



Acta de Otorrinolaringología & Cirugía de Cabeza y Cuello

www.revista.acorl.org.co



Trabajos originales

Minería de datos para la evaluación del riesgo de malignidad en pacientes con nódulos tiroideos con estudio citológico Bethesda IV

Risk of malignancy evaluation through data mining technic in patients with thyroid nodules with cytology study Bethesda IV

Rogers Leonardo Baquero G.*, Esteban Diazgranados G.*, Elizabeth León G.**, Juan de Francisco Zambrano*, Álvaro Eduardo Calixto G.*, Andrés Felipe Rey*, Cesar Alfonso Palencia***, Juan Fernando Castañeda***, Erika León G.****.

- * Médico especialista en Cirugía General; especialista en Cirugía de Cabeza y Cuello, Fundación Universitaria Ciencias de la Salud, Hospital de San José. Bogotá, Colombia.
- ** MSc, PhD Ciencias de la Computación e Ingeniería. Profesora asociada, Universidad Nacional de Colombia. Bogotá, Colombia
- *** Médico especialista en Cirugía General, fellow de Cirugía de Cabeza y Cuello, Fundación Universitaria Ciencias de la Salud, Hospital de San José. Bogotá, Colombia.
- **** MSc Epidemiología Clínica; Instructor asistente, división de Investigaciones, Fundación Universitaria de Ciencias de la Salud, Hospital de San José. Bogotá, Colombia.

Forma de citar: Baquero R, Diazgranados E, León E, Zambrano J, Calixto A, Rey A, Palencia C, Castañeda J, León E. Minería de datos para la evaluación del riesgo de malignidad en pacientes con nódulos tiroideos con estudio citológico Bethesda IV. Acta otorrinolaringol. cir. cabeza cuello. 2022;50(1): 36-44. DOI.10.37076/acorl.v50i1.618

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido: 3 de Mayo de 2021
Evaluado: 8 de Marzo 2022
Aceptado: 10 de Marzo 2022

Palabras clave (DeCS):

Bethesda categoría IV, modelo predictivo, minería de datos, Naive Bayes, riesgo de malignidad.

RESUMEN

Introducción: en el campo de la salud, cada decisión representa datos, y las técnicas de minería de datos han empezado a ser una metodología prometedora para el análisis de esta información, especialmente en el diseño de los modelos predictivos. **Métodos:** estudio observacional analítico de pacientes mayores de 15 años, con reporte de punción de aspiración con aguja fina con estudio Bethesda IV, sometidos a manejo quirúrgico en el Hospital de San José de Bogotá. Los datos recogidos de los pacientes se incluyeron en tres grupos: la información sociodemográfica y clínica, los hallazgos en la citología y los reportes de la ecografía. Se realizó el análisis mediante Naive-Bayes, árbol de decisión y redes neuronales. Se usó la herramienta Weka versión 3.8.2. **Resultados:** de los 427 pacientes, 195 tuvieron resultados de patología de carcinoma de tiroides (45,6 %). Se evidenciaron mejores resultados

Correspondencia:

Rogers Leonardo Baquero García
Email: rlbaquero@fucsalud.edu.co
Dirección: Calle 10 No. 18-75, Hospital de San José, segundo piso, oficina de Cirugía General, Cirugía de Cabeza y Cuello. Bogotá, Colombia.
Teléfono: 3212044151

usando la validación cruzada (10 fold) comparado con partición (66 %), la técnica de Bayes tuvo mejores resultados de clasificación correcta (91,1 %), comparado con la técnica de árbol (87,8 %) y la red neuronal (88,2 %). *Conclusiones:* el uso de la técnica de Naive Bayes muestra una importante exactitud para determinar la predicción de riesgo de malignidad en los pacientes con estudio citológico Bethesda IV, lo cual permitiría orientar de forma adecuada el manejo quirúrgico de los pacientes.

ABSTRACT

Key words (MeSH):

Thyroid neoplasms, data mining, predictive value of tests, theorem Bayes, Naive Bayes.

Introduction: In the health field, each decision represents data, and data mining techniques have begun to be a promising methodology for the analysis of this information, especially in the design of predictive models. *Methods:* Analytical observational study; patients older than 15 years with a report of Bethesda IV after a fine needle aspiration biopsy that undergoing surgical management at the Hospital de San José in Bogotá. The data collected from those patients were included in three groups: sociodemographic-clinical information, cytology findings, and ultrasound reports. Analysis was performed using three technics: Naive Bayes, decision trees, and neural networks. Weka tool version 3.8.2 was used. *Results:* 195 patients out of 427, had a thyroid carcinoma pathology (45.6%). Better results were evidenced using cross-validation (10 fold) compared with a partition (66%), the Bayes technique had better results of correct classification (91.1%), than the tree technique (87.8%) and neural network (88.2%). *Conclusions:* The use of the Naive Bayes technique shows an important accuracy to determine the prediction of risk of malignancy in patients with a Bethesda IV cytological study, which would allow an adequate guide to the surgical management of patients.

Introducción

En la actualidad, los sistemas informáticos constituyen una tendencia creciente en países desarrollados. En términos de automatización en la salud, estos han demostrado que podrían influir en la toma de decisiones terapéuticas, con base en estudios diagnósticos (1). Esto ha permitido el desarrollo de técnicas que agrupan la inteligencia artificial y los modelos estadísticos, que se han catalogado bajo el nombre de minería de datos (data mining) (2). Estas técnicas pretenden extraer la información relevante de bases de datos, utilizando algoritmos, que identifican la información denominada no trivial, que generalmente no se distingue en el manejo de grandes cantidades de datos (2). La definición de esta técnica hace referencia a un proceso de exploración, transformación y selección de datos de manera automatizada, que incorpora el reconocimiento de patrones y que ha mostrado un amplio campo de aplicación (3).

Es posible decir que en el campo de la salud, cada decisión representa datos, y la minería de datos ha empezado a ser una técnica prometedora para el análisis de esta información (4). Esta técnica tiene como requisito una adecuada calidad de los datos, para inferir conclusiones que ahorren tiempo en la investigación manual, con los esfuerzos centrados en la educación y en permitir su aplicabilidad en intervenciones del área de la salud (5, 6). La técnica bayesiana con el algoritmo de Naive Bayes es la más utilizada dentro de este modelo. En diversos estudios ha mostrado una exactitud en

las predicciones igual o superior a las demás técnicas de minería de datos (7-9).

La guía de la American Thyroid Association (ATA) del 2015 recomienda el manejo quirúrgico para las lesiones clasificadas como Bethesda IV (10). Sin embargo, el punto de discusión radica en la identificación adecuada de los pacientes con nódulos benignos, que se someten a cirugías innecesarias (11).

A pesar del buen rendimiento de la citología obtenida mediante la punción por aspiración con aguja fina (PAAF) en identificar el resto de las lesiones malignas de la glándula tiroidea, en su mayoría carcinoma papilar, su sensibilidad en el diagnóstico de la neoplasia folicular continúa siendo baja (12). Osorio y colaboradores realizaron un estudio en el que se incluyeron 95 pacientes con lesiones nodulares de la glándula tiroidea, a los cuales se les realizó citología mediante PAAF, y reportaron Bethesda IV en 10 %, con una sensibilidad de este sistema de clasificación en el diagnóstico de neoplasias foliculares calculada de 30,8 %, una especificidad del 92,7 %, un valor predictivo positivo (VPP) del 40 % y un valor predictivo negativo (VPN) del 89,4 % (13, 14).

En razón a lo anterior se han realizado estudios para identificar los factores predictivos de malignidad en los pacientes con nódulo tiroideo Bethesda IV (14, 15), como lo son un tamaño mayor de 4 cm, la presencia de microcalcificaciones, la hipoeogenicidad en el ultrasonido y el aumento en la vascularización del nódulo por medio de ultrasonido Doppler (11, 16). Los factores clínicos, como la edad, el género (sexo

masculino), las lesiones de consistencia dura a la palpación, las adenopatías en estaciones laterales del cuello, las características ecográficas de sospecha de malignidad adicionales, como la estructura sólida y la identificación de los nódulos solitarios, se han reportado igualmente como factores predictores de malignidad (11).

El objetivo de este trabajo es analizar los datos a través de las técnicas de minería de datos, que permitan identificar en la población colombiana los factores predictores de malignidad en los pacientes con resultados de citología Bethesda IV, con el fin de determinar la necesidad del manejo quirúrgico.

Métodos

Se recolectó información de una cohorte primaria de pacientes con reporte de citología Bethesda IV obtenida por PAAF, sometidos a manejo quirúrgico en el Hospital de San José de Bogotá, en el período comprendido entre el 1 de junio de 2010 al 30 de junio de 2018. El estudio se avaló por el comité de ética e investigaciones de la Facultad de Medicina de la Fundación Universitaria de Ciencias de la Salud (FUCS).

Se incluyeron pacientes de 15 años o más, con reporte citológico Bethesda IV sometidos a tiroidectomía total o parcial. Se excluyeron aquellos con un reporte histológico indeterminado o insatisfactorio o con pérdidas de información de datos superior al 50 %. Se revisaron las historias clínicas de cada paciente, que incluían informes hospitalarios, reportes citológicos, imagenológicos y quirúrgicos. El resultado del estudio histopatológico para cáncer de tiroides se clasificó como benigno o maligno, según el reporte histológico del procedimiento quirúrgico reportado por el servicio de patología del Hospital de San José.

Se obtuvieron 26 variables, que se incluyeron, en tres grupos, la información sociodemográfica y clínica, los hallazgos en la citología y los reportes de la ecografía. En el primer grupo se encuentran 11 variables predictoras, como la edad, el sexo, los antecedentes patológicos, alérgicos y toxicológicos, también los antecedentes familiares patológicos y de cáncer de tiroides. Se tuvieron en cuenta variables del

examen físico, como adenopatías y naturaleza de los nódulos a la palpación. Dicha información se tomó de los reportes de la historia clínica, elaborada por especialistas en cirugía de cabeza y cuello.

Además, se tuvo en cuenta las características citológicas del reporte histológico, como la proliferación folicular abundante, el patrón microfolicular, el material coloidal, la celularidad, la presencia de núcleos grandes y las pseudoinclusiones nucleares. Los hallazgos ecográficos fueron la presencia de microcalcificaciones, la naturaleza de los nódulos (hipoecoico, ovalado, bordes irregulares, único, hipervascularizado), la clasificación Thyroid Imaging Reporting and Data System (TIRADS) y la presencia de adenopatías.

Las estadísticas descriptivas se calcularon y compararon entre los dos grupos: pacientes con cáncer de tiroides y pacientes sin cáncer de tiroides. Se usaron la prueba de chi cuadrado y la prueba de Fisher, para las variables categóricas, y la prueba de Mann Whitney, para las variables continuas, previa verificación de la distribución de la normalidad de los datos.

Se realizaron exploración de datos con tres técnicas de minería de datos: Naive Bayes, árbol (trees J48) y red neuronal. A partir de la información se hizo un análisis con partición por defecto (66 %) y validación cruzada (10 fold) para cada técnica. Para el análisis se utilizó la herramienta Weka versión 3.8.2.

Resultados

Se identificaron 427 pacientes con reporte citológico previo Bethesda IV, de los cuales 195 tuvieron resultado de patología de carcinoma de tiroides (45,6 %) en el espécimen quirúrgico. En el grupo de pacientes con carcinoma de tiroides, la mediana de edad fue de 59 años, con un rango de edad que osciló entre 22 y 88 años. En ambos grupos, los pacientes fueron predominantemente mujeres (89,7 frente a 90,5 %; $p = 0,79$) (Tabla 1).

Tabla 1. Características sociodemográficas y clínicas de los pacientes con estudio citológico de la tiroides por Bethesda IV para el manejo quirúrgico, de acuerdo con el resultado histológico

Característica	Total n = 427	Cáncer de tiroides n = 195	Sin cáncer n = 232	Valor de p
Edad,				
Mediana RIC	59 (47-68)	58 (46-63)	61(48,5-72)	0,001
Rango (min-máx.)	22-88	22-83	23-88	
Sexo				
Mujeres, n (%)	385 (90,1)	175 (89,7)	210 (90,5)	0,79
Antecedentes patológicos, n (%)				
Ninguno	189 (44,2)	88 (45,1)	101 (43,5)	0,40
Oncológicos	15 (3,5)	5 (2,5)	10 (4,3)	
Metabólicos	94 (22)	39 (20)	55 (23,7)	
Cardiovasculares	79 (18)	43 (22)	36 (15,2)	

Característica	Total n = 427	Cáncer de tiroides n = 195	Sin cáncer n = 232	Valor de p
Cardiovasculares y metabólicos	42 (9.,8)	16 (8,2)	26 (11,2)	
Otros	8 (1.,8)	4 (2)	4 (1,7)	
Antecedentes alérgicos, n (%)				
Ninguno	391 (91.,5)	181 (92,8)	210 (90,5)	0,55
Antibióticos	22 (5.,1)	9 (4,6)	13 (5,6)	
AINE	8 (1,8)	4 (2)	4 (1,7)	
Opioides	3 (0.,7)	0 (0)	3 (1,2)	
Otros medicamentos	3 (0.,7)	1 (0,5)	2 (0,8)	
Antecedentes toxicológicos, n (%)				
Ninguno	372 (87.,1)	171 (87,7)	201 (86,6)	0,9
Consumo de cigarrillo	40 (9.,3)	18 (9,2)	22 (9,4)	
Radioterapia por patología maligna	5 (1.,1)	2 (1)	3 (1,2)	
Bebedor habitual	2 (0.,4)	1 (0,5)	1 (0,4)	
Otro	8 (1.,8)	3 (1,5)	5 (2,1)	
Antecedente familiar de cáncer de tiroides, n (%)	9 (2.,11)	7 (3,5)	2 (0,8)	0,05
Antecedentes familiares, n (%)				
Ninguno	379 (88.,7)	170 (87.,1)	209 (90.,1)	0,55
Oncológico ginecológico	5 (1.,1)	3 (1.,5)	2 (0.,8)	
Oncológico gastrointestinal	14 (3.,2)	6 (3.,1)	8 (3.,4)	
Oncológico mama	5 (1.,1)	3 (1.,5)	2 (0.,8)	
Oncológico cabeza y cuello	11 (2.,5)	8 (4.,1)	3 (1.,2)	
Oncológico- otros	4 (0.,9)	1 (0.,5)	3 (1.,2)	
Metabólicos, Cardiovasculares y otros	9 (2.,1)	4 (2.,0)	5 (2.,1)	0.,07
Disfonía, n (%)	35 (8.,2)	11 (5.,6)	24 (10.,3)	
Nódulo palpable, n (%)	268 (62.,7)	137 (70.,2)	131 (56.,4)	0.,003
Nódulo sólido, n (%)	269 (63)	141 (72.,5)	128 (55.,1)	0.,000
Adenopatías palpables, n (%)	15 (3.,5)	10 (5.,1)	5 (2.,1)	0.,09

AINE: antiinflamatorios no esteroideos; RIC: rango intercuartílico.

De los 195 pacientes con diagnóstico de malignidad se encontró predominante el carcinoma papilar de patrón clásico y folicular (91,3 %), seguido del carcinoma folicular de tiroides (6,7 %), en menor proporción el carcinoma medular (1,5 %) y el indiferenciado de tiroides (0,5 %) (**Figura 1**).

Los antecedentes patológicos personales y familiares se organizaron en varios grupos, en vista de los múltiples diagnósticos encontrados. Dentro del grupo de antecedentes oncológicos familiares, el carcinoma de tiroides en familia de primera línea mostró una asociación con el diagnóstico final de cáncer de tiroides (3,5 % frente a 0,8 %; $p = 0,05$).

Los antecedentes metabólicos personales, en el que se incluyeron diagnósticos, como diabetes mellitus, síndrome metabólico y obesidad, tuvieron una distribución homogénea en ambos grupos. Los antecedentes cardiovasculares personales, en el que se incluyeron diagnósticos, como falla cardíaca, infarto agudo de miocardio y enfermedad coronaria, también tuvieron distribución similar en ambos grupos.

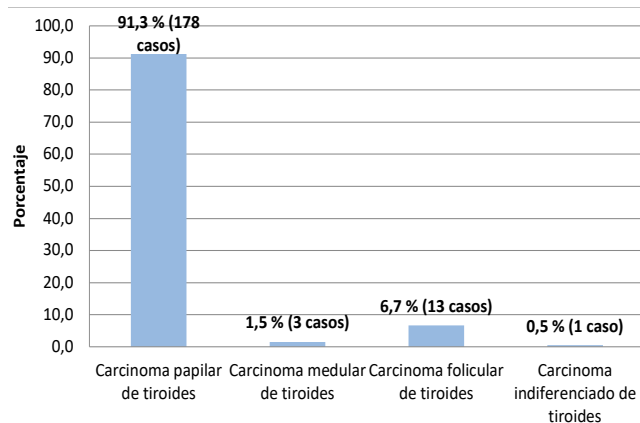


Figura 1. Distribución según el resultado histopatológico de los pacientes con carcinoma de tiroides.

No se hallaron diferencias entre los grupos en la manifestación de la disfonía (5,6 % frente a 10,3 %; $p = 0,07$); sin embargo, las características que se asociaron con el resultado de cáncer de tiroides fueron el nódulo palpable en el examen

físico (70,1 % frente a 56,4 %; $p = 0,003$) y la consistencia sólida del nódulo (72,5 % frente a 55,1 %; $p < 0,01$).

Con respecto a las características ultrasonográficas del nódulo tiroideo se identificaron rasgos fuertemente asociados con el cáncer de tiroides, como las microcalcificaciones (74,3 % frente a 14,2 %; $p < 0,01$), nódulo ovalado (22,5 % frente a 32,7 %; $p = 0,02$), nódulo hipoeoico (70,2 % frente a 37 %; $p < 0,01$), bordes irregulares (24,6 % frente a 3,4 %; $p < 0,01$) y la hipervascularización en el estudio Doppler (38 % frente a 6,4 %; $p < 0,01$). La clasificación TIRADS catalogó adecuadamente los pacientes con cáncer de tiroides a medida que aumentaba su denominación ($p < 0,01$), mien-

tras que tener un nódulo único no tuvo una asociación con malignidad estadísticamente significativa (**Tabla 2**).

En el estudio citológico mediante PAAF se identificaron factores asociados con el cáncer de tiroides, como la presencia de un patrón microfolicular (22,5 % frente a 10,7 %; $p < 0,01$), la identificación de los núcleos grandes (39,4 % frente a 6 %; $p < 0,01$), el aclaramiento de la cromatina (8,2 % frente a 3 %; $p = 0,01$) y la presencia de pseudoinclusiones nucleares (35,3 % frente a 0,8 %; $p < 0,01$), mientras la caracterización de la celularidad y la proliferación folicular abundante no tuvieron una asociación con la malignidad estadísticamente significativa (**Tabla 3**).

Tabla 2. Hallazgos imagenológicos de los pacientes con estudio citológico de tiroides Bethesda IV para el manejo quirúrgico de acuerdo con el resultado histológico

Característica	Total n = 427	Cáncer de tiroides n = 195	Sin cáncer n = 232	Valor de p
Microcalcificaciones, n (%)	178 (41,6)	145 (74,3)	33 (14,2)	< 0,01
Ovalado, n (%)	120 (28,1)	44 (22,5)	76 (32,7)	0,02
Hipoeoico, n (%)	223 (52,2)	137 (70,2)	86 (37)	< 0,01
Bordes irregulares, n (%)	56 (13,1)	48 (24,6)	8 (3,4)	< 0,01
Nódulo único, n (%)	147 (34,4)	69 (35,3)	78 (33,6)	0,7
Nódulo hipervascularizado, n (%)	89 (20,8)	74 (38)	15 (6,4)	< 0,01
TIRADS, n (%)				
2	128 (29,9)	2 (1)	126 (54,3)	< 0,01
3	110 (25,7)	23 (11,7)	87 (37,5)	
4	123 (28,8)	107 (54,8)	16 (7)	
5	58 (13,5)	55 (28,2)	3 (1,3)	
6	8 (1,8)	8 (4,1)	0 (0)	
Adenopatías presentes, n (%)	7 (1,6)	7 (3,6)	0 (0)	< 0,01

TIRADS: thyroid imaging reporting and data system

Tabla 3. Hallazgos en los estudios citológicos de tiroides por Bethesda IV para el manejo quirúrgico, de acuerdo con el resultado histológico

Característica	Total n = 427	Cáncer de tiroides n = 195	Sin cáncer n = 232	Valor de p
Proliferación folicular abundante, n (%)	262 (61,3)	127 (65,1)	135 (58,1)	0,14
Patrón microfolicular, n (%)	69 (16,1)	44 (22,5)	25 (10,7)	< 0,01
Escaso material coloide, n (%)	62 (14,5)	28 (14,3)	34 (14,6)	0,93
Celularidad, n (%)				
Abundante	139 (32,5)	66 (33,8)	73 (31,4)	0,32
Escasa	255 (59,7)	118 (60,5)	137 (59)	
Ausente	33 (7,7)	11(5,6)	22 (9,4)	
Núcleos grandes, n (%)	91 (21,3)	77 (39,4)	14 (6)	< 0,01
Aclaramiento de la cromatina, n (%)	23 (5,3)	16 (8,2)	7 (3)	0,01
Pseudoinclusiones nucleares, n (%)	71 (16,3)	69 (35,3)	2 (0,8)	< 0,01
Presencia de necrosis, n (%)	0	0	0	NA

NA: no aplica.

En el análisis de las técnicas de minería se evidenciaron mejores resultados usando la validación cruzada (10 fold), comparada con la partición (66 %). La técnica de Bayes tuvo

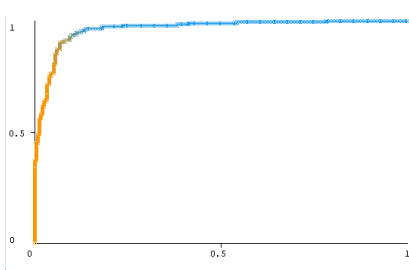
mejores resultados, dado por un área bajo la curva de característica operativa del recepto AUC-ROC de 0,964 (Tabla 4) y de clasificación con exactitud de 91,1 % (Figura 2).

Tabla 4. Resultados del análisis			
Estadísticas de exactitud	Modelos de clasificación		
	Prueba de validación cruzada 10 folds		
	Naive Bayes	Árbol (J48)	Red neuronal
Área ROC	0,964	0,878	0,948
Medida F	0,911	0,878	0,88
Exactitud	91,1 %	87,8 %	88,2 %
Estadísticas de exactitud	Modelos de clasificación Porcentaje de partición de 66 %		
	Naive Bayes	Árbol	Red neuronal
Área ROC	0,946	0,902	0,91
Medida F	0,89	0,857	0,876
Exactitud	88,9 %	85,5 %	87,5 %

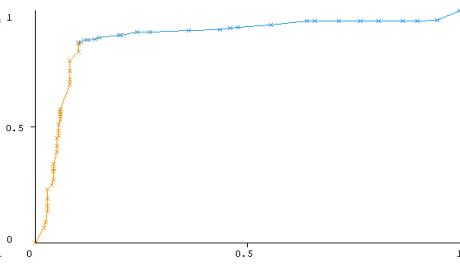
A. Modelo Naive Bayes

B. Modelo árbol

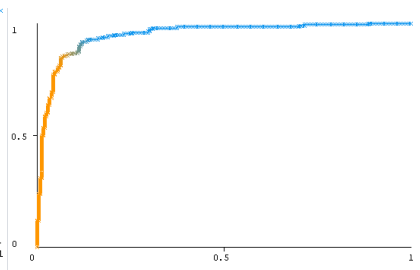
C. Redes neuronales



AUC: 0,96



AUC: 0,87



AUC: 0,94

Eje X: falsos positivos, eje Y: verdaderos positivos.

Figura 2. Área bajo la curva del modelo.

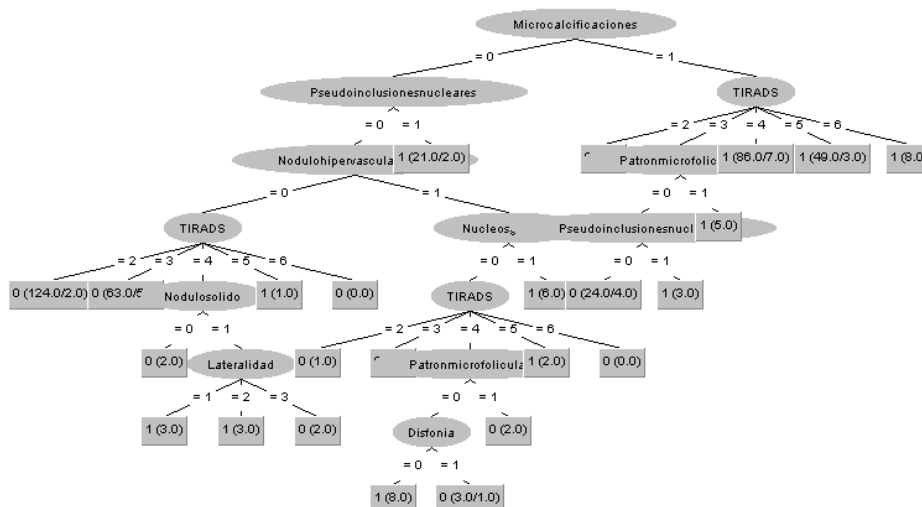


Figura 3. Árbol de decisión.

Con base en los anteriores resultados fue posible construir un árbol de decisión, de acuerdo con el hallazgo de las características identificadas por el modelo de Naive Bayes, para determinar la probabilidad de malignidad (**Figura 3**) de los pacientes con nódulo tiroideo por Bethesda IV.

Discusión

Estudios realizados en grandes poblaciones han documentado una prevalencia de nódulos palpables del 5 % en mujeres y del 1 % en hombres en áreas yodo-suficientes (17). Esta prevalencia asciende hasta el 68 %, cuando se tienen en cuenta los hallazgos ecográficos de pacientes que usualmente son asintomáticos (18, 19).

El cáncer de tiroides tiene un extenso espectro de presentaciones, las cuales abarcan desde lesiones foliculares benignas, hasta el cáncer anaplásico (20). La tiroidectomía parcial o total continúa siendo el pilar de tratamiento de los diferentes tipos de cáncer de tiroides (21). Los pacientes incluidos en el presente estudio se sometieron a un manejo quirúrgico con tiroidectomía parcial o total.

El sistema que se utiliza con mayor frecuencia para la descripción y clasificación de los hallazgos citológicos de los nódulos tiroideos (obtenidas por PAAF) es el sistema Bethesda, el cual se sustenta en 6 categorías, que permiten estimar el riesgo de malignidad.

Con base en los resultados de la ecografía y el PAAF, hasta 20 %-50 % de los pacientes con nódulos tiroideos se lleva a tiroidectomía; de los cuales, existe un grupo de pacientes con resultados falsos positivos que se lleva a tiroidectomía, que asumen innecesariamente el riesgo de presentar posibles complicaciones relacionadas con el procedimiento quirúrgico (sangrado, hipoparatiroidismo temporal o permanente, alteraciones de la voz, entre otros).

En un estudio de casos y controles realizado en el Hospital de San José, Bogotá, se planteó la evaluación de los factores asociados con la histopatología final benigna en los pacientes llevados a tiroidectomía, clasificados como Bethesda IV, V, VI, mediante PAAF. Se incluyeron 86 casos y 86 controles. En los casos de la categoría Bethesda IV se presentó el mayor número de falsos positivos (89 %). Como conclusión, este estudio sugirió el planteamiento de subclasificaciones citológicas en la categoría Bethesda IV, para minimizar el número de cirugías innecesarias. Los falsos positivos del Bethesda continúan siendo un reto, por lo que las pruebas moleculares y una segunda revisión cobran importancia (22).

En los últimos años se han venido desarrollando estudios que buscan identificar características clínicas, imagenológicas, citológicas y moleculares, que permitan precisar asertivamente los pacientes que requieren de manejo quirúrgico por patología maligna.

En un estudio retrospectivo publicado en 2016, Park y colaboradores evaluaron el rendimiento de los hallazgos ecográficos en un grupo de 35 pacientes, con reporte de ci-

tología Bethesda IV, y se identificó una sensibilidad del 50 % y un bajo VPP (50 %), comparado con otras categorías Bethesda, con una especificidad del 97 % y un VPN del 97 % (23). En este estudio, los datos mostrados identificaron las características ecográficas de presencia de microcalcificaciones e hipervascularización, como parte importante del árbol de decisiones, con el uso de la técnica de Naive Bayes de la minería de datos.

Con base en las anteriores consideraciones es posible afirmar que el hallazgo citológico Bethesda IV en pacientes que se encuentran en estudio de un nódulo tiroideo constituyen un reto para los médicos especialistas, que tendrán que tomar una decisión acertada de manejo quirúrgico debido al riesgo de malignidad, teniendo en cuenta que muchos pacientes podrían ser llevados a un manejo quirúrgico innecesario, con un riesgo de complicaciones que, a largo plazo, podría tener un impacto negativo en su calidad de vida.

En el presente estudio, el porcentaje de malignidad de los pacientes con estudio citológico Bethesda IV fue del 45,6 %, relativamente más alto en comparación con lo reportado en las guías de la ATA, publicadas en el 2015 (10).

El modelo previo de esta investigación identificó, mediante regresión logística en el análisis multivariado, 11 variables con características de potenciales factores predictivos de malignidad, como la presencia de nódulo sólido al examen físico (OR 2,5; IC 95% 1,03 a 6,5), de microcalcificaciones (OR 6,6; IC 95% 2,6 a 16,5), de hipervascularización en el estudio Doppler (OR 4,5; IC 95% 1,4 a 14,7), clasificación TIRADS, de patrón microfolicular, entre otras. Dicho modelo clasifica correctamente el 91,57 % de los datos, con una sensibilidad del 90,7 % y una especificidad del 92,2 %.

En nuestro medio se hace necesario la realización de estudios que, en términos de costo efectividad, permitan predecir la malignidad en pacientes con nódulo tiroideo, que tienen estudio citológico Bethesda IV. La realización de pruebas moleculares constituye una herramienta útil para determinar, con mayor precisión, la probabilidad de cáncer en este grupo de pacientes; sin embargo, son costosas, no se encuentran disponibles en todas las instituciones prestadoras de salud de nuestro país y ninguna tiene un VPP del 100 %.

El uso de la minería de datos utilizado en el presente estudio tuvo la finalidad de extraer la información más relevante de manera automatizada, para reconocer patrones (características identificadas preoperatoriamente) en los pacientes con cáncer de tiroides.

Las técnicas de minería de datos usaron 27 variables con resultados similares. Se evidencia que los mejores resultados de la clasificación se obtienen con la técnica de Naive Bayes, seguido por la de redes neuronales. La literatura reporta diferentes técnicas de minería de datos en el análisis de variables relacionadas con enfermedades tiroideas, sin embargo, no se ha reportado específicamente para pacientes con estudio citológico Bethesda IV (24).

Los métodos de Naive Bayes representan un conjunto de algoritmos de aprendizaje supervisado, basados en la apli-

cación del teorema de Bayes, con el supuesto “ingenuo” de independencia entre cada par de características, este clasificador es rápido de entrenar (25).

Este es un importante acercamiento a la predicción del riesgo de cáncer de tiroides de los pacientes con citología Bethesda IV en la población colombiana, utilizando características preoperatorias (clínicas, imagenológicas y citológicas), que no suponen un elevado costo de atención en salud (comparado con el uso de las pruebas moleculares); y podrían constituirse como una herramienta de bajo costo para identificar adecuadamente a los pacientes con cáncer de tiroides, para ofrecerles un manejo quirúrgico óptimo y disminuir la probabilidad de someter a los pacientes a procedimientos innecesario. Es necesario tener en cuenta que puede reportarse variabilidad en la literatura debido a las características particulares de cada población, por tanto, podrían tener valores de exactitud diferentes (25).

Conclusión

El uso de un modelo de predicción de malignidad en los pacientes con estudio citológico Bethesda IV permite orientar, de forma acertada, la decisión de la intervención quirúrgica, en razón a que se llevarían a cirugía aquellos pacientes con alta probabilidad de tener cáncer de tiroides. Los resultados con las técnicas de minería de datos son concordantes con los hallazgos de la regresión logística, y podrían constituirse como una herramienta costo efectiva en países en los que la realización de pruebas moleculares supone costos elevados, sin tener un VPP del 100 %.

Este estudio tuvo como limitaciones la pérdida de 70 pacientes debido a la ausencia de más del 50 % de la información en la historia clínica. Los estudios ultrasonográficos se realizaron en diversas instituciones, incluida la nuestra, y al ser un estudio imagenológico operador dependiente no obtuvimos información en relación con las características de los equipos utilizados, ni de los parámetros tomados en cuenta para catalogar los hallazgos dentro de la clasificación TIRADS. Otra limitación se relacionó con los reportes de citología, debido a que no se realizaron en su totalidad en nuestra institución, por lo que en algunos pacientes los datos de las características descriptivas del estudio no estaban completos.

Declaración de conflicto de intereses

Los autores del presente manuscrito no presentan ningún conflicto de interés.

Fuente de financiación

Este proyecto de investigación no contó con financiación externa a los participantes, autores y coautores descritos en este trabajo.

Agradecimientos

Agradecimientos especiales al departamento de Investigación, de la Fundación Universitaria Ciencias de la salud.

Contribución de cada autor

Elizabeth León Guzmán: análisis de datos. Rogers Leonardo Baquero García: recolección y análisis de datos, resultados, discusión, correcciones editoriales. Esteban Diazgranados Goenaga: supervisión del proyecto, discusión. Álvaro Eduardo Calixto Granados: supervisión del proyecto, discusión. Juan de Francisco Zambrano: supervisión del proyecto, discusión. Andrés Felipe Rey: recolección de datos. César Alfonso Palencia: recolección de datos, introducción. Juan Fernando Castañeda: recolección de datos, introducción. Erika León Guzmán: metodología

REFERENCIAS

1. Kumar R, Shaikh B, Chandio A, Ahmed J. Role of health management information system in disease reporting at a rural district of Sindh. *Pak J Public Health*. 2012;2(2):10-2.
2. Castro M. Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales. *Rev Esp Pedagog*. 2012;251:131-48.
3. Stilou S, Bamidis PD, Maglaveras N, Pappas C. Mining association rules from clinical databases: An intelligent diagnostic process in healthcare. *Stud Health Technol Inform*. 2001;84(2):1399-403.
4. Salazar J, Espinoza C, Mendiola A, Bermúdez V. Data mining and endocrine diseases: A new way to classify? *Arch Med Res*. 2018;49(3):213-5. doi: 10.1016/j.arcmed.2018.08.005.
5. Kim T, Zabaneh F, Holmes J, Disrude L, Price M, Gentry L. A practical data mining method to link hospital microbiology and an infection control database. *Am J Infect Control*. 2008;36(3):S18-S20. doi: 10.1016/j.ajic.2007.05.010.
6. Izzo CM, Prescott J, Peden A, Brown JF, Valencia F. Use of a data mining system to improve the process of public health reporting in a large teaching facility. *Am J Infect Control*. 2008;36(5):E186. doi: 10.1016/j.ajic.2008.04.216.
7. Michie D, Spiegelhalter DJ, Taylor CC, Campbell J (editores). *Machine learning, neural and statistical classification*. Upper Saddle River, NJ, EE. UU.: Ellis Horwood. 1994.
8. Koh HC, Tan G. Data mining applications in healthcare. *J Healthc Inf Manag*. 2005;19(2):64-72.
9. Kotsiantis S, Pierrakeas C, Pintelas P. Preventing student dropout in distance learning using machine learning techniques. Springer. 2003. p. 267-74.
10. Schreiner AM, Yang GC. Adenomatoid nodules are the main cause for discrepant histology in 234 thyroid fine-needle aspirates reported as follicular neoplasm. *Diagn Cytopathol*. 2012;40(5):375-9. doi: 10.1002/dc.21499.
11. Conzo G, Calò PG, Gambardella C, Tartaglia E, Mauriello C, Della Pietra C, et al. Controversies in the surgical management of thyroid follicular neoplasms. Retrospective analysis of 721 patients. *Int J Surg*. 2014;12(1):S29-34. doi: 10.1016/j.ijssu.2014.05.013.

12. Herrera F, Redondo K, Osorio C, Grice J, Fernández A. Utilidad de la citología obtenida mediante aspiración con aguja fina en el diagnóstico de las neoplasias foliculares de la glándula tiroides en la E.S.E Hospital Universitario del Caribe: un estudio retrospectivo. *Rev Colomb Cir.* 2015;30:12-7. doi: 10.30944/issn.2011-7582.
13. Jo VY, Stelow EB, Dustin SM, Hanley KZ. Malignancy risk for fine-needle aspiration of thyroid lesions according to the Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol.* 2010;134(3):450-6. doi: 10.1309/AJCP5N4MTHPAFXFB.
14. Osorio C, Fernández A, Herrera K, Marrugo Á, Ensuncho C, Redondo K, et al. Sensibilidad y especificidad de la citología obtenida mediante aspiración con aguja fina en el diagnóstico de las neoplasias foliculares de la glándula tiroides: un estudio prospectivo. *Rev Esp Patol.* 2016;49(3):144-50. doi: 10.1016/j.patol.2016.01.005.
15. Kuru B, Kefeli M. Risk factors associated with malignancy and with triage to surgery in thyroid nodules classified as Bethesda category IV (FN/SFN). *Diagn Cytopathol.* 2018;46(6):489-94. doi: 10.1002/dc.23923.
16. Lee SH, Baek JS, Lee JY, Lim JA, Cho SY, Lee TH, et al. Predictive factors of malignancy in thyroid nodules with a cytological diagnosis of follicular neoplasm. *Endocr Pathol.* 2013;24(4):177-83. doi: 10.1007/s12022-013-9263-x.
17. Yassa L, Cibas ES, Benson CB, Frates MC, Doubilet PM, Gawande AA, et al. Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer.* 2007;111(6):508-16. doi: 10.1002/cncr.23116.
18. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest.* 2009;39(8):699-706. doi: 10.1111/j.1365-2362.2009.02162.
19. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules--2016 update. *Endocr Pract.* 2016;22(5):622-39. doi: 10.4158/EP161208.GL.
20. Abdul-Jabar HB, Lynn J. The surgical management of thyroid cancer. *Nucl Med Commun.* 2004;25(9):869-72. doi: 10.1097/00006231-200409000-00002.
21. Rodríguez González H, Pava Marín R, Castaño Herrera LF, Valencia García LV, Pava Ripoll A. Evaluación de la precisión diagnóstica de la punción aspiración con aguja fina en pacientes con nódulo tiroideo. *Biosalud.* 2017;16(1):19-29. doi: 10.17151/biosa.2017.16.1.3.
22. Mejía M, Rivera Delgado M, Bonilla J, Melo M, Rojas García W. Factores asociados a histopatología final benigna en pacientes con biopsia por aspiración con aguja fina clasificados como Bethesda en IV-V-VI en Hospital de San José. Bogotá: Fundación Universitaria de Ciencias de la Salud. 2017.
23. Park SY, Hahn SY, Shin JH, Ko EY, Oh YL. The diagnostic performance of thyroid US in each category of the Bethesda system for reporting thyroid cytopathology. *PLoS One.* 2016;11(6):e0155898. doi: 10.1371/journal.pone.0155898.
24. Razia S, Narasinga MR. Machine learning techniques for thyroid disease diagnosis - A review. *Indian J Sci Technol.* 2016;9(28):1-9. doi: 10.17485/ijst/2016/v9i28/93705.
25. Boas FE, Liu LY, Kamaya A, Desser TS, Rubin DL. Estimating the probability of malignancy of thyroid nodules using adjusted naïve Bayes. 2012. Disponible en: <http://www.edboas.com/calc/thyroid.pdf>