

# Método automático de análisis y segmentación de imágenes para la detección de nódulos pulmonares a partir de radiografías de tórax

J. Naranjo Alcázar<sup>1</sup>, I. Bosch Roig<sup>2</sup>, R. Sanz Requena<sup>3</sup>, S. Vázquez Martínez<sup>2</sup>

<sup>1</sup> ETSIT, Universidad Politécnica de Valencia, Valencia, España, janaal@teleco.upv.es

<sup>2</sup> Departamento de Comunicaciones, Universidad Politécnica de Valencia, Valencia, España, igbosroi@dcom.upv.es

<sup>3</sup> Ingeniería Biomédica, Hospital Quirónsalud Valencia, Valencia, España, roberto.sanz@quironsalud.es

## Resumen

*El cáncer de pulmón es uno de los tipos de cáncer con mayor tasa de mortalidad, por lo que es necesaria la investigación en sistemas que permitan detectar la enfermedad de forma temprana y reducir la mortalidad de forma significativa. Por otro lado, la radiografía de tórax es una de las técnicas más habituales a la hora de evaluar el pulmón. Sin embargo, la utilización de esta prueba de forma masiva presenta una importante limitación: el volumen de trabajo de los servicios de radiología dificulta el análisis exhaustivo y personalizado de cada una de las radiografías. Por tanto, el objetivo planteado en este trabajo de investigación es el desarrollo de una metodología automática de análisis de imágenes que permita realizar un cribado de las radiografías, de forma que sólo lleguen al radiólogo aquellas imágenes con cierta probabilidad de presentar nódulos.*

## 1. Motivación

La motivación de este proyecto surge ante las nuevas temáticas de investigación de automatización de procesos médicos, en concreto mediante el tratamiento de imágenes. Así pues, surge la posibilidad de automatizar la detección de posibles nódulos pulmonares para que los servicios de radiología de los hospitales tengan menos volumen de trabajo y que así sólo trabajen sobre casos que presentan alguna posibilidad de poseer algún tipo de lesión.

El objetivo de este proyecto es aplicar ciertas técnicas de procesado de imagen para poder segmentar las zonas de interés en cada momento, hasta finalmente obtener únicamente el nódulo. Con esta metodología se intenta investigar si dichas técnicas son de provecho en este problema o no. Para la comprobación de dicha metodología, se ha trabajado sobre 8 casos con diferentes características, como pueden ser el fabricante del equipo de adquisición, calidad de la imagen o colocación del nódulo.

## 2. Metodología

Se han desarrollado diferentes algoritmos de procesado de imagen en función de las características de los diversos casos de estudio, para finalmente implementar una algorítmica estándar que permita la detección de los nódulos de forma independiente a cada caso.

En primer lugar, para poder detectar el nódulo debemos separar la información o parte de la imagen que nos interesa del resto. En este caso, el objetivo será segmentar

los dos pulmones del resto de la imagen. Una vez tengamos los pulmones segmentados, podremos realizar un procesado para poder detectar el nódulo.

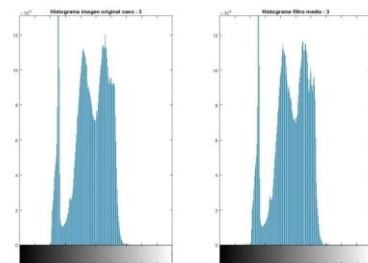
Así pues, la estrategia que se ha seguido para solucionar este problema es la segmentación de los pulmones basándonos en las intensidades de los píxeles.

### 2.1. Segmentación de los pulmones

En primer lugar el algoritmo extrae información del caso que este procesando. Hay que resaltar que se está trabajando con imágenes en formato DICOM, que es una estructura en la cual existe también información respecto al paciente, las condiciones de adquisición de la imagen, etc. La información relevante en nuestro caso es la relativa a la profundidad de bit, el equipo con el que se ha realizado la imagen y la distancia entre píxeles.

Para normalizar la imagen necesitamos la profundidad de bit y esta información la encontramos en la etiqueta *BitDepth* dentro de la estructura DICOM. Así pues, el primer procesado que realizamos sobre la imagen es su normalización, dividiéndola por su profundidad de bit. Así conseguiremos que el histograma se encuentre entre 0 y 1.

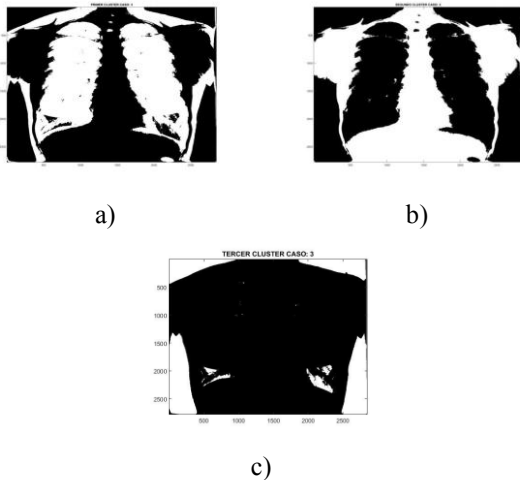
El siguiente paso es suavizar el histograma por si aparecieran picos no deseados reduciendo ruido. Para ello, se ha utilizado un filtro paso bajo (FPB) de media 3x3 para que el píxel en el cual se está aplicando el filtro obtenga el valor de la media de los píxeles que le rodea.



*Figura 1. Comparación de los histogramas de un caso pre y post procesado con el FPB.*

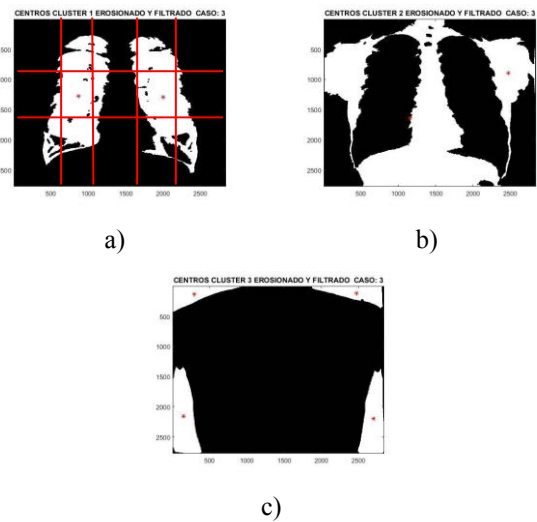
El histograma de este caso es el ideal, ya que podemos observar tres grandes máximos bien diferenciados. Cada uno de ellos corresponde a tres partes claramente separadas

en la radiografía que son: fondo (aire), pulmones y huesos y tejidos blandos. Esta propiedad se explotará mediante la separación de estas tres estructuras en *clusters* diferentes. La separación se aplicará por diferentes valores de intensidades. Esta separación de elementos se ha decidido realizarla de esta manera, ya que se ha observado en ciertos artículos la segmentación de estructuras observando los elementos que aparecen y el histograma de la imagen [1]. El método que se ha utilizado para la separación de la imagen en 3 clusters o grupos diferentes es *k-means*. Este método separa la imagen en diferentes clusters optimizado su nivel de intensidad [2]. También se utilizó la función *fuzzy c-means*, pero se descartó ya que el tiempo de computación requerido era superior y los clusters generados eran muy parecidos. Esta función nos devuelve el valor del cluster al que pertenece cada pixel. Los tres clusters que obtenemos son los mostrados en la Figura 2:



**Figura 2.** Segmentación del caso 3 a) Cluster 1, pulmones b) Cluster 2, huesos c) Cluster 3, fondo.

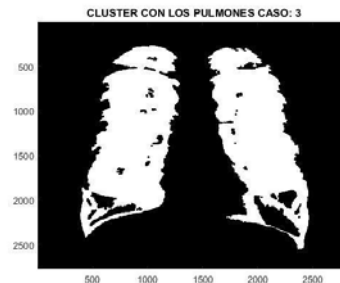
El siguiente objetivo es obtener el cluster de los pulmones de forma automática ya que, al ser una función de optimización, los clusters cambian de orden según la ejecución. Por lo tanto, el primer procesado que se realiza es una pequeña erosión para separar elementos de la imagen que quedan conectados en los clusters obtenidos. Por ejemplo, en ciertos casos, la intensidad de la región supraclavicular es muy parecida a la de los pulmones y dichos elementos aparecen en el mismo cluster. Así pues, realizamos este proceso para intentar separar ambos elementos. Posteriormente, se realiza una búsqueda de los centros de todas las estructuras presentes en la imagen para decidir qué centros pueden ser los de los pulmones según su posición. Así pues, se aplica un filtrado de área para eliminar todas las estructuras que presenten un área menor al pulmón, se comprobó en todos los casos el tamaño de los pulmones y el umbral se colocó en el más restrictivo. El siguiente paso será calcular los centros de las estructuras presentes en los tres clusters.



**Figura 3.** Centros tras la erosión y filtrado después de la segmentación del caso 3 a) Cluster 1, pulmones b) Cluster 2, huesos c) Cluster 3, fondo.

En la Figura 3, se puede apreciar cuál va a ser la tendencia de los centros de cada cluster. En el cluster 1 se pueden observar los centros de los pulmones (marcados con asteriscos rojos en la Figura 3 a)). Dichos centros se encuentran a la misma altura respecto al eje horizontal y aproximadamente a la misma distancia respecto a la mitad del eje vertical. Así pues, definimos unos márgenes en los cuales deberían encontrarse los centros de ambos pulmones. Estos márgenes se han representado en la Figura 3 a) mediante líneas rojas.

Una vez tenemos de forma lógica qué elemento corresponde a cada pulmón, debemos sumar ambas comprobaciones (pulmón izquierdo y derecho) y ver qué cluster tiene una suma de dos elementos, que corresponden a los dos pulmones. En caso de que exista más de un cluster seleccionado como el que posee los pulmones, se repite el proceso de los márgenes, pero de una forma más restrictiva, reduciendo los márgenes. Una vez sumados todos los elementos que se encuentran en dichos márgenes en cada cluster, tenemos que ver qué cluster tiene dos elementos (por eso la repetición del proceso si hay varios clusters que posean dos elementos que pudieran ser los pulmones).



**Figura 4.** Cluster en el cual se encuentran los pulmones detectados de forma automática, caso 3.

Una vez se ha seleccionado el cluster de los pulmones, es necesaria la segmentación de los pulmones para su procesado (en el caso 3 no es necesario este proceso, ver Figura 4). Ahora que tenemos el cluster y las coordenadas

de los centros de los pulmones identificadas, es muy fácil quedarnos únicamente con los pulmones. Este proceso es fundamental, ya que pueden existir casos en los cuales el filtrado por área no elimine el resto de elementos distintos a los pulmones. En la Figura 5, se muestra el caso 4 en el cual detectamos el cluster de los pulmones satisfactoriamente, pero no segmentamos únicamente los pulmones (ver Figura 5 a)). Para eliminar dichos elementos en todos los casos, se calcula la distancia de cada uno de los elementos del cluster respecto a los centros de los pulmones. Como conocemos los centros de los pulmones gracias a los márgenes, imponemos la condición de que, únicamente, se quede con las estructuras que presenten distancia 0 a dichos centros. El resultado es el mostrado en la Figura 5 b).

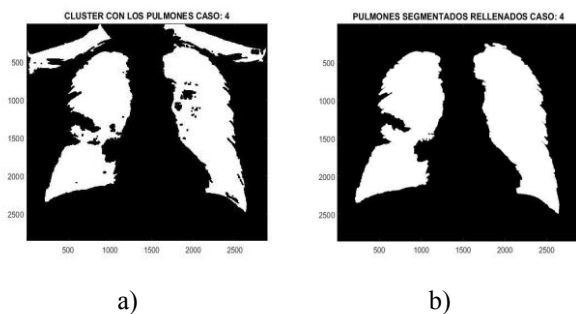


Figura 5. Pulmones finalmente segmentados caso 4, procesado de centros y de relleno para recuperar información.

## 2.2. Procesado morfológico y segmentación del nódulo

Como se ha visto anteriormente, a la hora de segmentar los pulmones en el caso 3, no se perdía información relevante sobre los mismos. Sin embargo, como el algoritmo tiene que ser estándar para todos los casos, vamos a realizar un procesado morfológico tipo *close*, es decir, dilatar la imagen para luego erosionarla y así recuperar elementos que se han perdido a la hora de la segmentación de imagen. El objetivo de este procedimiento es la recuperación del nódulo en caso de pérdida de éste. En la Figura 6, se muestra la mejora en caso de que el nódulo se encuentre en el borde del pulmón (caso 4). A la izquierda, se pueden observar los pulmones segmentados sin el procesado morfológico y a la derecha tras dicho procesado. Se puede apreciar cómo conseguimos recuperar el nódulo y elementos que se perdieron tras la segmentación.

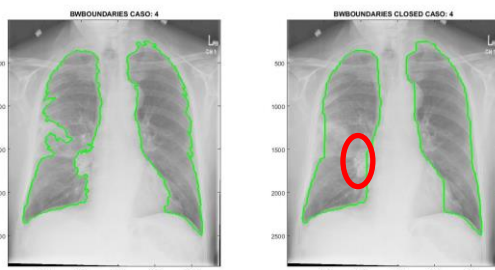


Figura 6. Pulmones segmentados pre y post procesado morfológico caso 4, el nódulo se encuentra resaltado con una elipse roja.

El siguiente paso tras el procesado de los pulmones es la separación de los diferentes elementos que aparecen en ellos. Lo que haremos será un clustering únicamente de los pulmones. El problema que tenemos ahora es que las intensidades que aparecen en los pulmones son muy parecidas, es decir, el histograma de los pulmones está muy concentrado en valores de intensidad similares. Por eso, estamos obligados a expandir y ajustar el histograma de los pulmones. Se ha decidido volver a utilizar *k-means*, pero en este caso que devuelva cinco clusters diferentes. Cada cluster está relacionado con los siguientes elementos: fondo de la imagen (también tiene que ser procesada), costillas, pulmón, ruido y nódulos. El proceso de optimización es el mismo que a la hora de segmentar los pulmones.

Una vez tenemos los elementos segmentados, debemos detectar de forma automática en qué cluster aparece el nódulo. Para ello, nos aprovecharemos del ajuste del histograma previamente realizado. Sabemos que los nódulos son elementos claros. Además, una vez expandido el histograma, forzamos a que el nódulo tenga una intensidad cercana al blanco. Por tanto, multiplicaremos cada cluster por la imagen de los pulmones con el histograma expandido y nos quedaremos con el cluster que presente la intensidad más clara. Una vez tenemos el cluster donde se encuentra el nódulo, vamos a eliminar aquellos elementos que no nos proporcionen información relevante. Para ello, eliminaremos todas las estructuras menores a 1 cm de diámetro (este margen viene definido por criterios médicos). Nuestro siguiente objetivo va a consistir en separar los bordes de los pulmones con los elementos que están en contacto con ellos, para así poder detectar los nódulos que se encuentran en la parte periférica del pulmón (caso 4, ver Figura 6). Podríamos calcular los dos mayores perímetros y eliminarlos, pero en ese caso podríamos estar perdiendo información referente al propio nódulo. Por lo tanto, vamos a realizar una pequeña erosión para poder separar los bordes de las estructuras contiguas.

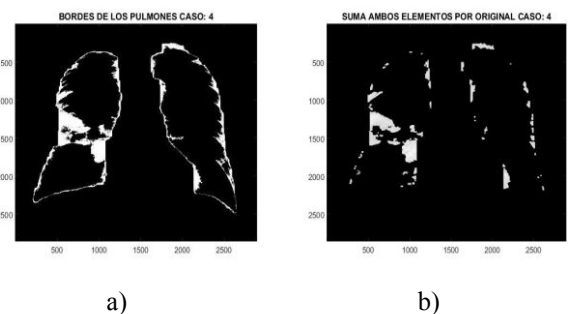


Figura 7. Detección borde de los pulmones caso 4, se aprecia cómo el nódulo se encuentra pegado a dichos bordes en la imagen de la izquierda. En la imagen de la derecha vemos como el nódulo es ya una estructura independiente.

Una vez ya tenemos los diferentes elementos por separados debemos buscar el nódulo. Para ello, lo primero que haremos será repetir el proceso de eliminación de elementos que poseen menos de un centímetro de diámetro. Lo que realizaremos sobre los elementos será un filtrado por forma y nos quedaremos con los elementos circulares (nódulos). El primer filtrado se realizará a partir de la

excentricidad (relación de eje mayor respecto eje menor) y de la solidez, propiedad que nos proporciona el cociente entre el área del casco convexo de los elementos y el área que poseen realmente [3]. Una vez que tenemos los posibles nódulos, vamos a utilizar otra propiedad que poseen los círculos para realizar un filtrado más exhaustivo. La propiedad que utilizaremos es la circularidad [4]. Por último, lo que se realiza es un cálculo de círculos mediante el algoritmo de Hough [5].

### 3. Resultados

En la siguiente tabla (Tabla 1) se detallan los resultados obtenidos en cada caso. Se puede observar la presencia de falsos positivos en casos con histogramas ruidosos o en los cuáles la segmentación de los pulmones no ha sido del todo satisfactoria. Por otro lado, la segmentación de los pulmones presenta resultados distintos, como la segmentación total, la pérdida del nódulo ya que se encuentra en la periferia o la adición de estructuras que no son el pulmón.

	Nódulo	Calidad imagen	Segmenta pulmones	Detecta nódulo
<b>Caso 1</b>	No	Histograma ruidoso	Sí, pérdida de información en un pulmón	Falso positivo
<b>Caso 2</b>	Sí	Histograma ruidoso	Sí, pérdida del nódulo	Detecta estructura diferente al nódulo
<b>Caso 3</b>	Sí	<b>Buena</b>	<b>Sí</b>	<b>Sí</b>
<b>Caso 4</b>	Sí	<b>Buena</b>	<b>Sí</b>	<b>Sí</b>
<b>Caso 5</b>	Sí	Colocación incorrecta del paciente	No	Falso negativo
<b>Caso 6</b>	No	<b>Buena</b>	<b>Sí</b>	<b>No</b>
<b>Caso 7</b>	Sí	<b>Buena</b>	<b>Sí</b>	Detecta estructura diferente al nódulo
<b>Caso 8</b>	No	Buena	Sí, incluye otras estructuras	Falso positivo

Tabla 1. Resultados obtenidos en los ocho casos trabajados.

En cuanto a los resultados obtenidos, podemos afirmar que la segmentación de los pulmones es una labor muy tediosa, ya que existen muchísimas variables que pueden modificar dicho examen. Entre estos factores, podemos destacar la radiación empleada (puede modificar la tonalidad de la imagen, es decir, el histograma), la colocación del paciente (pueden aparecer elementos no deseados, como la barbilla, entre otros), el fabricante del equipo (también puede modificar el histograma y que éste no sea suave) o si la imagen fue tomada en formato digital o fue escaneada, en este último caso, la imagen suele ser ruidosa. Por otro lado, en la detección de los nódulos, se han tenido que añadir ciertas consideraciones para detectar el mayor número de éstos, como la erosión de los bordes, en caso de que el nódulo estuviese en contacto con ellos (Figura 7), o el procesado de *close* morfológico necesario para poder

recuperar la información perdida tras la primera segmentación, ver Figura 6. Finalmente, para la selección del nódulo se ha decidido explotar la característica de la circularidad de éstos mediante tres criterios que comprueban dicha propiedad. En la Figura 8 se muestra la detección del nódulo en dos posiciones distintas del mismo (casos 3 y 4) y la no detección en un caso sano (caso 6).

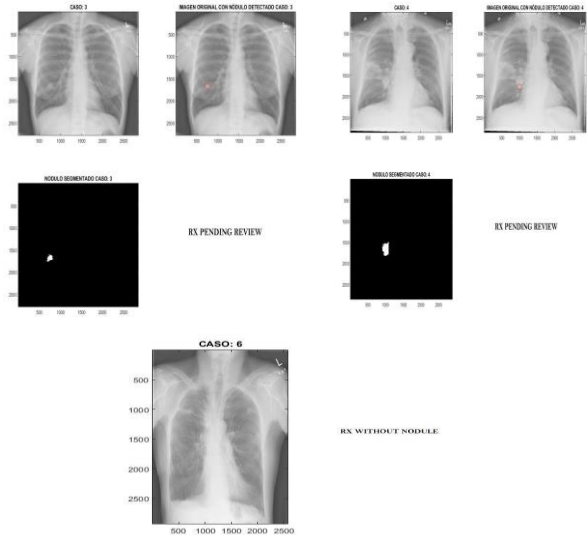


Figura 8. Resultados del algoritmo en tres casos, resueltos satisfactoriamente, caso 3 (nódulo rodeado de intensidad de grises, pulmón), caso 4 (nódulo en contacto con la columna) y caso 6 (sin nódulo, el algoritmo informa de que no existen nódulos)

### 4. Conclusiones

Como conclusión final, destacamos que el algoritmo diseñado es capaz de detectar nódulos correctamente en el caso de radiografías que cumplan una serie de características, como una buena colocación del paciente y una calidad de imagen mínima. Como principales líneas futuras, se plantean la independización de la metodología desarrollada de las diferentes características de calidad y la ampliación del número de estudios, para ir incorporando las singularidades de cada uno de ellos a la algorítmica desarrollada.

### Referencias

- [1] Dolejsi M. Detection of Pulmonary Nodules from CT Scans, CZECH TECHNICAL UNIVERSITY, 2007.
- [2] Abdel-Maksoud E, Elmogy M, and Al-Awadi R. Brain tumor segmentation based on a hybrid clustering technique. *Egypt. Informatics J.*, vol 16, sup 1, 2015, pp 71–81 (ISSN: 1110-8665).
- [3] Horejš J. Shape Analysis Using Global Shape Measures. *Proceedings of the 18th Annual. Conf. Tech. Computing. Bratislava*, Bratislava, 2010, pp. 1–6 (ISBN 978-80-970519-0-7).
- [4] Hirota K, Rosin Paul L, and Žunić J, A Hu moment invariant as a shape circularity measure, *Pattern Recognition*, vol 40, 2009, pp. 1–26 (ISSN: 0031-3203).
- [5] D. H. Ballard and C. M. Brown, “Hough Method for Curve Detection.pdf,” *Computer Vision*. pp. 123–131.