

**FACULTAD DE INGENIERÍA
UNIDAD DE POSGRADOS**

PRESENTACIÓN PROPUESTA

(Utilice o amplíe los espacios sugeridos según convenga – Tamaño Fuente: 12, Interlineado: 1,5)

TESIS DE MAESTRÍA: ☒ TRABAJO FINAL DE ESPECIALIZACIÓN: ☐

1. PROPONENTE: EDUARD LEONARDO SIERRA BALLÉN
2. CÓDIGO: 299614
3. PROGRAMA: MAESTRÍA EN INGENIERÍA - INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
4. DIRECTOR PROPUESTO: EDUARDO ROMERO Ph.D.
DEPARTAMENTO: FACULTAD DE MEDICINA – CENTRO DE TELEMEDICINA
TÍTULO: MORFOMETRÍA EN IMÁGENES DE CERVIX PARA LA DETERMINACIÓN DE LESIONES ESCAMOSAS
ÁREA: CIENCIAS DE LA COMPUTACIÓN
LÍNEA DE INVESTIGACIÓN: PROCESAMIENTO DE IMÁGENES
5. ANTECEDENTES Y JUSTIFICACIÓN: (Indicar los desarrollos previos, circunstancias y condiciones que llevaron a la conclusión de la necesidad y conveniencia del proyecto)

El cáncer constituye un problema de salud mayor en todo el mundo, se presenta cómo una división incontrolada de las células. Este crecimiento desordenado de células, denominadas células malignas, generalmente es tratable cuando se hace diagnóstico temprano. El uso de herramientas informáticas podría mejorar la calidad de este diagnóstico introduciendo conceptos objetivos en el mismo. Estos factores influyen directamente en que el posible tratamiento sea quirúrgico, radioterapia, quimioterapia o una combinación con el uso de medicamentos [1].

La citología clínica es el área que se encarga del estudio de la célula. El primer estudio de células malignas corresponde a un conjunto ilustraciones de células

observadas al microscopio hechas por Mueller en 1838 [1]. Desde entonces la citología se ha desarrollado como una disciplina y han venido apareciendo grandes avances en microscopía como los atlas de citología de diversos órganos. Uno de los trabajos más importante en patología es el del rumano Papanicolaou, en Nueva York, quién estudio el cáncer de cuello uterino e introdujo el examen de citología vaginal como un análisis para determinar la presencia de dicha patología [1].

Recientemente se han reportado diferentes proyectos que buscan introducir el diagnóstico automático de cáncer basado en imágenes histopatológicas. En el reporte técnico de Demir y Yener [2] se encuentran documentados algunos de los pasos del diagnóstico de cáncer, ilustrados en la figura 1, estos pasos incluyen: preprocesamiento para eliminar ruido y segmentar regiones de interés en imágenes microscópicas, selección y extracción de características. Para el caso del cáncer de origen cervical entre las características mencionadas se encuentra la textura, basada en fractales y la topología. Finalmente en el diagnóstico se emplean algoritmos de aprendizaje con clasificadores, métodos de evaluación y pruebas estadísticas.

Otro resumen importante es el de Loukas y Linney [3], el cual presenta el análisis de imágenes histológicas basado en tres factores biológicos que influyen en la respuesta de los tumores a la radiación: la proliferación, la hipoxia y la vasculatura. Al igual que la revisión de Demir y Yener [2] presentan diferentes pasos donde se destaca el problema de segmentar las células. Aquí hay muchos enfoques que incluyen técnicas tradicionales de imágenes usando umbralización, detección de bordes con máscaras de gradiente, hasta técnicas donde se aprovecha la información de color incluyendo métodos usando modelos donde se separa la componente de iluminación como el HSI (Hue Saturation Intensity) y el HSL (Hue Saturation Luminance). También uso de algoritmos de clasificación de píxeles en el espacio de color RGB, también usando análisis de componentes principales (ACP) fusión de regiones, emparejamiento de plantillas (contornos activos, transformada de Hough), análisis de texturas y clasificación (Redes neuronales, lógica difusa) [3].

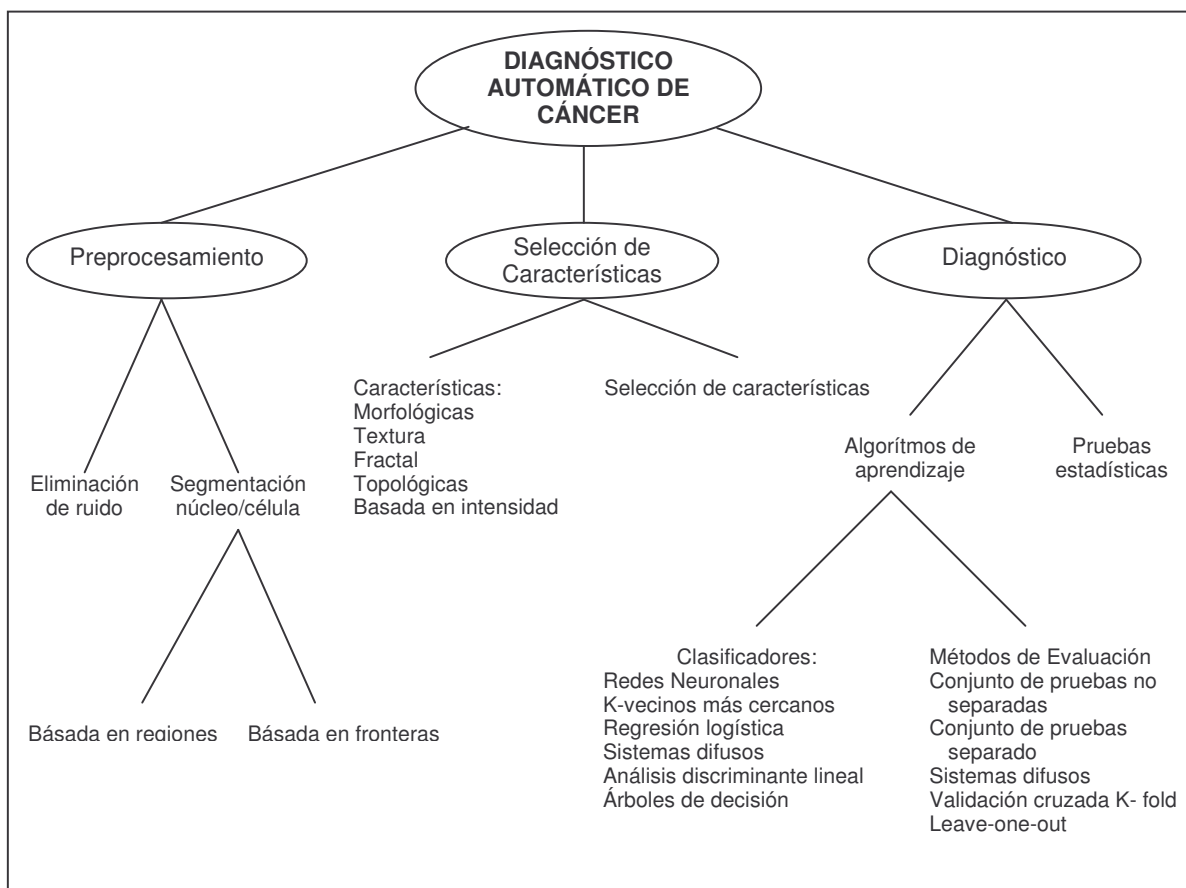


Figura 1. Revisión de pasos en diagnóstico automático de cáncer. Adaptado de Demir y Yener [1]

Por otro lado, Lezoray [4] en su tesis doctoral hace segmentación de células en imágenes en color con morfología matemática aplicada a imágenes de redes neuronales. Un trabajo anterior hecho por Walker et al. en 1994 [5], aplicado a células cervicales usa también morfología y extracción de características de textura.

Entre las técnicas de segmentación de células con aplicaciones en cancerología hay una comparación de métodos supervisados, basada en píxeles hecho por Maurie et al. [6], con imágenes de cáncer bronquial, en donde usa un esquema de segmentación en cinco etapas: simplificación de imágenes usando pre-tratamiento con suavizado, eliminación de ruido; seguido de clasificación de píxeles usando Bayes, kNN, SVM; extracción de marcadores basados en morfología matemática; luego watersheds para hacer crecimiento usando la información de color de la imagen obteniendo las regiones correspondientes al núcleo y al citoplasma. En este

trabajo se concluyó que es de importancia la escogencia del binomio compuesto por el modelo de color y del clasificador para obtener mejores resultados, hay diferentes combinaciones que dan mejores resultados ellos encontraron que para Bayes los mejores espacios de color son YCh_1Ch_2 , RGB y XYZ, para SVM son YCh_1Ch_2 , HSL y RGB, y para kNN los mejores son HSL, XYZ y RGB [6].

Otros enfoques tratan de solucionar problemas que se presentan con segmentación de células usando representaciones paramétricas para hacer emparejamiento de plantillas como contornos activos o transformada Hough [3]. Entre estas también hay técnicas que plantean el problema de segmentar las células como un problema de optimización de una representación paramétrica de la forma celular. Segmentación satisfactoria de células aisladas, múltiples células y células traslapadas se logra con la técnica de Wu, Barba y Gil [7]. Este método hace minimización sobre una representación paramétrica pero el costo computacional es prohibitivo para imágenes grandes donde hay muchas células. Otro esquema basado también en optimización usa un método evolutivo basado en búsqueda tabú de Jiang y Yang [8]. Este método usa memoria de corto plazo y tiene una lista para diversificar el espacio de búsqueda pero el problema está precisamente en la determinación del tamaño de esta lista. El número de familias es importante, si es pequeño se acerca a la búsqueda tabú, de otra manera la velocidad de convergencia empeora. Además el determinar al número de hijos también afecta la velocidad de convergencia para ese caso usaron valores entre 2 y 8. Además los autores encontraron que este método tenía casos donde no se obtenían segmentaciones óptimas.

El cáncer cuello uterino (cáncer cervical) es hoy el segundo más común en mujeres en todo el mundo [9], el diagnóstico temprano de esta enfermedad en una etapa precancerosa es muy importante para la efectividad de los tratamientos. Existen diversos exámenes para su diagnóstico como la citología vaginal, la colposcopia, la inspección visual directa, las espectroscopías fluorescente y reflectiva, entre otras; la mayoría de países del tercer mundo usan estudios de citología vaginal o inspección visual directa ya que son de muy bajo costo. En los sitios donde se ha implementado la citología vaginal se ha logrado bajar la incidencia de cáncer en un 70% pero esta prueba tiene inconvenientes ya que la sensibilidad de la misma es sólo del 51%, y la interpretación para dar un diagnóstico es altamente subjetiva [9].

Hay diversos proyectos que trabajan con el cáncer de cuello uterino, se encuentra por ejemplo un trabajo que busca construir una gran base de datos de imágenes colposcópicas vía web para articular una red de investigación donde se pueda estudiar la incidencia del virus del papiloma humano (VPH) en el desarrollo de este cáncer [9].

Hay también proyectos donde se busca dar un soporte más objetivo al diagnóstico asociado a los exámenes de patología; usando herramientas computacionales que permitan a patólogos expertos dar mejor sus diagnósticos, reproducirlos y que puedan ayudar a patólogos jóvenes en el aprendizaje. Un ejemplo de esto es un proyecto realizado en el Reino Unido donde se hizo un estudio completo, resultado del cual crearon un sistema computarizado para dar soporte de decisiones en la clasificación de lesiones cervicales escamosas [10] basado en redes de decisión bayesianas (bayesian belief networks). En este proyecto se desarrolló una herramienta donde se van dando ciertas características observadas y el sistema ayuda a afirmar el diagnóstico. Emplearon como referencia la clasificación de neoplasia intraepitelial cervical usada en Reino Unido (se usa la sigla CIN del inglés Cervical Intraepithelial Neoplasia). Según esta clasificación las anomalías son CIN I, CIN II y CIN III de acuerdo al grado de diferenciación [10]. La subjetividad en los diagnósticos fue una de las motivaciones en este proyecto, ya que estudios anteriores habían mostrado variaciones considerables inter-observador e intra-observador entre las clasificaciones CIN y la separación entre CIN y lesión no neoplásica, incluyendo coilocitosis. La evaluación de este sistema probó 50 casos que fueron evaluados inicialmente por 2 patólogos especialistas (uno de ellos patólogo ginecológico) y 2 patólogos junior; luego mezclaron los casos aleatoriamente y los evaluaron nuevamente a la siguiente semana. Esto se hizo una segunda y tercera vez y fueron evaluados por los 4 patólogos y 2 estudiantes de medicina usando el sistema de soporte de decisiones. Intra-observador la variación del diagnóstico usando el método convencional evaluando el mismo caso en 2 ocasiones separadas, se encontraron variaciones, siendo el patólogo ginecológico quien tuvo el menor número de variaciones, 11 de 50 (22%), usando el sistema de decisión esto no cambio mucho esta evaluación no tuvo en cuenta la precisión del diagnóstico. Inter-observador encontraron variación entre los especialistas y los

patólogos junior, usando el método de diagnóstico convencional, empleando el sistema esta variación fue menor y también se encontró que los estudiantes tuvieron alguna precisión en el diagnóstico.

6. IDENTIFICACIÓN DEL PROBLEMA:

En Colombia el cáncer cuello uterino es uno de los de más alta incidencia, un estudio realizado en el Instituto Nacional de Cancerología (INC) durante 2002 [11] revisó los casos nuevos de cáncer en dicho instituto y encontró que: “los tipos de cáncer más frecuentes fueron cuello uterino (14,4%), piel (13,3%), mama (12,2%), estómago (6,5%) y sistema hematopoyético (5,5%). El cáncer de pulmón tuvo una baja participación (2,2%). El 79,8% de los casos proviene de Bogotá y los departamentos vecinos”[11]. Otro estudio del mismo instituto hecho sobre causas básicas de defunción en el año 2002; muestra que el cáncer de cuello uterino es la segunda causa de defunción con 10.8% (el primero es el cáncer de estómago con 10.9%) y la primer causa de defunción en mujeres con un porcentaje del 18.8% [12].

La biopsia de cervix es uno de los estudios que se realizan para la determinación de alteraciones histopatológicas y permite ayudar en el diagnóstico de lesiones escamosas intraepiteliales (en [13] se encuentra una descripción detallada de estas lesiones). A las muestras se les aplica un método de tinción (Hematoxilina-Eosina) y se observan en el microscopio. El patólogo determina si hay alteraciones basándose en los siguientes criterios: pérdida de la polaridad de las células, mitosis anormales, alteración de la relación núcleo-citoplasma, hiperchromasia nuclear y alteración de la maduración. En general los patólogos tienen un alto grado de seguridad cuando la muestra es normal o cuando definitivamente tiene displasia, pero es muy difícil determinar el estado en algunos casos en los cuales se presenta un desorden celular temprano. Esto es muy importante, puesto que la efectividad de los tratamientos depende directamente de un diagnóstico temprano y de la posibilidad de hacer seguimiento y/o tratamiento a las pacientes que verdaderamente presentan patología pre-neoplásica.

Con este proyecto se busca encontrar métodos de cuantificación, que puedan ser usados como herramientas que permitan apoyar al patólogo en sus diagnósticos y hacerlos más objetivos.

7. OBJETIVO GENERAL Y OBJETIVOS ESPECÍFICOS: (Evaluables)

OBJETIVO GENERAL:

Desarrollar métodos de cuantificación que soporten el diagnóstico de lesiones escamosas basados de imágenes de cérvix.

OBJETIVOS ESPECÍFICOS:

- Seleccionar criterios de los empleados por los patólogos apropiados para construir una morfometría usando procesamiento digital de imágenes.
- Implementar algoritmos que permitan segmentar células en imágenes de cérvix.
- Validar los resultados de las segmentaciones de células en imágenes de cérvix con el apoyo de patólogos.
- Desarrollar una herramienta que pueda ser usada como apoyo por patólogos en el diagnóstico de lesiones escamosas en imágenes de cérvix.

8. METODOLOGÍA: (Descripción del procedimiento a emplear para obtener los objetivos propuestos).

Para el desarrollo del proyecto, se trabajará con una muestra de 10 imágenes en color tomadas de muestras de cérvix con método de tinción (Hematoxilina-Eosina) proveídas por el Departamento de Patología de la Universidad Nacional de Colombia, se cuenta con la asesoría de patólogos de dicho departamento los cuales darán el soporte en la determinación de las medidas de interés para ellos en este tipo de muestras y en la validación de resultados de las segmentaciones de núcleos celulares a lo largo del desarrollo del proyecto.

La investigación se desarrollará en las siguientes etapas:

1. Determinación de medidas morfométricas

En esta fase se realizará en conjunto con los patólogos un análisis que permita determinar medidas apropiadas para cuantificar lesiones escamosas, basadas en los criterios empleados por los patólogos a partir de la información que se pueda extraer de las imágenes. Por ejemplo cálculo de áreas y orientación de los núcleos.

2. Segmentación manual

Segmentaciones hechas por los expertos dibujadas a la mano que sirvan para validar las segmentaciones hechas por los algoritmos. Esta etapa comprende:

- Experimentación donde cada patólogo por separado dibujará las células de forma manual, determinando su núcleo y citoplasma. Esto se hará 5 veces sobre las imágenes con una separación entre cada experimento de una semana.
- Estudio de variabilidad intra-observador usando coeficiente de variabilidad.

3. Estudio de segmentación automática

El objetivo de esta fase es encontrar segmentaciones de las células. Para esto se establecen las siguientes actividades:

- Preprocesamiento para mejora de las imágenes.
- Estudio de segmentación usando técnicas basadas en aprendizaje: KNN, SVM y redes neuronales.
- Post procesamiento usando conocimiento a priori (pe. forma de las células) en modelo bayesiano para descartar objetos que no corresponden a células.
- Exploración de nuevas técnicas.

4. Validación de algoritmos de segmentación

Se realizará mediante un estudio de la variación entre los resultados de las segmentaciones obtenidas por los algoritmos y las obtenidas por los patólogos empleando un test Chi cuadrado.

5. Construcción de morfometría

La implementación de una herramienta computacional para ayuda de diagnóstico se ejecutará basándose en una metodología para el desarrollo del software asociado al proyecto se basa en la construcción evolutiva de prototipos de que permitan acercarse incrementalmente

Por otro lado el proyecto requiere la implementación de una herramienta computacional para ayuda de diagnóstico. La metodología para el desarrollo del software asociado al proyecto se basa en la construcción evolutiva de prototipos de que permitan acercarse incrementalmente a la solución. Esta se hará durante todo el desarrollo del proyecto, inicialmente se diseñará un prototipo que permita implementar y probar de manera rápida los métodos producto de la tesis y el cual inicialmente tendría métodos de apertura de imágenes, histograma y una herramienta de captura de regiones para determinación manual de células. Luego se construirán prototipos donde se integren los métodos de segmentación de células y la morfometría. Antes de empezar el desarrollo hay una etapa inicial donde se evaluará el uso de librerías ya constituidas como ImageJ, VTK, ImageMagick o la construcción de un aplicativo propio.

Durante el desarrollo de la tesis se seguirá recolectando bibliografía reciente acorde con el problema, y se confrontará con el fin de obtener la mayor cantidad de datos pertinentes al tema dinamizando de esta forma el proceso de desarrollo de la tesis.

Durante el desarrollo del proyecto se documentará el trabajo de manera que se vaya confeccionando adecuadamente el documento de tesis y los artículos producto del trabajo realizado.

9. ACTIVIDADES A DESARROLLAR:

- Recolección de información.
- Determinación de medidas morfométricas.
- Segmentación manual
 - Segmentación manual
 - Estudio de variabilidad intra-observador
- Estudio de segmentación de células.
 - Preprocesamiento.
 - Estudio de segmentación.
 - Postprocesamiento.
- Validación de algoritmos de segmentación.
- Construcción de la morfometría.
- Construcción de prototipos de software
 - Prototipo inicial.
 - Análisis de requerimientos incluye determinación de plataforma de desarrollo.
 - Diseño
 - Desarrollo de prototipo.
 - Pruebas
 - Prototipo con métodos de segmentación de células.
 - Análisis de requerimientos.
 - Diseño
 - Desarrollo de prototipo.
 - Pruebas
 - Prototipo con morfometría
 - Análisis de requerimientos.
 - Diseño
 - Desarrollo de prototipo.
 - Pruebas
- Documentación
- Escritura de documento de tesis
- Escritura de Artículo

10. CRONOGRAMA:

ACTIVIDAD		TIEMPO (SEMANAS)																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Recolección de información.																									
Determinación de medidas morfométricas																									
Segmentación manual																									
	Segmentación manual																								
	Estudio de variabilidad intra-observador																								
Estudio de segmentación de células.																									
	Preprocesamiento																								
	Estudio de segmentación.																								
	Postprocesamiento																								
Validación de algoritmos de segmentación																									
Construcción de la morfometría																									
Construcción de prototipos de software																									
	Prototipo inicial.																								
	Análisis de requerimientos incluye determinación de plataforma de desarrollo.																								
	Diseño																								
	Desarrollo de prototipo.																								
	Pruebas																								
Prototipo con métodos de segmentación de células																									
	Análisis de requerimientos.																								
	Diseño																								
	Desarrollo de prototipo.																								
	Pruebas																								
Prototipo con morfometría																									
	Análisis de requerimientos.																								
	Diseño																								
	Desarrollo de prototipo.																								
	Pruebas																								
Documentación																									
	Escritura de documento de tesis																								
	Escritura de Artículo																								

11. BIBLIOGRAFÍA BÁSICA:

- [1] O. Lezoray. "La réalisation d'une étude sur le tri informatique en cytologie". Maistre these. DEA d'Intelligence Artificielle et Algorithmique, Université de Caen. 1995.
- [2] C. Demir y B. Yener. "Automated cancer diagnosis based on histopathological images: a systematic survey". Reporte Técnico: *Rensselaer Polytechnic Institute, Department of Computer Science, TR-05-09*, 2005.
- [3] C. G. Loukas y A. Linney. "A survey on histological image analysis-based assesment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature". En: *computer Methods and Programs in Biomedicine*, 74:183 -199. 2004.
- [4] O. Lezoray. "Segmentation d'images couleur par morphologie mathématique et classification de données par réseaux de neurones : application à la classification de cellules en cytologie des séreuses". Ph. D. These. ECOLE DOCTORALE Structure, Information, Matière et Matériaux, U.F.R de Sciences, Université de Caen/Basse-Normandie. 2000.
- [5] R. Walker, P. Jackway, B. Lovell y I.D. Longstaff. "Clasifcation of cervical cell nuclei using morfological segmentation and textural feature extraction". En: *Second Australian and New Zealand Conference on Intelligent Information Systems*, Brisbane. 1994.
- [6] C. Meurie, G. Lebrun, O. Lezoray, A. Elmoataz. "A comparison of supervised pixel-based color image segmentation methods, Application in cancerology". En: *WEAS Transactions on Computers*, special issue on ICROSSIP'03, 02(03):739 – 744. 2003.
- [7] H-S. Wu, J. Barba y J. Gil. "A Parametric Algorithm for Segmentation of Cell Images". En: *IEEE Transactions on Biomedical Engineering*. 45 (3): 400 - 407 1998.
- [8] T. Jiang y F. Yang. "An Evolutionary Tabu Search for Cell Image Segmentation". En: *IEEE Transactions on Systems Man and Cybernetics-Part B: Cybernetics*, 32 (5): 675 – 678. 2002.
- [9] D. Ferris, "Analysis of digitized cervical images detect cervical neoplasia". En: *Medical Imaging 2004: Image Processing*, Proceedings of SPIE Vol. 5370. 2004.

- [10] G. J. Price, W.G. McCluggage, et al. "Computerized Diagnostic Decision Support System for the Classification of Preinvasive Cervical Squamous Lesions". En: Human Pathology, 34 (11):1193-1203,2003.
- [11]C. Pardo, R. Murillo, M. Piñeros y M. A. Castro. "Casos nuevos de cáncer en el Instituto Nacional de Cancerología, Colombia, 2002". En: Revista Colombiana de Cancerología, 07(03):4-19. 2003.
- [12]M. A. Castro, M. Piñeros y T. Martínez "Mortalidad en el instituto nacional de cancerología, Colombia, 2002" . En: Revista Colombiana de Cancerología, 07(03):20-13. 2003.
- [13]J. S. Torres. "Lesiones escamosas intraepiteliales cervicales. L.E.I.C". En: Revista Colombiana de Obstetricia y Ginecología, 49(04).1998.

12. RECURSOS FÍSICOS: (Especificar la disponibilidad y adjuntar carta de compromiso de la dependencia o empresa cuando sea necesario).

Se cuenta con siguientes recursos físicos:

- Recursos existentes en el Centro de Telemedicina de la Universidad Nacional (Servidores y equipos de cómputo desktop existentes).
- Equipo de cómputo del proponente.
- Acceso a internet por medio de la Universidad Nacional y a través del proponente.

13. COSTOS DEL TRABAJO Y FUENTES DE FINANCIACIÓN:

ITEM	Valor Unitario	Cantidad	VALOR FINANCIACION		Total
			Universidad	Estudiante	
1. RECURSOS DE INFORMACION					
a) Publicaciones	\$ 500,000	Global		\$ 500,000	\$ 500,000
b) Fotocopias	\$ 50	500		\$ 25,000	\$ 25,000
c) Internet	\$ 1,000	600 horas		\$ 600,000	\$ 600,000
d) Otros	\$ 200,000	Global		\$ 200,000	\$ 200,000
2. EQUIPO Y SOFTWARE					
a) Computador	\$ 4,000,000	Global		\$ 4,000,000	\$ 4,000,000
b) Otros equipos (red Universidad)	\$ 5,000,000	Global	\$ 5,000,000		\$ 5,000,000
c) Software	\$ 0	Global		\$ 0	\$ 0
d) Otros	\$ 400,000	Global		\$ 400,000	\$ 400,000
3. INSUMOS					
a) Papeleria	\$ 10,000	2 Resmas		\$ 20,000	\$ 20,000
b) Recursos de Almacenamiento	\$ 200,000	Global		\$ 200,000	\$ 200,000
c) Empastes	\$ 50,000	1		\$ 50,000	\$ 50,000
d) Toner impresora	\$ 100,000	2		\$ 200,000	\$ 200,000
d) Otros	\$ 200,000	Global		\$ 200,000	\$ 200,000
4. COSTO PERSONAL					
a) Director de proyecto	\$ 80,000	30 horas/mes	\$ 14,400,000		\$ 14,400,000
c) Elaboración	\$ 30,000	160 horas/mes		\$ 28,800,000	\$ 28,800,000
d) Evaluación (propuesta, proyecto terminado)	\$ 50,000	24 horas	\$ 1,200,000		\$ 1,200,000
5. COSTO PUBLICACIONES					
a) Simposio	\$ 4,000,000	Global		\$ 4,000,000	\$ 4,000,000
b) Articulo	\$ 300,000	Global		\$ 300,000	\$ 300,000
	TOTAL				\$ 60.095,000

14.COMENTARIO CON VISTO BUENO DEL DIRECTOR: (calificar los siguientes aspectos: organización, pertinencia, relevancia y originalidad).

Organización: Adecuada.

Pertinencia: Es un trabajo que requiere trabajo innovador desde el punto de vista computacional y la herramienta resultados es pertinente para el trabajo en patología.

Relevancia: Tiene gran importancia social para el país.

Originalidad: La propuesta es innovadora.

FIRMA DEL PROPONENTE

Ing. Eduard Leonardo Sierra Ballén

15. FIRMA DEL DIRECTOR (ASESORES)

Eduardo Romero Ph.D.
Director

16. FECHA: 17 de Febrero de 2006.