

# Actualización del asistente de evaluación clínica para la Estratificación de Riesgo Inferido Cardiovascular basado en redes bayesianas

Iris Sattolo<sup>1</sup>, Blas Cappello<sup>2</sup>, Jorge Ierache<sup>1</sup>,  
Instituto de Sistemas Inteligentes y Enseñanza Experimental de la Robótica FICCTE-UM<sup>1</sup>  
Experto en Medicina Cardiovascular<sup>2</sup>  
Laboratorio de Sistemas de Información Avanzados Facultad de Ingeniería U.B.A.<sup>3</sup>  
Facultad de Informática Ciencias de la Comunicación y Técnicas Especiales<sup>1</sup>  
Universidad de Morón, Cabildo 134, (B1708JPD) Morón, Buenos Aires, Argentina  
54 11 5627 2000 int 189  
iris.sattolo@gmail.com,cappelloblas@gmail.com.ar

**Abstract:** En este artículo se describe el trabajo de investigación que en la actualidad se está desarrollando dentro del área de inteligencia artificial aplicada a un dominio específico, la ciencia médica. Las enfermedades cardiovasculares constituyen un serio problema epidemiológico en el mundo, modificar esa tendencia es el objetivo de la prevención. Por tal motivo identificar los factores de riesgo sobre los cuales conviene enfocar la atención es la clave para poder intervenir en la manifestación de los eventos. Nuestro trabajo pretende contribuir con la estratificación de pacientes en un determinado riesgo, ofreciendo una herramienta que aplica la estructura del teorema de Bayes en una red causal, combinándola con la experiencia del experto.

**Keywords:** Biomedicina, Factores de Riesgo Cardiovascular, Redes Bayesianas, Aprendizaje Automático, Sistemas Basados en Conocimiento

## 1 Introducción

En este artículo se extiende nuestro trabajo realizado en el de estratificación de riesgo cardiovascular aplicando redes bayesianas [1], el cual se planteó para la prevención primaria cardiovascular, esto es, para personas que no posean ninguna enfermedad. Ampliamos el alcance del mismo al incorporar factores de riesgo que no fueron considerados en su momento, y que figuran como factores de riesgo en las tablas adoptadas por las distintas sociedades médicas. Incorporamos en esta oportunidad los factores Hipertensión arterial, Diabetes y edad de la persona.

Las enfermedades cardiovasculares (ECV) son la principal causa de muerte en todo el mundo[2]. Cada año mueren más personas por ECV que por cualquier otra causa. Según la Organización Mundial de la Salud, se calcula que en 2004 murieron por esta causa 17,3 millones de personas, lo cual representa un 30% de todas las muertes registradas en el mundo y para el 2030 morirán cerca de 23,6 millones. Las ECV representan una gran carga para la economía de los países [3]. En este momento la OMS cataloga a las ECV dentro de las enfermedades crónicas y como tal propone la prevención y control de las mismas.

Existen numerosas guías y recomendaciones de modelos de predicción de riesgo para las ECV. Las tablas cuantitativas son aquellas que, obtienen un número equivalente a la probabilidad de presentar una enfermedad cardiovascular en un periodo de tiempo, generalmente 10 años. Las tablas cualitativas consisten en la suma de factores de riesgo cardiovasculares que se presentan como variables dicotómicas (presente o ausente). En nuestro trabajo anterior construimos un asistente de evaluación clínica para estratificar el riesgo de padecer una ECV, basado en redes bayesianas. La propuesta de una nueva manera de estratificar riesgos aplicando la estructura del Teorema de Bayes en una red causal, surgió como novedad en el campo médico, constituyendo un aporte original en el tema de la estratificación del riesgo. Así fue que obtuvimos ERICA (Versión 1). El mismo fue planteado para clasificar a personas que no tenían hipertensión arterial, ni diabetes, y para personas de un rango de 20 años entre los 35 y 55 años. Debido a que tanto la hipertensión como la diabetes son enfermedades que pueden encontrarse presentes en el desencadenamiento de algún evento cardiovascular, hemos considerado en esta oportunidad, a éstas como Factores de Riesgo.

En la Tabla 1 se presenta un cuadro comparativo de los factores de Riesgo adoptados por las tablas propuestas por las más prestigiosas sociedades que estudian el tema. En las dos columnas finales, mostramos los factores de riesgo que se utilizaron para la construcción de ERICA (v1 y v2). Donde PAS: Presión arterial Sistólica, AHF:Antecedentes Heredo Familiares, DA: Dietas aterogénicas, AS: Aterosclerosis Subclínica, CCA: Condiciones clínica Asociadas, LOD: Lesión Órgano Diana, PCR: Proteína C Reactiva, X: Tabaquismo actual y pasado. Las Tablas ATP III, Framingham, SCORE, American Heart Association (AHA) son cuantitativas, a diferencia de la Sociedad europea 2003 que es cualitativa.

**Tabla 1.** Comparación de factores de riesgo de tablas y ERICA.

Tabla								
Factores	ATP III	Framingham	SCORE E.	Sociedad E. 2003	Guía SHAPE [4]	AHA	ERICA V1	ERICA V2
Tabaquismo	X	X	X		X	X	X	X
P.A.S	X	X	X	X	X	X	no	X
LDL alto				X	X		X	X
HDL < 40 mg	X	X		X		X	X	X
Colesterol Total	X	X	X	X	X	X		
AHF		X					X	X
Obesidad				X	X		X	X
Sedentarismo					X		X	X
DA							X	X
EMI/Placa					X		X	X
Calcio C.					X			
AS							X	X
Sexo	X	X	X		X	X	X	X
Edad	X	X	X		X	X	X	X

CCA		X		X			no	no
Diabetes	no	X	X	X		X	no	X
LOD				X			no	no
PCR				X				
Glucemia				X			X	X

## 2 Desarrollo

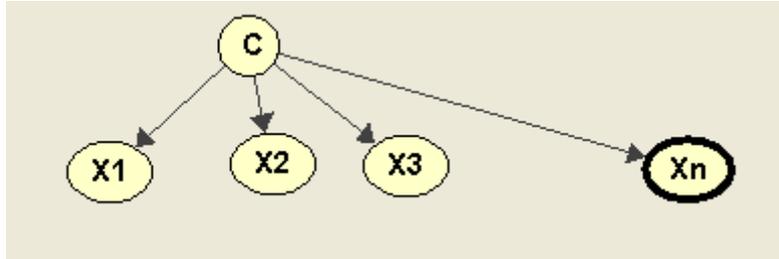
El objetivo fue obtener una red bayesiana con los nuevos factores propuestos, sin que estos modifiquen sustancialmente la red anterior, comparar el funcionamiento de ambas redes, y probar el funcionamiento de clasificador Naïve Bayes con los datos aportados de 120 historias clínicas.

Las redes bayesianas constituyen un formalismo que permite representar de manera compacta y eficiente un modelo probabilístico sobre un dominio concreto, se componen de dos partes, por un lado, la estructura, el modelo o parte **cuantitativa**: un grafo dirigido a cíclico (DGA), donde cada nodo representa una variable aleatoria y los arcos representan dependencias probabilísticas entre variables. Pero por otra, también se componen de una distribución condicional de probabilidades, esta parte de la red bayesiana se conoce como la parte paramétrica o **cuantitativa** de la red [5].

Existen dos formas de justificar los enlaces que introducimos u omitimos al construir nuestra red. La primera es de naturaleza teórica: formamos un modelo causal a partir de la experiencia de un especialista y trazamos los arcos correspondientes al modelo. ERICA fue concebida desde la experiencia del especialista, para el armado de la red causal y la obtención de los parámetros, los que debieron corregirse a medida que el sistema se puso en funcionamiento. Este proceso, conocido como *aprendizaje secuencial*, es en realidad un ajuste de parámetros.

El otro camino para justificar la red consiste en realizar una comprobación empírica a partir de un conjunto suficientemente amplio de casos, utilizando las herramientas estadísticas que se emplean para detectar correlaciones. Implementamos este método para crear una red y compara los resultados obtenidos. Utilizamos el clasificador Naïves Bayes, el cual predice una variable clase, dada cualquier configuración en el resto de las variables.

El modelo de este clasificador asume la independencia de los atributos  $x_1, x_2, \dots, x_n$  dada la clase  $C$ . Aunque esta asunción no se cumple la mayoría de las veces, generalmente, su buen comportamiento se debe a la exactitud con la que las distribuciones condicionales pueden calcularse, incluso con muestras relativamente pequeñas dado que el conjunto de padres de todas las variables de entrada  $x_1, x_2, \dots, x_n$  se limita tan solo a la clase:  $\Pi_{x_i} = \{C\}$ ,  $i=1..n$ . (Figura 1)

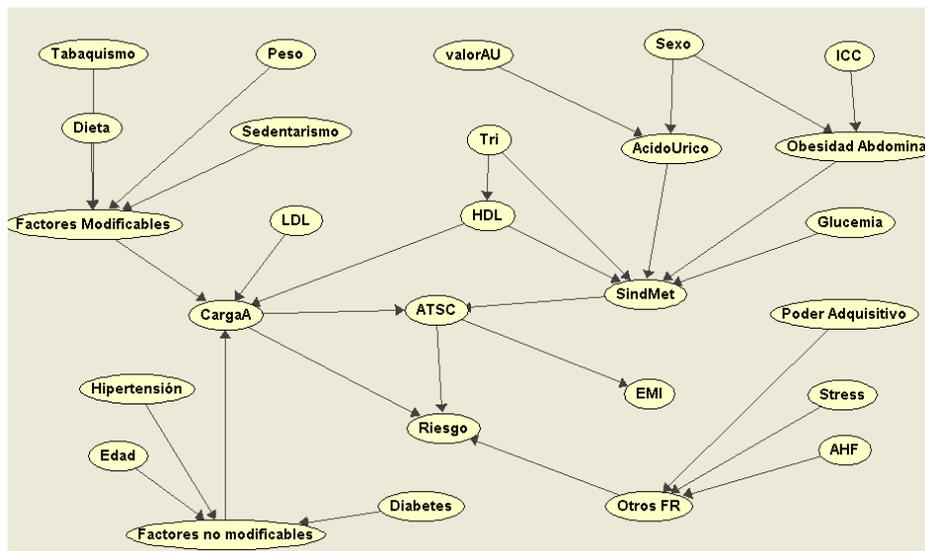


**Figura 1** “Clasificador Naives Bayes”

ERICA V1 Y ERICA V2 se construyeron con el programa ELVIRA [6]. El programa Elvira es fruto de un proyecto de investigación financiado por la CICYT y el Ministerio de Ciencia y Tecnología Español, en el que participan investigadores de varias universidades españolas y de otros centros. Este programa permite el ingreso de las redes Bayesianas de dos formas: (a) por un lado el ingreso manual, donde el usuario dibuja la red bayesiana en la pantalla y carga los valores de probabilidad asociados a cada nodo y (b) mediante la importación de archivos de casos.

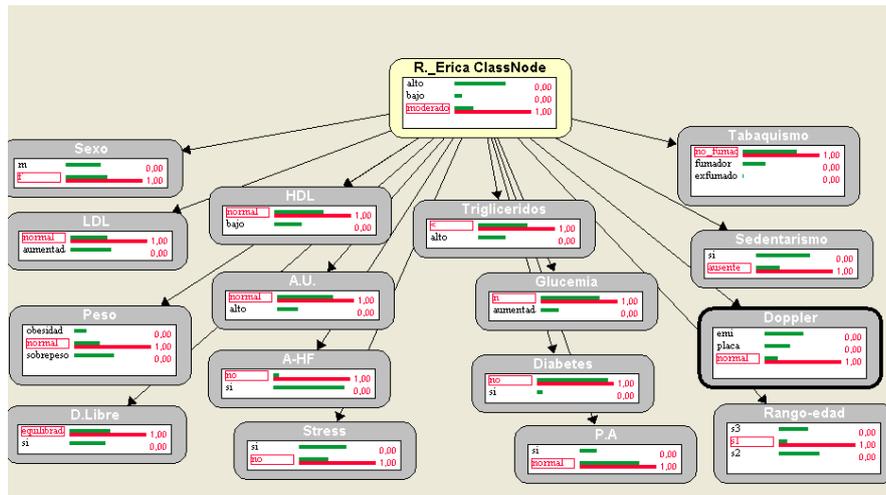
### 3 Solución propuesta

En la nueva red propuesta (figura2) consideramos Hipertensión, Edad y Diabetes [7], [8], [9] como factores de riesgo no modificables, se actualizaron los parámetros del nodo Carga Aterogénica el cual cuenta con 256 valores., se trabajó con el nodo factores no modificables como nuevo nodo intermedio de la red.



**Figura 2:** Red Bayesiana para estratificación de Riesgo Cardiovascular ERICA v2

Para armar la red con el Clasificador Naives Bayes se realizó la importación de datos desde el programa ELVIRA, donde se utilizó el algoritmo de imputación Zeros, ya que en el Nodo Doppler no contaba con todos los datos, para luego utilizar el clasificador Naives Bayes obteniéndose la red de figura 3 en la cual se muestra un caso donde el paciente no tiene factores de riesgo.



**Figura 3:** modo inferencia mostrado por el programa Elvira de la clasificación realizada por Naives Bayes

En relación a los datos utilizados para la verificación y validación de la red, se tomaron dos grupos de pacientes, el Grupo A 50 pacientes, todos con el estudio Doppler de los vasos del cuello realizados, edad entre 30 y 60 años sin hipertensión y sin diabetes, y el grupo B, 70 pacientes que pueden tener o no el estudio realizado, edad todos los rangos y, personas con o sin hipertensión y diabetes.

El grupo C (120) resulta de la suma de A y B y se utilizó para construir el clasificador. En el otro grupo de casos, grupo control, se utilizaron los mismos 46 casos que probamos en ERICA v1

**Tabla 2.** Datos utilizados para la verificación y validación de ERICA v2

CASOS	CANTIDAD	ERICA V1	ERICA V2	CLASIFICADOR
CONTROL	46 CASOS	SI	SI	SI
A	50 DATOS	SI	SI	
B	70 DATOS		SI	
C(A+B)	120 DATOS			SI

## 4 Resultados

Con las pruebas realizadas con el clasificador, los mismos no arrojaron los resultados esperados, en los casos que el factor estudiado tiene como tabla de probabilidad condicionada valores menores al 19 %. Son los nodos de Antecedente heredo familiar, Stress y Tabaquismo (ex fumador). Los valores que se tomaron para estos factores en ERICA v2 fueron, los de la prevalencia en la población Argentina, este es el mecanismo que se debe realizar con los nodos padres de las redes bayesianas. Esto podría confirmar lo antedicho cuando evaluamos ERICA v1 que “como ventaja de las redes Bayesianas es el poder transpolar el modelo a distintas poblaciones solo cambiando la prevalencia de las distintas variables que se toman en el trabajo.” [1]. Mostramos en la tabla 3 las diferencias encontradas entre el Clasificador Naives Bayes, el cual toma los datos proporcionados y clasifica según la población de muestra, en cambio ERICA (v1- v2) toma la prevalencia del factor de riesgo o enfermedad en la población, de cada nodo que no tenga ascendientes.

**Tabla 3:** Comparación de los Nodos AHF, Stress, Tabaquismo de ERICA v2 y Red obtenida por Naives Bayes.

Factor	ERICA v2	Clasificador
Antecedente HF	presente 54 % Ausente 46 %	Presente (si) 88 % Ausente (no) 12 %
Stress	Presente 70 % Ausente 30 %	Presente (si) 84 % Ausente (no) 16 %
Tabaquismo	Ex-fumador 16 % No fumador 46 % Fumador 39 %	Ex_fumador 2% No Fumador 84 % Fumador 14 %

Desarrollamos los casos de pruebas correspondientes a cada nodo, en este orden se presenta en la tabla 4 los resultados obtenidos de un caso de prueba del nodo riesgo, que contempla como resultados: (1) el comportamiento con los valores en la red sin el valor arrojado por el estudio Doppler, el cual mide el espesor mio-intimal de las paredes de la arteria carótida, EMI (Espesor mio-intimal) y (2) considerando el valor de EMI en la red de ERICA v2.

**Tabla 4: Caso de prueba** “Resultados del nodo riesgo con y sin el valor de EMI, aplicandos a la red ERICA v1, y ERICA v2

Elementos a probar: NODO RIESGO, NODO ATSC (Aterosclerosis Sub-Clínica)				
Entradas		Resultados		
Variable	Valor	Erica v1 sin doppler (1)	Erica v2 sin doppler (1)	ERICA v2 con doppler (2)
Sexo	Masculino			
LDL	aumentado			

HDH	bajo			
Triglicérido	alto			
A.U.	alto			
Glucemia	alta			
Tabaquismo	No fumador			
Sedentarismo	ausente			
Peso	normal			
Dieta	Equilibrada			
A.H.F.	no			
Stress	si			
P.A.	normal			
Edad	2			
Diabetes	no			
Doppler	emi			
Síndrome M		presente	presente	
ATSC		Avanzada 78 % Intermedio 19%	Avanzada 62% Intermedio 33%	Intermedia 97 %
Riesgo		Alto 72% Moderado 24 %	Alto 59% Moderado 33%	Moderado 84%

Sobre la base del caso de prueba presentado, se muestra en la figura 4 los valores en el modo inferencia de la red Causal propuesta en ERICA v2, empleando la herramienta ELVIRA. Los nodos coloreados en gris, son a los cuales se les introdujo los valores propuestos en nuestro caso de prueba. Los Nodos Riesgo y ATSC son los que inferimos. El nodo de Aterosclerosis sub-clínica (ATSC) solo se certifica una vez ingresado el valor del EMI. El nodo Síndrome M (síndrome metabólico) lo infiere la red. Los nodos Hipertensión, Diabetes, Edad y Factores no modificables, son los incorporados en esta nueva versión de ERICA.

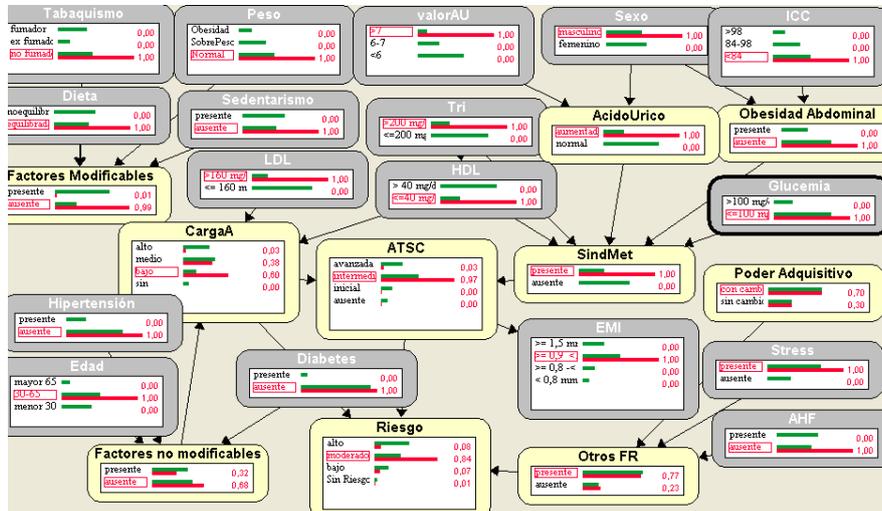


Figura 4. Pantalla arrojada en ERICA v2, sobre el caso de prueba de tabla 5

En relación a la comparación hecha anteriormente en ERICA v1 sobre los 50 casos en los cuales teníamos el valor del ECO Doppler realizado, y estos casos clasificados según la tabla propuesta por el American Heart Association (A.H.A) estudio SHAPE [4] sin introducir el valor del Eco Doppler Carotideo, el cual se utilizó como parámetro para certificar los resultados, realizamos la misma tarea con ERICA v2. y evaluamos los resultados, los cuales se muestran en la Tabla 5.

Tabla N° 5 Comparación de resultados A.H.A y ERICA v1 y ERIVA v2

		A.H.A			ERICA 1			ERICA2		
		B	M	A	B	M	A	B	M	A
DOPLER										
Normal	15	5	9	0	5	5	5	4	11	0
Aumentado	8	5	3	0	1	4	3	0	8	0
Placa	27	15	10	3	0	4	23	0	2	25
total	50	25	22	3	6	13	31	5	20	25

De los 15 pacientes con el estudio Doppler Normal, en los cuales ERICA v1 había clasificado en Bajo 5, Medio 5 y Alto 5, en ERICA v2 (sin introducir el valor EMI), re-estratificó en Moderado 11 y Bajo 4. Resulta interesante la disminución de los 5 casos en el valor alto. En los 8 casos con Doppler Aumentado, re-estratificó los 3 casos que fueran considerados altos en Erica v1 y el caso que fuera considerado bajo en moderado. De los 27 pacientes de los cuales el estudio Doppler dio como resultado el tener Placa Fibrolipídica, solo 3 fueron clasificados por el American Heart Association como de alto riesgo, contra 23 clasificados como de alto riesgo por ERICA. v1, y 25 por Erica v2. Sobre la base de estos casos, se visualiza una

mejor inferencia en la red, en el nodo Aterosclerosis sub-clínica y nodo riesgo. Al introducir el nodo del valor de Doppler, re-estratifica correctamente en todo los casos.

## 4 Conclusiones y Futuras Líneas de Investigación

Las redes bayesianas, modelos que combinan la teoría de grafos y de probabilidad, son aplicadas a la toma de decisiones en dominios donde la incertidumbre representa un papel importante, tal es el caso de la predicción de riesgos de contraer una determinada enfermedad cardiovascular.

En cuanto a la evaluación de la red propuesta por el Clasificador Naives Bayes con los casos reales no obtuvimos buenos resultados, pero esto se debió al hecho de haber realizado la clasificación con los datos tomados de un consultorio cardiológico, donde los casos son similares. No es una buena muestra de la población en general, si es una muestra para validar los resultados de casos reales.

En cuanto a la incorporación de nuevas variables, el resultado fue satisfactorio, se pudo modificar la red sin tener que variar las probabilidades generadas con ERICA v1. Solo se cambiaron las probabilidades condicionadas del Nodo Aterosclerosis Sub-Clínica.

Como futuras líneas de trabajo se plantea seguir refinando los nodos de probabilidades condicionadas revisando la bibliografía y consultando al experto, el cual está trabajando en un protocolo de control de pacientes con ERICA v2. Incorporar la red a un diagrama de influencia agregando los nodos de tratamiento y utilidad para mayor funcionalidad que consideren conductas terapéuticas, optimización de costos mediante la orientación en el tratamiento, y equilibrar la balanza costo / beneficio para los sistemas de salud.

## 5 Bibliografía

1. Ierache, J., Sattolo I., Cappello B, (2011) "Asistente de evaluación Clínica para la estratificación de riesgo Cardiovascular basado en redes Bayesianas" XVII Congreso Argentino de Ciencias de la Computación 5 al 9 octubre CACIC 2011, Universidad Nacional de la Plata, Facultad de Ingeniería
2. Khut y col. Prevalence of Convencional Risk Factors in Patients with Coronary Heart Decease. JAMA 2003; 290: 898-904
3. Revista argentina de cardiología *versión On-line* ISSN 1850-3748
4. Castillo, Gutiérrez y Hadi. Sistemas Expertos y Modelos de Redes Probabilísticas
5. Revista de la Facultad de Ciencias Médicas Universidad Nacional de Rosario. Maestría en Aterosclerosis. Volume 4 n 2 Módulo 15 Capítulo 21.
6. <http://www.ia.uned.es/~elvira>.
7. <http://materias.fi.uba.ar/7550/clasificadores-bayesianos.pdf> consultado 10/6/2012
8. [http://med.unne.edu.ar/revista/revista140/6\\_140.htm](http://med.unne.edu.ar/revista/revista140/6_140.htm) consultado 20/5/2012
9. <http://www.baires-salud.com.ar/diabetes-15/-la-prevalencia-de-diabetes-en-la-poblacion-argentina-es-de-85-239.html> consultado 20/5/2012
10. [http://www.indec.mecon.ar/principal.asp?id\\_tema=50](http://www.indec.mecon.ar/principal.asp?id_tema=50) consultado 20/5/2012