

# **SISTEMA EXPERTO PARA LA DETECCION DE CANCER A LA GLANDULA TIROIDES - SIECAT**

Sergio Ramírez Q.<sup>(1)</sup>, Cristina Ponce E.<sup>(2)</sup>, Indira Nolivos A.<sup>(3)</sup>

Facultad en Ingeniería Eléctrica y Computación

Escuela Superior Politécnica del Litoral (ESPOL)

Campus Gustavo Galindo, Km 30.5 vía Perimetral

Apartado 09-01-5863. Guayaquil-Ecuador

sframirezq@gmail.com<sup>(1)</sup>, cponce@espol.edu.ec<sup>(2)</sup>, inolivos@espol.edu.ec<sup>(3)</sup>

## **Resumen**

El objetivo del presente trabajo es el de desarrollar un Sistema Experto para el diagnóstico médico de cáncer de la glándula Tiroides (SIECAT) basado en la técnica de redes bayesianas que establezca la probabilidad de un paciente de obtener un diagnóstico positivo a partir de la valoración de su sintomatología. En la construcción del modelo bayesiano se incorporó el conocimiento de un experto oncólogo. Para la consecución del objetivo se desarrolló una interfaz de usuario que permite el acceso en línea de doctores en medicina -especialistas y no especialistas- a la información y conocimiento contenidos en el Sistema Experto, de tal forma que el SIECAT pueda ser utilizado como una herramienta de consulta.

**Palabras clave:** Sistema Experto, Redes Bayesianas, Software, Medicina.

## **Abstract**

The objective of the present work is to develop an expert system for the diagnosis of cancer in the Tiroidea gland (SIECAT) using the Bayesian Network technique. The SIECAT system permits to establish the probability of a patient having a positive diagnostic given the results from the medical evaluation of the present symptoms. The Bayesian model incorporates the knowledge of an expert oncologist. The system interface facilitates the access of medical doctors – specialist or not- to the information and knowledge contained in the expert system, to use SIECAT as an assessment tool.

**Keywords:** Expert system, Bayesian Networks, Software, Medicine.

## 1. Introducción

Los grandes progresos en la medicina están fundamentados en una mejor comprensión de la anatomía y fisiología humana, en el descubrimiento de las células y sus estructuras, el mejor entendimiento de la bioquímica y de la genética, los avances en la prevención y en la terapéutica de diversas enfermedades y finalmente en los logros relacionados con la aplicación de la estadística a la medicina. La medicina es un área en donde se requiere de mucho entrenamiento para ser un especialista, además, cuando existe una amplia diversidad de enfermedades, los síntomas de un paciente pueden llegar a confundir a un médico cuando se busca determinar rápidamente un diagnóstico oportuno, que puede significar la sobrevivencia o la muerte del paciente.

La Informática Médica, que agrupa los campos del software y el hardware para su uso en la Medicina, buscan reforzar y mejorar la toma de decisiones médicas y la atención al paciente. Como un lógico proceso de desarrollo, la Medicina ha ido asimilando la introducción de las computadoras para agilizar y mejorar los procesos de apoyo médico, teniendo una gran influencia, la que sigue aumentando más cada día con la introducción de la Inteligencia Artificial en la vigilancia del paciente con complejos equipos biomédicos, realización de procesamiento voluminoso de información para la toma de decisiones y muchas otras aplicaciones.

En los Estados Unidos, los estimados más recientes de la Sociedad Americana Contra El Cáncer en cuanto al cáncer de tiroides para el 2011 indican que:

- Alrededor de 48,020 nuevos casos de cáncer de tiroides (36,550 en mujeres y 11,470 en hombres) serán diagnosticados.
- Alrededor de 1,740 personas (980 mujeres y 760 hombres) morirán a causa de cáncer de tiroides.

El cáncer de tiroides se diagnostica comúnmente a una edad más temprana en comparación con la mayoría de los otros cánceres que afectan a los adultos. Casi dos de cada tres casos se encuentran en personas que tienen entre 20 y 55 años de edad.

La probabilidad de ser diagnosticado con cáncer de tiroides ha aumentado en los últimos años. Hoy día, la probabilidad es más del doble en comparación con 1990. En parte, esto se debe al aumento en el uso de ecografía de la tiroides, la cual puede detectar pequeños nódulos de tiroides que de otra manera no se hubiesen encontrado. De todos modos, al menos parte de este aumento también se debe a la detección de más tumores de mayor tamaño.

El cáncer de tiroides representa el 1 % de todos los cánceres. Hay 4 tipos principales de cáncer de tiroides: papilar, folicular, medular y anaplásico (Cifras de Ecuador) [1]. La causa del cáncer de tiroides no se conoce, pero se han identificado algunos factores de riesgo. Por ejemplo, las personas que hayan recibido radiación en la cabeza, el cuello o el tórax durante la infancia tienen un riesgo mayor que la población general de desarrollar cáncer de tiroides.

Entre los sistemas expertos desarrollados en el área de diagnóstico médico se puede mencionar el *Sistema HDDR* (Herramienta Didáctica para un Diagnóstico Reumatológico), sistema computacional automatizado pedagógico, diseñado para guiar a estudiantes de medicina en la obtención de un diagnóstico médico de afecciones reumáticas empleando agentes inteligentes basados en reglas utilizando un algoritmo, que opera mediante reglas heurísticas, que posee una base de conocimientos y de experiencia de un especialista en el área de reumatología [2]; y el sistema *PROSTANET*, una red bayesiana para el diagnóstico del cáncer de próstata. Su construcción manual se hizo con la ayuda de un urólogo, con un avance de 5 versiones hasta el momento [3].

## 2. Análisis del problema

La relación médico general y paciente comienza cuando se manifiestan las primeras etapas de la enfermedad, además la sintomatología del cáncer de tiroides representa una serie de síntomas que se asemejan a otras enfermedades tales como problemas con las amígdalas, bocio, entre otras. El médico realiza su proceso de decisión médica y aplica el tratamiento respectivo, a pesar de la constancia

del paciente en realizar las citas y seguir el tratamiento, los resultados no son los esperados.

El paciente invierte tiempo y dinero en vano, por consiguiente el cuadro clínico no mejora aunque el médico en consultas siguientes cambie el tratamiento. En vista de ello el paciente es remitido a otro especialista o el paciente cambia de médico, esto podría resultar en que se vuelva a repetir el cuadro anterior y el cuadro clínico siga desmejorando. Esto sumado a que existe la posibilidad de que exista una escasez de especialistas en el medio, sea por la ubicación geográfica del paciente o porque los demás médicos no conocen especialistas en oncología.

Por lo que un médico que no es especialista oncólogo o que tiene poca experiencia en el área, corre el riesgo de errar en su diagnóstico y aplicar un tratamiento equivocado, comprometiendo la salud del paciente.

### **3. Solución**

Dado que actualmente en el mercado nacional no existe un software especializado que ayude a un correcto diagnóstico para el cáncer de la glándula tiroides, se propone desarrollar un Sistema Experto que utilice un modelo de red bayesiana para determinar la probabilidad de obtener un diagnóstico positivo a partir de la valoración de la sintomatología del paciente.

En el modelo bayesiano se incorpora el conocimiento de un experto oncólogo, de tal forma que el proceso de decisión del diagnóstico médico es capturado. Para la consecución del objetivo se desarrolló una interfaz de usuario que permite el acceso en línea de médicos especialistas y no especialistas a la información y el conocimiento contenidos en el Sistema Experto.

Se plantea que el SISTEMA EXPERTO PARA LA DETECCIÓN DEL CÁNCER A LA GLÁNDULA TIROIDES (SIECAT), puede ser utilizado como una herramienta efectiva de consulta para mejorar la capacidad de diagnóstico de los usuarios médicos para esta patología.

### **4. Redes Bayesianas**

Una red bayesiana provee una forma compacta de representación del conocimiento, utilizando métodos flexibles de razonamiento basados en la teoría de probabilidad condicional de Bayes (Teorema de Bayes).

Entre las características que poseen las redes bayesianas, se puede destacar que permiten aprender sobre relaciones de dependencia y causalidad, permiten combinar conocimiento con datos, evitan el sobreajuste de los datos y pueden manejar bases de datos incompletas. Existen muchos modelos de redes bayesianas, para este trabajo se usará el modelo Naive Bayes (NB).

Naive Bayes es el modelo más simple de clasificación con redes bayesianas, ya que asume independencia entre todos los atributos dada una clase. NB es, por tanto, un modelo de atributos independientes. En este caso, la estructura de la red es fija y sólo es necesario aprender los parámetros (distribución de probabilidades). El fundamento principal del clasificador NB es la suposición de que todos los atributos son independientes conocido el valor de la variable clase.

A pesar de que asumir esta suposición en el clasificador NB es sin duda bastante fuerte y poco realista en la mayoría de los casos, se trata de uno de los clasificadores más utilizados. Además sus resultados son altamente competitivos contra los resultados de otras técnicas [4].

### **5. Proceso de diagnóstico de cáncer de glándula Tiroides**

Para resolver el problema planteado, se representó el modelo mental de un experto oncólogo, utilizando la técnica de redes bayesianas, a partir de la información obtenida en varias entrevistas presenciales. Los procesos de entrada y salida identificados se presentan a continuación:



- Computador Intel (R) Core TM 2 Duo CPU 1.67 GHz, 1 Gb Ram
- Disco Duro de 160 Gb.

**Software:**

- Sistema Operativo: Windows XP, Windows 7 Bussiness, Windows 7 Ultimate 32 y 64 bits.
- Software para implementar red bayesiana: GeNie 2.0

**Base de Datos:**

- My SQL.

**Lenguaje de Programación:**

- Netbean.

**Navegador:**

- IE versión 8 en adelante, Mozilla Firefox versión 8 en adelante.

**Gestor de servicio:**

- Wamp5.

## 7. Resultados

Para validar la red bayesiana se simularon 5 casos practicos, los cuales presentamos a continuación:

*El primer caso práctico está basado en el siguiente enunciado: Una persona adulta concurre a un chequeo médico para saber si por su condición de adulto (21 a 64 años de edad) es candidato a tener cáncer en la glándula tiroides, cual sería en porcentaje la distribución de la probabilidad.*

*El segundo caso práctico está basado en el siguiente enunciado: Una persona adulta de sexo masculino concurre a un chequeo médico para saber si por su condición de adulto y de su sexo es candidato a tener cáncer en la glándula tiroides, cual sería en porcentaje la distribución de la probabilidad.*

*El tercer caso práctico está basado en el siguiente enunciado: Una persona adulta de sexo femenino concurre a un chequeo médico para saber si por su condición de adulto y de su sexo es candidato a tener cáncer en la glándula tiroides, cual sería en porcentaje la distribución de la probabilidad.*

Los resultados obtenidos son los siguientes:

<b>Primer Caso</b>	
	<b>Rasgos Individuo</b>
<b>Candidato</b>	46%
<b>No es candidato</b>	54%
<b>Segundo Caso</b>	
<b>Candidato</b>	73%
<b>No es candidato</b>	27%
<b>Tercer Caso</b>	
<b>Candidato</b>	46%
<b>No es candidato</b>	54%

**Tabla 1** Distribución de probabilidad condicional para la variable rasgos del individuo, dados los casos 1, 2 y 3.

*El cuarto caso práctico está basado en el siguiente enunciado: Una persona concurre a un chequeo médico, el médico palpa la glándula tiroidea del paciente y nota que la consistencia de la glándula tiroidea es pétrea, bajo ese escenario, cuál sería en porcentaje la distribución de la probabilidad del diagnostico de la entrevista.*

Los resultados obtenidos son los siguientes:

	<b>Diagnostico de la entrevista</b>
<b>Positivo</b>	100%
<b>Negativo</b>	0%

**Tabla 2** Distribución de probabilidad condicional para la variable diagnóstico de la entrevista, dado que la consistencia de la glándula tiroidea es pétrea

*El quinto caso práctico está basado en el siguiente enunciado: Una persona concurre a un chequeo médico, el médico analiza el resultado de exámenes de laboratorio realizados y encuentra que los resultados indican que existe una inflamación de ganglios linfáticos, bajo ese escenario cual sería en porcentaje la distribución de probabilidad del diagnostico clínico.*

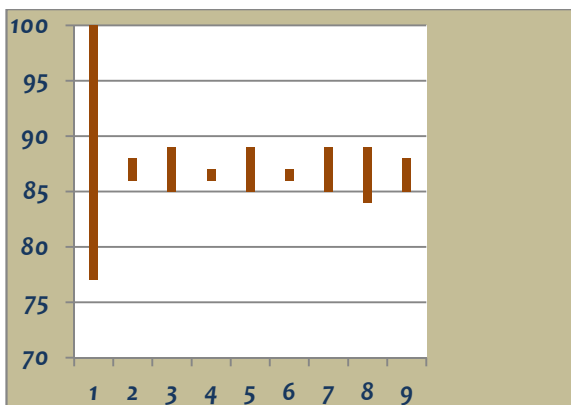
Los resultados obtenidos son los siguientes:

	Diagnostico Clínico
Positivo	84%
Negativo	16%

**Tabla 3 Distribución de probabilidad condicional para la variable diagnóstico clínico, dado que existe inflamación de ganglios linfáticos.**

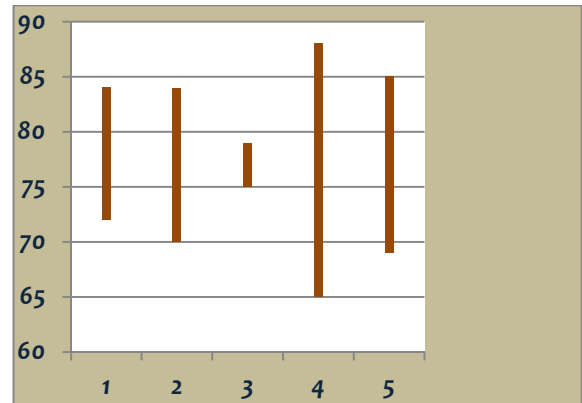
### 8. Análisis de sensibilidad

Presentamos la probabilidad de ocurrencia para cada proceso en el modelo bayesiano de la interface SIECAT:



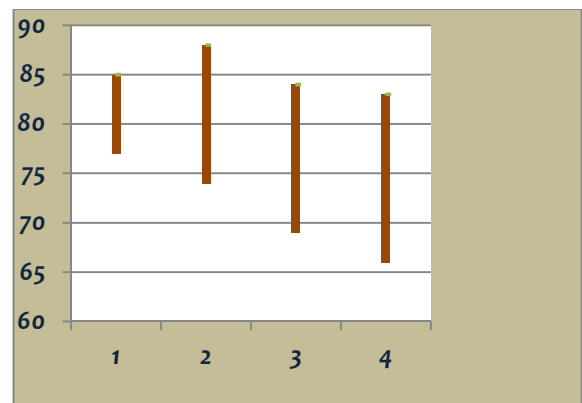
**Figura 4 Diagnostico de la entrevista – P (diag. entrevista=positivo para cáncer)**

En la figura 4 se muestra el análisis de sensibilidad de los síntomas que podría declarar el paciente cuando es entrevistado y se realizan exámenes físicos de parte del médico, las variables donde se ingreso evidencia son: Consistencia de la glándula tiroides (1), Agrandamiento de la glándula tiroides (2), Disfagia (3), Perdida de peso (4), Cambios de voz (5), Historia familiar con cáncer (6), Sexo (7), Edad (8), Riesgo Laboral (9) para el análisis de sensibilidad se ingreso evidencia a cada síntoma obteniendo que la probabilidad de que el Diagnostico de la Entrevista sea Positivo esta en un rango del 77 al 100%.



**Figura 5 Diagnostico clínico – P (diag. clínico=positivo para cáncer)**

En la figura 5 se muestra el análisis de sensibilidad de los datos que se extraen de los exámenes de laboratorio, las variables donde se ingreso evidencia son: Consistencia ganglios linfáticos (1), Inflamación ganglios linfáticos (2), Tipo inflamación glándula tiroides (3), Inflamación glándula tiroides (4), Ecografía (5), los resultados obtenidos son que la probabilidad de que el Diagnostico Clínico sea positivo va desde el 65 al 88 %.



**Figura 6 Diagnostico general – P (diag. general=positivo para cáncer)**

En la figura 6 se muestra el análisis de sensibilidad de las variables que representan los exámenes clínicos tales como: PAAF (1), ecografía (2), los diagnósticos de la entrevista (3) y clínicos (4); se obtuvo que la probabilidad de que el Diagnostico General sea positivo va desde el 66 al 88 %.

## 9. Conclusiones

Este trabajo nos ha permitido concluir lo siguiente:

Se cumplieron con los objetivos trazados en la tesis, de implementar una interface que permita a especialistas y no especialistas en oncología establecer la probabilidad de un paciente de padecer cáncer en la glándula tiroides a partir de la valoración de su sintomatología.

SIECAT puede ser utilizado como herramienta de referencia por expertos y no expertos en el área de oncología, pero debemos indicar que aun tiene ciertas limitaciones.

Hemos realizado un estudio crítico sobre la implementación de métodos para sistemas expertos y concluimos que la implementación mediante redes bayesianas es la más adecuada para implementar herramientas de diagnóstico médico; porque es el mejor método para evaluar la probabilidad de que dos eventos o sucesos diferentes ocurran simultáneamente.

El modelo sobre el que construyó la red bayesiana sin duda no es perfecto, pues recoge las creencias del experto oncólogo, dando lugar a que estas experiencias no sean tan precisas. El modelo podría ser mejorado realizando estudios más profundos sobre las variables definidas en la red bayesiana, y con un estudio clínico más detallado se podrían dar respuestas más categóricas en cuanto a la distribución de probabilidad conjunta de las variables que se están modelando. Las probabilidades dadas por el experto muestran un sesgo, este tema fue tratado cuando se hablo de las desventajas de usar una red bayesiana, se debe tener en cuenta que el modelo captura la forma de razonamiento del experto en cuanto al diagnóstico de la enfermedad.

## 10. Recomendaciones

SIECAT no está concluido, posiblemente siempre pueda verse sujeto a modificaciones debido a la naturaleza de

sus objetivos y al tipo de conocimiento que maneja.

La información que aporten casos clínicos de pacientes con historial médico que hayan presentado los síntomas de cáncer a la glándula tiroides, permitirá al sistema experto definir con un mayor grado de precisión las probabilidades condicionales de cada síntoma referido en nuestro trabajo. Se sugiere la creación y mantenimiento de una base de conocimientos junto con la interface con el objetivo de actualizar automáticamente la red bayesiana con nuevos síntomas y diagnósticos que manifieste la enfermedad con el fin de que la herramienta se encuentre actualizada.

## 11. Agradecimientos

Queremos agradecer a la PhD. Indira Nolivos y al Ing. Carlos Jordán, por su guía y soporte a lo largo del desarrollo de este trabajo. Un agradecimiento especial al Dr. Hollwyn Solórzano, por su contribución como experto oncólogo en la construcción del modelo experto en el cual se basa el sistema SIECAT, colaboración, su amabilidad, por dejarnos compartir su oficina y su tiempo durante las entrevistas realizadas.

## 12. Bibliografía

1. Montero Carpio Eduardo, "Protección Radiológica en el manejo de pacientes hospitalizados, sometidos a radioisoterapia con IODO-131 en el Instituto Oncológico Nacional Dr. Juan Tanca Marengo", <http://es.scribd.com/doc/38464767/proteccion-radiologica-en-el-manejo-de-pacientes-hospitalizados-sometidos-a-radioisototerapia-con-iodo-131,2003,pp.2>.
2. García Neri Héctor Manuel, "Herramienta Didáctica para un Diagnóstico Reumatológico", <http://www.slideshare.net/he>

ctorneri/herramienta-  
didactica-para-un-  
diganostico-reumatologico,  
2006,pp.1,38.

3. Lacave Rodero Carmen,  
"Explicación en redes  
bayesianas causales",

<http://www.cisiad.uned.es/tesis.html>,2003,pp. 166-167.

4. Sucar Luis Enrique, "Redes Bayesianas",  
<http://www.ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf>,  
2004,pp.1-5.