Document downloaded from http://www.elsevier.es, day 23/04/2024. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Clasificador para radiografías torácicas QUIBIM PRECISION: funcionamiento y validación

Descripción del clasificador para radiografías torácicas QUIBIM Precision (modelo de inteligencia artificial)

Modelo de aprendizaje profundo basado en las redes neuronales convolucionales (RNC) como herramienta informática de diagnóstico para el triaje de radiografías torácicas y la priorización de la lista de trabajo. El algoritmo incluye un conjunto de modelos de aprendizaje profundo que calculan la probabilidad de obtener 16 resultados diferentes y la probabilidad de anomalía en radiografías torácicas posteroanterior/anteroposterior (PA/AP). Estos 16 hallazgos patológicos cardiotorácicos distintos son los siguientes: atelectasia, cardiomegalia, hepatización, COVID-19, edema, enfisema, cardiomediastino aumentado, fibrosis, fractura, hernia, lesión pulmonar, opacidad pulmonar, derrame pleural, afectación pleural, engrosamiento pleural y neumotórax.

La estructura de las RNC que se utilizan en los clasificadores de datos de radiografías torácicas es DenseNet121 con 3 × 320 × 320 de tamaño de datos de la imagen. Los valores de los datos de las radiografías se normalizan mediante los pesos de ImageNet. Los modelos proporcionan la probabilidad de observar cada uno de los hallazgos en la radiografía introducida.
Los modelos de RNC en términos de hallazgos se agrupan mediante una red neuronal de referencia totalmente conectada que consta de tres capas densas compuestas por 8, 4 y 2 neuronas, respectivamente. El modelo proporciona la probabilidad de observar una anomalía en la radiografía introducida.

Datos e indicaciones de uso del clasificador para radiografías torácicas QUIBIM Precision (modelo de inteligencia artificial) (tabla 1)

Tabla 1 Datos e indicaciones de uso del clasificador para radiografías torácicas QUIBIM Precision

Protocolo de obtención de imágenes	**Modalidad**	Radiografía torácica
	Campo de visión	AP/AP
	Posición del paciente	Recta
	Contenido de la imagen	Tórax completo La imagen precisa cumplir los siguientes requisitos para el correcto funcionamiento del sistema: Radiografías torácicas AP/PA. El sistema realiza todo el análisis detallado en las siguientes secciones si hay al menos una radiografía AP/PA en los datos. En caso contrario, se desencadena una incidencia. ⚠ Radiografía torácica de un adulto. Las radiografías de las bases de datos utilizadas para entrenar a los modelos de inteligencia artificial procedían de pacientes mayores de 18 años.

Document downloaded from http://www.elsevier.es, day 23/04/2024. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Elaboración del clasificador para radiografías torácicas QUIBIM Precision

El conjunto de datos completo para entrenar y analizar el modelo de clasificador para radiografías torácicas estaba compuesto por 344 339 radiografías AP/PA procedentes de cuatro conjuntos de datos distintos:

• 112 120 muestras del conjunto de datos de 14 radiografías torácicas de los NIH (National Institutes of Health), que contenían imágenes etiquetadas con 14 patologías distintas obtenidas mediante técnicas de tratamiento del lenguaje natural aplicadas a los informes radiológicos de expertos.
• 191 229 muestras del conjunto de datos CheXpert, que contenía imágenes etiquetadas para la presencia de 14 observaciones como positivas, negativas o dudosas. Decidimos considerar que las dudosas también podían tratarse como negativas.
• 39 308 muestras del conjunto de datos PadChest, que contenía muestras a gran escala y resolución elevada para la exploración automatizada de imágenes médicas junto con sus informes asociados. Los informes se etiquetaron con 174 hallazgos radiográficos distintos y decidimos reducir el número de etiquetas y tratar de ajustarlas a las presentes en los conjuntos de datos de los NIH y de CheXpert.
• 1682 muestras de una base de datos no pública que contenía radiografías torácicas positivas para COVID-19.

La base de datos se dividió en dos grupos: entrenamiento y análisis.

Entrenamiento

Del total de 334 339 muestras, 330 939 se utilizaron en el proceso de entrenamiento.

Análisis

Con el fin de poder realizar una métrica de validación por clase, fue necesario dejar un conjunto de análisis compuesto de 3400 muestras. Este conjunto se organizó en 17 grupos (16 hallazgos más la clase anómala) de 200 muestras (100 correspondientes a las muestras con el hallazgo en cuestión y 100 a cualquier muestra que no presentó el hallazgo).

Para entrenar al clasificador de radiografías torácicas, se utilizó la biblioteca FastAI además de la biblioteca de aprendizaje profundo PyTorch. La estructura de RNC utilizada para entrenar a los modelos específicos de 16 hallazgos fue DenseNet121 previamente entrenada con ImageNet. La capa de salida se modificó con el fin de adaptar las RNC para una clasificación binaria. La estructura fue seleccionada por su profundidad y, por lo tanto, su capacidad para extraer patrones muy complejos y sutiles, necesarios para evaluar adecuadamente los distintos hallazgos de las radiografías torácicas. El tamaño seleccionado de las imágenes introducidas fue de 3 × 320 × 320,

Document downloaded from http://www.elsevier.es, day 23/04/2024. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

lo que permitió identificar y caracterizar pequeños hallazgos como nódulos pulmonares de milímetros de tamaño. Los valores de intensidad de píxeles de las radiografías se normalizaron mediante los estadísticos de ImageNet para mejorar la compatibilidad con pesos previamente entrenados de las RNC. Los bloques que componen la base DenseNet121 se enumeran en la figura 1.

Capas	Tamaño de salida	DenseNet-121
Convolución	112 × 112	
Agrupamiento	56 × 56	
Bloque denso (1)	56 × 56	[1 × 1 conv. / 3 × 3 conv.] × 6
Capa de transición (1)	56 × 56	
	28 × 28	
Bloque denso (2)	28 × 28	[1 × 1 conv. / 3 × 3 conv.] × 12
Capa de transición (2)	28 × 28	
	14 × 14	
Bloque denso (3)	14 × 14	[1 × 1 conv. / 3 × 3 conv.] × 24
Capa de transición (3)	14 × 14	
	7 × 7	
Bloque denso (4)	7 × 7	[1 × 1 conv. / 3 × 3 conv.] × 16
Capa de clasificación	1 × 1	

Figura 1 Bloque de construcción DenseNet121.

El proceso de entrenamiento de los 16 modelos de RNC consistió en ocho etapas. Durante las siete primeras etapas, que duraron cinco épocas, las capas convolucionales se congelaron en las etapas impares y se descongelaron en las pares. La velocidad de aprendizaje se adaptó a cada etapa y cada modelo de RNC de forma independiente, y varió de $1{,}00E^{-08}$ a $1{,}00E^{-03}$. El tamaño de la imagen introducida aumentó a lo largo de las etapas, desde 3 × 128 × 128 hasta 3 × 320 × 320. Durante el proceso de entrenamiento se aplicaron técnicas de aumento de los datos, que consistieron en rotaciones aleatorias en el intervalo de [–20, 20] grados, *zooms* en el intervalo [0,95; 1,05] y variaciones de iluminación en el intervalo [0%, 10%].

Las 16 probabilidades de salida de las RNC se agruparon mediante una red totalmente conectada de referencia para obtener la probabilidad de anomalía. La red de referencia consistió en tres capas densas compuestas de 8, 4 y 2 neuronas, respectivamente. La función de activación seleccionada para las capas ocultas fue ReLU y Softmax para la capa de salida. El modelo se entrenó durante 200 épocas, con una velocidad de aprendizaje de $1{,}00E^{-03}$.

Validación del clasificador para radiografías torácicas QUIBIM Precision

Los parámetros de validación del clasificador de radiografías torácicas se midieron mediante un conjunto de análisis descrito anteriormente. Los resultados obtenidos se muestran en la tabla 2.

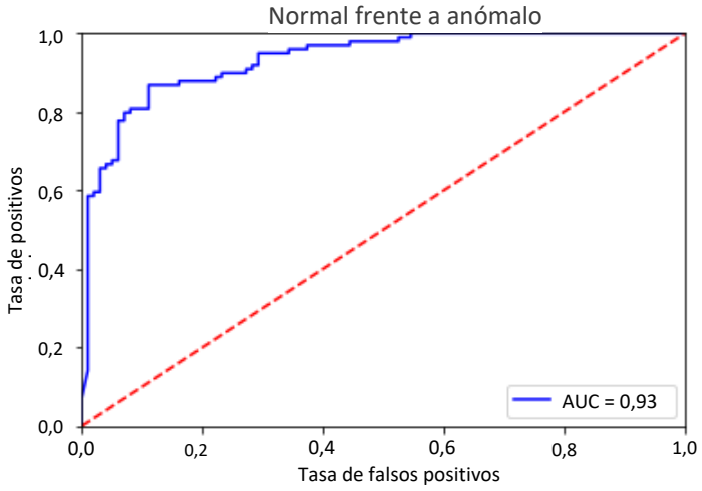
Document downloaded from http://www.elsevier.es, day 23/04/2024. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Tabla 2

Parámetros de los clasificadores de hallazgos							
Hallazgos	**Exactitud**	**Precisión**	**Sensibilidad**	**Especificidad**	**Puntuación F1**	**AUC**	**Umbrales**
Atelectasia	0,83	0,81	0,86	0,80	0,83	0,88	0,475
Cardiomegalia	0,92	0,89	0,96	0,88	0,92	0,96	0,275
Hepatización	0,79	0,75	0,87	0,72	0,81	0,84	0,430
COVID-19	0,96	0,93	1,00	0,92	0,96	0,95	0,961
Edema	0,85	0,84	0,88	0,83	0,86	0,92	0,447
Enfisema	0,94	0,94	0,95	0,94	0,94	0,98	0,580
Cardiomediastino aumentado	0,78	0,75	0,86	0,71	0,80	0,84	0,510
Fibrosis	0,85	0,82	0,91	0,80	0,86	0,93	0,510
Fractura	0,81	0,77	0,89	0,74	0,83	0,88	0,530
Hernia	0,92	0,92	0,93	0,92	0,92	0,97	0,467
Lesión pulmonar	0,77	0,75	0,83	0,72	0,78	0,84	0,480
Opacidad pulmonar	0,84	0,84	0,85	0,84	0,84	0,91	0,578
Derrame pleural	0,81	0,80	0,83	0,80	0,82	0,89	0,528
Pleural Otros	0,78	0,77	0,81	0,76	0,79	0,85	0,586
Engrosamiento pleural	0,86	0,85	0,89	0,84	0,87	0,93	0,520
Neumotórax	0,84	0,83	0,86	0,82	0,84	0,90	0,520
Normal frente a anómalo	**0,87**	**0,89**	**0,86**	**0,89**	**0,87**	**0,93**	0,550

El proceso que se siguió para determinar los umbrales mostrados en la tabla 2 fue el análisis de la exactitud, la sensibilidad y la especificidad, obtenidas con el uso de todos los umbrales entre 0 y 1, con una etapa de 0,001. El criterio para seleccionar el umbral consistió en elegir el punto más cercano al valor máximo de exactitud que presentó una sensibilidad superior a 0,80.

En la figura 2 se proporciona el AUC para la clasificación normal frente a anómalo.



El clasificador de radiografías torácicas proporciona los resultados que se muestran en la tabla 3

Document downloaded from http://www.elsevier.es, day 23/04/2024. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Tabla 3

Clasificador de hallazgos en radiografías torácicas	Probabilidad de anómalo
	Probabilidad de atelectasia
	Probabilidad de cardiomegalia
	Probabilidad de hepatización
	Probabilidad de COVID-19
	Probabilidad de edema
	Probabilidad de cardiomediastino aumentado
	Probabilidad de derrame pleural
	Probabilidad de enfisema
	Probabilidad de fibrosis
	Probabilidad de hernia
	Probabilidad de lesión pulmonar
	Probabilidad de opacidad pulmonar
	Probabilidad de engrosamiento pleural
	Probabilidad de afectación pleural
	Probabilidad de fractura
	Probabilidad de neumotórax

Todas las probabilidades de resultados proporcionadas por el algoritmo son valores de probabilidad entre 0 y 1. Estos valores representan la confianza del clasificador de radiografías torácicas para identificar esa imagen introducida como perteneciente a aquel hallazgo patológico. Un valor de 0 significaría que no hay ninguna probabilidad de pertenencia a ese hallazgo patológico y un valor de 1 significaría la certeza total de pertenencia a ese grupo patológico.

Una vez calculada la probabilidad de cada hallazgo detectado con el clasificador de radiografías torácicas a nivel individual, la herramienta las combina para cuantificar la probabilidad final de la anomalía. La probabilidad de anomalía se calcula mediante una red neuronal totalmente conectada que utiliza como datos introducidos las probabilidades de cada hallazgo, y proporciona como resultado la probabilidad global de la anomalía.

Tal como hemos explicado anteriormente, un valor de 0 significaría que no hay ninguna probabilidad de pertenencia a ese grupo de anomalía y un valor de 1 significaría la certeza total de pertenencia a ese grupo de anomalía.

Resultados obtenidos con el clasificador en la población con COVID-19

Document downloaded from http://www.elsevier.es, day 23/04/2024. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

De las 16 patologías mencionadas anteriormente, una de ellas es COVID-19. El modelo de RNC para la clasificación de COVID-19 se ha entrenado y validado como se detalla a continuación.

El modelo de RNC para la clasificación de COVID-19 se entrenó con 1582 muestras de una base de datos no pública de radiografías torácicas positivas para COVID-19 procedentes del Hospital Universitario y Politécnico La Fe, y el mismo número de muestras de radiografías torácicas no positivas para COVID-19 obtenidas de varios conjuntos de datos públicos: NIH Chest X-ray 14, CheXpert y PadChest.

La validación se realizó con un conjunto de análisis compuesto por 100 muestras de una base de datos no pública de radiografías torácicas positivas para COVID-19 procedentes del Hospital Universitario y Politécnico La Fe, y otras 100 muestras de radiografías torácicas no positivas para COVID-19 de varios conjuntos de datos públicos: NIH Chest X-ray 14, CheXpert y PadChest. Todas estas muestras no se utilizaron durante el proceso de entrenamiento.

Los resultados obtenidos en el conjunto de análisis se presentan en la tabla 4.

Tabla 4 Resultados obtenidos en el conjunto de análisis

Hallazgo	Exactitud	Precisión	Sensibilidad	Especificidad	Puntuación F1	AUC	Umbral
COVID-19	0,96	0,93	1,00	0,92	0,92	0,95	0,96

Todos estos parámetros se obtuvieron al seleccionar el umbral de 0,96. Si la probabilidad de resultado de la predicción de COVID-19 era superior a 0,96, la muestra se clasificó como COVID-19. Si no, es decir si la probabilidad del resultado era inferior a 0,96, la muestra se clasificó como no COVID-19. El criterio para seleccionar el umbral consistió en elegir el punto más cercano al valor máximo de exactitud que presentó una sensibilidad superior a 0,80.