

# Prototipo para la detección de cáncer de colon a través de análisis de imágenes

*Trabajo terminal No. 2021-A017*

*Alumnos: \*Hernández Escudero Luis Hugo, Tolentino Pérez Jazmin Yaneli.*

*Directores: Dr. Luna Benoso Benjamín, Dr. Silva García Víctor Manuel.*

*e-mail: [hhernandez.isc@gmail.com](mailto:hhernandez.isc@gmail.com)*

**Resumen** – En el presente protocolo de trabajo terminal se propone la realización de un prototipo para la detección de cáncer de colon a través del análisis de imágenes y reconocimiento de patrones en imágenes de cortes histológicos.

**Palabras clave** – Adenocarcinoma, análisis de imágenes, cáncer de colon, detección, reconocimiento de patrones.

## 1. Introducción

El cáncer, es una enfermedad cuyo origen es el crecimiento descontrolado de las células, las cuales modifican su forma y tamaño, entre otras características. El cáncer puede desarrollarse en cualquier parte del cuerpo [1].

Existen diversos tipos de cáncer, estos pueden originarse en los pulmones, en el seno, en el colon e incluso en la sangre. Los diferentes tipos de cáncer tienen algunas similitudes, pero son diferentes en la manera en que crecen y se propagan, además responden diferente a distintos tratamientos [2].

El cáncer de colon y recto es el tumor que se desarrolla en la pared interna del colon o del recto (intestino grueso) originándose inicialmente en la mucosa. La mayoría de estos tumores se originan en pólipos que son levantamientos de la mucosa, estos se denominan pólipos adenomatosos debido a su arquitectura histológica y requieren aproximadamente de 5 a 10 años para transformarse en lesiones tumorales [2].

Los tumores malignos colorrectales tienen origen en alguna de las tres capas: mucosa, muscular y serosa. El adenocarcinoma se origina en las glándulas, es el más frecuente, mientras que los linfomas, los sarcomas y los melanomas son infrecuentes [1].

Este tipo de cáncer predomina en personas mayores. La edad media de presentación es 70-71 años y la mayoría de los pacientes tienen más de 50 años en el momento del diagnóstico, sin embargo, también puede aparecer en personas jóvenes; además de que afecta a hombres y mujeres casi por igual [1].

Finalmente, para llegar al diagnóstico definitivo de cáncer colorrectal es necesaria la confirmación con biopsia que en la mayoría de los casos procede de la colonoscopia. En el estudio microscópico, el adenocarcinoma es el tipo histológico más frecuente (90-95%), seguido del adenocarcinoma coloide (10%) [1].

Existen diferentes artículos que exponen un análisis similar al propuesto:

Artículos	Descripción
<b>Microscopic Image Analysis for Quantitative Measurement and Feature Identification of Normal and Cancerous Colonic Mucosa [3].</b>	<p>Se analizaron 44 imágenes normales y 58 imágenes cancerosas de secciones del colon, las cuales se dividieron en dos subgrupos, un conjunto fue utilizado para el entrenamiento supervisado y el otro para probar el algoritmo de clasificación.</p> <p><b>Resultados:</b> El conjunto de entrenamiento, sensibilidad y especificidad del 93.1% y el 81.8% respectivamente con una precisión global del 88.2%, conjunto de entrenamiento, sensibilidad y especificidad del 93.1% y el 86.4% respectivamente, y una precisión global del 90.2%.</p> <p><b>Método:</b> Se utilizó el método de la matriz de coocurrencia de nivel gris, que consiste en dos pasos principales, el primero calculaba las matrices de coocurrencia intermedias y el segundo utiliza las matrices para deducir las características.</p>
<b>Texture Descriptions and Classification for Pathological Analysis of Cancerous Colonic Mucosa [4].</b>	<p>A partir del análisis de textura con parámetros de dimensión fractal, entropía y correlación, sobre muestras imágenes patológicas del colon humano, de las cuales 44 corresponden a un colon sano y 58 a un colon con cáncer y empleando clasificación no paramétrica.</p> <p><b>Resultados:</b> Se obtuvo una efectividad de 88% que se mejoró a 94% al emplear dimensión fractal en comparación de solo usar entropía y correlación.</p>
<b>Colon Cancer Diagnosis on Digital Tissue Images [5].</b>	<p>Desarrollo de un software que pueda ser una ayuda para el diagnóstico del cáncer de colon, utilizando este sistema se pueda ayudar a los pacientes con un diagnóstico temprano.</p> <p><b>Resultados:</b> La fracción de glándulas no detectadas es del 6.2% y la no detectada es del 18.8%.</p> <p><b>Método:</b> Código de estructura de color, el cual realiza una segmentación basada en el color y el resultado es utilizado para el siguiente proceso.</p>

Tabla 1. Metodologías y resultados obtenidos en trabajos similares

## 2. Objetivos

### Generales

Desarrollar un prototipo que puede ser de ayuda en la detección de casos de cáncer de colon a través del análisis de imágenes y reconocimiento de patrones en imágenes de cortes histológicos de donde se obtendrá el prediagnóstico [6].

### Particulares

- Seleccionar y emplear una técnica de segmentación, que nos permita extraer las características necesarias para la posterior clasificación de las muestras.
- Emplear un modelo de clasificación que nos ofrezca un nivel de precisión aceptable en un rango que pueda competir con los trabajos similares.
- Una aplicación web/escritorio que permita a los usuarios del prototipo cargar una imagen de un corte histológico del colon y obtener un prediagnóstico.

### 3. Justificación

Aproximadamente, el 70% de las muertes por cáncer suceden en países de ingresos medianos y bajos. En México, tan solo en el 2012 se registraron 14 millones de casos nuevos y 8.2 millones de personas fallecidas. A nivel mundial el cáncer colorrectal es el tercer tipo de cáncer más común en pacientes y es el segundo tipo de cáncer que más muertes causa. Afortunadamente, muchos cánceres (seno, colorrectal, cervical, etc.) tienen altas probabilidades de curarse si se detectan a tiempo, adicionalmente algunos de ellos pueden prevenirse evitando los factores de riesgo [8, 9].

En la batalla contra el cáncer, el tiempo suele jugar un papel muy importante, pues la detección oportuna puede ser la diferencia entre, poder y no poder hacer algo, con ayuda de este prototipo los pacientes podrían tener acceso a un diagnóstico más oportuno a menor costo en cuanto a tiempo y esfuerzo por parte de los centros de salud.

En la siguiente tabla se muestra la probabilidad de sobrevivir dependiendo de la etapa del cáncer colorrectal.

Clasificación TNM <sup>1</sup>	Clasificación de Duke <sup>2</sup>	Sobrevivir (%)
Etapa 0-Carcinoma in situ		
Etapa I-Sin afección ganglionar, sin metástasis, el tumor invade la capa muscular de la mucosa hasta la submucosa (T1, N0, M0); también es posible que haya crecido hasta la muscular propia (T2, N0, M0)	A	90-100
Etapa II-Sin afección ganglionar, sin metástasis, el tumor invade la subserosa (T3, N0, M0); el tumor invade otros órganos (T4, N0, M0)	B	75-85
Etapa III-Ganglios linfáticos regionales involucrados (cualquier T, N1, M0)	C	30-40
Etapa IV- Metástasis distante	D	< 5

Tabla 2. Estado y probabilidad de supervivencia al cáncer colorrectal [7].

El proceso manual de detección del cáncer colorrectal se realiza a través del análisis patológico de cortes histológicos del tejido del colon bajo un microscopio, este proceso requiere mucho tiempo y los patólogos suelen evaluar grandes cantidades de muestras al día, generalmente las técnicas de evaluación son subjetivas, algunas de las medidas cuantitativas, como la etapa del cáncer, están sujetas principalmente a la evaluación visual de los patólogos, estas deficiencias en el procedimiento hacen notoria la necesidad de automatizar el proceso de detección.

Es por ello, que en este trabajo se presenta la propuesta de una metodología que permita la detección de casos de cáncer de colon mediante el análisis de imágenes de cortes histológicos haciendo uso de reconocimiento de patrones, para realizar una clasificación entre un adenocarcinoma o un lipoma de colon.

Los beneficios que podría ofrecer a los patólogos y a los centros de salud de la misma manera son muy importantes, en primer lugar el tiempo, un clasificador automático podría clasificar una muestra en segundos, adicionalmente si el prototipo se despliega como un servicio web estaría alcance de cualquier centro de salud durante cualquier hora del día, no es el objetivo del prototipo sustituir a los patólogos en el trabajo de detección, se sabe que algunos casos son difíciles de diagnosticar, se pueden llegar a presentar distintos diagnósticos de diferentes especialistas respecto a la

<sup>1</sup> Sistema de estadificación TNM, es un recurso para que los médicos puedan determinar la etapa del cáncer según ciertas normas, en este sistema a cada cáncer se le asigna una letra o número para describir el tumor, los ganglios y la metástasis. La **T** se refiere al tumor original (primario), la **N** se refiere a los ganglios linfáticos; indica si el cáncer se ha propagado a los ganglios linfáticos cercanos. La **M** se refiere a la metástasis; indica si el cáncer se ha propagado a partes distantes del cuerpo.

<sup>2</sup> Sistema de clasificación usado para describir la extensión del cáncer colorrectal. Los estadios van de A (estadio temprano) a D (estadio avanzado).

misma muestra, es donde nuestro prototipo podría intervenir, como una ayuda para los especialistas en los casos difíciles.

#### 4. Productos o resultados esperados

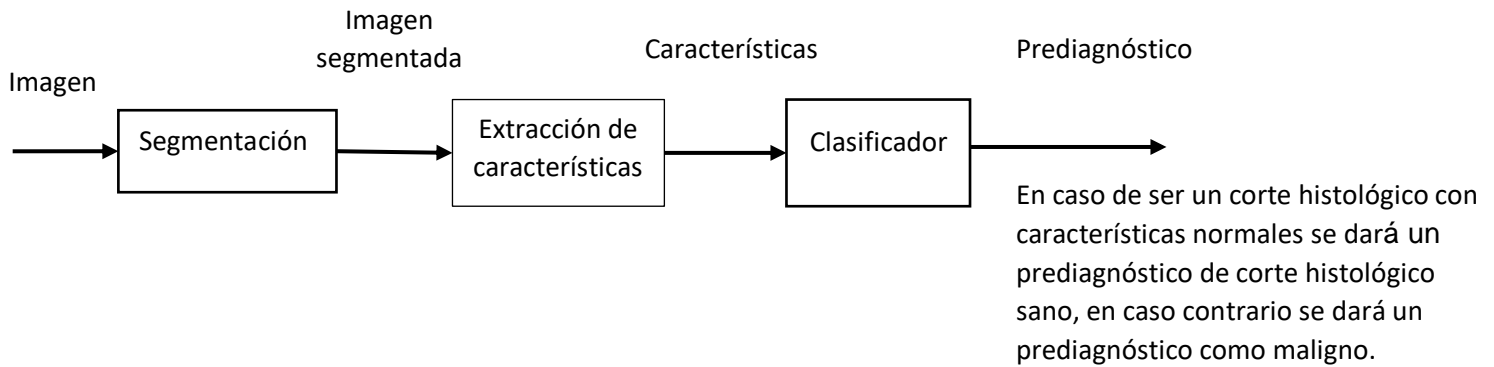


Figura 1. Diagrama de caja negra del prototipo planteado.

El prototipo que planteamos realizar tiene tres módulos principales, el primero de ellos para la segmentación de la imagen, el segundo de ellos para la extracción de las características de la imagen y el último de ellos es un clasificador binario, las categorías en las que una imagen puede clasificarse son: normal o maligno.

El primero de ellos, tiene como entrada una imagen y como salida un vector con las características (correspondientes a la imagen). Dentro de este componente se realizarán todas las tareas necesarias de procesamiento digital de imágenes que nos faciliten el objetivo principal de este componente, segmentar la imagen.

El siguiente componente es un clasificador binario que tiene como entrada el vector de características de la imagen y como salida el prediagnóstico para la imagen analizada, este diagnóstico tiene como posibles resultados “normal” (para un corte histológico de colon normal) y “maligno” (para un corte histológico de colon anormal).

Manual de usuario, que tiene como propósito servir de guía a los usuarios que pretendan usar el prototipo, aquí se especificarán los requisitos que deben cumplir las imágenes a ser clasificadas, los pasos que se deben seguir al interactuar con la interfaz, entre otros detalles técnicos como los tiempos de espera y el porcentaje de eficacia.

#### 5. Metodología

Se propone hacer uso del modelo incremental, debido a que es un proceso del desarrollo de software que permite que los requisitos se dividan en diferentes iteraciones independientes del ciclo del desarrollo de software.

El modelo incremental permite que pueda adaptarse a cualquier tipo de proyecto, y cuenta con al menos 7 fases que se describen a continuación:

1. Requerimientos: Definición de objetivos generales y específicos.
2. Definición de las tareas y las iteraciones: Una vez que se ha definido lo que se busca, la siguiente fase es crear una lista de tareas y agruparlas en las iteraciones que conformaran al proyecto.
3. Diseño de los incrementos: Es necesario definir la evolución de los incrementos, tomando en cuenta que cada iteración debe de superar a la anterior.

4. Validación de incrementos: Buscar la aprobación de los incrementos o en su defecto buscar las causas de que el incremento no fuera aprobado.
5. Integración de incrementos: Cuando son validados los incrementos, dan forma a lo que se denomina evolución del proyecto, donde cada incremento contribuye al resultado final.
6. Entrega del producto: Una vez que el producto completo ha sido validado en su totalidad y corresponde con los objetivos iniciales, se procede a la entrega final.

Aunque el modelo incremental divida las tarea en iteraciones, (pequeños lapsos en los cuales se trabaja para conseguir los objetivos específicos), no se trata de iteraciones independientes, estas están vinculadas para que cada una suponga un avance con respecto a la iteración anterior y de esta forma se obtenga el producto completo al final de las iteraciones [10].

## 6. Cronogramas

CRONOGRAMA Nombre del alumno: Hernández Escudero Luis Hugo

TT No.: 2021-A017

Título del TT: Prototipo para la detección de cáncer de colon a través de análisis de imágenes.

Actividad	AGO	SEP	OCT	NOV	DIC	ENE	FEB	MAR	ABR	MAY	JUN
Definición de la técnica de segmentación.											
Definición del modelo de clasificación.											
Diseño y planeación de los incrementos.											
Evaluación de TT I.											
Implementación de los incrementos.											
Validación y pruebas.											
Desarrollo de aplicación web/escritorio.											
Evaluación de TT II.											

CRONOGRAMA Nombre del alumno: Tolentino Pérez Jazmin Yaneli

TT No.: 2021-A017

Título del TT: Prototipo para la detección de cáncer de colon a través de análisis de imágenes.

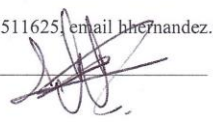
Actividad	AGO	SEP	OCT	NOV	DIC	ENE	FEB	MAR	ABR	MAY	JUN
Definición de Requerimientos.											
Definición de tareas e iteraciones.											
Diseño y planeación de los incrementos.											
Evaluación de TT I.											
Implementación de los incrementos.											

Validación y pruebas.												
Integración de los incrementos.												
Manual de usuario.												
Evaluación de TT II.												

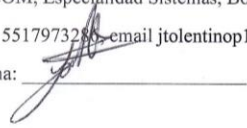
## 7. Referencias

- [1] Gonzalez F. E., “Cáncer de colon y recto,” [En línea]. Disponible: <https://seom.org/info-sobre-el-cancer/colon-recto?showall=1>. [Último acceso: 09 03 2021].
- [2] I. N. d. C. ESE, “Manual para la detección temprana del cáncer de colon y recto,” 2015. [En línea]. Disponible: <https://www.cancer.gov.co/files/libros/archivos/Colon>. [Último acceso: 08 03 2021].
- [3] Nasser E. A. , Naguib N.G. R, “Microscopic Image Analysis for Quantitative Measurement and Feature Identification of Normal and Cancerous Colonic Mucosa,” IEEE Transactions on information technology in biomedicine, VOL. 2, NO. 3 Septiembre 1998,
- [4] Nasser E. A. , Naguib N.G. R. , Sharif B. S. , Benett M. K. , and Murray A. “Texture Descriptions and Classification for Pathological Analysis of Cancerous Colonic Mucosa,” Image Processing and its Applications, Conference Publication No. 465 IEE 1999
- [5] Zoltán K., Zoltán T., Sandor S., Zoltán V., Szabolcs S., “Colon Cancer Diagnosis on Digital Tissue Images, “IEEE 9th International Conference on Computational Cybernetics Julio 8-10, 2013
- [6] Borkowski AA, Bui MM, Thomas LB, Wilson CP, DeLand LA, Mastorides SM. Lung and Colon Cancer Histopathological Image Dataset (LC25000). arXiv:1912.12142v1 [eess.IV], 2019
- [7] A. Young, R. Hobbs, and D. Kerr, eds., ABC of Colorectal Cancer, second ed. Wiley-Blackwell, 2011
- [8] WHO, “World Health Organization: Cancer,” 2018. [En línea].  
Disponible: <http://www.who.int/en/news-room/fact-sheets/detail/cancer>.
- [9] INCAN, “Programa Institucional 2020-2024 del Instituto Nacional de Cancerología.” 2020. [En línea].  
Disponible: [http://dof.gob.mx/nota\\_detalle.php?codigo=5605754&fecha=25/11/2020](http://dof.gob.mx/nota_detalle.php?codigo=5605754&fecha=25/11/2020).
- [10]JPérez A., “Características y fases del modelo incremental,” 2016. [En línea]. Disponible:  
<https://www.obsbusiness.school/blog/caracteristicas-y-fases-del-modelo-incremental> [último acceso: 17 03 2021].

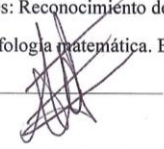
## 8. Alumnos y Directores

*Luis Hugo Hernández Escudero.* – Alumno  
de la carrera de Ing. En Sistemas Computacionales en  
ESCOM, Especialidad Sistemas, Boleta: 2016310335  
Tel. 5585511625, email [lhernandez.isc@gmail.com](mailto:lhernandez.isc@gmail.com)  
Firma: 

CARACTER: Confidencial  
FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos  
108, 113 y 117 de la Ley Federal de Transparencia y Acceso  
a la Información Pública.  
PARTES CONFIDENCIALES: Número de boleta y teléfono.

*Jazmin Yaneli Tolentino Pérez.* – Alumno  
de la carrera de Ing. En Sistemas Computacionales en  
ESCOM, Especialidad Sistemas, Boleta: 2015121247  
Tel. 5517973288, email [jtolentinop1400@alumno.ipn.mx](mailto:jtolentinop1400@alumno.ipn.mx)  
Firma: 

*Dr. Benjamín Luna Benoso.* – Licenciatura en Matemáticas  
por la ESFM del IPN. Maestría y doctorado en ciencias de  
la computación del CIC-IPN. Actualmente profesor de tiempo  
completo de la ESCOM.

Áreas de interés: Reconocimiento de patrones, Análisis de  
imágenes, morfología matemática. Email: [blunab@ipn.mx](mailto:blunab@ipn.mx)  
Firma: 

*Dr. Victor Manuel Silva García.* – Licenciatura en Matemáticas  
por la ESFM del IPN, Doctorado por el CIC-IPN. Actualmente  
profesor del CIDETEC-IPN.

Áreas de interés: Análisis de imágenes, criptografía. Email: [vsilvag@ipn.mx](mailto:vsilvag@ipn.mx)  
Firma: 