

Clasificación de enfermedades del tórax usando aprendizaje profundo y aumento de datos de calidad en el conjunto de datos ChestX-ray8

Amhed A. Ortiz-Preciado, Juan Antonio Vega-Fernández, Gilberto Ochoa-Ruiz

Universidad Autónoma de Guadalajara,
Departamento de Ciencias Computacionales, Guadalajara, México
antonio.vega@edu.uag.mx, gilberto.ochoa@edu.uag.mx

Resumen. Presentamos un método novedoso para la clasificación de enfermedades del tórax en el conjunto de datos estándar y ampliamente utilizado ChestX-ray8, el cual contiene más de 100,000 imágenes de vista frontal con 8 enfermedades. Las principales contribuciones de este artículo son tres: 1. La introducción de una evaluación de la calidad de la imagen para el pre-filtrado del conjunto de datos, ya que ha sido demostrado que los ejemplos de entrenamiento de baja calidad afectan el rendimiento de las soluciones basadas en aprendizaje profundo. 2. Un enfoque de aumento de datos basados en la calidad de la imagen, y 3. Una arquitectura simplificada de red convolucional basada en AlexNet con solamente 200,000 parámetros. Con estas modificaciones estructurales a la arquitectura de la red y el aumento de datos basados en calidad, fuimos capaces de obtener resultados comparables con otras propuestas del estado del arte que usan el mismo conjunto de datos ChestX-ray8 en 4 de 8 categorías y logramos superar los resultados en las otras 4 categorías o enfermedades, tal como lo demuestran los resultados experimentales, todas usando una red convolucional que puede ejecutarse en dispositivos pequeños.

Palabras clave: aprendizaje profundo, CNN, aprendizaje débilmente-supervisado, chestX-Ray8.

Thoracic Disease Classification Using Deep Learning and Quality-Aware Augmentation on the ChestX-Ray8 Dataset

Abstract. We present a novel method for thoracic disease classification on the now standard and widely used ChestX-ray8 dataset, which contains over 100,000 frontal view images with 8 diseases. The main contributions of this paper are threefold: 1. The introduction of a quality image assessment for prefiltering the dataset, as it has been demonstrated that training examples with low quality undermine the performances on deep learning-based solutions. 2. A quality aware data augmentation approach and 3. A simplified ConvNet Architecture based on AlexNet with only 200,000 parameters. With these architectural modifications and via the use of quality-aware data augmentation, we were able to attain comparable results to other state of the art proposals that use the ChestX-ray8

dataset on 4 of the 8 categories and surpass them in the other 4, as validated by thorough experimental results, all using a ConvNet that can run on constrained devices.

Keywords: deep learning, CNN, weakly-supervised learning, chestX-Ray8.

1. Introducción

El campo de la visión computacional ha visto un progreso exponencial en una amplia gama de aplicaciones debido al uso de aprendizaje profundo y especialmente a la existencia de grandes conjuntos de datos de imágenes anotadas [1]. Se han mostrado mejoras significativas en el rendimiento de problemas previamente considerados difíciles, tales como el reconocimiento de objetos, detección y segmentación sobre los enfoques basados en obtener las características de la imagen a mano [2].

En años recientes, se ha demostrado que el aprendizaje profundo puede producir incrementos en rendimiento similares en el dominio del análisis de imágenes médicas para tareas de detección y segmentación de objetos [3]. Trabajos recientes notables incluyen aplicaciones médicas importantes, por ejemplo, en el dominio de la neumología (clasificación de enfermedades de los pulmones [4] y detección de nódulos pulmonares en imágenes de TC [5]). Comparado con otras aplicaciones en visión computacional, la principal limitante en las aplicaciones médicas de aprendizaje automático es que la mayoría de los métodos propuestos son evaluados utilizando un conjunto de datos mas bien pequeño con cientos de pacientes cuando mucho, pero se han hecho avances en los últimos años con la introducción de conjuntos de datos disponibles públicamente. En este artículo, enfocamos nuestros esfuerzos en la identificación y clasificación de enfermedades del tórax, un problema que presentamos en las siguientes secciones.

1.1. Uso de aprendizaje profundo en la identificación de enfermedades del tórax

Más de un millón de adultos son hospitalizados con neumonía y cerca de 50,000 mueren de esta enfermedad solamente en los EE. UU. cada año [6]. Las radiografías o placas del tórax actualmente son el mayor método para diagnosticar la neumonía [7], y juegan un papel crucial en los cuidados clínicos y los estudios epidemiológicos. Sin embargo, detectar la neumonía en las radiografías es una tarea difícil que depende de la disponibilidad de especialistas radiólogos, por lo que existe una amplia oportunidad de ayudar a los especialistas con nuevas tecnologías de Inteligencia Artificial para facilitar su trabajo y mejorar el sistema de salud.

Como se mencionó anteriormente, la placa del tórax se utiliza en el cribado y diagnóstico de muchas enfermedades de los pulmones. Un gran número de estudios de radiografías y reportes radiológicos típicamente son almacenados en los sistemas PACS (Sistema de Comunicación y Archivo de Imágenes) de los hospitales. Se han realizado una buena cantidad de investigaciones para aprovechar el conocimiento contenido en estas bases de datos, pero existe un reto muy específico: los registros médicos frecuentemente están débilmente etiquetados y no contienen anotaciones. Esto significa que la radiografía tiene asociado el diagnóstico del médico especialista -como negativo

(paciente sano) o positivo (identificando la enfermedad). Pero la radiografía no contiene delimitada la región de la imagen que lo llevo a tal conclusión o diagnóstico. Así que la investigación reciente se ha enfocado en explotar estos conjuntos de datos con los paradigmas de aprendizaje profundo, los cuales requieren de grandes cantidades de datos o imágenes, construyendo sistemas CAD de gran escala para propósitos médicos.

Detectar las enfermedades del tórax y especialmente la neumonía en las radiografías puede ser difícil para el especialista. La apariencia de varias enfermedades en la placa es frecuentemente vaga, se puede sobreponer con otros diagnósticos, y puede imitar otras anomalías benignas u otras enfermedades.

En la figura 1, se muestran algunas de las enfermedades más comunes y el enfoque de este y otros enfoques del estado del arte. Como se puede observar, la calidad de muchas de estas imágenes es subóptima y no es adecuada para los algoritmos de aprendizaje automático.

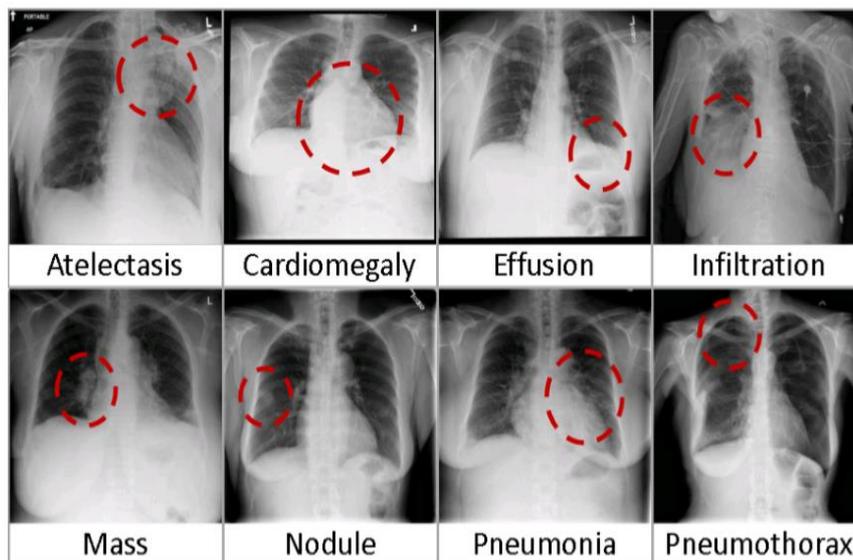


Fig. 1. Ocho enfermedades comunes del tórax observadas en radiografías que validan la dificultad del reto de automatizar el diagnóstico, tomado de [11].

Además, el etiquetado anatómico y patológico genérico y abierto, no se puede obtener por medio del *crowdsourcing* o subcontratación pública, dado que es implausible para anotadores no entrenados médicamente. También, la anotación completa al nivel de la región en cajas delimitantes para localizar la patología específica, lo cual sería necesario en conjuntos de datos para visión computacional [8] es actualmente inviable.

Consecuentemente, la mayoría de la investigación en la identificación de enfermedades del tórax es formulada como un problema de clasificación multi etiquetas débilmente supervisado, para superar las dificultades antes mencionadas y utilizando la información de la imagen completa en tándem con anotaciones débiles de la imagen y explorando diferentes avenidas de investigación.

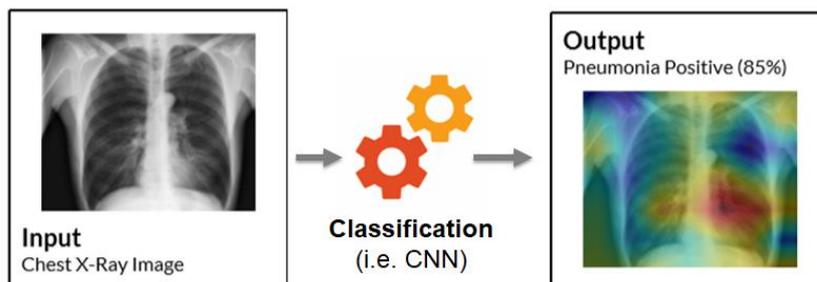


Fig. 2. El Modelo propuesto de CNN toma una radiografía como entrada y obtiene la probabilidad de la patología y la región en la imagen más probable donde se localiza esta patología como salida [9].

Con el fin de aliviar la escasez de conjuntos de datos disponibles públicamente, se han llevado a cabo varios esfuerzos recientes para crearlos [9], con números de pacientes que varían entre unos pocos cientos y hasta dos mil. Particularmente para las imágenes de placas del tórax, el conjunto de datos público más grande era Open-I [10], el cual contenía 3,955 reportes radiológicos de la Indiana Network for Patient Care y 7,470 placas del tórax asociadas de los sistemas PACS de hospitales.

Más recientemente, Wang et al. (2017) presentaron una nueva base de datos muy grande de radiografías, “ChestX-ray8” [11]. Esta base de datos se compone de 108,948 imágenes de rayos-X de vista frontal de 32,717 pacientes únicos con etiquetas de 8 enfermedades obtenidas usando minería de textos sobre el reporte radiológico y en donde cada imagen podía tener múltiples enfermedades. Este conjunto de datos se ha utilizado desde entonces en varios estudios utilizando enfoques débilmente supervisados para el entrenamiento de modelos de aprendizaje profundo.

Debido a la cuidadosa curación y a su gran volumen, ChestX-ray8 representa un estudio clínico más realista y completo que Open-I, y consecuentemente más adecuado para el desarrollo y comparación de modelos. Además, se ha convertido en el estándar de facto sobre el cuál, otros métodos han entrenado y validado sus modelos. Varios trabajos han tratado de resolver el problema de la anotación de la imagen en las placas del tórax [12], incluyendo el estudio inicial presentado junto con el conjunto de datos.

La mayoría de estos estudios, incluyendo el presentado en [13] hacen uso de modelos ConvNet RNN con el objetivo de modelar las dependencias entre etiquetas, las cuales son muy fuertes cuando se evalúan enfermedades torácicas. Adicionalmente, la mayoría de estos trabajos detectan neumonía y localizan las áreas en la imagen más indicativas de esta patología, creando un mapa de calor en el proceso, como se ilustra en la figura 2 arriba.

Comparado con trabajos previos, nuestro enfoque difiere en tres maneras principales: 1. Hacemos uso de aumento de datos para incrementar el tamaño del conjunto de datos de ChestX-ray8. 2. Implementamos un esquema de filtrado de datos para eliminar imágenes que pudieran impactar negativamente el proceso de entrenamiento, y 3. Diseñamos una red convolucional mucho más pequeña y reducida que logra mejores resultados que los presentado por Wang et al [2017]. En la siguiente sección, explicaremos en detalle cada una de estas contribuciones y en las secciones 3 y 4, discutiremos a fondo los resultados obtenidos por nuestra metodología. Finalmente, discutiremos las limitaciones de nuestro enfoque y las líneas de investigación futura.

2. Formulación del problema

El enfoque presentado aquí es un problema de clasificación múltiple, ya que identificamos y clasificamos la imagen de la radiografía en 8 posibles clases o enfermedades. Tomamos los datos del conjunto de ChestX-ray8 y lo entrenamos con una red convolucional para hacer la clasificación, y obtener un modelo después de algunos ajustes finos como el *dropout* o eliminación de pesos y la regularización L2, entre otras estrategias. Sin embargo, uno de los mayores problemas del aprendizaje profundo, cuando se trabaja con conjuntos de datos como el utilizado en este estudio es el problema del sobre entrenamiento o *overfitting*.

Una solución ampliamente usada es el aumento de datos (A.D.), esto es, generar más datos de entrenamiento a partir de las muestras existentes. aplicando a estos ejemplos una variedad de transformaciones aleatorias que produzcan imágenes creíbles. El proceso de aumento de datos puede reducir significativamente la pérdida de validación, como se ilustra en la figura 3, en donde se muestra la gráfica con datos de entrenamiento y validación para dos redes convolucionales idénticas, una sin usar aumento de datos y otra usando esta estrategia. Utilizando aumento de datos claramente combate el problema de sobre entrenamiento que se puede observar en el lado izquierdo de la figura, en donde se observa que el valor de la función de pérdida con los datos de validación inicialmente decremanta y luego se incrementa de nuevo. mientras que continúa decremantando con los datos de entrenamiento. El lado derecho de la figura muestra como la pérdida de validación se puede regularizar usando el aumento de datos, entre otras estrategias.

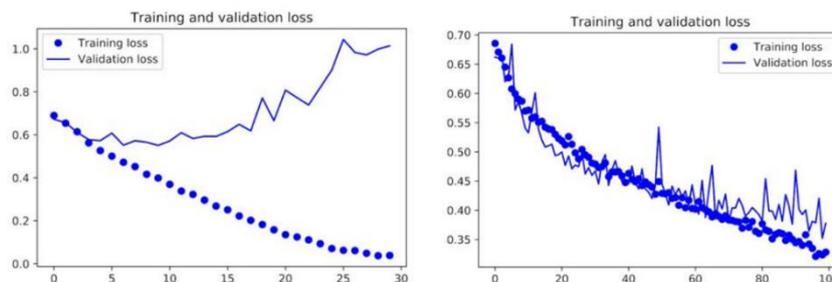


Fig. 3. Función de pérdida por entrenamiento y validación para dos redes convolucionales idénticas, a) sin usar aumento de datos y b) con aumento de datos. La función de pérdida claramente muestra como el A.D. previene el sobre entrenamiento.

Sin embargo, como veremos en la siguiente sección, si las estrategias de aumento de datos se implementan sin cuidado de la calidad de las imágenes, algunos ejemplos pobres en el entrenamiento nos pueden llevar a un modelo subóptimo de la red convolucional [14,15].

2.1. El conjunto de datos ChestX-ray8 y sus limitaciones

El enfoque de aumento de datos ha probado ser un método exitoso en diferentes aplicaciones de visión computacional, pero como veremos en la siguiente sección, hay algunas consideraciones importantes cuando se aplica este método a imágenes médicas. El

problema surge de la pobre calidad de algunas de las imágenes en el conjunto de datos, el cual puede afectar el entrenamiento y producir un modelo subóptimo. Estos modelos pueden introducir un gran número de falsos positivos o aún peor, clasificar equivocadamente un nuevo ejemplo de prueba, como lo documentan exhaustivamente los autores en [14]. Los autores de este estudio aplicaron una batería de pruebas en las cuales evaluaron el rendimiento de algunas redes convolucionales bien conocidas como VGG16 y GoogLeNet utilizando imágenes modificadas con diferentes tipos de distorsiones de la calidad de la imagen: compresión, bajo contraste, desenfoco y ruido gaussiano. Los autores encontraron que las redes son resilientes a distorsiones de compresión pero que se ven grandemente afectadas por las otras tres. Esto ha motivado una gran cantidad de investigación en esta área, con algunos trabajos concentrándose en los efectos que producen estas distorsiones en la calidad de la imagen para distintas tareas de visión computacional basadas en redes de aprendizaje profundo [15].

Esto es especialmente importante en el dominio de las imágenes médicas, particularmente con el conjunto de datos ChestX-ray8, el cual contiene 108,948 radiografías frontales de 32,717 pacientes únicos recolectados entre 1992 y 2015, con etiquetas de 8 enfermedades comunes obtenidas por mineo de texto de los reportes radiológicos utilizando técnicas de procesamiento de lenguaje natural. Hasta donde sabemos, no se ha reportado ninguna evaluación de la calidad de las imágenes en los trabajos previos con ese conjunto de datos, el cual incluye un gran número de imágenes con calidad pobre, como se observa en la figura 4, probablemente introducidos por la manera en la que los autores crearon el conjunto de datos. Algunas de las imágenes tienen un contraste muy bajo, mientras que otras están desenfocadas o carecen de definición, y otras sufren de saturación extrema.

Además de la estrategia de aumento de datos descrita anteriormente, la segunda contribución de este artículo es la implementación de una evaluación de la calidad de la imagen para determinar si el ejemplo de entrenamiento puede considerarse un buen candidato para la fase de aumento de datos y posterior procesamiento en la fase de entrenamiento. Algunas de las consideraciones para evaluar la calidad de la imagen, así como los algoritmos que se utilizaron para determinarla, serán presentados en la siguiente sección.



Fig. 4. Ejemplos de imágenes de baja calidad consideradas no útiles para clasificación, las cuales fueron automáticamente filtradas en nuestra fase de preprocesamiento.

2.2. Pre-proceso y filtrado de imágenes

Como se mencionó con anterioridad, el rendimiento de los modelos de redes convolucionales puede ser afectado severamente por la calidad de las imágenes de entrada; los autores de [14], realizaron pruebas individuales para ruido gaussiano, desenfoco, compresión y bajo contraste. Individualmente, aún para cantidades

moderadas de ruido o desenfoque, la exactitud de la red decreta significativamente, y uno solo puede asumir que la combinación de varias de estas distorsiones en las imágenes puede producir resultados más pobres. Así, en este artículo implementamos un “filtro” o proceso de selección basado en varias métricas¹ tradicionales del procesamiento de imágenes para estimar si una imagen puede ser considerada para el procesamiento posterior o no, tales como, la tasa de señal a ruido (SNR), el índice de desenfoque, y la varianza de la imagen para determinar el contraste. Típicamente, estos parámetros están correlacionados: una imagen con altos niveles de ruido presenta un bajo contraste y un alto desenfoque.

Las imágenes en los datos de ChestX-ray8 fueron primero pre-procesados usando técnicas de procesamiento de imágenes tradicionales: 1. Calculamos el índice de desenfoque utilizando un operador de estimación de desenfoque espectral. 2. Para estimar el contraste hacemos uso de la varianza como medida de la calidad de la imagen, usando la tasa de contraste de Michelson¹, como se aprecia en la figura 5, dejando fuera las imágenes con una varianza alta. 3. Adicionalmente, realizamos una prueba SNR utilizando los métodos propuestos por Legendijk y Biemond, para determinar si la imagen es adecuada para las fases de aumento de datos y entrenamiento.

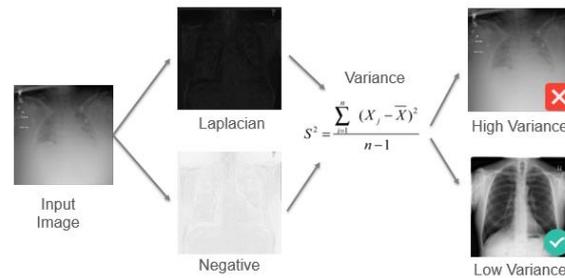


Fig. 5. Proceso para descartar imágenes con bajo contraste usando el cálculo de la varianza.

Una vez que estas pruebas se han llevado a cabo, realizamos los procedimientos de restauración de la imagen removiendo las distorsiones como el desenfoque (usando la convolución de Wiener), y bajo contraste (usando CLAHE, Contrast Limited Adaptive Histogram Equalization) y repitiendo la prueba SNR para determinar si la imagen corregida puede ser utilizada en el procesamiento posterior. Como veremos en la siguiente sección, este proceso produjo un conjunto de datos aumentado en donde las métricas de calidad de la imagen fueron tomadas en cuenta.

2.3. Resultados del proceso de aumento de datos en ChestX-ray8

En esta sección, discutiremos brevemente los resultados obtenidos con aumento de datos vía la evaluación de calidad de la imagen. Como se mencionó anteriormente, ChestX-ray8 contiene 108,948 radiografías frontales de 32,717 pacientes únicos. Debido a consideraciones de espacio, no incluimos la tabla completa aquí, y nos concentramos solamente en las patologías que hemos decidido enfrentar en este

¹ Legendijk and Biemond, “Basic methods for image restoration and identification”

artículo. Para una revisión más completa del conjunto de datos, el lector es dirigido a [9, 11] pero presentamos aquí una breve introducción.

Primero, el tamaño del conjunto de datos se incrementó a 136,526 imágenes, aumentando el número de ejemplos de entrenamiento para algunas de las clases que previamente estaban pobremente representadas, (por ejemplo, cardiomegalia y neumonía, por citar dos casos). Además, la mayoría de los ejemplos corresponden a pacientes sanos (63,832) y el resto fueron divididos en 8 clases (como se mostró en la figura 1). El número de imágenes para cada una de estas clases se muestra en la primera fila de la Tabla 1 (nombrada “original”) y en la segunda fila mostramos el número de imágenes que se obtuvieron después del proceso de filtrado descrito en la sección 2.2 y el aumento de datos (nombrado “modificado”).

Tabla 1. Comparación de los datos de ChestX después del aumento de datos y filtrado.

	Cardiomegalia	Efusión	Infiltrado	Masa	Nódulo	Neumonía
Original	9,986	10,235	18,018	3,681	4,965	1,513
Nuestro	10,977	11,217	19,722	4,086	5,541	1,656

Nótese que después de aplicar el aumento de datos, obtuvimos 221,760 imágenes, de las cuales 70,908 fueron descartadas después del filtrado que evaluaba la calidad de la imagen. Por lo tanto, nuestro conjunto de datos modificado, creció a un tamaño de 150,852 imágenes, las cuales fueron utilizadas para el entrenamiento de la red convolucional que se presenta en la siguiente sección.

3. Solución propuesta

Comparado a otros conjuntos de datos para clasificación como COCO e ImageNet, la extensión espacial de las enfermedades en ChestX-ray8 ocupan una región más bien pequeña de la imagen, lo cual puede afectar el rendimiento de la red convolucional cuando se cuenta con poder de cómputo limitado, como es el caso aquí. Los autores de [11] extrajeron las imágenes de rayos-X del archivo DICOM y las ajustaron a un tamaño de 1024 x 1024 píxeles (de su tamaño original de 3,000 x 2,000) sin perder detalles significativos. En nuestro caso, las imágenes del conjunto de datos modificados fueron reducidos todavía más a un tamaño de 224 x 224 píxeles, debido a limitantes del hardware disponible, pero fuimos capaces de obtener resultados satisfactorios, como veremos a continuación.

En las siguientes secciones veremos que tuvimos que hacer compromisos respecto a la profundidad de la red convolucional, para acomodarnos al hardware utilizado. Sin embargo, en la sección 4 veremos que fuimos capaces de obtener y en algunos casos superar los resultados reportados por Wang et al. [11].

3.1. Arquitectura propuesta

La red convolucional implementada en nuestro trabajo se puede ver en la parte superior de la figura 6. Es una arquitectura más bien mínima, ya que está conformada

por 5 capas convolucionales y 3 capas completamente conectadas, con una cuenta de parámetros baja de solamente 254,138.

La arquitectura está basada en AlexNet que tiene un total de 62.4 millones de parámetros y fue simplificada gradualmente hasta que el modelo se ajustó al hardware disponible, un GPU Nvidia GeForce 1050 GTX. Compare con la arquitectura usada en [11] que realizó pruebas en varios modelos pero que obtuvo los mejores resultados con ResNet-50, mostrada en la parte inferior de la figura 6, y la cual tiene 25.6 millones de parámetros.

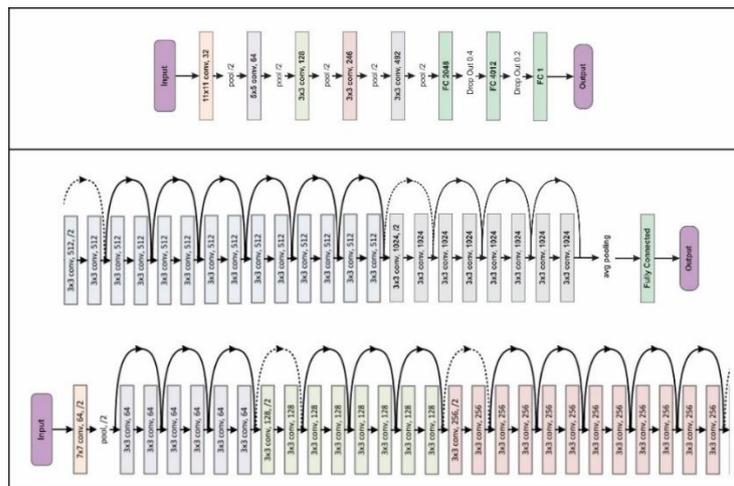


Fig. 6. Arquitecturas para a) la red convolucional propuesta, y b) el sistema en [11] basada en ResNet50.

3.2. Proceso de entrenamiento

Debido al tamaño de la imagen y al límite de memoria del GPU, fue necesario reducir el tamaño de la imagen y del batch para poder cargar la red completa en el GPU e incrementamos el número de iteraciones para acumular los gradientes. La combinación de ambos puede variar en diferentes redes convolucionales, pero en nuestro caso mantuvimos el tamaño del batch constante a 32.

La red fue entrenada de principio a fin usando el optimizador de Adam con los parámetros estándar ($\beta_1 = 0.9$ y $\beta_2 = 0.999$). Usamos un valor inicial de 0.001 para la tasa de aprendizaje que disminuíamos por un factor de 10 cada vez que la pérdida de validación se estancaba después de una época y tomábamos el modelo con la pérdida de validación más baja.

En total, 150,852 imágenes de rayos-X de vista frontal fueron incluidas en la base de datos, de las cuales 87,020 imágenes contenían una o más patologías. Las restantes 63,832 imágenes eran casos normales. Para la tarea de clasificación de la patología, dividimos de manera aleatoria el conjunto de datos completo en tres grupos: entrenamiento (70%), validación (10%), y pruebas (20%). En entrenamiento y validación el modelo fue ajustado finamente con *Stochastic Gradient Descent*. En nuestros experimentos, solo reportamos los resultados obtenidos en la clasificación de las 8 enfermedades sobre los datos de prueba.

4. Resultados y discusión

A continuación, presentamos los resultados obtenidos de nuestros experimentos y los comparamos con enfoques similares, específicamente, los modelos presentados por Wang et al. [11].

4.1. Resultados de entrenamiento y validación

Entrenamos a nuestra CNN para clasificar las 8 enfermedades del tórax presentadas en el conjunto de datos de ChestX-ray8. En la mayoría de los casos, los modelos convergieron rápidamente, debido a la baja complejidad del modelo de nuestra arquitectura CNN, requiriendo menos de 3,000 iteraciones o épocas y obteniendo mejores resultados que la investigación base reportada en [11]. Las gráficas de la figura 7 muestran la exactitud obtenida con los datos de validación en 4 de las 8 enfermedades, la línea roja representa la prueba de nuestro modelo usando los datos originales, no aumentados, mientras que la línea azul muestra los resultados en el conjunto de datos modificado, con aumento de datos.

La exactitud obtenida con los datos de validación para las 8 clases se muestra en la Tabla 2, para ambos casos, los datos originales no aumentados, y con los datos aumentados con imágenes de calidad en el conjunto de datos ChestX-ray8. Como se mencionó anteriormente, podemos observar claramente una mejora significativa en todas las 8 clases, lo cual es especialmente sorprendente dado que nuestra red convolucional es más simple.

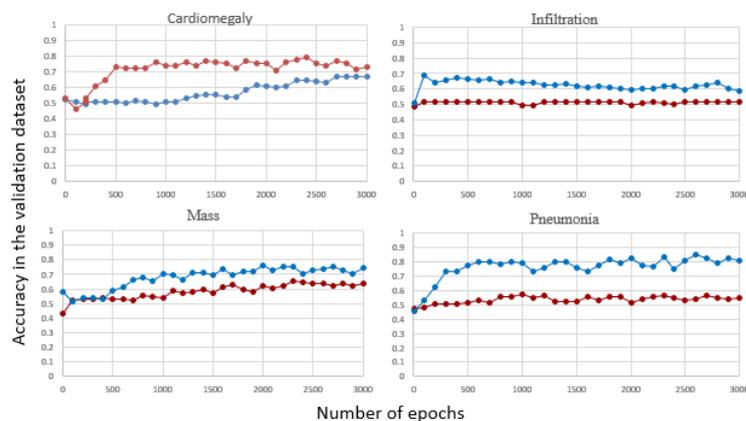


Fig. 7. Resultados de las pruebas de exactitud usando los datos de validación de ChestX-Ray, sin usar aumento de datos (línea en rojo) y con aumento de datos y filtrado (en azul), usando la red propuesta.

Tabla 2. Resultados para exactitud en el mayor de los casos para clasificación múltiple en diferentes modelos de DCNN model settings.

	Cardiomegalia	Efusión	Infiltrate	Mass	Nodule	Pneumonia
No DA	0.7065	0.5966	0.5598	0.6875	0.6449	0.6051
W/ DA	0.8071	0.6471	0.6957	0.7898	0.6989	0.8636

4.2. Comparación con el modelo original de ChestX-ray8

Obtuvimos muy buenos resultados usando una red convolucional relativamente simple inspirada en AlexNet, como se muestra en la Tabla 3. Los autores de [11] probaron 4 redes diferentes con ChestX-ray8: AlexNet, GoogLeNet, VGG-Net-16 y ResNet50. Los mejores resultados se obtuvieron con ResNet50, excepto para “Masa” donde AlexNet obtuvo los mejores resultados. Cuando comparamos nuestro modelo con los resultados reportados para AlexNet y ResNet50, vemos una mejora significativa para todos los casos del primero, excepto “Efusión” y mejores resultados en “Cardiomegalia”, “Infiltración”, “Masa” (una clase particularmente difícil) y “Neumonía” con el segundo. Creemos que el aumento de datos con calidad jugó un papel significativo para alcanzar estos resultados y para un futuro, deseamos investigar si las clases que no obtuvieron los mejores resultados se pueden mejorar con otros modelos de redes o modificando el proceso de filtrado.

Tabla 3. Comparación en términos de exactitud de nuestros modelos con otros modelos del estado del arte.

Red-Clase	Cardiomegalia	Efusión	Infiltrado	Masa	Nódulo	Neumonía
AlexNet	0.6458	0.6925	0.6041	0.5644	0.6487	0.5493
ResNet	0.8071	0.7362	0.6090	0.5609	0.7164	0.6333
Nuestra	0.8141	0.6471	0.6957	0.7898	0.6989	0.8636

4.3. Discusión sobre limitantes y extensiones del enfoque propuesto

Como hemos discutido a fondo en este artículo, el enfoque presentado de aumento de datos basado en la calidad de la imagen ha producido resultados muy interesantes con los datos de ChestX-ray8. Hasta donde sabemos, este enfoque no se ha reportado en la literatura y consideramos que esta es la principal contribución de este trabajo. Como se mencionó en la sección 3.1, hemos hecho uso de una arquitectura relativamente simple de red convolucional, debido principalmente a la disponibilidad de hardware con la que contábamos. Esta limitante, también nos llevó a utilizar imágenes más pequeñas que las reportadas en la literatura (224 x 224, en contraste a imágenes de 1024 x 1024 como se menciona en [11] y otros trabajos).

Sin embargo, los resultados reportados en la subsección anterior son muy prometedores y creemos que nuestro enfoque de aumento de datos se puede poner a prueba usando otras redes convolucionales más complejas, como ResNet-50, una línea que podemos explorar en el futuro utilizando herramientas como Google Colab o los Kernels de Kaggle.

5. Conclusiones

El diagnóstico médico se ha convertido en un dominio más interesante y viable para el aprendizaje automático. En particular, las placas del tórax son el tipo de estudio radiológico más común en el mundo y un ejemplo particularmente retador de clasificación múltiple para el diagnóstico médico. Acumulando cerca del 45% de todos

los estudios radiológicos, la placa del tórax ha alcanzado una ubicuidad global como una herramienta de cribado de bajo costo para una gran cantidad de patologías incluyendo el cáncer de pulmón, la tuberculosis y la neumonía.

Sin embargo, la escasez general de conjuntos de datos médicos, disponibles públicamente, han inhibido el desarrollo de soluciones basadas en aprendizaje profundo en este campo. Además, el preprocesamiento sobre conjuntos de datos como ImageNet o COCO pueden introducir sesgos no intencionados los cuales pueden ser inaceptables en ambientes clínicos, haciendo la transferencia de aprendizaje más difícil de alcanzar en este contexto.

Igualmente, la mayoría de los ambientes clínicos demandan modelos que puedan predecir exactamente múltiples diagnósticos. Esto transforma muchos problemas médicos en problemas de clasificación múltiple, en donde una sola imagen pueda tener un gran número de resultados, los cuales pueden ser ambiguos o pobremente definidos y es probable que la imagen sea etiquetada de manera inconsistente. Estos problemas son expuestos debido a la introducción del conjunto de datos ChestX-ray8 en 2017, ya que los conjuntos de datos disponibles previamente, tales como Open-I eran considerados demasiado pequeños.

Desde entonces, varios trabajos han mejorado los resultados de manera progresiva al diagnosticar enfermedades torácicas basados en este conjunto de datos. Sin embargo, estos trabajos no toman en cuenta la calidad de la imagen de los ejemplos de entrenamiento, lo cual ha producido resultados pobres en trabajos previos, tal como lo demostramos experimentalmente aquí utilizando el enfoque de aumento de datos con imágenes de calidad. Los resultados de nuestros experimentos fueron capaces de igualar los alcanzados en [11] en 4 enfermedades y superarlos en las otras 4, utilizando una red más pequeña. Estamos planeando un estudio más completo donde se evalúen otras redes más complejas tales como ResNet-50 para validar todavía mejor nuestros resultados.

Referencias

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, pp. 1097–1105 (2012)
2. Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Presa Reyes, M., Shyu, M.-L., Chen, S.-C., Iyengar, S.S.: A Survey on Deep Learning: Algorithms, Techniques, and Applications. *ACM Comput. Surv* 51(5), Article 92, 36 pages (2018)
3. Litjens, G., Kooi, T., Ehteshami Bejnordi, B., Adiyoso Setio, A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Medical Image Analysis*, vol. 42, pp. 60–88 (2017)
4. Shin, H., Roth, H., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learnings. *IEEE Trans. Medical Imaging* 35(5), 1285–1298 (2016)
5. Setio, A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S., Wille, M., Naqibullah, M., Snchez, C., van Ginneken, B.: Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks. *IEEE Trans. Medical Imaging* 35(5), 1160–1169 (2016)
6. CDC, 2017, <https://www.cdc.gov/features/pneumonia/index.html>. Último acceso: 2019/05/11

7. WHO. Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children (2001)
8. Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.J., Shamma, D. A., Bernstein, M., Fei-Fei, L.: Visual genome: Connecting language and vision using crowdsourced dense image annotations (2016)
9. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Ball, R.L., Langlotz, C., Shpanskaya, K., Lungren, M. P., Ng, A.Y.: CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. <https://arxiv.org/abs/1711.05225>
10. Demner-Fushman, D., Kohli, M.D., Rosenman, M., Shooshan, B., Sonya, E., Rodriguez, L., Antani, S., Thoma, G.R., McDonald, C.J.: Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association* 23 (2), 304–310 (2015)
11. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *arXiv preprint arXiv:1705.02315* (2017)
12. Chen, S.-F., Chen, Y.-C., Yeh, C.-K., Wang, F., Yu-C.: Order-free rnn with visual attention for multi-label classification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017)
13. Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D., Lyman, K.: Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv preprint arXiv:1710.10501* (2017)
14. Dodge, S., Karam, L.: Understanding how image quality affects deep neural networks. In: *Quality of Multimedia Experience (QoMEX), Eighth International Conference on, IEEE, arXiv:1604.04004v2*, pp. 1–6 (2016)
15. Diamond, S., Sitzmann, V., Boyd, S., Wetzstein, G., Heide, F.: Dirty Pixels: Optimizing Image Classification Architectures for Raw Sensor Data. *arXiv:1701.06487*