

Segmentación automática del hígado en imágenes de tomografía computarizada

Automated liver segmentation on computed tomography images

Lázaro Ramón Caraballo Rodríguez¹, Marlen Pérez Díaz^{2*}, Rubén Orozco Morales³

Resumen La segmentación automática de hígado facilita la detección de anomalías hepáticas a partir de imágenes de tomografía computarizada. La segmentación manual de estructuras y órganos no alcanza los niveles de precisión con los que cuentan los modernos sistemas automáticos 3D. Esta investigación se basa en encontrar una alternativa eficaz y eficiente para la realización de esta tarea en rutina clínica. El objetivo es comparar tres métodos de segmentación automática del hígado, a partir de imágenes de tomografía. Los métodos comparados fueron *k-means plus*, campos aleatorios ocultos de Markov y red neuronal de mapas auto-organizados. Para la comparación se realizó un análisis estadístico y computacional de los resultados de la segmentación. Se utilizaron como métricas de calidad de la segmentación el coeficiente de Dice, la distancia promedio de superficie promedio y la distancia superficial simétrica máxima en 20 tomografías abdominales. De los tres métodos analizados, se comprobó que todos lograron una correcta segmentación, pero *k-means plus* consiguió una mejor precisión con menos tiempo de cómputo y menor tasa de fallos. Se concluye que *k-means plus* puede ser adecuado para su utilización en la rutina clínica.

Palabras Clave: hígado, procesamiento digital de imágenes, segmentación, tomografía computarizada de abdomen.

Abstract *Automatic liver segmentation facilitates the detection of liver abnormalities from computed tomography images. Manual segmentation of structures and organs does not reach the precision levels of modern 3D automatic systems. This research is based on finding an effective and efficient alternative to perform this task in clinical routine. The objective is to compare three methods of automatic liver segmentation from tomography images. The methods compared were k-means plus, hidden Markov random fields and neural network of self-organizing maps. Statistical and computational analysis of the segmentation results was performed for comparison. Dice coefficient, average average surface distance and maximum symmetric surface distance in 20 abdominal tomographies scans were used as segmentation quality metrics. Of the three methods analyzed, all were found to achieve correct segmentation, but k-means plus achieved better accuracy with less computation time and lower failure rate. It is concluded that k-means plus may be suitable for use in clinical routine.*

Keywords: *liver, digital image processing, segmentation, abdomen's computed tomography.*

Mathematics Subject Classification: 5804, 68Q07, 68Q09, 6804, 68N30.

¹Departamento de Automática, Facultad de Ingeniería Eléctrica, Universidad Central "Marta Abreu" de las Villas, Villa Clara, Cuba. Email: ramyduro@gmail.com.

²Departamento de Automática, Facultad de Ingeniería Eléctrica, Universidad Central "Marta Abreu" de las Villas, Villa Clara, Cuba. Email: mperez@uclv.edu.cu, marlen20012001@gmail.com.

³Departamento de Automática, Facultad de Ingeniería Eléctrica, Universidad Central "Marta Abreu" de las Villas, Villa Clara, Cuba. Email: rorozco@uclv.edu.cu.

*Autor para Correspondencia (*Corresponding Author*)

Editado por (*Edited by*): Damian Valdés Santiago, Facultad de Matemática y Computación, Universidad de La Habana, La Habana, Cuba.

Maquetado por (*Layout by*): Lázaro Daniel González Martínez, Facultad de Matemática y Computación, Universidad de La Habana, La Habana, Cuba.

Citar como: Caraballo Rodríguez, L.R.; Pérez Díaz, M.;& Orozco Morales, R. (2024). Segmentación automática del hígado en imágenes de tomografía computarizada. *Ciencias Matemáticas*, 38(1), 31–40. DOI: <https://doi.org/10.5281/zenodo.14577046>. Recuperado a partir de <https://revistas.uh.cu/rcm/article/view/9343>.

Introducción

El procesamiento digital de imágenes continúa siendo una línea de investigación abierta en la práctica de la imagenología médica. Pese a la modernización que se ha alcanzado en la tecnología de la tomografía computarizada (TC), que incluye software para procesar las imágenes, todavía se continúa investigando en métodos de mejoramiento de la calidad de la imagen para facilitar el diagnóstico médico. Los métodos para la segmentación de diferentes estructuras, órganos y lesiones, son parte de ese quehacer.

Así, por ejemplo, la visualización completa del hígado, se hace muy compleja por su ubicación. Se encuentra en la parte superior derecha de la cavidad abdominal, debajo del diafragma y por encima del estómago, el riñón derecho y los intestinos. Al estar cerca de tantos elementos es inevitable la superposición de estructuras sobre las imágenes médicas [27]. Debido a esto, se acude al procesamiento digital de imágenes para auxiliar en la tarea de localización y segmentación de este órgano.

El objetivo de este trabajo es implementar y comparar tres métodos modernos de segmentación 3D del hígado, a partir de imágenes de TC, para buscar una alternativa eficaz y eficiente computacionalmente para la tarea, que pueda utilizarle en la rutina clínica, una vez que sea incorporada a un software libre.

Los métodos de segmentación de estructuras y órganos son parte de las mejoras progresivas que se han introducido. En el caso específico del hígado, son una pieza clave para la detección de enfermedades hepáticas. Son también el paso inicial para posibles trasplantes, por la facilidad que presenta esta técnica para la determinación del volumen hepático, tanto del donante como del receptor. También es crucial para la detección de tumores hepáticos y la planificación de radioterapia interna selectiva, así como para la reducción del riesgo de dosis de radiación excesiva o insuficiente a suministrar al paciente enfermo.

De manera general, no existe una solución global de segmentación para todos los problemas que se plantean. Los métodos y técnicas de segmentación abarcan desde las más sencillas, basadas en diferencias en niveles de gris, hasta las más complejas basadas en textura, imágenes vectoriales y modelos deformables.

Durante las últimas décadas, se han realizado varias investigaciones sobre los métodos de segmentación hepática, obteniéndose métodos automáticos y semiautomáticos para segmentar el hígado como órgano y los tumores de hígado [26, 18, 17, 7, 5]. La mayoría de estos estudios incorporan como información la ubicación espacial y la forma de este órgano. Además, se incluyen varios pasos de procesamiento previo y posterior para mejorar o refinar el resultado. Se pueden citar los métodos que se basan en modelos deformables [13, 31], atlas probabilísticos y crecimiento por regiones [24], contornos activos [14, 6], redes neuronales artificiales [28], que incluyen las redes profundas convolucionales [19, 10, 20], sistemas basados en reglas [4], mapeo no lineal aplicando Monte Carlo [21], aprendizaje y codificación dispersa [29],

entre otros. La mayoría de estos métodos han sido desarrollados y publicados a nivel de laboratorio, por lo que aún no tienen implementación clínica, o han conducido a un software propietario.

En este sentido, para el presente trabajo se escogieron tres enfoques de segmentación diferentes para ser comparados con fines de implementación en una futura rutina clínica: uno con un método morfológico, otro con un método estadístico y, finalmente, un método de aprendizaje de máquinas.

En todos los casos, la segmentación subdivide cada imagen en sus partes constituyentes u objetos, con el fin de separar las partes de interés del resto de la imagen. Por lo tanto, el nivel al que se lleva a cabo esta subdivisión depende del problema a resolver y del método en cuestión a aplicar.

Se debe expresar que la segmentación automática es una de las tareas más difíciles del procesamiento de imágenes y rara vez llega a alcanzar una solución perfecta, por lo que se debe buscar un método alternativo, llamado “estándar dorado” de comprobación (*ground-truth*), para la verificación de los resultados. Para realizar las comparaciones, en este estudio se utilizan métricas de calidad de la segmentación de la imagen real con respecto a su máscara de segmentación.

Relevancia del estudio

La visualización del hígado en una TC es compleja por la ubicación del órgano y la superposición de otras estructuras. Debido a esta razón, se acude al procesamiento digital de imágenes para auxiliar en la localización y visualización de su forma exacta y las posibles anomalías. Sin embargo, no existe hasta el presente un método de segmentación que resuelva el 100% de casos. Los métodos incluidos en los software de los tomógrafos son limitados. Por eso es un problema abierto a la investigación. En este trabajo se comparan tres métodos de origen muy diverso para su posible implementación en rutina clínica a partir de procesamiento en una computadora.

1. Materiales y métodos

1.1 Descripción de los métodos

Se implementaron tres métodos de segmentación en *Matlab* 20 para compararlos en cuanto al desempeño en la tarea de segmentar el hígado a partir de imágenes de TC.

El primero fue el método morfológico *k-means plus*. Este es un método de aprendizaje no supervisado que se utiliza cuando se tienen datos no etiquetados. El objetivo es encontrar grupos en los volúmenes de datos según la similitud en sus características. El algoritmo divide los objetos en *k* grupos mutuamente excluyentes. Lo hace de tal manera que los centros de cada grupo permanecen lo más cerca posible de cada objeto, pero lo más lejos posible de los centros de otros grupos.

En el problema de investigación se entiende por grupos o clases a los píxeles que pertenezcan a la región de interés del hígado, con respecto a los píxeles que pertenezcan a órga-

nos aledaños. Cada grupo se caracteriza por su punto central (centroide) [3].

Antes de aplicar el método, cada imagen fue preprocesada. Para esto, primero se realizó un ajuste de su intensidad con una tolerancia desde $0,8 \times 10^{-6}$ hasta $0,9 \times 10^4$ para ajustar el contraste, y un valor de σ de la gaussiana igual a 0,520, para suavizar la imagen y atenuar el ruido. Esto se realizó mediante prueba y error, con el objetivo de maximizar los parámetros de calidad de la segmentación.

Para lograr lo anterior, se expandió el rango de intensidades hasta los valores máximos y mínimos en formato de 16 bits mediante el siguiente código de *Matlab*:

```
Imin = double ( Imin );
Imin = Imin - min ( Imin ( : ) );
Imin = Imin / max ( Imin ( : ) );
Imin = uint16 ( 2^16 * Imin );
```

Para la reducción del ruido se utilizó un filtro gaussiano (promedio móvil ponderado), que asume ruido blanco gaussiano aditivo (AWGN, por sus siglas en inglés) mediante la función de *Matlab* `imgaussfilt`, la cual asume por defecto que el tamaño del filtro es $2 * \text{ceil}(2 * \text{sigma}) + 1$.

Se decidió en este caso no utilizar filtro de mediana porque, aunque preserva los bordes mejor que el filtro gaussiano, es más apropiado para ruido de impulso. Tampoco se utilizaron ninguno de los diversos tipos de filtros morfológicos (alternantes, geodésicos, entre otros) por ser más costosos computacionalmente y porque el ruido a atenuar no responde a estructuras morfológicas, como es el caso de las interferencias o los artefactos.

Durante la etapa de lectura de la imagen se obtuvieron las imágenes de referencia DICOM, que se tomaron como entrada al algoritmo. Además, se creó un volumen 3D, concatenando todos los cortes de imágenes 2D donde el hígado estuvo visible.

Para la implementación del algoritmo *k-means plus*, primeramente, se declararon 18 clases o grupos, en las que se segmentaría la imagen luego de 6 iteraciones, determinadas también de manera experimental. Se utilizó un elemento estructurante de radio de 10 píxeles para un suavizado morfológico. En el siguiente paso se utilizó la función `kmeans_plus` para realizar la segmentación. El elemento estructurante se aplicó después de segmentar el hígado para eliminar irregularidades en su contorno mediante un cierre morfológico (`imclose`). El tamaño del elemento está en correspondencia con las dimensiones de la imagen (512×512 píxeles) y de las estructuras a suavizar.

Seguidamente, como parte del posprocesamiento, se aplicaron un grupo de operaciones morfológicas de encogimiento, erosión y dilatación, esta vez con elemento estructurante de radio igual a 7 píxeles [1]. Estas operaciones se realizaron en el siguiente orden: dilatación (`imdilate`), reconstrucción (`imreconstruct`), cierre (`imclose`) y llenado (`imfill`). Todas estas, como operaciones binarias, contribuyen a perfilar el contorno del hígado y a eliminar otras estructuras no deseadas.

El segundo método fue el de campos aleatorios ocultos de Markov con maximización de expectativas (HMRF-EM, por sus siglas en inglés) [22]. Este es un método estadístico que utiliza las medidas de probabilidad para modelar secuencias de datos representados por vectores de observación.

Para realizar el etiquetado de píxeles, el modelo asignó una etiqueta del conjunto de etiquetas L a cada estado S . Así, en la detección de los bordes del hígado sobre una imagen, se asignó una etiqueta f_i dentro del conjunto L , para cada estado, con $i \in S$, donde S está formado por los píxeles de la imagen del hígado. Este método modela las interacciones espaciales entre píxeles cercanos. Sea $F = (f_1, \dots, f_n)$ una familia de variables aleatorias definidas en S , donde cada variable aleatoria F_i toma un valor f_i en L . Se denomina campo aleatorio a la familia de variables aleatorias F .

Se dice que esta familia cumple que:

$$P(F = f) > 0, \quad (1)$$

$$P(F_i = f_i | F_j = f_j, j \in d, i \neq j) = P(F_i = f_i | F_j = f_j, j \in N_i). \quad (2)$$

En la ecuación (2), $P(\cdot)$ y $P(\cdot | \cdot)$ son las funciones de densidad de probabilidad conjunta y condicional, respectivamente.

La esencia del algoritmo de maximización de expectativas (EM, por sus siglas en inglés) es utilizar los datos observados disponibles para estimar los datos faltantes y luego utilizar esos datos para actualizar los valores de los parámetros.

En la práctica, la implementación de este método comenzó con la declaración de 14 clases o grupos para segmentar la imagen, mediante prueba y error. Se utilizaron 10 iteraciones del algoritmo de maximización de expectativas y 6 iteraciones como parte de la auto-organización de grupos. Se declaró también una σ gaussiana de 0,30 y un elemento estructurante de radio de 11 píxeles, para realizar un suavizado morfológico, con semejante justificación a la del método anterior. Todos estos parámetros se seleccionaron de manera experimental, lo que busca lograr una buena calidad de segmentación. Se cargaron las imágenes a utilizar, así como su referencia médica y se montaron sobre el algoritmo.

Durante el preprocesamiento se realizó un ajuste de intensidad, un ensanchamiento de contraste y un suavizado de intensidad, usando un filtro gaussiano, a fin de aprovechar en gran medida toda la información de la imagen, semejante a como se realizó en el método anterior.

En la etapa de segmentación, primeramente, se realizó un método del tipo *k-means* para agrupar las clases, realizando operaciones morfológicas como encogimiento, erosión, dilatación y reconstrucción geodésica (explicadas anteriormente). Luego se aplicó la función `HMRF_EM` de *Matlab* para lograr la segmentación de la imagen, mediante el método de los mapas ocultos de Markov. Por último, se aplicó la EM para utilizar las mejores etiquetas.

El tercer método es el mapa auto-organizado o red de Kohonen (SOM, por sus siglas en inglés) [30]. Este es un modelo de red neuronal con aprendizaje no supervisado. Este tipo de red se basa en el mapa topológico existente en la corteza cerebral. Las neuronas topológicamente cercanas entre

sí tienden a producir respuestas al mismo tipo de estímulo en la corteza cerebral. La formación de este mapa puede hacerse por el método estándar o el de lotes.

En el entrenamiento por el método estándar, el proceso comienza con la inicialización de los pesos sinápticos, mediante la asignación de valores aleatorios. Luego se realizan tres pasos llamados: competitivo, cooperativo y adaptación sináptica. En el primero, para cada vector de patrón de entrada presentado en el mapa, todas las neuronas calculan valores de una función discriminante. La neurona que es más similar al vector de patrón de entrada es la ganadora (unidad de mejor coincidencia, BMU en inglés).

En el segundo paso, la neurona ganadora (BMU) encuentra la ubicación espacial en un vecindario topológico de neuronas excitadas. Entonces, las neuronas de este vecindario pueden cooperar. Por último, en el tercer paso, las neuronas excitadas pueden modificar los valores de la función discriminante, relacionada con el vector de patrón de entrada, presentado mediante el proceso de ajustes de pesos. Así, esta red neuronal obtiene una auto-organización en patrones, a partir únicamente de los datos de entrada, o propiedades de la realidad, lo que expresa las relaciones de vecindad y agrupamiento de los objetos mediante una estructura topológica.

El método SOM sigue dos ecuaciones básicas: la de encontrar la neurona ganadora determinada por la mínima distancia euclidiana a la entrada (3) y la actualización de la posición de las neuronas dentro del grupo (4):

$$d_{ij} = \|x(t) - w_{ij}(t)\|, \quad (3)$$

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)[x(t) - w_{ij}(t)], i \in N_c, \quad (4)$$

donde x es la entrada, N_c es la vecindad de la neurona ganadora, w_{ij} es cualquier peso de nodo y d_{ij} es la distancia euclidiana.

Se puede notar a partir de (4) que el proceso de actualización es una variación de la ubicación del nodo, proporcional a la distancia euclidiana del nodo a la entrada, multiplicada por la secuencia de ganancia, si el nodo se encuentra dentro de la vecindad. Si no está dentro de la vecindad, su posición permanece sin alteración.

Para implementar este método en *Matlab* se declararon 16 clases en las que se va a segmentar la imagen, igualmente de manera experimental, mediante prueba y error. Para el suavizado de la imagen se utilizó $\sigma = 0,15$ como parte del preprocesamiento de la imagen. Se declaró un elemento estructurante de radio de 9 píxeles para operaciones morfológicas como el suavizado y la erosión, como se explicó en el primer método.

Luego se comenzó la etapa de lectura de imágenes, donde se construyó un volumen 3D a partir de cortes 2D con las imágenes a segmentar y su respectiva referencia. En el bloque de preprocesamiento se realizó el ajuste de intensidad, luego un ensanchamiento de contraste con el fin de aprovechar más características de la imagen, y, finalmente, un suavizado de intensidad con un filtro gaussiano, tal y como ya se explicó.

Para la segmentación se utilizó la función *ClustSom*. Dentro de esta función, primeramente se definió el mapa auto-organizado y se entrenó la red neuronal con el 80% de los datos, para después segmentar. Seguidamente en el posprocesamiento, se realizaron un grupo de operaciones morfológicas de encogimiento, erosión, dilatación y reconstrucción, para mejorar todo el proceso, cuya descripción es semejante a la explicada para *k-means plus*.

1.2 Imágenes

Para probar el desempeño de los métodos se utilizaron imágenes de la base de datos de TC abdominal (CHAOS_Train_Sets) y de CHAOS (*Combined Healty Abdominal Organ Segmentation*), que se encuentra disponible en línea (<https://doi.org/10.5281/zenodo.3362844>). Esta es una base de datos anotada, que cuenta con las máscaras de segmentación para poder utilizarlas como referencia. Son imágenes DICOM anonimizadas.

Se utilizaron las TC de 20 pacientes, que cuentan con un total de 2874 cortes 2D. Existe una gran variabilidad en la cantidad de cortes utilizados por cada paciente, debido a la diversidad en cuanto a tamaño y morfología del hígado. El promedio fue de 90 cortes por paciente. Por esta razón, para aplicar cada método se determinó manualmente el intervalo de cortes de TC donde aparecía el hígado, en cada TC de cada paciente.

Todas las imágenes fueron determinadas en el área superior del abdomen de los pacientes en la fase venosa portal, después de la inyección de un agente de contraste. En esta fase, el parénquima hepático aumenta al máximo, a través del suministro de sangre por la vena porta hepática.

Los conjuntos de imágenes DICOM de 16 bits por píxel (bpp) presentaron una resolución de 512×512 píxeles, con un espesor de corte entre 0,7-0,8 mm y una distancia entre cortes de 3 a 3,2 mm.

En cada imagen se tuvo en cuenta que los valores de Hounsfield fueran similares en órganos cercanos y diferentes para el mismo tejido por la acción de los medios de contraste, así como la variación significativa en cuanto a la forma anatómica. Con estos elementos se pudo hacer la delimitación original del número de clases necesarias por método, lo cual se probó mediante prueba y error.

1.3 Evaluación de la calidad de la segmentación

Para evaluar la calidad de la segmentación y realizar la comparación de los métodos se utilizaron el coeficiente de Dice, la distancia promedio de superficie promedio (ASSD, por sus siglas en inglés) y la distancia superficial simétrica máxima (MSSD, por sus siglas en inglés), por comparación con la referencia que traía la base de datos [12]. Las ecuaciones de estas métricas se describen a continuación.

El coeficiente de Dice mide la similitud entre dos elementos, en este caso sería la imagen de referencia y la imagen segmentada:

$$Dice = \frac{2|V_{Ref} \cap V_{Seg}|}{|V_{Ref}| + |V_{Seg}|}, \quad (5)$$

donde V_{Ref} es el conjunto de vóxeles identificados como primer plano en el volumen de referencia y V_{Seg} es el conjunto de vóxeles del primer plano en el volumen segmentado.

El coeficiente de Dice está entre 0 y 1. Una segmentación perfecta tiene el valor 1, mientras que una segmentación sin intersección alguna de los primeros planos de la referencia con imagen segmentada alcanza el valor 0.

La métrica ASSD calcula la distancia entre los primeros planos segmentados de dos volúmenes 3D, es decir, A y B . Para calcularla se determinan los vóxeles de borde de las dos segmentaciones. Las posiciones relativas de los vóxeles en la imagen se transforman en coordenadas reales (mm) mediante el uso de parámetros en las etiquetas DICOM. Para cada vóxel de borde del volumen A , se determina el vóxel de borde más cercano en el volumen B y se calcula la distancia entre los dos. Después de esto, se calcula el promedio de todas las distancias. La distancia de un vóxel x a un conjunto de vóxeles A o B se define como:

$$d(x, A) = \min_{y \in A} d(x, y). \quad (6)$$

donde $d(x, y)$ es la distancia euclidiana de las coordenadas de los vóxeles $x; y$ del espacio real.

El promedio de todas las distancias almacenadas da como resultado la métrica ASSD, que se define como:

$$ASSD = \frac{\sum_{x \in B_{Seg}} d(x, B_{Ref})}{|B_{Ref}| + |B_{Seg}|} + \sum_{y \in B_{Ref}} d(y, B_{Seg}), \quad (7)$$

donde B_{Ref} y B_{Seg} son los conjuntos de vóxeles de borde de la referencia y el volumen segmentado, respectivamente.

Para una segmentación perfecta, la ASSD alcanza un valor de 0 mm y aumenta en la medida en que la calidad de la segmentación disminuye.

Finalmente, la métrica MSSD (o distancia de Hausdorff) se calcula como el máximo de todas las distancias de vóxeles simétricos $d(x, B_{Ref})$ y $d(y, B_{Seg})$. El valor de MSSD es 0 mm para una segmentación perfecta y aumenta en la medida en que la calidad de la segmentación disminuye.

1.4 Hardware

Para la implementación de los métodos se utilizó como hardware una laptop HP- Pavilion con las siguientes características: CPU Intel® Core™ i5-8250U, 3.40 GHz, 6 MB cache, con 4 núcleos; tarjeta gráfica Intel® UHD Graphics 620 (Integrada) y RAM 8 GB DDR4-2133 SDRAM (2,4 GB).

2. Resultados

2.1 Resultados visuales y análisis de fallos

En la Figura 1 se aprecia el desempeño de los tres métodos para un mismo corte 2D (corte 38) de la imagen tomográfica del paciente 1 de la BD. Por comparación con su máscara de referencia puede valorarse visualmente el resultado de la segmentación y sus métricas asociadas a cada método.

Los resultados se pueden considerar favorables con los tres métodos para el conjunto de imágenes de TC empleado, pues

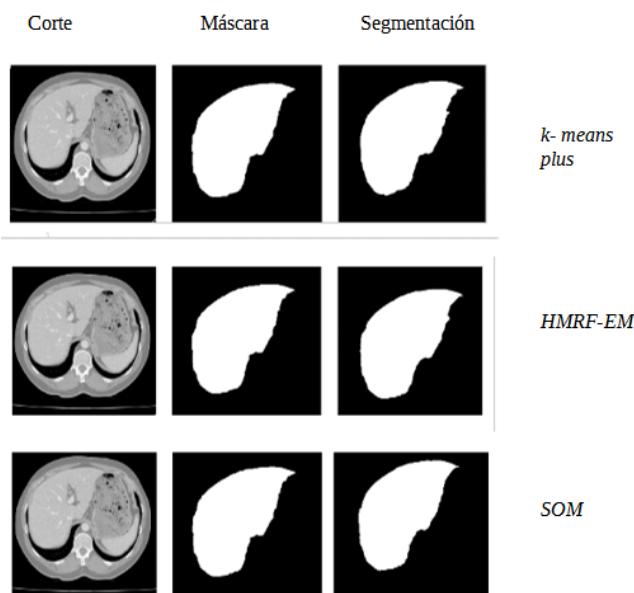


Figura 1. Desempeño de los tres métodos [Performance of the three methods].

en casi todos los casos se segmentó la región deseada. Incluso, cuando la imagen no tenía buena calidad, se apreció gran similitud entre la segmentación y la máscara de referencia en cuanto a forma, bordes y detalles, para todos los cortes de cada imagen 3D.

Las estadísticas de los fallos se describen en la Tabla 1. Aunque los métodos no fallaron en todos los cortes del vóxel 3D de cada paciente, basta que falle en segmentar un corte 2D para que la segmentación del vóxel se cuente como fallida. Las causas de los fallos se vinculan a hígados con una morfología muy compleja, imágenes con bajo contraste entre estructuras aledañas y ruido [27]. En la Figura 2 se muestra un ejemplo donde se produjo un fallo en los tres métodos.

Tabla 1. Estadística descriptiva de los fallos de los tres métodos en un mismo corte de un mismo paciente [Descriptive statistics of the failures of the three methods in the same slice of the same patient].

Métodos (cantidad de fallos)	Métricas (media y desviación \pm estándar)		
	Dice	ASSD	MSSD
<i>k-means plus</i> (3)	0,704 \pm 0,216	16,693 \pm 2,787	84,579 \pm 14,639
HRMF-EM (5)	0,608 \pm 0,151	18,243 \pm 7,210	78,312 \pm 16,007
SOM (4)	0,317 \pm 0,094	29,433 \pm 12,907	86,843 \pm 15,139

2.2 Análisis objetivo de calidad de la segmentación

En la Figura 3 se muestra cómo se comportó el coeficiente de Dice para cada método. Se aprecia que hubo una mayor fluctuación del coeficiente respecto a la referencia con el método SOM, seguido del método HMRF-EM y, por último, el *k-means plus*. Por tanto, este último fue más preciso al detectar el contorno del hígado y presentar la mayor similitud de la imagen segmentada con su máscara de referencia, según esta métrica [12].

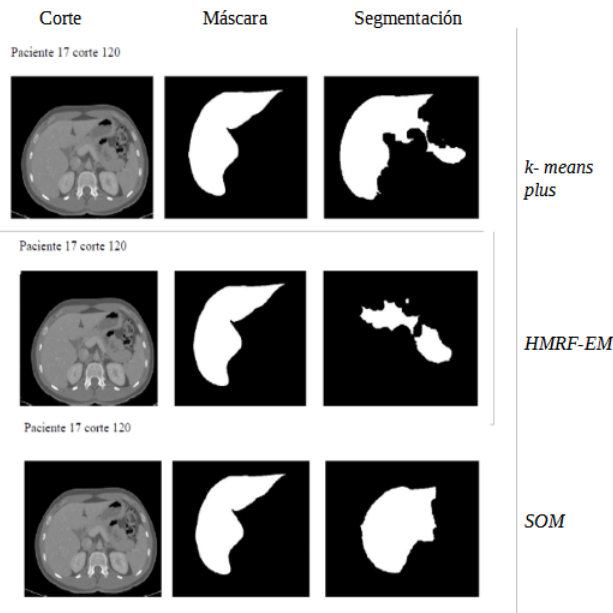


Figura 2. Desempeño fallido de los tres métodos en un mismo corte de un mismo paciente [Failed performance of the three methods in the same specimen from the same patient].

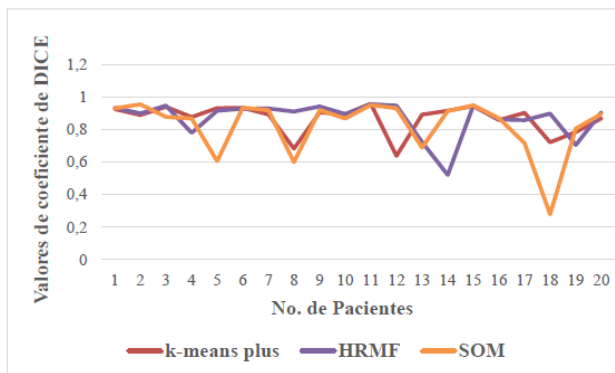


Figura 3. Comportamiento del coeficiente de Dice para cada paciente por método [Behavior of the Dice coefficient for each patient by method].

En la Tabla 2 se presentan las medias del coeficiente de Dice para las segmentaciones de imágenes de hígados de los 20 pacientes por cada método. El coeficiente es mejor, mientras más próximo a 1 se encuentre [12].

Tabla 2. Estadística descriptiva del coeficiente de Dice en el conjunto de pacientes por cada método [Descriptive statistics of the Dice coefficient in the set of patients for each method].

Métodos	Valor mínimo	Valor máximo	Media	Desviación típica
<i>k-means plus</i>	0,639	0,960	0,870	0,090
HRMF-EM	0,522	0,957	0,871	0,110
SOM	0,279	0,956	0,825	0,170

En la Figura 4 se muestra la calidad de la segmentación para los tres métodos según la métrica ASSD. Nuevamente, el método *k-means plus* se muestra más estable (con menor fluctuación) con respecto a la máscara de referencia. En el caso del HRMF-EM presenta resultados admisibles, aunque se observan varios picos no deseados. En comparación con los métodos anteriores, el SOM presentó mayor fluctuación.

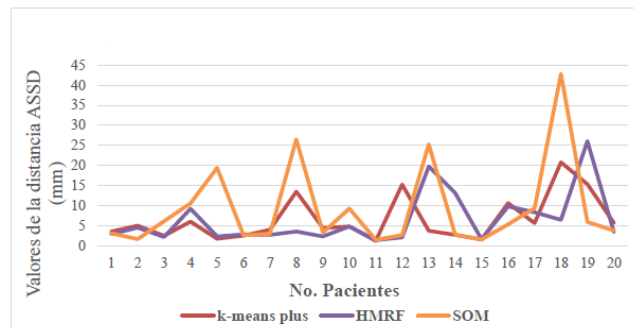


Figura 4. Comportamiento de la métrica ASSD para cada paciente por método [Behavior of the ASSD metric for each patient by method].

En la Tabla 3 se muestran las medias de la métrica ASSD para los tres métodos. Esta métrica es mejor mientras menor es su valor [12]. Por esta razón, tanto *k-means plus* como HRMF-EM se encuentran en el entorno de los 6,5 mm, siendo menores y con menor desviación típica que la métrica para SOM. Para los tres métodos hubo casos donde la métrica ASSD superó los 15 mm y que, por tanto, se consideraron fallidos.

Tabla 3. Estadística descriptiva de la métrica ASSD para todo el conjunto de pacientes con los métodos de segmentación estudiados [Descriptive statistics of the ASSD metric for the entire set of patients with the segmentation methods studied].

Métodos	Valor mínimo	Valor máximo	Media	Desviación típica
<i>k-means plus</i>	1,271	20,821	6,576	6,208
HRMF-EM	1,59	26,089	6,517	6,537
SOM	1,527	42,825	9,375	10,926

En la Figura 5 y Tabla 4 se presenta el comportamiento de la calidad de la segmentación según la MSSD. Tanto HRMF-EM como *k-means plus* presentan una media en el rango de aceptado de 0 a 60 mm. Mientras que SOM supera el intervalo requerido.

Tabla 4. Estadística descriptiva de la métrica MSSD del conjunto de pacientes por cada método [Mean for each method of the patient set for the MSSD metric].

Métodos	Valor mínimo	Valor máximo	Media	Desviación típica
<i>k-means plus</i>	15,213	148,777	59,726	36,916
HRMF-EM	18,608	138,074	56,601	35,947
SOM	16,025	143,770	64,420	43,842

Los resultados obtenidos en el cálculo de la MSSD se

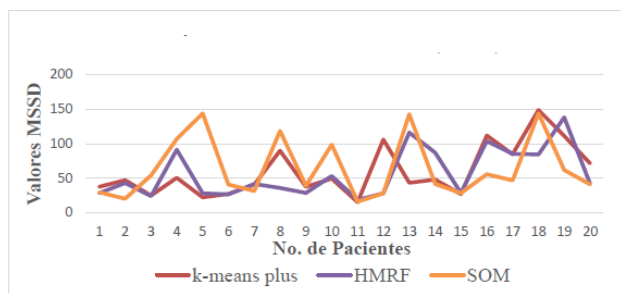


Figura 5. Comportamiento de la métrica MSSD para cada paciente por método [*Descriptive statistics of the MSSD metric for the set of patients by each method*].

interpretan de manera similar a la métrica ASSD, donde los valores óptimos son los más pequeños. Por consiguiente, se puede afirmar que, según esta métrica, el método HMRF-EM presentó los mejores resultados para segmentar el hígado, seguido de *k-means plus*.

2.3 Eficiencia computacional de los métodos

El método *k-means plus* tuvo un tiempo de ejecución ligeramente por debajo de los 14 minutos por paciente, para las prestaciones computacionales utilizadas, mientras que HMRF-EM tardó 15,33 minutos y SOM 118,19 minutos. Es apreciable que el método *k-means plus* fue 1,11 veces más veloz que HMRF-EM y 8,7 veces más rápido que SOM.

La decisión de escoger un método eficaz y eficiente para una rutina clínica depende de varios factores como la capacidad de tener mayor o menor precisión, según el objetivo médico que se persigue y, además, de las condiciones de procesamiento disponibles. Con un procesador más avanzado y la ayuda de una moderna tarjeta gráfica se reduciría significativamente el tiempo de cómputo de los tres métodos, sin tener que sacrificar precisión.

3. Discusión general

Para obtener una buena segmentación es necesario contar con una buena calidad de la imagen a analizar, o sea, que posea bajos niveles de ruido, alto contraste, alta resolución espacial, poca distorsión y que no contenga artefactos [27]. Las imágenes procesadas presentan una resolución espacial estándar de 512×512 píxeles a 16 bpp, sin la presencia de artefactos visibles. En cada método, antes de realizar la segmentación, se realizó un suavizado de la imagen para disminuir ruido y facilitar todo el procedimiento posterior.

En el caso de las imágenes con contraste muy bajo entre estructuras, esto influyó en el rendimiento de los métodos, pues en algunos casos se segmentaron estructuras no deseadas junto al hígado, como el bazo, el cual presenta un rango de valores de Hounsfield similar al del hígado [27].

En los pacientes con una alta complejidad morfológica del órgano, como el caso donde el hígado estaba particionado, los métodos probados solo segmentaron una parte. Este as-

pecto se considera una limitación de la presente investigación. En la actualidad, los software o equipos profesionales dan la posibilidad al especialista médico de sintonizar algunos parámetros.

En este caso se trabajó solamente con imágenes de 20 pacientes, para un total de 2874 cortes, logrando una eficacia comparable a métodos en el estado del arte sin necesidad de entrenamiento. Los presentes resultados aún no están concebidos para uso comercial ni público. Tan solo buscan una alternativa eficaz y eficiente a partir de comparar tres métodos de segmentación no supervisada. Sin embargo, ante la gran diversidad de imágenes reales, estos resultados pueden tomarse como una segmentación inicial, que pueda utilizarse en el futuro como dato de entrada a otros métodos con mayor poder de discriminación.

Como posible solución se plantea optimizar todos los parámetros de cada método, para su mejor desempeño, incluido el número de iteraciones. Esto, por supuesto, requiere de controlar el tiempo de cómputo, porque para rutina clínica estos dos factores deben tener una buena relación de compromiso.

Otra limitación se relaciona con el relativo pequeño tamaño de la muestra empleada. Aunque se segmentaron más de 2000 cortes 2D por método, pertenecen a solo 20 pacientes, por lo que pudiera esperarse que no hay una alta variabilidad de casos en la muestra. Deben realizarse en el futuro implementaciones de todos los métodos para volúmenes de datos mucho mayores.

Por otra parte, al añadir los métodos en un software libre para rutina clínica, se deben incluir todos los métodos que sean factibles en un tiempo de cómputo aceptable. De este modo, ante la falla de uno, se puede probar otro para buscar eficacia en la tarea. La implementación, además, debe hacerse en un software de código libre como, por ejemplo, *Python*. La presente implementación constituye un ejercicio de laboratorio realizado en *Matlab*, que es un software propietario sujeto a licencia, por lo que su empleo en rutina clínica no es lo más recomendable. Toda la programación de los métodos estudiados en el presente trabajo será próximamente realizada en *Python* para su posible inclusión en un software libre.

Es interesante hacer notar que el método que incorporó aprendizaje de máquinas fue el de peor desempeño y mayor tiempo de cómputo. Este aspecto guarda relación, no solo con los parámetros de desempeño mínimos establecidos para que cumpliera mínimamente con la tarea, que se alejan de los óptimos, sino con las prestaciones computacionales utilizadas en CPU. Este método, para lograr un desempeño cercano a *k-means plus* y HMRF-EM, tardó 8 veces más tiempo. Es de esperar que, de utilizarse una GPU para su ejecución, pueda incrementarse su tasa de aprendizaje y, con ello, su eficacia, para un tiempo de cómputo que sea competitivo con los otros dos métodos. Hacia esta meta deberán dirigirse los esfuerzos futuros.

A partir de la complejidad de la segmentación del hígado, se han realizado muchas investigaciones con distintos niveles de éxito, sin embargo, problemas como los apreciados en la

presente investigación (imposibilidad de que un método sea capaz de segmentar el 100% de los casos) son comunes en esta área de investigación [26, 18, 17, 7, 5, 13, 31, 24, 14, 6, 28, 19, 10, 20, 4, 21, 29, 3]. Por esta razón puede afirmarse que aún no se ha logrado un método exitoso para todos los casos. Por eso, es útil disponer de varios métodos de segmentación para cada tipo de tarea, de modo que ante la falla de un método, otro pueda resolver la misma.

Comparar los resultados obtenidos con la implementación de otros autores es complejo, puesto que, al revisar los artículos no se utiliza la misma base de datos ni las mismas métricas de evaluación de desempeño. Por ejemplo, en [9] se presenta un modelo de convolución basado en un algoritmo adaptativo *k-means* modificado, para obtener una mejor respuesta a la segmentación del método *k-means* tradicional. El artículo prueba cómo mejorar el método tradicional. Sin embargo, en [16] obtienen el excelente resultado de 97,5% de exactitud utilizando el *k-means* tradicional para segmentar y luego clasificar lesiones en hígados. Mientras que en [11], aplican *k-means* con el auxilio de un vector de características de entrada para disminuir su tasa de fallos.

A modo de comparación de los presentes resultados, existe un trabajo de [15] donde toman la propia base de datos CHAOS, segmentan manualmente los hígados de varios subconjuntos de 20 TC abdominales y evalúan la calidad de la segmentación manual utilizando las mismas métricas del presente artículo. La Tabla 5 muestra los valores de Dice, ASSD y MSSD para dos de los subconjuntos del artículo citado, donde en la tercera columna se colocan los relativos al método *k-means plus* del presente trabajo. No existe certeza que exista correspondencia entre casos de los subconjuntos *A* y *B* con los del presente trabajo.

Tabla 5. Comparación de resultados de las métricas con la segmentación manual y la obtenida por *k-means plus* [Comparison of metrics results with the manual segmentation and the one obtained by *k-means plus*].

Métrica	Subconjunto A segmentado manualmente	Subconjunto B segmentado manualmente	Subconjunto segmentado mediante <i>k-means plus</i>
Dice	0,979 ± 0,013	0,982 ± 0,011	0,870 ± 0,090
ASSD	2,89 ± 1,08	2,57 ± 1,02	6,576 ± 2,098
MSSD	57,83 ± 21,54	57,56 ± 20,98	59,73 ± 36,91

La segmentación manual realizada por expertos humanos siempre es más exacta y precisa que la obtenida por un método automático, pero para efectos prácticos no es factible por el consumo de tiempo y cansancio físico que produce en el radiólogo o especialista ante grandes volúmenes de datos. Como se puede apreciar, los resultados del método *k-means plus* se encuentran en el entorno de lo logrado manualmente por expertos humanos en la misma base de datos.

En [30] se aplicó el método HMRF-EM y el autor sugiere que se utilice, principalmente, para refinar la salida de la segmentación directa de algunos otros métodos como el *k-means*. Esta conclusión también es refrendada por [2]. Otros

autores, como [25], plantean que los resultados de este método pueden mejorarse utilizando remuestreo de arranque, lo cual acelera su desempeño y hace que disminuya la tasa de fallos.

En [8] se utilizó el método SOM, que arrojó resultados satisfactorios en varios conjuntos de prueba, pero siempre se presentaron fallos. Esto sugiere que su implementación futura en rutina clínica es aún un campo abierto a la investigación, coincidiendo con lo que se plantea en el presente trabajo. Trabajos como [23] muestran, sin embargo, resultados muy alentadores al probar que este método mejora notablemente el reconocimiento de las lesiones hepáticas y abre el camino para la cuantificación precisa del realce del contraste, por lo que vale la pena continuar el trabajo para su perfeccionamiento.

Como continuación de este trabajo se prevé no solo la programación en *Python* de los métodos y su incorporación a un software, sino su refinamiento con la optimización de parámetros. Igualmente, se deben estudiar otros métodos que a criterio de la comunidad científica también estén aportando buenos resultados para la tarea planteada.

Conclusiones

Los tres métodos estudiados han mostrado potencial para su implementación en la rutina clínica, principalmente, *k-means plus* y HMRF-EM, al cumplir la tarea de segmentar hígados en imágenes de TC con una eficacia aceptable.

En la investigación realizada, el método *k-means plus* presentó el mejor desempeño desde el punto de vista de la mayoría de las métricas objetivas que evalúan la calidad de la segmentación, así como mediante análisis visual comparativo con respecto a la referencia considerada como estándar dorado y por su menor número de fallos.

El método *k-means plus* presentó, además, la mayor eficiencia computacional para las prestaciones utilizadas. El resto de los métodos también pueden ser aplicados, no obstante, el método SOM solo resultaría eficiente si se utilizan mejores prestaciones computacionales.

Suplementos

Este artículo no contiene información suplementaria.

Conflictos de interés

Los autores declaramos que no tenemos ningún conflicto de interés en relación con este artículo. No hubo ninguna subvención para este artículo.

Contribución de autoría

Conceptualización R.O.M., M.P.D.

Curación de datos L.R.C.R.

Análisis formal L.R.C.R.

Investigación L.R.C.R., R.O.M.

Metodología R.O.M., L.R.C.R.

Administración de proyecto M.P.D.

Recursos M.P.D.

Software R.O.M.

Supervisión M.P.D.

Validación L.R.C.R.

Visualización L.R.C.R.

Redacción: preparación del borrador original M.P.D.

Redacción: revisión y edición R.O.M.

Referencias

- [1] Bachem, O., M. Lucic, H. Hassani, and A. Krause: *Fast and provably good seedings for k-means*. Advances in Neural Information Processing Systems, 29, 2016. <https://dl.acm.org/doi/pdf/10.5555/3157096.3157103>.
- [2] Bahendwar, Y.S. and G.R. Sinha: *Segmentation of computed tomography images using HMRF-EM algorithm with K-means Clustering*. Research and Reviews Journal of Computational Biology, 4(3):14–17, 2015. https://www.researchgate.net/profile/Professor-G-Sinha/publication/292502347_Segmentation_of_Computed_Tomography_Images_Using_HMRF-EM_Algorithm_with_K-Means_Clustering/links/56aefb1b08ae43a3980eab07/Segmentation-of-Computed-Tomography-Images-Using-HMRF-EM-Algorithm-with-K-Means-Clustering.pdf.
- [3] Bahmani, B., B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii: *Scalable k-means++*. arXiv preprint arXiv:1203.6402, 2012. <https://doi.org/10.48550/arXiv.1203.6402>.
- [4] Botella, J.S.: *Estudio y desarrollo de un preproceso basado en la difusión no lineal para la segmentación en imágenes*. Tesis de Doctorado, Universidad Politécnica de Madrid, 2013. https://oa.upm.es/14616/1/JAVIER_SANGUINO_BOTELLA.pdf.
- [5] Chowdhary, C.L., M. Mittal, K. P. P.A. Pattanaik, and Z. Marszalek: *An efficient segmentation and classification system in medical images using intuitionist possibilistic fuzzy C-mean clustering and fuzzy SVM algorithm*. Sensors, 20(14):3903, 2020. <https://doi.org/10.3390/s20143903>.
- [6] Chuah, T.K., J.H. Lim, and C.L. Poh: *Group average difference: a termination criterion for active contour*. Journal of Digital Imaging, 25:279–293, 2012. <https://doi.org/10.1007/s10278-011-9405-y>.
- [7] Conze, P.H., A.E. Kavur, E. Cornec-Le Gall, N.S. Gezer, Y. Le Meur, M.A. Selver, and F. Rousseau: *Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks*. Artificial Intelligence in Medicine, 117:102109, 2021. <https://doi.org/10.1016/j.artmed.2021.102109>.
- [8] Cunha, A., C. Watanabe, and C. Carvalho: *Medical image segmentation using the Kohonen neural network*. IEEE Latin America Transactions, 17(02):297–304, 2019. <https://doi.org/10.1109/TLA.2019.8863176>.
- [9] Debelee, T.G., F. Schwenker, S. Rahimeto, and D. Yohannes: *Evaluation of modified adaptive k-means segmentation algorithm*. Computational Visual Media, 5:347–361, 2019. <https://doi.org/10.1007/s41095-019-0151-2>.
- [10] Dou, Q., L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.A. Heng: *3D deeply supervised network for automated segmentation of volumetric medical images*. Medical Image Analysis, 41:40–54, 2017. <https://doi.org/10.1016/j.media.2017.05.001>.
- [11] Foruzan, A.H., Y.W. Chen, R.A. Zoroofi, A. Furukawa, Y. Sato, M. Hori, and N. Tomiyama: *Segmentation of liver in low-contrast images using k-means clustering and geodesic active contour algorithms*. IEICE Transactions on Information and Systems, 96(4):798–807, 2013. <https://doi.org/10.1587/transinf.E96.D.798>.
- [12] Heimann, T., B. Van Ginneken, M.A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, and I. Wolf: *Comparison and evaluation of methods for liver segmentation from CT datasets*. IEEE Transactions on Medical Imaging, 28(8):1251–1265, 2009. <https://doi.org/10.3390/s24061752>.
- [13] Jiménez, D., L. Fernández de Manuel, J. Pascau González-Garzón, J.M. Tellado, E. Ramón, D. Menéndez y M.J. Ledesma Carbayo: *Algoritmo Level-Set para segmentación hepática en TAC con restricciones de curvatura local*. En Libro de actas del XXIX Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB), páginas 207–210, 2011. <http://www2.die.upm.es/im/papers/CASEIB2011.pdf>.
- [14] Kaur, A., Y. Singh, N. Neeru, L. Kaur, and A. Singh: *A survey on deep learning approaches to medical images and a systematic look up into real-time object detection*. Archives of Computational Methods in Engineering, pages 1–41, 2021. <https://doi.org/10.1007/s11831-021-09649-9>.
- [15] Kavur, A.E., N.S. Gezer, M. Barış, S. Aslan, P.H. Conze, V. Groza, and M.A. Selver: *CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation*. Medical Image Analysis, 69:101950, 2021. <https://doi.org/10.1016/j.media.2020.101950>.

- [16] Krishan, A. and D. Mittal: *Effective segmentation and classification of tumor on liver MRI and CT images using multi-kernel K-means clustering*. Biomedical Engineering/Biomedizinische Technik, 65(3):301–313, 2020. <https://doi.org/10.1515/bmt-2018-0175>.
- [17] Le, D.C., K. Chinnasarn, J. Chansangrat, N. Keeratibharat, and P. Horkaew: *Semi-automatic liver segmentation based on probabilistic models and anatomical constraints*. Scientific Reports, 11(1):6106, 2021. <https://doi.org/10.1038/s41598-021-85436-7>.
- [18] Li, Y., Y.Q. Zhao, F. Zhang, M. Liao, L.L. Yu, B.F. Chen, and Y.J. Wang: *Liver segmentation from abdominal CT volumes based on level set and sparse shape composition*. Computer Methods and Programs in Biomedicine, 195:105533, 2020. <https://doi.org/10.1016/j.cmpb.2020.105533>.
- [19] Long, J., E. Shelhamer, and T. Darrell: *Fully convolutional networks for semantic segmentation*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015. <https://doi.org/10.48550/arXiv.1411.4038>.
- [20] Lu, F., F. Wu, P. Hu, Z. Peng, and D. Kong: *Automatic 3D liver location and segmentation via convolutional neural network and graph cut*. International Journal of Computer Assisted Radiology and Surgery, 12:171–182, 2017. <https://doi.org/10.1007/s11548-016-1467-3>.
- [21] Lu, X., J. Wu, X. Ren, B. Zhang, and Y. Li: *The study and application of the improved region growing algorithm for liver segmentation*. Optik, 125(9):2142–2147, 2014. <https://doi.org/10.1016/j.ijleo.2013.10.049>.
- [22] Mohamed, M.A. and P. Gader: *Generalized hidden Markov models. I. Theoretical frameworks*. IEEE Transactions on Fuzzy Systems, 8(1):67–81, 2000. <https://doi.org/10.1109/91.824772>.
- [23] Paire, J., V. Sauvage, A. Albouy-Kissi, V. L. Marcus, C. Marcus, and C. Hoeffel: *Fast CEUS image segmentation based on self organizing maps*. In *Medical Imaging 2014: Image Processing*, volume 9034, pages 261–266. SPIE, March 2014. <https://doi.org/10.1117/12.2043459>.
- [24] Pla Alemany, S.: *Desarrollo de un método de segmentación automática del hígado en sus segmentos de Couinaud y los vasos hepáticos mediante un atlas probabilístico a partir del análisis de imágenes de Tomografía Computarizada*. Tesis de Doctorado, Universitat Politècnica de València, 2020. <http://hdl.handle.net/10251/148598>.
- [25] Saïd, T. B., O. Azaiz, F. Chaieb, S. M'hiri, and F. Ghorbel: *Segmentation of liver tumor using HMRF-EM algorithm with Bootstrap resampling*. In *2010 5th International Symposium On IV Communications and Mobile Network*, pages 1–4. IEEE, 2010. <https://doi.org/10.1109/ISVC.2010.5656429>.
- [26] Sengun, K. E., Y. T. Cetin, M. S. Guzel, S. Can, and E. Bostanci: *Automatic liver segmentation from CT images using deep learning algorithms: a comparative study*. arXiv preprint, arXiv:2101.09987, 2021. <https://doi.org/10.48550/arXiv.2101.09987>.
- [27] Suetens, P.: *Fundamentals of medical imaging*. Cambridge University Press, 2017. <https://doi.org/10.1017/9781316671849>.
- [28] Tang, F.: *The Intelligent Medical Imaging System: The use of cloud technology and computer-aided methods*. International Journal of Radiology & Medical Imaging, 3(120):1–1, 2017. <https://doi.org/10.15344/2456-446X/2017/120>.
- [29] Tong, T., R. Wolz, Z. Wang, Q. Gao, K. Misawa, M. Fujiwara, and D. Rueckert: *Discriminative dictionary learning for abdominal multi-organ segmentation*. Medical Image Analysis, 23(1):92–104, 2015. <https://doi.org/10.1016/j.media.2015.04.015>.
- [30] Wang, Q.: *HMRF-EM-image: implementation of the hidden Markov random field model and its expectation-maximization algorithm*. arXiv preprint, arXiv:1207.3510, 2012. <https://doi.org/10.48550/arXiv.1207.3510>.
- [31] Zhou, X., T. Kitagawa, K. Okuo, T. Hara, H. Fujita, R. Yokoyama, and H. Hoshi: *Construction of a probabilistic atlas for automated liver segmentation in non-contrast torso CT images*. In *International Congress Series*, volume 1281, pages 1169–1174. Elsevier, May 2005. <https://doi.org/10.1016/j.ics.2005.03.079>.

