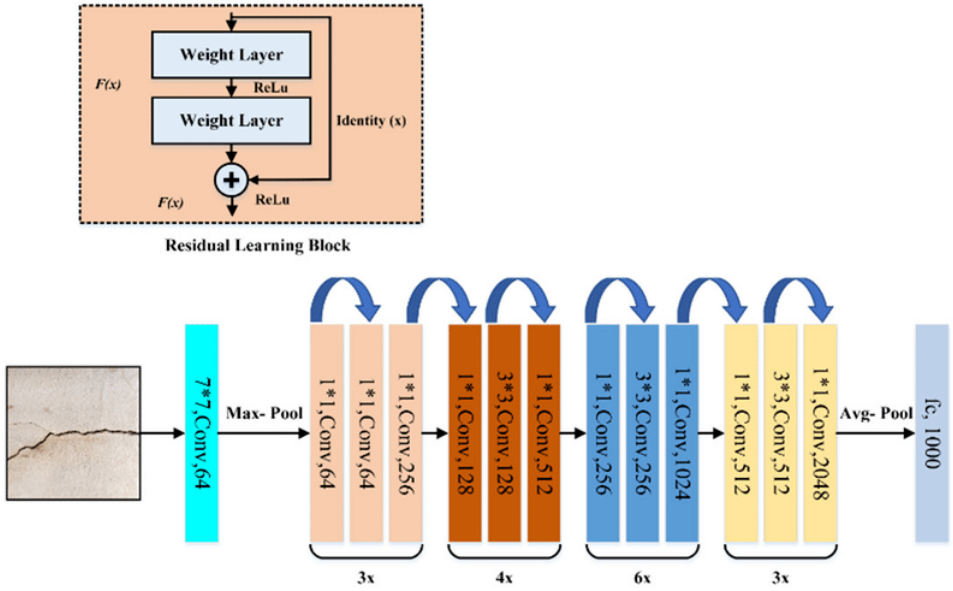

Figura 2.9: Esquema de la arquitectura de la ResNet-50 (obtenido de [32])

AlexNet

Esta CNN, presentada por Alex Krizhevsky, Ilya Sutskever y Geoffrey Hinton en 2012, fue una de las primeras redes que demostró la eficacia de las CNN en la tarea de clasificación de imágenes. Como se puede ver en la Figura 2.10, su arquitectura consta de 8 capas de profundidad, siendo las primeras cinco capas convolucionales y las últimas tres capas *fully connected*. El tamaño de entrada es 227x227 píxeles y puede clasificar entre 1000 clases. Entre los bloques convolucionales se lleva a cabo una reducción de dimensionalidad. En el primero se realiza mediante la operación de la convolución con *stride* de tamaño 4x4 y en el resto con *maxpooling*. El éxito de esta CNN se debe a la implementación de ciertos elementos como son las capas ReLU no-lineales y la técnica de regularización con *dropout*, que acelera el entrenamiento y reduce el sobreentrenamiento [33]. A diferencia de lo realizado con las otras arquitecturas, en este caso solo se ha utilizado AlexNet inicializada con los pesos de ImageNet reentrenando todas sus capas.

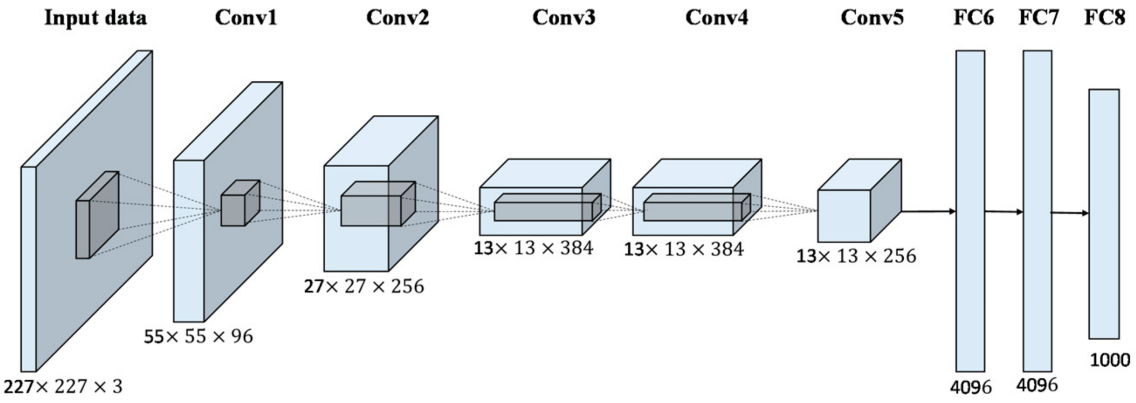

Figura 2.10: Esquema de la arquitectura de la AlexNet (obtenido de [33])

Inceptionv3

InceptionV3 es un modelo de CNN que se utiliza para el reconocimiento de imágenes y la clasificación de objetos. Fue desarrollado por Szegedy y otros como parte del proyecto Inception y es una mejora del modelo original Inception. Utiliza una arquitectura, la cual se puede ver en la Figura 2.11, en la que las capas convolucionales están conectadas en paralelo, permitiendo que la red aprenda diferentes características en diferentes escalas y niveles de abstracción. Esto permite que la red capture detalles finos y características globales de una imagen al mismo tiempo. Al igual que con AlexNet, solo se ha utilizado inicializada con los pesos de ImageNet y reentrenando todas sus capas.

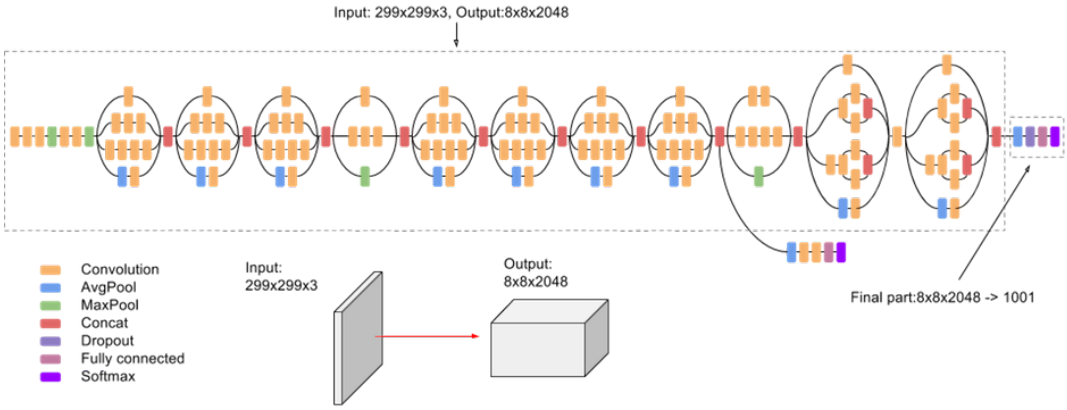

Figura 2.11: Esquema de la arquitectura de la Inceptionv3 (obtenido de [34])

2.2.3. Aprendizaje múltiples instancias

Uno de los avances en el entrenamiento de redes con aprendizaje profundo ha sido el algoritmo de aprendizaje débilmente supervisado. Este permite utilizar grandes cantidades de datos que no están completamente etiquetados o lo están de forma imprecisa. En este tipo de aprendizaje los datos están agrupados en bolsas, en cada una de ellas hay un número de instancias. Si dentro de una bolsa hay una instancia positiva, se considerará toda la bolsa positiva (Figura 2.12), independientemente de la cantidad de instancias negativas que haya. Así, el modelo aprende a reconocer las instancias que son más representativas para poder asignar una etiqueta a cada bolsa.

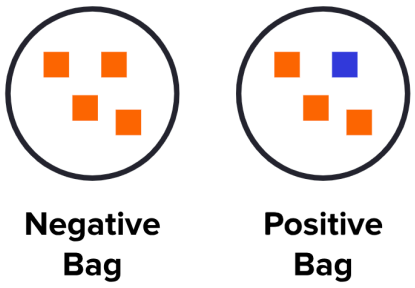

Figura 2.12: Ejemplo de bolsa positiva y negativa (obtenido de [35])

En este proyecto se trabaja con imágenes histológicas que contienen distintas estructuras, como células cancerosas, tejido sano y colágeno. El objetivo principal es detectar un tipo específico de tejido, que son las células de SE. En este contexto, el enfoque de MIL se considera una opción prometedora, ya que nos permite entrenar con la etiqueta global sin necesitar anotar cada tejido. Su funcionamiento se basa en clasificar una bolsa si al menos una instancia de las que contiene es positiva, sin necesidad de entrenar el modelo indicando si cada una de ellas es positiva o negativa. Así, las bolsas serán imágenes histológicas y las instancias dentro de ellas parches de la imagen que contienen distintos tejidos, pudiendo ser SE o no. El clasificador buscará clasificar una bolsa completa como positiva si al menos una de las instancias contiene células de SE. Debido al potencial de este tipo de aprendizaje, se ha considerado de interés comparar el aprendizaje supervisado con el MIL para ver cuál de los dos ofrece mejores resultados.

En la Figura 2.13 podemos ver una forma muy común de implementar MIL, en esta se intenta determinar el máximo margen entre instancias de manera que haya al menos una instancia positiva en cada bolsa positiva. Así, el margen es cambiado para que cumpla esta condición.

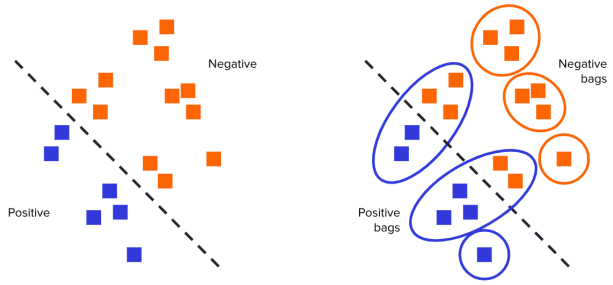

Figura 2.13: Ejemplo de implementación comparado con aprendizaje supervisado tradicional (obtenido de [35])

La implementación del aprendizaje de múltiples instancias se puede dividir en tres partes:

1. Creación de las bolsas: Se crea una bolsa por cada de una de las imágenes que se van a utilizar. La bolsa está compuesta de la imagen de esa bolsa recortada en parches de un tamaño determinado, en nuestro caso de 512x512 píxeles, y un vector etiquetas de la longitud del número de parches que se han creado. Así, la bolsa resulta en un vector que contiene un número de imágenes de un tamaño determinado y otro vector con el mismo número de etiquetas que imágenes hay y todas las etiquetas siendo iguales, ya que cada bolsa es positiva con tal de contener una muestra de la clase positiva y solo es negativa en caso de no contener ninguna imagen positiva.

2. Extracción de características: Se obtiene un vector de características para cada una de las instancias dentro de cada bolsa. La extracción de estos vectores la realiza la CNN.

3. Agregación: La agregación es la representación de las características de una bolsa para tenerlas a nivel de bolsa en lugar de a nivel de instancia y así poder hacer predicciones para la bolsa entera. Para obtener esta representación se pueden realizar distintas operaciones de agregación de características, como son el máximo, el mínimo, la media y *MILAttention*.

    - Máximos: Este tipo de agregación implica seleccionar el valor máximo de cada característica entre todas las instancias en una bolsa. Para cada característica, se toma el valor

más grande encontrado entre todas las instancias y se forma un vector resultante que contiene los valores máximos. Este enfoque resalta las características más destacadas o dominantes presentes en las instancias de la bolsa.

- Mínimos: El enfoque de agregación mediante el mínimo es similar al método del máximo, pero en lugar de seleccionar el valor máximo, se selecciona el valor mínimo de cada característica entre todas las instancias en una bolsa. Es útil cuando se busca capturar las características mínimas o menos representativas en la bolsa.
- Media: Este tipo de agregación se ha utilizado para el entrenamiento y en ella se calcula el promedio de las características de todas las instancias en una bolsa. Se suman todas y se dividen por el número total de instancias. La media proporciona una representación promedio de las características dentro de la bolsa. Se utiliza cuando se desea una medida que refleje las características típicas de las instancias en la bolsa.
- *MILAttention* [36]: Este tipo de agregación utiliza la atención para asignar pesos a cada instancia dentro de un conjunto, enfocándose en las más relevantes. Estos pesos indican la importancia relativa de cada instancia en el conjunto y se utilizan para calcular la agregación de características. Para implementar este tipo de agregación se representa cada instancia con un vector de características, se aplica la función de atención a las instancias para obtener unos pesos, que indican la importancia relativa de cada característica, y se realiza la agregación ponderando los pesos con las características. Esta agregación resulta muy útil en MIL debido hay una relación compleja entre las instancias y las etiquetas. No todas las instancias contribuyen por igual a la clasificación de la bolsa, algunas pueden ser más informativas o más representativas de la clase que otras.

4. Clasificación: Se realiza la clasificación de las bolsas utilizando la representación de las características a nivel de bolsa y su etiqueta. Así, se consigue asignar una etiqueta que clasifique a cada bolsa completa.

Capítulo 3

Resultados y discusión

3.1. Métricas de evaluación

Para el cálculo de las métricas de evaluación del funcionamiento del modelo, en la tarea de clasificación se han utilizado los valores de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN). Para el problema de clasificación binaria que se tiene en este proyecto se ha considerado al SE como la clase positiva (asignado un valor de 1 a las etiquetas) y al rabdomiosarcoma como la negativa (asignado un valor de 0 a las etiquetas). Así, cuando la etiqueta verdadera coincide con la etiqueta predicha se tiene un valor verdadero y si esa es 1 será VP y si es 0 VN. De manera inversa, si la etiqueta verdadera y la predicha no coinciden, se tendrá un valor falso y si la etiqueta obtenida es 1 será FP y si es 0 FN. Estos valores son los que se presentan en una matriz de confusión que se representa de la forma que se puede ver en la Figura 3.1.

Figura 3.1: Matriz de confusión (obtenido de [37])

Con estos valores se pueden calcular distintas métricas que son de utilidad para medir la calidad del funcionamiento de nuestro sistema. Las métricas que se pueden calcular y que se han utilizado en este proyecto son:

- Precisión (prec): Indica la proporción de elementos clasificados correctamente.

$$\frac{VP + VN}{n} \tag{3.1}$$

  Siendo VP y VN la suma de todas las predicciones correctas y n el número total de imágenes que se han usado.

- Sensibilidad (sens): Indica la proporción de valores positivos bien predichos por el modelo respecto al total de valores positivos reales.

$$\frac{VP}{VP + FN} \tag{3.2}$$

- Especificidad (esp): Indica la proporción de valores negativos bien predichos por el modelo respecto al total de valores negativos reales.

$$\frac{VN}{VN + FP} \tag{3.3}$$

- F1-score: Calcula la armónica media ponderada de la precisión y la sensibilidad del modelo.

$$\frac{2 * \text{Precisión} * \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \tag{3.4}$$

- Área bajo la curva ROC (AUC): ROC es la *Receiver Operating Characteristic Curve* y representa la capacidad de diagnóstico de un clasificador binario. Cuanto mayor es su valor, mayor es la capacidad del modelo para diferenciar entre clases. Un valor cercano a 1 significaría que su funcionamiento es muy bueno y un valor cercano a 0.5 que es muy malo.

Además de estas métricas que evalúan como de bien clasifica un modelo, existen técnicas para obtener las regiones discriminatorias utilizadas por una CNN para identificar una clase específica en la imagen. Estas técnicas son los *class activation maps* (CAM) y para imágenes médicas resulta de gran utilidad debido a que se puede ver en que zona de la imagen está fijándose el modelo para tomar decisiones de clasificación. Así, los resultados se pueden contrastar con patólogos expertos que puedan confirmar que son regiones de interés. Para obtener los CAMs se ha utilizado el método Grad-CAM [38]. Este funciona calculando los gradientes de las predicciones con respecto a la salida de la última capa convolucional, la cual contiene las características más relevantes y de alto nivel aprendidas por el modelo. Cada activación de la última capa convolucional se pondera por

su gradiente calculado y de su resultado se hace la media, que se presenta en un mapa de calor que resalta las regiones más relevantes. Su resultado se superpone con la imagen de entrada original. Como se puede ver en el ejemplo de la Figura 3.2, la parte superior de la imagen está siendo en la que más se está fijando el modelo a la hora de clasificar, esto lo identificamos porque esta parte se presenta mucho más roja que el resto.

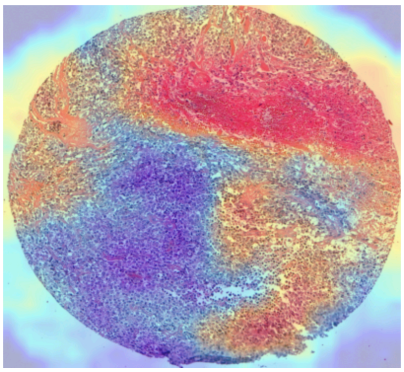

Figura 3.2: Ejemplo de class activation map

A continuación se presentan los resultados tanto para entrenamiento supervisado como para entrenamiento MIL. Los conjuntos de datos de entrenamiento, validación y test utilizados para el entrenamiento en ambos casos ha sido el mismo. Las métricas indicadas en las Tablas 3.1, 3.2 y 3.3, que han sido utilizadas para evaluar el funcionamiento del modelo, han sido calculadas para el conjunto de test. Estando este compuesto por 115 imágenes para el entrenamiento supervisado y 115 bolsas para el MIL.

3.2. Resultados entrenamiento supervisado

En esta sección se presentan las distintas métricas calculadas para evaluar el rendimiento del modelo tras realizar un aprendizaje de transferencia con redes previamente entrenadas. Se han explorado cuatro arquitecturas y para cada una se indica si se ha llevado a cabo un reentrenamiento completo, *fine-tuning* o extracción de características. Para las primeras dos se han realizado pruebas de las tres opciones mencionadas y sus resultados se pueden ver en la Tabla 3.1.

Modelo	Prec	Sens	Esp	F-Score	AUC
ResNet-50: extracción de características	0,345	0,588	0,095	0,434	0,342
ResNet-50: *fine-tuning*	0,782	0,843	0,810	0,811	0,826
ResNet-50: reentrenada	0,944	**1**	0,952	**0,971**	**0,976**
VGG16: extracción de características	0,495	0,902	0,254	0,639	0,578
VGG16: *fine-tuning*	0,645	0,961	0,571	0,772	0,766
VGG16: reentrenada	0,885	0,902	0,905	0,893	0,903

Tabla 3.1: Resultados de clasificación en test para entrenamiento supervisado

Analizando los resultados se puede apreciar que de los tres enfoques abordados el que mejores resultados ofrece es el reentrenemiento completo, seguido del *fine-tuning* y acabando con extracción

de características. Estos resultados parecen lógicos, ya que los mejores resultados ocurren cuanto más se reentrena la red para los datos concretos del modelo.

Aunque es cierto que la extracción de características ha demostrado ofrecer buenos resultados para otros enfoques, se puede ver que para imágenes histológicas resulta en un funcionamiento un poco pobre. Esto puede ser debido a la complejidad de este tipo de imágenes y la necesidad de aprender características mucho más específicas para poder realizar buenas predicciones.

El *fine-tuning* ha resultado en una precisión de 0.782 para la arquitectura de la ResNet-50 y 0.645 para la de la VGG16. Estos resultados, aunque son mejores que los que nos da la extracción de características, siguen siendo demasiado bajos como para considerarse óptimos para algo tan crítico como es la detección de un tumor altamente agresivo. A pesar de haber obtenido resultados relativamente bajos de precisión, ambas han logrado una sensibilidad mayor de 0.84, lo cual es un muy buen resultado, ya que basándonos en como se realiza su cálculo (3.2) vemos que esto significa que hay un nivel bajo de FN predichos. Por otra parte, la especificidad ofrecida por la ResNet-50 es 0.810, lo cual no es un mal resultado, pero la de VGG16 es 0.571. Esto es deficiente, ya que como se puede ver en la ecuación 3.3, quiere decir que hay una cantidad alta de FP. Además, en los modelos de clasificación se busca tener un equilibrio entre la sensibilidad y la especificidad y en este caso estaría muy desbalanceado.

A la vista de que para las primeras dos arquitecturas las métricas con un reentrenamiento completo resultan significantemente mejores, se ha optado por implementar únicamente este enfoque para las siguientes dos. Sus resultados, junto con los de las dos arquitecturas probadas anteriormente, se presentan en la Tabla 3.2. Se puede ver que de las cuatro arquitecturas con las que se ha realizado este tipo de aprendizaje, las dos que mejor funcionan son AlexNet y ResNet-50, teniendo cada una mejores resultados para una métrica en concreto.

Modelo reentrenado	Prec	Sens	Esp	F-Score	AUC
ResNet-50	0,944	**1**	0,952	**0,971**	**0,976**
VGG16	0,885	0,902	0,905	0,893	0,903
AlexNet	**0,979**	0,922	**0,984**	0,949	0,953
InceptionV3	0,943	0,980	0,952	0,962	0,966

Tabla 3.2: Resultados de clasificación en test para entrenamiento supervisado

Debido a que la finalidad de este proyecto es la detección de tejidos de SE para el diagnóstico correcto de una patología, entre que el sistema detecte más o menos casos positivos de los que debería, es aconsejable que detecte de más, ya que si detectase de menos podría estar pasando por alto casos de SE que deben ser tratados. Es por eso que lo más deseable para este sistema será tener una sensibilidad alta, puesto que si nos fijamos en como se realiza su cálculo (3.2) esto querrá decir que la cantidad de FN es baja. Este criterio ayuda a elegir cuál de las dos arquitecturas es óptima para el entrenamiento del modelo del proyecto y la que se va a utilizar para el MIL. Mirando la Tabla 3.2 se puede ver remarcado que para la arquitectura de la ResNet-50, además de tener los mejores resultados de F-score y AUC, se obtienen una sensibilidad de 1. Esto hace que se considere la arquitectura óptima para este proyecto. Además, que esta arquitectura tenga un buen funcionamiento es algo esperado, ya que en el estudio del estado del arte se ha visto que las arquitecturas basadas en aprendizaje residual han proporcionado buenos rendimientos para clasificación de imágenes histológicas.

Aunque este resultado sea satisfactorio, no hay que pasar por alto el análisis de la calidad de la clasificación del modelo basado en los CAMs. Esto nos da la información de en que partes del core se está fijando la atención y así nos permite determinar si se está realizando la clasificación de forma correcta. En la Figura 3.3 se muestra un ejemplo de CAMs de imágenes de SE (izquierda) y rabdomiosarcoma (derecha) que han sido clasificadas correctamente. Se puede ver como el modelo fija la atención en distintas partes del tejido tumoral del core. Posiblemente, estas zonas contendrán células que tendrán una morfología característica del SE y rabdomiosarcoma que las diferenciaran entre ellas y, por lo tanto, podrán ser usadas para la diferenciación. Con estos resultados, podemos confirmar que para realizar la clasificación el modelo se está fijando en regiones del core que son de interés, ya que gracias a la confirmación del patólogo podemos saber que son células cancerosas. Si las zonas de calor (zonas rojas) estuviesen en el fondo o en tejidos como colágeno, sí que habría que replantearse como se está haciendo la clasificación y posiblemente modificar ciertos parámetros.

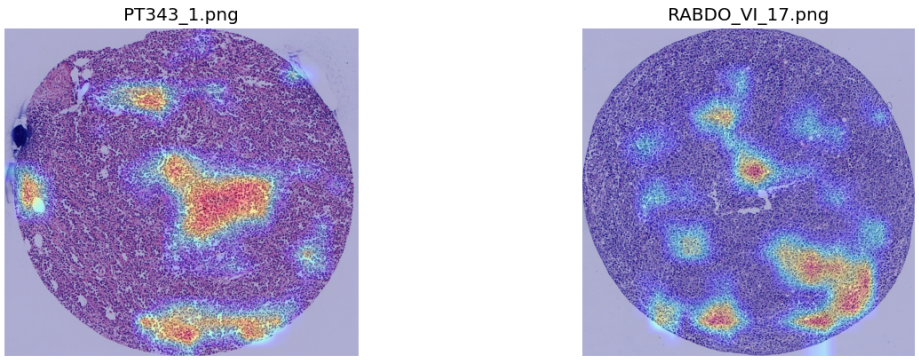

Figura 3.3: CAMs de imágenes de SE (izquierda) y rabdomiosarcoma (derecha) clasificadas correctamente

3.3. Resultados entrenamiento MIL

En la Tabla 3.3 se presentan las distintas métricas para evaluar el rendimiento del modelo tras realizar MIL con distintas agregaciones de características. Las agregaciones que se han explorado son la de máximos, la de mínimos y *MILAttention*. Se han remarcado los valores más altos de cada una de las métricas y se puede ver que todos han sido obtenidos cuando se ha utilizado la agregación de la media.

Agregación	Prec	Sens	Esp	F-Score	AUC
Máximos	0,746	0,940	0,750	0,832	0,845
Media	**1**	**0,955**	**1**	**0,977**	**0,978**
MILAttention	0,905	0,950	0,889	0,927	0,919

Tabla 3.3: Resultados de clasificación en test para MIL con ResNet-50

Aunque se podría haber esperado que *MILAttention* lograra un rendimiento superior, por su capacidad de atención selectiva, este enfoque ha mostrado resultados muy buenos y cercanos a los

obtenidos con la media, pero no ha alcanzado el mejor rendimiento. Esto puede ser por su sensibilidad al ruido en las etiquetas, por su sensibilidad al tamaño de las bolsas o por otros factores que podrían ser explorados.

En los resultados obtenidos mediante la agregación de la media, se observa que la sensibilidad, una métrica crucial que representa el número de falsos negativos (FN), supera el valor de 0.95, lo cual indica un rendimiento muy satisfactorio. Además, tanto la precisión como la especificidad llegan a un valor de 1. Estos valores son altamente positivos. Además, se puede ver que todas las métricas, excepto la sensibilidad, son mejores que para aprendizaje supervisado tradicional.

Al igual que en el aprendizaje supervisado, vamos a analizar los CAMs para verificar si las métricas obtenidas representan adecuadamente el nivel de calidad de la clasificación del modelo. De esta manera, podremos identificar en qué parte de la imagen se enfoca el modelo para llevar a cabo la clasificación y evaluar si se está identificando correctamente los tejidos de interés. Como se puede ver en la Figura 3.4, observando el CAM de la izquierda, podemos ver que el modelo está siendo capaz de detectar los límites del tejido y fijarse en la zona de la imagen en la que hay core. Con el CAM de la derecha podemos ver que además ha aprendido a seleccionar correctamente en que tipo de tejidos fijarse, ya que se puede ver que no ha fijado su atención en las células rosas de la esquina inferior izquierda, las cuales no son características de ninguna de las dos patologías que se quieren detectar.

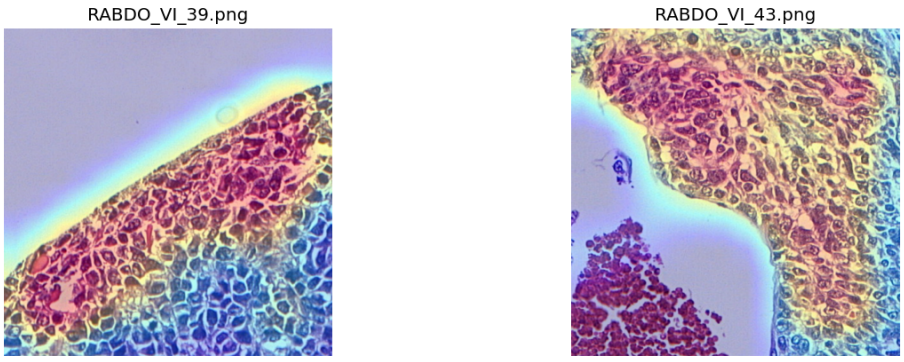

Figura 3.4: CAMs de imágenes clasificadas correctamente mediante MIL

Capítulo 4

Conclusiones y propuesta de trabajo futuro

Llevar a cabo este proyecto nos ha permitido explorar el potencial que tienen los modelos basados en aprendizaje profundo para la clasificación de imágenes médicas, en particular las histológicas. Hemos observado que estos sistemas de ayuda pueden brindar un gran apoyo a los profesionales médicos en el diagnóstico de patologías.

Hemos implementado modelos de aprendizaje profundo con diferentes arquitecturas y hemos encontrado que ResNet-50 y AlexNet ofrecen los mejores resultados. Entre las dos, ResNet-50 ha logrado resultados ligeramente superiores, lo que demuestra que el aprendizaje residual es una técnica útil que permite el entrenamiento de la red, evitando el problema del desvanecimiento del gradiente en las primeras capas.

En cuanto al transfer learning, hemos descubierto que cuanto más reentrenamos las capas de la red, mejores resultados obtenemos. La extracción de características por sí sola no ha superado el 50 % de precisión, mientras que con el *fine-tuning* hemos alcanzado precisiones de hasta el 78 %. Sin embargo, los mejores resultados se han logrado al reentrenar completamente la red, llegando al 94 % de precisión con la arquitectura de ResNet-50.

En cuanto a MIL, hemos encontrado que tanto la agregación de la media de las características como la técnica *MILAttention* proporcionan resultados muy buenos para imágenes histológicas, siendo ligeramente mejores en el caso de la media. Esto puede resultar sorprendente, ya que el resultado esperado era un mejor rendimiento para *MILAttention*. Aun así, se puede entender, puesto que los resultados son muy similares y la agregación de la media también es conocida por su excelente rendimiento. Esta es un caso particular de *MILAttention* en la que todas las instancias tienen el mismo peso y puede haber ocurrido que en este caso todas las instancias hayan sido igual de significativas. Las métricas obtenidas para la agregación de la media han sido muy positivas y permiten concluir que este tipo de enfoque tiene la oportunidad de ofrecer un alto rendimiento para aplicaciones similares. Además, se ha visto que con MIL se pueden llegar a hacer mejores predicciones que con el aprendizaje supervisado tradicional, ya que se han obtenido resultados ligeramente mejores en todas las métricas, excepto la sensibilidad, que ha sido ligeramente inferior.

Una de las limitaciones del sistema clasificador diseñado es que está entrenado para diferenciar imágenes de SE y rabdomiosarcoma, en vez de diferenciar las de SE frente a cualquier otro tipo de tumor de células redondas y pequeñas. Conseguir la diferenciación con cualquier tipo de tumor de

células redondas y pequeñas sería una línea de trabajo futuro de mucho provecho, ya que, como se ha explicado en los objetivos del proyecto, la diferenciación total marcaría una gran diferencia en la esperanza de supervivencia de los pacientes gracias a la posibilidad de ofrecer un diagnóstico certero.

Otra de las propuestas de trabajo sería profundizar en el uso de la técnica MIL con la agregación de características de *MILAttention* para obtener resultados aún mejores. Se considera que esta combinación tiene un gran potencial y que aún no se ha alcanzado su máximo rendimiento. Una oportunidad de mejora sería utilizar imágenes de mayor tamaño que las empleadas en este proyecto, que fueron de 512x512 píxeles. Sin embargo, es importante tener en cuenta que esto implica un costo elevado, ya que se requieren muchos recursos para entrenar modelos con todas las imágenes necesarias para implementar MIL y cuanto mayor tamaño tienen más se necesitan.

Bibliografía

[1] Instituto nacional del cáncer. *¿Qué es el cáncer?* 2021. URL: https://www.cancer.gov/espanol/cancer/naturaleza/que-es.

[2] American Cancer Society. *Si usted tiene un sarcoma de tejidos blandos*. 2023. URL: https://www.cancer.org/es/cancer/tipos/sarcoma-de-tejidos-blandos/si-usted-tiene-sarcoma.html#:~:text=El%5C%20sarcoma%5C%20es%5C%20un%5C%20tipo,tejidos%5C%20profundos%5C%20de%5C%20la%5C%20piel..

[3] Miranda Arellano Diana Belén Cuenca Mora Tatiana Karolina. *Ewing's Sarcoma: A Case Report*. 2021. URL: file:///Users/anarubiogil/Downloads/9664-Article%5C%20Text-45177-2-10-20210930.pdf.

[4] Hospital Sant Joan de Déu. *Se descubre un gen crucial para el desarrollo del sarcoma de Ewing*. 2020. URL: https://www.sjdhospitalbarcelona.org/es/noticias/descubre-gen-crucial-desarrollo-del-sarcoma-ewing.

[5] American Society of Clinical Oncology. *Sarcoma de Ewing en la niñez y adolescencia: Estadísticas*. 2021. URL: https://www.cancer.net/es/tipos-de-c%C3%A1ncer/sarcoma-de-ewing-en-la-ni%C3%B1ez-y-adolescencia/estad%C3%ADsticas.

[6] Revista Española de Cirugía Ortopédica y Traumatología. *Sarcoma de Ewing, análisis de supervivencia a los 6 años con terapia multidisciplinar*. 2018. URL: https://www.elsevier.es/es-revista-revista-espanola-cirugia-ortopedica-traumatologia-129-avance-resumen-sarcoma-ewing-analisis-supervivencia-los-S188844151830170X#:~:text=En%5C%20Espa%5C%C3%5C%B1a%5C%2C%5C%20el%5C%20sarcoma%5C%20de,infancia%5C%2C%5C%20por%5C%20delante%5C%20del%5C%20osteosarcoma..

[7] Francisco José Paz-Gómez. *Tumor de células pequeñas redondas y azules: abordaje diagnóstico*. 2004. URL: https://www.medigraphic.com/pdfs/medsur/ms-2004/ms041c.pdf.

[8] American Society of Clinical Oncology. *Sarcoma de Ewing en la niñez y adolescencia: Diagnóstico*. 2022. URL: https://www.cancer.net/es/tipos-de-c%5C%C3%5C%A1ncer/sarcoma-de-ewing-en-la-ni%5C%C3%5C%B1ez-y-adolescencia/diagn%5C%C3%5C%B3stico.

[9] Mayo Clinic. *Biopsia: Algunos tipos de biopsia que se utilizan para diagnosticar el cáncer*. 2021. URL: https://www.mayoclinic.org/es-es/diseases-conditions/cancer/in-depth/biopsy/art-20043922.

[10] Cancer care of western new york. *Sarcoma de Ewing*. 2023. URL: https://www.cancercarewny.com/content.aspx?chunkiid=247778.

[11] Instituto nacional del cáncer. *tinción con hematoxilina y eosina*. 2023. URL: https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/tincion-con-hematoxilina-y-eosina.

[12] Tristar technology group. *Lung Cancer (NSCLC) 250 Tissue Microarray*. 2020. URL: https://tristargroup.us/product/lung-cancer-nsclc-250-tissue-microarray/.

[13] Hamza Mousa. *10 Open-source Whole-Slide Image Viewers and Analysis Programs; Redefining Digital Pathology*. 2019. URL: https://medevel.com/10-os-whole-slide-image/.

[14] Noticias Parlamento Europeo. *¿Qué es la inteligencia artificial y cómo se usa?* 2021. URL: https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa.

[15] BeSmart Corp. *¿Diferencia entre Artificial Intelligence, Machine Learning y DP?* 2021. URL: https://besmartcorp.com/blog/f/%5C%C2%5C%BFdiferencia-entre-artificial-intelligence-machine-learning-y-dp.

[16] Alejandro Cartas. *Diagrama de un perceptron con cinco señales de entrada*. 2021. URL: https://web.archive.org/web/20170509011136/https://commons.wikimedia.org/wiki/File:Perceptr%C3%B3n_5_unidades.svg.

[17] Alexis Alulema. *Funciones de Activación*. 2022. URL: https://alexisalulema.com/es/2022/09/23/funciones-de-activacion-en-tensorflow/.

[18] Shiksha Online. *A Comprehensive Guide to Convolutional Neural Networks*. 2023. URL: https://www.shiksha.com/online-courses/articles/a-comprehensive-guide-to-convolutional-neural-networks/.

[19] Sergio Gonzalez. *Redes neuronales para clasificar imágenes: Tensorflow y Keras*. 2021. URL: https://www.modeldifferently.com/2021/10/image_classification/.

[20] Computer science wiki. *Max-pooling / Pooling*. 2018. URL: https://computersciencewiki.org/index.php/Max-pooling_/_Pooling.

[21] Juan Domingo Farnos. *El aprendizaje por transferencia es la idea de superar el paradigma de aprendizaje aislado y utilizar el conocimiento adquirido para una tarea para resolver los relacionados.* 2019. URL: https://juandomingofarnos.wordpress.com/2019/05/10/el-aprendizaje-por-transferencia-es-la-idea-de-superar-el-paradigma-de-aprendizaje-aislado-y-utilizar-el-conocimiento-adquirido-para-una-tarea-para-resolver-los-relacionados/.

[22] Han R Liao H Long Y. *Deep learning-based classification and mutation predictionfrom histopathological images of hepatocellular carcinoma*. 2021. URL: https://onlinelibrary.wiley.com/doi/epdf/10.1002/ctm2.102.

[23] Luiz Eduardo Oliveirac Caroline Petitjeanb Fabio Spanhol. *Multiple instance learning for histopathological breast cancer image classification*. 2018. URL: https://www.sciencedirect.com/science/article/pii/S0957417418306262?ref=cra_js_challenge&fr=RR-1.

[24] JAN NEUMANN SARAH CONSALVO FLORIAN HINTERWIMMER. *Two-Phase Deep Learning Algorithm for Detection and Differentiation of Ewing Sarcoma and Acute Osteomyelitis in Paediatric Radiographs*. 2022. URL: https://ar.iiarjournals.org/content/anticanres/42/9/4371.full.pdf.

[25] Isidro Machado Puerto. *Histological heterogeneity of Ewing's sarcoma/PNET: an immunohistochemical analysis of 415 genetically confirmed cases with clinical support*. 2009. URL: https://link.springer.com/article/10.1007/s00428-009-0842-7.

[26] Isidro Machado Puerto. *Tumores de células redondas y pequeñas de hueso y partes blandas con especial referencia al sarcoma de ewing/pnet y su diagnóstico diferencial. Un estudio de 841 casos*. 2011. URL: https://dialnet.unirioja.es/servlet/tesis?codigo=254330.

[27] Francisco Vargas-Bonilla Maribel Arroyave-Giraldo Alejandro Restrepo-Martínez2. *Incidencia de la Segmentación en la Obtención de Región de Interés en Imágenes de Palma de la Mano*. 2011. URL: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-77992011000200008.

[28] Pontificia Universidad Católica del Perú. *¿Qué es MATLAB?* 2021. URL: https://proyecto-matlab.pucp.edu.pe/que-es-matlab/index.htm#:~:text=MATLAB%5C%2C%5C%20el%5C%20lenguaje%5C%20de%5C%20c%5C%C3%5C%A1lculo,sistemas%5C%20din%5C%C3%5C%A1micos%5C%20multidominio%5C%20e%5C%20integrados..

[29] OpenWebinars. *Qué es Visual Studio Code y qué ventajas ofrece*. 2022. URL: https://openwebinars.net/blog/que-es-visual-studio-code-y-que-ventajas-ofrece/.

[30] Daniel. *VGG: ¿Qué es este modelo? ¡Daniel te lo cuenta todo!* 2023. URL: https://datascientest.com/es/vgg-que-es-este-modelo-daniel-te-lo-cuenta-todo.

[31] Luis Casaverde Pacherrez Daniel Pérez Aguilar Redy Risco Ramos. *Transfer learning en la clasificación binaria de imágenes térmicas*. 2021. URL: https://www.researchgate.net/publication/356119399_Transfer_learning_en_la_clasificacion_binaria_de_imagenes_termicas.

[32] Hamad Al Jassmi Luqman Ali Fady Alnajjar. *Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures*. 2021. URL: https://www.researchgate.net/publication/349717475_Performance_Evaluation_of_Deep_CNN-Based_Crack_Detection_and_Localization_Techniques_for_Concrete_Structures.

[33] Liqin Cao Xiaobing Han Yanfei Zhong y Liangpei Zhang. *Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification*. 2017. URL: https://www.mdpi.com/2072-4292/9/8/848.

[34] Google Cloud. *Guía avanzada de Inception v3*. 2023. URL: https://cloud.google.com/tpu/docs/inception-v3-advanced?hl=es-419.

[35] Paulo Maia. *An Introduction to Multiple Instance Learning*. 2021. URL: https://nilg.ai/202105/an-introduction-to-multiple-instance-learning/.

[36] Max Welling Maximilian Ilse Jakub M. Tomczak. *Attention-based Deep Multiple Instance Learning*. 2018. URL: https://arxiv.org/pdf/1802.04712.pdf.

[37] Juan Ignacio Barrios Arce. *La matriz de confusión y sus métricas*. 2019. URL: https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/.

[38] François Chollet. *Grad-CAM class activation visualization*. 2021. URL: https://keras.io/examples/vision/grad_cam/.

Listado de siglas empleadas

ANN Artificial neural network.

AUC Area under curve.

CAM Class activation maps.

CNN Convolutional Neural Network.

DL Deep learning.

FN Falsos negativos.

FP Falsos positivos.

HE Hematoxilina Eosina.

IA Inteligencia artificial.

IVO Instituto Valenciano de Oncología.

ML Machine learning.

MLP Multilayer perceptron.

MSE Mean square error.

RMN Resonancia Magnética Nuclear.

SE Sarcoma de Ewing.

TAC Tomografía Axial Computerizada.

TCRP Tumor de células redondeadas y pequeñas.

TIFF Formato de archivo de imágenes con etiquetas.

TMA Tissue microarrays.

VN Verdaderos negativos.

VP Verdaderos positivos.

WSI Whole slide images.

Parte II

Anexos

Apéndice A

Tabla de relación del trabajo con los Objetivos de Desarrollo Sostenible de la agenda 2030

ODS	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza				x
ODS 2. Hambre cero				x
ODS 3. Salud y bienestar	x			
ODS 4. Educación de calidad				x
ODS 5. Igualdad de género				x
ODS 6. Agua limpia y saneamiento				x
ODS 7. Energía asequible y no contaminante				x
ODS 8. Trabajo decente y crecimiento económico				x
ODS 9. Industria, innovación e infraestructuras		x		
ODS 10. Reducción de las desigualdades			x	
ODS 11. Ciudades y comunidades sostenibles				x
ODS 12. Producción y consumo responsables				x
ODS 13. Acción por el clima				x
ODS 14. Vida submarina				x
ODS 15. Vida de ecosistemas terrestres				x
ODS 16. Paz, justicia e instituciones sólidas				x
ODS 17. Alianzas para lograr objetivos				x

Tabla A.1: Objetivos de Desarrollo Sostenible

Descripción de la alineación del TFG/TFM con los ODS con un grado de relación más alto:

- ODS 3. Salud y bienestar: Este proyecto está altamente relacionado con este Objetivo de Desarrollo Sostenible. Esto es porque la implementación de esta tecnología tiene el potencial de mejorar significativamente los pronósticos para pacientes con sarcoma de Ewing, proporcionando diagnósticos más precisos y tempranos. Además, la tecnología desarrollada

podría ser aplicada a otros tipos de enfermedades, ampliando su utilidad y beneficios a un espectro más amplio de pacientes para que más personas puedan beneficiarse de ella.

- ODS 9. Industria, innovación e infraestructuras: El uso de la inteligencia artificial aplicado a aspectos médicos es algo aún poco extendido. Es por eso que se considera que este proyecto es algo innovador y que, por lo tanto, tiene una relación con este Objetivo de Desarrollo Sostenible.
- ODS 10. Reducción de las desigualdades: Se considera que este proyecto está ligeramente relacionado con este Objetivo de Desarrollo Sostenible, puesto que al impulsar avances en la detección del SE, posibilita que en un futuro sea más accesible un diagnóstico y, por lo tanto, cualquier persona que lo necesite tenga la posibilidad de recibirlo.