

Detección y clasificación de enfisema pulmonar en imágenes de TAC mediante Redes Neuronales Convolucionales Multiescala

D. Bermejo Peláez^{1,2}, R. San José Estépar³, M.J Ledesma-Carbayo^{1,2}

¹ Biomedical Image Technologies, Universidad Politécnica de Madrid, Madrid, España; ² CIBER-BBN, Madrid, España

³ Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Resumen

En este trabajo proponemos y validamos una herramienta para el reconocimiento de patrones de enfisema pulmonar, fenotipo principal de la Enfermedad Pulmonar Obstructiva Crónica (EPOC) en imágenes de TAC. El método propuesto se basa en un Red Neuronal Convolutiva (CNN) Multiescala diseñada para la detección y clasificación de 6 clases de tejido pulmonar, incluyendo 5 patrones de enfisema y tejido normal. La red propuesta consta de 4 capas convolucionales y 3 de submuestreo, y la entrada a la misma corresponde a una representación multiescala de la imagen pulmonar a clasificar. Dicho método ha sido entrenado y validado con un conjunto de datos de 1337 muestras provenientes de 267 escáneres de TAC pulmonar.

1. Motivación

La Enfermedad Pulmonar Obstructiva Crónica (EPOC) es actualmente la cuarta causa de muerte en el mundo, y será la tercera en unos años dada su creciente prevalencia, lo que ha supuesto un creciente interés en estudios poblacionales que permitan entender mejor su fisiopatología y alternativas terapéuticas. La EPOC está caracterizada por una limitación persistente del flujo de aire. Se puede dividir en dos fenotipos principales: la bronquitis crónica y el enfisema pulmonar, caracterizado por la pérdida de la elasticidad pulmonar y el agrandamiento anormal de los espacios aéreos distales a los bronquiolos terminales, juntamente con una destrucción de las paredes alveolares y de los lechos capilares. La progresión de enfisema tradicionalmente ha sido evaluada mediante técnicas espirométricas (FEV1), aunque hoy en día se considera un parámetro no específico y poco sensible en los estadios tempranos de la enfermedad.

Los cambios en la densidad pulmonar, y por tanto, la presencia y progresión del enfisema, pueden ser medidos usando tomografía axial computarizada (TAC) [1]. El análisis densitométrico de TAC está ampliamente aceptado como técnica estándar para la cuantificación de enfisema. Esta técnica consiste en elegir cierto umbral Hounsfield en la máscara del pulmón para discriminar el tejido enfisematoso del no enfisematoso. Esta técnica, a pesar de su gran sensibilidad al ruido y a los parámetros de adquisición, representa el método escogido en la mayoría de los estudios clínicos [2].

Aunque un enfoque densitométrico sí podría cuantificar la extensión de la enfermedad, no podría llevar a cabo una clasificación de la misma. Es por eso que se han propuesto diversos métodos para la clasificación de enfisema basado en información de textura [3-5].

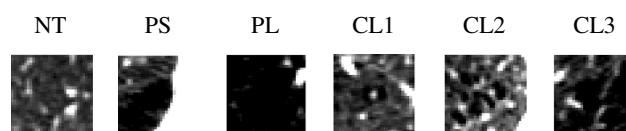


Figura 1. Ejemplo de regiones de los distintos patrones de enfisema a clasificar, en una ventana de $[-945,700]$ (HU). Regiones de tamaño de 31×31 píxeles.

Existen patrones radiológicos prototípicos del tejido enfisematoso correspondiente a enfisema centrilobular, paraseptal y panlobular. Estos patrones llevan a una clasificación de seis patrones y tipos de tejido enfisematoso: tejido normal (NT), enfisema paraseptal (PS), panlobular (PL), y centrilobular leve, moderado y severo (CL1,CL2,CL3). En la Figura 1 se muestran ejemplos de imágenes de TAC de cada una de las clases descritas, con sus correspondientes patrones radiológicos.

2. Métodos

2.1. Aproximación al problema

La metodología propuesta está basada en el etiquetado de regiones de interés (ROIs) bidimensionales. La definición de la extensión física de dichas regiones es crítica para el éxito de la clasificación, ya que una ROI demasiado pequeña puede no contener un lóbulo secundario completo (unidad fundamental del parénquima pulmonar), y una ROI demasiado grande puede corresponder a una región con dos clases de tejidos diferentes. Tras llevar a cabo diferentes experimentos para la correcta definición de la extensión apropiada de estas regiones, se ha corroborado el tamaño óptimo propuesto por L. Sørensen et al. [6] – $24,18 \times 24,18$ mm².

2.2. Método de clasificación de referencia

El método de referencia de este trabajo está basada en el propuesto por Mendoza et al. en [7], que fue implementado y mejorado en [8]. Este método se basa en la discriminación de las distintas clases de enfisema mediante las distribuciones locales de intensidad, estimadas mediante el método de Estimación Kernel de la función de Densidad (KDE), y clasificadas posteriormente mediante un clasificador KNN.

En [7] los autores mostraron que la discriminación entre clases de enfisema se puede llevar a cabo mediante la distribución de intensidad de la imagen a clasificar, obteniendo siempre mejores resultados que cuando se usaban características más complejas como LBP, LBPINT o INT.

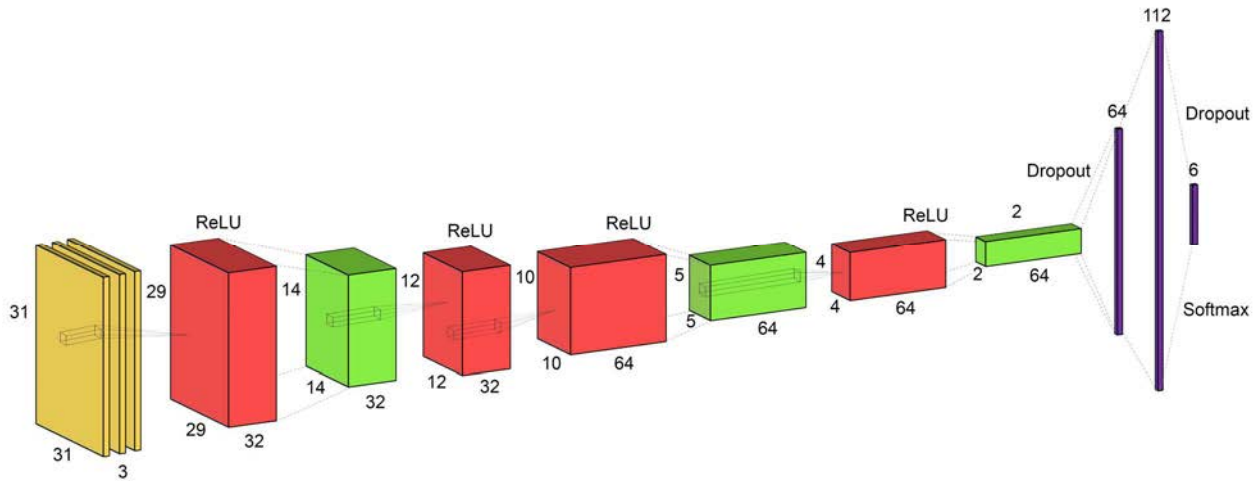


Figura 2. Arquitectura de la CNN multiscala propuesta con 4 capas convolucionales (rojo) y 3 de submuestreo (verde).

2.3. Red Neuronal Convolucional Multiescala

Las CNNs desde sus primeras implementaciones ya demostraron un excelente rendimiento en tareas como el reconocimiento y clasificación de imágenes. Estas arquitecturas profundas están inspiradas en procesos biológicos como el sistema visual, y están diseñadas para llevar a cabo un reconocimiento de patrones a partir de los propios datos de entrada incorporando tanto la etapa de extracción automática de características como la de clasificación. Gracias a este aprendizaje y extracción automática de las características, no es necesario una definición y extracción manual de las mismas.

Arquitectura

Como sabemos, el sistema visual humano es un proceso multiescala. En este trabajo proponemos una Red Neuronal Convolucional Multiescala (M-CNN), donde la red pueda aprender los descriptores o características óptimas de los datos de entrada no sólo al nivel de escala de la imagen original, sino también a otros niveles. De esta forma la entrada a la red se compone de una representación multiescala a diferentes niveles de la imagen a clasificar. Se hará uso de dos escalas adicionales a parte de la determinada por la imagen original. Las versiones escaladas corresponden a versiones filtradas mediante filtros gaussianos según la siguiente ecuación:

$$I_{\sigma} = I * G_{\sigma}, \sigma \in \{0.3, 0.8\}$$

La arquitectura de la CNN Multiescala propuesta se muestra en la Figura 2, compuesta de 4 capas convolucionales y 3 de submuestreo.

La entrada a la red es una imagen de tamaño 31x31x3, donde se reflejan las 3 representaciones a escalas diferentes, que es convolucionada en la primera etapa con un banco de 32 filtros de tamaño 3x3 para captar información local de la imagen. Estas primeras capas serán las encargadas de aprender las características de más bajo nivel de los datos de entrada. A esta capa convolucional, después de aplicarle la transformación no lineal ReLU $f(x) = \max(0, x)$, le sigue una capa de agrupamiento de tipo *maxpooling* con un campo receptivo de 2x2, donde reduciremos la dimensión de las características a la mitad. A estas capas les siguen una sucesión de capas

convolucionales y de agrupamiento, donde se extraen características distribuidas en una jerarquía, desde niveles locales hasta niveles con mayor abstracción.

Las características resultantes son introducidas en una sucesión de 3 capas densamente conectadas compuestas de 64, 112 y 6 neuronas respectivamente, ya que 6 es el número de clases considerado.

El entrenamiento de la red se basa en un problema de optimización para la minimización de la función de coste, en este caso la entropía cruzada. En este trabajo usamos el método de Descenso de Gradiente Estocástico (SDG) con una actualización de momentos de Nesterov [9].

Prevención del sobreajuste

El problema del sobreajuste de los parámetros de la red a los datos de entrenamiento se ha abordado mediante la implementación de cuatro técnicas diferentes.

En primer lugar se ha llevado a cabo una regularización de la función de coste. Mediante la regularización L2 se penaliza la magnitud al cuadrado de todos los pesos de la red (ω) directamente en la función de coste añadiendo un término regularizador de valor $\frac{1}{2} \lambda \omega^2$, donde λ es el parámetro regularizador. Este método penaliza los cambios abruptos dando preferencia a cambios suaves.

La segunda técnica empleada es el aumento de datos. A mayor número de datos con los que entrenar la red, menor será el sobreajuste debido a la mayor variabilidad. Se ha llevado a cabo una técnica de aumento de datos artificial usando 7 transformaciones diferentes sobre el conjunto de datos de entrenamiento gracias a transformaciones de rotación, inversión y combinación de ambas.

Otra técnica es la interrupción anticipada, donde se detiene la fase de entrenamiento antes de que comience el proceso de sobreajuste.

Por último, se ha reducido el sobreajuste de la red mediante la técnica de *dropout* [10], que consiste en desconectar aleatoriamente ciertas neuronas a lo largo de la arquitectura con una probabilidad p , o dejarlas activadas con probabilidad $1-p$ durante la etapa de entrenamiento. Esta práctica es equivalente a entrenar un conjunto de redes que

contenga todas las posibles configuraciones de dicha red. Durante la etapa de validación se hará uso de la arquitectura original con todas las neuronas activas, de forma que el resultado equivaldrá a un promediado de los resultados de todas las posibles configuraciones encontradas.

3. Experimentos y resultados

El conjunto de datos usado en este trabajo para el entrenamiento y validación del método propuesto se compone de 1337 ROIs etiquetadas manualmente entre las 6 clases bajo estudio por neumoradiólogos experimentados. Todas estas muestras provienen de 267 estudios de TAC pulmonares del estudio COPDGene.

3.1. Prevención del sobreajuste

La Figura 3 muestra los resultados obtenidos en la minimización de la función de coste durante el entrenamiento (*epochs*) tanto en el conjunto de datos de entrenamiento como el de validación para el método propuesto aplicando y sin aplicar las técnicas descritas para la prevención del sobreajuste. Como se observa, obtenemos un mejor desempeño de la red eliminando el sobreajuste cuando estas técnicas son aplicadas.

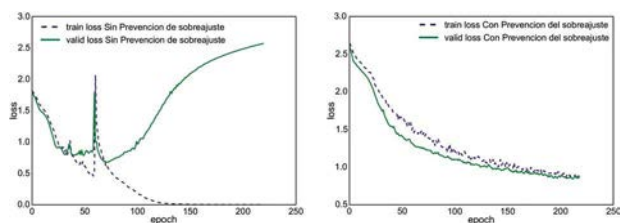


Figura 3. Diferencia en la minimización de la entropía cruzada al aplicar las técnicas de prevención del sobreajuste. (a) Sin aplicarlas. (b) Aplicadas.

3.2. Evaluación

La evaluación del método propuesto se basa en un esquema de entrenamiento-validación-test, donde 1003 muestras se usarán para el entrenamiento y validación del método, y 334 para el test final. Concretamente evaluaremos el rendimiento del método mediante validación cruzada aleatoria de 100 iteraciones, donde un 75% de las 1003 muestras componen el conjunto de entrenamiento, y el 25% restante el de validación.

Red Convolutiva Multiescala

En primer lugar se ha querido evaluar y comprobar que la introducción de tres representaciones a distintas escalas de un mismo parche a clasificar mejora el rendimiento de la red. Como se desprende de la Tabla 1, la CNN Multiescala obtiene una mayor precisión global de clasificación en las 6 clases consideradas frente a la misma arquitectura del método propuesto con entradas monoescala.

Arquitectura	Precisión [media (sd)]
CNN Monoescala	0.825 (0.025)
CNN Multiescala	0.854 (0.020)

Tabla 1. Diferencia entre la precisión de clasificación de la CNN monoescala y la multiescala para el problema de clasificación de enfisema.

Análisis del rendimiento del método

En la Tabla 2 se muestra la comparación del rendimiento del método de referencia (KDE-KNN) y de la CNN Multiescala (M-CNN) propuesta en este trabajo para el mismo problema de clasificación con los mismos datos y la misma metodología de evaluación. La Figura 4 muestra las matrices de confusión para ambos métodos. Como podemos observar, el método propuesto aumenta la precisión de clasificación en todas las clases bajo estudio y reduce considerablemente el error cometido por el método de referencia en la clasificación de las clases CL1, CL2 y CL3.

Método	Precisión [media (sd)]	95% CI [LL,UL]
KDE-KNN	0.672 (0.015)	[0.648532, 0.696318]
M-CNN	0.854 (0.020)	[0.849476, 0.858548]

Tabla 2. Diferencia entre la precisión de clasificación del método propuesto y del método de referencia.

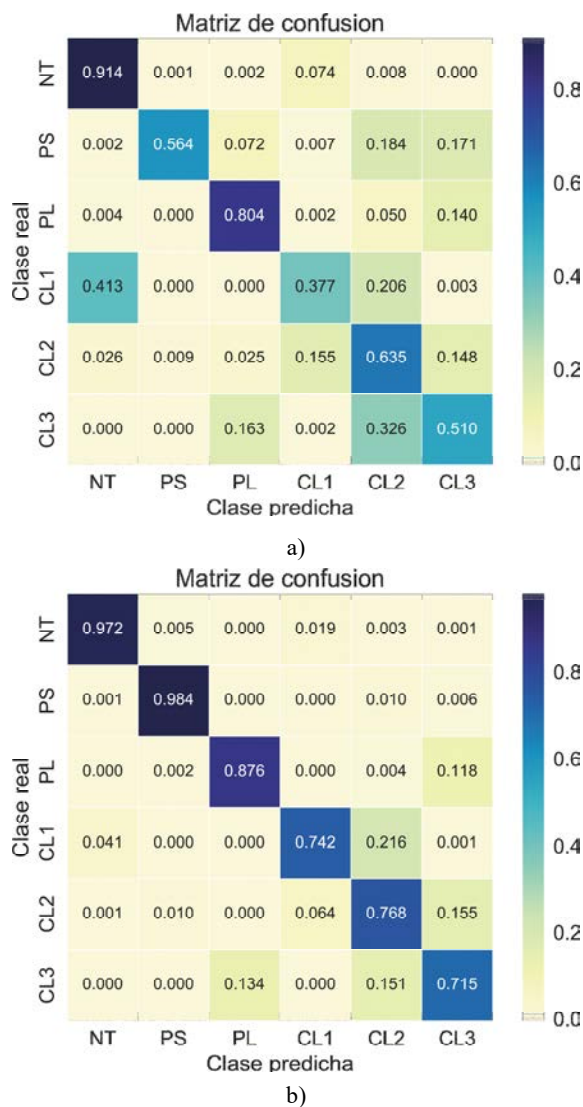


Figura 4. Matrices de confusión del método de referencia (a) y del propuesto (b).

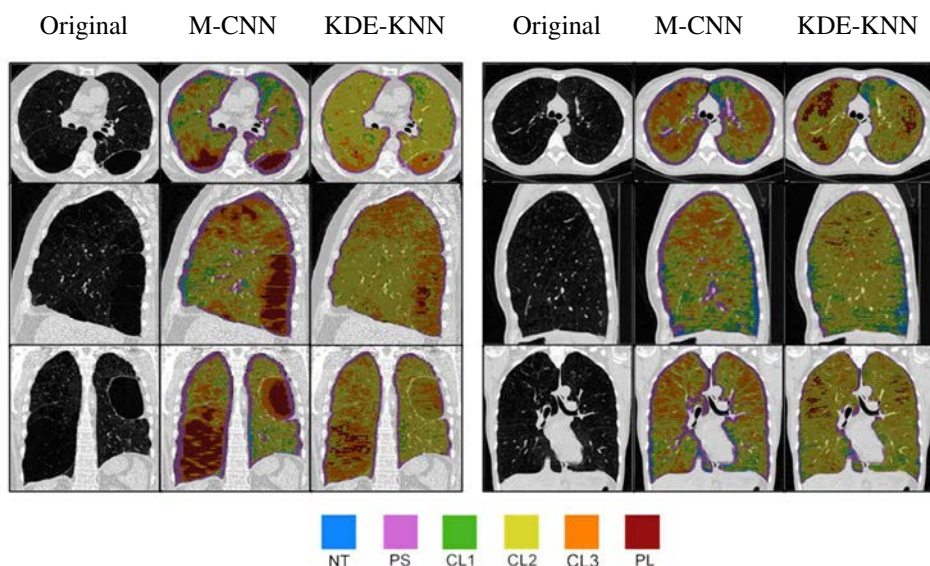


Figura 5. Resultados de clasificación para dos casos distintos de TAC pulmonar. (Primera columna) Imágenes TAC originales. (Segunda Columna) Resultados con el método propuesto. (Tercera columna) Resultados con el método de referencia.

3.3. Clasificación de pulmón completo

Para un mayor análisis, debido a que en problemas de clasificación de imágenes tiene un valor adicional evaluar los resultados a un nivel global analizando cómo se comportan los algoritmos clasificando imágenes completas, en la Figura 5 se muestran los resultados obtenidos usando ambas metodologías, el método de referencia y el propuesto en este trabajo, en la clasificación de dos casos de imágenes de TAC completas. La clasificación se ha llevado a cabo en una rejilla de muestreo de la imagen original de 5x5 píxeles en cada corte axial, donde el resto de vóxeles se han clasificado usando una interpolación del vecino más cercano. Se ha comprobado visualmente por expertos el mejor desempeño del método propuesto en la clasificación de pulmones completos.

4. Conclusiones

En este trabajo proponemos una Red Neuronal Convolutiva Multiescala para la clasificación de imágenes de TAC pulmonares en 6 clases, incluyendo 5 patrones de enfisema y tejido normal. Una nueva arquitectura se ha diseñado para capturar características de textura del tejido pulmonar en varias escalas diferentes. La arquitectura propuesta consta de 4 capas convolucionales con activaciones ReLU y 3 de submuestreo tipo *maxpooling*. El entrenamiento de la red se ha llevado a cabo minimizando la entropía cruzada mediante SDG con actualización de momentos de Nesterov. El enfoque propuesto proporciona resultados prometedores, consiguiendo una precisión global de clasificación del 85.4% en las 6 clases bajo estudio, habiendo sido evaluado en una amplia base de datos. La validación del método se ha llevado a cabo en un conjunto de 1337 muestras y mediante clasificación de imágenes de pulmón completo.

Agradecimientos

Este trabajo está parcialmente financiado por el proyecto TEC2013-4851-C2-R2 del Ministerio de Economía y Competitividad.

Referencias

- [1] A. Dirksen, et al. Exploring the role of CT densitometry: a randomised study of augmentation therapy in 1-antitrypsin deficiency. *European Respiratory Journal*, vol. 33, no. 6, pp. 1345-1353, 2009.
- [2] E. Cavigli, et al. Whole-lung densitometry versus visual assessment of emphysema. *European radiology*, vol. 19, no. 7, pp. 1686-1692, 2009.
- [3] R. Uppaluri, et al. Quantification of pulmonary emphysema from lung computed tomography images. *American Journal of Respiratory and Critical Care Medicine*, vol. 156, no. 1, pp. 248-254, 1997.
- [4] A. Depeursinge, et al. Lung tissue classification using wavelet frames, *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007*, pp. 6259-6262, 2007.
- [5] Y. S. Park, et al. Texture-based quantification of pulmonary emphysema on high-resolution computed tomography: Comparison with density-based quantification and correlation with pulmonary function test. *Investigative radiology*, vol.43, no. 6, pp.395-402, 2008.
- [6] L. Sørensen, et al. Quantitative analysis of pulmonary emphysema using local binary patterns, *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 559-569, 2010.
- [7] C. S. Mendoza, et al. Emphysema quantification in a multi-scanner HCRT cohort using local intensity distributions. *Proc IEEE Int Symp Biomed Imaging*, pp. 474-477, 2012.
- [8] D. Bermejo, et al. Cuantificación de enfisema pulmonar a partir de imágenes de TAC para la determinación de calcio coronario, *Actas del XXXIII Congreso Anual de la Soc. Esp. Ing. Biomédica*, pág. 285-288, 2015.
- [9] Y. Bengio, et al. Advances in optimizing recurrent networks, *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8624-8628, 2013.
- [10] N. Srivastava, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning*, pp. 1929-1958, 2014.