

Universidad de las Ciencias Informáticas



Facultad 2

**Algoritmo para la disminución de falsos positivos en la
identificación de nódulos pulmonares
solitarios en imágenes de tomografía de tórax**

Trabajo de Diploma para optar por el Título de Ingeniero en Ciencias
Informáticas

AUTORES

Marien Díaz Ruiz
Emilio Enrique Cardero Álvarez

TUTORES

MSc. Arellys Rivero Castro
Ing. Luis Manuel Cruz Correa

La Habana, junio de 2017
“Año 59 de la Revolución”



"...el futuro de nuestra Patria tiene que ser necesariamente un futuro de hombres de ciencia, tiene que ser un futuro de hombres de pensamiento, porque precisamente es lo que más estamos sembrando; lo que más estamos sembrando son oportunidades a la inteligencia..."

Fidel Castro Ruz

Declaración de autoría

Marien Díaz Ruiz, con carné de identidad 94121524733 y Emilio Enrique Cardero Álvarez, con carné de identidad 92060718605 declaramos que somos los autores de la investigación titulada “Algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax”, para optar por el título de Ingeniero en Ciencias Informáticas. Autorizamos a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio. Para que así conste, firmamos la presente declaración jurada de autoría en La Habana a los 22 días del mes de junio del año 2017.

Autores:

Marien Díaz Ruiz

Emilio Enrique Cardero Álvarez

Tutores:

MSc. Arellys Rivero Castro

Ing. Luis Manuel Cruz Correa

DATOS DE CONTACTO DE LOS TUTORES

MSc. Arellys Rivero Castro: graduada de ingeniera en Ciencias Informáticas, egresada de la UCI en el año 2009. En diciembre de 2014 defendió con resultados satisfactorios su tesis de maestría. Ha impartido las asignaturas de Sistemas Operativo y Seguridad Informática. Fue profesora de la Facultad 7 donde realizaba la función de Asesora de Seguridad Informática. Actualmente se desempeña como Analista en el Departamento de Desarrollo de Aplicaciones del Centro de Informática Médica de la Universidad de las Ciencias Informáticas.

Correo Electrónico: arcastro@uci.cu

Ing. Luis Manuel Cruz Correa: graduado de ingeniero en Ciencias Informáticas en la UCI en el año 2015. Desde su graduación se encuentra trabajando en el Centro de Informática Médica, donde desempeña el rol de Desarrollador. Durante este periodo ha colaborado en dos publicaciones relacionadas con el área de Clasificación de Nódulos Pulmonares Solitarios en imágenes de tomografía de tórax.

Correo Electrónico: lmcorrea@uci.cu

AGRADECIMIENTOS

Marien Díaz Ruiz:

A mi Mamá Longina y mi Papá Félix; porque confiaron en mí y me han dado el amor y la fuerza para seguir adelante cada día. Sin ellos no podría ser la persona que soy. Gracias a los dos por existir, los amo.

A mis hermanos Daimé y Yandy. A mi hermanita y amiga Ana Fé, por ayudarme y estar presente siempre, de la cual recibí todo tipo de apoyo; a su mamá Virtudes.

A mis tíos, en especial a mi segundo padre, mi Tío y Padrino Mateo, por ser la persona tan maravillosa que es, por haberme apoyado, querido e inculcado buenos valores siempre.

A mi abuela Josefa, a mis otros abuelos y mi Tío Alfonso que hoy no están presentes físicamente pero siempre lo estarán en mis corazones, por sus buenas enseñanzas, amor y cariño.

A mis primos, en especial a Isabella, Dianelis, Dunelis, Zoila, Cari, Marcos, Isbrandy, por darme cariño y apoyo.

A mis dos tutores y en especial a Arelys: Gracias por tu ayuda durante todo este año. Agradezco tu paciencia, tus regaños y tu apoyo incondicional. Ya sabes a que me refiero. Has sabido escucharme y ayudarme siempre que te lo he pedido. Sin ti, estos resultados no hubieran sido posible. Eres una gran persona y quiero que lo sepas.

A este tribunal y al de los cortes, porque sus críticas constructivas posibilitaron obtener buenos resultados.

Por último y no menos importante: A mis compañeros del barrio y la universidad, por su ayuda y presencia, en especial a mi primo Reiman, a Dariel y a Yanira.

Emilio Enrique Cardero Álvarez:

A mis padres, por ser quienes son y también con mucho esfuerzo hacerme quien soy yo.

Al Msc. Rafael L. Cardero Álvarez, sin su apoyo y ayuda esta tesis no hubiera sido posible.

Al Doctor Tomas, por mostrarme que es lo que realmente espera un radiólogo de un producto de este tipo y mostrarme el proceso básico de detección de tumores.

A la profe Rosa y al profe Sergio Menéndez, por ser ejemplos de pedagogía en nuestra universidad.

A mis tutores, por su inestimable ayuda, y por haberme introducido en el mundo del procesamiento de imágenes digitales.

A mis amigos y vecinos del barrio, por su preocupación y estar pendientes.

A mis compañeros de la universidad, por su sincera amistad y camaradería.

A todos los profesores de la facultad 2, y en especial al decano Darvis Dorvigni, por brindarme cada uno su conocimiento en su especialidad.

A toda mi familia materna, por quererme como a un hijo.

A toda mi familia paterna, por su atención y preocupación sin límites por mis resultados.

A ti Marien, gracias por el esfuerzo que pusimos juntos, el cual nos permitió completar esta titánica tarea.

A mi oponente y al jurado, gracias por sus correcciones constructivas, las cuales permitieron mejorar la calidad de esta tesis.

DEDICATORIA

Marien Díaz Ruiz:

A mis padres por ser mis guías y ejemplos a seguir, por su inmenso amor y cariño.

A la Revolución y a su eterno comandante en jefe Fidel Castro Ruiz, por darme la oportunidad de formarme como Ingeniera.

A todas las personas que padecen de cáncer de pulmón, sepan que la comunidad científica está haciendo todo lo posible para mitigar sus efectos.

Emilio Enrique Cardero Álvarez:

A todo aquel que padezca de esa terrible enfermedad que es el cáncer de pulmón. Espero todo el sudor y trabajo que derrochamos permita, de alguna manera, ayudarle a obtener una cura.

RESUMEN

La comunidad internacional ha realizado un esfuerzo para dar solución a la identificación del cáncer de pulmón en fases tempranas de la enfermedad, a partir del análisis de imágenes médicas. Para ello se han desarrollado sistemas de diagnóstico asistido por computador para la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax. Dichos sistemas proporcionan a los radiólogos una segunda opinión en la interpretación de los resultados diagnósticos.

En la presente investigación se exponen los resultados obtenidos con el desarrollo de la fase de identificación de un sistema de diagnóstico asistido por ordenador, utilizando para ello métodos de segmentación de imágenes. Para la obtención de la región pulmonar se utilizan técnicas de umbralización y componentes conexas. Para la inclusión de los nódulos periféricos se hace uso de la técnica *rolling-ball*. Con el objetivo de determinar las estructuras candidatas a ser nódulos pulmonares solitarios se utilizó la técnica de agrupamiento *K-Means* y con el objetivo de disminuir falsos positivos se utilizaron criterios de tamaño, circularidad y gradiente. Para el desarrollo del algoritmo se utilizó como Entorno de Desarrollo Integrado Visual Studio 2015 y como lenguaje de programación C# 4.0. Se obtuvieron resultados de sensibilidad de un 83% y un promedio de 4,75 falsos positivos por imagen médica analizada, al utilizar imágenes previamente diagnosticadas, obtenidas de la base de datos internacional *The Lung Image Database Consortium Image Collection*.

Palabras clave: cáncer de pulmón, nódulos pulmonares solitarios, procesamiento de imágenes médicas, tomografía de tórax.

ÍNDICE

INTRODUCCIÓN 1

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA DEL ALGORITMO PARA LA DISMINUCIÓN DE FALSOS POSITIVOS EN LA IDENTIFICACIÓN DE NÓDULOS PULMONARES SOLITARIOS EN IMÁGENES DE TOMOGRAFÍA DE TÓRAX..... 8

 1.1. Representación de nódulos pulmonares solitarios en imágenes de tomografía computarizada 8

 1.2. Segmentación de imágenes 10

 1.3. Análisis del desarrollo de sistemas CAD para la detección de nódulos pulmonares solitarios 12

 1.4. Nódulos Periféricos 14

 1.5. Resultados del estudio sobre las tendencias en el desarrollo de sistemas CAD para la detección de nódulos pulmonares solitarios..... 16

 1.6. Descripción de los algoritmos para identificar nódulos pulmonares solitarios identificados en el análisis de tendencias 16

 1.7. Comparación entre LDM y K-Means 18

 1.8. Elección del modelo para el algoritmo de identificación de nódulos pulmonares solitarios a utilizar ... 19

 1.9. Metodologías, tecnologías y herramientas..... 20

 1.10. Conclusiones del capítulo 23

CAPÍTULO 2: CARACTERÍSTICAS DEL ALGORITMO PARA LA DISMINUCIÓN DE FALSOS POSITIVOS EN LA IDENTIFICACIÓN DE NÓDULOS PULMONARES SOLITARIOS EN IMÁGENES DE TOMOGRAFÍA DE TÓRAX. 24

 2.1. Identificación de nódulos pulmonares solitarios 24

 2.2. Modelo de Dominio 24

 2.3. Análisis de los requisitos funcionales 26

 2.4. Análisis de los requisitos no funcionales 28

 2.5. Propuesta de solución 29

 2.6. Descripción del filtro Iris 32

 2.7. Funcionamiento del algoritmo K-Means 33

2.8. Definición de los actores.....	34
2.9. Diagrama de Casos de Uso del Sistema.....	35
2.10. Descripción del Caso de Uso del Sistema Identificar estructuras nodulares.....	35
2.11 Conclusiones del capítulo.....	39
CAPÍTULO 3: ARQUITECTURA Y DISEÑO DEL ALGORITMO PARA LA DISMINUCIÓN DE FALSOS POSITIVOS EN LA IDENTIFICACIÓN DE NÓDULOS PULMONARES SOLITARIOS EN IMÁGENES DE TOMOGRAFÍA DE TÓRAX.....	40
3.1. Modelo Arquitectónico.....	40
3.2. Diseño del algoritmo.....	41
3.3. Patrones de diseño.....	47
3.4. Conclusiones del capítulo.....	49
CAPÍTULO 4: IMPLEMENTACIÓN DEL ALGORITMO PARA LA DISMINUCIÓN DE FALSOS POSITIVOS EN LA IDENTIFICACIÓN DE NÓDULOS PULMONARES SOLITARIOS EN IMÁGENES DE TOMOGRAFÍA DE TÓRAX.....	50
4.1. Diagrama de Componentes del algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de tomografía de tórax.....	50
4.2. Estándar de codificación.....	52
4.3. Pseudocódigo de los principales métodos del algoritmo para la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax.....	52
4.4. Resultados alcanzados por el algoritmo de disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax.....	57
4.5. Conclusiones del capítulo.....	62
CONCLUSIONES GENERALES.....	63
RECOMENDACIONES.....	64
REFERENCIAS BIBLIOGRÁFICAS.....	65

INTRODUCCIÓN

En la actualidad, la realización de diagnósticos utilizando imágenes ha tenido un gran auge en todo el mundo, específicamente en el sector de la medicina. La aplicación de técnicas como el ultrasonido (US, por sus siglas en inglés), la tomografía computarizada (CT, por sus siglas en inglés) y la resonancia magnética (MR, por sus siglas en inglés), ha propiciado un medio efectivo y no invasivo en la identificación de diferentes patologías, incentivando las investigaciones médicas. Con el incremento de la utilización de imágenes digitales, ha surgido la necesidad de usar el computador para facilitar su almacenamiento, procesamiento y posterior análisis; con el objetivo de tomar decisiones diagnósticas y terapéuticas (1).

Una imagen en dos dimensiones (2D), puede ser definida como una función $f(x, y)$ donde x e y son las coordenadas de un punto en el plano y el valor de f en cualquier par (x, y) es la intensidad, nivel de gris o color existente en dicha posición. Cuando los valores de x , y y de f en cada par (x, y) son cantidades finitas y discretas, se dice que la imagen es digital (1).

El procesamiento de imágenes digitales se realiza a través de un equipo de cómputo digital (1). En el campo de la medicina, dicho procesamiento se realiza con disímiles objetivos. Algunos de ellos son medir, modificar o mejorar la visualización de ciertas regiones de interés del cuerpo humano representadas en la imagen digital.

La utilización de métodos de procesamiento de imágenes médicas ha propiciado la detección de diferentes enfermedades entre las que se encuentran los cánceres, hemorragias internas y signos de enfermedad cardiaca. El cáncer es la principal causa de muerte en todo el mundo. Según la Organización Mundial de Salud (OMS) (2), en 2015 se atribuyen a esta enfermedad 8,8 millones de defunciones. Los cinco tipos de cáncer que causaron mayor número de fallecimientos en ese año son (3):

- ✓ Pulmonar (1,69 millones de defunciones)
- ✓ Hepático (788 000 defunciones)
- ✓ Colorrectal (774 000 defunciones)
- ✓ Gástrico (754 000 defunciones)
- ✓ Mamario (571 000 defunciones)

La tasa de mortalidad en Cuba por tipo de cáncer más elevada en el período del 2013-2016 según el Anuario Estadístico de Salud (4), emitido por el Ministerio de Salud Pública (MINSAP), corresponde a los tumores

malignos de tráquea, bronquios y pulmón. En la Figura 1 se evidencia el crecimiento del número de casos en los últimos cuatro años.

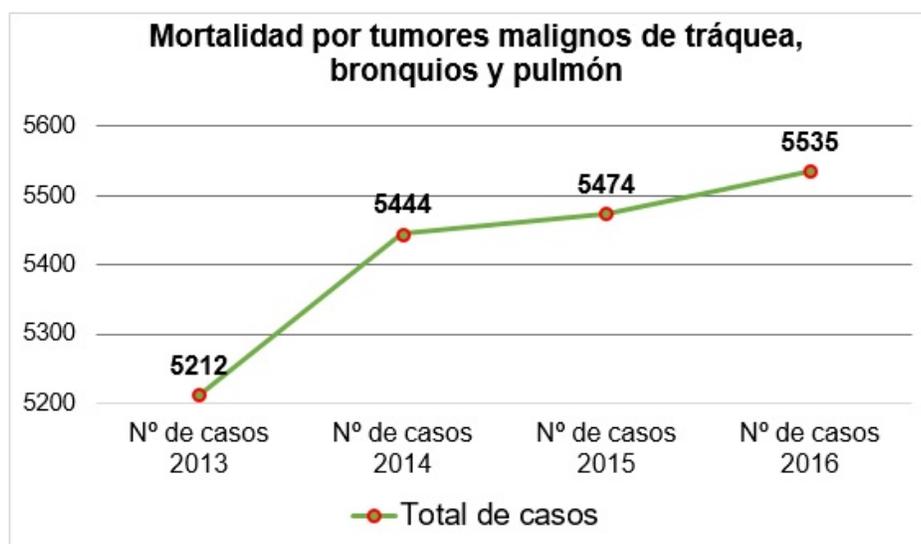


Figura 1. Mortalidad en Cuba por tumores malignos de tráquea, bronquios y pulmón en el período del 2013-2016 (4).

Para la detección temprana del cáncer de pulmón se utiliza con alta efectividad el diagnóstico por imágenes médicas. Las imágenes de CT son especialmente útiles para la identificación y clasificación de estructuras nodulares, propiciado por la resolución espacial y densidad que presentan. Permiten detectar estructuras nodulares menores a 1 mm de diámetro. Al emplear las imágenes para el diagnóstico se ha demostrado que tienen una tasa de detección de 85,1% con un valor predictivo positivo del 96,1% y una tasa de falsos positivos de 3,9% (5).

El rápido aumento de la manifestación del cáncer de pulmón en la población mundial, ha propiciado que un elevado número de investigadores a nivel internacional, desarrollen diversos trabajos encaminados a la detección temprana y eficaz de dicha enfermedad, el aumento de la capacidad diagnóstica del personal médico y la disminución del tiempo de evaluación de las imágenes. A partir de la diversidad de criterios y técnicas de procesamiento de las imágenes médicas, se han desarrollado numerosos Sistemas de Diagnóstico Asistido por Ordenador (CAD, por sus siglas en inglés) (6–9), con el objetivo de ayudar al radiólogo en el diagnóstico médico a partir del análisis de las imágenes. Estos sistemas cuentan con una

estructura dividida en cuatro fases: Preprocesamiento, Segmentación o Identificación, Extracción de rasgos y Clasificación.

En la fase de Segmentación se utilizan enfoques basados en umbralización¹, basados en bordes y orientados a regiones (10–12). Para la determinación de estos enfoques ha sido relevante el conocimiento del dominio, siendo fundamental en este caso, la forma de representación de las estructuras nodulares en las imágenes de CT.

Los nódulos pulmonares solitarios se representan en las imágenes de CT como estructuras redondeadas de alta intensidad al igual que la camilla del paciente y algunas estructuras internas del pulmón como los alveolos, bronquios, bronquiolos, vasos sanguíneos y las venas (13, 14). El principal inconveniente se encuentra en que, a pesar de haber claridad respecto a las características morfológico-radiográficas de los nódulos, diversos autores coinciden en la existencia de complejidades presentes durante la identificación y/o segmentación de las estructuras nodulares. Algunas de estas complejidades son:

- ✓ Diversos nódulos pulmonares son de muy pequeño tamaño, lo que propicia que sean confundidos con estructuras ruidosas (6).
- ✓ Las características de algunos nódulos no son completamente representativas, en ocasiones presentan complejos patrones de calcificación, textura o forma (11).
- ✓ Poseen características morfológico-radiográficas similares a otras estructuras como los bronquios y los vasos sanguíneos que también se encuentran dentro de la región pulmonar (12, 15).

La mayoría de los trabajos que abordan el tema de la identificación de nódulos pulmonares solitarios se enfocan en la obtención de altos valores de sensibilidad, despreciando en alguna medida, los valores de falsos positivos. El objetivo de los investigadores es obtener un resultado tangible luego de la fase de segmentación de las estructuras, para solo perfeccionar el proceso en las fases de reconocimiento de formas y clasificación. Si no es menos cierto que es relevante poder identificar la mayor cantidad de estructuras nodulares contenidas dentro de la región pulmonar durante la fase de identificación, la presencia de falsos positivos no contribuye de forma efectiva en la recomendación realizada por los sistemas informáticos a los médicos que hacen uso de él. Enfocar la atención del especialista hacia una estructura que no se corresponde con sus objetivos, puede propiciar la pérdida de tiempo, la falta de acierto en el

¹ *Consiste en eliminar los valores superiores o inferiores (dándoles valor de cero) respecto a un valor conocido como umbral (10).*

diagnóstico médico y por tanto la disminución de la calidad de vida del paciente; siendo un enfoque contrario a los resultados que se esperan obtener con el desarrollo de un CAD.

En el Centro de Informática Médica (CESIM), de la Universidad de las Ciencias Informáticas (UCI), se desarrolló la fase de identificación de estructuras nodulares de un CAD, que es capaz de identificar en una imagen de tomografía de tórax, estructuras nodulares con una certeza de un 79%. Estos resultados son aceptables, pero el promedio de falsos positivos por imagen médica es de 5,02 para estructuras de 3 a 30 mm de diámetro. En dicho trabajo se da solución al problema planteado utilizando un enfoque basado en técnicas de umbralización. El criterio utilizado para la selección es el bajo costo computacional de estas técnicas de procesamiento de imágenes y la sensibilidad demostrada por el algoritmo Local Density Maximum (LDM), en la identificación de estructuras altamente densas. Si bien, las técnicas de umbralización arrojan resultados relevantes en las primeras fases del proceso de identificación y contribuyen con la segmentación de las estructuras contenidas dentro de la región pulmonar:

- ✓ No facilitan la diferenciación entre todas las estructuras que son altamente densas, propiciando la aparición de un alto número de falsos positivos.
- ✓ Las estructuras nodulares no manifiestan una representación homogénea de sus valores de intensidad, evidenciándose más densas hacia el interior. Al aplicar solo técnicas de umbralización es posible que estas queden reducidas y se propicie su eliminación por considerarse estructuras ruidosas.
- ✓ El criterio descrito antes, integrado a la diversidad de formas puede propiciar que una estructura sea considerada circular, cuando era espiculada (pérdida de información durante la umbralización), influyendo posteriormente de forma negativa en el proceso de clasificación del CAD.

Los elementos anteriormente descritos inciden directamente en la identificación de estructuras que no son precisamente nódulos pulmonares solitarios, repercutiendo de forma negativa con la aparición de un elevado número de falsos positivos.

Por lo anteriormente planteado se define como **problema a resolver**: ¿cómo disminuir el número de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax?

El problema se enmarca en el **objeto de estudio**: técnicas de procesamiento de imágenes de tomografía de tórax, siendo el **campo de acción**: técnicas de segmentación de imágenes para la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax.

Se plantea como **objetivo general**: desarrollar un algoritmo para disminuir el número de falsos positivos en la detección de nódulos pulmonares solitarios en imágenes de tomografía de tórax.

Para dar cumplimiento al objetivo del presente trabajo, como guía se definieron las siguientes tareas de la investigación:

1. Análisis de las necesidades del proceso de disminución de falsos positivos en la detección de nódulos pulmonares en imágenes de tomografía de tórax.
2. Determinación de los principales referentes teóricos de las técnicas de procesamiento de imágenes médicas para la disminución de falsos positivos en la detección de nódulos pulmonares en imágenes de tomografía de tórax, estableciendo similitudes con la investigación en curso.
3. Desarrollo del algoritmo para disminuir el número de falsos positivos en imágenes de tomografía de tórax.
4. Validación de los resultados alcanzados por el algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax.

Durante el desarrollo de la investigación se utilizaron diferentes métodos científicos, todos bajo la concepción dialéctico–materialista como método general.

El método científico de la investigación aporta la orientación correcta al trabajo realizado por el investigador, permite el logro de los objetivos, el alcance de buenos resultados y condiciona el surgimiento de nuevos conocimientos. Los métodos científicos que se utilizan durante el desarrollo de la investigación son (16):

- **Teóricos**

- ✓ **Histórico–Lógico**: utilizado para la determinación de antecedentes, tendencias y regularidades de las técnicas de procesamiento de imágenes médicas, específicamente las técnicas de segmentación para la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax.
- ✓ **Analítico–Sintético**: empleado para la determinación de las generalidades y la selección de las técnicas de procesamiento a utilizar en la elaboración del algoritmo.
- ✓ **Inductivo–Deductivo**: utilizado para adquirir conocimientos generalizados acerca del procesamiento de imágenes médicas, así como determinar aquellos métodos más específicos implicados en los cambios que se necesitan para el desarrollo del algoritmo que posibiliten la disminución de falsos positivos en imágenes de tomografía de tórax.

- ✓ **Modelación:** utilizado para representar gráficamente los diferentes elementos que componen el diseño del algoritmo, confeccionar los modelos y diagramas asociados a su desarrollo.
- **Empíricos**
 - ✓ **Observación científica:** método utilizado en distintos momentos de la investigación en cuestión, para recopilar información necesaria de conceptos y funcionalidades, de forma tal que permita caracterizar el algoritmo para la detección de nódulos pulmonares en imágenes de tomografía de tórax, logrando así una visión general del algoritmo.
 - ✓ **Experimento:** para validar los resultados obtenidos por el algoritmo de disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de CT mediante la evaluación de los indicadores sensibilidad y promedio de falsos positivos.
- **Estadísticos**
 - ✓ **Estadística Descriptiva:** utilizada para recolectar, organizar, representar y describir los elementos relacionados con el impacto social del cáncer de pulmón en la población mundial, las características de las estructuras nodulares y los resultados obtenidos en los experimentos realizados.

Al utilizar el algoritmo desarrollado para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax, se esperan obtener los siguientes beneficios:

- ✓ Al integrar el algoritmo al Sistema para el Almacenamiento, Transmisión y Visualización de Imágenes Médicas (PACS) desarrollado por el CESIM, pudiera contribuir a la toma de decisiones para el diagnóstico y a la formación de nuevos especialistas en el área de la Imagenología.
- ✓ La utilización del algoritmo pudiera contribuir a la reducción del tiempo de análisis de los estudios médicos, mejorando así la calidad de vida del paciente.

El documento de investigación estará estructurado de la siguiente forma:

Capítulo 1. Fundamentación teórica. Se efectúa un análisis de los principales referentes teóricos existentes a nivel internacional y nacional relacionados con la detección de nódulos pulmonares solitarios en imágenes de tomografía de tórax. Se realiza una breve descripción de los conceptos fundamentales asociados al tema de investigación. Se realiza un análisis de los principales métodos de procesamiento de imágenes médicas y se detalla la secuencia en la que se van a utilizar para la obtención de buenos

resultados en la solución del problema planteado en la investigación. Se describen las herramientas, tecnologías y metodologías empleadas en la solución propuesta.

Capítulo 2. Características del algoritmo. Se describe como se realiza la identificación de estructuras nodulares en imágenes de tomografía de tórax y a partir de ello se realiza el Modelo de Dominio correspondiente. Se especifican los requisitos funcionales y no funcionales a partir de los cuales se realiza una propuesta de solución para el problema planteado en la investigación. Se modela el Diagrama de Casos de Uso del sistema relacionado con la identificación de nódulos pulmonares solitarios y se realiza su descripción.

Capítulo 3: Análisis y Diseño del algoritmo. Se abordan los elementos de diseño del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de CT. Se propone un modelo arquitectónico para el desarrollo del algoritmo y se modelan el Diagrama de Clases de Diseño y el Diagrama de Secuencia del Diseño. Se describen los patrones de diseño utilizados.

Capítulo 4: Implementación del algoritmo. Se describe y modela el Diagrama de Componentes referente a la presente investigación. Se muestran ejemplos en pseudocódigo de la implementación de las técnicas relevantes utilizadas para el desarrollo del algoritmo propuesto para dar solución a la problemática planteada. Se describen los indicadores a partir de los cuales se realizará la validación del algoritmo. Se valida el algoritmo, utilizando un conjunto de imágenes previamente diagnosticadas y publicadas en *The Lung Image Database Consortium Image Collection*.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA DEL ALGORITMO PARA LA DISMINUCIÓN DE FALSOS POSITIVOS EN LA IDENTIFICACIÓN DE NÓDULOS PULMONARES SOLITARIOS EN IMÁGENES DE TOMOGRAFÍA DE TÓRAX.

En este capítulo se describen los principales conceptos asociados al tema de investigación, así como los principales métodos de procesamiento de imágenes médicas que pudieran ser utilizados en la investigación. Se exponen las principales características radiográficas de los nódulos pulmonares solitarios (NPS). Se definen las tres primeras fases de un CAD (Preprocesamiento, Segmentación y Extracción de rasgos). Se incluyen los resultados del estudio de las tendencias del desarrollo de este tipo de sistemas, con el objetivo de identificar los pasos a seguir para la realización del algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de CT.

1.1. Representación de nódulos pulmonares solitarios en imágenes de tomografía computarizada

El diagnóstico por imágenes médicas es una de las formas más rápidas y efectivas para la detección del cáncer de pulmón. Algunas modalidades como la CT y la MR, son útiles para la diferenciación entre tipos de tejidos, debido a las diversas formas en que el tejido libera la energía absorbida durante el proceso de realización del estudio médico. Este elemento posibilita la identificación de tejido patológico (6, 17) y por tanto de tejido cancerígeno en las diferentes estructuras anatómicas.

1.1.1. Nódulo pulmonar solitario

Se define como una lesión única, visible radiológicamente y que se encuentra rodeada completamente por parénquima pulmonar, sin otras alteraciones que pudieran sugerir alguna otra patología. Tradicionalmente se ha definido como nódulo a lesiones menores de 3 cm en diámetro; lesiones mayores son clasificadas como masas pulmonares, las que tienen mayor riesgo de ser malignas. Se estima que son benignos el 90% de los nódulos que miden menos de dos centímetros de diámetro, y que el 60% de todos los NPS son benignos (18).

Algunas de las características más significativas que permiten la identificación de los NPS a partir de imágenes de CT como se evidencia en la Tabla 1 son: contorno, localización, calcificación y densidad.

Tabla 1. Clasificación de los NPS según contorno, localización, calcificación y densidad (13, 19).

Características morfológico-radiográficas de los NPS			
Contorno	Localización	Calcificación	Densidad
<ul style="list-style-type: none"> ✓ Lisos ✓ Lobulados ✓ Espiculados 	<ul style="list-style-type: none"> ✓ Centrales ✓ Periféricos ✓ Subpleurales 	<ul style="list-style-type: none"> ✓ Laminares ✓ Centrales ✓ Sólidos o globales ✓ Palomita de maíz ✓ Dispersos ✓ Excéntricos 	<ul style="list-style-type: none"> ✓ Sólidos (sólidos no calcificados) ✓ Calcificados (sólidos calcificados) ✓ Vidrios deslustrados (esmerilados).

1.1.2. Tomografía Computarizada

Procedimiento computarizado de imágenes por rayos X en el que se proyecta un haz de rayos X a un paciente y se gira rápidamente alrededor del cuerpo, produciendo señales que son procesadas por la computadora de la máquina, para generar imágenes transversales del cuerpo. Estos cortes se llaman imágenes tomográficas y contienen información más detallada que las técnicas convencionales. Una vez que la computadora de la máquina recolecta varios cortes sucesivos, se pueden juntar digitalmente para formar una imagen tridimensional (3D) del paciente, que permita fácilmente la identificación y ubicación de las estructuras básicas, así como de posibles tumores (20). La CT posibilita la detección de un alto porcentaje de nódulos pulmonares con mayor precisión que otras técnicas imagenológicas, debido a su calidad, determinada por su alta resolución y las especificidades que presupone la utilización de los rayos X (6, 7).

1.1.3. Diagnóstico Asistido por Computador

Se ha convertido en una de las mayores áreas de investigación en la imagenología médica y radiología diagnóstica. El ordenador señala zonas sospechosas dentro de las imágenes, utilizándose como una segunda opinión para ayudar al radiólogo a la hora de interpretar la imagen, mejorando así el rendimiento, la consistencia del diagnóstico y reduciendo el tiempo de lectura de las imágenes (7). Existen disímiles sistemas CAD encaminados a la detección de diferentes enfermedades, tales como cáncer de pulmón (21, 22), cáncer de próstata (23, 24) y cáncer de mama (25, 26). Para poder realizar su trabajo cuentan con una estructura dividida en cuatro fases, donde cada fase tiene un objetivo bien definido en el proceso de transformar la imagen médica original en una imagen médica modificada (6). En la presente investigación

se pretenden abarcar las fases: Segmentación y Extracción de Rasgos. El objetivo de los autores es seleccionar las estructuras candidatas a NPS en imágenes de CT.

En la fase de **Preprocesamiento** se realizan diferentes transformaciones a la imagen con el objetivo de eliminar ruido, se resaltan bordes para que en fases posteriores se puedan delimitar con mayor facilidad las diferentes estructuras representadas. Diversos autores (27, 28), han empleado varias técnicas para realizar el preprocesamiento a imágenes que no tienen bien definidas la región pulmonar, con el objetivo de propiciar su posterior análisis. Durante la etapa de **Segmentación** se divide la imagen en varias regiones de interés con el objetivo de detectar estructuras importantes y definir las como regiones para analizar en fases posteriores. Con los resultados obtenidos en la fase de segmentación se realiza la **Extracción de Rasgos** de las regiones. Los rasgos descriptivos están determinados por las estructuras que se quieren procesar y están asociados a la forma, tamaño, textura, relación con estructuras próximas, entre otras. Al concluir este proceso se tienen identificadas y caracterizadas las regiones de interés, lo cual posibilita su **Clasificación** (6).

1.2. Segmentación de imágenes

La segmentación es el particionamiento de una imagen en un conjunto de regiones homogéneas y significativas, de tal manera que los píxeles en cada región particionada poseen un conjunto similar de propiedades o atributos. Estos conjuntos de propiedades de la imagen pueden incluir los niveles de grises, el contraste o propiedades texturales (29). Los métodos de segmentación a emplear están determinados por el tipo de imagen, la estructura anatómica u objeto a identificar y el objetivo específico que se quiere lograr con el proceso. Elementos como el ruido, volúmenes parciales y movimientos en la imagen, también pueden tener consecuencias significativas en el desempeño de los algoritmos de segmentación, porque cada imagen y tipo de tejido tiene un conjunto de características propias (1). Existen diversos métodos de segmentación (1, 10, 29) y no es factible la utilización de una técnica genérica que se aplique a todas las situaciones que requieren de procesamiento de la imagen médica. Se utiliza una combinación de las técnicas de segmentación existentes, en función de la imagen y lo que se quiera alcanzar como resultado.

1.2.1. Umbralización binaria

La umbralización binaria es un método que busca segmentar imágenes creando una partición binaria de las intensidades de la misma. La segmentación se logra agrupando todos los píxeles con mayor intensidad al umbral en una clase y los restantes píxeles en otra clase. La umbralización es una técnica efectiva para

obtener la segmentación de imágenes donde estructuras diferentes tienen intensidades contrastantes u otras características diferenciales. Esta técnica posee un alto grado de aceptación (1, 6).

1.2.2. Agrupamiento

Las técnicas de agrupamiento (*clustering*), son técnicas de reconocimiento de patrones que buscan particionar un espacio característico derivado de cada imagen, usando datos con etiquetas conocidas. Un espacio característico es un rango espacial de cualquier función de la imagen, siendo las intensidades de la imagen el más común de los espacios característicos (10, 30).

Las técnicas de agrupamiento no utilizan datos de entrenamiento, por lo que son métodos no supervisados. Para compensar la falta de los datos de entrenamiento, iteran entre segmentar la imagen y caracterizar las propiedades de cada clase. En este sentido, se entrenan a sí mismos usando los datos disponibles. Una técnica de agrupamiento muy común es la técnica *K-Means* (30, 31). Esta agrupa datos calculando iterativamente la media de la intensidad para cada clase y segmentando la imagen mediante la clasificación de cada píxel en la clase con la media más cercana (1).

1.2.3. Crecimiento de regiones

La técnica de segmentación de imágenes conocida como crecimiento de regiones o de semilla, es una versión de la técnica de agrupamiento, donde las regiones iniciales de la imagen son cada uno de los píxeles. La segmentación se inicia en un punto concreto de la imagen, marcado por el usuario. A partir de esa semilla se añaden los píxeles vecinos que cumplan un cierto criterio. Esta técnica tiene en cuenta las relaciones que existen entre los objetos de la imagen, intentando crear regiones uniformes agrupando píxeles que satisfacen ciertos criterios de similitud. El resultado depende fuertemente del criterio y de la definición de conectividad empleada. Intenta hacer crecer una región conectada uniformemente, originada en cada semilla (10).

1.2.4. Watershed

La transformación *watershed* (1, 32) es una técnica morfológica de segmentación de imágenes en nivel de gris. Es una técnica basada en regiones, que divide todo el dominio de la imagen en conjuntos conexos. Se les aplica a imágenes en 2D y 3D. Esta técnica encierra varios de los conceptos de otras técnicas como son: umbralización y crecimiento de regiones antes mencionadas. *Watershed* a menudo produce resultados estables en la segmentación, incluyendo bordes de segmentación continuos. El algoritmo de *watershed* usa

conceptos de matemática morfológica para particionar la imagen en regiones homogéneas. Este método sufre de sobresegmentación, la cual ocurre cuando la imagen es segmentada en un número innecesario de regiones. Los algoritmos de *watershed* en imágenes médicas por lo general son procesados posteriormente para mezclar regiones separadas que pertenecen a la misma estructura.

1.3. Análisis del desarrollo de sistemas CAD para la detección de nódulos pulmonares solitarios

Se analizaron diversos sistemas CAD existentes a nivel nacional e internacional, dedicados a la identificación de estructuras candidatas a NPS en imágenes de CT. Durante la investigación se determina realizar el estudio de sistemas que utilizan técnicas de procesamiento en 2D, por considerarse más sencillas y menos costosas en tiempo de ejecución. Como parte de la descripción realizada, se resumen algunas de las técnicas de segmentación utilizadas por otros autores para la identificación de NPS.

En (33) utilizan un primer umbral en -800 Unidades *Hounsfield*² (HU, por sus siglas en inglés) para excluir la región externa del tórax. Como la densidad del pulmón en la imagen de CT varía según la profundidad de la inspiración, el ancho del haz y la calibración del escáner. Este umbral no es suficiente por lo que a continuación utilizan el histograma de la imagen dentro de la región torácica y mediante el algoritmo *K-Means* segmentan cada región pulmonar dentro del tórax utilizando únicamente los niveles de gris de la imagen. Para la reducción de falsos positivos se diseñaron clasificadores basados en reglas y se realizó un análisis discriminante lineal (LDA).

En (34) realizan el preprocesado de la imagen para mejorar su contraste y eliminar el ruido, utilizando la filtración lineal y el recorte de fondo, con el objetivo de que la imagen sea adecuada para el posterior procesamiento. Posteriormente se realiza una binarización a la imagen y se rellenan los agujeros presentes en la región pulmonar. Una vez obtenida la máscara pulmonar, el proceso de segmentación se realiza utilizando el algoritmo *K-Means*. Las imágenes de CT segmentadas resultantes, se utilizaron para extraer la región del tumor de la imagen del pulmón, introduciendo el algoritmo crecimiento de regiones.

En (35) realizan una umbralización mediante el método de *Otsu* para separar las regiones de alta y baja densidad. Para la identificación de la región pulmonar se utiliza la técnica de detección de componentes conexas. Como esta técnica se basa en la identificación de las estructuras de color blanco que se encuentran dentro de la imagen, es necesario invertir la imagen antes de buscar las componentes conexas.

² Unidades de medida correspondiente a las distintas densidades de los tejidos en una tomografía computarizada.

Con el resultado obtenido se procede a tomar las componentes que representasen a los pulmones y formar con ellas la máscara pulmonar. Luego se aplica una corrección de esa máscara pulmonar para incluir a los nódulos periféricos o subpleurales, los cuales se pierden tras la umbralización. Con la máscara corregida se efectúa una comparación píxel por píxel con la imagen original con el objetivo de quedarse con la región pulmonar. Se aplica el algoritmo LDM (6, 15) con la finalidad de detectar las estructuras nodulares. Al concluir este paso se tienen los candidatos a NPS. Realizan la reducción de falsos positivos atendiendo a criterios de tamaño y circularidad. La Figura 2 muestra paso a paso las transformaciones que sufre la imagen original de tomografía al aplicar el algoritmo descrito.

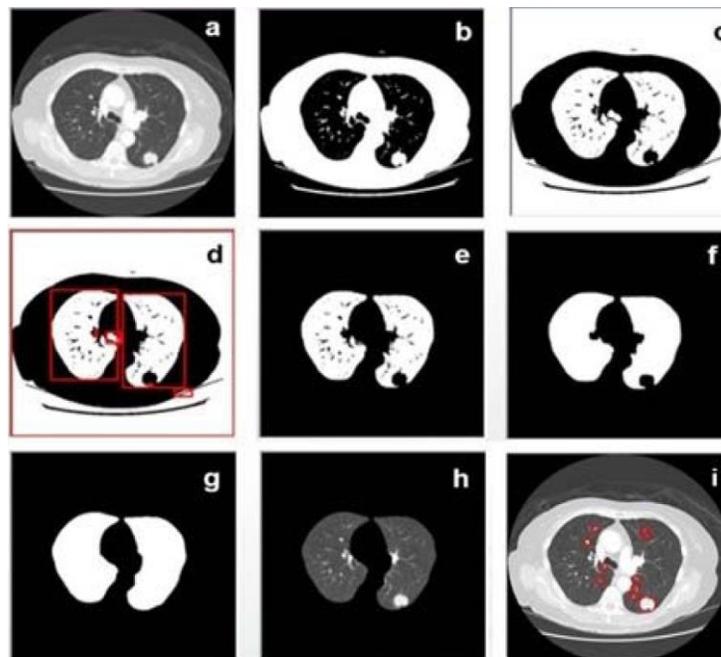


Figura 1. Secuencia de pasos para la identificación de nódulos pulmonares solitarios: (a) imagen original, (b) umbralización adaptativa, (c) inversión de la imagen umbralizada, (d) detección de componentes conexas en (c), (e) identificación de las regiones correspondientes a los pulmones, (f) primera aproximación de la máscara pulmonar, (g) máscara pulmonar corregida luego de haber aplicado cierre con disco, (h) delimitación de la región pulmonar mediante una comparación píxel a píxel entre las Figuras (a) y (g), (i) región de interés sobre la que se aplicara el algoritmo LDM para identificar las estructuras candidatas a NPS (35).

En (6) realizan la umbralización mediante el método de *Otsu*. Posteriormente se realiza la inversión de la imagen para aplicar la técnica de detección de componentes conexas e identificar la región pulmonar. Luego se aplica cierre con disco para suavizar los bordes de los pulmones e incluir las estructuras que pudieran ser de interés para el especialista. Se utiliza LDM para determinar las estructuras candidatas a NPS. Aplican

el filtro iris para determinar en cuáles de las estructuras que fueron previamente identificadas como candidatas sus valores de intensidad crecen hacia el centro.

Diversos autores en sus investigaciones (7, 15, 36, 37), utilizan técnicas de umbralización durante la fase de segmentación de estructuras, para identificar NPS en imágenes de tomografía de tórax. El principio de creación de las imágenes de tomografía, a partir de la representación de la cantidad de rayos X absorbida por el paciente durante la realización del estudio médico, propicia la obtención de buenos resultados al aplicar dichas técnicas. Los resultados obtenidos son satisfactorios debido a la facilidad de implementación y el bajo costo computacional al procesar la imagen médica.

Al analizar investigaciones que abordan temas relacionados con la identificación de estructuras en imágenes, se identifica una tendencia en el uso de algoritmos de agrupamiento para la identificación de objetos a partir de un grupo de características estructurales o de homogeneidad. Se evidencia claramente en la bibliografía consultada, el uso de *K-Means* con resultados favorables a la hora de clasificar objetos o estructuras en diferentes contextos o áreas de la ciencia, tales como la determinación de plantas enfermas (38), la identificación de cáncer de hueso (39) y la identificación de cáncer de pulmón (21, 33, 34). Si no es menos cierto que el contexto de aplicación de las técnicas de segmentación, es un elemento relevante en la obtención de buenos resultados, las características morfológico-radiográficas de los nódulos pulmonares en imágenes de CT, hace factible el uso de algoritmos de agrupamiento. Si se tiene presente, que en la imagen de CT se representan las estructuras o elementos, en función de su valor de intensidad y que los NPS son estructuras de alto valor de intensidad, es posible obtener buenos resultados al utilizar técnicas de agrupamiento con un enfoque basado en homogeneidad por valores de intensidad de los píxeles.

1.4. Nódulos Periféricos

Como los nódulos se representan en la imagen de CT como zonas de alta densidad al igual que el cuerpo y la camilla del paciente, Figura 3(a), si existen nódulos periféricos (estructura nodular que hace contacto con la pleura³), al utilizar técnicas de umbralización para la segmentación del área pulmonar, se originan errores en la máscara pulmonar resultante, ya que el nódulo se confunde con la estructura que se encuentra fuera del área pulmonar, quedando excluido de la máscara, como se evidencia en la Figura 3(b). Para dar solución a este inconveniente se han utilizado numerosas técnicas tales como: cierre morfológico con disco

³ Membrana que recubre el exterior de los pulmones (40).

(6, 35), *rolling-ball* (41), determinación de puntos frontera o técnicas de relleno (33) y contornos activos (42), dando resultados similares a los que se evidencian en la Figura 3(c).

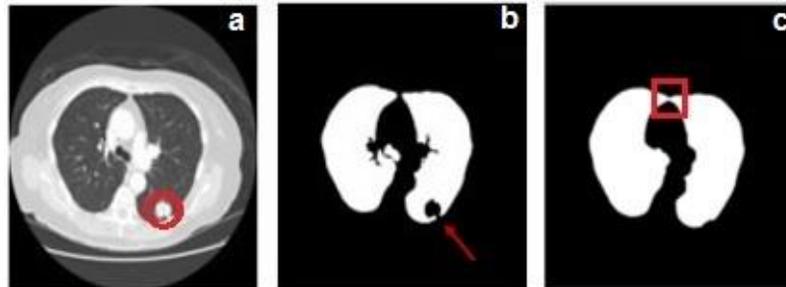


Figura 2. Nódulo periférico o subpleural excluido de la máscara pulmonar al aplicar técnicas de umbralización (fuente: elaboración propia).

Las técnicas mencionadas para realizar esta operación, a partir de su funcionamiento presentan un grupo de ventajas y desventajas que propician el relleno del área eliminada durante la umbralización con mayor o menor dificultad.

La técnica **cierre con disco** (10,12,30), consiste en aplicar las operaciones de dilación y erosión morfológica con un elemento estructural en forma de disco, con el objetivo de rellenar orificios en pequeños fragmentos. Ofrece resultados favorables de forma automática y es de bajo costo computacional. A pesar de ello, si los pulmones se encuentran a una distancia muy pequeña, la dilación puede unirlos, Figura 3(c), dando la posibilidad de generarse falsos positivos.

La utilización de la **determinación de los puntos frontera** (7), se lleva a cabo mediante la comparación de las curvaturas en los puntos de la frontera de pulmón. Esta técnica deja trazos, ángulos y bordes rectos, provocando que muchas veces la zona donde se aloje el nódulo periférico no sea rellenada completamente, aunque es suficiente para su detección.

Los **contornos activos** (14), han demostrado ser una herramienta útil para diferentes aplicaciones en el desarrollo de sistemas CAD; para la segmentación del área pulmonar y la segmentación de los nódulos contenidos dentro de dicha estructura. Esta técnica puede ser implementada de forma automática y semiautomática. La variante semiautomática ofrece mejores resultados pero requiere la intervención del radiólogo para establecer los puntos de control y construir la estructura.

Rolling-ball (7, 43), consiste en hacer rodar una circunferencia alrededor del contorno del objeto, para corregir las deformaciones e incluir los nódulos subpleurales dentro de la máscara pulmonar. Se ha decidido seleccionar esta técnica, utilizada en (41, 44, 45), debido a que ofrece resultados favorables en la corrección de la máscara pulmonar y es de fácil implementación.

1.5. Resultados del estudio sobre las tendencias en el desarrollo de sistemas CAD para la detección de nódulos pulmonares solitarios

Una vez realizados los estudios pertinentes a los procedimientos desarrollados por varios autores para la realización de sistemas CAD, útiles para la identificación de NPS en imágenes de CT, se evidencia que existen intenciones de darle solución a la problemática planteada, dado por el aumento considerable del número de muertes por este tipo de patología. Existen disímiles soluciones para la identificación de estructuras nodulares en las imágenes de CT, lo que evidencia que son variadas las combinaciones de técnicas de segmentación de imágenes que se pueden utilizar para facilitar la disminución del número de falsos positivos. Se evidencian tendencias en la obtención de buenos resultados al emplear técnicas de umbralización como LDM y de agrupamiento como *K-Means*, en diversas áreas de ciencia que hacen uso del procesamiento de imágenes, entre las que se incluyen con resultados destacados la identificación de NPS en imágenes de tomografía de tórax.

1.6. Descripción de los algoritmos para identificar nódulos pulmonares solitarios identificados en el análisis de tendencias

A continuación, se presenta una descripción simplificada del funcionamiento de los algoritmos LDM y *K-Means*. Se mencionan algunos de los parámetros que son necesarios tener en cuenta para su diseño e implementación.

1.6.1. LDM

La técnica LDM (12, 15) se enfoca en la determinación de estructuras como los vasos sanguíneos, las paredes bronquiales y los NPS que manifiestan en la imagen de tomografía, valores de densidad superiores a los del parénquima pulmonar. Por lo tanto, esta técnica se centra en encontrar estructuras que tienen alta densidad en su vecindario.

Utiliza los siguientes valores, los cuales se pueden determinar experimentalmente:

1. Valor de umbral inicial: especifica el valor de la intensidad donde empieza el umbral.

2. Valor de paso del umbral: especifica el valor de paso en que el umbral se reduce en cada momento.
3. Umbral de parada: es el valor de la intensidad en que se detiene el algoritmo.
4. Densidad mínima de máximo local.
5. Relación de valor local: especificando el cambio de volumen del objeto con respecto al volumen de su caja circundante.

Para iniciar la ejecución de LDM (6, 12, 46) se aplica un umbral inicial, que pudiera ser el máximo valor de densidad de la imagen. Luego irá disminuyendo según el valor de paso con que cambiará al próximo nivel de umbral, hasta alcanzar el valor del umbral de parada. Compara las estructuras identificadas en el nivel de umbral actual, con las estructuras identificadas en el nivel de umbral previo y analiza cuáles de estas estructuras pueden ser incluidas como posibles estructuras nodulares.

1.6.2. *K-Means*

K-Means (47), comienza con un parámetro de entrada k y parte un conjunto de n objetos en k grupos, tal que la similitud resultante dentro de un grupo es alta, pero la similitud con otros grupos es baja. Busca una partición óptima de los datos minimizando el criterio de la suma del error cuadrado con un procedimiento iterativo de optimización.

El algoritmo, consiste en tomar k píxeles aleatorios y buscar para los demás píxeles con cual de esos k píxeles tiene menor diferencia de intensidad, para agruparlo en su clase. Luego en cada clase se buscará el píxel más representativo y con ellos se procederá nuevamente a buscar cual de cada uno de estos píxeles tiene menor diferencia de intensidad respecto de los demás píxeles, para agregarlo a su clase. Si todos los píxeles quedaron en la primera clase en que estaban, se asume que el algoritmo es estable y esa es la respuesta, sino se repite el algoritmo desde el comienzo, usando como los k objetos a los que se había obtenido como los más representativos antes de hacer la última comprobación.

A continuación, se desglosa la secuencia de pasos utilizada por el algoritmo *K-Means* apoyándose en la Figura 4:

1. Inicialmente se seleccionan k objetos del conjunto de entrada (k parámetro a priori del método), en este caso son A1, A4 y A7, Figura 4(a).
2. Estos k objetos serán los centroides iniciales de los k grupos (media de los puntos del grupo, prototipo).

- Se calculan las distancias de los objetos (datos) a cada uno de los centroides, en este caso los objetos son A2, A3, A5, A6, A8. Los objetos se asignan a aquellos grupos cuya distancia es mínima con respecto a todos los centroides.

Criterio de agrupamiento: minimizar la suma de las distancias al cuadrado de todos los puntos al centro (prototipo) de su grupo: suma error cuadrático.

- Se actualizan los centroides como el valor medio de todos los objetos asignados a ese grupo, ver Figura 4(b).
- Se repiten los pasos del 2 al 4 hasta que se satisface algún criterio de convergencia.
- Elegir el modelo (k centroides, prototipos) que minimiza el error cuadrático total.

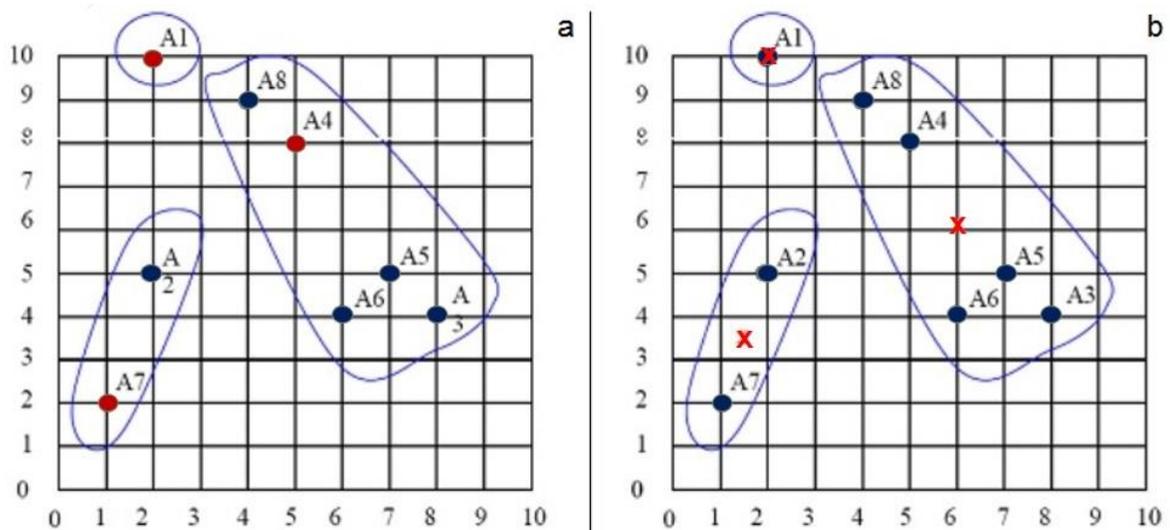


Figura 3. Funcionamiento del algoritmo K-Means (48).

1.7. Comparación entre LDM y K-Means

Una vez identificado que existe una tendencia en el uso de técnicas como LDM y *K-Means* en la identificación de NPS, se analizaron algunas ventajas y desventajas de estos métodos con el objetivo de seleccionar el algoritmo a utilizar en la presente investigación.

En los estudios donde se aborda el tema referente a la implementación del filtro LDM (7, 10), se evidencia que el método presenta altos valores de sensibilidad al ser utilizado en las imágenes de tomografía para la identificación de NPS. Estas plantean que la definición de los parámetros necesarios para su ejecución, se

pueden determinar de forma experimental y no se especifican los parámetros utilizados. Los autores que utilizan LDM, han determinado los valores de sus parámetros de forma experimental, algunos han utilizado una data pequeña y no completamente representativa de las estructuras nodulares, lo que propicia en muchos casos la obtención de resultados no generalizables al análisis de las imágenes para detectar las estructuras, siendo una desventaja importante y una limitante que impide su selección. Si bien es cierto que la utilización de varias combinaciones de estos parámetros aumenta el valor de sensibilidad en la identificación de nódulos pulmonares, también aumenta el número de falsos positivos.

En las investigaciones donde se utiliza el algoritmo *K-Means* (33, 34), se evidencia que el método arroja buenos resultados al ser utilizado en las imágenes de CT para la identificación de NPS. Tiende a arrojar en su primera ejecución menor número de falsos positivos y facilita la diferenciación entre estructuras que son y no son nódulos. Como es sensible a la elección de centroides iniciales, se recomienda realizar varias ejecuciones con varios centroides y comparar resultados para determinar el valor óptimo.

1.8. Elección del modelo para el algoritmo de identificación de nódulos pulmonares solitarios a utilizar

Existen varios autores que en sus investigaciones obtienen buenos resultados en cuanto a valores de sensibilidad y reducción de falsos positivos. Para el desarrollo de la presente investigación se considera necesario la utilización de varias técnicas que permitan obtener un elevado valor de sensibilidad con la menor cantidad de falsos positivos posible.

Cuando se analiza la imagen médica es necesario transitar por un grupo de etapas para evitar el procesamiento innecesario de algunas regiones y la reducción de falsos positivos. Para el desarrollo de la fase de segmentación, generalmente se desglosan los procedimientos realizados a la imagen médica en tres partes fundamentales: Obtención de la región pulmonar, Identificación de las estructuras y Reducción de falsos positivos. En la Tabla 2 se desglosan los pasos a seguir durante el desarrollo del algoritmo para la identificación de NPS, tomando como punto de partida el algoritmo desarrollado por (6). Para una mejor comprensión, en la columna dedicada a la especificación de pasos a seguir por los autores de la presente investigación, se marcan de forma diferenciada, los cambios más significativos realizados al algoritmo existente.

Tabla 2. Técnicas de segmentación a emplear en el desarrollo del algoritmo de identificación de NPS (fuente: elaboración propia).

Etapas para el análisis de la imagen	Algoritmo desarrollado por (7)	Algoritmo Propuesto
Obtención de la región pulmonar	a) Umbralización adaptativa b) Componentes conexas c) Cierre con disco	a) Umbralización adaptativa b) Componentes conexas c) <u>Rolling-ball</u>
Identificación de las estructuras	d) Filtro <i>Local Density Maximum</i> e) Filtro Iris	d) <u>Algoritmo K-Means</u> e) Filtro Iris
Reducción de falsos positivos	En ambos casos se realiza atendiendo a: ✓ Tamaño ✓ Circularidad	

Para la obtención de la máscara pulmonar se utilizará una umbralización adaptativa, donde el umbral se adaptará a la densidad de la imagen. Para la identificación de la región pulmonar se hará uso de la técnica de componentes conexas, la cual permite identificar las estructuras aisladas y descartar aquellas que no cumplan las exigencias de tamaño para clasificar como regiones pulmonares. Luego se utilizará *rolling-ball* para mejorar la máscara pulmonar incluyendo las estructuras nodulares que hagan contacto con la pleura a la región pulmonar. Con el objetivo de identificar las regiones candidatas a nódulos pulmonares se hará uso del algoritmo *K-Means*. Ya identificadas las regiones, se realizará la reducción de falsos positivos utilizando el filtro de tamaño y filtro de circularidad para seleccionar las estructuras que cumplan ciertos criterios de tamaño y circularidad. Posteriormente se aplicará el filtro iris ya que las estructuras candidatas a nódulos presentan un patrón en el cual, los valores de intensidad disminuyen desde el centro hacia los bordes.

1.9. Metodologías, tecnologías y herramientas

Uno de los pasos fundamentales en la creación de un software es la selección adecuada de las metodologías, tecnologías y herramientas a utilizar. En el epígrafe se describen las principales características y funcionalidades que serán utilizadas en el desarrollo del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares en imágenes de CT. El empleo de las metodologías, tecnologías y herramientas a utilizar fueron definidas por el CESIM para el desarrollo de software en el proyecto PACS-RIS.

1.9.1. Metodología de Desarrollo: *Agile Unified Process (AUP)*

AUP (49), es una versión simplificada de *Rational Unified Process (RUP)*. Este describe de una manera simple y fácil de entender la forma de desarrollar aplicaciones de software de negocio usando técnicas ágiles y conceptos que aún se mantienen válidos en RUP. Por políticas de la universidad se realizó una variación de AUP que está dividida en 3 fases:

Inicio: se llevan a cabo las actividades relacionadas con la planeación del proyecto. En esta fase se realiza un estudio inicial de la organización que permite obtener información fundamental acerca del alcance del proyecto, realizar estimaciones de tiempo, esfuerzo y costo y decidir si se ejecuta o no el proyecto.

Ejecución: en esta fase se ejecutan las actividades requeridas para desarrollar el software, incluyendo el ajuste de los planes del proyecto considerando los requisitos y la arquitectura. Durante el desarrollo se modela el negocio, obtienen los requisitos, se elaboran la arquitectura y el diseño, se implementa y se libera el producto.

Cierre: en esta fase se analizan tanto los resultados del proyecto como su ejecución y se realizan las actividades formales de cierre del proyecto.

De los cuatro escenarios de esta metodología, los autores de la presente investigación deciden utilizar el escenario número dos; donde se modela el negocio mediante el modelo conceptual y el sistema se describe utilizando casos de uso.

1.9.2. Lenguaje Unificado de Modelado: *Unified Modeling Lenguaje 2.1*

Permite generar diseños que capturen ideas en una forma convencional y fácil de comprender para comunicarlas a otras personas. Su ventaja radica principalmente en que es un lenguaje que permite desarrollar un plan bien analizado que pueda ser comprensible tanto para el cliente como para el equipo de desarrollo (35).

La utilización de este lenguaje gráfico permitirá visualizar, especificar, construir y documentar el algoritmo para la disminución de falsos positivos en imágenes de tomografía de tórax, posibilitando además la descripción de diferentes aspectos relacionados con el lenguaje de programación utilizado.

1.9.3. Herramienta CASE *Enterprise Architect 7.5*

Las herramientas de Ingeniería de Software Asistida por Computadora (CASE, por sus siglas en inglés), comprenden diferentes programas que se utilizan para auxiliar el análisis y modelado del proceso de software (50).

Enterprise Architect es una herramienta que ofrece salida de documentación de alta calidad. Posee un entorno fácil de usar, rápido y flexible. Facilita realizar ingeniería inversa de código fuente en 10 lenguajes. Se integra fácilmente con *Visual Studio*. Por estas características es que se decide utilizar como herramienta CASE para el modelado del algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de CT (35).

1.9.4. Entorno Integrado de Desarrollo *Visual Studio 2015*

Un entorno integrado de desarrollo (IDE, por sus siglas en inglés), es un programa informático compuesto por un conjunto de herramientas de programación, contiene un editor de código, un compilador, un depurador y un constructor de interfaz gráfica, utilizado por los programadores para la implementación de productos de software (51). *Visual Studio* es un IDE que permite escribir código de manera precisa y eficiente. Soporta el lenguaje de programación C#. Provee funcionalidades para refactorizar, identificar y corregir problemas de código (52).

1.9.5. Lenguaje de programación C# 4.0

C# es un lenguaje de programación orientado a objetos que permite a los desarrolladores crear un código fácil de mantener. Esta versión incorpora tipos genéricos, clases parciales y otras construcciones de lenguaje que son útiles para generar código en menos tiempo y de fácil entendimiento. Su sintaxis permite el empleo de encapsulación, herencia y polimorfismo en la creación de aplicaciones (53).

1.9.6. Librería *EmguCV 2.4.2*

EmguCV es un adaptador de la librería de procesamiento de imágenes *OpenCV*, que permite realizar invocaciones directas a funciones de *OpenCV*. Es una librería de código abierto y multiplataforma diseñada con el lenguaje de programación C#. Esta librería ha sido seleccionada ya que incluye funcionalidades esenciales para el procesamiento de imágenes médicas y por su infraestructura fácil de usar. Ha sido utilizada por varios investigadores para la identificación de NPS (35).

1.9.7. Librería CALIB

Es una solución de software para llevar a cabo los procesos de lectura, procesamiento y visualización de imágenes médicas, provenientes de múltiples equipos y de diversas modalidades. Permite a los desarrolladores de la presente investigación obtener el espaciado de cada píxel en la imagen DICOM. Esta librería fue desarrollada en Centro de Informática Médica perteneciente a la Universidad de las Ciencias Informáticas (54).

1.10. Conclusiones del capítulo

Después de analizar los métodos de segmentación y las tendencias del desarrollo de sistemas CAD para la detección de nódulos pulmonares en imágenes de CT, se arribaron a las siguientes conclusiones:

- ✓ La identificación de nódulos pulmonares está dividida en tres etapas principales: extracción de la región pulmonar, detección de los nódulos pulmonares y reducción de falsos positivos.
- ✓ Se evidencian tendencias en el uso de técnicas de umbralización y de agrupamiento como LDM y *K-Means*, para la identificación de NPS en imágenes de tomografía de tórax.
- ✓ Se decide utilizar los algoritmos *rolling-ball* para la corrección de la máscara pulmonar y *K-Means* para la identificación de los nódulos pulmonares, teniendo en cuenta que su combinación pudiera propiciar la reducción del número de falsos positivos.

CAPÍTULO 2: CARACTERÍSTICAS DEL ALGORITMO PARA LA DISMINUCIÓN DE FALSOS POSITIVOS EN LA IDENTIFICACIÓN DE NÓDULOS PULMONARES SOLITARIOS EN IMÁGENES DE TOMOGRAFÍA DE TÓRAX.

En el capítulo se especifican los requisitos funcionales y no funcionales asociados al algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de CT, queda de forma precisa una representación y descripción de los casos de usos con los que contará el algoritmo. Se realiza la descripción del Modelo de Dominio correspondiente y de la solución propuesta.

2.1. Identificación de nódulos pulmonares solitarios

Para realizar la identificación de NPS, el médico hace uso de un sistema de visualización de imágenes médicas contenido en una estación de trabajo. Debe seleccionar la imagen de CT que se quiera analizar, para aplicarle las transformaciones necesarias que permitan identificar las estructuras nodulares. Se conoce que, para la determinación del diagnóstico médico, el especialista evalúa un grupo de características estructurales, de contenido interno e incluso de representación de los NPS.

A pesar de que el especialista puede hacer uso de un grupo de herramientas informáticas para determinar la naturaleza de la estructura a analizar, en su mayoría, este hace uso de su conocimiento tácito. Este último elemento presupone una alta complejidad de representación de los elementos del negocio, por lo que se decide la realización de un Modelo de Dominio. El objetivo, es la representación de los conceptos claves asociados al negocio en que se enmarca la identificación de los NPS en imágenes de CT. Dicho modelo permite la representación de la realidad física del sistema utilizando un modelado orientado a objetos en la solución, en este caso en la identificación de NPS en imágenes de CT.

2.2. Modelo de Dominio

En el Modelo de Dominio se muestran las clases del dominio y las relaciones entre ellas. En la Figura 5 se representa el Diagrama de Clase del Modelo de Dominio realizado, correspondiente a la identificación de estructuras nodulares en imágenes de CT, donde el especialista desde una estación de trabajo tiene que visualizarlas mediante un sistema de visualización de imágenes médicas. A cada una de las imágenes debe aplicárseles un conjunto de transformaciones que permitan identificar las estructuras candidatas a NPS.

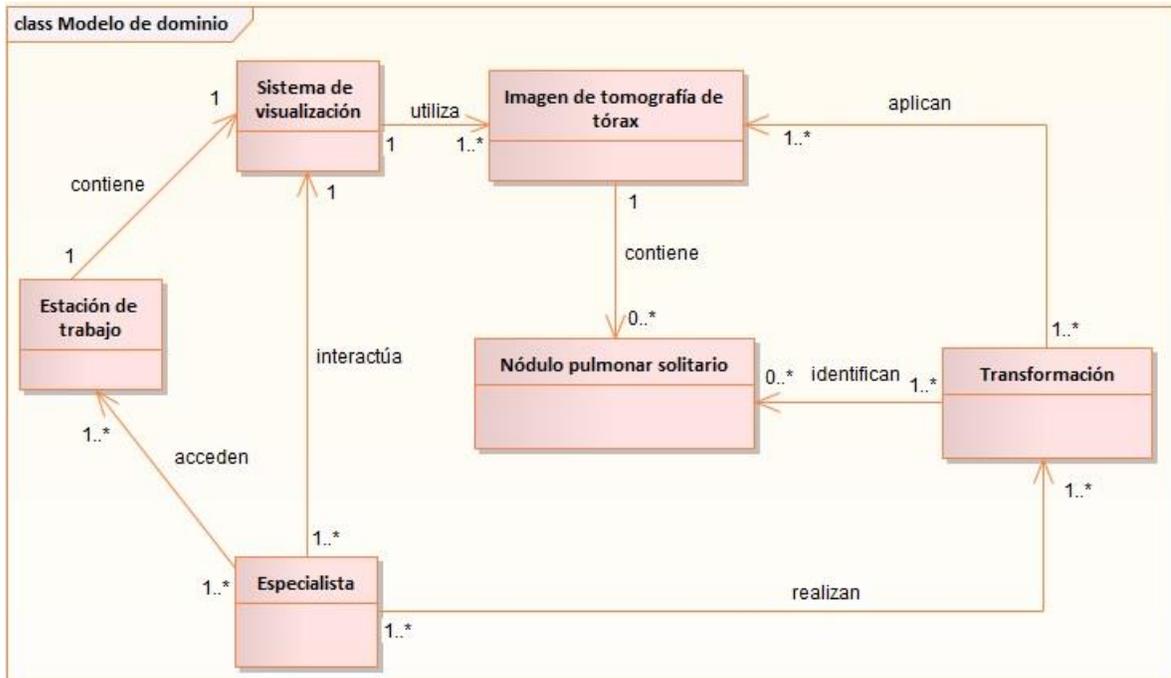


Figura 4. Modelo de Dominio del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax (fuente: elaboración propia).

Para un mejor entendimiento del modelo de dominio en la Tabla 3 se describen brevemente los conceptos que lo conforman.

Tabla 3. Clases del Modelo de Dominio del algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de tomografía de tórax (fuente: elaboración propia).

Clases	Descripción
Especialista	Médico que realiza la identificación de los NPS.
Sistema de visualización	Sistema informático que utiliza el especialista para visualizar y procesar las imágenes médicas.
Estación de trabajo	Computadora que contiene al sistema de visualización.
Imagen de tomografía de tórax	Imagen generada por un tomógrafo axial computarizado.
Nódulos pulmonares solitarios	Estructuras identificadas como parte del diagnóstico médico.

Transformaciones	Modificaciones aplicadas a las imágenes de tomografía con el objetivo de ayudar al especialista en la identificación de los NPS.
------------------	--

2.3. Análisis de los requisitos funcionales

Los requisitos del sistema son descripciones detalladas de las funciones que debe realizar el sistema, sus propiedades esenciales y deseables con el objetivo de cumplir las necesidades de los usuarios finales (50). Los requisitos funcionales para el algoritmo se muestran en la Tabla 4.

Tabla 4. Requisitos funcionales del algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de tomografía de tórax (fuente: elaboración propia).

Requisitos	Descripción
RF 1 Umbralizar imagen	Luego de haber leído la imagen se realiza una umbralización con el fin de distinguir entre las regiones pertenecientes al pulmón y las regiones externas a él.
RF 2 Invertir imagen	Invierte los valores de cada píxel de la imagen, donde los píxeles con valores 1 recibirán valores 0 y viceversa.
RF 3 Analizar componentes conexas	Identifica las componentes conexas dentro de la imagen y elimina las más pequeñas y las más grandes.
RF 4 Obtener la máscara pulmonar	Se procede a detectar los dos objetos más grandes que estén cercanos al centro de la imagen, obteniéndose así una primera máscara pulmonar.
RF 5 Corregir la máscara pulmonar	Corrige la máscara pulmonar utilizando <i>rolling-ball</i> . Donde las estructuras subpleurales que quedan excluidas tras la umbralización, se recuperan con este paso.
RF 6 Identificar la región pulmonar	Utiliza la máscara corregida para identificar la región perteneciente a los pulmones dentro de la imagen original.
RF 7 Aplicar el algoritmo <i>K-Means</i>	Se realiza el agrupamiento por clases utilizando el algoritmo <i>K-Means</i> , para separar cada píxel en clases según su intensidad.
RF 8 Detectar estructuras internas del pulmón	Con la imagen resultado del <i>K-Means</i> , se procede a buscar la clase con mayor intensidad media.
RF 9 Separar estructuras del borde y del centro del pulmón	Se separan en dos grupos las estructuras del borde y del centro del pulmón.

RF 10 Aplicar filtro de tamaño	Se seleccionan las estructuras que tengan un diámetro entre 3mm y 30mm.
RF 11 Aplicar filtro de circularidad	Se seleccionan aquellas estructuras que tengan forma aproximadamente circular.
RF 12 Aplicar filtro Iris	Buscar aquellos que cumplan con las características de ser oscuros en el borde, claros en el centro, y el cambio sea progresivo, no brusco.
RF 13 Mostrar estructuras candidatas	Se señalan en la imagen original los candidatos a nódulos con círculos rojos a su alrededor.

En el proyecto PACS-RIS existe una tendencia a la organización de requisitos por paquetes funcionales, con el objetivo de facilitar la comprensión de los elementos modelados. Un paquete es un mecanismo de propósito general para organizar un modelo de manera jerárquica (55). A pesar de ser sencilla y breve la descripción de los requisitos para la identificación de NPS en imágenes de CT, es idóneo mantener la estructura de representación utilizada en el proyecto. El objetivo de realizar el diagramado de esta forma, es facilitar la incorporación de nuevos paquetes funcionales, relacionados con fases posteriores a la identificación de estructuras nodulares, como es el caso de la Extracción de Rasgos y la Clasificación. En la Figura 6 se muestra el Diagrama de Requerimientos Funcionales, agrupados por paquetes.

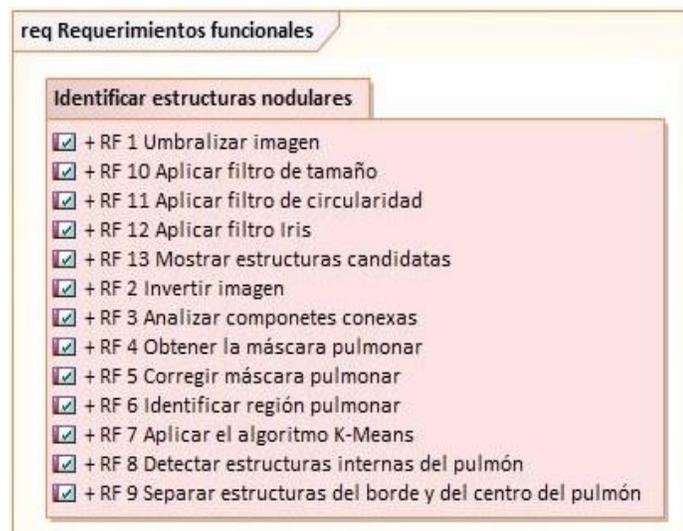


Figura 5. Requisitos Funcionales del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax (fuente: elaboración propia).

2.4. Análisis de los requisitos no funcionales

Los requisitos no funcionales no están directamente relacionados con los servicios específicos prestados por el sistema a sus usuarios, sino que especifican o restringen las características del sistema en su conjunto. Pueden relacionarse con propiedades emergentes del sistema tales como fiabilidad y tiempo de respuesta y definir restricciones en la implementación del sistema, tales como la capacidad de almacenamiento (50).

Diseño e implementación

- ✓ C# como lenguaje de programación.
- ✓ Visual Studio 2015 como IDE.
- ✓ Utilizar Enterprise Architect 7.5 como herramienta para el modelado del algoritmo.

Funcionamiento

- ✓ La estación de trabajo debe utilizar sistema operativo Windows 7, Windows 8, Windows 8.1 o Windows 10.
- ✓ Utilizar 1 GB de RAM o superior.
- ✓ Utilizar 80 GB de disco duro o superior.
- ✓ Utilizar imágenes de tomografía de tórax.
- ✓ Framework .Net versión 4.5.

En la Figura 7 se muestra el Diagrama de Requerimientos no Funcionales, agrupados por paquetes.



Figura 6. Requisitos no Funcionales del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax (fuente: elaboración propia).

2.5. Propuesta de solución

A partir del estudio de las técnicas de segmentación utilizadas por varios autores para la identificación de NPS, se propone la realización de un algoritmo que sea capaz de mantener o mejorar los valores de sensibilidad alcanzados por dichos autores, obteniendo un menor número de falsos positivos por imagen médica analizada. Teniendo en cuenta las características morfológicas-radiológicas que presentan los NPS en las imágenes de tomografía, el algoritmo propuesto realiza un análisis de la imagen de CT en busca de estructuras de alto valor de densidad dentro de la región pulmonar y luego selecciona solo aquellas estructuras que cumplan con ciertos criterios de tamaño y circularidad. Posteriormente de las estructuras resultantes de este proceso son seleccionadas aquellas cuyos valores de intensidad crecen hacia el centro. Se muestran los resultados obtenidos en la imagen original mediante el señalamiento de las estructuras nodulares que pudieran ser de interés para el especialista.

Obtención de la región pulmonar

Con el objetivo de obtener la región pulmonar, la imagen original de CT Figura 8(a) transita por una serie de pasos, comenzando con una umbralización binaria. Los valores de densidad de las imágenes de CT son variables, por lo que se decide la implementación de un umbral adaptativo. Con este procedimiento se separan las regiones de alta y baja densidad representadas en la imagen, Figura 8(b). Mediante un valor inicial realiza una segmentación de los píxeles de la imagen en dos clases y con ellas se realizan los siguientes pasos:

- a) Cálculo de la media para cada clase.
- b) Cálculo de las medias entre las dos clases.
- c) Si el valor obtenido es igual al valor inicial, el algoritmo termina retornando ese umbral, sino se repite el procedimiento utilizando el valor obtenido como nuevo umbral.

Para identificar la región pulmonar, se aplica la técnica detección de componentes conexas a la imagen umbralizada. Debido a que esta técnica se basa en la identificación de estructuras de color blanco que se encuentran dentro de la imagen, es necesario invertir la imagen umbralizada Figura 8(c). Al concluir la inversión se procede a aplicar correctamente la técnica de detección de componentes conexas, Figura 8(d). Dicha técnica permite identificar las estructuras aisladas. Luego se eliminan aquellas que no cumplan los requerimientos de tamaño para clasificar como regiones pulmonares, Figura 8(e).

Luego de haber identificado las regiones correspondientes a los pulmones, se aplica el método *rolling-ball* para incluir los nódulos contenidos en el borde de los pulmones, Figura 8(f). Posteriormente se procede a buscar todos los píxeles de color blanco, para cada uno, se selecciona el píxel que este en su misma posición en la imagen original. El resultado de esta operación permite obtener una primera aproximación a la máscara pulmonar, Figura 8(g).

Identificación de las estructuras nodulares

Se aplica el algoritmo *K-Means* a la región pulmonar para separar cada píxel en clases según su intensidad. Se utilizan cuatro clases donde estarán contenidos las estructuras de mayores valores de intensidad del pulmón, las estructuras con menores valores de intensidad del pulmón, el parénquima pulmonar y el exterior de los pulmones respectivamente, Figura 8(h). Con la imagen resultado del *K-Means*, se procede a buscar la clase con mayor intensidad media.

Reducción de falsos positivos

Las estructuras de la clase se separan en dos grupos, el grupo que contiene las del borde del pulmón y el que contiene las del interior del pulmón. Una vez que se tienen aquellas estructuras con altos valores de intensidad, se lleva a cabo un proceso de reducción de falsos positivos, utilizando criterios de tamaño y circularidad, atendiendo a las características morfológicas-radiográficas de los NPS, Figura 8(i).

✓ **Tamaño**

Se seleccionan las estructuras que poseen un diámetro que se encuentra en el intervalo de 3 a 30 mm. Las que se encuentran por encima del límite superior no son consideradas NPS y por las características de los datos y las imágenes con las que se cuenta para la validación del algoritmo desarrollado se decide restringir el límite inferior.

✓ **Circularidad**

A partir de que las estructuras nodulares se consideran aproximadamente circulares, se realiza una discriminación de aquellas estructuras que tengan un valor de circularidad fuera del rango [0,11; 0,3]. Se seleccionan aquellas estructuras que tengan criterio de circularidad (56):

$$metric = 4\pi * A/P^2 \quad [1]$$

metric: nivel de circularidad, es un número entre 0 y 1, donde uno indica un objeto perfectamente circular y cero uno que no tiene absolutamente nada circular.

Constante pi (π)= 3,1415926535897932384626433832795

A = área del objeto a evaluar

P = Perímetro del objeto a evaluar

✓ Filtro de gradiente

Al obtener las estructuras resultantes del análisis de los valores de tamaño y circularidad, las candidatas a ser nódulos pulmonares presentan un patrón en el cual, los valores de intensidad disminuyen desde el centro hacia los bordes. Los nódulos se representan en un mapa de gradiente como zonas donde los gradientes convergen; es por ello que para la identificación de las estructuras nodulares se escoge como método el análisis del campo de vectores gradiente. Se aplica a la imagen resultante una variante de estos filtros de gradiente, denominada filtro Iris. Los resultados se muestran sobre la imagen original, Figura 8(j).

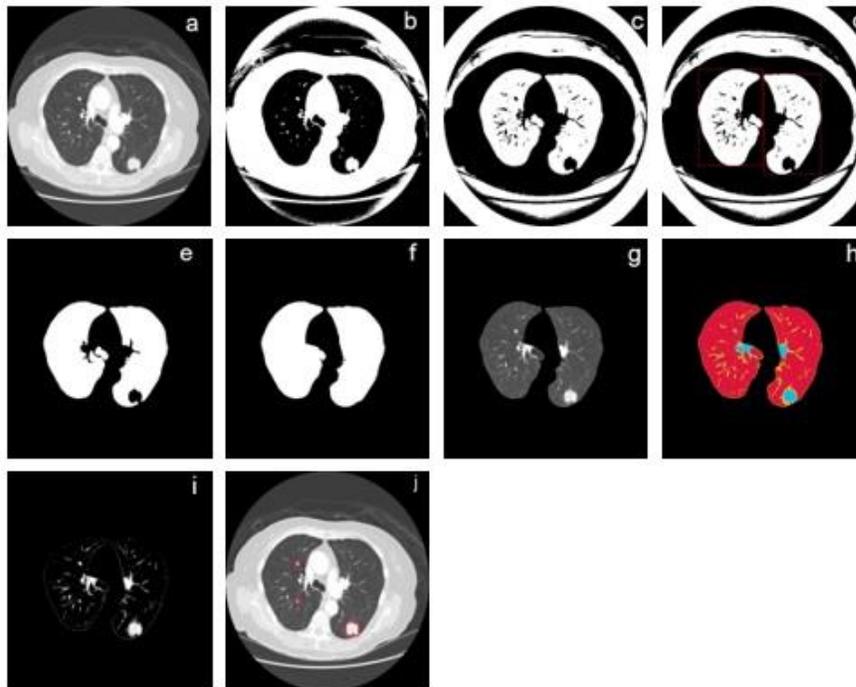


Figura 7. Pasos del algoritmo para la disminución de falsos positivos en imágenes de tomografía de tórax. a) Imagen original, b) imagen original umbralizada, c) inversión de la umbralización, d) componentes conexas

cercanas al centro, e) máscara pulmonar sin corregir, f) máscara pulmonar corregida, g) región pulmonar, h) aplicación del algoritmo *K-Means* a la región pulmonar, i) estructuras internas detectadas, j) candidatos a nódulos señalados con cuadrados rojos (fuente: elaboración propia).

2.6. Descripción del filtro Iris

Para aplicar el filtro Iris en la presente investigación, se realiza un análisis de las estructuras contenidas en la región pulmonar, que fueron resultantes al aplicar el algoritmo *K-Means* y luego los filtros de tamaño y circularidad. En dicho proceso se emplean tanto la imagen resultado de *K-Means*, como la imagen original. Para obtener el mapa de vectores gradientes de la imagen con los candidatos a nódulos, se usan operadores *Prewitt* (57, 58) por cada fila $G_r(x, y)$ y columna $G_c(x, y)$. Luego se calcula la orientación de cada vector gradiente respecto al eje x mediante la ecuación:

$$a = \tan^{-1} \frac{G_c(x', y')}{G_r(x', y')} \quad [2]$$

Para analizar la convergencia del campo de vectores hacia un punto P en específico, el filtro analiza su entorno. Mientras más vectores converjan en P , mayor será la salida del filtro en él. Este análisis se realiza mediante el uso de N líneas a modo de radios de circunferencia que convergen en el punto. El radio máximo de esas líneas viene dado por el parámetro R_{max} (radio máximo), lo cual conlleva a que se analicen a todos los puntos desde P hasta R_{max} .

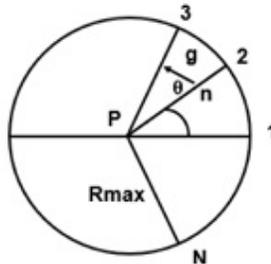


Figura 8. Punto de aplicación del filtro (P) y N líneas de análisis (l).

Para obtener el índice de convergencia (IC) del vector \vec{g} en un punto Q_i de coordenadas (x', y') con respecto a P , se debe calcular el coseno del ángulo que forma el vector con la línea que une a ambos puntos. En la Figura 9 se puede apreciar como la línea 2 intercepta al vector \vec{g} , formando al ángulo θ . Para cada punto Q_i el filtro calcula la media de convergencia de todos los índices de convergencia a lo largo del segmento $\overline{PQ_i}$.

Esto es denominado índice de convergencia modificado (ICM) (59) y se determina con la siguiente expresión:

$$ICM = \frac{1}{PQ_i} \sum_P^{Q_i} \cos \theta(x', y') dl \quad [3]$$

La salida del filtro iris en el punto P se define como:

$$O(P) = \frac{1}{N} \sum_{i=0}^{N-1} \max_{Rmin < n < Rmax} \frac{1}{n} \sum_{m=1}^n \cos \theta_{im} \quad [4]$$

donde $\frac{1}{n} \sum_{m=1}^n \cos \theta_{im}$ es la media de los índices de convergencia desde el punto P hasta el punto Q_i , a lo largo de la i -ésima línea, y el índice n representa la distancia en píxeles que separa a los puntos P y Q_i .

2.7. Funcionamiento del algoritmo *K-Means*

K-Means es un algoritmo cuyo objetivo es agrupar un conjunto de objetos en k clases a partir de un criterio de semejanza dado. En esta investigación se emplea para poder separar en clases los píxeles dentro de la región pulmonar, donde el criterio de semejanza será que la diferencia entre la intensidad de los píxeles de una misma clase sea pequeña comparada con la diferencia de intensidades entre píxeles de diferentes clases. El parámetro que se le debe suministrar es el número de clases en que se hará el agrupamiento, conocido como k . Cuando se ha completado el agrupamiento, se procede a buscar aquella clase que posea la mayor intensidad media, valor que se calcula mediante la ecuación:

$$I_k = \frac{1}{N} \sum_{n=1}^N i_n \quad [5]$$

Donde:

I_k : Valor de intensidad media de la clase k .

N : Número de píxeles agrupados en la clase k .

i_n : Valor de la intensidad del n -ésimo píxel de la clase k .

Determinación del parámetro k

Para encontrar el valor de las k clases que se deben usar para agrupar, se realizó un experimento, el cual consistió en evaluar el promedio de falsos positivos por imagen médica analizada y la sensibilidad para un conjunto de 5 series seleccionadas, utilizando en cada iteración un número diferente de clases. Las series que se usaron fueron LIDC-IDRI-0003, LIDC-IDRI-0005, LIDC-IDRI-0006, LIDC-IDRI-0028 y LIDC-IDRI-0030. Los resultados se muestran en la Tabla 5.

Tabla 5. Resultados del experimento para determinar el valor de k que mejores resultados de sensibilidad y promedio de falsos positivos arroja (fuente: elaboración propia).

Valor del parámetro k	Sensibilidad	Promedio de falsos positivos por imagen médica analizada
$k = 3$	100%	144.79
$k = 4$	100%	54.40
$k = 5$	22%	7.26

No se aplica *K-Means* con $k = 1$ ni $k = 2$ porque el resultado en ambos casos serían las mismas regiones pulmonares, sin llegar a separarse las estructuras más intensas, que es lo que se desea. No se aplica para $k \geq 6$, pues tantas clases provocarían bajos valores de sensibilidad.

El valor óptimo de clases a emplear, es aquel que propicie alcanzar valores de sensibilidad cercanos a un 100% con el menor número de falsos positivos posible. Teniendo como premisa la condición antes planteada y partiendo de los resultados del experimento ejecutado, se evidencia que el valor óptimo es $k = 4$.

2.8. Definición de los actores

Un actor representa a cualquier ente externo que interactúa con el sistema, pueden ser humanos u otros sistemas (50). El actor asociado a la presente investigación se muestra en la Tabla 6.

Tabla 6. Definición del actor referente al algoritmo de clasificación de nódulos pulmonares solitarios (elaborada por los autores).

Actor	Descripción
Especialista	Médico de cualquier especialidad que interactúa con el algoritmo para la disminución de falsos positivos en la identificación de estructuras nodulares en imágenes de CT.

2.9. Diagrama de Casos de Uso del Sistema

Los diagramas de casos de uso representan las interacciones entre el sistema y sus usuarios. Determinan las funciones que el sistema puede ejecutar (50). Este diagrama representa el comportamiento y la interacción de los usuarios con el algoritmo.

En la Figura 10 se muestra el Diagrama de Casos de Uso del Sistema referente al algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de CT.

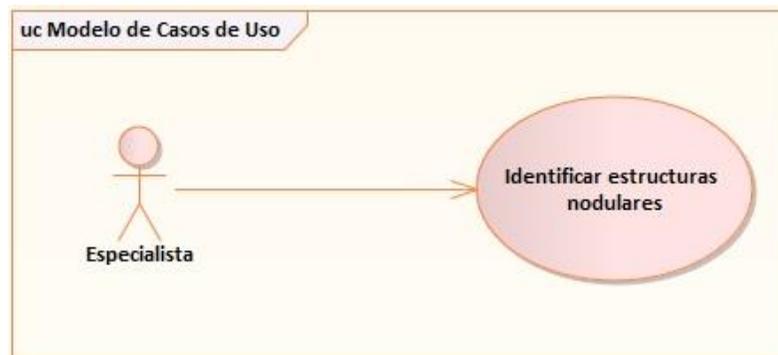


Figura 9. Diagrama de Casos de Uso del Sistema correspondiente al algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax (fuente: elaboración propia).

2.10. Descripción del Caso de Uso del Sistema Identificar estructuras nodulares

En la Tabla 7 se muestra la descripción del Caso de Uso del Sistema Identificar estructuras nodulares. La descripción se realiza con el objetivo de especificar cada uno de los elementos que componen el caso de uso, así como los flujos de eventos por los que está compuesto. Aparece reflejada la valoración de la complejidad y la prioridad del caso de uso.

Capítulo 2: Características del sistema

Tabla 7. Descripción del Caso de Uso del Sistema Identificar estructuras nodulares del algoritmo de disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de CT (fuente: elaboración propia).

Objetivo	Mostrar la imagen original con las posibles estructuras nodulares señaladas.
Actores	Especialista
Resumen	El caso de uso comienza cuando el especialista selecciona la opción Identificar estructuras nodulares en el sistema de visualización de imágenes médicas. Como resultado se mostrará los posibles NPS en la imagen analizada.
Complejidad	Alta
Prioridad	Crítica
Precondiciones	La imagen seleccionada debe ser de tomografía de tórax.
Postcondiciones	No aplica.
Flujo de eventos	
Flujo básico Identificar estructuras nodulares	
1.	El especialista selecciona la opción Identificar estructuras nodulares en el sistema de visualización de imágenes médicas.
2.	El sistema procesa la imagen y señala en ella las posibles estructuras nodulares.
3.	Termina el caso de uso.
Flujo alternativo 1 No seleccionó ninguna imagen	
1.	El especialista selecciona la opción identificar estructuras nodulares sin haber cargado la imagen.
2.	El algoritmo muestra un mensaje de error, Debe seleccionar una imagen.
Sección 1: Umbralizar imagen	
Flujo básico	
1.	El sistema aplica una umbralización basado en un umbral adaptativo para separar los píxeles de la imagen en dos clases.
Flujo alternativo	

No aplica	
Sección 2: Invertir imagen	
Flujo básico	
1.	El algoritmo invierte los valores de la imagen binaria, donde los valores en 0 toman valor 1 y viceversa.
Flujo alternativo	
No aplica	
Sección 3: Analizar componentes conexas	
Flujo básico	
1.	El algoritmo aplica la técnica de detección de componentes conexas para analizar las estructuras aisladas.
Flujo alternativo	
No aplica.	
Sección 4: Obtener máscara pulmonar	
Flujo básico	
1.	El algoritmo utiliza el resultado tras el análisis de componentes conexas y traza una región para eliminar las estructuras más pequeñas y quedarse con las dos más grandes dentro de la región pulmonar.
Flujo alternativo	
No aplica.	
Sección 5: Corregir máscara pulmonar	
Flujo básico	
1.	El algoritmo aplica la técnica de <i>rolling-ball</i> sobre la máscara pulmonar obtenida previamente.
Flujo alternativo	
No aplica.	
Sección 6: Identificar región pulmonar	

Flujo básico	
1.	El algoritmo identifica la región de interés usando la máscara pulmonar corregida y la compara píxel a píxel con la imagen original.
Flujo alternativo	
No aplica.	
Sección 7: Aplicar el algoritmo <i>K-Means</i>	
Flujo básico	
1.	Se aplica el algoritmo <i>K-Means</i> que permitirá dividir el fondo de la imagen, la parénquima pulmonar, los vasos sanguíneos de intensidad media y las estructuras internas del pulmón.
Flujo alternativo	
No aplica.	
Sección 8: Detectar estructuras internas del pulmón	
Flujo básico	
1.	Permitirá separar las estructuras contenidas en los pulmones del resto de la imagen.
Flujo alternativo	
No aplica.	
Sección 9: Aplicar filtro iris	
Flujo básico	
1.	Permite seleccionar aquellas estructuras internas del pulmón cuya intensidad aumente progresivamente hacia el centro.
Flujo alternativo	
No aplica.	
Sección 10: Aplicar filtro de circularidad	
Flujo básico	
1.	Selecciona las estructuras internas del pulmón que se representen de forma circular.
Flujo alternativo	

No aplica.		
Sección 11: Mostrar estructuras candidatas		
Flujo básico		
1.	El algoritmo muestra una imagen en la cual se señalan las posibles estructuras nodulares.	
Flujo alternativo No se identificaron estructuras candidatas		
1.	No se encuentra ninguna estructura nodular.	
2.	El algoritmo muestra la imagen original.	
Relaciones	CU incluidos	No aplicable.
	CU extendidos	No aplicable.
Requisitos no funcionales	RNDI 1, RNDI 2, RNFO 4.	
Asuntos pendientes	No aplicable.	

2.11. Conclusiones del capítulo

Luego de realizada la propuesta de solución para el desarrollo del algoritmo para la disminución de falsos positivos en la identificación de NPS, se llegaron a las siguientes conclusiones:

- ✓ Como resultado de la utilización de los métodos Analítico-Sintético e Inductivo-Deductivo, se realizó el Modelo de Dominio, a partir de la determinación y síntesis de los principales conceptos del negocio, relacionados con la identificación de NPS en imágenes de tomografía de tórax.
- ✓ Se determinó mediante el análisis de los resultados obtenidos en el experimento realizado, que el valor óptimo de clases a emplear por el algoritmo *K-Means* es $k = 4$.
- ✓ Se identificó la idoneidad de los filtros de tamaño, circularidad e iris en la reducción de falsos positivos, teniendo en cuenta a las características morfológico-radiográficas de los NPS y los resultados obtenidos por otros investigadores.

CAPÍTULO 3: ARQUITECTURA Y DISEÑO DEL ALGORITMO PARA LA DISMINUCIÓN DE FALSOS POSITIVOS EN LA IDENTIFICACIÓN DE NÓDULOS PULMONARES SOLITARIOS EN IMÁGENES DE TOMOGRAFÍA DE TÓRAX.

En el presente capítulo se describe el modelo arquitectónico utilizado en el desarrollo del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios. Se representa el diseño a partir de los Diagramas de Clase del Diseño y el Diagrama de Secuencia del Diseño. Se describen los elementos de estos diagramas y los patrones de diseño utilizados en el desarrollo del algoritmo.

3.1. Modelo Arquitectónico

El diseño arquitectónico se ocupa de comprender cómo debe organizarse un sistema y diseñar la estructura general de ese sistema. Es el vínculo crítico entre el diseño y la ingeniería de requisitos, ya que identifica los principales componentes estructurales de un sistema y las relaciones entre ellos. La salida del proceso de diseño arquitectónico es un modelo arquitectónico que describe cómo el sistema está organizado como un conjunto de componentes que se comunican (50). Como modelo arquitectónico del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares en imágenes de tomografía de tórax se utilizó Tuberías y Filtros (*Pipes and Filter*).

La arquitectura basada en tuberías y filtros es utilizada cuando se requiere el procesamiento de un objeto mediante un conjunto de transformaciones comprendidas en fases secuenciales, donde los datos fluyen de uno a otro y se transforman a medida que se mueven a través de la secuencia. Cada paso de procesamiento se implementa como una transformación. Los datos de entrada fluyen a través de estas transformaciones hasta convertirlas en salida. Las transformaciones pueden ejecutarse secuencialmente o en paralelo (50, 55). Esto se evidencia en el conjunto de transformaciones que se le deben aplicar a la imagen de CT con el objetivo de identificar los NPS. En la Figura 11 se muestra la secuencia de técnicas aplicadas a las imágenes de CT para identificar las estructuras nodulares.

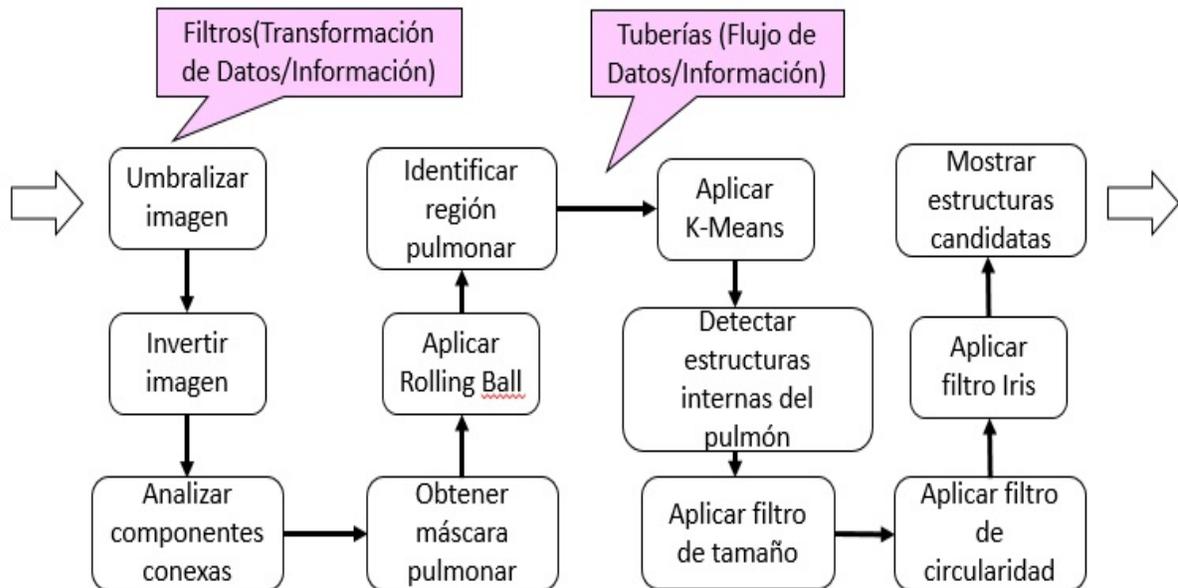


Figura 10. Arquitectura basada en Tuberías y Filtros del algoritmo para la identificación de nódulos pulmonares solitarios en imágenes de CT (fuente: elaboración propia).

3.2. Diseño del algoritmo

Una de las etapas más importantes del proceso de desarrollo de software es la etapa de diseño. Esta etapa tiene como propósito transformar los requisitos en un diseño que detalle lo que debe realizar el sistema, diseñar sus funcionalidades y adaptarlas al entorno de implementación, con el objetivo de obtener un producto de alta calidad (50).

Un diagrama de clases de diseño muestra la estructura de un software. Contiene información sobre las clases, atributos, información acerca del tipo de los atributos, métodos y relaciones entre las clases. El diagrama de clases de diseño posibilita un mejor entendimiento de la estructura del sistema (55). Cada clase del diagrama cuenta con atributos y métodos bien definidos facilitando a los programadores el desarrollo del mismo. En la Figura 12 se muestra el Diagrama de Clases del Diseño perteneciente al algoritmo, para una mayor comprensión de las clases del diagrama, se realizó una descripción de las mismas, ver [Anexo 1](#).

class Diagrama de clases



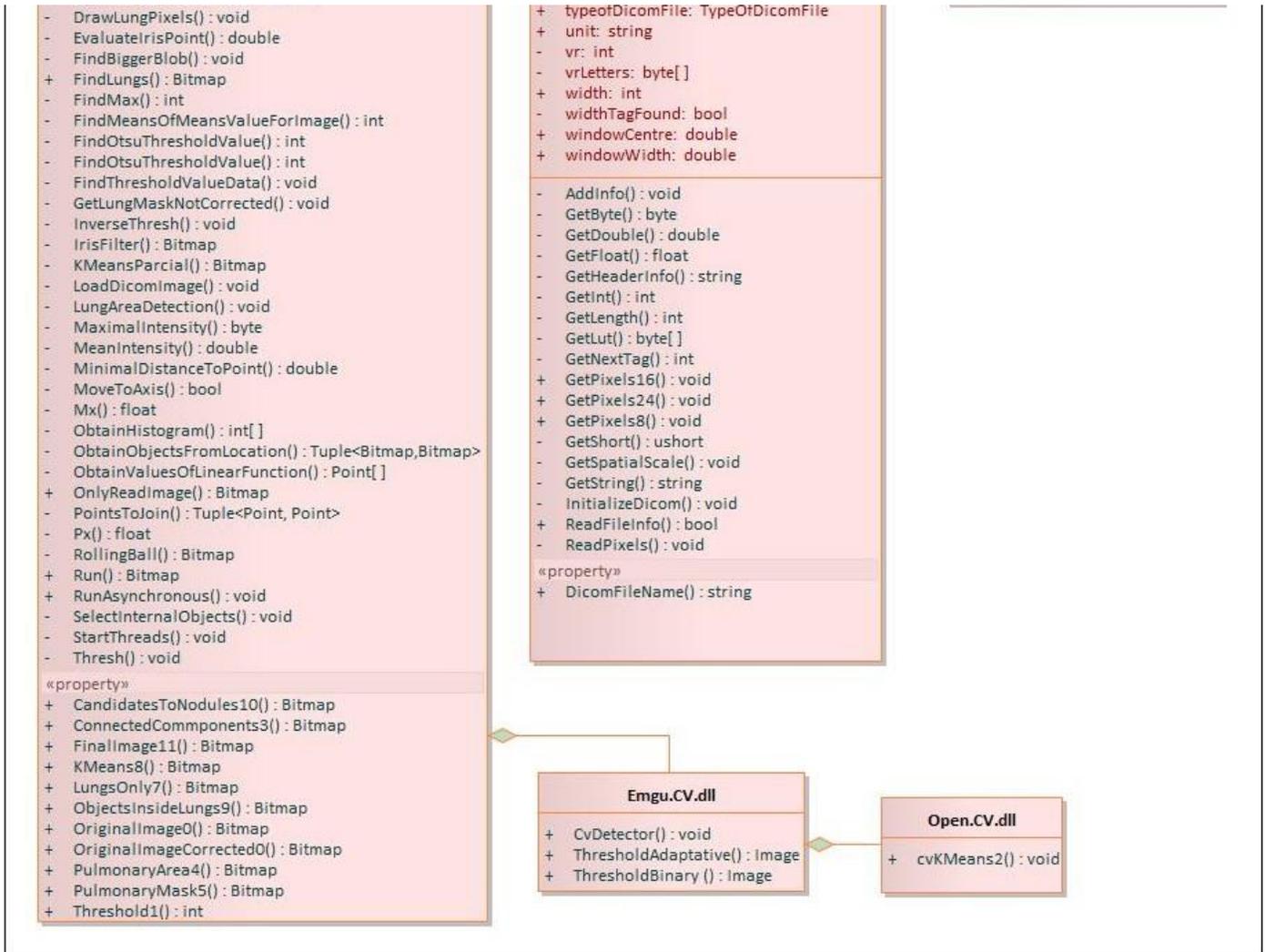


Figura 11. Diagrama de Clases del Diseño asociado al algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax (fuente: elaboración propia).

Con el objetivo de alcanzar un mayor entendimiento de las actividades que se llevan a cabo en el caso de uso referente al algoritmo, se realizó el Diagrama de Secuencia del Diseño. Estos diagramas permiten representar las interacciones que existen entre un conjunto de objetos y sus relaciones, así como los mensajes que son enviados entre ellos en orden temporal (60). En la Figura 13 se muestra el Diagrama de Secuencia del Diseño correspondiente al algoritmo.

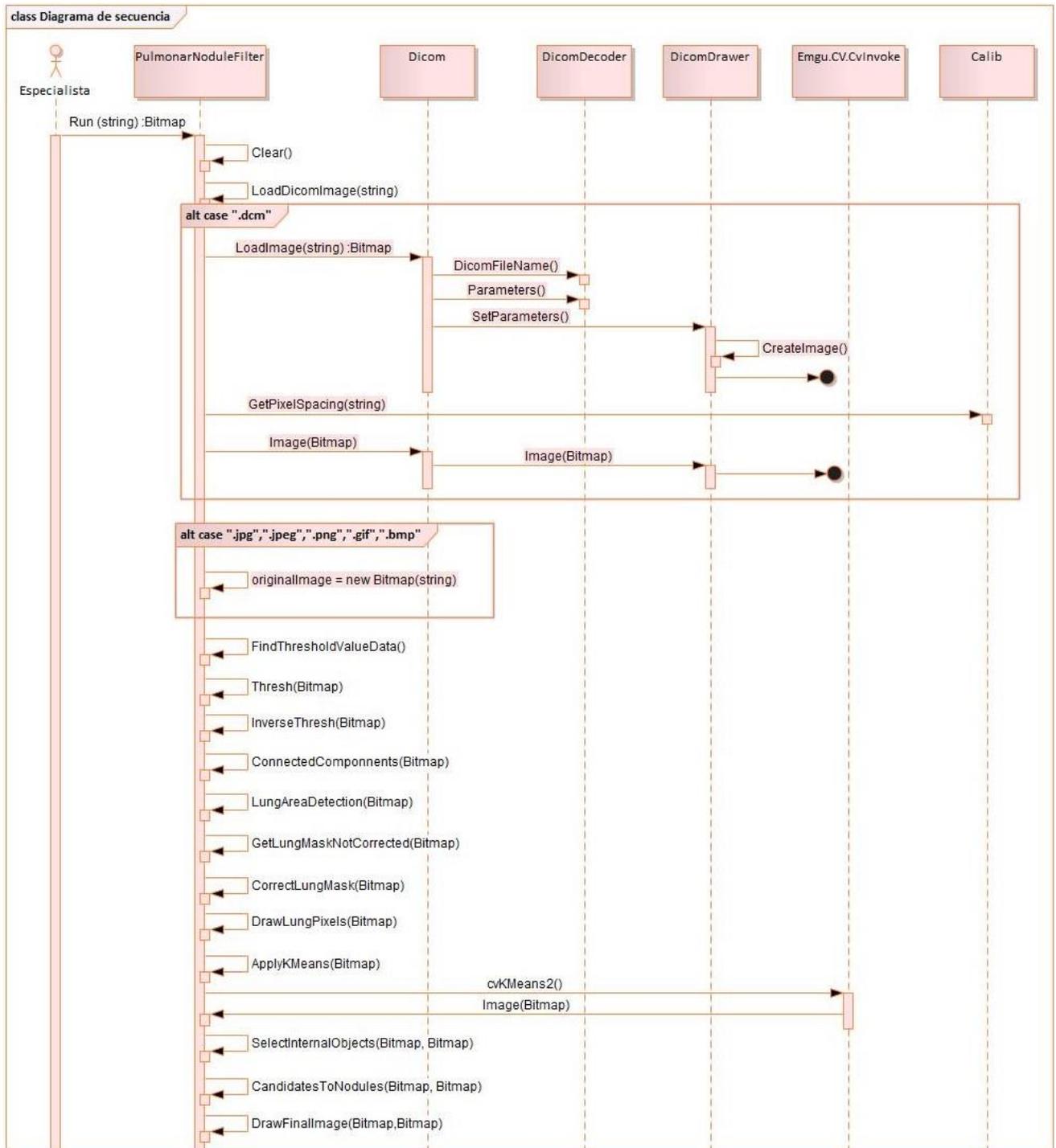


Figura 12. Diagrama de Secuencia del Diseño del algoritmo para la identificación de nódulos pulmonares solitarios (fuente: elaboración propia).

3.2.1. Descripción del diagrama de secuencia

Para hacer uso del algoritmo, el especialista debe utilizar un sistema de visualización de imágenes médicas, encargado de cargar una imagen de tomografía de tórax. El especialista debe seleccionar la opción “Identificar nódulos pulmonares”, para que el sistema permita visualizar las estructuras candidatas a nódulos, la imagen en análisis transcurre por una serie de transformaciones que se describen a continuación:

Si la imagen no se encuentra en formato DICOM se obvian los pasos del 4 al 6.

1. Se ejecuta el método *Run* que posee en su interior todos los pasos a ejecutar. Este método se encuentra en la clase *PulmonarNoduleFilter*.
2. Se ejecuta el método *Clear* para inicializar los valores de la clase.
3. Se ejecuta el método *LoadDicomImage*, que se implementa en la misma clase *PulmonarNoduleFilter*.
4. La función *LoadDicomImage* invoca a *LoadImage* de la clase *Dicom*, que utiliza la propiedad *DicomFileName*. Dicha propiedad analiza el tipo de imagen DICOM que se encuentra en la clase *DicomDecoder*.
5. La clase *DicomDecoder* tiene la propiedad *Parameters* que retorna los parámetros a *Dicom*; clase encargada de llamar a la función *SetParameters* de la clase *DicomDrawer* pasándole los parámetros obtenidos de *DicomDecoder*. *DicomDrawer* mediante el método *CreateImage* crea la imagen a partir de estos parámetros.
6. La clase *DicomDrawer* mediante la propiedad *Image* devuelve a *Dicom* la imagen creada y *Dicom* retorna dicha imagen a la clase *PulmonarNoduleFilter*.
7. Mediante la funcionalidad *FindThresholdValueData* se calcula el umbral óptimo para segmentar la imagen.
8. Para segmentar la imagen se utiliza el método *Thresh* de la clase *PulmonarNoduleFilter* y el umbral obtenido en el paso anterior.
9. Se invierte la imagen con la funcionalidad *InverseThresh* que se encuentra en la clase *PulmonarNoduleFilter*.

10. Al concluir la inversión de la imagen se aplica la técnica de componentes conexas mediante la funcionalidad *ConnectedComponents* de la clase *PulmonarNoduleFilter*.
11. Se usa el método *LungAreaDetection* para determinar las componentes conexas que representan a la región pulmonar.
12. Para obtener una primera aproximación a la máscara pulmonar se utiliza la funcionalidad *GetLungMaskNotCorrected* contenida en *PulmonarNoduleFilter*.
13. Para corregir la máscara obtenida en el paso 11 se emplea la función *CorrectLungMask*.
14. Para obtener los píxeles de la imagen original que corresponden a la región pulmonar se emplea el método *DrawLungPixels*.
15. Para obtener el agrupamiento mediante *K-Means* se emplea el método *ApplyKMeans* de la clase *PulmonarNoduleFilter*; pero requiere de la funcionalidad *cvKMeans2* que se encuentra en la librería *Emgu.CV.CvInvoke*.
16. El método *SelectInternalObjects* permite obtener la clase cuyos píxeles en la imagen original poseen la mayor intensidad media, en esta clase estarán candidatos a nódulos iniciales.
17. En el método *CandidatesToNodules* se identifican varias estructuras que por sus características no pueden ser NPS, por lo que se precede a la eliminación de las mismas.
18. Cuando se han identificado las estructuras candidatas a ser NPS, el algoritmo las visualiza haciendo uso del método *DrawFinalImage*.

Para una mejor comprensión de la lectura de la imagen, en la Figura 14 se muestra el pseudocódigo del método *LoadDicomImage* de la clase *PulmonarNoduleFilter*.

```
1 FUNCTION LoadDicomImage ( string: source)
2 INICIO
3     Si File.Exists(source) Hacer
4         var ext= Path.GetExtension(source);
5         switch (ext.ToLower())
6         {
7             Caso ".dcm":
8                 this.originalImage = this.reader.LoadImage(source);
9                 this.scale = GetPixelSpacing(source);
10                Salir;
11
12                Caso ".jpg":
13                Caso ".jpeg":
14                Caso ".png":
15                Caso ".gif":
16                Caso ".bmp":
17                    this.originalImage = new Bitmap(source);
18                    Salir;
19
20                default:
21                    throw new Exception("Archivo en formato invalido");
22                }
23            Si (this.originalImage == null)
24            {
25                Image<Bgr, byte> i = new Image<Bgr, byte>(new Size(500, 500));
26                this.originalImage = i.ToBitmap();
27            }
28
29            Si (this.ReadOriginalImageEvent != null) this.ReadOriginalImageEvent(this, this.originalImage);
30        }
31        Si no throw new InvalidOperationException("La direccion suministrada no es valida");
32        Si (this.BeginAnalisisEvent != null) this.BeginAnalisisEvent(this);
33    }
```

Figura 13. Pseudocódigo del método LoadDicomImage de la clase PulmonarNoduleFilter (fuente: elaboración propia).

3.3. Patrones de diseño

Un patrón de diseño es una descripción de las clases y los objetos comunicándose entre sí. Los patrones ayudan a capturar conocimiento y hacen el diseño orientado a objetos más flexible y reutilizable, posibilitando obtener un software de calidad (61). En el desarrollo de la presente investigación se utilizan varios patrones de diseño orientado a objetos para la asignación de responsabilidades, por ser una buena práctica en el diseño de software (55). Entre los patrones utilizados se encuentran el Controlador, perteneciente a los Patrones de Principios Generales para Asignar Responsabilidades (GRASP, por sus siglas en inglés) y de los estructurales el Adaptador.

3.3.1. Controlador

El patrón Controlador sirve como intermediario entre una determinada interfaz y el algoritmo que la implementa, recibe los datos del usuario y los envía a las distintas clases según el método que sea llamado.

Este patrón sugiere que la lógica de negocios debe estar separada de la capa de presentación para aumentar la reutilización de código y el control sobre la aplicación, así como facilitar las actividades de validación y seguridad (62). Esto se evidencia cuando el especialista ejecuta el algoritmo desde una interfaz, la cual invoca al método *Run* contenido en la clase *PulmonarNoduleFilter*.

Problema: ¿quién debe ser el responsable de gestionar las entradas al algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de CT?

Solución: con el objetivo de dar solución al problema se utiliza el patrón Controlador.

Aplicación: la clase *PulmonarNoduleFilter* cuenta con la funcionalidad *Run*, la cual es la encargada de ejecutar todas las transformaciones a la imagen de CT con el fin de detectar las estructuras nodulares.

3.3.2. Adaptador

Se utiliza para transformar una interfaz en otra, de tal modo que una clase que no pudiera utilizar la primera, haga uso de ella a través de la segunda (35). Esto se evidencia cuando se requiere la utilización de la librería OpenCV para el procesamiento de imágenes médicas pero esta librería no es compatible con Visual Estudio 2015. Para dar solución a este inconveniente se utilizó el adaptador EmguCV, con el objetivo de utilizar las funcionalidades con que cuenta OpenCV.

Problema: ¿cómo se puede utilizar algo que es útil y no es compatible con el sistema?

Solución: se utiliza el patrón Adaptador.

Aplicación: en la presente investigación se hace necesario la utilización de la librería OpenCV pero no es compatible con Visual Estudio 2015 por lo que se utilizó el adaptador (*wrapper*) EmguCV.

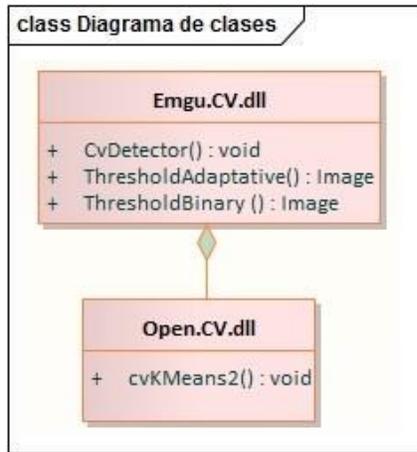


Figura 14. Utilización del patrón Adaptador en el algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de CT (fuente: elaboración propia).

3.4. Conclusiones del capítulo

Con la realización de este capítulo se llegaron a las siguientes conclusiones:

- ✓ La realización del Diagrama de Clases del Diseño facilitó la implementación del algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de CT.
- ✓ Se definió que el modelo arquitectónico Tuberías y Filtros es el más adecuado para la organización del algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de CT debido a la ejecución secuencial de las técnicas empleadas, teniendo en cuenta que la salida de una actividad es la entrada de la siguiente.
- ✓ Se determinó el empleo del patrón Controlador para la asignación de responsabilidades y del patrón Adaptador para utilizar las funcionalidades con que cuenta *OpenCV*.

CAPÍTULO 4: IMPLEMENTACIÓN DEL ALGORITMO PARA LA DISMINUCIÓN DE FALSOS POSITIVOS EN LA IDENTIFICACIÓN DE NÓDULOS PULMONARES SOLITARIOS EN IMÁGENES DE TOMOGRAFÍA DE TÓRAX.

En el presente capítulo se describe el Diagrama de Componentes utilizados para la implementación del algoritmo. Se relacionan los principales métodos en pseudocódigo desarrollados para lograr la disminución de falsos positivos en la identificación de las estructuras nodulares en las imágenes de CT, así como el estándar de codificación utilizado. Mediante el diseño de un experimento se presentan los resultados alcanzados por el algoritmo en cuanto a los valores de sensibilidad y promedio de falsos positivos, realizando una comparación con el algoritmo desarrollado por (6) con imágenes previamente diagnosticadas disponibles en The Lung Image Database Consortium Image Collection (LIDC-IDRI).

4.1. Diagrama de Componentes del algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de tomografía de tórax

El Diagrama de Componentes incluye una especificación de cómo los componentes deben empaquetarse para su despliegue como entidades independientes y ejecutables. Representa la estructura física del código, es utilizado para estructurar el Modelo de Implementación en términos de subsistemas y modelar la vista estática de un sistema, describiendo sus elementos físicos y las relaciones entre estos. Representa cómo un sistema es dividido en componentes y evidencia la organización y las dependencias lógicas entre estos (50, 63). En la Figura 16 se muestra el Diagrama de Componentes referente al algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de tomografía de tórax.

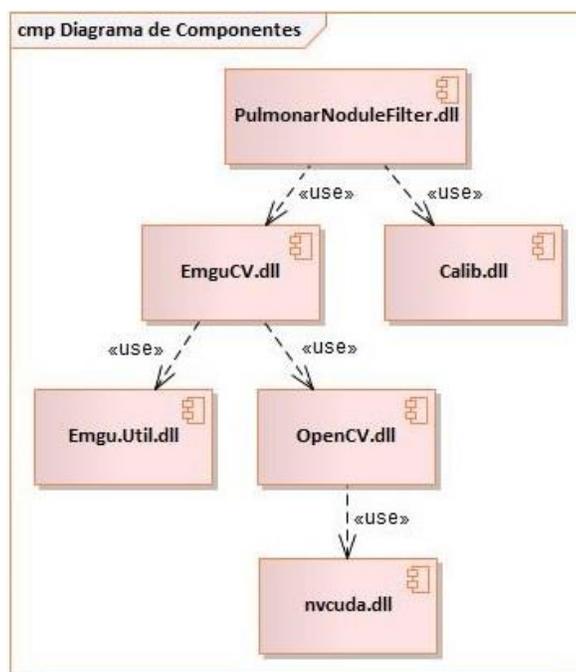


Figura 15. Diagrama de Componentes del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de CT (fuente: elaboración propia).

Para alcanzar una mejor comprensión se realiza una breve descripción de cada uno de los componentes representados en el diagrama, como se evidencia en la Tabla 8.

Tabla 8. Descripción de los elementos que conforman en el Diagrama de Componentes del algoritmo para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax (fuente: elaboración propia).

Componente	Descripción
PulmonarNoduleFilter.dll	Librería para la disminución de falsos positivos en la detección de nódulos pulmonares en imágenes de tomografía de tórax. Resultado de la presente investigación.
EmguCV.dll	Adaptador de la librería OpenCV, para el trabajo en .NET.
OpenCV.dll	Librería para el procesamiento de imágenes médicas.
Emgu.Util.dll	Librería que ofrece las utilidades para el trabajo con .NET.
nvcuda.dll	Librería utilizada por OpenCV, para el procesamiento gráfico.

Calib.dll	Librería utilizada para obtener el espaciado de cada píxel en la imagen DICOM.
-----------	--

4.2. Estándar de codificación

Para lograr la uniformidad y una mayor comprensión del código perteneciente al algoritmo para la disminución de falsos positivos en la identificación de NPS en imágenes de tomografía de tórax por otros desarrolladores, se utilizó un estándar de codificación que se muestra a continuación, utilizado en el proyecto PACS-RIS perteneciente al CESIM (64):

- ✓ Usar nombres descriptivos para los nombres de clases, propiedades y métodos. Asignar los nombres en idioma inglés, siempre garantizando que estén escritos correctamente.
- ✓ Se utiliza la notación *CamelCase* para declarar los nombres de los atributos de las clases con letra inicial minúscula. Si es un nombre compuesto se utiliza minúscula y mayúscula. Ejemplos: *angles*, *bodyArea*.
- ✓ Se utiliza la notación *PascalCase* para declarar los nombres de los métodos, clases y valores enumerables en mayúscula. Si son palabras compuestas se escribe mayúscula al inicio de cada palabra. Ejemplos: *Run*, *PulmonarNoduleFilter*, *TypeOfDicomFile*.
- ✓ Escribir comentarios en las clases y en los métodos, así como a las instrucciones, que sean de mediano a alto nivel de complejidad.

4.3. Pseudocódigo de los principales métodos del algoritmo para la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax

El propósito que persiguen los autores de esta investigación con la especificación del pseudocódigo de los principales métodos del algoritmo desarrollado, es facilitar la comprensión y reproducción de las técnicas y mecanismos utilizados para futuras implementaciones. Se hace énfasis en las técnicas y algoritmos utilizados para la segmentación de imágenes médicas e identificación de estructuras con altos valores de densidad, por la relevancia que representan en la investigación desarrollada.

4.3.1. Calcular umbral para la segmentación

El algoritmo que permite calcular el umbral empleado en la segmentación evoluciona del modo siguiente:

1. Se da un valor inicial al umbral.

2. Se segmentan los píxeles de la imagen en dos clases a partir de ese umbral.
3. Se calcula la media de los píxeles de cada clase por separado y luego se calcula la media entre estas dos clases.
4. Si el umbral es igual a la media resultante del paso anterior, se devuelve ese umbral como resultado, sino, se repite el proceso usando como umbral a la media del paso 3.

```
1  FUNCTION CalcularUmbralMediaDeMedias (Imagen: imagenOriginal, entero:ancho, entero: alto)
2  INICIO
3      var umbral, sal: enteros;
4      medPul, medNoPul, t1: decimales;
5      segmentación: matriz [alto, ancho];
6      umbral=251;
7      sal=0;
8      Mientras sal = 0 Hacer
9          segmentación = Segmentar (imagenOriginal, alto, ancho, umbral);
10         medPul = CalcularMedia (imagenOriginal, alto, ancho,255, segmentación);
11         medNoPul = CalcularMedia (imagenOriginal, alto, ancho,0, segmentación);
12         t1= (medNoPul + medPul) / 2;
13         Si (t1 = umbral)
14             sal = 1;
15             Sino umbral = t1;
16         FinSi
17     FinMientras
18     Retornar umbral;
19 FIN
```

Figura 16. Pseudocódigo correspondiente al método desarrollado para calcular el umbral (fuente: elaboración propia).

4.3.2. Aplicar rolling-ball a la máscara pulmonar

Para corregir la máscara pulmonar se ejecutan los siguientes pasos.

1. Se aplica a todos los contornos de la máscara sin corregir.
2. Se recorre punto por punto el borde del contorno que se analiza.
3. Al punto actual, junto al anterior, se calcula la pendiente de la recta que los une.
4. Construir varias esferas de forma tal que la recta comprendida por el radio que une al centro de dicha esfera con el punto actual tenga como pendiente la inversa de la pendiente que se calculó anteriormente.
5. Si el borde de dicha esfera se intersecta con el contorno en otro punto, además del punto que se analiza actualmente, estos puntos deben ser unidos por una recta.
6. Rellenar los huecos que queden entre el contorno y la recta que se dibujó.

```

1  FUNCTION RollingBall (Imagen: mascara, entero: alto, entero: ancho)
2  INICIO
3      var answer: Imagen [alto, ancho];
4      array: Point [];
5      sphere: Circle;
6      deep, jump, x1, y1: enteros;
7      points: Tupla<Point, Point>;
8      inverseDegree: decimal;
9      deep = 5;
10     jump = 3;
11     contour: Contour<Point>;
12     Desde contour Hasta contour = null contour = contour. Next Hacer
13         array = contour.ToArray();
14         Desde radio = 1 Hasta 8 radio++ Hacer
15             Desde x = jump Hasta contour. Length x = x + jump Hacer
16                 answer[array[x]. Y, array[x].X,0] = 255;
17                 inverseDegree = -1 / CalculateDegree (array [x - jump], array[x]);
18                 sphere = this. BuildSphere (inverseDegree, array[x], currentContour, radio * deep);
19                 Si sphere. Radius > 0
20                     Si contour. InContour (sphere. Center) > 0
21                         Continuar;
22                     FinSi
23                     x1 = sphere. Center.X;
24                     y1 = sphere. Center. Y;
25                     points = PointsToJoin (sphere, contour, Imagen [alto, ancho]);
26                     answer. Draw (Segmento (points. Item1, points. Item2), Gray (255));
27                 Sino
28                     Continuar
29                 FinSi
30             FinDesde
31         FinDesde
32         answer. Draw (contour, Gray (255));
33     FinDesde
34     Retornar answer;
35 FIN

```

Figura 17. Pseudocódigo correspondiente al método desarrollado para aplicar la técnica rolling-ball a la máscara pulmonar (fuente: elaboración propia).

En la implementación del método *rolling-ball*, como las operaciones no dependen una de las otras, los autores de la presente investigación deciden realizarlas en paralelo para agilizar el proceso. Se utilizó la clase *Parallel* (65) de C#, para evaluar en paralelo las diversas circunferencias en cada punto, quedando de la línea 14 a la 31 mostrada en la Figura 18, como se evidencia en la Figura 19.

```

14 Parallel.For (1, 8, radio=> {
15 Desde x = jump Hasta contour. Lenght x = x + jump Hacer
16 answer[array[x]. Y, array[x].X,0] = 255;
17 inverseDegree = -1 / CalculateDegree (array [x - jump], array[x]);
18 sphere = this. BuildSphere (inverseDegree, array[x], currentContour, radio * deep);
19 Si sphere. Radius > 0
20     Si contour. InContour (sphere. Center) > 0
21         Continuar;
22     FinSi
23     x1 = sphere. Center.X;
24     y1 = sphere. Center. Y;
25     points = PointsToJoin (sphere, contour, Imagen [alto, ancho]);
26     answer. Draw (Segmento (points. Item1, points. Item2), Gray (255));
27 Sino
28     Continuar
29 FinSi
30 FinDesde
31 });

```

Figura 18. Pseudocódigo correspondiente al método desarrollado para aplicar la técnica rolling-ball a la máscara pulmonar con radios asíncronos (fuente: elaboración propia).

4.3.3. Algoritmo K-Means

El algoritmo *K- Means* se aplica de siguiente manera:

1. Se selecciona la imagen con la región pulmonar separada del resto.
2. Se aplica el algoritmo *K- Means* a esta imagen, con cuatro grupos o clases.
3. Devolver la imagen resultante de etiquetar los píxeles de cada clase con un color que represente a su clase y lo diferencie de los píxeles de las demás clases.

```

1 FUNCTION AplicarKMeans (Imagen: RegionPulmonar, entero: alto, entero: ancho)
2 INICIO
3     var centroides: entero;
4     resultado: Imagen [alto, ancho];
5     centroides = 4;
6     resultado = KMeans (RegionPulmonar, alto, ancho, centroides);
7     Retornar resultado;
8 FIN

```

Figura 19. Pseudocódigo correspondiente al método desarrollado para aplicar la técnica K-Means (fuente: elaboración propia).

4.3.4. Obtener estructuras internas del pulmón

Para separar las estructuras internas del pulmón, se siguen los siguientes pasos:

1. Se crea un arreglo de imágenes, con tamaño igual a la cantidad de clases usadas en *K-Means* llamado *derivadasOriginal* y un arreglo de colores unidimensional donde en cada posición se

almacenará uno de los colores usados en el etiquetado de *K-Means*, el cual se llamará “colores”. Se debe ignorar tanto en el arreglo de imágenes, como en el arreglo de colores, al grupo que representa la parte que se encuentra fuera de los pulmones. Este color se guarda en la variable “ignorado”.

- Al recorrer la imagen de *K-Means*, cuando se encuentre un píxel de color diferente a ignorado, buscar su posición en colores, almacenarlo en la variable “índice”. A la imagen que esté en la posición igual al valor de índice en *derivadasOriginal*, guardar el píxel de la máscara pulmonar que se encuentra en la misma posición que el que se está evaluando en la imagen de *K-Means*. Dicho píxel también estará en la misma posición que se encontraba en la imagen original.
- Calcular la media de las intensidades de los píxeles de cada una de las imágenes de *derivadasOriginal*. En esta media se deben ignorar los píxeles que fueron agregados en el paso anterior, pues el resto son píxeles ruidosos en esta imagen.
- Devolver la imagen con mayor media.

```

1  FUNCION ObtenerEstructurasInternas(Imagen: ResultadoKMeans, Imagen: MascaraPulmonar, entero: alto, entero: ancho)
2  INICIO
3      var ignorado: Color;
4      derivadasOriginal: Lista<Imagen [alto, ancho]>;
5      sumaTotalPíxeles, cantidadTotalPíxeles: Listas<enteros>;
6      colores: Lista<Color>;
7      índice, cant: enteros;
8      cant = 0;
9      ignorado = ResultadoKMeans [0,0];
10     Desde y = 0 Hasta alto y++ Hacer
11         Desde x = 0 Hasta ancho x++ Hacer
12             Si ResultadoKMeans [x, y] != ignorado
13                 Si colores. Contains (ResultadoKMeans [x, y])
14                     índice = colores. GetIndex (ResultadoKMeans [x, y]);
15                     derivadasOriginal[índice] [x, y] = MascaraPulmonar [y, x];
16                     sumaTotalPíxeles[índice] += MascaraPulmonar [y, x]. Intensidad;
17                     cantidadTotalPíxeles[índice]++;
18                 Sino
19                     colores. Add (ResultadoKMeans [y, x]);
20                     índice = cant;
21                     derivadasOriginal. Add (Imagen [alto, ancho]);
22                     sumaTotalPíxeles. Add (MascaraPulmonar [y, x]. Intensidad);
23                     cantidadTotalPíxeles. Add (1);
24                     derivadasOriginal[índice] [y, x] = MascaraPulmonar [y, x];
25                     cant++;
26             FinSi
27         FinSi
28     FinDesde
29     índice = 0;
30     cant = sumaTotalPíxeles[índice] / cantidadTotalPíxeles[índice];
31     Desde x = 0 Hasta x < sumaTotalPíxeles. Length x++ Hacer
32         Si sumaTotalPíxeles[x] / cantidadTotalPíxeles[x] > cant
33             cant = sumaTotalPíxeles[x] / cantidadTotalPíxeles[x];
34             índice = x;
35         FinSi
36     FinDesde
37     Retornar derivadasOriginal[índice];
38
39  FIN

```

Figura 20. Pseudocódigo correspondiente al método desarrollado para obtener las estructuras internas del pulmón (fuente: elaboración propia).

4.4. Resultados alcanzados por el algoritmo de disminución de falsos positivos en la identificación de nódulos pulmonares solitarios en imágenes de tomografía de tórax

Los índices de incidencia y prevalencia del cáncer a nivel mundial han propiciado que diversos investigadores potencien el desarrollo de CAD. Con el objetivo de facilitar las investigaciones realizadas y el proceso de comprobación de los resultados obtenidos, se han creado bases de datos que concentran el criterio de un grupo de especialistas médicos. Enmarcadas en el tema de identificación de NPS, utilizando imágenes de tomografía de tórax, destacan bases de datos como:

- ✓ *The Lung Image Database Consortium Image Collection-Image Database Resource Initiative (LIDC-IDRI)* (40, 66, 67).
- ✓ *Japanese Society of Radiological Technology (JSRT)* (68, 69).
- ✓ *Early Lung Cancer Action Program Public Lung Image Database (ELCAP)* (67).

Entre las bases de datos relacionadas anteriormente, destaca la LIDC-IDRI por su elevado número de estudios y su especificidad en la descripción de las estructuras nodulares contenidas en las imágenes médicas. Para que los resultados de un algoritmo desarrollado sean comparables con otros, deben utilizarse iguales criterios o aspectos de medición (en este caso, sensibilidad y falsos positivos), los mismos datos y ambiente de prueba y validación. Por lo anteriormente planteado, se decide la utilización de la base de datos LIDC-IDRI para la validación del presente algoritmo, teniendo en cuenta de forma adicional que fue la base de datos seleccionada por los autores de las investigaciones precedentes.

4.4.1. Indicadores para evaluar el desempeño del procedimiento de diagnóstico

Para verificar el adecuado funcionamiento del presente algoritmo, se hace necesario el cálculo de los indicadores para evaluar el rendimiento del proceso de diagnóstico. El resultado de una prueba diagnóstica puede ser positivo o negativo, pero estos pueden ser correctos e incorrectos, dando lugar a cuatro tipos de resultados como se evidencia en la Tabla 9.

Tabla 9. Relación entre los resultados de una prueba diagnóstica (70).

Resultados de la prueba	Es un nódulo	No es un nódulo
El algoritmo señala la estructura	verdadero positivo (VP)	falso positivo (FP)
El algoritmo no señala la estructura	falso negativo (FN)	verdadero negativo (VN)

- ✓ VP: como resultado de la clasificación el algoritmo señala el objeto en presencia de una anomalía clínica.
- ✓ VN: como resultado de la clasificación el algoritmo no señala el objeto en ausencia de la anomalía clínica.
- ✓ FP: como resultado de la clasificación el algoritmo señala el objeto en ausencia de la anomalía clínica.
- ✓ FN: como resultado de la clasificación el algoritmo no señala el objeto en presencia de la anomalía clínica.

El promedio de falsos positivos (*PFP*), representa la cantidad de falsos positivos encontrados en total en una prueba, dividido por la cantidad de imágenes (*CI*) analizadas en dicha prueba.

$$PFP = \frac{FP}{CI} \quad [6]$$

La validez se define como la capacidad de un instrumento para medir lo que intenta medir. Esta sólo puede determinarse si existe un procedimiento de referencia, también conocido como estándar de oro, el cual, es considerado como un procedimiento definitivo para establecer si alguien tiene la característica de interés. Los parámetros que miden la validez de una prueba diagnóstica son la sensibilidad y la especificidad (70).

La sensibilidad (*S*) es la proporción de los individuos que tienen la enfermedad y el resultado de la prueba realizada da positivo, es decir, la proporción de verdaderos positivos, o de enfermos diagnosticados, respecto del total de enfermos en la población de estudio. Cuanto más alto es el valor de *S*, hay una mejor capacidad en la detección de enfermos por medio de la prueba (71):

$$S = \frac{VP}{VP + FN} \quad [7]$$

La especificidad (*E*) es la habilidad para identificar aquellos individuos que no tienen cáncer. Es la proporción de los verdaderos negativos, respecto al total de individuos sanos en la población de estudio. Mientras mayor sea el valor de *E*, hay una mejor capacidad en la detección de individuos sanos:

$$E = \frac{VN}{VN + FP} \quad [8]$$

Para evaluar la efectividad de un CAD, se utiliza el indicador de precisión (P), el cual se calcula a partir de los indicadores sensibilidad y especificidad:

$$P = \frac{VP + VN}{VP + VN + FP + FN} \quad [9]$$

En la Tabla 10 se evidencia una relación de valores de precisión, que especifica qué tan buenos son los resultados alcanzados por un CAD, a partir de los resultados obtenidos al medir este indicador.

Tabla 10. Relación entre los valores de precisión y los conjuntos que definen (72).

Precisión	Etiqueta
90%-100%	Excelente
80%-90%	Muy Bien
70%-80%	Bien
60%-70%	Suficiente
50%-60%	Malo
< 50%	No es de utilidad

En la presente investigación no es posible la medición del indicador especificidad, teniendo en cuenta que no es factible la contabilización de los VN. A pesar de ello, a partir del análisis de los indicadores implicados en el cálculo de la precisión, se identifica que es necesaria la obtención de altos valores de sensibilidad, con un reducido número de falsos positivos por imagen médica analizada. De esta forma se garantiza desde la fase de identificación o segmentación de estructuras del CAD, la obtención de altos valores de precisión.

Al realizar un desglose de los indicadores S y E se evidencia que la aparición de FN y FP propician la disminución de los mencionados indicadores (S y E), favoreciendo que se hagan menores de 1 y por tanto que dichos valores sean menores al 100%. En dicho análisis se evidencia la relevancia de los valores de precisión en el diagnóstico médico. Para el paciente es crucial que el resultado de su prueba sea lo más

acertado posible, con el menor número de estructuras mal identificadas, ya sea porque son nódulos y no se detectaron, o porque no lo era y se les indicaron como nódulos.

4.4.2. Composición del banco de pruebas utilizado para la validación del algoritmo desarrollado

Las imágenes contenidas en la base de datos LIDC-IDRI están en formato DICOM (*Digital Imaging and Communications in Medicine*) y se encuentran anonimizadas para garantizar la privacidad de los pacientes. Para cada una de las series de imágenes que contiene la base de datos del consorcio, se dispone de un fichero que describe en formato XML, los hallazgos de los especialistas. Se especifican las estructuras que son consideradas NPS, a través de la descripción de un conjunto de coordenadas que representan su borde. Se puntualiza la imagen médica en la que está contenida; así como características de espiculación, lobulación, textura, densidad, entre otras características distintivas relativas a dichas estructuras (67).

Para señalar las lesiones presentes en las imágenes se utilizaron los criterios de 12 imagenólogos, que colaboran con el consorcio. Dichos especialistas realizaron una revisión de las imágenes, utilizando una interfaz de ordenador, en la cual se les permitió realizar mediciones y variar los parámetros de ancho y centro de ventana, para finalmente marcar las lesiones identificadas.

4.4.3. Selección de la muestra utilizada en el experimento para la validación del algoritmo desarrollado

El muestreo es una herramienta de la investigación científica, cuya función básica es determinar qué parte de una población debe examinarse, con la finalidad de hacer inferencias sobre dicha población. El Muestreo Intencional es un tipo de muestreo no probabilístico, el investigador selecciona los elementos que son representativos o con posibilidades de brindar mayor información (73).

Para la validación del algoritmo se realizó un experimento de tipo Intencional. La población seleccionada son las imágenes de LIDC-IDRI. La muestra utilizada son las series LIDC-IDRI-0001, LIDC-IDRI-0011, LIDC-IDRI-0012, LIDC-IDRI-0014, LIDC-IDRI-0016, LIDC-IDRI-0018, LIDC-IDRI-0021, LIDC-IDRI-0024, LIDC-IDRI-0027, siendo estas las nueve series utilizadas para validar el algoritmo desarrollado por (6). La muestra se compone por un total de 766 imágenes.

4.4.4. Análisis de los resultados obtenidos

Para la automatización del cálculo de los indicadores de validez del algoritmo desarrollado se realizó una correspondencia entre los resultados obtenido por el algoritmo y las anotaciones realizadas en el fichero

XML de la LIDC-IDRI, como mismo se realiza en (6). Se realizó una comparación entre los resultados obtenidos por el algoritmo desarrollado y el precedente en función de los valores de sensibilidad y falsos positivos por imagen médica analizada. En la Tabla 11, se muestran los detalles relacionados con dicha comparación.

Tabla 11. Resultados alcanzados por el algoritmo propuesto (AP) para la disminución de falsos positivos en la identificación de nódulos pulmonares solitarios conjuntamente con uno de los analizados en el estado del arte (fuente: elaboración propia).

Indicadores Series	Algoritmo precedente (7)		Algoritmo desarrollado	
	S	FPF	S	FPF
LIDC-IDRI-0001	0,89	5	1	4,66
LIDC-IDRI-0011	0,64	4,85	0,88	3,66
LIDC-IDRI-0012	0,69	4,88	0,87	6
LIDC-IDRI-0014	1,00	5,17	0,83	5
LIDC-IDRI-0016	0,65	5,03	0,85	4,66
LIDC-IDRI-0018	0,83	4,61	0,80	4,4
LIDC-IDRI-0021	0,83	4,80	0,80	8,4
LIDC-IDRI-0024	0,77	6,20	0,77	4,66
LIDC-IDRI-0027	0,76	4,98	0,66	1,75
Resultado Final	79%	5,02	<u>83%</u>	<u>4,75</u>

El algoritmo desarrollado por (6) obtuvo un valor de sensibilidad de 79% con un promedio de falsos positivos de 5,02. Al ejecutar el algoritmo desarrollado por los autores, se obtuvo un valor de sensibilidad de un 83% con un promedio de 4,75 falsos positivos por imagen médica analizada, evidenciándose una mejoría en los resultados. A pesar de que el objetivo de la investigación es la disminución del promedio de falsos positivos,

se evidencia un aumento representativo de los valores de sensibilidad, siendo mejor en un 4%, respecto a las investigaciones precedentes.

4.5. Conclusiones del capítulo

Con la realización de este capítulo se arribaron a las siguientes conclusiones:

- ✓ La utilización de las técnicas seleccionadas en el análisis del referente teórico de la investigación, facilitó el aumento de los indicadores sensibilidad y falsos positivos, respecto a las investigaciones precedentes.
- ✓ Se validaron los resultados alcanzados por el algoritmo desarrollado en esta investigación, utilizando 9 series contenidas en The Lung Image Database Consortium Image Collection. Se obtuvo una sensibilidad de un 83% y un promedio de 4,75 falsos positivos por imagen médica, para estructuras de 3-30 mm de diámetro.

CONCLUSIONES GENERALES

Al término de la investigación los autores arriban a las siguientes conclusiones:

- ✓ El análisis de la bibliografía consultada durante el desarrollo de la investigación, facilitó la selección de las técnicas *rolling-ball* y *K-Means* con el objetivo de alcanzar valores aceptables de sensibilidad.
- ✓ Utilizar *K-Means* como técnica de agrupamiento para la identificación de estructuras pulmonares dentro de la región pulmonar, permitió el aumento de los valores de sensibilidad respecto a las técnicas de umbralización utilizadas en investigaciones precedentes.
- ✓ Se constató la idoneidad del uso de las características morfológico-radiográficas de tamaño, circularidad y gradiente para la reducción de los valores de falsos positivos por imagen médica analizada.
- ✓ Al ejecutar el algoritmo utilizando un conjunto de imágenes públicas en la base de datos *The Lung Image Database Consortium Image Collection*, se obtuvo una sensibilidad de un 83% y un promedio de 4,75 falsos positivos por imagen médica, para estructuras de 3-30 mm de diámetro; siendo superior a los resultados alcanzados por las investigaciones precedentes.

RECOMENDACIONES

Para futuras investigaciones se recomienda:

- ✓ Realizar el preprocesamiento de las imágenes médicas a analizar; para mejorar su calidad de representación y por consiguiente facilitar el análisis e identificación de las estructuras nodulares.
- ✓ Realizar el procesamiento de las series de imágenes utilizando algoritmos en tres dimensiones para aumentar los resultados de sensibilidad alcanzados.
- ✓ Evaluar resultados utilizando técnicas de agrupamiento difuso como *fuzzy K-Means*, para mejorar la clasificación de los nódulos en cuanto a su probabilidad de pertenencia a una clase determinada.

REFERENCIAS BIBLIOGRÁFICAS

1. COTO, ERNESTO. Métodos de Segmentación de Imágenes Médicas. Marzo 2003.
2. Cáncer, principal causa de muerte a nivel mundial [en línea]. 2016. [Accedido 29 octubre 2016] Disponible en: <http://eleconomista.com.mx/entretenimiento/2016/02/09/cancer-principal-causa-muerte-nivel-mundial>.
3. Organización Mundial de la Salud [en línea]. 2017. [Accedido 29 octubre 2016] Disponible en: <http://www.who.int/mediacentre/factsheets/fs297/es/>.
4. BESS CONSTANTÉN, Sonia. Anuario Estadístico de Salud 2016. MINISTERIO DE SALUD PÚBLICA, 2017.
5. CALVA ARCOS, Mario y ACEVEDO TIRADO, María T. Revisión y actualización general en cáncer colorrectal. Anales de Radiología México. Marzo 2009. Vol. 1, p. 99–115.
6. RIVERO CASTRO, ARELYS. Algoritmo basado en técnicas de segmentación de imágenes de tomografía de tórax para aumentar el acierto en la identificación de nódulos pulmonares solitarios. MSc. La Habana: Universidad de las Ciencias Informáticas, 2014.
7. SUÁREZ CUENCA, JORGE JUAN. Desarrollo de un sistema de diagnóstico asistido por computador para detección de nódulos pulmonares en tomografía computarizada multicorte. PhD. Santiago de Compostela, 2009.
8. DOI, KUNIO. Current status and future potential of computer-aided diagnosis in medical imaging. The British Journal of Radiology. 2005. Vol. 78. DOI 10.1259/bjr/82933343.
9. LARA TAMBURRINO, MARÍA DEL CARMEN y OLMEDO ZORILLAB, ÁLVARO. Detección temprana y diagnóstico del cáncer mamario. Revista de la Facultad de Medicina de la UNAM. 2011. Vol. 54, no. 1.
10. GONZÁLEZ MARCOS, ANA; MARTÍNEZ DE PISÓN, FRANCISCO; ASCACÍBAR, ALPHA, JAVIER; PERNÍA ESPINOZA, VERÓNICA; ALBA ELÍAS, FERNANDO; CASTEJÓN LIMAS, MANUEL; ORDIERES MERÉ, JOAQUÍN y ELISEO VERGARA GONZÁLEZ. Técnicas y algoritmos básicos de visión artificial. Universidad de La Rioja, 2006. ISBN 84-689-9345-X.
11. BAE, KYONGTAE T; KIM, JIN-SUNG; NA, YONG-HUM; KIM, KWANG GI y KIM, JIN-HWAN. Pulmonary Nodules: Automated Detection on CT Images with Morphologic Matching Algorithm—Preliminary Results. Radiology. 2005. Vol. 236, p. 286–294. DOI 10.1148/radiol.2361041286.
12. ZHAO, BINSHENG; GAMSU, GORDON; GINSBERG, MICHELLE S; JIANG, LI y SCHWARTZ, LAWRENCE H. Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. Journal Of Applied Clinical Medical Physics. 2003. Vol. 4, no. 3, p. 248-260. DOI 10.1120/1.1582411.

13. COSTA MONTANÉ, DANIEL MARINO; PRADO LAGE, YULIEN; LOZANO SALAZAR, JORGE LUIS; PLASENCIA ASOREY, CAROLINA y RIESGO COSME, YALILI DE LA CARIDAD. Principales aspectos clinicoepidemiológicos del cáncer de pulmón. 2011. Vol. 15.
14. SUÁREZ CUENCA, JORGE JUAN; TAHOCES, PABLO G; LADO, M.J; SOUTO, M y VIDAL, J.J. Detección automática del nódulo pulmonar subpleural basada en el empleo de contornos activos. 2008. Vol. 1, p. 585–591.
15. EJAZ, NAVEED; JAVED, SHAZIA y SAJID, ZEESHAN. Implementation of Computer Aided Diagnosis System for Lung Cancer Detection. Lecture Notes on Software Engineering. Noviembre 2013. Vol. 1, no. 4. DOI 10.7763/LNSE.2013.V1.78.
16. HERNÁNDEZ SAMPIERÍ, Roberto, FERNÁNDEZ COLLADO, Carlos y BAPTÍSTA LUCIO, Pilar. Metodología de la investigación. 4. México: McGRAW-HILL, 2010. ISBN 970-10-5753-8.
17. CALZADO, A y GELEIJNS, J. Tomografía computarizada. Evolución, principios técnicos y aplicaciones. Revista Física Médica. 2011. Vol. 11, no. 3, p. 163-180.
18. SEPÚLVEDA, CRISTIÁN; SEPÚLVEDA, ALFREDO y FUENTES, ESTEBAN. Nódulo pulmonar solitario. Revista Chilena de Cirugía. Febrero 2008. Vol. 60, p. 71-78.
19. JEONG, Y. J; YI, C.A y LEE, K.S. Nódulos pulmonares solitarios: detección, caracterización y guías para su diagnóstico y tratamiento. 2008. Vol. 50, no. 3. DOI 10.1016/S0033-8338(08)71964-7.
20. National Institute of Biomedical Imaging and Bioengineering [en línea]. 2012. [Accedido 20 enero 2017]. Disponible en: <https://www.nibib.nih.gov/espanol/temas-cientificos/tomograf%C3%ADa-computarizada-tc>.
21. YANAGIHARA, TAKANOBU y TAKIZAWA, HOTAKA. Pulmonary Nodule Detection from X-ray CT Images Based on Region Shape Analysis and Appearance-based Clustering. Algorithms 2015. Mayo 2015. p. 209-223. DOI 10.3390/a8020209.
22. MO GOO, JIN. A Computer-Aided Diagnosis for Evaluating Lung Nodules on Chest CT: the Current Status and Perspective. Korean J Radiology. 2011. p. 145-155. DOI 10.3348/kjr.2011.12.2.145.
23. LLOBET, RAFAEL; PEREZ-CORTÉS, JUAN C; TOSELLI, ALEJANDRO H y ALFONS, JUAN. Computer-aided detection of prostate cancer. International Journal of Medical Informatics. 2006. p. 547–556. DOI 10.1016/j.ijmedinf.2006.03.001.
24. VOS, P C; BARENTSZ, J O; KARSSEMEIJER, N y HUISMAN, H J. Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis. Physics in Medicine and Biology. 2012. p. 1527-1542. DOI 10.1088/0031-9155/57/6/1527.
25. TANG, JINSHAN; MEMBER, SENIOR; RANGAYYAN, RANGARAJ M; XU, JUN; NAQA, ISSAM y YANG, YONGYI. Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent

- Advances. IEEE Transaction on Information Technology in Biomedicine. 2009. Vol. 3, no. 2, p. 236-251. DOI 10.1109/TITB.2008.2009441.
26. FREER, TIMOTHY W y ULISSEY, MICHAEL J. Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center. *Radiology*. 2001. p. 781-786. DOI 10.1148/radiol.2203001282.
27. VIJAYA, G y SUHASINI, A. An Adaptive Preprocessing of Lung CT Images with Various Filters for Better Enhancement. *Academic Journal of Cancer Research*. 2014. p. 179-184. DOI 10.5829/idosi.ajcr.2014.7.3.84231.
28. SHYAMALA, S y PUSHPARANI, M. Research Article Pre-Processing And Segmentation Techniques For Lung Cancer On Ct Images. *International Journal of Current Research*. 2016. Vol. 8, p. 31665-31668.
29. TINKU, ACHARYA y AJOY K, RAY. *Image Processing. Principles and Applications*. A JOHN WILEY & SONS, 2005. ISBN 13 978-0-471 -71998-4.
30. BENÍTEZ SÁNCHEZ, IGNACIO JAVIER. *Técnicas de Agrupamiento para el Análisis de Datos Cuantitativos y Cualitativos*. Valencia: Universidad Politécnica de Valencia, 2006.
31. MARTÍN-LANDROVE, MIGUEL. *Modelo de Sistema de Información para el Manejo de Imágenes Cerebrales Tomográficas de Apoyo al Diagnóstico*. 2015. p. 1-65. DOI 10.13140/RG.2.2.12958.13124.
32. MIRALLES CHAPLE, ANDRÉS. *Módulo De Segmentación De Imágenes Médicas Digitales Mediante Técnicas De Watershed*. La Habana: Universidad de las Ciencias Informáticas, 2011.
33. GURCAN, METIN N; SAHINER, BERKMAN; PETRICK, NICHOLAS; CHAN, HEANG-PING; KAZEROONI, ELLA A; CASCADE, PHILIP N y HADJIISKI, LUBOMIR. Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system. *Medical Physics*. Noviembre 2002. Vol. 29.
34. MAHMOOD, FALEH H; ABBAS, WAFAA A y ALI, S. M. Estimating the Lung Tumor size in CT images using Image Segmentation Techniques. *International Journal of Emerging Technology and Advanced Engineering*. Julio 2014. Vol. 4, p. 509-517.
35. BORGES GONZÁLEZ, YOSUANI y NARANJO GORRÍN, YOANNY. *Algoritmo para la identificación de nódulos pulmonares en imágenes de tomografía de tórax*. La Habana: Universidad de las Ciencias Informáticas, 2014.
36. AGUILLÓN PÉREZ, JULIÁN MAURICIO; DUARTE PACHECO, SERGIO ALEJANDRO y HERRERA GARCÍA, RODRIGO JAVIER. Realce de candidatos a nódulo pulmonar en radiografías de tórax por medio de filtros de convergencia. *INGENIERÍA*. 2014. Vol. 19, no. 2, p. 85-104.

37. K, BHAVANISHANKAR y SUDHAMANI, M.V. Techniques For Detection Of Solitary Pulmonary Nodules In Human Lung And Their Classifications-A Survey. *International Journal on Cybernetics & Informatics*. Febrero 2015. Vol. 4, no. 1. DOI 10.5121/ijci.2015.4103.
38. DL PUGOY, REINALD ADRIAN y Y MARIANO, VLADIMIR. Automated rice leaf disease detection using color image analysis. *SPIE*. 2011. Vol. 8009. DOI 10.1117/12.896494.
39. P, SINTHIA y SUJATHA, K. A Novel Approach To Detect Bone Cancer Using K-Means Clustering Algorithm And Edge Detection Method. *ARPN Journal of Engineering and Applied Sciences*. 2016. Vol. 11, no. 13, p. 8002-8007.
40. ARMATO, SAMUEL G; MCNITT-GRAY, MICHAEL F; REEVES, ANTHONY P; MEYER, CHARLES R; MCLENNAN, GEOFFREY; ABERLE, DENISE R; KAZEROONI, ELLA A; MACMAHON, HEBER; VAN BEEK, EDWIN J.R y YANKELEVITZ, DAVID. The Lung Image Database Consortium (LIDC): An Evaluation of Radiologist Variability in the Identification of Lung Nodules on CT Scans. *Academic Radiology*. noviembre 2007. Vol. 14, no. 11, p. 1409-1421. DOI 10.1016/j.acra.2007.07.008.
41. ARMATO III, SAMUEL G; L GIGER, MARYELLEN y MACMAHON, HEBER. Automated detection of lung nodules in CT scans: Preliminary results. *American Association of Physicists in Medicine*. 2001. Vol. 28, no. 8. DOI <http://dx.doi.org/10.1118/1.1387272>.
42. SHAHVARAN, ZAHRA; KAZEMI, KAMRAN; HELFROUSH, MOHAMMAD SADEGH y JAFARIAN, NASSIM. Region-based Active Contour Model based on Markov Random Field to Segment Images with Intensity Non-Uniformity and Noise. *Journal of Medical Signals and Sensors*. 2012. Vol. 2, no. 1, p. 17-24.
43. LIEBOWITZ, BURT H y CARSON, JOHN H. *Computer Architectures for image processing*. Computer society. 1981.
44. RETICO, A; DELOGU, P; FANTACCI, M.E; GORI, I y PREITE MARTINEZ, A. Lung nodule detection in low-dose and thin-slice computed tomography. *Computers in Biology and Medicine*. 2008. p. 525 – 534. DOI 10.1016/j.compbiomed.2008.02.001.
45. ARMATO, SAMUEL G; L GIGER, MARYELLEN; MORAN, CATHERINE J; BLACKBURN, JAMES T; DOI, KUNIO y MACMAHON, HEBER. Computerized Detection of Pulmonary Nodules on CT Scans. *Radiographics*. 1999. Vol. 19, no. 5. DOI <http://dx.doi.org/10.1148/radiographics.19.5.g99se181303>.
46. EJAZ, Naveed; JAVED, Shazia y SAJID, Zeeshan. Implementation of Computer Aided Diagnosis System for Lung Cancer Detection. *Lecture Notes on Software Engineering*. Noviembre 2013. Vol. 1, no. 4. DOI 10.7763/LNSE.2013.V1.78.
47. GIMÉNEZ, YANINA. *Clasificación no supervisada: El método de k-medias*. Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, 2010.
48. XU, RUI y WUNSCH, DONALD C. *Clustering*. IEEE Press, 2007.

49. RODRÍGUEZ SÁNCHEZ, TAMARA. Metodología de desarrollo para la Actividad productiva de la UCI. La Habana: Universidad de las Ciencias Informáticas, 2015.
50. SOMERVILLE, IAN. Software Engineering. 9. 2011. ISBN 978-0-13-703515-1.
51. Entorno Integrado de Desarrollo [en línea]. 2013. [Accedido 10 diciembre 2016]. Disponible en: <https://fergarciaac.wordpress.com/2013/01/25/entorno-de-desarrollo-integrado-ide/>.
52. Visual Studio [en línea]. 2017. [Accedido 10 diciembre 2016]. Disponible en: <https://www.visualstudio.com>.
53. New C# Features in the .NET Framework 4. [en línea]. 2010. [Accedido 10 diciembre 2016]. Disponible en: <http://msdn.microsoft.com/es-es/magazine/ff796223.aspx>.
54. SÁNCHEZ DIEGUEZ, MAIKEL y VALLÉS GÁMEZ, ANTONIO ENRIQUE. Framework CALIB. Procesos de lectura, procesamiento y visualización de imágenes médicas. La Habana: Universidad de las Ciencias Informáticas, 2010.
55. LARMAN, CRAIG. UML y Patrones. 2. 2003.
56. MathWorks [en línea]. 2017. [Accedido 8 febrero 2017]. Disponible en: <https://www.mathworks.com/help/images/examples/identifying-round-objects.html>.
57. PAJARES MARTÍNEZ, GONZALO y DE LA CRUZ GARCÍA, JESÚS M. Visión por Computador Imágenes digitales y aplicaciones. 2. RA-MA, 2008. ISBN 978-84-7897-831-1.
58. C GONZALEZ, RAFAEL y WOODS, RICHARD E. Digital Image Processing. Second. New Jersey : Jody McDonnell, 2002. ISBN 0-201-18075-8.
59. WEI, JUN; HAGIHARA, YOSHIHIRO y KOBATAKE, HIDEFUMI. Detection of cancerous tumors on chest X-ray images - candidate detection filter and its evaluation. International Conference on Image Processing. 2002. Vol. 3, p. 397-401. DOI 10.1109/ICIP.1999.817143.
60. KUCHANA, PARTHA. Software Architecture Design Patterns in Java. New York: CRC/AUERBACH PUBLICATIONS, 2004. ISBN 0-8493-2142-5.
61. PAVÓN MESTRAS, JUAN. Patrones de diseño orientado a objetos. Madrid, 2004.
62. SPIEGEL, JHON. GRASP Design Patterns: Designing Objects with Responsibilities. 2004.
63. PRESSMAN, ROGER S. Software Engineering. A Partitioner's Approach. 7ma. 2010. ISBN 978-0-07-337597-7.
64. Microsoft, Capitalization Styles [en línea]. 2017. [Accedido 28 abril 2017]. Disponible en: [https://msdn.microsoft.com/en-us/library/x2dbyw72\(v=vs.71\).aspx](https://msdn.microsoft.com/en-us/library/x2dbyw72(v=vs.71).aspx).

65. Microsoft [en línea]. 2016. [Accedido 12 abril 2017]. Disponible en: [https://msdn.microsoft.com/es-es/library/system.threading.tasks.parallel\(v=vs.110\).aspx](https://msdn.microsoft.com/es-es/library/system.threading.tasks.parallel(v=vs.110).aspx).
66. KURAVATI, R; SASIDHAR, B y RAMESH, D. A Novel Method for Classification of Lung Nodules as Benign and Malignant using Artificial Neural Network. *International Journal of Engineering and Computer Science*. 2014. Vol. 3, no. 8, p. 7641-7645.
67. ARMATO III, SAMUEL G; MCLENNAN, GEOFFREY; BIDAUT, LUC; MCNITT-GRAY, MICHAEL F; MEYER, CHARLES R; REEVES, ANTHONY P; ZHAO, BINSHENG; ABERLE, DENISE R; HENSCHKE, CLAUDIA I y ERIC A HOFFMAN. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) : A Completed Reference Database of Lung Nodules on CT Scans. *Medical Physics*. Febrero 2011. Vol. 38, no. 2, p. 915-931.
68. S, HAMADA; B, BRAHIM y SUZIAH S. A Computer Aided Diagnosis System for Lung Cancer based on Statistical and Machine Learning Techniques. *Journal of Computers*. 2014. Vol. 9, no. 2, p. 425-431.
69. Shiraishi, J; Katsuragawa, S; Ikezoe, J; Matsumoto, T; Kobayashi, T; Komatsu, K; Matsui, M; Fujita, H; Kodera, Y y Doi, K. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*. 2000. Vol. 174, no. 1, p. 71-79.
70. CUEVAS RENAUD, CORINA y ALEJO MARTÍNEZ, AMALIA. Sensibilidad y Especificidad de una prueba [en línea]. 2010. Disponible en: <http://www.psicol.unam.mx/Investigacion2/pdf/SENSIBILIDAD%20Y%20ESPECIFICIDAD.pdf>.
71. FLORKOWSKI y M, CHRISTOPHER. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. *The Clinical Biochemist Reviews*. 2008. Vol. 29, no. 1, p. 83-87.
72. ŠIMUNDIĆ, ANA MARIA. Measures of diagnostic accuracy : basic definitions [en línea]. Department of Molecular Diagnostics University Department of Chemistry, Sestre milosrdnice University Hospital , Zagreb , Croatia, 2008. Disponible en: www.ifcc.org/ifccfiles/docs/190404200805.pdf.
73. HERNÁNDEZ LEÓN, ROLANDO ALFREDO y COELLO GONZÁLEZ, SAYDA. *El Proceso De Investigación Científica*. La Habana: Editorial Universitaria del Ministerio de Educación Superior, 2011. ISBN 978-959-16- 1307-3.