Reconocimiento automático de signos radiológicos de atelectasia en radiografías de tórax con CNN

Máster en Física Médica

Trabajo de Comienzo de Investigación

Tutores: Dr. Ángel Pérez de Madrid y Pablo

Dr. Miguel Romero Hortelano

Autor: D. José Luis Pastor Sanz

Resumen

En este estudio se aplican algunos de los métodos automáticos de detección de imágenes en la especialidad radiológica de diagnóstico a partir de radiografía de tórax de pulmón, especialidad que ha cobrado relevancia durante la reciente pandemia de coronavirus. En particular se centra en la localización de atelectasias por medio de redes neuronales de convolución (CNN). Tras un ensayo inicial con una red básica sobre la base de datos de partida, donde se estudia el impacto de las diferentes configuraciones en los resultados, se ensayan métodos de mejora simplificando la base de imágenes en términos del tipo de atelectasia que se pretende descubrir. Por otra parte, se propone un método de mejora basado en la reeducación de los parámetros de la red con incremento de ejemplos en la base de entrenamiento a partir de imágenes que clasifican mal en la fase de validación.

Summary

This study explores some of the automatic image detection methods applied in radiological specialty of diagnosis from lung chest radiography, a specialty that has become relevant during the recent coronavirus pandemic. It focuses on the localization of atelectasis by means of convolution neural networks (CNN). After an initial test with a basic network over the initial database, where the impact of the different configurations on the results is studied, methods to improve are tested by simplifying the database in terms of the type of atelectasis to be discovered. On the other hand, an improvement method based on the updating of the network parameters with an increase of examples in the training base from images that misclassify in the validation phase is proposed.

Contenido

Introducción

1.1. Anatomía torácica e interpretación de la atelectasia en radiografías de tórax.

1.1.1. Anatomía de la región y principios de radiología torácica.

En un trabajo cuyo objetivo es la identificación de un patrón pulmonar específico de forma automática a partir de una radiografía de tórax es conveniente primero introducir la forma en que los radiólogos comúnmente interpretan dichas imágenes en busca de signos, así como detallar los espacios habituales del pulmón y aledaños que pueden ser observados en dicho formato. Y para ello, son precisas unas cuantas nociones básicas de anatomía de la región implicada.

En este punto es preciso señalar que son varios los sistemas implicados en una RX de tórax. El espacio radiado comprende a las dos cavidades corporales principales del cuerpo, la cavidad dorsal y la ventral. Dentro de esta última tenemos la cavidad torácica completa, y la parte superior de la cavidad abdominal. De la cavidad dorsal, la parte implicada es la correspondiente del canal vertebral. La cavidad torácica, protagonista indiscutible de nuestra imagen, se encuentra rodeada de las costillas, los músculos pectorales, esternón, y porción torácica de la columna vertebral.

Dentro de la cavidad torácica, encontramos tres cavidades menores. En primer lugar, la cavidad pericárdica, que comprende el corazón, las dos cavidades pleurales, que contienen sendos pulmones, y la cavidad mediastínica, que contiene todas las vísceras de la cavidad torácica excepto los pulmones: esófago, tráquea, timo, corazón y algunos vasos sanguíneos importantes (Figura 1) [1].

El diafragma separa la cavidad torácica de la abdominal. De esta última, tenemos que la parte superior del estómago y el hígado son parte de la sombra radiológica, aunque no son objeto de estudio por este método de diagnóstico.

La tráquea, que ocupa el espacio entre los dos pulmones, y se sitúa en posición anterior sobre el esófago, se subdivide en la carina en dos bronquios primarios, visibles en RX, el derecho un poco mayor que el izquierdo. Una vez entran en los pulmones, se dividen en los secundarios, uno por cada lóbulo pulmonar, donde dejan de ser visibles en condiciones normales. A su vez en 10 bronquios terciarios, en lo que se conoce como segmento broncopulmonar, que acaban en lo que se conoce como lobulillos, dotados de vénula, arteriola, vaso linfático y rama de bronquiolo terminal. Estos derivan en los bronquiolos respiratorios, que se subdividen por fin en conductos alveolares, tras 25 órdenes de ramificación desde la tráquea (Figura 2) [1].

Así pues, los pulmones se subdividen en lóbulos a partir de las fisuras, segmentos lobulares, y lobulillos. En apartados siguientes, se relatará la posición de los elementos visibles más importantes de estas estructuras en una RX de tórax, así como sus contornos normales más representativos.

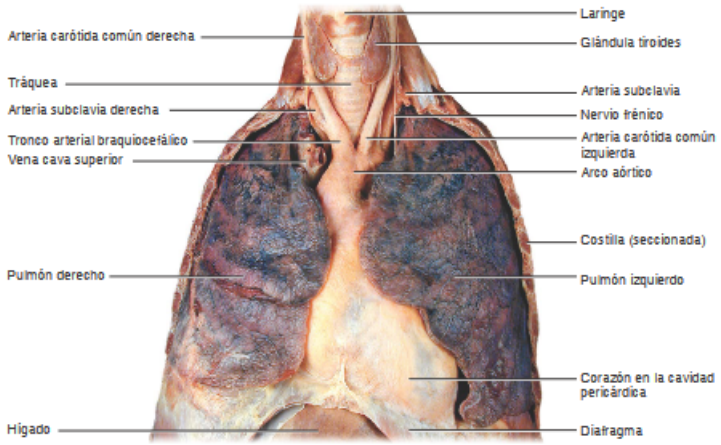


Figura 1. Anatomía básica de los pulmones en foto realista.

(Fuente: Principios de Anatomía y Fisiología, Tortora y Grabowsky, 2005)

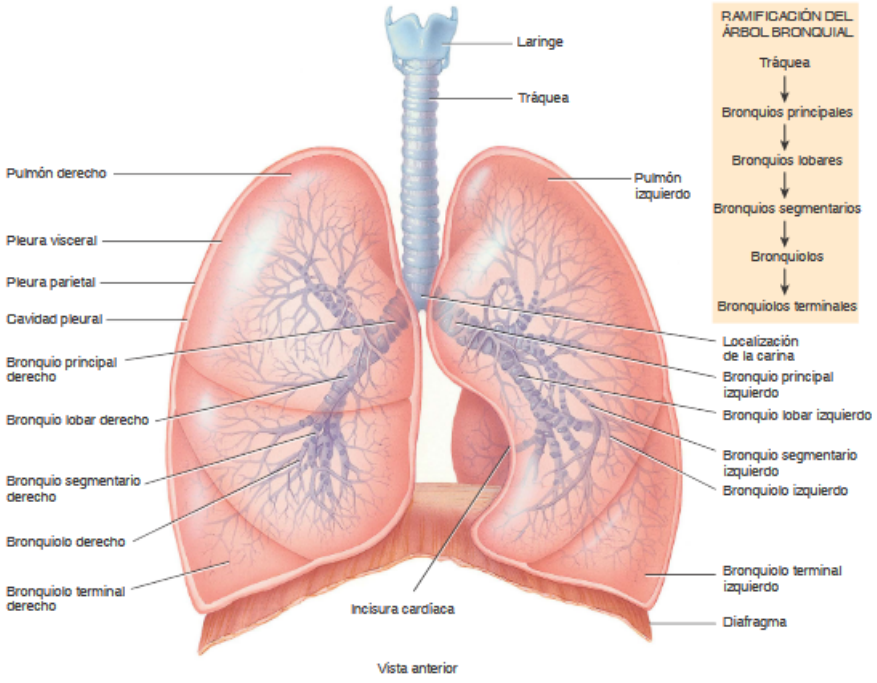


Figura 2. Lóbulos y vasos de los pulmones, principales localizaciones.

Principios de radiología torácica.

Las radiografías de las que partimos están en principio tomadas por un equipo no portátil en dirección postero-anterior (PA), y se radian en bipedestación e inspiración forzada para aumentar el tamaño de los objetos, en situación de mínima magnificación y máxima nitidez, esto es, con el paciente pegado a la placa. Asumimos entonces una dificultad añadida en la detección de signos que queden por detrás del corazón, que requieren una orientación lateral de la radiografía.

Una exposición alta de la placa a RX, por ejemplo, a través de un tejido poco absorbente, como el aire, precipita mucha plata y la ennegrece (entiéndase el fenómeno digital equivalente actual). Y a la inversa, si llega poca cantidad de RX, porque la radiación ha atravesado un tejido denso, no precipita compuesto y permanece blanca. Entre los dos extremos blanco-negro, tenemos un rango de densidades que podremos normalizar para obtener información extra sobre la naturaleza de los tejidos. Los distintos grados corresponden a aire, grasa, tejidos blandos (músculo, líquido), y metal (hueso). El corazón y el diafragma son de densidad agua. Tejidos blandos, el pulmón, la luz de la tráquea (los cartílagos tienen una densidad similar a los huesos) y la burbuja gástrica son de densidad aire. Las costillas son de densidad metal.

La radiografía es una imagen en dos dimensiones de la que se requiere extraer información tridimensional. En ella se superponen todas las estructuras que están en la dirección en la que inciden los rayos X, en este caso, dirección anteroposterior. Los puntos clave en toda radiografía de tórax, son los siguientes, como se aprecia en la Figura 3 [2]:

- A. Presencia de aire en otras estructuras externas.
- B. Ángulo costofrénico.
- C. Corazón.
- D. Aorta descendente.
- E. Tráquea
- F. Carina.
- G. Hilio.
- H. Botón aórtico.
- J. Aorta ascendente.

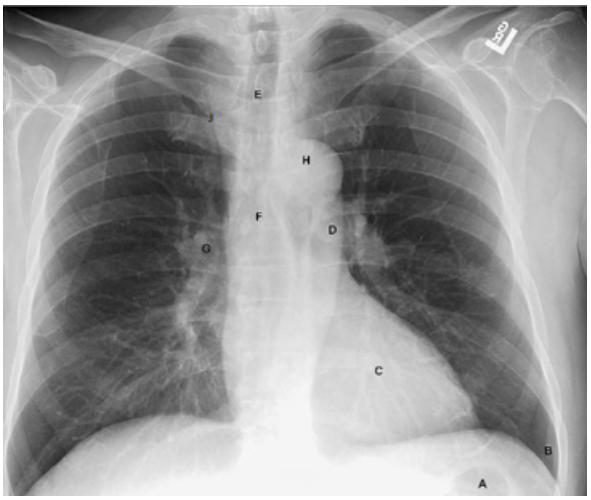


Figura 3. Signos radiográficos fundamentales en una RX de tórax [2].

Una exploración exhaustiva de una RX de tórax requiere empezar chequeando la parte abdominal (densidad homogénea del hígado, posible bazo, espacio gástrico y ángulo hepático y esplénico del colon, hemidiafragma), posteriormente la parte torácica (mama derecha, arco posterior de la costilla, escápula, clavícula, arco anterior de la costilla), la región mediastínica (tráquea, carina, botón aórtico, aorta ascendente, corazón, hilio derecho más bajo que el izquierdo).

En cuanto al parénquima, esto es, la parte interior de los pulmones, los alveolos son radiotransparentes. Los vasos hiliares son lo único visible del intersticio (formado por la red de soporte de los alveolos, esto es, vasos sanguíneos, linfáticos, bronquios y tejido conectivo), y sólo son visibles en la parte proximal al hilio. Luego los vasos se ramifican y se hacen demasiado pequeños para la resolución de una RXT.

Anatomía Lobar

Es necesario entender los patrones de la anatomía lobar para conocer las diferentes patologías asociadas al pulmón que afectan específicamente a los lóbulos. El pulmón derecho tiene tres lóbulos (superior, medio e inferior), y el izquierdo dos (superior o língula, e inferior). Los lóbulos vienen delimitados por las cisuras lobulares o septos, recubiertos por pleura visceral. Dichas cisuras sólo son visibles, dado su pequeño grosor, si el haz de rayos incide perpendicular a su superficie, así pues, no todas las cisuras son visibles en proyección PA (por ejemplo, la cisura mayor oblicua del pulmón izquierdo sólo es visible en radiografía lateral, o la cisura mayor en el pulmón derecho que separa los lóbulos superior y medio del lóbulo inferior). La cisura consta entonces de aire y queda delimitada por pleura. Si el líquido pleural penetra en la fisura, esta aumenta su grosor.

Entonces sólo es visible en bipedestación y proyección frontal la cisura menor horizontal, siempre y cuando la cisura mantenga en el paciente dicha orientación paralela a los rayos (no siempre ocurre). Las cisuras ayudan a localizar la afección, y son el signo radiológico más fiable en la atelectasia

lobar.

El signo de la silueta

Los contrastes entre regiones de diferente densidad sirven para delimitar las siluetas. Las curvas que se forman son un instrumento indispensable para localizar afecciones. Se denomina signo de la silueta a la pérdida de la curva por anomalías. Por ejemplo, una consolidación en un lóbulo inferior borraría la línea del hemidiafragma. Esto sería un signo de la silueta.

Es importante conocer su localización precisa de cada elemento si queremos detectar el signo de la silueta. Por ejemplo, los bordes cardíacos tienen localización anterior, la aorta descendente, localización posterior, la aorta ascendente, localización anterior, el botón aórtico, medio posterior. El lóbulo medio y la língula son estructuras anteriores en contacto con el corazón. Si se observan estas estructuras, se trata de un signo de la silueta debido seguramente a consolidaciones. Los lóbulos inferiores, están en posición inferior y posterior con respecto a la cisura mayor. No están en contacto con los bordes cardíacos, y reposan sobre los diafragmas.

Si está borrado el hemidiafragma derecho, el lóbulo con afección tiene que ser el inferior. Si además se borra el borde cardíaco, la patología también reside en lóbulo medio. Su enfermedad se superpone al hilio y al borde cardíaco, pero no borran su silueta, por no estar en contacto directo. Una consolidación en el lóbulo superior derecho, por encima de la cisura menor, y anterior a la cisura mayor, produce el signo de la silueta en la aorta ascendente y la línea paratraqueal derecha. Si el afectado es el lóbulo superior izquierdo, lo que desaparece es la aurícula izquierda y el botón aórtico. Estos hechos serán de gran utilidad para la mejora del etiquetado de las imágenes de atelectasia para el diseño de algoritmo de reconocimiento automático.

El broncograma aéreo

Los bronquios, de densidad aire, son visibles en el entorno del mediastino, de densidad tejidos blandos o agua. Sin embargo, los bronquios periféricos, cuando penetran en los pulmones dejan de ser visibles puesto que llegan a un entorno de densidad similar y tienen paredes muy finas. En los pulmones los únicos vasos visibles son los vasos sanguíneos, de densidad agua en entorno aire.

Si el entorno de los bronquios cambia de densidad, de densidad aire a densidad agua, los vasos ahora se encuentran rodeados de un medio de densidad diferente y son visibles en radiografía de tórax. Se detecta entonces el llamado signo del broncograma aéreo en pulmones con patologías. Igualmente, si no se observan los vasos sanguíneos, es igualmente signo de patología, pues se dejan de observar espacios de densidad agua, hecho que indica que el entorno es de una densidad similar y por tanto anormal.

Sin embargo, y es un caso frecuente en la afección que es objeto de estudio en este proyecto, el

broncograma dejará de ser visible y el bronquio tiene algún tipo de obstrucción (por secreciones, mucosa u objetos, como es el caso de asma o cáncer de origen bronquial), pues dejará de contener aire. Así mismo hay que hacer constar que una fibrosis intersticial cursa sin detección del broncograma aéreo. Tampoco aparece este signo en caso de enfisema o si la consolidación del pulmón no es completa.

Si la patología es en el lóbulo inferior izquierdo, el corazón en ocasiones enmascara el signo radiográfico de dicha patología. En estos casos, adquieren especial relevancia la presencia de broncograma aéreo en esa zona.

Como veremos en secciones posteriores un broncograma apretado es un claro signo de atelectasia pulmonar no obstructiva.

Signo de colapso lobar y pulmonar

Aunque ambos suponen pérdida general de volumen en los pulmones, la diferencia entre colapso y atelectasia es difusa. En general se espera de la atelectasia una disminución menor que en el caso de colapso.

Para determinar el grado de colapso, es conveniente observar los desplazamientos en los elementos que rodean al pulmón o lóbulo desplazado. Por ejemplo, si la tráquea está desplazada de su línea media, o el corazón pierde su silueta porque se desplaza en dirección al pulmón colapsado. El mejor dato, sin embargo, lo tenemos con el desplazamiento de la cisura, que se mueve en dirección al lóbulo colapsado. Recordemos que, en una radiografía frontal, no todas las cisuras son visibles, en cualquier caso.

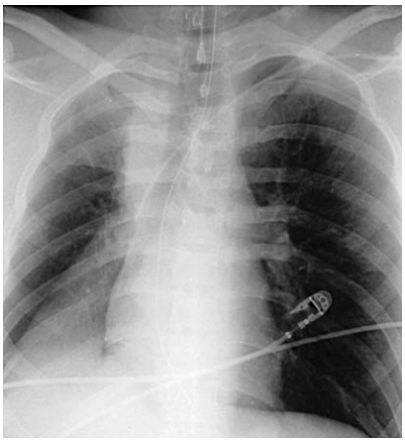


Figura 4. Atelectasia de lóbulo superior e inferior derecho: Cisura menor elevada, signo de la silueta en diafragma derecho, lóbulo medio aireado, que permite observar cisura y borde cardíaco. (Fuente: Felson, Principios de radiología torácica, 2007)

Otro signo habitual de pérdida de volumen pulmonar es el desplazamiento hacia arriba del diafragma, aun manteniendo su silueta habitual, por colapso de lóbulo medio o superior. Si la obstrucción es a nivel endobronquial en el lóbulo izquierdo, es probable el colapso de lóbulo superior izquierdo y língula,

pues comparte esa vía respiratoria. Igualmente pasaría con los lóbulos superior y medio derechos.

Otro signo de pulmón atelectático no obstructivo, como se ha comentado en el apartado anterior, es la aparición de broncograma aéreo apretado. Igualmente, si la consolidación no es suficiente, pueden observarse marcas vasculares de área reducida. Si la consolidación es debida al colapso del lóbulo inferior izquierdo, tras el corazón, se puede detectar si aparece broncograma en el espacio cardíaco.

Signos indirectos de atelectasia, son el desplazamiento del hilio y diafragma. El desplazamiento es en dirección al lóbulo colapsado. Si el lóbulo medio o la língula son los fragmentos que reducen su volumen, el hilio permanece en su lugar. Normalmente el hilio izquierdo está por encima del derecho, así que, si esto no ocurre, puede ser un dato que contrastar. El diafragma derecho está normalmente mas elevado que el izquierdo. Si esto no ocurre, puede estar ocurriendo un colapso lobar. Otro tanto puede decirse de las estructuras medianísticas o el corazón.

Como queda claro entonces, el pulmón reducido es más radiopaco. Cambia su densidad y aparece como consolidado en la radiografía. Físicamente, el pulmón adyacente tiende a ocupar parte del espacio que deja el pulmón reducido, especialmente si la afección es crónica. Se dice entonces que el pulmón no enfermo está hiperinsuflado de forma compensatoria.

1.1.2. La atelectasia.

Cuando fallan los mecanismos fisiológicos que mantienen al pulmón distendido, se colapsa. Hay cinco mecanismos básicos que causan pérdida de volumen:

- La obstrucción de un bronquio reabsorbe el aire presente tras la obstrucción. La presión externa producto de un hemotórax.
- Si hay aire o líquido en el espacio pleural, sea este el parietal o el visceral.
- Como resultado de una intervención o de una fibrosis se producen cicatrices, que producen una contracción del pulmón. Igualmente, la sedación profunda afecta a las bases pulmonares.
- Si disminuye el surfactante que reduce la distensibilidad pulmonar (atelectasias adhesivas)
- Hipoventilación debida a depresión del sistema nervioso central o a dolor. Atelectasias laminares, de forma segmentaria.

La atelectasia es una de las enfermedades características del espacio aéreo del parénquima pulmonar. Se entiende por atelectasia una expansión incompleta del pulmón por obstrucción intrínseca o extrínseca de las vías respiratorias [3].

Una radiografía de tórax es una imagen fija del espacio torácico, que, precisamente por ser fija, no nos puede dar dicha información histórica temporal. Es por tanto que, el análisis de dicho patrón patológico se debe centrar en la localización de densidades extrañas en torno a espacios comúnmente aéreos, la pérdida de las siluetas habituales del pulmón, así como tamaños perceptiblemente anormales de dichos órganos en sus paradigmas habituales.

Por espacio aéreo en el pulmón se entiende en primer lugar los alveolos y vasos bronquiales, que son las estructuras de dicho espacio que comúnmente contienen gas. Para entender el conjunto de propiedades características de la imagen de la atelectasia pulmonar, es crucial en primer lugar explicar el concepto del broncograma aéreo. El broncograma aéreo está formado por los bronquios que contienen aire, rodeados por alveolos no aireados. Y se expresa en la radiografía como un sistema de ramas oscuras en torno a un conjunto homogéneo opaco, esto es, blanco. Se trata entonces, de una marca posible de atelectasia, pues el colapso de los alveolos se traduce en una menor expansión de los pulmones.

En los alveolos, en este caso, se sustituye el aire por sustancia líquida de diverso origen. Aunque en la tráquea y en los bronquios proximales, se observa perfectamente el aire, por estar rodeado del mediastino, que es un tejido de densidad cercana a la del agua, una vez que los vasos aéreos penetran en el espacio pulmonar, estos se hacen radiotransparentes y no aparecen de forma natural en las radiografías, por estar rodeados de los alveolos, que a su vez contienen aire. El conjunto bronquios distales alveolos no dejan huella radiológica, el espacio aparece oscuro. Sólo se hacen visibles en

determinadas circunstancias, como es el caso de la patología que centra nuestro estudio, en la que el espacio en torno a dichos conductos se rodea de forma persistente de una sustancia radiopaca. Si se observa el aire, es que está rodeado de un tejido más denso, que por tanto, no es normal. Dicha sustancia puede ser sangre (contusión o hemorragia), células (linfoma o carcinoma de células alveolares), o pus (neumonía). Toda vez que se observa este signo, se infiere generalmente una enfermedad asociada al espacio aéreo.

El broncograma aéreo es por tanto símbolo de lesión que afecta al parénquima, por tanto indicativa de condensación pulmonar. Los patrones intersticiales, generalmente no causan opacidad suficiente como para observar dicha ramificación de los vasos. La morfología de dicha estructura puede dar referencias importantes para el diagnóstico diferencial de una enfermedad. Una forma de distinguir precisamente la atelectasia de una neumonía es la presencia del árbol bronquial completo expandido en el caso de la neumonía. En la atelectasia, los bronquios se juntan adoptando una disposición paralela estrujada, indicando pérdida de volumen. Esto es así, porque en el caso de dicha patología, los alveolos están vacíos y colapsados.

Si la obstrucción es a nivel bronquial proximal, puede ser que los bronquios distales no contengan aire, y no se aprecie el broncograma. En este caso, la imagen será indistinguible de un derrame pleural, es decir, una opacidad homogénea total. Habría entonces que complementar con otros signos, como el de un desplazamiento mediastínico homolateral, que no concordaría con un derrame masivo.

La atelectasia no es exclusivamente el resultado del colapso alveolar por una obstrucción bronquial (también llamada atelectasia reabsortiva). También puede producirse por comprensión extrínseca del pulmón (atelectasia pasiva), fibrosis del parénquima (cicatrización), o aumento de la tensión superficial del alveolo por deficiencia del surfactante (atelectasia adhesiva). En estos últimos casos se conserva intacta la vía bronquial.

1.1.3. Signos radiográficos de atelectasia en RX de tórax.

La pérdida de volumen en el pulmón va acompañada de una serie de signos identificativos directos sin los cuales pueden confundirse con otras patologías: cambio de posición de las fisuras o agrupamiento de los grupos bronquiales intrapulmonares. Si esto ocurre, puede ser que otras estructuras se vean afectadas, convirtiéndose esto en signos secundarios de un pulmón con atelectasia: desplazamiento del mediastino y el hilio, elevación del hemidiafragma, o la hiperaireación del pulmón adyacente, que es un signo de atelectasia persistente. El desplazamiento del hilio se observa a partir del hecho de que el hilio izquierdo se encuentra más elevado que el derecho. Si esto no ocurre, es posible que haya ocurrido un desplazamiento de uno de los dos por culpa de esta afección.

Veamos algunos ejemplos: Si el colapso viene del lóbulo superior derecho (LSD) se observa como la cisura menor, y parte de la mayor, se elevan. Así mismo se aprecia dicho lóbulo superior derecho aplastado contra el mediastino, y una consecuente elevación del hilio. Si el colapso ocurre en el lóbulo superior izquierdo (LSI) la fisura principal se desplaza hasta un plano casi paralelo a la pared torácica anterior. Si se comprime el lóbulo medio (LM) ambas fisuras se aproximan. Si el colapso se extiende hasta el lóbulo inferior (Inferiores) surge una opacidad que se extiende hasta el ángulo costofrénico, desdibujándose la silueta cardiaca. Puede simular una elevación del arco hemidiafragmático. Un esquema de los movimientos antes descritos puede verse en la Tabla 1.

Manifestaciones radiológicas de las atelectasias lobares					
	DIRECCIÓN DEL COLAPSO	POSICIÓN DEL HILIO	REORIENTACIÓN DE LAS CISURAS	POSICIÓN DEL DIAFRAGMA	REORIENTACIÓN TRAQUEAL
LSD	Superior y medial	Elevación	Craneomedial (menor) Anteromedial (mayor)	Elevación ipsilateral	Desviación hacia la derecha
Inferiores	Posteromedial	Descenso	Posteroinferior (mayor) Posteroinferior(menor)	Aproximación de las costillas	Verticalización del bronquio principal
LM	--	--	Posteroinferior (menor) Anterosuperior(mayor)	--	--
LSI	Anterior y superior	Elevación	Anterior (mayor)	Elevación ipsilateral	Desviación hacia la izquierda

Tabla 1. Cuadro resumen de los signos radiográficos de atelectasia lobular (Fuente: SERAM)

1.2. Redes de convolución

A finales de los años 80 aparece el concepto de red convolucional como una variante de las redes neuronales construidas a base de perceptrones. Las redes convolucionales son redes neuronales a las que se le añade una o más capas previas llamadas de convolución. Dichas capas procesan la entrada antes de ser tratada por la red. El proceso es similar a la operación misma entre dos funciones dadas. Cuando dos funciones *convolucionan* generan una nueva función que es el resultado de multiplicar una sobre otra a lo largo de un dominio que se va desplazando hasta que la primera función recorre completamente la segunda. Cuando aplicamos dicho concepto a las redes neuronales la convolución se realiza a partir de unos operadores denominados *kernels,* que son comúnmente matrices regulares de pequeña dimensión inicializadas de forma específica. Los *kernels,* también conocidos de forma más intuitiva como filtros, operan por el conjunto de datos de la entrada de la red generando una serie de salidas filtradas. El operador va recorriendo el conjunto de datos y se generan tantas salidas como filtros, como se puede apreciar en la Figura 5 [4]. En dicha figura se observa como el filtro, denominado *feature detector* en la figura, siendo en realidad el *kernel*, pues en el fondo va a resaltar ciertas propiedades de la imagen una vez ejecutado, opera sobre la imagen original para producir un *feature map*, o mapa de características de la imagen. Se observa en este ejemplo que la dimensión del mapa de características se ha reducido. Es esta una opción de la convolución que se deriva de la manera en la que el filtro opera sobre la imagen.

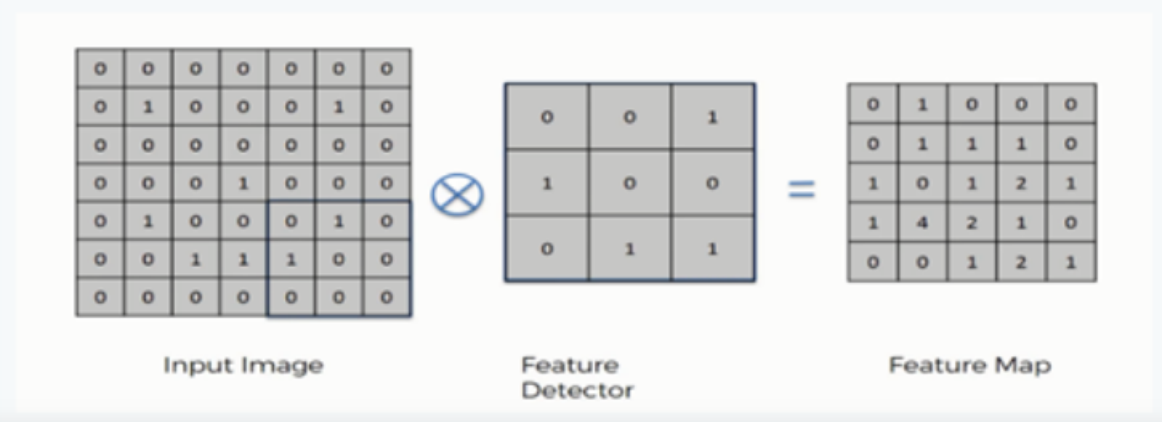


Figura 5. Proceso de convolución en redes neuronales de convolución. Aplicación del *kernel*.

Las capas convolucionales de la red extraen una serie de mapas de características de la imagen que serán las que alimenten la parte de esta consistente en capas densamente conectadas de perceptrones. Se generarán tantas imágenes como filtros se apliquen. Es decir, una capa de convolución de 64 filtros

producirá 64 imágenes filtradas nuevas basadas en la original que seguirán su proceso por la red, normalmente hacia otras capas de convolución. Dichos mapas contienen desde características de detalles de la imagen de "bajo nivel", tales como filos y esquinas, hasta detalles de "alto nivel", como grandes estructuras. Los valores de los píxeles en características de bajo nivel es muy probable que estén correlacionados, es decir que tengan valores próximos. Este hecho es lo que define las fronteras. Las fronteras son entornos con grandes diferencias de valor respecto a los píxeles de su vecindad.

Para ilustrar este hecho podemos mostrar algunas imágenes filtradas representativas de una radiografía de tórax. Se aprecia en la Figura 6 como la aplicación de los filtros, del que se ha puesto uno como ejemplo cada vez, va generando las distintas imágenes (una por cada filtro). Unas recuerdan a la imagen original, otras están a oscuras, unas contienen bordes, y algunas manchas significativas. Estas imágenes alimentaran la red de clasificación densa en el proceso de entrenamiento de la red. A medida que pasen los ciclos, los kernels irán corrigiendo sus valores y dichas imágenes se irán diferenciando hacia las categorías de clasificación.

En el modelo de programación que vamos a utilizar, y que explicaremos en el apartado *Tecnologías Utilizadas,* los valores asociados a los filtros, y el mapa de imágenes que resulta de aplicar la convolución, son entidades recogidas cuya visualización es inmediata con los comandos propios de visualización de imágenes. En la Figura 6 se incluye también el nombre de la capa correspondiente (conv2d_1…) y los valores asociados al *kernel*. normalizados entre 0 y 1. En cuanto al vector de dimensiones, los primeros dígitos atienden al tamaño del filtro, el tercero al número de inputs, y el cuarto al número de outputs. Así por ejemplo el vector (3,3,32,64) viene a decir que se procesarán 64 filtros de 3x3 en cada una de las 32 imágenes de entrada, resultando en 64 imágenes cada vez que se procesa cada una de ellas.

A parte de las etapas de convolución, que son el elemento más característico, estas redes se distinguen por incluir una serie de procesos intermedios que las definen igualmente, y que resulta necesario describir sucintamente en esta introducción

- Unidad de rectificación lineal (ReLU): Se trata de una función aplicada a las salidas de los nodos que pretende ajustar las imágenes filtradas a las discontinuidades de la imagen de entrada.
- Agrupación (Pooling): Asociar píxeles contiguos de tal forma que se reduce la dimensión de la imagen sin perder sus características (ver Figura 7 [4]).
- Aplanamiento: Disponer el mapa de características en columna para la alimentación de la parte densa de la red.
- Capa densa o conexión completa: Un conjunto de perceptrones interconectados, el concepto tradicional de red neuronal.

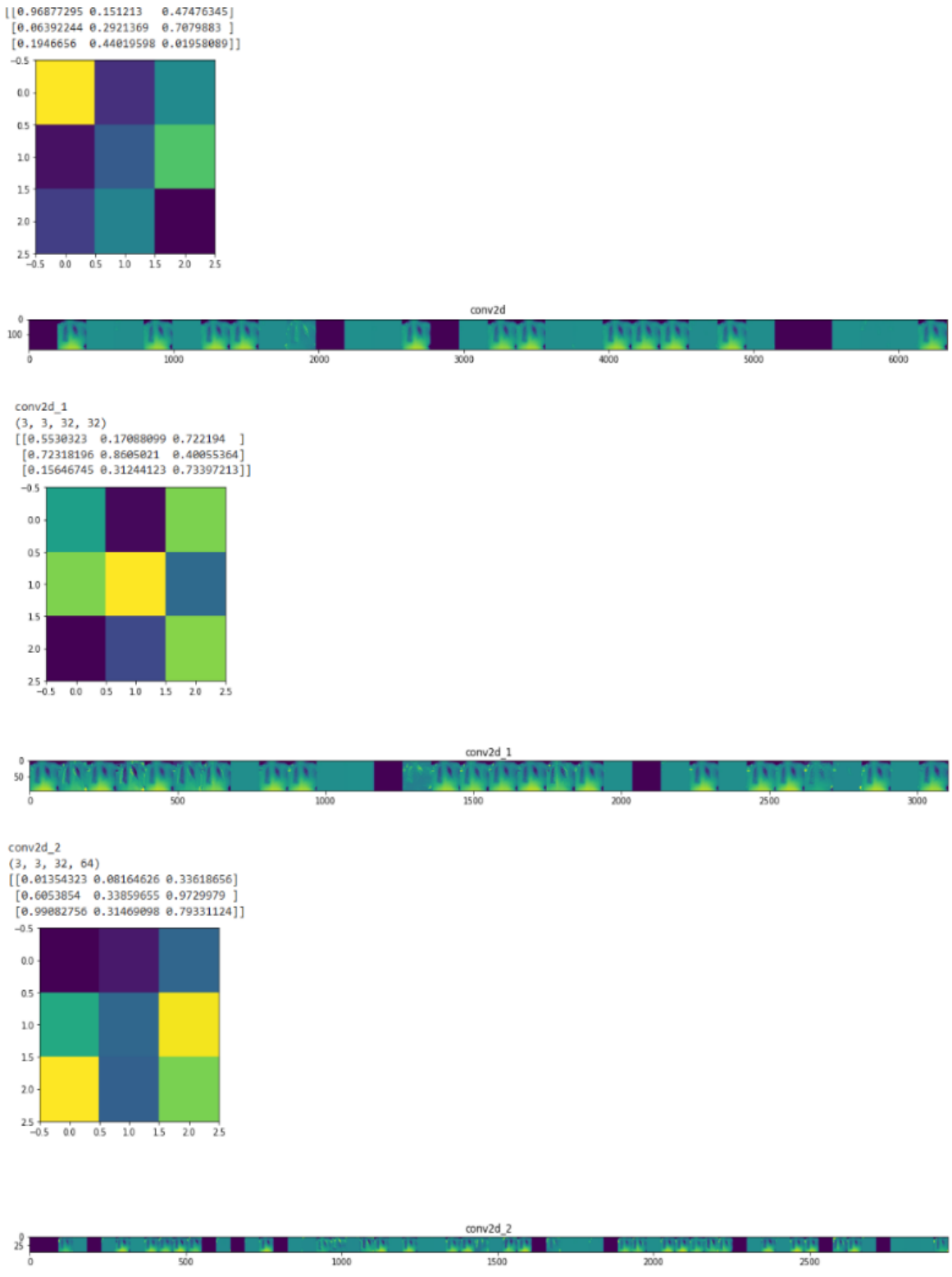


Figura 6. Aplicación de los kernels en las 3 capas de convolución y la generación de las imágenes filtradas sucesivas

El perceptrón es la unidad lógica de la red en la fase de capa densa. A imagen de una neurona, el perceptrón actualiza una serie de entradas en base a unos pesos editables, una función sesgo, y una función de activación que no transmite la entrada hasta que se alcanza cierto valor umbral.

Además de las capas que constituyen la estructura de la red de convolución, es fundamental como fluyen los datos para el proceso de ajuste que posibilita la clasificación por *retropropagación* en base a un entrenamiento de la red, esto es, como se van editando los parámetros que definen las operaciones matemáticas implícitas en el proceso, y como la red regula dicho refinamiento progresivo de forma supervisada, es decir, por comparación con la etiqueta que define el valor real de la imagen, para lo que se utiliza una función llamada de *entropía cruzada.*

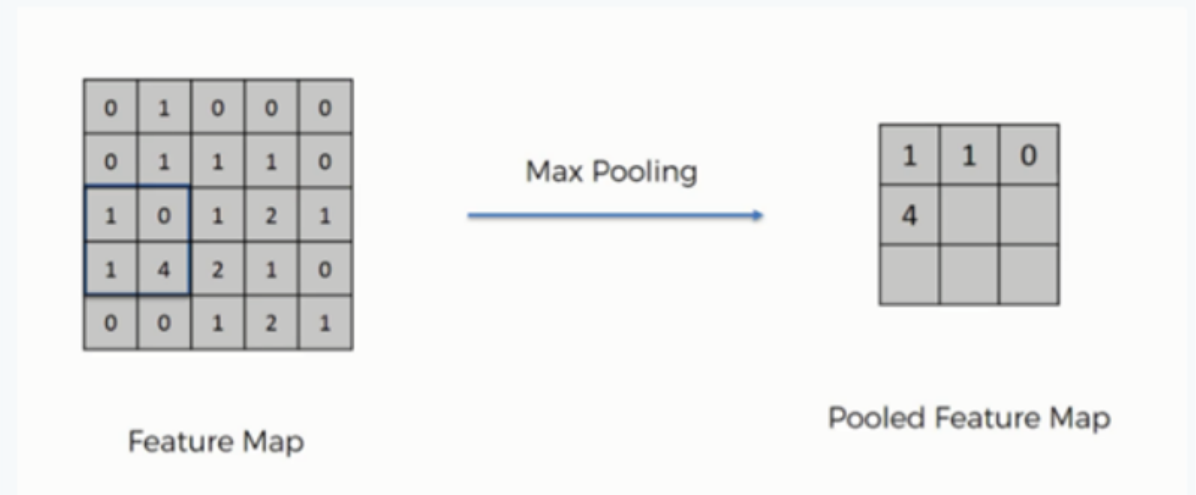


Figura 7. Aplicación de agrupamiento. Se reduce la dimensión escogiendo el máximo valor dentro de la matriz 2x2 que recorre la imagen.

Motivación, objetivos y planificación.

Las redes de convolución como medio para la clasificación de radiografías de tórax han experimentado un auge reciente como resultado de la última pandemia del virus causante del síndrome respiratorio SARS-CoV-2, cuya confirmación se basa mayoritariamente en métodos bioquímicos (PCR). Se prefiere la imagen simple sobre los CT por su bajo coste y facilidad de implementación a costa de la precisión en el diagnóstico.

Sin embargo, cuando se inicia un estudio de clasificación de este estilo sobre una base de datos de naturaleza médica aparecen una serie de dificultades específicas que pudieran no surgir cuando la base de datos es de otra especie. La primera y fundamental es la escasez de ejemplos, y de ellos, la escasez de imágenes que habitualmente conforman. Esta es una queja generalizada en varios de los estudios consultados [5],[6]. Los lugares que alojan imágenes médicas asocian dichas imágenes a estudios concretos con alguna finalidad alejada de la clasificación, y son pocos los estudios sobre redes neuronales de clasificación que permiten el acceso a la base de datos utilizada. Cuando por fin se encuentra una base de datos propicia, es habitual que el conjunto de imágenes no se encuentre correctamente catalogada y preparada para el entrenamiento de redes.

El proyecto de diseño y análisis de radiografías de tórax, llevado a cabo por la *National Library of Medicine,* perteneciente al *National Institutes of Health de EE. UU.*, por el *departamento de Radiología y ciencias de la imagen*[1], iniciado en 2017, sin embargo, es un buen ejemplo del que partir [7], por una serie de razones:

1. La base de datos de dicho proyecto, denominada *ChestX-ray8,* se ha realizado a partir de los diagnósticos médicos presentes en los archivos digitales de numerosos hospitales americanos y se encuentra clasificada previamente por patologías
2. Los autores de dicho trabajo han dejado escrito que ceden el uso de la base de datos para la investigación.
3. Contiene un número suficiente de imágenes ordenadas en carpetas por patologías como para iniciar estudios de clasificación de CNN significativos.

No obstante, a pesar de reunir estas cualidades esenciales para el tratamiento por una red de convolución, ya que reúne un número suficiente de casos y están adecuadamente catalogados, creemos que se puede optimizar la misma para mejorar los resultados de la clasificación.

[1]https://nihcc.app.box.com/v/ChestXray-NIHCC/file/256057377774

Por lo tanto, en este trabajo se plantean los siguientes objetivos:

1. Elaborar una red neuronal de convolución con los elementos mínimos necesarios que desarrollen los algoritmos que permitan demostrar los mecanismos que se van a a probar para optimizar la base de datos. La CNN que se elabore debe ser sencilla, con los elementos indispensables para que realice su función, pero lo suficientemente ligera como para poder realizar las pruebas con eficacia con la tecnología mínima con la que se cuenta.
2. Optimizar la base de datos introduciendo mejoras en la clasificación de las patologías. Las patologías estás clasificadas, pero dicha clasificación atiende exclusivamente a criterios médicos asociados al historial del paciente y no visuales. Se piensa que es posible refinar tal clasificación atiendo a características que puedan extraerse únicamente de la imagen. Igualmente se centrará análisis en una sola patología, la atelectasia. Siendo esta el punto de partida de una subclasificación posterior en ramas de tipos de atelectasia. Se cree que al definir de forma más precisa la tipología de la enfermedad, los criterios visuales en las que se establece la clasificación se simplifican, mejorando los resultados de esta.
3. Optimizar la base de datos por medio de algoritmos que mejoren el entrenamiento de la red. La base de datos se subdivide en datos de entrenamiento y validación de forma estándar, normalmente 80%-20%, antes de comenzar el proceso de forma aleatoria, sin atender a ningún criterio de conveniencia, únicamente en base a un porcentaje de división. Se piensa que esta división se puede efectuar durante el propio proceso para optimizar la fase de entrenamiento consultando previamente que imágenes son más difíciles de clasificar e incluyendo en el entrenamiento a dichas imágenes.
4. Optimizar la base de datos utilizando el mecanismo anterior simultáneamente a un proceso de *data augmentation* con preprocesamiento previo de la imagen para aumentar el tamaño de los datos de entrenamiento. Para ello se establecerá qué tipo de procesamiento previo es el que más encaja en este tipo de clasificación. Se analizará que tipo de transformación permite mejorar los datos de validación y por qué.

El objetivo es introducir los mecanismos de aprendizaje habituales en el cerebro humano para aprender de manera efectiva: aprendizaje significativo y constructivismo. Esto es, aprender a partir de lo aprendido. Que la red, en este caso, vaya construyendo la semántica de la clase de objetos que está clasificando. Para sello, deberán de tener todos los elementos de la base de datos la misma representatividad, como para que la red en el entrenamiento establezca el valor de los parámetros de forma precisa. El tratamiento de las imágenes de mala predicción requiere que se tenga presente el valor predictivo, que habrá que ir calculando durante el proceso.

Igualmente, si este es demasiado malo, a lo mejor conviene desecharlo para no alterar demasiado la predicción del resto de imágenes, y no desplazar los pesos demasiado.

Las fases del proyecto, una vez localizada la base de datos de trabajo, serán las siguientes:

1. Una fase introductoria de documentación en dos partes. Por un lado, se revisarán manuales de radiología de tórax para comprender los criterios visuales a los que recurren los facultativos para establecer diagnósticos sobre esta patología pulmonar en este formato. Para ello, se realizará un estudio somero de la anatomía general de la zona radiada. Por otra parte, se estudiará la herramienta informática utilizada en este estudio, las redes de convolución. En particular, redes de convolución para el reconocimiento de imágenes, y dentro de este gran subgrupo, aquellas con finalidad clínica.

2. Decidir la tecnología a utilizar: lenguaje de programación y entorno de ejecución. No se trata de un trabajo de investigación sobre programación, sino un trabajo que recurre a la programación para demostrar una serie de hipótesis de naturaleza médica. Por ello se buscará aquel recurso que contenga la suficiente flexibilidad como para introducir las cuestiones que a probar, pero que permita el uso de funciones de programación ya diseñadas para otros fines.

3. Estudio del estado de la cuestión, diseño y resultados. Lectura de aquellos artículos que utilizan redes de convolución para clasificación de imágenes médicas de radiografías de tórax, que además introduzcan alguna particularidad encaminada al proceso de optimización. En particular aquellos estudios que basen esta mejora en el tratamiento de los ejemplos de entrenamiento, y no en la mejora de la propia red, que dejaremos para trabajos futuros. Se diseñará la red de convolución con las características enunciadas y los programas que pongan a prueba los objetivos del proyecto, así como la forma en la que evaluar los resultados de la red.

Estado de la cuestión

Para el diseño de bases de datos, la tendencia más arraigada, es la de utilizar los textos facultativos que acompañan las imágenes en archivos DICOM [2] [7], y obtener, a partir de algoritmos, segmentaciones, detecciones de frontera, análisis sintácticos y semánticos, negaciones y otros procedimientos sobre los textos, una clasificación de los diagnósticos. A partir de aquí, se desarrolla una correspondiente catalogación de sus imágenes correspondientes que poder etiquetar con patologías de manera oportuna. Para tal objetivo, uno de los procesos usados es el denominado *NLP, Natural Language Processing*. En última medida, no es sino el propio radiólogo, desde su conocimiento y experiencia, el que categoriza cada radiografía, y no el algoritmo a partir de cierta característica concreta de la propia imagen. La razón principal es la de que el diagnóstico no depende exclusivamente de la imagen, si no, además, de la historia clínica del paciente. Esto limita, a su vez, el alcance de la herramienta digital de reconocimiento basada exclusivamente en la imagen.

Por otra parte, el uso de redes neuronales para el reconocimiento de imágenes médicas tiene ya cierta tradición en el ámbito de la denominada *Computer Aided Diagosis* (CAD). Y parece que el procedimiento habitual es el de partir de redes ya diseñadas, bien a partir de otras redes conocidas para la clasificación general de imágenes, como AlexNet[3], o GoogleLeNet[4], o redes para clasificación de imágenes médicas asociadas a las radiografías de tórax, como Resnet50 [8]. A partir de dichas CNN, las propuestas habituales son, bien partir desde cero integrando y refinando la arquitectura de la red, bien heredar la configuración de pesos y parámetros de una red ya entrenada con una base de datos abundante pero diferente a la del problema, lo que se conoce como *Transfer Learning*.

Las redes neuronales de éxito que hoy en día tienen valores predictivos del orden del 99% están basado en conjuntos de imágenes claramente diferenciables, tal es el caso de *VGG16* con la base de imágenes *ImageNet* [9].

Son varios los estudios que han combinado el preprocesamiento de las imágenes con la aplicación de la CNN [10] para la mejora de resultados. Se demuestra que, si sometemos al conjunto de imágenes a un tratamiento previo que mejore la ROI, o el etiquetado de catalogación del conjunto de datos, mejoran los parámetros característicos que nos da la métrica de la clasificación.

---

[2] DICOM (**D**igital **I**maging and **Co**mmunication **I**n **M**edicine) es un estándar de transmisión de imágenes médicas en las que las imágenes se acompañan de la información médica correspondiente.

[3]https://en.wikipedia.org/wiki/AlexNet

[4]https://papers.withcode.com/method/googlenet

Otros trabajos [11], realizan diferentes experimentos con una serie de tratamientos previos de las imágenes. Así distinguen entre imágenes segmentadas, centrado en región de interés e imágenes recortadas con las que nutrir la red, con resultados dispares con las que conformar una interpretación del comportamiento de la herramienta, para poder dirigir los esfuerzos al interés clínico.

Estudios recientes con radiografías de tórax que buscan este objetivo [12], demuestran que transformaciones tales como pequeñas rotaciones, zoom zonal, cambios en la iluminación de la imagen o deformaciones, mejoran generalmente los resultados.

Las precisiones de los estudios más relevantes dedicados a la detección automática de neumonías derivadas del COVID-19 rondan el 90% con redes de eficacia contrastada en amplias bases de datos de radiografías de tórax. Sirva como ejemplo la arquitectura COVID-net con entrenamiento de 13.000 imágenes. El procedimiento habitual en estos estudios es el de multiplicar el número de ejemplos con afecciones con *data-augmentation* para equilibrar el entrenamiento y evitar sesgos [13]. Para la evaluación de la propia red, las medidas más habituales en los estudios citados [12,13,14], y que utilizaremos en este trabajo, son la precisión, el área bajo la curva (ROC) o la matriz de confusión.

Tecnologías utilizadas

Para decidir qué recursos tecnológicos se van a adaptar mejor a los objetivos que persigue este trabajo hay que estudiar aspectos de los datos tales como la tipología y la cantidad de estos, así como la capacidad de procesamiento de nuestro entorno de ejecución, que va a permitir realizar los ciclos de entrenamiento en tiempos asequibles dentro del periodo de trabajo en el que vamos a desarrollar nuestro proyecto.

Por las circunstancias en las que se ha programado, pequeños terminales y bases de datos discretas, se ha centrado en algoritmos y procedimientos asociados con el mundo *Machine Learning* más comúnmente utilizados, que son los que describimos a continuación.

Lenguaje de programación

Los programas han sido codificados con Python sobre la API de bajo nivel TensorFlow, con la API Keras de alto nivel para el uso de funcionalidades especialmente creadas para las aplicaciones de Deep Learning por Google en 2017. La combinación de Keras con TensorFlow permite la construcción de redes de forma sencilla, desde el diseño de las capas hasta las especificaciones de los parámetros de compilación y entrenamiento: optimizadores, activaciones... Es decir, nos facilita un acceso suficientemente fino a las especificaciones de la red sin necesidad de codificar desde el principio. No es deseable una codificación desde cero porque lo que se pretende en este trabajo no es en sí investigar sobre la propia estructura de la red, sino más bien estudiar el efecto de la optimización de la base de datos en los resultados de una red neuronal estándar de clasificación.

TensorFlow ejecuta líneas de flujo consistentes en operaciones matemáticas sobre elementos de entrada en forma de tensores, que son distribuciones multidimensionales de números que representan un ente determinado. En nuestro caso, imágenes médicas de dimensiones 1.024 x 1.024 píxeles que en seguida redimensionamos para su uso optimizado por la red en entidades de tipo 200 x 200. Dichas líneas de flujo se suman en nudos que conforman la topología de la red. Keras, por otra parte, está diseñado para su uso en redes neuronales, y articula lenguaje de alto nivel con funcionalidades creadas en formato TensorFlow para un uso más amable.

Se puede ilustrar este aspecto con ejemplos: las librerías de Keras, del tipo *Keras.models*, o *Keras.layers*, agrupan un conjunto de instrucciones sobre TensorFlow que obtienen la funcionalidad necesaria para configurar en este caso de forma directa el modelo de compilación de la red, por un lado, o las especificaciones de la capa de convolución con la introducción de los parámetros pertinentes, por otro.

Además, Python incorpora librerías especializadas que se han usado para:

- El tratamiento rápido y creación de arrays multidimensionales de imágenes (*numpy*).
- La visualización de imágenes y gráficos (*matplotlib*).
- Conectar con las funcionalidades del sistema operativo (*os*).
- Edición de imágenes (*pil*)
- Uso de algoritmos de clasificación y de evaluación *(sklearn).*
- Uso de estructuras de datos (*pandas*).

Entorno de ejecución

Para la compilación, el entrenamiento y la evaluación de resultados de la red se ha recurrido a la utilidad de *Google Colaborativo* en su versión *Pro*, que proporciona varios entornos de ejecución: CPU (*central process unit),* GPU (*graphical process unit)* y TPU (*tensor process unit).* Para el peso de nuestros programas, que direccionan bases de datos de pocos elementos, y tienen relativamente pocos parámetros que editar en la fase de entrenamiento, ha sido suficiente trabajar con el entorno de ejecución *GPU*. Dicho entorno, similar al que proporciona una tarjeta gráfica, es comúnmente utilizado en la ejecución de CNN con el lenguaje de programación antes descrito. El tipo de *GPU* utilizada en cada ejecución no es en principio suministrada al usuario, que queda a expensas de cómo el entorno administra los recursos entre sus clientes. Por el tipo de red utilizada y el peso de la base de datos, el ritmo de procesamiento y la memoria ram asociada han sido suficientemente ágil como para permitir las pruebas que exigían los objetivos del proyecto.

En concreto, un procesador Intel(R) Xeon(R) CPU @ 2.30GHz, con una memoria RAM asociada durante la ejecución de 37,8 Gigabytes. Para el uso de este entorno de ejecución externo, se ha tenido que subir el conjunto de imágenes de trabajo a la nube de Google.

Diseño

5.1 Diseño de la CNN para la clasificación.

a. Estructura.

La red consta de los siguientes elementos, que trataremos de justificar paso por paso. La imagen de partida ha sido escalada a un tamaño de 200x200[5] píxeles. El número de capas de convolución va a determinar el número de parámetros actualizables. El criterio para determinar su número es práctico y se determina por ensayo y error. Se pretende una red ágil que demuestre las hipótesis de partida sobre refinamiento y procesamiento de la base de datos de imágenes. Si se fija una red demasiado profunda, con un número de capas, y nodos por capas, elevado, quizá mejoremos los resultados de la red por sí misma, pero ralentizaría demasiado el proceso de prueba. Por otra parte, los estudios consultados [12], [13],[14], usan como mínimo 3 niveles de convolución. Esta profundidad de convolución, de tres pisos, en base a nuestras pruebas, realiza de forma óptima el proceso sin ralentizar el mismo y arroja resultados de precisión aceptables. El resto de los parámetros editables, tamaño de los *kernels, stride, padding, capa densa,* entre otros, también han sido ensayados con distintos valores y no resultan en mejoras significativas en la precisión de la red. Como se observa en la tabla 3, la estructura es la siguiente:

- 1ª capa de convolución de 32 nodos con filtros de tamaño 3x3. Genera 32 matrices de salida de 198x198 píxeles. Para calcular el tamaño de la salida, se ha utilizado la fórmula,

$$((W - K + 2P)/\mathrm{S} + 1)$$

Con:

a. W el tamaño de la imagen de entrada, 200.

b. K el tamaño del filtro o *kernel*, 3. El tamaño del *kernel* es ajustable. Se ha probado a cambiar su tamaño a 5, 7 y 9. El tipo de *kernel* utilizado y su tamaño, así como su incidencia en la exactitud de la red, es un estándar de las CNN.

c. P el *padding,* es decir, el número de bordes añadidos a la matriz a la hora de aplicar la convolución. En nuestro caso, 0. Añadir *padding* a la imagen original puede aportar exactitud a la hora de aplicar el *kernel*, pues añade la información de bordes de la imagen. En nuestro

[5] El tamaño de la imagen incide en la cantidad de parámetros ajustables. Si aplicamos el tamaño original de 1024x1024, el procesamiento se hace imposible con los recursos que disponemos. Hemos probado a cambiar el tamaño de la imagen en el entorno de ese tamaño y la precisión de la red no cambia.

caso, como los pulmones están en el centro de ésta, no tiene sentido añadir nada.

d. *S* el *stride*, esto es, el número de píxeles que se traslada el filtro cada vez que lo aplicamos por la matriz imagen. Si vamos píxel a píxel, el *stride* es 1. Si avanzamos 2 píxeles, aplicamos

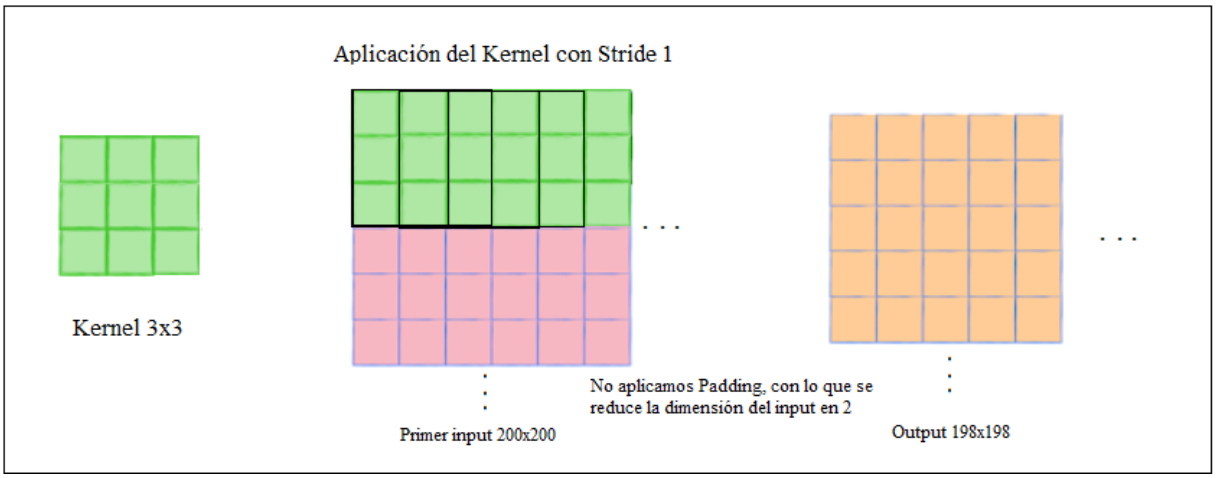


Figura 8. Aplicación del *kernel* 3x3 a la imagen de entrada con *stride* 1 y no *padding* (borde)

la mitad de las veces, de ahí que haya que dividir entre *S* el tamaño de la salida. Cuanto mayor es *S*, menor es la salida. Esto derivará en menos parámetros ajustables, pero también limitará la exactitud.

Con todo ello, el tamaño de la salida es de (198,198,32), como se observa en la Figura 8.

- 1ª capa "*maxpooling*" que reduce la dimensión de cada una de las 32 matrices de salida de la capa anterior, a la mitad, a partir del píxel de mayor valor de una matriz 2x2 que recorre la imagen con un *stride* de 2. No tiene un efecto apreciable en bordes. Podría ocurrir que dicha matriz 2x2 recorriera un borde y se quedara con el valor alto, despreciando la textura vecina. Como el filtro se desplaza a la textura vecina, ésta quedaría representada en el siguiente movimiento. El algoritmo de reducción, en este caso, tampoco es significativo. Tamaño de la salida (99,99,32).
- 2ª capa de convolución de 32 nodos con *kernel* de tamaño 3x3. Genera 32 matrices de salida de 97x97. Tamaño de salida (97,97,32).
- 2ª capa *maxpooling* con salida (48,48,32).
- 3ª capa de convolución de 64 nodos con *kernel* de 3x3. Genera 64 matrices de salida de 46x46. Tamaño de salida (46,46,64).
- 3ª capa *maxpooling* con salida (23,23,64).
- Capa *flatten*, que prepara la entrada a la capa densa, disponiendo todos los píxeles en una matriz vertical unidad para poder ser procesados por la red. Tamaño de la salida

23x23x64=33856.

- Primera capa densa de 64 nodos con 33856 entradas cada una, con 64 salidas a la segunda capa densa.
- Segunda capa densa, con 1 nodo, 64 entradas, y una salida.

Todos estos datos quedan resumidos en la Tabla 3.

b. Cálculo de parámetros editables.

Los parámetros son los pesos que multiplican a las entradas en los nodos en el proceso de entrenamiento para ajustar a la salida deseada minimizando el error predictivo de la red. Las redes CNN potencian el proceso de extracción de características de la imagen reduciendo el número de pesos editables.

- Cálculo de parámetros ajustables en las capas de convolución:

La gran diferencia entre las capas de convolución y las capas densas es que en el primer caso se actualizan los parámetros de los *kernels* con tantas capas como entradas (convolución "multicanal"), para generar tantas salidas como número de *kernels*, reduciéndose considerablemente el número de parámetros a actualizar en el aprendizaje. Así pues, en nuestro caso, como se observa en la Tabla 3:

a. 1ª red de convolución: 32 filtros de 3x3 (hay que multiplicar por 3 el número de inputs para ajustar al paquete de programación del *keras* "CONV2D", preparado para imágenes en color con tres canales. En total 3x9x32 + 32 = 896 parámetros para generar 32 imágenes filtradas.
b. 2ª red de convolución: le llegan 32 imágenes. Se actualizarán 32 *kernels* de 9 elementos de 32 canales de profundidad por cada una de las 32 imágenes de entrada. A cada uno de los 32 filtros hay que sumarle el sesgo. Así pues: 32x9x32+32=9.248 parámetros para generar 32 imágenes filtradas.
c. 3ª red de convolución: le llegan 32 imágenes. A cada imagen se le aplica el procedimiento anterior, pero esta vez con 64 filtros. Así pues: 32x9x64+64=18.496 parámetros para generar 64 imágenes filtradas.

- Cálculo de parámetros ajustables en las capas densas:
    a. 1ª capa densa: Los parámetros editables son tantos como el producto de los nodos que forman la capa, por el número de entradas a cada nodo, más los sesgos, que son tantos como nodos. En este caso 64x33.856 + 64= 2.166.848 parámetros.
    b. 2ª capa densa: 1 nodo con 64 entradas y un sesgo, total 65 parámetros editables.

c. Activación de los nodos.

Distinguimos las capas de convolución de las capas densamente conectadas. El concepto de nodo es sutilmente diferente en uno u otro tipo de capas. Un nodo, por establecer una definición general, es aquella parte de la red que será activada en función de una entrada operada por unos pesos editables, transmitiendo un estímulo o no. En el caso de una capa densa de perceptrones, está claro el concepto. El nodo es el lugar donde confluyen las sumas de las salidas de la capa anterior. En el caso de una capa de convolución, los nodos están constituidos por los filtros, entendidos estos con toda su profundidad multicanal, que son la parte de la red editable en esta primera sección. Entonces tenemos:

a. Capas de convolución: con activación ReLU, descrita de la siguiente manera:

   ReLU(convolución(y)),

   donde *y* es la imagen de salida de la red tras aplicar el filtro a la imagen de entrada. La activación de la imagen pretende devolver la no linealidad en la imagen tras el suavizado de la convolución. La función ReLU devuelve "0" hasta que la entrada supera cierto umbral. De esta forma, la salida no sigue linealmente a la entrada sino a partir de ese número, rompiendo la proporcionalidad.

b. Primera capa densa: activa aquellos nodos que superan en su entrada cierto umbral:

   ReLU(Ax + B),

   donde A y B representan los pesos y sesgos de la entrada de cada nodo. De esta forma, la red sólo activa ciertos nodos al paso de hacia delante de los datos, discriminando el resultado.

c. Nodo de salida de la red: transforma la entrada a un valor entre 0 y 1 según la llamada función *Sigmoide.* Los valores altos tienden de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a 0.

d. Agrupamiento (explicación del *pooling)*

El objetivo del agrupamiento es doble. Por un lado, minimiza el peso de la imagen al rescatar los valores esenciales, por otro lado, ayuda a combatir el sobreajuste a los datos de entrada eliminando complejidad de la imagen. Si la imagen es demasiado compleja, la red aprenderá a reconocerla, y esto supondrá perder generalidad. La mayoría de las imágenes que queremos clasificar como de este tipo, no tienen este exceso de complejidad específica de esta imagen en particular.

El tipo de agrupamiento escogido en este algoritmo es el de valor máximo. Este tipo de agrupamiento mantiene las divergencias altas entre píxeles de la imagen, específica de bordes.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 198, 198, 32)	896
activation (Activation)	(None, 198, 198, 32)	0
max_pooling2d (MaxPooling2D)	(None, 99, 99, 32)	0
conv2d_1 (Conv2D)	(None, 97, 97, 32)	9248
activation_1 (Activation)	(None, 97, 97, 32)	0
max_pooling2d_1 (MaxPooling2D)	(None, 48, 48, 32)	0
conv2d_2 (Conv2D)	(None, 46, 46, 64)	18496
activation_2 (Activation)	(None, 46, 46, 64)	0
max_pooling2d_2 (MaxPooling2D)	(None, 23, 23, 64)	0
flatten (Flatten)	(None, 33856)	0
dense (Dense)	(None, 64)	2166848
activation_3 (Activation)	(None, 64)	0
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 1)	65
activation_4 (Activation)	(None, 1)	0

Tabla 3. Cuadro resumen de la estructura de la red CNN utilizada en este estudio inicial

e. Entrenamiento.

- Función de coste (de error): al tratarse de una red de clasificación binaria (atelectasias vs normal), la función de coste recomendada es la llamada de *entropía cruzada.* Esta función de coste (C), de definición:

$$C = -t \cdot \log(y),$$

donde *t* es el valor etiquetado, e *y*, el valor obtenido por la red tras aplicar la función *softmax*[6], tiene un

[6] La función *softmax* convierte un valor en un valor de formato probabilístico.

gradiente:

$$\frac{\partial C}{\partial z} = y - t,$$

que es grande para valores de *t* e *y* muy diferentes, lo que acelera la optimización de la red, siendo C la función de coste.

- Optimizador: El objetivo del entrenamiento de la red, es minimizar la función de coste con los valores de los parámetros editables necesarios, procurando además una buena generalización. El algoritmo que se aplica es el conocido como *retropropagación* basado en el descenso del gradiente de la función de error. Son varios los optimizadores utilizados, y su conveniencia depende del tipo de red y su cometido. Para una red clasificadora binaria, los optimizadores que muestran mejor convergencia son los tipos RMSProp y ADAM, que son los utilizados en este ejemplo.

5.2 Optimización de la base de datos Chest-Ray8 para atelectasias.

El uso de redes neuronales para el reconocimiento de imágenes médicas basadas en radiografías depende de una construcción de base de datos suficientemente abundante como para que se alimente la red con suficiente rigor.

El proyecto de diseño y análisis de radiografías de tórax, llevado a cabo por la *National Library of Medicine,* perteneciente al *National Institutes of Health*, por el *departamento de Radiología y ciencias de la imagen,* iniciado en 2017 [7], es el punto de partida, como queda dicho (Sección 2). La base de datos de dicho proyecto, denominada *ChestX-ray8,* se ha realizado a partir de los diagnósticos médicos presentes en los archivos digitales de numerosos hospitales americanos. Consiste en 108.948 radiografías agrupadas en 8 distintas patologías tomadas a un conjunto de 32.717 pacientes afectados de patologías pulmonares.

- Atelectasia, cardiomegalia, efusión, infiltración, masas, nódulos, neumonía y neumotórax.

De ellas, prácticamente la mitad son radiografías "normales", sin patología registrada. El resto se agrupa de forma diversa tal y como podemos ver en la tabla 2.

Atelectasia	11.349
Cardiomegalia	2.403
Efusión	7.992
Infiltración	11.786
Masas	2.924
Nódulos	3.006
Pneumonía	325
Pneumotórax	2.200

Tabla 2. Distribución de imágenes de patologías en la base de datos Chest-Xray8

De ello sacamos dos conclusiones iniciales:

- El desequilibrio en el número de imágenes utilizadas en función de cada patología, recordando que, para un entrenamiento correcto de clasificación, se requiere una entrada equilibrada de datos, para no sobrentrenar la red en un tipo específico.
- La mezcla entre patologías y patrones patológicos (pneumonía frente a infiltraciones).
- Los diagnósticos utilizados no son unívocos. Gran parte de ellos implican varias patologías (por ejemplo: efusión + infiltración...).
- La carpeta *No findings* recoge todas aquellas radiografías que el radiólogo en su informe concluyó no contenían afección ninguna. Sin embargo, si analizamos detalladamente la carpeta, comprobamos que existen imágenes de pulmones sanos de gran opacidad o incluso sin silueta característica definida. Deducimos que estas imágenes corresponden a casos en los que la toma no ha sido realizada en la mejor de las condiciones, con la inspiración completada, pulmones completamente expandidos y aireados.
- La carpeta que más nos interesa, la carpeta de atelectasias mezcla de forma indiscriminada todo tipo de atelectasia.

Un rápido recorrido nos dice que existen a simple vista archivos cuya catalogación no responde a un criterio puramente visual. La experiencia del radiólogo y el conjunto de datos del paciente que acompañan el documento han sido los elementos clave en el diagnóstico de atelectasia. Sin embargo, en una red cuyo objeto es la de clasificar de forma automática en base a los píxeles de la imagen, no tienen cabida este tipo de ejemplos. Es necesario, si queremos que la red modelice la función que estamos buscando, eliminar cosas tales como medias radiografías, o radiografías parciales, como demasiado poca penetración, y otras, que pueden ser objeto de divergencia de datos al modelo. Igualmente encontramos en la carpeta de atelectasias demasiada diversidad como para que la red pueda ajustar los datos. A falta de ellos, pues creemos que con el suficiente número la red aprendería a reconocerlos, debemos de tomar medidas subsidiarias, como la de especializar la catalogación. Una vez que se ha reclasificado la base de imágenes en imágenes “claramente reconocibles como patológicas”, y de una patología concreta, creemos que es buena idea estudiar aquellas imágenes que, a pesar de ofrecer visualmente una afección clara, la red no consigue clasificarla con éxito. Tras el análisis de estos casos, propondremos aumentar el peso específico de ellas en el conjunto de imágenes para que la red la incluya en su entrenamiento, si así pudiera ser útil para la mejora de la métrica de nuestra CNN.

Procederemos en orden según tres principios:

a. **Etiquetado de las imágenes:** Todas las imágenes etiquetadas como “normales”, y todas las etiquetadas como “atelectasias”, no lo son de manera evidente. Y si no lo son de manera evidente para un ojo, claro está, poco entrenado como el nuestro, tampoco será fácil para la red establecer un modelo que las clasifique de forma automática. Recordemos que el proceso de construcción de la base de datos *Chest Xray8*, cuya documentación y estudios anexos son públicos y disponibles para su comprobación [7], ha sido a partir de los expedientes médicos de los pacientes. En uno y en otro caso, dichos expedientes contienen una información vital para el patólogo que en última medida los animan a inclinarse por un diagnóstico. Esto a veces lleva al hecho de que muchas de las imágenes que responden a pulmones con atelectasias, no lo son de forma evidente, pero si acaban de serlo para un ojo adiestrado con el respaldo de un informe médico. A falta de un criterio médico que respalde nuestro trabajo, comenzaremos entonces clasificando dichas imágenes en subcarpetas con patologías más evidentes desde nuestro punto de vista, y las someteremos a entrenamiento, ya que el objetivo de este trabajo no es el de profundizar en el diagnóstico de atelectasias a partir de sus radiografías, si no el de mejorar los mecanismos informáticos de clasificación de imágenes con esta afección. Por otra parte, los resultados obtenidos en el apartado anterior constituyen el resultado de una red clasificatoria binaria con sólo dos posibles resultados: atelectasia o pulmones sin patologías observables. Una rápida prospección que enfrenta a dichos pulmones normales con una selección de colapsos totales de pulmón sea izquierdo o derecho, arroja resultados mucho mejores. Esto nos lleva a formular la hipótesis de si el hecho de profundizar en la forma de clasificar la imagen con la que entrenar la red, mejora el resultado global de clasificación de atelectasia frente a pulmones normales.

b. **Preprocesado de las imágenes:** Estudios recientes [15] demuestran que, modificando las propiedades de la imagen previo a la entrada en la red, en particular ecualizando el histograma y aplicando filtros, mejoran ostensiblemente los resultados de la medida de la red clasificatoria. Para ello será necesario aplicar alguna técnica de procesamiento de imagen que automatice la escritura de un nuevo archivo de imágenes tratadas con la que alimentar de nuevo la CNN.

c. **Aumentando el número de ejemplos de la base de datos**: Añadiendo especialmente aquellas que clasifican mal con el proceso de *data augmentation*. Nos centraremos especialmente en algunas imágenes que obtienen un mal resultado en las clasificaciones anteriores. Como se trata de un conjunto de imágenes minoritaria, someteremos a este conjunto reducido de imágenes a un proceso de multiplicación que equilibre su presencia en el conjunto de datos, y así favorecer su sesgo final.

5.2.1 Refinamiento de la catalogación:

Clasificación visual de los signos radiológicos.

La base de datos proporciona un conjunto etiquetado de imágenes de atelectasias, pero no proporciona ningún detalle del tipo de obstrucción u otra afección causante. Es por ello, que no parece mala idea hacer una búsqueda de ciertas propiedades en la imagen que confirmen el diagnóstico, como por ejemplo un broncograma aéreo comprimido perceptible, una opacidad homogénea acompañada de un desplazamiento de volumen, o un truncamiento perceptible de un bronquio secundario en el seno del parénquima pulmonar, que generen una alteración en el tamaño del pulmón, o una alteración de las siluetas características. De todas las atelectasias, vamos a realizar una selección de patologías claras para el entrenamiento y testeo de la red. Insisto, habría que remarcar que la imagen no ha sido la única referencia del patólogo en el establecimiento del diagnóstico, pero si debe ser la única referencia con la que cuente el algoritmo. Es por ello por lo que se hace necesaria esta clasificación previa.

Extracción de características visuales que ayuden en la clasificación práctica

En la Figura 9 tenemos ejemplos de atelectasias lobulares. Un ojo poco entrenado podrá encontrar rápidamente las características que definan uno u otro tipo de atelectasia si es acusada.

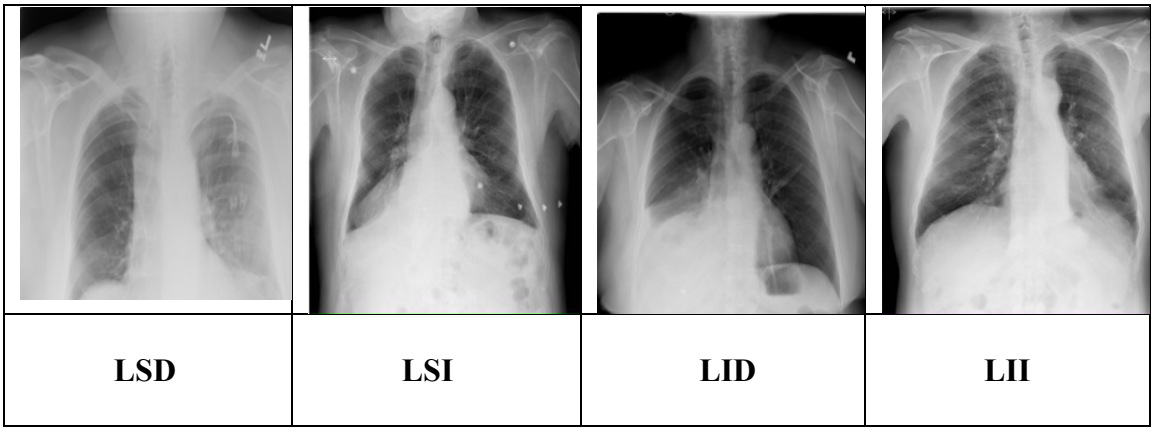


Figura 9. Cuadro resumen de los signos radiográficos de atelectasia lobular, LSD,LSI,LID,LII (Fuente: Chest Xray8)

(LSD: Lóbulo superior derecho; LSI Lóbulo superior izquierdo; LID Lóbulo inferior derecho, LII Lóbulo inferior izquierdo)

En este apartado, vamos a tratar de enfocar cada una de las atelectasias posibles, y extraer visualmente las características más definitorias. Este trabajo no puede ser tan exhaustivo como debiera ser. Para que lo fuera, se requeriría del apoyo de radiólogos experimentados, o de una base

de datos debidamente clasificada. Sin embargo, ello se podría obviar, ya que en este trabajo únicamente se pretende demostrar como el refinamiento en la etiquetación de las imágenes mejora la clasificación de estas en redes de entrenamiento.

En la Figura 9 se presentan ejemplos evidentes de atelectasias lobulares. Si es un colapso del lóbulo superior derecho (LSD), se observa un espacio de radiopacidad disminuida en ese pulmón en dirección vertical paralela al mediastino. Si es de lóbulo inferior derecha (LID), tenemos ese pulmón muy contraído y la línea del diafragma desdibujada y muy superior. Si es de lóbulo inferior izquierda (LII) es colapso queda por detrás del espacio del corazón, por lo que se percibe el broncograma aéreo. Otras atelectasias serán especialmente reconocibles, como aquellas que suponen un colapso total de uno de los pulmones.

Procedimiento

Se ha comenzado clasificando aquellas atelectasias más recurrentes y fácilmente identificables, y se deja para trabajos futuros un estudio más exhaustivo con los casos más complejos. Concretamente, se abordarán las atelectasias producidas por colapso total, como consecuencia de un carcinoma broncogénico intrínseco o extrínseco. Se reconocen con claridad por la falta completa de uno de los dos pulmones. En segundo lugar, se abordan aquellas radiografías que hayan perdido el borde cardiaco izquierdo y el hilio, propio de las LSI. Finalmente, se cubren aquellas LID motivadas por afectación de bronquio intermediario.. Una vez que tengamos clasificadas así las imágenes, se procede a su clasificación con la CNN diseñada para su evaluación.

5.2.2 *Data augmentation* con imágenes error

El objetivo de este apartado es probar si al enriquecer la base de datos de entrenamiento con imágenes que clasifican mal en la validación se consigue aumentar la precisión.

El primer paso ha consistido en una búsqueda de aquellas imágenes que no obtienen buena calificación en las operaciones realizadas hasta ahora. Una vez seleccionadas, se han sometido a un proceso de aumentación y luego se han añadido al conjunto de datos que alimenta la trama de nodos.

El proceso de incremento del número de ejemplos en la base de datos para mejorar los valores de exactitud de la red es un recurso contrastado que adquiere especial validez cuando las imágenes no son fáciles de obtener [15]. Las técnicas habituales incluyen rotación, espejo, escalado, recorte y otros, siempre que el contenido semántico de la imagen permanezca inalterado.

El equilibrado de la base de datos con imágenes que dan mala calificación en la clasificación será en cantidad inversa a su valor de red obtenido. Cuanto más lejos de su etiqueta, mayor será el enriquecimiento de la base de datos con imágenes obtenidas a partir de ella por transformaciones. La idea última es modificar el aprendizaje de la red para que admita estas imágenes. Se pretende ver si la adaptación de la red a las nuevas imágenes generadas altera de alguna forma el aprendizaje previo, modificando los resultados obtenidos.

Análisis de “imágenes error”

En este apartado se trabaja con aquellas imágenes mal clasificadas por los algoritmos de clasificación de la CNN. Se analizan las posibles causas por las que el algoritmo de aprendizaje arroja un mal resultado, y se detectan aquellas imágenes que peor resultado obtienen en la red de clasificación. Con ellas se vuelve a reentrenar la red después de aumentar su número y eventualmente procesarlas.

Se comienza con los falsos negativos, esto es, por ejemplo, imágenes etiquetadas como afectadas por atelectasia LID con etiqueta “1”, pero que son clasificadas como imágenes normales, con etiqueta en torno a “0” (ver Figura 10), y luego se procede a la inversa, analizando los falsos positivos, imágenes normales, pero la red interpreta como pulmones con patología.

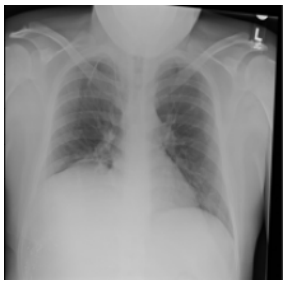
		
0,0005	0,01	0,01

Figura 10. Falsos negativos y sus valores predictivos. En realidad, son pulmones con patología que la red clasifica como normales

Una vez identificados, se procede a incorporar los datos generados a los entrenamientos de forma inversa a su valor predictivo.

Automatización del proceso

Realizar una búsqueda de falsos negativos y falsos positivos de forma manual revisando su valor predictivo imagen por imagen es una tarea ardua. Se puede automatizar el proceso antes descrito y además hacerlo en base al mejor umbral de catalogación, el ideal. Es decir, que en cada fase del entrenamiento se decida qué umbral es el que menos falsos positivos y falsos negativos produzca, y en base a él, obtener las imágenes "erróneas" para enriquecer la base de datos. De esta forma, la próxima vez que entrenemos a la red, tendrá en cuenta estas nuevas adquisiciones.

Por mejor umbral se entiende aquel valor predictivo que maximiza sensibilidad y especificidad, es decir, que hace máximo el número de verdaderos positivos, por encima de dicho umbral, y de verdaderos negativos, por debajo de dicho umbral. También puede definirse como aquel que minimice el número de falsos positivos y negativos. Toda vez que la red pasa por un ciclo de entrenamiento, genera un modelo predictivo cuyos valores representativos mencionados anteriormente pueden oscilar. Una selección al azar que no responda a ningún modelo predictivo generaría estadísticamente los mismos valores verdaderos que falsos.

Es decir, el algoritmo puede realizar el proceso de entrenamiento con inclusión de errores de forma progresiva hasta que maximice su exactitud minimizando su error de validación. En dicho proceso es importante que se tenga en cuenta que el umbral de clasificación ideal, es decir, el valor del umbral que maximiza la suma de valores verdaderos positivos y negativos, va a ir cambiando cada vez, y habrá que ir actualizando ese valor cada vez que se inicie el proceso de búsqueda de imágenes erróneas. El concepto de imagen errónea será cambiante a medida que se vaya aproximando al mejor modelo posible.

Para la búsqueda de dicho umbral ideal, el algoritmo realiza una consulta de todos los valores

verdaderos para cada umbral y toda vez que lo hace realiza la operación ya conocida como:

$$Verdaderos\ positivos - (1 - falsos\ positivos)$$

(ver Figura 11) La primera línea del código rellena un *dataframe* con el conjunto de valores de la operación antes descrita para todos los *thresolds* posibles, la segunda línea localiza el *thresold* que más acerca a cero dicha operación). Para la obtención de la serie de umbrales en términos del número de falsos valores, utilizamos la curva ROC. La curva ROC[7] representa los pares verdaderos positivos frente falsos positivos para diferentes valores del umbral de clasificación. Cuanto mayor sea el área bajo esta curva, mayor será el poder predictivo de la red.

```
roc = pd.DataFrame({'tf' : pd.Series(tpr-(1-fpr), index=i), 'thresholds' : pd.Series(thresholds, index=i)})
ideal_roc_thresh = roc.iloc[(roc.tf-0).abs().argsort()[:1]]  #Localiza el punto que haga mas próximo a 0 esta operación
print("Ideal threshold is: ", ideal_roc_thresh['thresholds'])
```

Figura 11. Algoritmo que localiza el umbral ideal a partir de la lista de umbrales que proporciona la curva ROC

Así pues, el algoritmo realiza los siguientes pasos (ver Figura 12, *X_test* contiene las imágenes a validar):

a. Encontrar los falsos negativos o falsos positivos: hace un recorrido por todas las imágenes del conjunto de imágenes de validación en búsqueda de imágenes de etiqueta "1", con atelectasia, pero clasificada como pulmones normales. Lo hace en base al mejor umbral de clasificación posible, que sea ha determinado con anterioridad.

b. Incrementar el conjunto de imágenes de entrenamiento con las imágenes obtenidas del apartado anterior, eliminándolas del grupo de validación.

c. Volver a entrenar la red.

d. Validar de nuevo.

e. Repetir el ciclo hasta que el proceso converja a un valor fijo o hasta que el grupo de imágenes de validación reduzca demasiado su número.

[7] Responden a las siglas en inglés de *Receiver Characteristic Operator.*

```
for n in range(len(X_test)):
    img = X_test[n]
    input_img = np.expand_dims(img, axis=0)
    if ((model.predict(input_img) > ideal_roc_thresh) and (y_test[n] == 0)):
        X_train_list.append(X_test[n])
        y_train_list.append(y_test[n])
        m = m+1

X_train = np.array(X_train_list)
y_train = np.array(y_train_list)
```

Figura 12. Algoritmo que incluye en la lista de imágenes de entrenamiento aquellas imágenes que clasifican mal según un umbral dinámico

El algoritmo de la Figura 12 se buscan las imágenes con valor predictivo mayor al umbral ideal, que es el que mejor clasifica, que además tiene etiqueta cero. Es decir, se buscan falsos positivos. En ocasiones, si el umbral ideal es muy bajo, resulta más lógico localizar los falsos negativos (serán menos), y hay que intercambiar el mayor por el menor, y el cero por un uno, cosa que de realizar manualmente en el progreso de aprendizaje y validación de la red.

La idea es incluir los errores lo antes posible en el proceso de aprendizaje para modelar con más generalización. El concepto es entonces un proceso dinámico de transferencia de aprendizaje. Toda vez que se actualice la base de datos con los elementos mal validados, se hereda el aprendizaje del proceso anterior y se aplica a la nueva base así modificada.

Preprocesamiento de las imágenes para *data-augmentation*

Como último paso, se han aplicado técnicas de transformación de la imagen como paso previo o intermedio dentro del proceso explicado en el apartado anterior, con el objeto no ya de enriquecer la base de entrenamiento con imágenes de "calidad para tal fin", sino que además se añadan otras derivadas de estas. El objetivo inicial es añadir no sólo el peso significativo que supone incrementar la base con ejemplos característicos, si no mejorar en la generalización de la red. Y hacerlo sin perder la exactitud adquirida hasta ahora.

El procedimiento habitual es incrementar el número de ejemplos de entrenamiento a partir de transformaciones básicas sobre las imágenes tales como filtros, ruidos, escalamientos, recortes o giros. De entre todas las posibilidades, se ha tenido en cuenta aquellas que mantengan las formas y bordes de la imagen de partida, por considerar es el punto de partida de la abstracción en las sucesivas capas de convolución. Obviamente, se han descartado aquellas transformaciones de la imagen que la red nunca encontraría en el proceso de validación de una atelectasia.

Se ha tenido muy en cuenta que, en el proceso de transformación de la imagen para incluirla en la base de entrenamiento o en la base general de datos, es primordial decidir que transformaciones de la imagen son adecuadas. Transformaciones tales como cambios en el brillo, contraste o nitidez, no aportan datos con contenido significativo, pues se incrementa la base de entrenamiento con imágenes de propiedades muy uniformes. Es bien sabido que la transformación espejo, por ejemplo, añadiría a la base de entrenamiento un conjunto de imágenes que no responden a la realidad, pues dicho cambio genera una imagen que nunca corresponderá a un caso real, nunca la forma de un pulmón izquierdo con atelectasia será similar a un derecho.

Estudios recientes con radiografías de tórax que buscan este objetivo [16] demuestran que transformaciones tales como pequeñas rotaciones, zoom zonal, cambios en la iluminación de la imagen o deformaciones, generan buenos resultados.

Así que las transformaciones que se han realizado, dada la discusión del apartado anterior, tal como se puede apreciar en la Figura 13, son:

- Pequeñas rotaciones de hasta 10 grados.
- Zoom general, sin ocultar ROI: se ha aplicado un recorte con *resize* (zoom) de la zona de interés, y se han incluido este nuevo tipo de imágenes en la base de datos de entrenamiento a partir de las imágenes que clasifican mal en la base de datos.

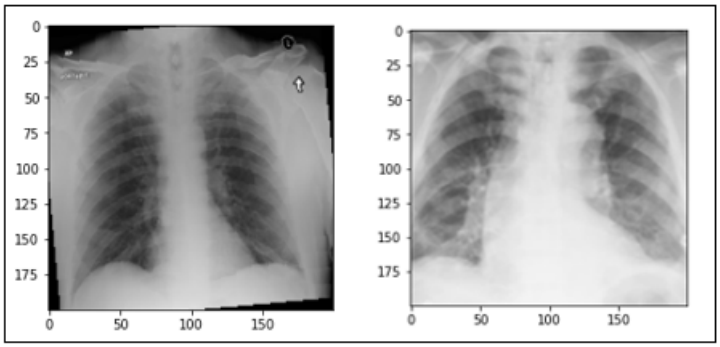


Figura 13. Imagen con preprocesamiento: una rotación de 10 grados y un zoom sobre la ROI.

Así pues, el algoritmo ha realizado los siguientes pasos (ver Figura 14*, X_test* contiene las imágenes a validar):

a. Ha encontrado los falsos negativos o falsos positivos: hará un recorrido por todas las imágenes del conjunto de imágenes de validación, en busque de imágenes de etiqueta "1", con atelectasia, pero clasificada como pulmones normales. Lo hará en base al mejor umbral de clasificación posible, que se habrá determinado con anterioridad.

b. Ha incrementado el conjunto de imágenes de entrenamiento y su etiqueta correspondiente con las imágenes obtenidas del apartado anterior, eliminándolas del grupo de validación y aplicando previamente una transformación de rotación o zoom sobre la ROI.

c. Ha vuelto a entrenar la red.

d. Ha validado de nuevo.

e. Ha repetido el ciclo hasta que el proceso converja a un valor fijo o hasta que el grupo de imágenes de validación reduzca demasiado su número.

```
for n in range(len(X_test)):
    img = X_test[n]
    input_img = np.expand_dims(img, axis=0)
    if ((model.predict(input_img) < ideal_roc_thresh) and (y_test[n] == 1)):
        img = Image.fromarray((X_test[n] * 255).astype(np.uint8))
        rotate_img= img.rotate(5)
        X_test_f = np.array(rotate_img)
        X_train_list.append(X_test_f)
        y_train_list.append(y_test[n])
        m=m+1

    else:
        X_test_list.append(X_test[n])
        y_test_list.append(y_test[n])
        r = r + 1
```

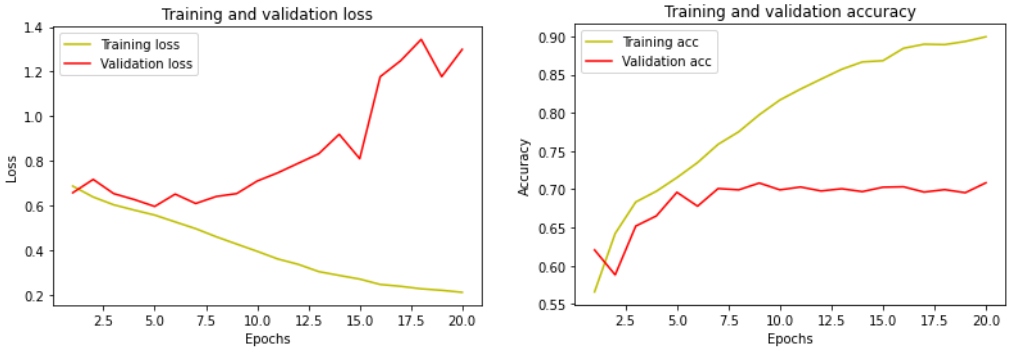
Figura 14. Algoritmo que incluye en la lista de imágenes de entrenamiento aquellas imágenes que clasifican mal según un umbral dinámico

Resultados

6.1 Resultados de la CNN de diseño propio con la base de datos sin refinar.

En esta propuesta inicial se ha alimentado la red con 17.305 imágenes de radiografías de tórax sin patologías de ningún tipo apreciable, y con 11.349 atelectasias de diversa índole, unas más evidentes que otras. Aproximadamente mismo número de cada categoría. El 80% se ha utilizado para el entrenamiento de la red, y el resto para su validación.

La evolución de la métrica obtenida queda recogida en los siguientes gráficos de resultados:



Gráfica 1. Evolución del error y de la exactitud en el entrenamiento y validación de la red de prueba

Como se observa en la Gráfica 1, a partir de unos pocos *epochs* la curva *validation loss* evoluciona de forma muy divergente a la curva equivalente de *training*. Alcanza un mínimo en torno al epoch 6, donde se produciría un *early stopping*, y progresa abruptamente elevándose a partir de aquí. Esto podría ser un caso de *underfitting*, esto es, el problema de clasificar esta imagen es demasiado complejo, no se puede modelizar por esta red. Le cuesta mucho diferenciar una imagen sin patologías de una imagen con atelectasia. Esto parece lógico pues las imágenes se han seleccionado a partir de una base de datos que incluye textos con diagnósticos, y es difícil en muchos casos discernir las diferencias a simple vista.

Matriz de confusión, exactitud, precisión, sensibilidad y especificidad.

La matriz de confusión es la siguiente con un umbral de 0,51 es muy simétrica (figura 15).

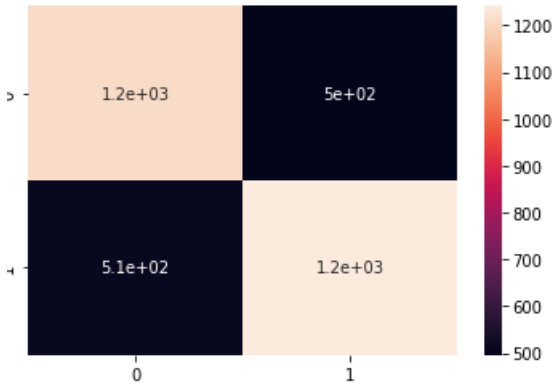


Figura 15. Matriz de confusión del algoritmo de clasificación diseñado con CNN

Es decir, de las 3.410 imágenes utilizadas en la validación de la red, un 20% del total, y entendiéndose por *imágenes negativas*, 1.700, de etiqueta "0" en la matriz de confusión, pulmones sin afecciones, y por *imágenes positivas,* 1.710, de etiqueta "1" en la matriz de confusión, pulmones con atelectasias:

- Verdaderos negativos: 1.200 imágenes negativas fueron clasificadas como negativas.
- Verdaderos positivos: 1.200 imágenes positivas fueron clasificadas como positivas.
- Falsos negativos: 500 imágenes positivas fueron clasificadas como negativas.
- Falsos positivos: 510 imágenes negativas fueron clasificadas como positivas.

Eso arroja los siguientes parámetros de evaluación, las cuatro lecturas posibles de la matriz de confusión:

- Exactitud: este parámetro mide cómo se comporta la red generalizando los datos a los datos de validación, es decir, dentro del entorno de la base de datos escogida para el estudio. Mide los aciertos de la red. En forma gráfica:

$$Exactitud = \frac{Valores\ verdaderos}{Todos\ los\ valores} = \frac{2400}{3410} = 0{,}70$$

- Precisión: mide lo buena que es la red detectando positivos, esto es, patología. En otras palabras:

$$Precisión = \frac{Valores\ verdaderos\ positivos}{Valores\ clasificados\ positivos} = \frac{1200}{1700} = 0{,}71$$

- Sensibilidad: es un valor que determina lo buena que es nuestra red discriminando casos positivos. Es la tasa de verdaderos positivos frente a positivos reales. Dicho de otro modo, la capacidad de la red de detectar pacientes enfermos de entre todos los enfermos.

$$Sensibilidad = \frac{Valores\ verdaderos\ positivos}{Todos\ los\ positivos\ reales} = \frac{1200}{1710} = 0{,}70$$

- Especificidad: es un valor que determina lo buena que es nuestra red discriminando casos negativos, esto es, sanos. Es la tasa de verdaderos negativos frente a negativos reales. En términos médicos, la capacidad de reconocer pacientes sanos de entre todos los pacientes sanos.

$$Especificidad = \frac{Valores\ verdaderos\ negativos}{Todos\ los\ negativos\ reales} = \frac{1200}{1700} = 0{,}71$$

Medición de la precisión de la prueba

Otros valores interesantes que pueden ser calculados y sirven para validar la red son [17]:

- Tasa de falsos positivos (TFP)

  Es la tasa de falsa alarma, esto es, la probabilidad de que se dé un falso positivo (en realidad es negativo).

$$TFP = \frac{Valores\ falsos\ positivos}{Todos\ los\ negativos\ reales} = \frac{500}{1700} = 0{,}29$$

- Valor predictivo positivo (VPP)

  Es la probabilidad de que un valor positivo sea realmente positivo.

$$VPP = \frac{Valores\ verdaderos\ positivos}{Todos\ los\ positivos} = \frac{1200}{1710} = 0{,}71$$

- Valor predictivo negativo (VPN)

  Es la probabilidad de que un valor negativo sea realmente negativo, es decir, no tenga patología.

$$VPN = \frac{Valores\ verdaderos\ negativos}{Todos\ los\ negativos} = \frac{1200}{1700} = 0{,}70$$

Con los datos de las cuatro posibles lecturas de la matriz, resumidas en la tabla 4, y las relaciones entre ellos, es posible de forma más o menos objetiva establecer la validez la red utilizada. La tasa de precisión y la sensibilidad nos dirá hasta qué punto la red sirve para predecir. Un parámetro que mide conjuntamente ambos parámetros es el llamado *f1-score.* Se calcula de la siguiente manera:

$$F1 = \frac{2 \cdot Precisión \cdot Sensibilidad}{Precisión + Sensibilidad} = \frac{2 \cdot 0{,}7 \cdot 0{,}71}{0{,}7 + 0{,}71} = 0{,}71$$

Exactitud	70
Precisión	71
Sensibilidad	70
Especificidad	70

Tabla 4. Cuatro lecturas en % de la matriz de confusión de la CNN

La tasa f1 de 0,71 es baja como para valorar positivamente el poder predictivo de la red. La sensibilidad es mayor que la precisión. Esto viene a decir que la red reconoce los positivos, pero incluye en esta clase demasiados elementos de la otra. Confunde demasiados casos de pulmones sin patologías como pulmones con Atelectasias.

6.2 Resultados de la CNN con la base de datos optimizada.

Revisión de la catalogación

Tras revisar las imágenes de la carpeta “atelectasia”, las bases de datos serán las siguientes:

a. Se ha analizado la carpeta de pulmones sin patologías, y eliminamos aquellas que no responden con claridad a este criterio, es decir eliminamos confusas.

b. Se han seleccionado algunas imágenes, las que creemos suficientes para elaborar nuestras pruebas con un mínimo de validez, con los criterios que definen las patologías analizadas (colapsos totales y atelectasias lobulares superior derecha en inferior izquierda). Las nuevas carpetas filtradas por patologías son las siguientes:

- Atelectasias con eliminación de imágenes confusas (EC): 17.056 imágenes.
- Atelectasias con colapso total de un pulmón (CT): 95 imágenes.
- Atelectasias lobulares del lóbulo superior derecho (LSD): 163 imágenes.
- Atelectasias lobulares del lóbulo inferior derecho (LID): 93 imágenes.

Tras probar la red con estas bases de datos seleccionadas según los criterios citados, se presentan en la Tabla 5:

	Inicial	**EC**	**CT**	**LSD**	**LID**
Exactitud	70	73	94	83	89

Tabla 5. Resultados en % tras refinar el etiquetado de las carpetas iniciales, comparado con el análisis anterior

Si se comparan los resultados con nuestra red inicial sin eliminación de imágenes confusas (tal como se especifica en los resultados del apartado justamente anterior) y con eliminación de imágenes confusas de la carpeta de “imágenes sin patologías”, así como con una selección de imágenes por patologías, se comprueba que los resultados son sensiblemente mejores (Tabla 5), como se explica a continuación:

- Los resultados en la clasificación de la base de datos con todos los elementos antes y después de eliminar imágenes confusas no cambia demasiado, antes del 70%, y después al 73%.

- Los resultados en la clasificación una vez separadas por patologías mejoran hasta el 94% en colapsos totales, 83% en lobulares superior derecho, y 89% inferior izquierdo.

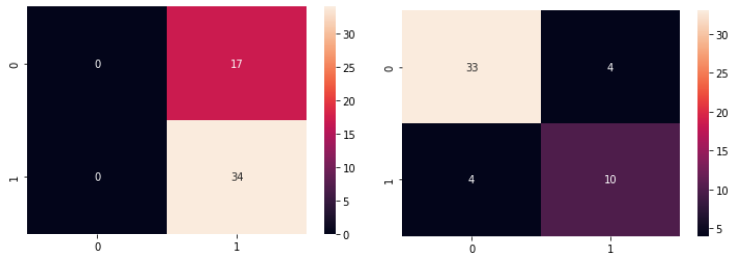


Figura 16. Matriz de confusión cuando enfrentamos imágenes de la misma clase y de las clases LSD y LID

Resultados con *data-augmentation* y umbral ideal

Para comprobar la hipótesis de este apartado, se ha recurrido a las bases de datos que se ha utilizado en el refinamiento de etiquetado del apartado anterior, que ya suponía una mejora significativa de los primeros resultados obtenidos con todas las atelectasias mezcladas. Evidentemente, se conocen las limitaciones que surgen cuando en los entrenamientos se utilizan bases de datos escasas y fácilmente "clasificables" en cuanto a una posible generalización de la red. Sin embargo, tiene valor teórico y es un buen punto de partida del que podemos sacar conclusiones iniciales.

Cuando se ha aplicado el algoritmo de enriquecimiento de la base de datos con imágenes que clasifican mal, variando progresivamente el umbral ideal cada cierto número de *epochs,* los resultados mejoran ostensiblemente, como se aprecia en la Tabla 6. Los elementos de la base de datos que se recogen en la primera columna en dicha tabla, parte son para el entrenamiento, parte para la validación. El número de ellos para uno u otro proceso varían un poco según el proceso de enriquecimiento de la base de entrenamiento se va enriqueciendo con las imágenes que clasifican mal. El número de elementos finalmente validados, en todo caso, son los que figuran en las matrices de confusión de la Figura 16.

	Elementos	Exactitud anterior	Exactitud posterior
LID	148	89	95
LSD	320	83	94
Colapso	148	93	97

Tabla 6. Resultados de las redes antes y después de aplicar la revisión progresiva de la base de datos de forma dinámica

No se han seguido exactamente los mismos pasos en los tres procesos. En los dos primeros, se ha reeducado la red con falsos negativos. Es decir, imágenes de atelectasias clasificadas como normales. En el segundo, sin embargo, dado el éxito de la red identificando colapsos, ha sido mucho más eficiente proceder al contrario. Ampliar la base de datos con falsos positivos: imágenes normales que eran mal clasificadas. Sólo de esta manera la red ha convergido al valor reflejado en la Tabla 6.

Ya sólo incluyendo directamente las imágenes no validadas correctamente se obtiene un salto significativo en la exactitud de la red. Como se aprecia en la Figura 17, el número de falsos positivos y falsos negativos (diagonal en negro) se ha reducido considerablemente. Al analizar los errores de validación, comprobamos que convergen homogéneamente con los errores de entrenamiento, como se espera de una CNN que funcione apropiadamente.

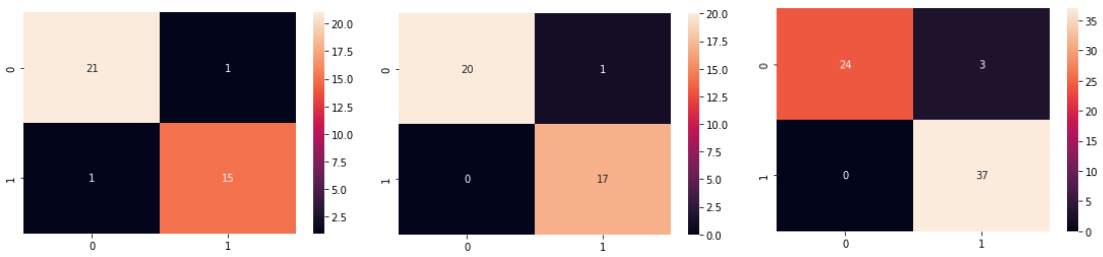


Figura 17. Matriz de confusión de las clases LID, colapso y LSD con ampliación dinámica del paquete de entrenamiento

Resultados con *data-augmentation*, umbral ideal y preprocesamiento

Las transformaciones realizadas y los efectos observados han sido:

- Pequeñas rotaciones de hasta 10 grados: como se observa en la Figura 18 (falsos positivos en diagonal en negro), pequeñas rotaciones en la imagen, no alteran la esencia de la imagen, cargando la base de datos con más ejemplos futuribles que a la postre mejoran la validación hasta el 95% en exactitud si se va conformando la misma con los datos que peor clasifican, en la primera de las modalidades que queremos mostrar en este apartado.
- Zoom general, sin ocultar ROI: El resultado se puede apreciar en la Figura 19. De las 49 imágenes sólo dos han resultado falsos positivos.

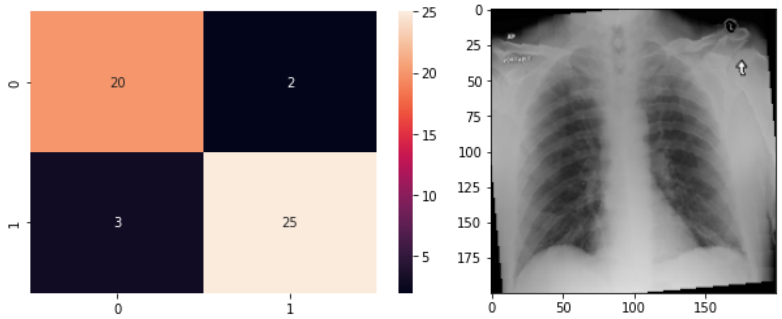


Figura 18. Matriz de confusión y ejemplo de las clases LSD con incremento con imágenes rotadas rotación 5º

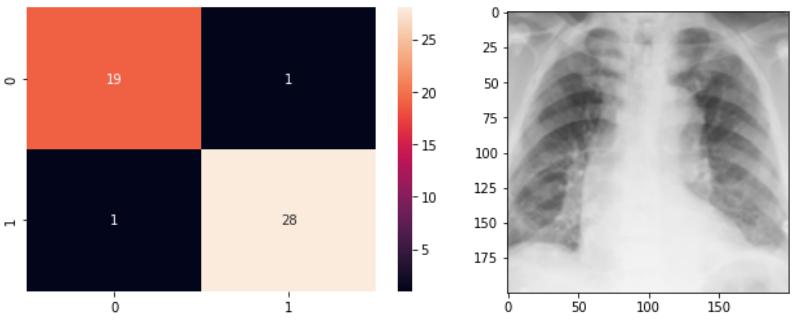


Figura 19. Matriz de confusión y ejemplo de las clases LSD con incremento con imágenes con zoom en ROI

Se ha considerado que, el siguiente paso natural sería probar la base entera de entrenamiento y obtener la validación con un preprocesamiento previo de transformaciones similares a las efectuadas hasta ahora, como paso previo al paso por la red. Así, al realizar un procesamiento previo

de todas las imágenes con un zoom en la zona de interés, se han obtenido obtienen resultados sensiblemente mejores, de hasta 96%, como se observa en la Figura 20.

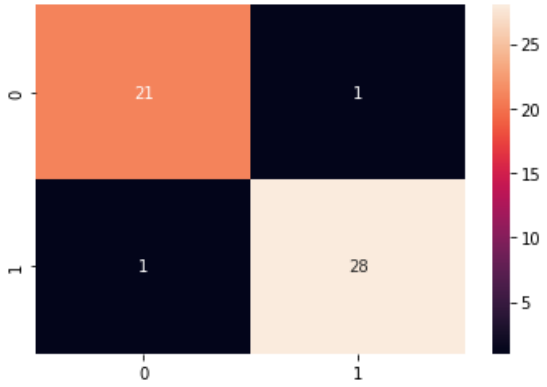


Figura 20. Matriz de confusión y ejemplo de las clases LSD con preprocesamiento previo en ROI

Discusión de resultados

Las bases de datos médicas consultadas se han elaborado con criterios no exclusivamente visuales, sino en base al historial médico del paciente y la experiencia del radiólogo. Así, cuando se ha ensayado la red de partida de diseño propio con la base de datos inicial, que contenía el conjunto de todas las atelectasias, los resultados de la red han sido muy pobres (70%). Esto era en principio predecible. En el diagnóstico de atelectasias a partir de una radiografía de tórax por una red neuronal de convolución las únicas referencias son los píxeles.

Cuando se ha ordenado la base de datos la red ha encontrado mucho mejor un modelo que se ajuste al problema. Si se separan las imágenes con patologías evidentes al ojo no entrenado según el tipo de patología en particular, los resultados de la red mejoran un 10% de media con muchas menos imágenes en la base de datos para entrenar.

Es de destacar el dato del alto reconocimiento de imágenes con colapsos totales (94%), que viene a decir que cuanto más diferentes sean dos imágenes, mejor encajan en dos clases diferenciables por la red. Si se enfrentan dos clases de atelectasias entre sí arbitrarias, por ejemplo, imágenes con atelectasia de LSD (Lóbulo superior derecho) con imágenes con atelectasias de LID (Lóbulo inferior izquierdo), los resultados son compatibles: la red clasifica ambas patologías con una exactitud del 80%. Igualmente, si se enfrentan dos carpetas con pulmones de la misma clase, la red no ofrece ningún valor predictivo.

Se ha encontrado un factor significativo de mejora en el aprendizaje de la red cuando se ha incluido en la base de entrenamiento los mejores ejemplos posibles. Se ha probado que, si durante el ciclo de entrenamiento se seleccionan las imágenes que constituyen falsos positivos o falsos negativos del grupo de imágenes de validación, mejoran ostensiblemente los resultados. Al incluir estas imágenes se ha conseguido en varios casos resultados muy satisfactorios, de hasta el 97% en casos particulares, con muy pocas imágenes de partida.

Para determinar qué imágenes clasifican mal se ha recurrido al cálculo del umbral ideal de clasificación, que se va decidiendo en el propio proceso de aprendizaje. Dicho umbral, que minimiza el número de falsos resultados, decide qué imágenes son las que peor clasifican. Al incluir dichas imágenes en el entrenamiento es cuando los resultados de la red han mejorado sensiblemente.

Cuando hemos ampliado la base de datos con las imágenes que constituyen falsos positivos o falsos negativos con preprocesamiento previo, esto es, pequeñas rotaciones y zoom de la ROI, los resultados siguen siendo buenos y las curvas de error y precisión de entrenamiento y validación convergen.

Conclusiones

1. Se pueden realizar pruebas en redes sencillas sobre bases de datos (no muy abundantes) con medios modestos y sin coste, al alcance del investigador interesado en el tema.
2. La base de datos de partida [7] mezcla diferentes tipos de atelectasia, donde muchas de las radiografías no ofrecen un diagnóstico evidente. Se demuestra en este estudio que los resultados de nuestra red mejoran con la especialización de su aprendizaje: si refinamos la tipología de ejemplos hasta el diagnóstico último se obtienen resultados interesantes con bases de datos del orden de cientos de imágenes. La red obtiene un 70% de acierto en el reconocimiento 11.349 imágenes de atelectasias generales frente a un número similar de pulmones sin patologías, y hasta un 94% con 95 imágenes de atelectasias de colapso total frente a un número similar de pulmones sin patologías. Se demuestra que es mas provechoso especializar la red que incrementar indiscriminadamente el número de datos, que es la tendencia más habitual. Para especializar la red se debe refinar la base de datos de entrenamiento con criterios exclusivamente visuales de la patología concreta.
3. El aprendizaje de la CNN está ligado al conjunto de imágenes de entrenamiento. Si se incluyen en dicha base de entrenamiento a los mejores ejemplos para dicha tarea los resultados de la validación mejoran hasta un 10% en algunos casos. En este estudio se ha considerado los mejores ejemplos posibles como aquellos que la red no logra identificar en sus primeras fases de entrenamiento, esto es, los falsos positivos y los falsos negativos.

Bibliografía

1. Gerald J. Tortora, Sandra Reynolds Grabowski (2005). *Principios de anatomía y fisiología*.
2. Lauren R. Goodman, 3ª ed (2007). *Felson. Principios de radiología torácica*. Mc Graw Hill.
3. Isabel Belda González, Daniel Soliva Martínez, Pedro Fernández Iglesias, Lourdes Hernández Muñóz, Vilbrun Jean-Pierre. *Atelectasias Pulmonares: Aprende a verlas para poder encontrarlas,* Sociedad Española de Radiología Médica (SERAM).
4. Juan Gabriel Gomila (28 de enero de 2020, última consulta octubre 2022). *La guía definitiva de las redes convolucionales*. Frogames. https://frogames.es/la-guia-definitiva-de-las-redes-neuronales-convolucionales/.
5. Wang, J.; Lee, S. *Data Augmentation Methods Applying Grayscale Images for Convolutional Neural Networks in Machine Vision*. *Appl. Sci.* **2021**, *11*, 6721. https://doi.org/10.3390/app11156721
6. Monshi MMA, Poon J, Chung V, Monshi FM. *CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR*. Comput Biol Med. 2021 Jun;133:104375. doi: 10.1016/j.compbiomed.2021.104375. Epub 2021 Apr 15. PMID: 33866253; PMCID: PMC8048393.
7. Summers, Ronald (NIH/CC/DRD) (Creado en sept 2017, actualizado abril 2022, última consulta octubre 2022). National Institutes of Health - Clinical Center: *CXR8.* https://nihcc.app.box.com/v/ChestXray-NIHCC
8. Morteza Heidari, Seyedehnafiseh Mirniaharikandehei, Abolfazl Zargari Khuzani, Gopichandh Danala, Yuchen Qiu, Bin Zheng, *Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms*, International Journal of Medical Informatics, Volume 144, 2020, https://doi.org/10.1016/j.ijmedinf.2020.104284.
9. Avinash Thite. (2021, Octubre, última consulta septiembre 2022). Great Learning: *Introduction to VGG16 | What is VGG16?* https://www.mygreatlearning.com/blog/introduction-to-vgg16/
10. Narim, Ali; Kaya, Ceren; Pamuk, Ziynet. *Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Analysis and Applications*, 2021, vol. 24, no 3, p. 1207-1220.

11. Monshi MMA, Poon J, Chung V, Monshi FM. *CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR*. Comput Biol Med. 2021 Jun;133:104375. doi: 10.1016/j.compbiomed.2021.104375. Epub 2021 Apr 15. PMID: 33866253; PMCID: PMC8048393

12. Arias-Londono JD, Gomez-Garcia JA, Moro-Velazquez L, Godino-Llorente JI. *Artificial Intelligence Applied to Chest X-Ray Images for the Automatic Detection of COVID-19. A Thoughtful Evaluation Approach.* IEEE Access. 2020 Dec 14;8:226811-226827. doi: 10.1109/ACCESS.2020.3044858. PMID: 34786299; PMCID: PMC8545248.

13. Wang L., Lin Z. Q., and Wong A., *COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images*, Sci. Rep., vol. 10, no. 1, Dec. 2020, Art. no. 19549.

14. Giełczyk A, Marciniak A, Tarczewska M, Lutowski Z (2022) *Pre-processing methods in chest X-ray image classification*. PLoS ONE 17(4): e0265949. https://doi.org/10.1371/journal.pone.0265949

15. Shorten, C., Khoshgoftaar, T.M. *A survey on Image Data Augmentation for Deep Learning. J Big Data* 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

16. Wang, J.; Lee, S. *Data Augmentation Methods Applying Grayscale Images for Convolutional Neural Networks in Machine Vision*. *Appl. Sci.* 2021, *11*, 6721. https://doi.org/10.3390/app11156721

17. Barrios Arce, J.I. (2019, 26 de Julio, última consulta agosto 2022). Health with big data: *La matriz de confusión y sus métricas*. https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/