

Sistema experto probabilístico para el diagnóstico de adenocarcinoma gástrico

Probabilistic expert system for diagnosing gastric adenocarcinoma

Sistema especialista probabilístico para diagnosticar adenocarcinoma gástrico

Jorge Gutiérrez Gutiérrez¹, Carlos Castillo Diestra¹, Dayán Acosta Aparicio¹

Resumen

El objetivo del estudio fue construir un sistema experto probabilístico para diagnosticar la enfermedad adenocarcinoma gástrico. El sistema usa base de datos consistente en historias clínicas de pacientes que han tenido los síntomas relacionados a esa enfermedad independiente si padecieron o no de esta enfermedad. El sistema es probabilístico porque aplica el teorema de Bayes para calcular la probabilidad de que un paciente nuevo que pase consulta con un médico tenga la enfermedad dado que se conoce que tiene o no algunos o todos o ningún síntoma. Para la construcción del software para el sistema experto se ha utilizado una metodología en cascada, considerando análisis, diseño, implementación y pruebas, programado en el lenguaje prolog y conectado con una base de datos donde se tenga procesado la información de las historias clínicas las cuales son la entrada al sistema experto. Los resultados comprenden la recolección de la información pertinente al diagnóstico mediante análisis de documentos, la elaboración de la base de datos resumen en MySQL, el desarrollo del software que implementa el teorema de Bayes para el cálculo determinístico elaborado en SWI-prolog, así como las pruebas respectivas, las que indican que los resultados son satisfactorios.

Palabras clave: adenocarcinoma gástrico, diagnostic, sistema probabilístico.

Abstract

The objective of this study is to build a probabilistic expert system to diagnose gastric adenocarcinoma disease. The system uses database consisting of medical records of patients who have had symptoms related to the disease independently whether they suffered from this disease. The system is probabilistic because it applies the Bayes theorem to calculate the probability of passing a new patient consultation with a physician has the disease because it is known to have or not some or all or no symptoms. To build the software for the expert system has been used a cascade methodology, considering analysis, design, implementation and testing, scheduled in the prolog language and connected to a database where you have processed the information from medical records. The results of the research include the collection of relevant information to diagnosis through analysis of documents, the preparation of the database in MySQL summary, the development of software that implements the Bayes theorem for deterministic calculation prepared in SWI-Prolog, and the respective tests, indicating satisfactory results.

Keywords: gastric adenocarcinoma, diagnostic, system probabilistic.

Resumo

O objetivo deste estudo é a construção de um sistema especialista probabilística para diagnosticar a doença adenocarcinoma gástrico. O sistema utiliza banco de dados que consiste em prontuários de pacientes que tenham os sintomas relacionados com a doença, independentemente se eles sofreram com esta doença. O sistema é probabilística porque se aplica o teorema de Bayes para calcular a probabilidade de passagem de um novo paciente consulta com o médico tem a doença porque é conhecida por ter ou não alguns ou todos ou nenhuns sintomas. Para construir o software para o sistema especialista foi usado uma metodologia de cascata, considerando-se a análise, projeto, implementação e testes, programado na linguagem Prolog e conectado a um banco de dados onde você ter processado as informações dos prontuários médicos. Os resultados da pesquisa incluem a recolha de informações relevantes para o diagnóstico por meio de análise de documentos, a elaboração do banco de dados MySQL, em resumo, o desenvolvimento de software que implementa o teorema de Bayes para o cálculo determinista preparado em SWI-Prolog, e os respectivos testes, indicando que obteve resultados satisfatórios.

Palavras chave: adenocarcinoma gástrico, diagnostic, sistema probabilístico.

¹Universidad San Pedro, Facultad de Ingeniería, jlgutgut@yahoo.es

Recibido, 2 de febrero de 2016

Aceptado, 8 de marzo de 2016

Introducción

El adenocarcinoma gástrico es un tumor epitelial maligno, originado en el epitelio glandular de la mucosa gástrica. Invade la pared gástrica, infiltrándose en las mucosas de los uscularis, la submucosa y por lo tanto en la lámina propia de los muscularis.

Histológicamente, hay dos tipos importantes de cáncer gástrico según la clasificación de Lauren (Lauren, 1965): tipo intestinal y tipo difuso. En el adenocarcinoma tipo intestinal las células del tumor describen las estructuras tubulares irregulares, abrigando la pluriestratificación, lúmenes múltiples, tejido conectivo reducido. A menudo se junta la metaplasia intestinal en la mucosa vecina. Es asociado más a metaplasia y a un proceso de gastritis crónica. En cambio en el adenocarcinoma tipo difuso se tiene su origen en células mucosas gástricas no asociado a gastritis crónica y es mal diferenciado.

Según estadísticas del Instituto Nacional de Enfermedades Neoplásicas (Inen, 2011) en el Perú, de acuerdo a los resultados de los registros de cáncer poblacionales de Lima, Trujillo y Arequipa, se ha estimado que en el año 2004 habían ocurrido 4,541 casos nuevos. En las ciudades de Trujillo y Arequipa el cáncer de estómago es la primera causa de cáncer en ambos sexos (Inen, 2011). La casuística del INEN revela que en los últimos años se recibe en promedio 820 casos de cáncer de estómago, con tendencia al incremento.

La inteligencia artificial es una disciplina de la ciencia de la computación que tiene por finalidad emular el comportamiento y el razonamiento del ser humano. Tal que pueda por ejemplo imitar sus movimientos, hablar, escuchar, leer, escribir, razonar y resolver problemas para los que se necesita inteligencia.

Los investigadores Russell y Norvic (2004) plantean una interesante teoría llamada Agentes Inteligentes, en el que todo sistema de hardware y software es capaz de poder captar percepciones del medio ambiente interactuante y luego procesarlas inteligentemente mediante cualquier técnica para finalmente actuar, es llamado agente inteligente. Así la inteligencia artificial es la disciplina consistente en crear agentes inteligentes.

Uno de estos agentes son los sistemas expertos los que son capaces de recoger el conocimiento humano y trasladarlo a un ambiente digital para que sea utilizado por usuarios no expertos. Diferentes definiciones de Sistemas Expertos han sido propuestas hasta la fecha en la literatura de la IA. Sin embargo, a pesar de los diferentes enfoques y caracterizaciones utilizados en éstas, una base o fundamento común puede ser extraído de todas ellas: “un Sistema Experto es un sistema con pericia en la solución de problemas. Esto es, un sistema que posee conocimientos acerca de un dominio particular, comprensión de problemas del dominio y métodos de inferencia o razonamiento para manipular este conocimiento y resolver los problemas en la misma forma en que lo haría un experto humano”(Kemper, 2004).

En la teoría de la probabilidad el teorema de Bayes es un resultado enunciado por Thomas Bayes en 1763 que expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de sólo A.

La Inteligencia Artificial tiene mucho campo de aplicación tanto en la industria, en la empresa y en la sociedad. Sin embargo son pocas las aplicaciones que calan en las empresas locales, siendo las más populares los sistemas de reconocimiento de huellas,

las aplicaciones de reconocimiento de voz –frases cortas o palabras-, o los sistemas de control.

Esta investigación se justifica porque brinda a los médicos una herramienta inteligente que les permite diagnosticar enfermedades en este caso el adenocarcinoma gástrico. La investigación será una referencia obligatoria para la implementación de otros sistemas expertos para otras enfermedades, así como para la implementación de nuevos sistemas expertos. Los resultados del sistema pueden servir a médicos jóvenes para contrastar sus propios diagnósticos con los presentados por el sistema. El fundamento científico del sistema descansa en el teorema de Bayes. En esta investigación el problema se formula: ¿Cómo diagnosticar la enfermedad Adenocarcinoma Gástrico? La hipótesis postula que: El desarrollo de un sistema experto probabilístico permitirá diagnosticar la enfermedad Adenocarcinoma Gástrico. El objetivo general de la investigación se enuncia: Desarrollar un sistema experto probabilístico para el diagnóstico de la enfermedad Adenocarcinoma Gástrico.

Los objetivos específicos son recolectar y analizar información sobre el adenocarcinoma gástrico, en pacientes de hospitales de la localidad, elaborar una base de datos con la información recolectada, aplicar el teorema de Bayes en este contexto, construir un software para el sistema experto, probar el sistema para identificar inconsistencias.

Material y métodos

El tipo de investigación es tecnológica. La población objeto de estudio constituyeron las historias clínicas de 200 pacientes de un centro hospitalario de la ciudad de Chiclayo. Se aplicaron las técnicas del análisis documental, análisis heurístico, entrevistas. Se procesó y analizó la información de las historias clínicas para luego extraer los síntomas necesarios para el modelo estadístico que se está proponiendo. Luego se implementó una base de datos con el resultado del procesamiento de datos. Luego se desarrolló la interfaz gráfica del sistema. Por último se evaluó el sistema validando las salidas del mismo.

Resultados

Para alcanzar los objetivos propuestos en esta investigación se ha seguido una metodología de desarrollo de sistemas expertos de Weiss y Kulikowski (1984) la cual consta de las siguientes fases (figura 1):

1. Planteamiento del problema. La primera etapa en cualquier proyecto es normalmente la definición del problema a resolver. Puesto que el objetivo principal de un sistema experto es responder a preguntas y resolver problemas, esta etapa es quizás la más importante en el desarrollo de un sistema experto. Si el sistema está mal definido, se espera que el sistema suministre respuestas erróneas.
2. Encontrar expertos humanos que puedan resolver el problema. En algunos casos, sin embargo, las bases de datos pueden jugar el papel del experto humano.
3. Diseño de un sistema experto. Esta etapa incluye el diseño de estructuras para almacenar el conocimiento, el motor de inferencia, el subsistema de explicación, la interfaz de usuario, etc.

4. Elección de la herramienta de desarrollo. Debe decidirse si realizar un sistema experto a medida utilizando lenguaje de programación.
5. Desarrollo y prueba de un prototipo. Si el prototipo no pasa las pruebas requeridas, las etapas anteriores (con las modificaciones apropiadas) deben ser repetidas hasta que se obtenga un prototipo satisfactorio.
6. Refinamiento y generalización. En esta etapa se corrigen los fallos y se incluyen nuevas posibilidades no incorporadas en el diseño inicial.
7. Mantenimiento y puesta al día. En esta etapa el usuario plantea problemas o defectos del prototipo, corrige errores, actualiza el producto con nuevos avances, etc.



Figura 1. Metodología para el desarrollo del sistema experto.

El sistema experto debe asistir a un usuario final presentándole el cálculo de la probabilidad que tiene un paciente auscultado de tener la enfermedad Adenocarcinoma Gástrico, aunque el sistema se puede ensayar sobre cualquier otra enfermedad porque se utilizaría el mismo modelo científico, solo cambiarán las estadísticas históricas que requiere el modelo bayesiano.

Para tal fin es necesario recolectar estadísticas de pacientes que han presentado los síntomas asociados a esta enfermedad, entre otros tenemos: Pérdida del apetito, pérdida de peso (sin tratar de bajar de peso), dolor abdominal, malestar impreciso en el abdomen, justo encima del ombligo, sensación de llenura en el abdomen superior después de comer una comida pequeña, acidez (agruras) o indigestión, náusea, vómitos, con o sin sangre, hinchazón o acumulación de líquido en el abdomen, recuento bajo de glóbulos rojos (anemia)

Luego con esta información debemos relacionar en número cuantos pacientes han tenido la enfermedad Adenocarcinoma Gástrico y cuántos no, en un periodo de tiempo; de acuerdo a nuestra muestra se analizarán 200 casos. La información debe ser organizada en una base de datos.

El sistema experto probabilístico debe conectarse a la base de datos y recuperar la información pertinente es decir el número de pacientes enfermos que presentaron síntomas y el número de pacientes sanos que presentaron síntomas, luego se aplica el teorema de Bayes y el sistema reporta el valor de la probabilidad. Este cálculo asiste a

un usuario final por ejemplo un médico en la toma de decisiones respecto al paciente que pasa consulta. La información en las estadísticas permite inferir de una u otra forma si un paciente nuevo tiene la enfermedad o que tan probable es que la tenga. El software para el sistema experto debe ser visual, confiable y de preferencia debe correr en internet.

Para desarrollar el software del sistema experto se ha utilizado:

a) La IDE swi-prolog, es un entorno desarrollado por la Universidad de Amsterdam; esta IDE es un software libre por lo que no se ha incurrido en costos adicionales para su uso licenciado. Esta IDE permite editar, compilar y ejecutar código prolog.

El lenguaje prolog -programming in logic- es un lenguaje representativo de la Inteligencia Artificial, área a la cual pertenece este proyecto. Prolog es un lenguaje basado en lógica de predicados de primer orden y el motor de inferencia que utiliza infiere usando encadenamiento hacia adelante de reglas, administrando al mismo tiempo una pila para los cálculos intermedios.

Los autores hemos seleccionado esta herramienta sobretodo por tener la suficiente experiencia y dominio de la misma.

b) El motor de base de datos MySQL, es un software libre fácilmente portable que permite manejar de manera robusta e íntegra una base de datos, como la que se requiere en esta investigación. Mediante la tecnología ODBC la IDE swi-prolog se conecta con MySQL para acceder a través de consultas SQL a la información necesaria para el cálculo de la probabilidad bayesiana.

c) XPCE no es prolog, es una extensión a modo de librería escrita en lenguaje prolog y que permite manejar la interfaz gráfica, que es un componente importante en la arquitectura de un sistema experto.

El teorema de Bayes se aplica de la siguiente manera para esta investigación:

$$P_{bayes} = \frac{NumEnfermos \cap NumSintomas}{NumEnfermos \cap NumSintomas + NumSanos \cap NumSintomas}$$

La base de datos se presenta en la tabla 1 que resume la información histórica y la consolida con los campos de los síntomas y el estado del paciente de la muestra, el estado puede ser enfermo o sano. El campo número indica cuántos pacientes tuvieron el perfil de síntomas y estado de enfermedad respectivo a cada fila.

Tabla 1. Información histórica de los campos de los síntomas y estado del paciente

Id	dolor	vómito	pérdida peso	estado	número
1	Si	Si	Si	e	22
2	Si	Si	No	e	10
3	Si	No	Si	e	6
4	Si	No	No	e	2
5	No	Si	Si	e	6
6	No	Si	No	e	3
7	No	No	Si	e	1
8	No	No	No	e	0

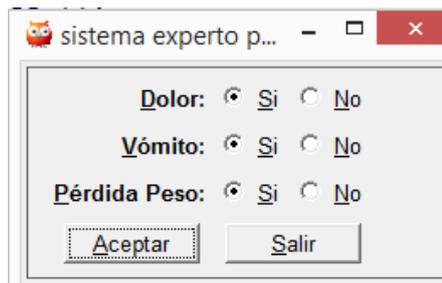


Figura 2. Interfaz de captura de datos

La conexión a la base de datos se ha manejado mediante ODBC por lo que se ha creado un origen de datos llamado ConectaBD con alias bd (figura 2), el mismo que conecta a una base de datos origen llamada Hospital y a una tabla llamada resumen. Sobre esta tabla se ensayan dos consultas requeridas para el modelo de probabilidades de Bayes, por un lado los pacientes enfermos y que presentaron síntomas de la enfermedad, y por otro lado los pacientes que no estuvieron enfermos pero sí tuvieron los mismos síntomas. En la lógica del código se llama a un predicado conexión, luego se carga la consulta y por último se ejecuta.

Finalmente el código prolog (figura 3) procesa las respuestas de las dos consultas hechas sobre la base de datos y calcula la probabilidad bayesiana mediante una simplificación de la fórmula a efectos de aligerar la línea de código.

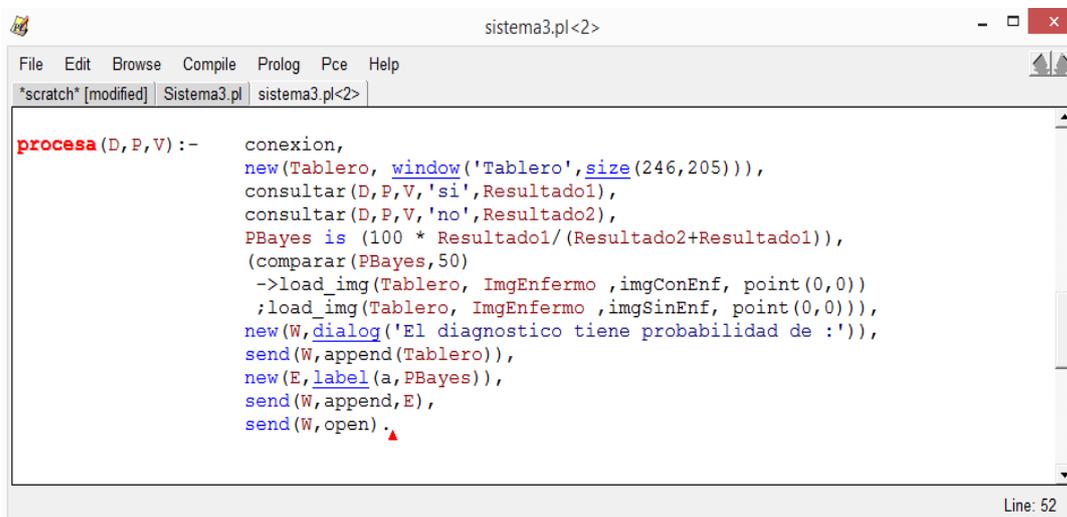


Figura 3. Código prolog para el cálculo de la probabilidad bayesiana.

Las interfaces de salida de resultados se presentan a continuación discriminando la imagen referencial presentada en función de un umbral mayor a 50% (figura 4).



Figura 4. A la izquierda, imagen de salida con alta probabilidad de tener la enfermedad. A la derecha, imagen de salida con baja probabilidad de tener la enfermedad.

Discusión

Se tuvieron limitaciones en cuanto al acceso a la información por cuanto no se lleva un registro estadístico del mismo en las instituciones de salud de Chimbote, el cual relacione los síntomas con su respectiva enfermedad; hemos tenido acceso a cuadros sumarios de personas y sus respectivas enfermedades. Hemos completado la información con fuentes externas y entrevistas a médicos especialistas, los cuales nos manifestaron que si bien tienen conocimiento de que hay buenos trabajos científicos que están probando y validando modelos estadísticos y formales sobre el diagnóstico de enfermedades como el adenocarcinoma gástrico, no es todavía suficiente como para que ellos opten por estas nuevas alternativas, toda vez que consideran que el médico tiene un conocimiento heurístico que es alimentado por una casuística producto de auscultar día a día a pacientes. Mientras mayor es el número de pacientes que son auscultados, pues mayor será su repositorio de conocimientos.

Nuestros resultados son alternativos a los antecedentes revisados del tema donde otras tecnologías y/o procedimientos son mayormente utilizados a la hora de hacer un diagnóstico médico. Así nuestra propuesta sigue en buena cuenta la aplicación de teorías probabilísticas que aún se están desarrollando y validando en la actualidad.

Se hicieron las pruebas de caja negra al software y se verificó que realiza los cálculos correctamente; sin embargo el sistema acepta entradas discretas es decir en términos de si y no. Para mejorar la fiabilidad de los resultados y puede ser más útil es necesario que los datos se puedan manejar de manera difusa ampliando en todo caso el número de respuestas por síntomas a una pequeña escala. Así podría ser mucho más refinada y precisa la probabilidad calculada.

Conclusiones

Se aplicó la metodología de Weiss y Kulikowski para construir el sistema experto probabilístico.

Se usó el gestor de base de datos MySQL para guardar la tabla que resume la información recolectada.

Se creó un origen de datos en Windows y se desarrolló en swi-prolog el código respectivo para acceder al origen, hacer las consultas a la base de datos, y presentar los resultados al usuario final.

Se han elaborado las pruebas de caja negra para verificar que el sistema experto es confiable; en tal sentido las pruebas demostraron que el sistema funciona correctamente estableciendo los cálculos correctos.

Referencias bibliográficas

- Kemper, N. (2004). *Apuntes curso sistemas inteligentes para la toma de decisiones*. Escuela de Posgrado UNT.
- Lauren P. (1965). The Two Histological main types of gastric carcinoma. Diffuse and so-called Intestinal type carcinoma: an attempt at histoclinical classification. *Acta PathoMicrobiol Scand*, 64: 31–49
- Registro de Cáncer de Lima Metropolitana, 2004-2005. (2011). *Informe Ejecutivo*. Instituto Nacional de Enfermedades Neoplásicas, Departamento de Epidemiología y Estadística.
- Rusell B. & Norvig U. (2004). *Inteligencia Artificial. Un enfoque práctico* (2º ed.). Mc Graw Hill.
- Weiss, S. M. & Kulikowski, C. A. (1984), *A Practical Guide to Designing Expert Systems*. Rowman and Allanheld, Totowa, N.J.