

00020121208

Algoritmo de reconocimiento de nódulos pulmonares en imágenes computarizadas (TACs)

Víctor Augusto Pedraza Bonilla

PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERIA  
MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN  
BOGOTÁ, D.C.  
2017

00020121208

Algoritmo de reconocimiento de nódulos pulmonares en  
imágenes computarizadas (TACs)

Victor Augusto Pedraza Bonilla

MEMORIA DEL TRABAJO DE GRADO REALIZADO PARA CUMPLIR UNO  
DE LOS REQUISITOS PARA OPTAR AL TÍTULO DE  
MAGÍSTER EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

**Director**

Leonardo Flórez Valencia

**Comité de Evaluación del Trabajo de Grado**

Mariela Josefina Curiel Huérfano

Jaime Andrés Pavlich Mariscal

**Página web del Trabajo de Grado**

<http://pegasus.javeriana.edu.co/~PI1710-2-ARNPIM>

PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERIA  
MAESTRÍA EN INGENIERIA DE SISTEMAS Y COMPUTACIÓN  
BOGOTÁ, D.C.  
Mayo 2017

**PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERIA  
MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

**Rector Magnífico**

Jorge Humberto Peláez, S.J.

**Decano Facultad de Ingeniería**

Ingeniero Jorge Luis Sánchez Téllez

**Director Maestría en Ingeniería de Sistemas y Computación**

Ingeniera Angela Carrillo Ramos

**Director Departamento de Ingeniería de Sistemas**

Ingeniero Efraín Ortiz Pabón

**Artículo 23 de la Resolución No. 1 de Junio de 1946**

*“La Universidad no se hace responsable de los conceptos emitidos por sus alumnos en sus proyectos de grado. Sólo velará porque no se publique nada contrario al dogma y la moral católica y porque no contengan ataques o polémicas puramente personales. Antes bien, que se vean en ellos el anhelo de buscar la verdad y la Justicia”*

## AGRADECIMIENTOS

Primero quiero dar gracias a mis padres, por haberme brindado todo el apoyo en este y cada paso de mi vida, a mis hermanos por ese acompañamiento incondicional, y por último a mi asesor por las horas de trabajo dedicadas a este proyecto

## Contenido

<b>INTRODUCCIÓN.....</b>	<b>14</b>
<b>1. DESCRIPCIÓN GENERAL .....</b>	<b>15</b>
OPORTUNIDAD Y PROBLEMÁTICA.....	15
<b>2. DESCRIPCIÓN DEL PROYECTO .....</b>	<b>17</b>
2.1. OBJETIVO GENERAL .....	17
2.2 OBJETIVOS ESPECÍFICOS .....	17
2.3 FASES DE DESARROLLO .....	17
<b>3. MARCO TEÓRICO / ESTADO DEL ARTE .....</b>	<b>21</b>
<b>4. ALGORITMO PROPUESTO .....</b>	<b>40</b>
<b>5. PROCESO DE EVALUACIÓN.....</b>	<b>46</b>
<b>6. CONCLUSIONES.....</b>	<b>46</b>
<b>REFERENCIAS .....</b>	<b>52</b>

## Índice de Figuras

Figura 1. Fases de la metodología propuesta para el desarrollo del proyecto.....	19
Figura 2. Tendencia de mortandad de los principales cáncer en hombres en el mundo.....	23
Figura 3. Tendencia de mortandad de los principales cáncer en mujeres en el mundo.....	23
Figura 4. Lesión en VE puro de 6 mm.....	27
Figura 5. Nódulo de 12 mm mixto .....	27
Figura 6. Nódulo con componente sólido cercano al 90% de 14 mm: Adenocarcinoma Invasor.....	28
Figura 7. Pasos para detección y clasificación de nódulos.....	30
Figura 8. Metodología de detección mediante extracción de características.....	31
Figura 9. División de la imagen y asignación a un árbol cuaternario.....	34
Figura 10. Definición de sobre ajuste y subajuste .....	39
Figura 11. Ejemplo grafico de clases linealmente no separables.....	40
Figura 12. Macroalgoritmo propuesto.....	43
Figura 13. Figura de resultados con diferentes valores de sigma.....	57

## Índice de Tablas

Tabla 1. Clases de nódulos pulmonares .....	26
Tabla 2. Índice de confiabilidad del coeficiente k.....	51
Tabla 3. Matriz de confusión coeficiente kappa.....	52
Tabla 4. Algoritmos propuestos por otros autores y el Algoritmo Propuesto.....	55



## Índice de Ecuaciones

Ecuación 1. Ecuación de la entropía .....	32
Ecuación 2. Ecuación de la unión de las regiones con más de una semilla.....	35
Ecuación 3. Condición de la fusión de las regiones.....	36
Ecuación 4. Ecuación de selección del umbral.....	36
Ecuación 5. Ecuación de la Esfericidad .....	37
Ecuación 6. Ecuación de la normal entre la ecuación de la recta y el primer punto de cada clase.....	40
Ecuación 7. Condición de la ecuación de la normal.....	40
Ecuación 8. Función de Kernel Lineal.....	41
Ecuación 9. Función de Kernel Polinomial.....	41
Ecuación 10. Función de Kernel Radial Basis Function.....	41
Ecuación 11. Calculo de la varianza y de la media de la intensidad de los pixeles.....	44
Ecuación 12. Ecuación de distribución estadística normal.....	45
Ecuación 13. Ecuación de la prueba estadística de t-student.....	45
Ecuación 14. Ecuación de la esfericidad del nódulo segmentado.....	46
Ecuación 15. Ecuación del área de la esfera.....	46
Ecuación 16. Ecuación del radio de la esfera .....	46
Ecuación 17. Ecuación de la suavidad.....	47
Ecuación 18. Ecuación de desviación estándar de la estadística de la intensidad del nódulo.....	47

---

Ecuación 19. Ecuación de la Curtosis de la Región de interés.....	47
Ecuación 20. Función del kernel Selecciona-do.....	48
Ecuación 21. Función que divide las clases del SVM .....	48
Ecuación 22. Ecuación de sensibilidad.....	49
Ecuación 23. Ecuación de Especificidad .....	49
Ecuación 24. Ecuación de Exactitud .....	50
Ecuación 25. Ecuación de Precisión.....	50
Ecuación 36. Ecuación del coeficiente k.....	50

## **ABSTRACT**

The lung cancer is a disease with a high rate of a death, with 1200000 deaths each year. For this reason, this work wants bring a support to make a better early diagnosis of the disease. This support going to make with a lung nodule classification from Computer Tomography, for the classification will be make with a support vector machine, having as input a shape descripts like the sphericity and volume and texture descripts like smoothness and kurtosis

## **RESUMEN**

El cáncer de pulmón es una enfermedad con una alta tasa de mortalidad con 1200000 muertes cada año [kioskea,2014]. Es por esto que con este trabajo se desea brindar un apoyo y mejorar un diagnóstico temprano a la enfermedad. Este apoyo se realizará por medio de la clasificación de nódulos pulmonares a partir de tomografías computarizadas, esta clasificación se realizará por medio de máquinas de soporte vectorial, teniendo como entradas, descriptores geométricos como la esfericidad y volumen de los nódulos, y un descriptor de textura como la suavidad de los nódulos y la curtosis de la intensidad de los nódulos.

## RESUMEN EJECUTIVO

El cáncer de pulmón fue la enfermedad con más alta tasa de mortandad en el año 2012, con 1200000 muertes en todo el año, este cáncer tiene origen en unas lesiones que se producen en el pulmón llamadas Nódulos pulmonares, estos nódulos se pueden observar a temprana edad de la enfermedad en las tomografías computarizadas. Los nódulos pulmonares pueden variar de tamaño, desde 2mm hasta 10mm, cuando sobrepasa de 3 mm se considera que este nódulo es de tipo maligno, también afecta otras variables en la toma de decisiones de la malignidad del nódulo, como lo es la regularidad de la forma y la textura, es por esto que se busca tener estos parámetros para realizar la clasificación de los nódulos entre malignos o benignos, es decir si con el tiempo generarán cáncer o no generaran cáncer inicialmente en el pulmón y posteriormente en el resto del cuerpo.

Para realizar esta clasificación se diseñó un algoritmo que tiene tres pasos principales que son: la adquisición y lectura de la imagen, posteriormente se realiza una segmentación y por último se realiza la clasificación. Para realizar este algoritmo se hizo uso de la librería Insight Segmentation and Registration Toolkit (ITK) que brinda una amplia gama de herramientas para realizar lectura de imágenes 3D, segmentación y preprocesamiento de imágenes médicas. Para la segmentación de la imagen se hará uso de un algoritmo de crecimiento de regiones llamado ConfidenceConnected, este algoritmo requiere como entrada las coordenadas de donde se encuentra el nódulo y la imagen de la tomografía computarizada. Luego toma como semilla las coordenadas donde se encuentran los nódulos y realiza el proceso de crecimiento de regiones que consiste en tomar un umbral definido por el valor de la estadística de los voxeles que se encuentran en la región de los voxeles que son definidos como posibles nódulos que en este caso será definido como región de interés. Para este proceso de segmentación se realiza varias iteraciones hasta que todos los voxeles sean analizados y hayan sido aceptados o rechazados por el umbral. Al final de cada iteración se modifica el valor del límite del umbral que se va a segmentar. Para la definición del umbral se hizo uso de la distribución de probabilidad T-Student ya que la cantidad de voxeles que abarcan un nódulo no superan el número límite de la población que requiere T-Student que es el valor de la muestra de 30. Una vez se ha obtenido segmentados los nódulos pulmonares se realizará el proceso de clasificación, para este proceso se necesita tener como entrada los valores de los descriptores de características, que en este caso se hará uso de 4 descriptores los cuales son:

**Volumen:** es la cantidad de espacio que ocupa el nódulo, por lo tanto, si el volumen es mayor a  $85\text{mm}^3$  se considera como un nódulo maligno.

**Esfericidad:** es la similitud que tiene el nódulo con una esfera, por lo tanto, si la similitud a una esfera es alta, se considera como un nódulo benigno.

Curtosis: es el grado de concentración que tienen los valores de la intensidad de la región de interés alrededor de una zona central de distribución de frecuencias. Con este método se podrá saber que tan uniforme es la intensidad de cada pixel en la región de interés.

Suavidad: Es la uniformidad de la intensidad de los pixeles en la región de interés, con esto se puede saber que tan liso o corrugado es el nódulo el cual se está analizando.

Luego de tener estos descriptores se insertan en un vector de características para ser insertado en la máquina de soporte vectorial. Primero se realiza el entrenamiento definiendo a que clase pertenece cada Nódulo, para este entrenamiento se tomaron una cantidad de 100 imágenes tomadas de la base de imágenes de LIDC-IDRI Creado por Instituto Nacional de Cáncer en Estados Unidos junto a Imaging Technology Development Branch en el año 2000 con una cantidad de 1025 tomografías computarizadas en formato dicom, que es en un formato definido por el campo medico con un escáner de Tomografías Computarizadas específico al igual que el formato y calidad de las imágenes resultantes, además las Tomografías Computarizadas deben estar previamente etiquetadas por médicos expertos. Luego de este entrenamiento se procede a realizar la clasificación, donde se busca tener como resultado la clase de Nódulo que se encuentra, en este caso se etiqueto el nódulo como de clase 1, 2, 3 o 4 donde las clases 1 y 2 son consideradas como nódulo benigno y la clase 3 y 4 como nódulo maligno. Para la clasificación se realizó la modificación de la constante de sigma que consiste en la tolerancia del error en el momento de realizar la clasificación, el valor de sigma se tomó entre los intervalos de 0.1 y 0.9 donde el valor más optimo fue el valor de 0.4 que se obtuvieron los valores de clasificación más altos.

Para la evaluación de la clasificación se utilizó el coeficiente kappa donde se desea observar el grado de concordancia en la evaluación de un médico experto y el algoritmo propuesto donde se tuvo un valor de 0.80 de concordancia de 1 que es el valor máximo de concordancia. También se tomaron los siguientes estimadores para la evaluación: especificidad con un valor de 0,12, sensibilidad con un valor de 0.86, Exactitud con un valor de 0.89, y precisión con un valor de 0.91.

Como conclusión se pudo observar que este algoritmo toma un tiempo promedio de 10 minutos en tener una respuesta lo cual se podría realizar algún trabajo de reducción de tiempo para un trabajo a futuro.

También se puede realizar para trabajos futuros la clasificación por medio de redes neuronales convolucionales ya que estos métodos han sido utilizados en trabajos de visión computacional y reconocimiento de patrones, teniendo muy buenos resultados en el momento de la clasificación.

El método de segmentación de Confidence Connected con distribución de T-Student, demostró resultado en el momento de la segmentación de los nódulos pulmonares.

## INTRODUCCIÓN

El cáncer pulmonar es una enfermedad que ha generado una alta tasa de mortalidad en el mundo, y que puede llegar a tratarse si se detecta a tiempo. Para este tratamiento se deben detectar los nódulos pulmonares, que son pequeñas lesiones en el pulmón que se pueden observar confiablemente a través de imágenes CT y que se observan como pequeñas manchas en las tomografías que pueden ser benignos (lesiones normales) o malignos (lesiones que presentan cáncer). Es por esto que se desea diseñar y desarrollar un algoritmo que apoye al diagnóstico de esta enfermedad de manera temprana realizando una detección y clasificación de estos nódulos.

Para poder realizar la clasificación es necesario hacer una serie de tratamiento de preprocesamiento de la imagen, para este proyecto se hará preprocesamiento y segmentación por medio del algoritmo de confidence connected, haciendo uso de la distribución de T-Student para la segmentación de los nódulos, requiriendo como entrada la imagen y las coordenadas de las regiones candidatas a segmentar, y teniendo como salida la imagen segmentada.

Estos nódulos se detectarán y clasificarán haciendo uso de las máquinas de soporte vectorial (SVM) teniendo como entrada del SVM diferentes descriptores de forma y textura como la suavidad, la esfericidad, el volumen y la curtosis para obtener tres tipos de nódulos los cuales son: difuso, central, laminar, cartilaginoso; donde los dos primeros nódulos corresponden a nódulos benignos, mientras los dos últimos pertenecen a nódulos malignos.

Para el núcleo del SVM se hará uso del Radial Basis Function Gaussiano, ya que es adaptable para dos o más clases, además de tener un vector de características que se construirá a partir de los descriptores anteriormente mencionados.

Para la evaluación se hará uso de la evaluación por parte de un experto que será usado para comparar el resultado con el algoritmo y se tendrán como criterios de evaluación la sensibilidad, especificidad y precisión, además del uso del coeficiente k.

## 1. DESCRIPCIÓN GENERAL

### Oportunidad y problemática

El cáncer pulmonar es una enfermedad con una tasa de mortalidad alta, que puede llegar a controlarse mediante diferentes tratamientos si las anomalías son detectadas de manera temprana de la enfermedad. Una forma de detectar el cáncer de pulmón en sus primeras etapas, consiste en encontrar nódulos en el pulmón [Santos,2014]. Los nódulos son pequeñas lesiones que se forman en el pulmón y que se observan como manchas ovaladas irregulares en tomografías computarizadas o radiografías de tórax. Los nódulos se forman a causa de enfermedades respiratorias mal tratadas, consumo de cigarrillo, contaminación ambiental, etc. Estos nódulos inicialmente no generan ningún síntoma en la persona, pero con el paso del tiempo van creciendo y generando consecuencias como el cáncer pulmonar. Por lo general, estos nódulos no se pueden detectar a simple vista durante un diagnóstico, debido a que son pequeños y pueden confundirse con otros tejidos en el pulmón. Por lo anterior es pertinente crear herramientas de software para el análisis de las imágenes diagnósticas, mediante las cuales se pueda analizar el pulmón y encontrar estos nódulos para llevar a cabo control y acciones pertinentes [Gae,2014]. Actualmente estos nódulos sólo pueden ser clasificados por los médicos de forma manual, mediante tomografías axiales computarizadas (TAC) ya que al tratarse de una imagen 3D se caracteriza con mayor claridad la textura y forma del nódulo [Ying,2012]. La detección de los nódulos de manera manual necesita un alto nivel de experticia del médico, debido a que en el pulmón se encuentran diferentes tejidos con forma y textura similar.

Este proyecto propone un método para la detección y clasificación de nódulos pulmonares a partir de imágenes de tomografías axiales computarizadas (TAC).

Actualmente existen diferentes algoritmos de clasificación y detección de nódulos pulmonares a partir de imágenes de TAC. Por ejemplo, Tong [Tong,2007] propone un método que analiza diferentes características como el área, forma y centro de masa a través de redes neuronales. El algoritmo de Narayanan [Narayanan,2000] utiliza el método para la detección mediante matriz hessiana, que solamente caracteriza el nódulo por la forma, con un bajo nivel de detalle de los vóxeles para analizar la lesión. El algoritmo de Soltaninejad [Soltaninejad,2012] utiliza un método robusto de máquina de soporte vectorial como el algoritmo de KNN, que solamente analiza el contorno del nódulo. Debido que estos algoritmos usan solamente la caracterización del contorno y forma, además del poco nivel de detalle en los vóxeles usados para analizar la lesión hace que tengan diversas fallas en el momento de realizar la clasificación. Adicionalmente se suman otro tipo de dificultades como los diferentes tipos de textura que se encuentran en los tejidos dentro del pulmón y la extracción del parénquima. Estos inconvenientes anteriormente mencionados generan dificultades que hacen que el promedio de error en el momento de la clasificación de nódulos pulmonares sea del 90 %. Además del promedio de error, existen los falsos positivos que son clasificaciones que se realizan y que no son reales, estas

clasificaciones suceden por la confusión entre nódulos pulmonares y otros tejidos pulmón [Santos,2014], es por esto que se debe tener en cuenta características diferentes a la forma y la morfología. Debido a que existen otras características importantes como la textura que pueden ayudar mucho a realizar una detección y clasificación de los nódulos más confiable, ya que actualmente no existen suficientes trabajos de investigación en el campo de algoritmos de clasificación de nódulos pulmonares [Ying,2012]. Este proyecto propone un método para la clasificación de los nódulos pulmonares que permita reducir el porcentaje de error su clasificación, a partir de imágenes TAC, teniendo en cuenta características del nódulo como la textura y el nivel de detalle de los vóxeles en la lesión para obtener una clasificación. Adicionalmente, se pretende hacer uso de un método robusto como el uso de descriptores y las máquinas de soporte vectorial, empleada para la detección de los nódulos y que ha tenido resultados confiables con un promedio de 85% de imágenes correctamente clasificadas [Berrington,2000]. Para la interfaz gráfica se hará uso de herramientas como Paraview y tree slicer.



## 2. DESCRIPCIÓN DEL PROYECTO

### 2.1. Objetivo general

Desarrollar un algoritmo para la clasificación de nódulos pulmonares a partir de tomografías computarizadas.

### 2.2 Objetivos específicos

1. Caracterizar el contexto clínico de las enfermedades pulmonares.
2. Analizar las precondiciones y pos condiciones del algoritmo de análisis de nódulos pulmonares.
3. Implementar el algoritmo de análisis de nódulos pulmonares a partir de tomografías computarizadas.
4. Probar el algoritmo de análisis de nódulos pulmonares a partir de tomografías computarizadas.

### 2.3 Fases de desarrollo

#### 2.3.1 Caracterización del contexto clínico

Esta es la fase de la etapa de la observación, por lo tanto, se debe tener una revisión de las enfermedades causadas por los nódulos pulmonares y las características de los mismos. Se busca generar diversos comparativos para observar en qué disciplina está ubicado el proyecto, cómo se va a resolver el proyecto y qué aportes se brindarán al campo médico. También se buscarán y analizarán los algoritmos realizados hasta el momento, se observarán ventajas y desventajas de cada uno.

Las actividades a realizar para ejecutar esta fase son:

1. Búsqueda del estado del arte.
2. Realización de comparativos de los trabajos encontrados.
3. Búsqueda de errores comunes en los trabajos encontrados.
4. Conclusiones de cómo se va a realizar el algoritmo.

### 2.3.2 Analizar Precondiciones y pos condiciones del algoritmo

En esta fase se analiza qué se tiene antes y que se obtendrá después de que se haya construido el algoritmo, se formula la hipótesis con el objetivo de observar lo que se desea lograr con el algoritmo que se va a desarrollar, también se analiza cómo se va a realizar, los métodos que serán reutilizados, los métodos que se van a crear y los resultados a obtener al final con el producto.

### 2.3.3 Diseño y Desarrollo del algoritmo

El modelo metodológico que se va a llevar a cabo para esta fase de diseño y desarrollo es el modelo de prototipo rápido, donde se busca realizar el proyecto de la manera más ágil posible, reduciendo al máximo el número de iteraciones de cada ciclo de ejecución de la metodología de tal forma que primero se crea un prototipo del algoritmo que sean fácilmente adaptables al cambio y a reestructuración de diferentes módulos para luego realizar pruebas, aprobación por el usuario y por último pasar a la fase final.

1. Comunicación: en este proceso se realiza la reunión donde se analizarán los avances en el diseño del algoritmo.
2. Plan Rápido – Modelado del diseño rápido: se realizan los pasos a seguir para la construcción del prototipo además de un modelo de construcción de información que refleje el esquema conceptual del prototipo.
3. Construcción del prototipo: Se realiza el desarrollo componente del algoritmo.
4. Entrega Retroalimentación: se realizan pruebas por parte del usuario utilizando las imágenes propuestas (Tomografías Etiquetadas)

### 2.3.4 Evaluar el algoritmo y análisis

Esta fase se realizará la evaluación de los resultados que se tiene luego de la construcción del algoritmo donde se busca cumplir las siguientes actividades:

- A. Ejecución del algoritmo con las 250 tomografías computarizadas analizadas y etiquetadas por un médico experto.
- B. Comparar los resultados obtenidos por el médico experto y los resultados obtenidos por el algoritmo de Clasificación por medio del análisis de coeficiente de capa.
- C. Realizar análisis de confiabilidad y rendimiento del algoritmo.

## 2.4 Metodología del Proyecto

Como metodología se escogió el método científico, ya que es la que más se acopla al proyecto en sus diferentes fases como se puede ver en la siguiente imagen.

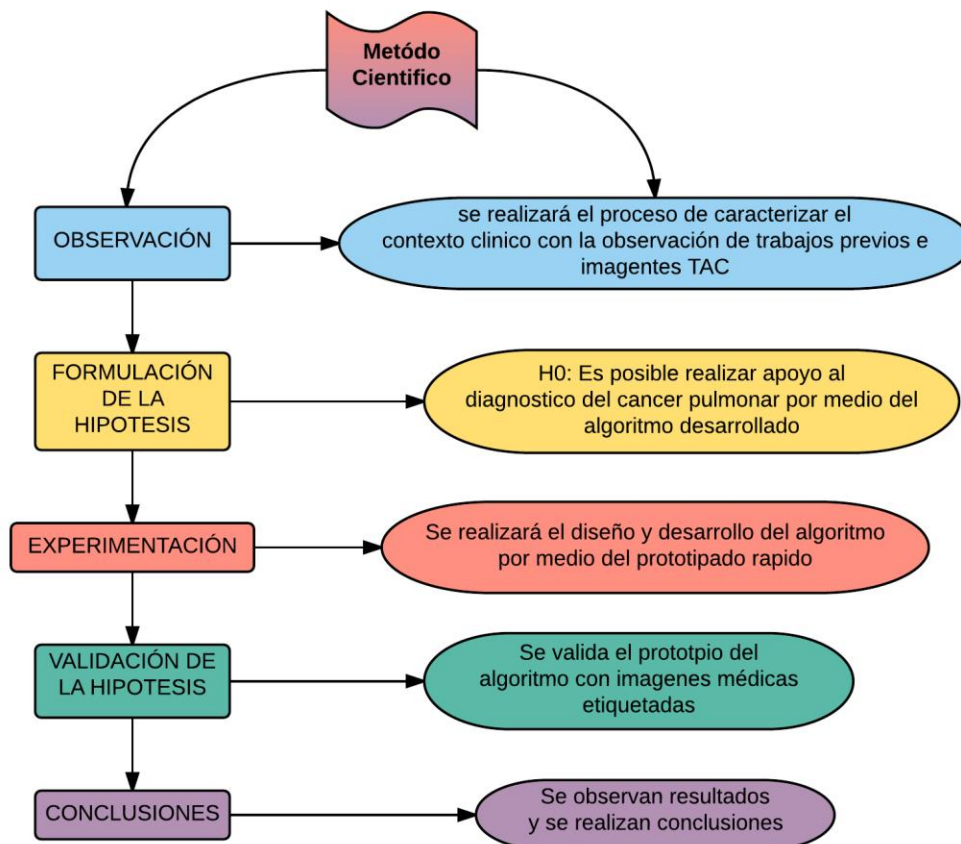


Figura 1. Fases de la metodología propuesta para el desarrollo del proyecto

**Precondiciones del Algoritmo**

Este algoritmo tiene como precondiciones las imágenes computarizadas etiquetadas por expertos, esto se debe a que el algoritmo no realiza la detección de los nódulos pulmonares, por esta razón los nódulos deben ser etiquetados de manera manual

Las imágenes de entrada deben ser Tomografías computarizadas, esto se debe a que en la tomografía computarizada se puede realizar un diagnóstico más claro a temprana edad de la enfermedad, además que se observan con más detalle de los nódulos pulmonares

El algoritmo solo es válido para el cáncer de pulmón ya que en otro cáncer del cuerpo la manera de desarrollarse es diferente a este tipo de cáncer.

**Postcondiciones del Algoritmo**

El algoritmo tiene como resultado la imagen segmentada como archivo externo, junto al resultado de diagnóstico, entre nódulo de tipo maligno y no maligno arrojado desde la consola de ejecución

### 3. MARCO TEÓRICO / ESTADO DEL ARTE

#### 3.1. Contexto Clínico

##### A. Pulmones

Los pulmones son los órganos incluidos en el sistema respiratorio encargados de realizar la función respiratoria, es decir son los que suministran oxígeno a todo el cuerpo por medio de la sangre y expulsa el dióxido de carbono (producto desecho por las células del cuerpo) también participa en otras funciones importantes como son las funciones metabólicas y cardiovasculares.

##### B. Cáncer Pulmonar

El cáncer pulmonar es el conjunto de enfermedades resultantes del crecimiento de células malignas en el sector del pulmón, especialmente en el tejido pulmonar o los bronquios, suele nacer a partir de células epiteliales generando irritación e inflamación crónica del epitelio bronquial, Producto de esta irritación crónica y de factores genéticos se producen mutaciones que llevan al crecimiento rápido y descontrolado de algunas células, lo que se denomina transformación neoplásica, generándose así un cáncer puede derivar en metástasis que consiste en que las células cancerígenas se desplazan a diferentes partes del cuerpo multiplicando células cancerígenas a su paso.

Las causas del cáncer de pulmón suelen ser las siguientes [Berrington,2000]:

- **Tabaco:** Es la principal causa. El humo de tabaco, con su elevada concentración de carcinógenos, va a parar directamente al aire y es inhalado tanto por los fumadores, como por los no fumadores.
- **Exposición a carcinógenos:** El amianto es tal vez la más conocida de las sustancias industriales relacionadas con el cáncer de pulmón, pero hay muchas otras como el uranio, arsénico, ciertos productos derivados del petróleo, etc.
- **Predisposición genética:** Se sabe que el cáncer puede estar causado por mutaciones (cambios) del ADN, que activan oncogenes o provocan que los genes supresores de tumores permanezcan inactivos. Algunas personas heredan mutaciones del ADN de sus padres, lo que aumenta en gran medida el riesgo de desarrollar cáncer.
- **Agentes causantes de cáncer en el trabajo:** las personas con riesgo de desarrollar esta enfermedad son los mineros que tienen posibilidad de inhalar minerales radiactivos, como el uranio, y los trabajadores que manipulan diferentes productos químicos como el arsénico, el cloruro de vinilo, los cromatos de níquel, los productos derivados del carbón, el gas de mostaza y los éteres clorometílicos.

- **Mariguana:** Los cigarrillos de marihuana contienen más alquitrán que los de tabaco. Igualmente, por inhalar el humo profundamente y mantenerlo aumenta el riesgo de contraer esta enfermedad.
- **Inflamación recurrente:** La **tuberculosis** y algunos tipos de neumonía a menudo dejan cicatrices en el pulmón. Estas cicatrices aumentan el riesgo de que la persona desarrolle el tipo de cáncer de pulmón llamado **adenocarcinoma**.
- **Contaminación del aire:** En algunas ciudades, la contaminación del aire puede aumentar ligeramente el riesgo del cáncer de pulmón. Esta posibilidad es mucho menor que la que provoca el hábito tabáquico. [Ramiro,2013]

El cáncer pulmonar es una enfermedad con alto índice de mortalidad, siendo el segundo que más muertes causa en el mundo. En el 2012 la cifra de cáncer reportados fue de 14 millones con una tasa de 1.8 millones de víctimas [Berrington,2000].

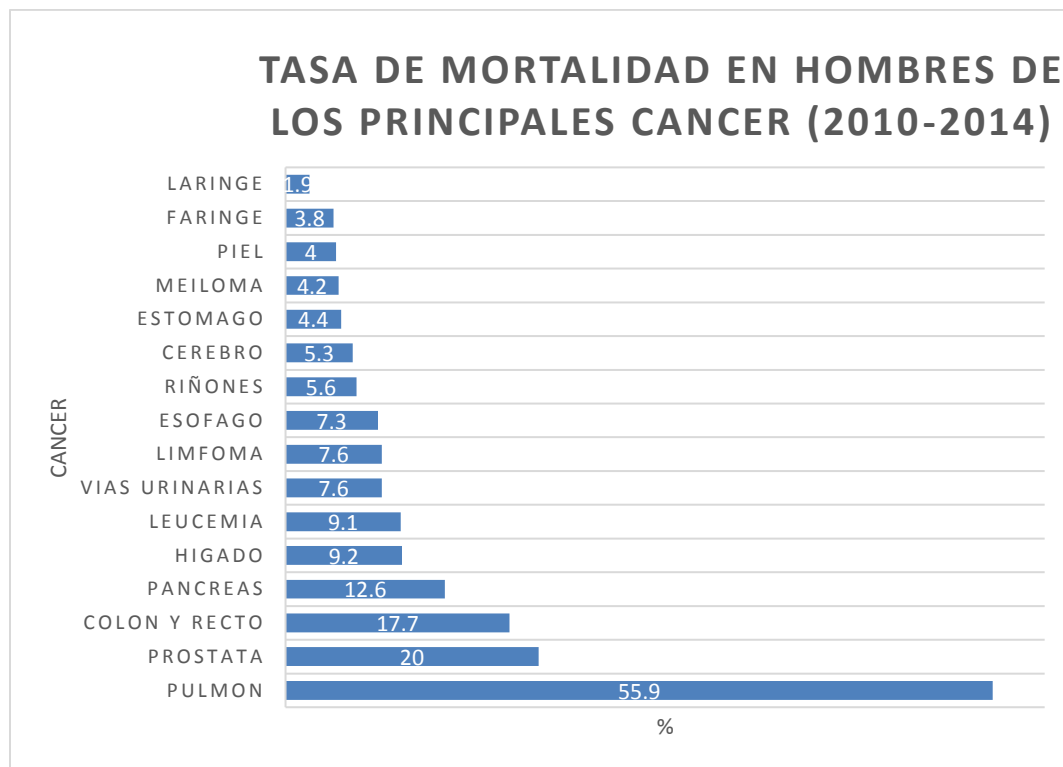


Figura 2. Tendencia de mortandad de los principales cáncer en hombres en el mundo [INC,2013]

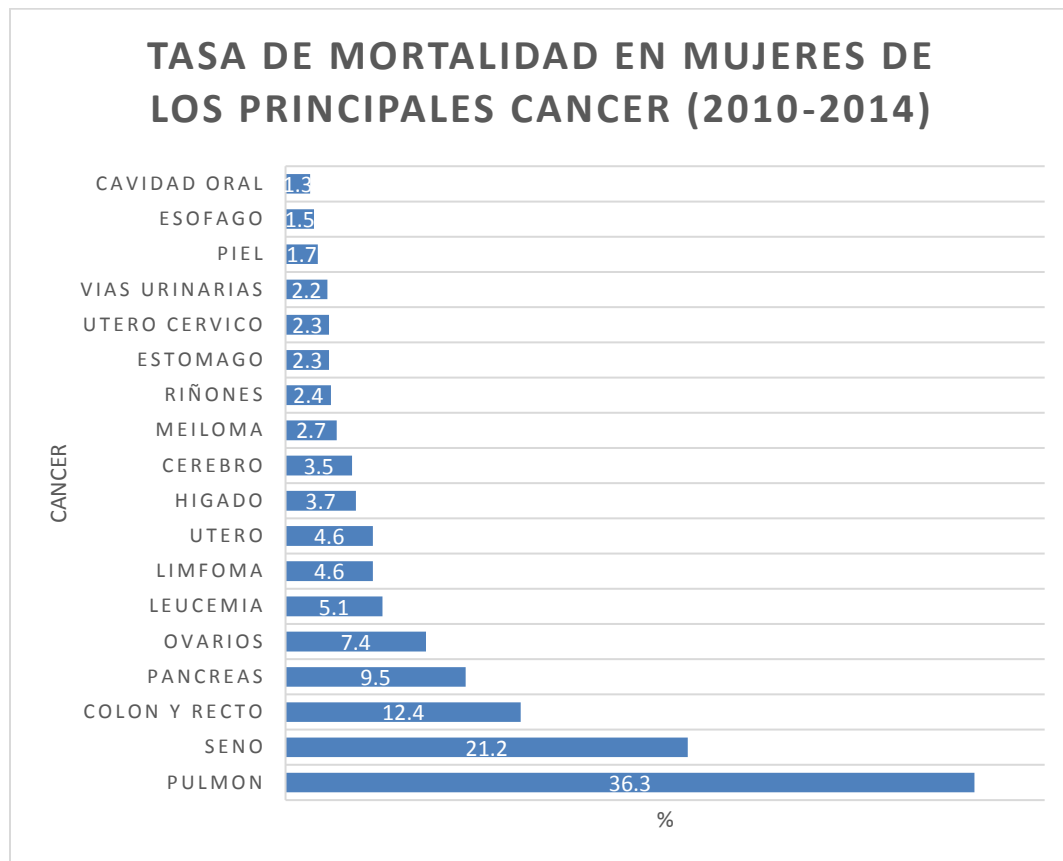


Figura 3. Tendencia de mortandad de los principales cáncer en mujeres en el mundo [INC,2013]

Para el año 2017, los cálculos de la Sociedad Americana Contra El Cáncer para este cáncer en los Estados Unidos son:

- Se diagnosticarán alrededor de 222,500 nuevos casos de cáncer de pulmón (116,990 hombres y 105,510 mujeres).
- Alrededor de 155,870 personas morirán a causa de cáncer de pulmón (84,590 hombres y 71,280 mujeres).

### C. Tipos de Cáncer pulmonar

Según la apariencia de las células al ser examinadas a través del microscopio, los cánceres de pulmón pueden dividirse en dos tipos:

**Cáncer de pulmón de células no pequeñas:** Este tipo se desarrolla en personas fumadores, ex fumadores, fumadores pasivos o personas que han estado expuestas al radón. Los tipos principales de cáncer de pulmón de células no pequeñas reciben su nombre dependiendo del tipo de células encontradas en el cáncer: carcinoma escamocelular (también llamado carcinoma epidermoide), adenocarcinomas, carcinoma de células grandes, carcinoma adenoescamoso y carcinoma no diferenciado.

**Cáncer de pulmón de células pequeñas:** Sólo se desarrolla en fumadores y ex fumadores.

### D. diagnostico

Debido a que los síntomas del cáncer de pulmón a menudo no se manifiestan hasta que la enfermedad está avanzada, solamente un 15 por ciento de los casos se detectan en sus etapas iniciales. Muchos casos de cáncer de pulmón en etapa precoz se diagnostican accidentalmente, como resultado de pruebas médicas que se llevan a cabo por otro problema de salud no relacionado con el cáncer.

### Biopsias

Una biopsia del tejido del pulmón sirve para confirmar o desmentir un posible diagnóstico de cáncer, además de proporcionar información valiosa para determinar el tratamiento adecuado. Si finalmente se detecta un cáncer de pulmón, se realizarán pruebas adicionales para determinar hasta qué punto se ha propagado la enfermedad, entre ellas:

### Estudios radiológicos

Estos estudios utilizan rayos X, campos magnéticos, ondas sonoras o sustancias radiactivas para crear imágenes del interior del cuerpo.

Con frecuencia se utilizan varios estudios radiológicos para detectar el cáncer de pulmón y determinar la parte del cuerpo adonde haya podido propagarse. La radiografía de tórax se puede ver si existe alguna masa o mancha en los pulmones. [Ramiro,2013]

### Tomografía computarizada (TC)

Es el método que da información más precisa acerca del tamaño, la forma y la posición de un tumor, y puede ayudar a detectar ganglios linfáticos aumentados de tamaño que podrían contener un cáncer procedente del pulmón.

Las tomografías computarizadas son más sensibles que las radiografías de tórax ya que detecta los tumores en su etapa inicial (nódulos).



**Exámenes de imágenes por resonancia magnética (RM)**

En este método se utilizan poderosos imanes, ondas radiales y modernos ordenadores para tomar imágenes transversales detalladas. Estas imágenes son similares a las que se producen con la tomografía computarizada, pero son aún más precisas para detectar la propagación del cáncer de pulmón al cerebro o a la médula espinal.

**Biopsia con aguja**

Se introduce una aguja en la masa cancerosa mientras se visualizan los pulmones en un tomógrafo computarizado. Después se extrae una muestra de la masa y se observa en el microscopio para ver si contiene células cancerosas.

**Análisis de sangre**

Con frecuencia el especialista puede realizar ciertos análisis de sangre para ayudar a detectar si el cáncer de pulmón se ha extendido al hígado o a los huesos, así como para diagnosticar ciertos síndromes que aparecen lejos del tumor cancerígeno[Berrington,2000]..

*E. Nódulos Pulmonares*

Son pequeñas agrupaciones de células que se forman en diferentes partes del pulmón y pueden ser visualizadas en las tomografías computarizadas. Los nódulos pulmonares pueden ser benignos (inofensivos) o malignos, es decir que representan a algún tipo de cáncer. Se estima que el 90 por ciento de los nódulos pulmonares que miden menos de dos centímetros de diámetro (0.8 pulgadas) son benignos, y cuando aparece una sola mancha (nódulo pulmonar solitario), se estima que el 60% de todos los nódulos pulmonares solitarios son benignos. [Ramiro,2013].

## F. Tipos de Nódulos pulmonares

Existen las siguientes clases de nódulos pulmonares [Clavero,2013].

<b>Lesiones pre-invasoras</b>	<ul style="list-style-type: none"> <li>• Lesiones menores a 5 mm</li> </ul>
<b>Mínimamente Invasores</b>	<ul style="list-style-type: none"> <li>• Lesión de &gt; a 3mm con componente invasor</li> <li>• De predominio lepidico (tegido liso) con &lt; 5 mm invasión (previamente BAC no mucinoso (no cancerígeno )</li> </ul>
<b>Lesiones Invasores</b>	<ul style="list-style-type: none"> <li>• Otros tipos: textura corrugada, sólido</li> <li>• Lesiones irregulares que inicia en las células que producen moco</li> </ul>
<b>Otras variantes de Lesiones invasores</b>	<ul style="list-style-type: none"> <li>• Grandes tamaños</li> <li>• Solidos e irregulares</li> </ul>

Tabla 1. Clases de nódulos pulmonares

Las características de dichos nódulos se pueden ver en la siguiente imagen

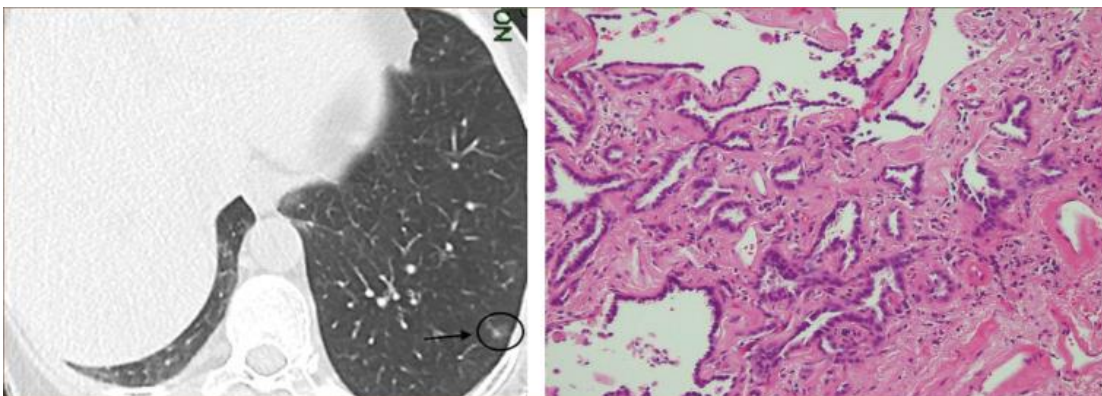


Figura 4. Lesión en VE puro de 6 m

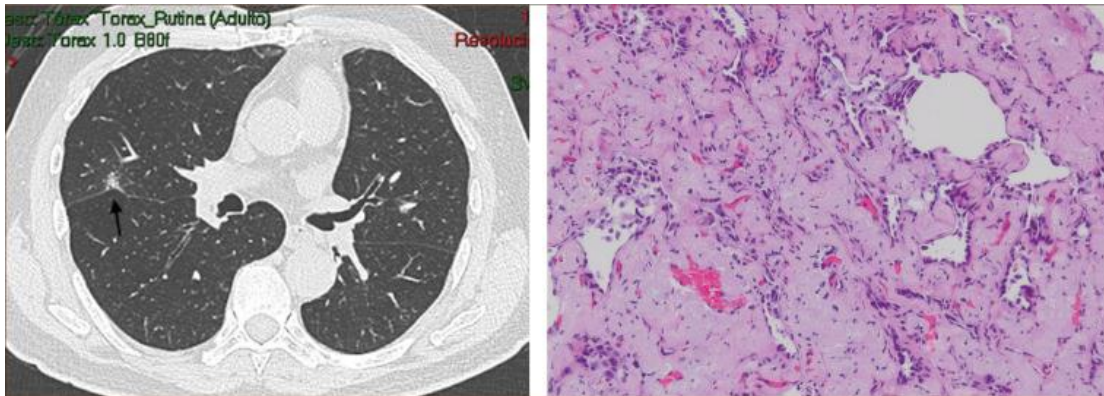


Figura 5. Nódulo de 12 mm

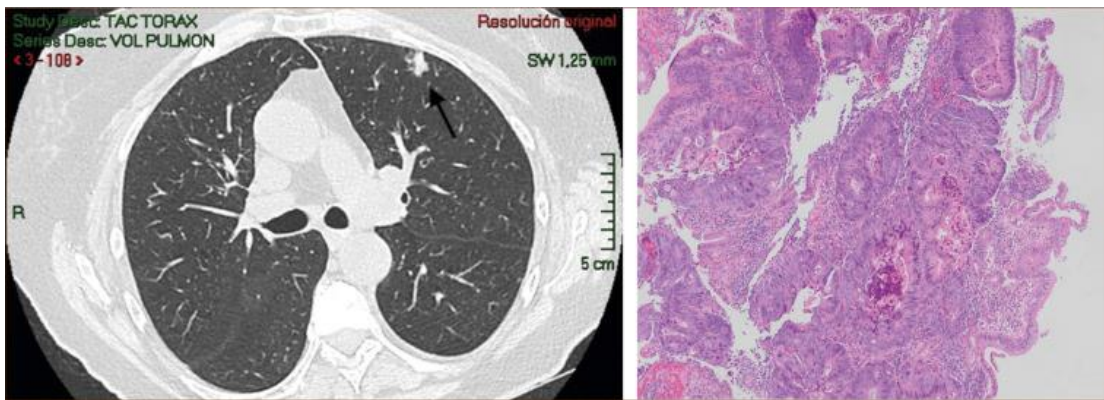


Figura 6. Nódulo con componente sólido cercano al 90% de 14 mm

### 3.2. Contexto grafico

#### A. Pixel y Voxel

Un píxel o pixel, plural píxeles (acrónimo del inglés picture element, elemento de imagen) es la menor unidad homogénea en color que forma parte de una imagen digital, ya sea esta una fotografía, un fotograma de vídeo o un gráfico.

En las imágenes de mapa de bits o en los dispositivos gráficos cada píxel se codifica mediante un conjunto de bits de longitud determinada (la llamada profundidad de color); por ejemplo, puede codificarse un píxel con un byte (8 bits), de manera que cada píxel admite 256 variaciones (28 variaciones con repetición de 2 valores posibles en un bit tomados de 8 en 8) [GAE, 2015].

El vóxel (del inglés volumetric pixel) es la unidad cúbica que compone un objeto tridimensional. Constituye la unidad mínima procesable de una matriz tridimensional y es, por tanto, el equivalente del píxel en un objeto 2D.

Para crear una imagen en tres dimensiones, los vóxeles tienen que sufrir una transformación de opacidad. Esta información da diferentes valores de opacidad a cada vóxel. Esto es importante cuando se han de mostrar detalles interiores de una imagen que quedaría tapada por la capa exterior más opaca de los vóxeles.

Las imágenes con vóxeles se usan generalmente en el campo de la medicina y se aplican, por ejemplo, en la tomografía axial computarizada o para las resonancias magnéticas. De este modo, los profesionales pueden obtener un modelo preciso en tres dimensiones del cuerpo humano [GAE, 2015].

### 3.3. Estado del arte

A partir de diferentes documentos enunciados y analizados anteriormente, una forma de prevenir el cáncer pulmonar es detectando los nódulos pulmonares, que son lesiones que se generan en el pulmón debido a células en mal estado. Estos nódulos se pueden detectar por medio de diferentes tipos de imágenes médicas, como la resonancia magnética, los rayos x y la tomografía axial computarizada (TAC). [INC,2013]. Los TAC tienen ventajas sobre las otras imágenes médicas, ya que proveen una visión más clara para el médico al realizar el diagnóstico y brinda una mayor resolución para poder detectar diferentes y pequeños nódulos que se encuentren en el pulmón. Con un buen diagnóstico realizado a tiempo y con un tratamiento exitoso es posible poder combatir el cáncer ya que el 80 % de los pacientes que se les detectaron a tiempo el cáncer de pulmón en el año 2010 y tuvieron un buen tratamiento superaron con satisfacción dicha enfermedad.

Actualmente existen investigaciones para apoyar el diagnóstico en la detección y clasificación de los nódulos de manera computarizada. A través de algoritmos de procesamiento de imágenes médicas se brinda un apoyo en la detección y clasificación de nódulos pulmonares. Para la realización de estos algoritmos es necesario realizar una serie de pasos y así poder realizar la detección y clasificación. Según Santos [Santos,2014] los pasos son:

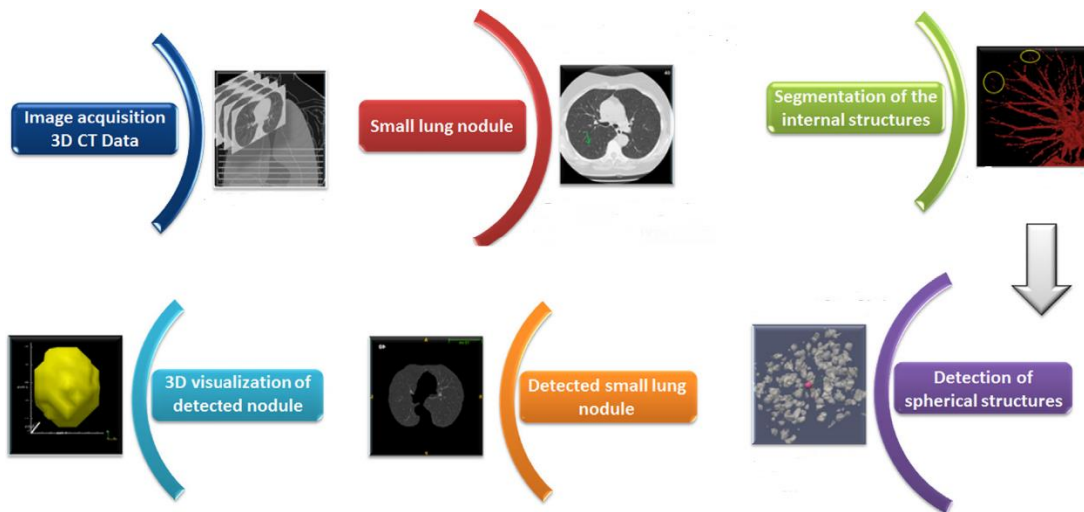


Figura 7. Pasos para detección y clasificación de nódulos. Tomado de [Santos,2014]

La Figura 5 presenta los pasos para la detección y clasificación de nódulos. Se observa que es necesario realizar una adquisición de la imagen, una segmentación de estructuras internas luego la selección y clasificación de los candidatos como nódulos y no nódulos por medio de la comparación de los candidatos con estructuras esféricas.

Por otro lado, Talebpour [Talebpour, 2013] propone la metodología presentada en la gráfica 2, donde se utiliza una extracción de características para la toma de decisiones si la región de interés es un nódulo o no es un nódulo.

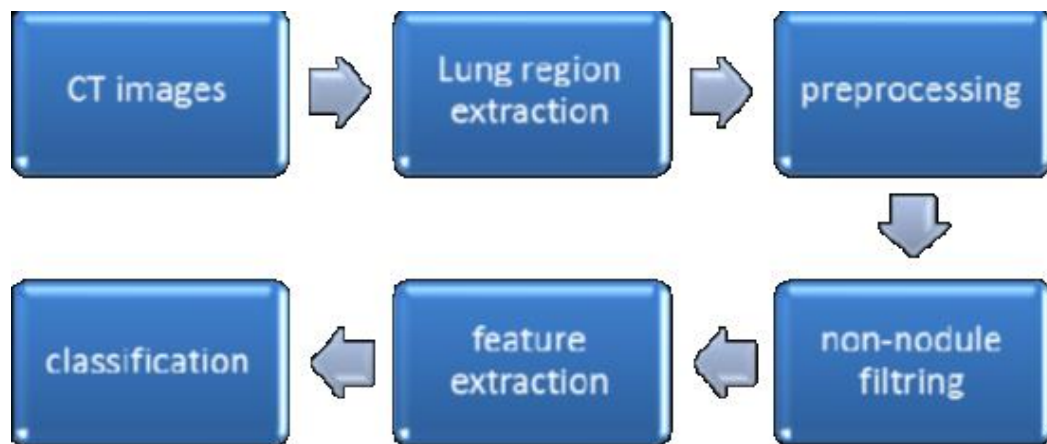


Figura 8. Metodología de detección mediante extracción de características [Ramiro,2013].

Usualmente los proyectos de reconocimiento de patrones utilizan esta metodología para realizar la detección y clasificación de nódulos en los pulmones, donde en la parte del preprocesamiento se debe reducir al máximo el ruido de las imágenes ya que según los autores [Bouma,2013] [Cuifang, 2013] para una buena detección y clasificación es muy importante reducir en lo máximo posible los defectos y el ruido que se encuentra en la imagen para así poder llevar a cabo una buena segmentación.

A continuación, se describen cada fase de la metodología:

#### *Colección de datos*

La colección de datos es el proceso donde se toman todas las imágenes médicas (TAC) para generar una base de datos de múltiples imágenes para realizar el entrenamiento del algoritmo en caso de ser necesario [Gonzales,2003].

#### *Segmentación*

La segmentación es el proceso donde se realiza un enfoque en la región de interés del usuario (en este caso donde es probable que se encuentre el nódulo). La segmentación tiene diversos métodos que son importantes para poder llevar a cabo una buena detección y clasificación, es por esto que se debe realizar de manera minuciosa y con alto nivel de detalle en el momento de análisis de los vóxeles [Lakshmi,2013]. Para dichas segmentaciones se realizan diversos métodos como la conectividad difusa que consiste en tomar una semilla y generar un conjunto de reglas, una vez obtenidas estas reglas se realiza la comparación con otro punto del pulmón y la toma de decisiones si cumple o no con las reglas propuestas. Estos puntos se definen a través de vóxeles, que son píxeles proyectados en 3D, y a partir de la similitud con ese otro punto se realiza la conectividad difusa para la toma de decisiones de la segmentación. El análisis de los puntos se realiza a partir del control de la variación de la intensidad de la imagen

y con estos comparativos se utiliza la conectividad difusa, la cual fue creada con el objetivo de poder realizar operaciones a partir de reglas lingüísticas y así tener resultados más precisos. [Gao,2007].

Otro método que se utiliza para la segmentación son las máquinas de soporte vectorial que son algoritmos basados en el aprendizaje de la máquina y que son usados para la segmentación y análisis de patrones los cuales realizan una segmentación de diferentes tipos como lo son: Ground Glass, árbol de entradas, nódulos encontrados, consolidación de textura y extracción de parénquima ,que consiste en tomar la textura de los fluidos del órgano del que se está realizando la segmentación, esto se realiza por medio de las máquinas de aprendizaje y la comparación de los puntos de la segmentación.[Soltaninejad,2012] El algoritmo de Vessel es usado para detectar bordes por medio de la diferencia en la intensidad de la imagen y de esta forma es usado en algunos casos para detectar la embolia pulmonar, este algoritmo genera una frontera de vessel y de esta manera forma el árbol vascular y permite realizar la detección de la embolia en los pulmones [Lakshmi,2013] . En el algoritmo de Henry Bouma [Ramiro,2013] se observa la importancia del proceso de segmentación. Este autor propone realizar la segmentación por medio de valores propios, haciendo uso de las características de la imagen para mejorar la sensibilidad, en el momento de la clasificación de las embolias pulmonares que al igual que los nódulos son anormalías en el pulmón. También muestra cómo se deben tomar diversas características como: la intensidad de la imagen, contorno, localización y tamaño de la anomalía. Estos factores contribuyeron a generar un avance en el momento de obtener una sensibilización del algoritmo en el periodo de la detección. Aunque el porcentaje de error que tuvo como resultado fue de 22%.

### *La detección*

La detección sirve para encontrar los nódulos en el pulmón. Para la detección existen diversos algoritmos que extraen las características del nódulo, como el algoritmo de Fantacci [Fantacci,2011], que observa diferentes características del nódulo como el volumen y radio, luego a estos dos valores les aplica una varianza y una desviación estándar de todo lo detectado y así logra llegar a tomar una decisión en la detección por medio de redes neuronales basada en vóxeles, este algoritmo presenta una sensibilidad del 80% y un 10% de falsos positivos. Otro algoritmo propuesto es el de Soltaninejad que utiliza el método de máquinas de aprendizaje, es también llamado algoritmo KNN, que es un algoritmo robusto y que no necesita de entrenamiento como las redes neuronales, al no haber entrenamiento se genera un resultado de manera más ágil. Esta propuesta toma como característica la forma del contorno del nódulo, y por medio del análisis de píxeles y vóxeles del algoritmo KNN se realiza una comparación. Por último, por medio de métodos estocásticos realiza una toma de decisiones y muestra los nódulos segmentados. A pesar que presenta una fiabilidad del 90% los autores concluyen que los resultados cambian dependiendo de la cantidad de nódulos y otros tejidos que se encuentren en el pulmón, lo que hace que el algoritmo sea menos fiable en el momento de ser utilizado, también se concluye que manejar las imágenes en 3D ayuda a realizar una detección de los nódulos con resultados más fiables. Tong [Tong,2007] presenta un algoritmo que no solo detecta, sino que también clasifica los nódulos lo cual es una tarea que pocos autores han propuesto, para este algoritmo se tiene en cuenta las siguientes características del nódulo:

- Volumen: lo toma de la suma en el eje x, eje y, y eje z de la cantidad de pixeles en la región de interés.
- Perímetro: lo detecta haciendo una suma de pixeles de todo el borde del área.
- Diámetro equivalente: es el diámetro aproximado del nódulo se toma por la siguiente ecuación. Diámetro equivalente  $D = \sqrt[4]{4 * Volumen/\pi}$ .
- Índice de Irregularidad: es para hallar la morfología del nódulo, es decir la cantidad de irregularidades que tiene en la forma.  $I = 4\pi P^2$ .

Al igual que Narayanan, Alex Santos [Santos,2014] usa descriptores de tamaño para realizar la detección, además de diversos descriptores de textura como lo son la entropía (S) definido por:

$$S = -\sum_i \sum_j p(i, j) \log(p(i, j))$$

Ecuación 1. Ecuación de la entropía

Donde  $p(i, j)$  es la matriz de coocurrencia de los niveles de gris del nódulo.

Finalmente, estas características son ingresadas en una red neuronal para ser entrenadas con el 75% de la base de datos y ser probadas con el 15% de la base de datos. Al igual que el trabajo anterior de redes neuronales este método tiene el riesgo de que la red neuronal sea sobreentrenada y genere conflictos en el momento de tener la confiabilidad, a pesar de lo anteriormente mencionado este algoritmo mostró una confiabilidad del 90% en el momento de realizar la clasificación.

Otro algoritmo de Clasificación es el de Ying Li [Ying,2012] que propone una clasificación utilizando máquinas de soporte vectorial haciendo uso del método Manifold Learning donde realiza un análisis de los pixeles por medio de la distancia euclidiana para observar el contorno y la morfología del nódulo. Este algoritmo tuvo diferentes resultados ya que como conclusión demostraron que el nódulo es ms fácil de detectar y clasificar cuando este se encuentra aislado de otros tejidos en el pulmón; suceso que es poco común, en los nódulos pulmonares. Al final este algoritmo mostró un error mínimo del 27% en el caso de los nódulos aislados, mientras que en el otro caso mostró un error del 35.6%. Mukherjee [Mukherjee,2014] propone un algoritmo de clasificación haciendo uso de filtros gaussianos y de clasificadores bayesianos para realizar la clasificación de los nódulos entre benignos y malignos. Como características toma el centro de masa, varianzas, desviación estándar, intensidad de los pixeles, área, perímetro, tamaño y diámetro efectivo. El porcentaje de error que muestra este algoritmo es del



15% y para la clasificación toma como nódulo maligno a aquellos nódulos que tengan tamaño cercano a los 20mm, este tamaño fue elegido por medio de un método de máximos y mínimos para observar cual era el tamaño promedio y que tamaño tomar para llegar a la conclusión de que sea un tumor maligno o benigno, como se puede observar en la imagen [Bouma,2013]. [Krzysztof,2013]

A partir de estos trabajos encontrados se puede observar las características que se toman en el momento de realizar la detección y en el momento de realizar la clasificación, también se puede observar que actualmente se encuentran pocos algoritmos que se enfoquen en la clasificación, así mismo se puede observar que el porcentaje de confiabilidad de los trabajos no superan el 90%.

### *Segmentación de los Nódulos Pulmonares*

La segmentación es un procedimiento en el que se extrae una región de interés (ROI) de una imagen representada como  $f(x,y)$  y se elimina el fondo y las otras regiones sin interés en la imagen.

Uno de las técnicas de segmentación más usada y que cumple con estas condiciones es el algoritmo de crecimiento de regiones, consiste unir cada pixel a los vecinos o a un grupo de pixeles, para unir cada pixel a un grupo este pixel debe cumplir con una serie de condiciones dadas por medio de distribuciones estadísticas. [Perez,2012]

El crecimiento de regiones sin semillas consiste en dividir toda la imagen en subimágenes y se toma la región que no sea homogénea y se vuelve a dividir, así sucesivamente hasta que todas las subimágenes sean homogéneas y cumplan con las condiciones anteriormente mencionadas, este método se llama partición o división. Usualmente este método cumple la regla que si la imagen tiene una dimensión  $N \times N$  y  $N$  es potencia de 2, es decir,  $N=2^n$ , el procedimiento recursivo se puede representar mediante un árbol cuaternario (cada vértice tiene 4 descendientes) como se puede ver en la siguiente figura [Xujiong,2014].

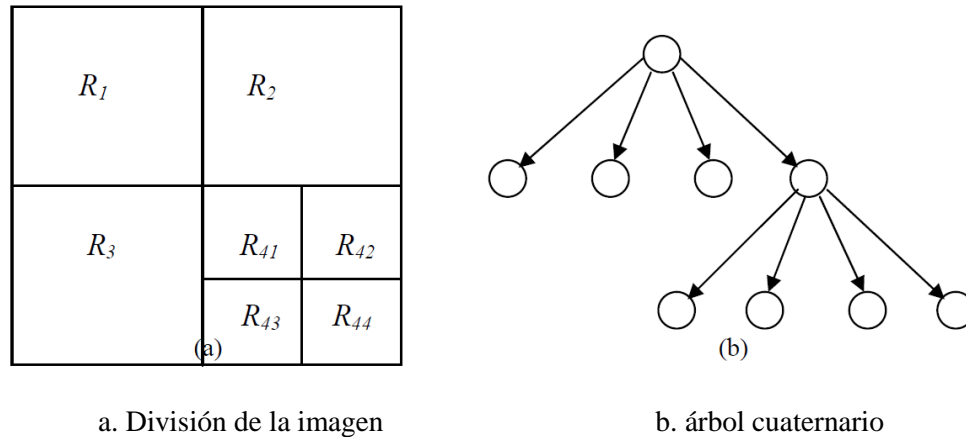


Figura 9. División de la imagen y asignación a un árbol cuaternario.

Para la toma de decisiones de la evaluación de las regiones y cálculo de la distancia entre las diferentes regiones se hace uso de dos heurísticas bidireccionales, este método fue propuesto aml [Amol,2010]:donde una heurística se utiliza para evaluar los parámetros de cada región y establecer si pertenece a una u otra región y a partir de este paso se puede saber el límite de cada región y es guardado en un vector, luego se denota  $W$  como el número de vectores menores del umbral establecido  $T1$  y el perímetro establecido  $P1$ , luego de este paso se crean múltiples regiones que son denotadas como regiones atómicas mientras que la segunda heurística es usada para refinar las decisiones tomadas por la heurística 1. [Zhiwei,2014].

$$|m_i - m_j| < k\sigma_i \quad \text{y} \quad |m_i - m_j| < c\sigma_j$$

Ecuación 2. Unión de las regiones con más de una semilla.

El éxito de una buena segmentación depende de una correcta selección de los umbrales y la elección de más de una semilla. En el caso del crecimiento de regiones se usa la selección del umbral por medio del análisis de las estadísticas entre la región de interés y los vecinos. Es decir, si los valores de la diferencia de la estadística de la región de interés y sus vecinos es menos del umbral entonces se acepta los valores, en caso contrario se rechaza. Existen muchos criterios de evaluación para segmentar las imágenes, por ejemplo, existe la evaluación por iluminación de la imagen, esto significa que las secciones de la imagen que sea más iluminado serán segmentadas y los valores más oscuros serán rechazados [Xujiong,2014].

### Definición de los vectores de la clasificación de los nódulos

Los nódulos son lesiones que tienen bastante similitud con una esfera a comparación con los otros órganos internos del pulmón como los vessel y los arboles bronquiales, es por esto que se toma como vectores de características la esfericidad y la matriz hessiana, ya que nos puede brindar una clasificación de los candidatos que son nódulos y que no son nódulos. [Zhiwei,2014]

**Esfericidad:** es la similitud que tiene la región de interés con una esfera y se representa con la siguiente ecuación.

$$\text{Esfericidad} = 6\sqrt{\pi} V A^{-3/2}$$

Ecuación 5. Ecuación de la Esfericidad.

**Matriz Hessiana:** esta una matriz cuadrada de nxn de segundas derivadas parciales de cada valor, donde se tiene como resultado si la región de interés tiende a una concavidad o a una convexidad, es decir a encontrar que tan uniforme es la forma del nódulo que se está detectando. [Clavero,2013].

**Área:** Es el número de píxeles que se encuentran sobre un eje en el área de interés [Rendon,2016].

**Elipticidad:** es la medida de compresión del diámetro de una esfera para tener una similitud a una elipse [Zhang,2004].

### Máquinas de Soporte Vectorial

Son algoritmos que se basan en la teoría del aprendizaje y fueron creados Vapnik en los años 90 con el objetivo de solucionar problemas de clasificación binaria actualmente se utilizan para resolver otros tipos de problemas (regresión, agrupamiento, multclasificación), también se usa en diferentes campos como la inteligencia artificial, la visión artificial, reconocimiento de caracteres, entre otros. [Krzysztof,2013]

El fundamento básico de las máquinas de soporte vectorial se basa en un hiperplano que se para dos clases, con la condición de que sean separables o cuasi separables (si tienen ruido), la búsqueda de este hiperplano se basa por medio de funciones denominadas kernels que se explicara más adelante.

La diferencia con los otros algoritmos de autoaprendizaje consiste en que mientras los otros algoritmos se basan en tener un buen entrenamiento y reducir los errores (error empírico) las

máquinas de soporte vectorial se enfocan en la minimización del denominado riesgo estructural, definiendo el hiperplano por medio de los valores del límite de cada clase, de esta forma en el momento de realizar el entrenamiento se obtiene una buena generalización. [Taleb-pour,2014].

Durante el entrenamiento se debe mostrar a las máquinas de soporte vectorial elementos de cada clase y definir cada elemento a que clase pertenece.

La buena generalización se obtiene evitando lo siguiente:

- **Sobreajuste:** es cuando los parámetros obtenidos en el entrenamiento se ajustan demasiado a los elementos de cada clase por lo que se pierde generalización y al introducir una nueva muestra no la clasifica bien.
- **Subajuste:** el hiperplano definido es demasiado sencillo y no logra una buena generalización debido a que se han utilizado muy pocas muestras de entrenamiento, por lo que la máquina no generara resultados con muestras nuevas

En la siguiente grafica se podrá observar los diferentes ejemplos de sobreajuste y subajuste

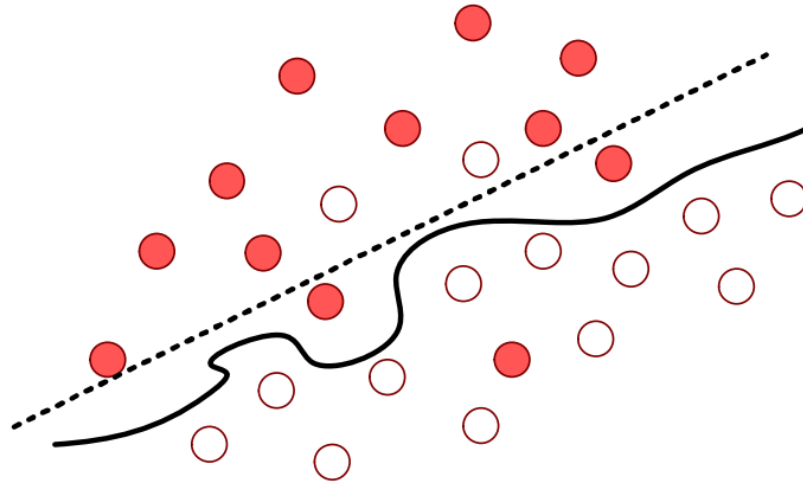


Figura 10: Definición de sobre ajuste y subajuste

Existen diferentes tipos de máquinas de soporte vectorial, pero en este tipo de trabajos se utiliza la máquina de soporte vectorial con diferentes clases, por lo tanto, no son separables linealmente como se puede ver en la siguiente figura [Gao, 2007].

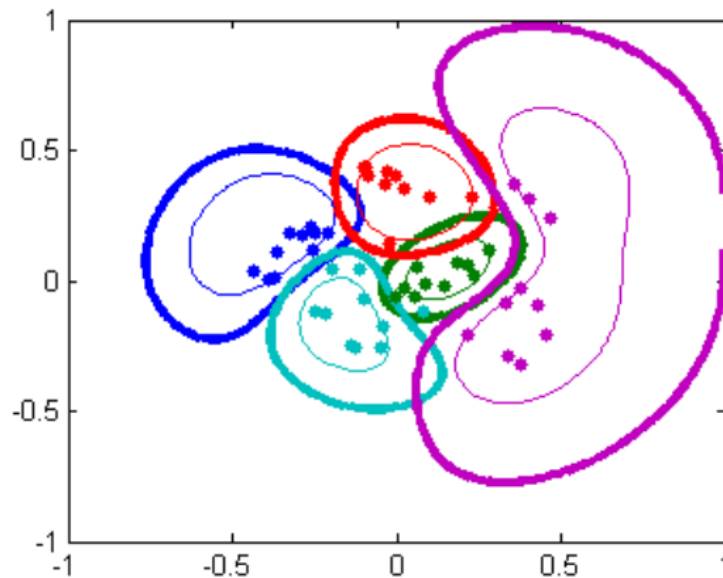


Figura 10. Ejemplo de clases linealmente no separables [Gao, 2007].

Lo ideal es encontrar una función que pueda separar los máximos de clases y que tenga una distancia equidistante entre la función y las clases, para esto se halla la ecuación de la normal.

$$\frac{|D(x')|}{\|w\|}$$

Ecuación 6. Ecuación de la normal entre la ecuación de la recta y el primer punto de cada clase.

Donde  $\| \cdot \|$  pertenece al calor absoluto del elemento de la clase y  $\| \cdot \|$  pertenece al valor unitario de  $w$ ,  $w$  es el vector que define el hiperplano  $D(x')$ , además de tener la propiedad de ser perpendicular al plano, por lo tanto, todos los elementos del entrenamiento deben cumplir la siguiente función: [Gao,2007].

$$\frac{y_i D(x_i)}{\|w\|} \geq T, \quad i=1, \dots, n$$

Ecuación 7. Condición de la ecuación de la normal.

### Definición de los *Kernel*

Un *kernel* es una función simétrica  $k : X \times X \rightarrow \mathbb{R}$  que toma como entrada dos objetos (por ejemplo, vectores, textos o árboles sintácticos) y obtiene como salida un número real caracterizando su similitud [Talebpour,2014].

Los *kernel* fueron usados para manejar las máquinas de soporte vectorial en casos que las clases no fueran linealmente separables transformando el espacio de los elementos de cada clase para poder encontrar un espacio donde los elementos sean separables linealmente.

Lo que se puede observar es que la función de *kernel* nos define una función de similitud entre dos muestras y este es el valor que necesitamos tanto para encontrar la solución del problema como después aplicarlo en la función de clasificación.

Existen tres tipos diferentes de kernel

Lineales:

$$K(x_i, x_j) = x_i^T x_j$$

Ecuación 8. Funcion de Kernel Lineal.

Polinomiales:

$$K(x_i, x_j) = (x_i^T x_j)^d$$

Ecuación 9. Funcion de Kernel Polinomial.

Radial basis function (RBF):

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$$

Ecuación 10. Funcion de Kernel Radial Basis Function.

Para este proyecto se utilizó el kernel *radial basis function* debido a que es utilizado en los casos de múltiples clases y en funciones linealmente no separables. Para el proceso de la clasificación de las múltiples clases realiza una validación cruzada entre todas las clases y realiza la toma de una decisión por medio de la distancia euclidiana entre los resultados esperados[Aberle,2016].

## 4. ALGORITMO PROPUESTO

Este proyecto tiene como objetivo la clasificación, pero se decidió experimentar con un algoritmo de segmentación para obtener los nódulos aislados de las regiones que no son de interés. A partir del estado del arte se propuso un algoritmo de segmentación y clasificación de los nódulos pulmonares haciendo uso de algoritmos de crecimiento de regiones y máquinas de soporte vectorial. Como entrada del algoritmo se requieren imágenes etiquetadas por expertos donde esté definido las coordenadas de la ubicación de los nódulos pulmonares que van a ser segmentados y posteriormente clasificados. Para este algoritmo se utilizan las librerías Insight Segmentation and Registration Toolkit (ITK) que es utilizada para el procesamiento y segmentación de imágenes médicas, también brinda herramientas para tratamiento de imágenes 3D que son las imágenes utilizadas en este proyecto y LibSVM que es la librería diseñada para el trabajo de las máquinas de soporte vectorial (SVM) implementadas en C++, ya que son unas librerías implementadas para este lenguaje de programación, además que este lenguaje es de bajo nivel y por esto ofrece un trabajo más accesible a diferentes características del a programación a comparación de los lenguajes de alto nivel.

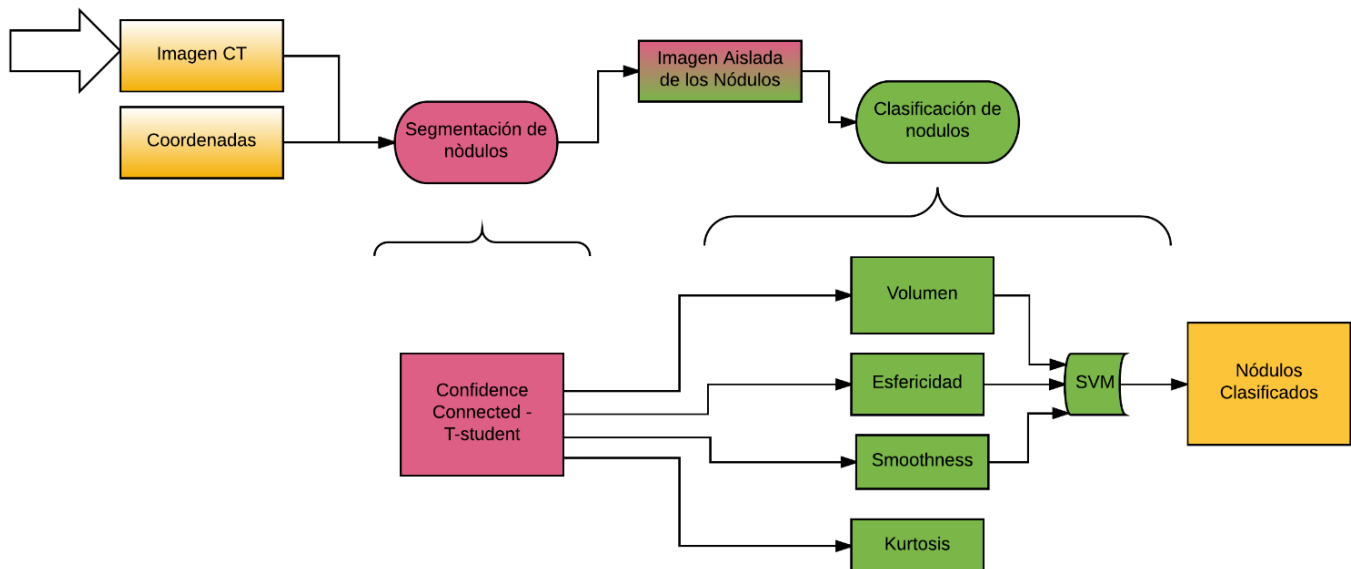


Figura 11. Macroalgoritmo propuesto.



#### 4.1 Segmentación de los Nódulos Pulmonares

A pesar de que el proyecto estaba enfocado en la clasificación de los nódulos pulmonares, fue necesario diseñar un algoritmo de segmentación de los nódulos para aislarlos de las regiones y otras estructuras que no son de interés.

Se eligió el algoritmo de *confidence connected* es basado en el método de crecimiento de regiones donde se realiza la segmentación haciendo uso de las estadísticas de la intensidad de los píxeles que se encuentran en la región. En primer lugar, el algoritmo calcula la media y la desviación estándar de la intensidad de todos los píxeles que se encuentran incluidos en la región. Como entrada este algoritmo requiere las coordenadas del pixel que pertenece a la región a segmentar, a partir de este pixel se crea la región conformada por los vecinos, a esta región se calcula la desviación estándar  $\sigma$  y la media dadas por las siguientes ecuaciones.

$$\sigma = \frac{\sqrt{\sum(x_i - m)^2}}{n-1} \quad \bar{x} = \frac{\sum x_i}{n}$$

Ecuación 11. Calculo de la varianza y de la media de la intensidad de los píxeles.

Donde  $x_i$  es el valor de la intensidad del pixel sub  $i$  y  $n$  es el número de píxeles que se encuentran en la región.

Luego con estas ecuaciones se crea un rango y todos los píxeles vecinos que sus valores de intensidad se encuentren dentro de este rango son aceptados e incluidos en la región, Cuando no se encuentran más píxeles vecinos que satisfagan el criterio, se considera que el algoritmo ha terminado su primera iteración. En ese punto, la media y desviación estándar de los niveles de intensidad se recalculan usando todos los píxeles actualmente incluidos en la región. Esta media y desviación estándar se recalcula dando como resultado un nuevo rango de intensidad que se usa para analizar a los nuevos vecinos de la región actual y evaluar si su intensidad cae dentro del nuevo rango. Este proceso es iterativo y se repite hasta que no se puedan agregar más píxeles o se alcanza el número máximo de iteraciones definido por el usuario. El número de iteraciones se debe especificar en base a la homogeneidad de las estructuras anatómicas las cuales se van a segmentar entre más homogénea es la imagen menos iteraciones van a ser requeridas.

El algoritmo *confidence connected* genera un umbral de segmentación por medio de procesos estadísticos como es del de la distribución normal y todos los voxels que entren en este umbral pertenecen a la sección de segmentación.

$$Z = \frac{x - \mu}{\sigma}$$

Ecuación 12. Ecuación de distribución estadística normal.

Donde  $X$  es la media,  $\mu$  es el valor del pixel,  $\sigma$  es el valor de la desviación estándar.

A pesar de que se utilizó el algoritmo de confidence connected que se utiliza en la segmentación de imágenes de tomografías computarizadas e imágenes en 3D, pero en este caso se necesitaba segmentar regiones pequeñas como lo son los nódulos pulmonares. La solución que se obtuvo para realizar esta segmentación fue realizar una modificación al algoritmo que consistió en trabajar con la distribución t-student a cambio de usar la distribución normal, ya que la distribución t-student utiliza un número de muestras de la población en vez de utilizar el total de la población, esto brinda que la campana de la distribución sea más estrecha al utilizar un número de muestras menor o igual a 30, además de hacer uso de la desviación estándar ( $s$ ) a cambio de la varianza como se puede observar en la siguiente ecuación. De esta forma se ha logrado realizar la segmentación de objetos pequeños como lo son los nódulos pulmonares.

$$t = \frac{x - \mu_0}{\frac{s}{\sqrt{n}}}$$

Ecuación 13. Ecuación de la prueba estadística de t-student.

Al igual que el algoritmo de confidence connected con la distribución normal, este algoritmo requiere como entrada las coordenadas de los nódulos en el eje  $x$ ,  $y$ ,  $z$  y como salida se obtiene la imagen de los nódulos segmentados como se puede observar en la siguiente imagen.

#### 4.2 Clasificación de los Nódulos Pulmonares

Según lo propuesto por Orta [INC,2013], existen 4 tipos de nódulos pulmonares, por este motivo se realizará la clasificación según 4 tipos diferentes de nódulos pulmonares, los cuales dos clases son malignos y dos clases son benignos, es por esto que se van a utilizar 4 clases en la clasificación de los nódulos en este proyecto.

Para esta clasificación se hará uso de 4 descriptores de forma y de textura para los nódulos los cuales son:

**Volumen:** se calcula a partir de la sumatoria de todos los píxeles en el eje  $x$ ,  $y$ ,  $z$  de la región de interés (región del nódulo), se utiliza para saber qué tan grande es el nódulo y a partir de ahí tomar una decisión si los nódulos son malignos o benignos.

Según Talebpour [Cuifang, 2013] los nódulos mayores a 5mm son nódulos que tienen un alto grado de malignidad y riesgo de cáncer.

**Esfericidad:** es la similitud que tiene la región de interés con una esfera y se representa con la siguiente ecuación, es importante este descriptor por que los nódulos que son similares a una esfera tienden a ser benignos, mientras que los nódulos que no son similares a una esfera tienen a ser malignos.

$$\text{Esfericidad: } 6\sqrt{\pi} V A^{-3/2}$$

Ecuación 14. Ecuación de la esfericidad del nódulo segmentado.

Donde el área se obtiene de la ecuación.

$$A = \pi r^2$$

Ecuación 15. Ecuación del área de la esfera.

El radio se obtiene a partir de la ecuación.

$$r = \sqrt[3]{\frac{3v}{4\pi}}$$

Ecuación 16. Ecuación del radio de la esfera.

**Suavidad:** La suavidad es la uniformidad que tiene el valor de la intensidad de los píxeles de la región de interés, es importante para la toma de decisiones.

En la clasificación de los nódulos, se ha observado que los nódulos más uniformes en cuanto a la intensidad son nódulos que tienen tendencia a ser benignos, por este motivo se usa como característica en la entrada de la máquina de soporte vectorial.

$$\text{Suavidad} = 1 - \frac{1}{1 + \sigma^2}$$

Ecuación 17. Ecuación de la suavidad.

Donde  $\sigma$  se obtiene de la ecuación.

$$\sigma^2 = \frac{1}{N} \sum_{i=1, j=1}^N [p(i, j) - \mu]^2$$

Ecuación 18. Ecuación de desviación estándar de la estadística de la intensidad del nódulo.

Donde  $\mu$  es la media de la intensidad del nódulo que se está analizando.

**Kurtosis:** es la medida de concentración que tiene los valores de intensidad de cada pixel de la región de interés, con esto se puede calcular la similitud o disimilitud del valor de intensidad de los pixeles de la región de interés.

Con la kurtosis se obtiene la medida del error de la estadística de la intensidad de los pixeles, por lo tanto, se obtiene un valor que define que tan regular es la imagen en términos de intensidad.

$$\text{Curtosis} = \frac{1}{N} \sum_{i=1, j=1}^N \left[ \frac{p(i, j) - \mu}{\sigma^2} \right]^4$$

Ecuación 19. Ecuación de la Kurtosis de la Región de interés.

Luego estos descriptores serán insertados en el vector de características que será insertado en la máquina de soporte vectorial (SVM) para realizar la el entrenamiento y clasificación. Para el entrenamiento se tomaron una muestra de 100 nódulos previamente etiquetados y tomados de la misma base de imágenes de la tomografía computarizada.

Para este proyecto se utilizó el kernel *radial basis function* debido a que se necesita cuando el resultado tiene más de una clase además que los elementos de cada clase no son linealmente separables, ya que usualmente emplean algún método de determinar un subconjunto de centros. También utiliza la validación cruzada que consiste en validar todos los elementos de la clase con otros los otros elementos de las diferentes clases, de esta forma se puede clasificar los elementos esperando resultados válidos. Intrínsecamente los vectores de características contribuyen a un local de la función de Gauss.

Además [Gonzales, 2016] Menciona las siguientes importancias del kernel RBF

1. Son computacionalmente simples y rápidos de ejecutar.
2. Son fácilmente escalables a datos de alta dimensionalidad, ya que la selección de atributos sólo necesita ser aplicada una vez, para luego evaluar el desempeño de diferentes métodos de clasificación.
3. Estos métodos ignoran las interacciones con el método predictivo, y, por ende, las relaciones entre las distintas variables.

Por los anteriores motivos se eligió este kernel para la clasificación de los nódulos pulmonares

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{(2\sigma)^2}}$$

Ecuación 20. Función del kernel Seleccionado.

A partir del entrenamiento se obtuvo la siguiente función la cual divide las clases

$$\sum_{i=1}^N \alpha_i y_i K(x_i, x_j)$$

Ecuación 21. Función que divide las clases del SVM

Donde  $K(x_i, x_j)$  pertenece al kernel elegido donde se varió el valor de  $\sigma$ , y el valor de  $\alpha$  es dado por la máquina de soporte vectorial que es el grado de modificación de los datos de cada clase para hacerlos separables y,  $y_i$  es el vector de características insertado anteriormente.

Posteriormente se realiza el proceso de clasificación donde se muestran los resultados obtenidos

## 5. PROCESO DE EVALUACIÓN

Para este proyecto se seleccionaron las TC de la base de imágenes LIDC-IDRI desarrollado por el Instituto Nacional de Cancer en Estados Unidos junto a Imaging Technology Development Branch Chief en el año 2000 [Armato, 2006]. Esta base de imágenes propuestas por Armato et al. [Perez, 2012], disponibles en NCI's Cancer Imaging Archive2 y de libre acceso. Esta base de imágenes contiene un total de 1018 CT bajo el estándar DICOM, tomadas por medio de un escáner contemporary spiral CT bajo el método estándar de captura. Luego de la captura las TC fueron etiquetadas por 4 radiólogos expertos. Este proceso de etiquetado se realizó mediante dos fases: la primera fase cada radiólogo de manera independiente etiqueta todas las imágenes tomando todas las lesiones candidatas a nódulos mayor a 3mm de diámetro como nódulo y todas las lesiones candidatas a nódulos menor a 3mm de diámetro como no nódulo Aberle [ITK, 2004]. Luego las evaluaciones realizadas anteriormente por los radiólogos expertos, son evaluadas por los demás radiólogos de manera independiente, las evaluaciones que tuvieron coincidencia tres radiólogos o más radiólogos fueron fusionadas teniendo en cuenta el resultado en común.

En la sección de la segmentación se realizó la validación del algoritmo por medio de comparaciones entre las imágenes segmentadas por el algoritmo e imágenes segmentadas manualmente tomadas de la base de imágenes LIDC-IDRI, por otro lado esta base de imágenes se encontraba etiquetado el tamaño (diámetro) cada nódulo pulmonar y por otro lado en el algoritmo se halló el volumen de cada nódulo, lo que facilitó para hallar el diámetro de cada nódulo y realizar la comparación con los resultados en la base de imágenes y de las 1018 imágenes que tiene la base de imágenes se tomaron 350 imágenes donde los resultados coincidían para poder realizar el proyecto.

Para las pruebas se tomaron un set de 100 imágenes para realizar el entrenamiento de la máquina de soporte vectorial, posteriormente se tomó un set de 250 imágenes donde en el momento de realizar la clasificación se obtienen 4 clases de nódulos etiquetados como clase 1, 2, 3 o 4 donde las clases 1 y 2 son consideradas como nódulo benigno y la clase 3 y 4 como nódulo maligno. Para realizar la validación del algoritmo por medio de las siguientes ecuaciones.

$$\text{Sensitividad} = \frac{TP}{TP+FN} = 0,918918919$$

Ecuación 22. Ecuación de sensitividad.

$$\text{Especificidad} = \frac{TN}{FP+TN} = 0,3125$$

Ecuación 23. Ecuación de Especificidad.

$$\mathbf{Exactitud} = \frac{TP+TN}{TP+FP+FN+TN} = 0,842519685$$

Ecuación 24. Ecuación de Exactitud.

$$\mathbf{Precisión} = \frac{TP}{TP+FP} = 0,902654867$$

Ecuación 25. Ecuación de Precisión.

Donde

**TP:** verdaderos positivos; son las predicciones correctas en el cual la instancia es positiva.

**FN:** falsos negativos: es el número de las predicciones incorrectas cuando la instancia es negativa.

**FP, falsos positivos:** es el número de predicciones incorrectas cuando la instancia es positiva.

**TN, verdaderos negativos:** es el número de predicciones incorrectas cuando la instancia es negativa.

Los resultados obtenidos fueron los siguientes:

tp	199
fn	17
fp	19
tn	15

La evaluación de estos nodulos se realiza por medio del coeficiente kappa, donde existen dos evaluadores, en este caso será el algoritmo y el experto, las respuestas de ambos evaluadores se insertarán de la siguiente forma.

a partir de esta matriz se halla la siguiente ecuación.

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Ecuación 26. Ecuación del coeficiente k.

donde  $Pr(a)$  es la suma de las diagonales dividido entre el total de elementos insertados en la matriz.

$Pr(e)$  es la probabilidad de acuerdo entre los evaluadores.

$Pr(A) * Pr(B) = Pr \text{ caso 1.}$

$Pr(a) * Pr(b) = Pr \text{ caso 2.}$

$Pr(e) = PR \text{ caso 1} + Pr \text{ caso 2.}$

Luego se ha obtenido  $Pr(a)$  y  $Pr(b)$  se procede a hallar el valor de  $k$

Donde  $k$  tiene que tender a 1 para que haya un 100% de efectividad del algoritmo como se puede ver la tabla a continuación.

kappa	grado de acuerdo
< 0	sin acuerdo
0 - 0,2	insignificante
0,2 - 0,4	bajo
0,4 - 0,6	moderado
0,6 - 0,8	bueno
0,8 - 1	muy bueno

Tabla 2. Índice de confiabilidad del coeficiente kappa.

A partir de estos datos se obtuvo la siguiente matriz de confusión.



ARNP/ MED	1	2	3	4
1	50	4	4	2
2	7	44	3	5
3	2	2	50	4
4	2	3	3	55

Tabla 3. Matriz de confusión coeficiente kappa.

Donde se obtuvo un valor de  $k = 0,750855756$ .

Y se obtuvo la siguiente Grafica con los diferentes valores de Sigma entre 0.1 y 0.9 que fueron valores óptimos en trabajos similares y en este proyecto el valor más optimo fue el valor de sigma de 0.4.

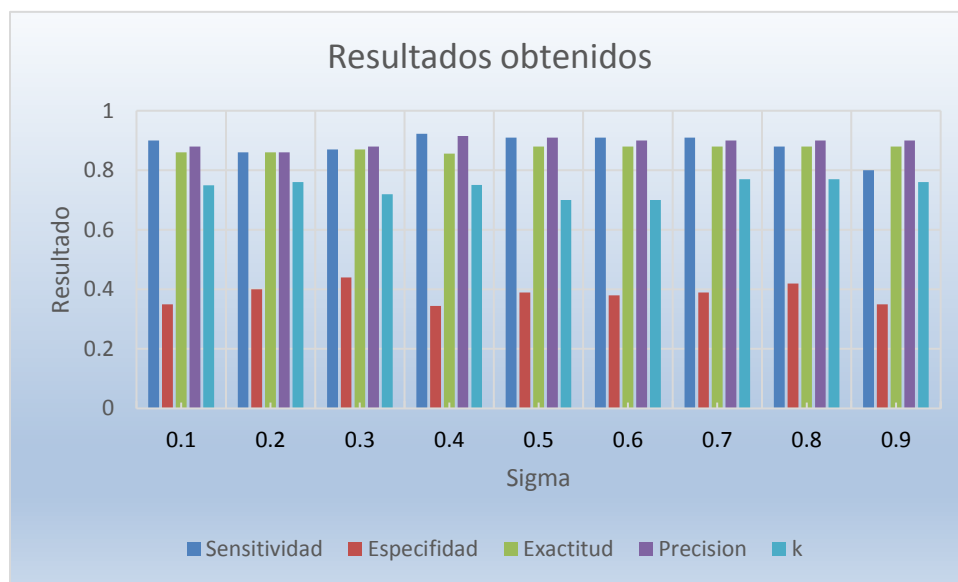


Figura 12. Grafica de resultados con diferentes valores de sigma.

Los anteriores resultados se combinaron por medio de porcentajes de exito donde el más importante es la sensitividad, ya que es el valor que más se encuentra en los trabajos relacionados y se puede realizar una comparación con el trabajo desarrollado.

En la evaluación de complejidad del algoritmo se encontró que el crecimiento de regiones es un algoritmo de complejidad  $O(n)$  o de orden lineal, donde  $n$  es el número de píxeles que tiene la imagen, mientras que en la clasificación el SVM es de complejidad  $O(n^4)$  es decir polinomial, donde  $n$  es el número de elementos que tiene cada clase del SVM.

Con lo anterior se puede notar que lo único que se puede paralelizar es el entrenamiento de la SVM, pero el entrenamiento solo se realiza una vez del tiempo de vida del algoritmo, por lo tanto, no es necesario realizar una paralelización ya que lo complejo del algoritmo es la alta cantidad de píxeles que tiene la tomografía computarizada.

En comparación con otros trabajos propuestos se puede ver como en valores de la sensibilidad se logró el 91% igualando el trabajo de Farahani, pero se realizó un mayor número de pruebas con un mayor número de imágenes. Esta comparación se realizó con la sensibilidad debido a que fue el único valor que se encontraba en todos los trabajos relacionados.

Algoritmo	Detección	Clasificación	Descriptores de forma	Descriptores de textura	Máquinas de aprendizaje	Nº de imágenes	Resultados
[Mukherjee,2014]	No	Si	Si	No	K-NN	20	100%
[Santos,2014]	Si	No	Si	Si	SVM	140	90,60%
[Taleb-pour,2014]	Si	No	Si	Si	RN	144	90%
[Li,2012]	No	Si	Si	No	SVM	210	90%
[Orban,2014]	Si	No	Si	No	SVM	71	81%
[Soltani-nejad,2012]	Si	No	Si	No	K-NN	80	90,30%
[Farahani,2015]	No	Si	Si	No	K-NN	60	91%
[Farahani,2015]	No	Si	Si	No	SVM	60	90%
[Silva,2009]	Si	No	Si	No	SVM	33	96%
[Gonzales, 2016]	No	Si	Si	Si	SVM	70	85%
Algoritmo Propuesto	No	Si	Si	Si	SVM	250	91%

Tabla 4. Algoritmos propuestos por otros autores y el Algoritmo Propuesto

## 6. CONCLUSIONES

Como conclusiones se pudo observar que este algoritmo toma un tiempo promedio de 10 minutos en tener una respuesta lo cual se podría realizar algún trabajo de reducción de tiempo para un trabajo a futuro.

También se puede realizar para trabajos futuros la clasificación por medio de redes neuronales convolucionales ya que estos métodos han sido utilizados en trabajos de visión computacional y reconocimiento de patrones, teniendo resultados confiables (90% de efectividad) en el momento de la clasificación.

El método de segmentación de Confidence Connected con distribución de T-Student, mostro confiabilidad en el momento de la segmentación de los nódulos pulmonares ya que se pudo obtener las 350 imágenes con una correcta segmentación.

## 7. REFERENCIAS

- [kioskea,2014] Kioskea, Cáncer de pulmón y mortalidad., 2014
- [Santos,2014] Santos Alex, Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM, Elsevier, 2014.
- [Gae,2014] Gael Dournes., Computed Tomographic Measurement of Airway Remodeling and Emphysema in Advanced Chronic Obstructive Pulmonary Disease Correlation with Pulmonary Hypertension., American Journal, 2015.
- [Ying,2012] Ying Li, Lung Nodule Classification Using Supervised Manifold Learning Based on All-Class and Pairwise-Class Feature Extraction., IEEE, 2012
- [Tong,2007] JIA Tong Computer-Aided Lung Nodule Detection Based on CT Images., IEEE, 2007
- [Narayanan,2000] Lakshmi Narayanan, A Computer Aided Diagnosis for detection and classification of lung nodules., IEEE, 2015
- [Berrington,2000] A Berrington, 100 years of observation on British radiologists: mortality from cancer and other causes 1897–1997., british intitutr of rhadiology, 2000
- [INC,2013] Instituto nacional de cáncer, Informe Anual a la Nación sobre el Estado del Cáncer, de 1975 a 2010, con una sección especial sobre la prevalencia de comorbilidad y su efecto en la supervivencia de personas con cáncer de pulmón, seno, próstata o colorrectal., Instituto Nacional de Cancer, 2013
- [Ramiro,2013] Orta Ramiro, Nódulos pulmonares subsólidos en tomografía computada multicorte. Características y diagnósticos diferenciales con el adenocarcinoma de pulmón., Hospital Italiano de Buenos Aires, 2013

- [Clavero,2013] Clavero Jose, Nódulos pulmonares.,Clinica las condes 2015
- [Gao,2007] GaoYuanyuan, A New Method for Detection of Pulmonary Nodules., IEEE, 2007
- [Soltaninejad,2012] Soltaninejad Sarah, Mohsen Keshani, Lung Nodule Detection by KNN Classifier and Active Contour Modelling and 3D Visualization., IEEE, 2012
- [Lakshmi,2013] Lakshmi Narayanan, A Computer Aided Diagnosis for detection and classification of lung nodules., IEEE, 2015
- [Bouma,2013] Henri Bouma, Automatic Detection of Pulmonary Embolism in CTA Images., IEEE, 2013.
- [Cuifang, 2013] Li Cuifang, Experimental Investigation of Fuzzy Enhancement for Nonsolid Pulmonary Nodules, IEEE, 2013
- [Krzysztof,2013] Krzysztof Chris Ciesielski, Joint graph cut and relative fuzzy connectedness image segmentation algorithm., Elsevier., 2013.
- [Xujiong,2014] Xujiong Ye, Xinyu LinQian Yiming, Shape-Based Computer-Aided Detection of Lung Nodules in Thoracic CT Images, 2014.
- [Zhiwei,2014] Zhiwei Zhai, Daifeng Shi, Computer-aided Detection of LungNodules with Fuzzy Min-max Neural Network., IEEE, 2014.
- [Jhilar,2014] Jhilar Mukherjee Amlan Chakrabarti, Automatic Detection andClassi\_cation of Solitary Pulmonary Nodules from Lung CT Images., IEEE,2014.
- [Talebpour,2014] Talebpour A RAutomatic Lung Nodules Detection In ComputedTomography Images Using Nodule Filtering AndNeural Networks., IEEE, 2014
- [Perez,2012] M. Perez, Segementación de imágenes., Universidad de Malaga

- [ITK,2004] ITK tutorial., <https://itk.org/ITK/help/tutorials.html>, 2004
- [Cristina,1998] M. Cristina., DETECTOR MULTIUSUARIO PARA DS-CDMA BASADO EN SVM., universidad E.T.S.I, 1998
- [Gonzales,2003] L. Gonzales Modelos de Clasificación basados en Maquinas de Vectores Soporte., Dialnet, 2003
- [Zhang,2004] Zhang Jing,Lung Nodule Classification Combining Rule-based and SVM., IEEE, 2010
- [Rendon,2016] Rendon Elmal, Automatic Lung Nodule Segmentation and Classification in CT Images Based on SVM., IEEE, 2016
- [Aberle,2016] Aberle, D.R, Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine* 365, 395–409. doi:10.1056/NEJMoa1102873. American Cancer Society, 2016. Cancer facts and figures 2016.
- [Mukherjee,2014] Jhila Mukherjee, Automatic Detection and Classification of Solitary Pulmonary Nodules from Lung CT Images., 2014, IEEE
- [Li,2012] Ying Li, Lung Nodule Classification Using Supervised Manifold Learning Based on All-Class and Pairwise-Class Feature Extraction., 2012, IEEE
- [Orban,2014] Gergely Orbán, Lung Nodule Detection On Digital Tomosynthesis Images: A Preliminary Study., 2014, IEEE
- [Farahani,2015] Farzad Vasheghani Farahani, Lung Nodule Diagnosis from CT Images Based on Ensemble Learning., 2015, IEEE
- [Silva,2009] João Rodrigo Ferreira da Silva Sousa, Methodology for automatic detection of lung nodules in computerized tomography images., 2009, Elsevier
- [Gonzales, 2016] Elmar Rendon Gonzalez, Automatic Lung Nodule Segmentation and Classification in CT Images Based on SVM., 2016, IEEE