Asistente de evaluación clínica para la Estratificación de Riesgo Inferido Cardiovascular basado en redes bayesianas

Iris Sattolo[1], Blas Cappello[2], Jorge Ierache[1].

[1]Universidad de Morón-Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales
[1]Instituto de Sistemas Inteligentes y enseñanza experimental de la robótica
[2] Experto en Medicina Cardiovascular

iris.sattolo@gmail.com, cappelloblas@gmail.com, jierache@yahoo.com.ar

Resumen: El propósito de éste trabajo es presentar el desarrollo de una herramienta de software orientada a la salud basada en Redes Bayesianas, la cual permite estratificar el riesgo de padecer algún evento cardiovascular.

Palabras claves: Biomedicina, Factores de Riesgo Cardiovascular, Redes Bayesianas, Aprendizaje Automático, Sistemas Basados en Conocimiento.

1. Introducción:

Las enfermedades cardiovasculares constituyen una de las principales causas de morbilidad y mortalidad en el mundo. Por tal motivo es de gran interés para la ciencia médica desarrollar modelos de predicción de riesgo a padecer este tipo de enfermedades. Existen en la actualidad, diferentes tablas y algoritmos que se ocupan del tema, no obstante, los eventos cardiovasculares siguen ocurriendo en pacientes que fueron considerados de bajo riesgo y, por otro lado, no ocurren en pacientes considerados de alto riesgo. Por tal motivo surgió la necesidad de considerar una herramienta que: contribuya a mejorar la identificación de personas potencialmente vulnerables (en riesgo) en la prevención primaria cardiovascular, considere un número mayor de factores de riesgo representativos de la población en general y puedan ser extrapolables a distintas comunidades, incorpore una nueva manera de inferir el riesgo utilizando inferencia probabilística, permita agregar biomarcadores de información diagnóstica que puedan aportar valor a la estratificación del riesgo cardiovascular, facilite la estratificación del riesgo del sujeto evaluado (considerando al mismo sin riesgo, bajo riesgo, moderado riesgo y alto riesgo).
Por las razones anteriormente citadas, se ha propuesto un asistente de Estratificación de Riesgo Cardiovascular (ERICA) basado en redes bayesianas, el que incluye una mayor cantidad de variables e incorpora biomarcadores, que no se tuvieron en cuenta hasta el momento.

2. Descripción del problema.

Las enfermedades cardiovasculares constituyen un serio problema epidemiológico en el mundo contemporáneo; aproximadamente 17 millones de personas mueren cada año por infarto agudo de miocardio y accidente cerebro vascular [1].A lo largo de las últimas décadas se ha avanzado en las estrategias de tratamiento de los episodios agudos. Terapéuticas farmacológicas, quirúrgicas, hemodinámicas, han modificado el panorama de pacientes que sufren de algún episodio cardiovascular. Los resultados han sido menos alentadores en las estrategias destinadas a evitar que estos cuadros ocurran. *La prevención* es el escenario del presente y del futuro. Es allí, donde se verán concentrados los esfuerzos humanos y económicos en controlar esta expansión que parece no detenerse. El estudio Framingham [2]. permitió por primera vez identificar y definir los factores de riesgo cardiovasculares (FRCV) como precursores de las distintas manifestaciones clínicas de la enfermedad cardiovascular. Desde su publicación surgieron distintas tablas ATP III, SCORE, American Heard Association, Sociedad europea 2003, entre otras. La construcción de dichas tablas surge de estudios realizados en poblaciones cerradas. Sin embargo, las variables que pueden ser significativas para dicha población, pueden no serlo cuando son aplicadas en otras, pudiendo aparecer en las mismas, factores propios que pueden ser considerados de riesgo. No existe un consenso en la adopción de los factores de riesgo para la construcción de las tablas que proponen las distintas sociedades. En la práctica médica, tampoco reemplazan la evaluación personalizada, la contextualización y el juicio del médico. Como la evaluación del riesgo es una operación clínica que parte de tomar elementos de la historia del paciente, antecedentes heredo familiares y genéticos, determinaciones de laboratorio y otros exámenes complementarios, la intervención del experto en el tema fue imprescindible en este trabajo.
En la tabla N° 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 la columna final se muestran los factores de riesgo que se utilizaron para la construcción de ERICA. Donde PAS: Presión arterial Sistólica, AHF: Antecedentes Heredo Familiares, DA: Dietas aterogénicas, AS: Aterosclerosis Subclínica, CCA: Condiciones Clínicas Asociadas, LOD: Lesión Órgano Diana, PCR: Proteína C Reactiva, Xa: Tabaquismo actual, Xp: Tabaquismo actual y pasado, No: Excluidos los pacientes por considerarse enfermos. Las Tablas ATP III, Framingham, SCORE, American Heard Association son cuantitativas, a diferencia de la Sociedad europea 2003 que es cualitativa.

Tabla / Factores	ATP III	Framingham	SCORE Europa	Sociedad Europea 2003	Guía SHAPE	American Heard Association	ERICA
Tabaquismo	X a	Xp	X a		Xp	X a	Xp
P.A.S	X	X	X	X	X	X	no
LDL alto				X	X		X

HDL < 40 mg	X	X		X		X	X
Colesterol Total	X	X	X	X	X	X	
AHF		X					X
Obesidad				X	X		X
Sedentarismo					X		X
DA							X
EMI/Placa					X		X
Calcio Coronario					X		
AS							X
Sexo	X	X	X		X	X	X
Edad	X	X	X		X	X	X
CCA		X		X			no
Diabetes	no	X	X	X		X	no
LOD				X			no
PCR				X			
Glucemia				X			X

Tabla 1 "Cuadro comparativo de los factores de Riesgo"

Las tablas propuestas por los distintos estudios (Framingham, escore de España) se pueden dividir en dos categorías: *Tablas cuantitativas*: son aquellas que obtienen un número que equivale a la probabilidad de presentar una Enfermedad Cardiovascular (ECV) en un periodo de tiempo, generalmente 10 años. La mayoría de los métodos cuantitativos aceptan como riesgo bajo aquel que es < 10%, medio cuando el resultado es 10-20% y alto cuando ≥ 20%. *Las tablas cualitativas*: consisten en la suma de factores de riesgo cardiovasculares (FRCV), que se presentan como variables dicotómicas (presente o ausente). Clasifica a los individuos en leve, moderado o grave.

El principal inconveniente de las tablas cuantitativas es que al manejar umbrales de intervención según la edad de los pacientes y el riesgo absoluto de presentar un evento a corto y medio plazo hace que muchos jóvenes con importante carga de FRCV no alcanzarían los umbrales de tratamiento, a pesar de tener un mayor riesgo respecto a sus similares en edad y sexo, mientras que por el contrario la mayoría de los hombres mayores de 60 años alcanzarían umbrales de tratamiento, aún a pesar de presentar tan solo un ligero incremento del riesgo relativo**.**

Por otro lado, en el campo de la ciencia médica continúan los debates sobre utilizar el razonamiento probabilístico en la estratificación de riesgo cardiovascular [3], [4], sin que exista hasta el momento ninguna propuesta en el tema. Sin embargo, las redes bayesianas están siendo utilizadas en otras áreas de la ciencia médica PROSTANET, DIAVAL [5], TRIAGE [6], [7]

Como el núcleo central del enfoque bayesiano radica en determinar cómo actualizar cuantitativamente los conocimientos probabilísticos que tenemos de un fenómeno, parece lógico que en el campo de los meta-análisis, en el que se combina información sobre diferentes estudios, también sea de utilidad ese planteamiento.

En este contexto es que este trabajo propone una nueva manera de tratar el tema de riesgo cardiovascular. Surge así ERICA como un asistente de evaluación clínica de estratificación de riesgo inferido cardiovascular mediante redes bayesianas.

3. Fundamentos teóricos

Una red bayesiana es un grafo conexo acíclico dirigido, donde las uniones entre los nodos tienen definidas una dirección, en el que los nodos representan variables aleatorias que pueden ser continuas o discretas; y las flechas (arcos) representan influencias causales, el que un nodo sea padre de otro implica que es causa directa del mismo.[9].Los estados que puede tener una variable deben cumplir con dos propiedades: (1) ser mutuamente excluyentes, es decir, un nodo sólo puede encontrarse en uno de sus estados en un momento dado. (2) ser un conjunto exhaustivo, es decir, un nodo no puede tener ningún valor fuera de ese conjunto. Existen dos formas de justificar los enlaces que introducimos u omitimos al construir nuestra red. La primera es de naturaleza teórica: se forma un modelo causal a partir de la experiencia de un especialista y se trazan los arcos correspondientes al modelo (este es el caso que se plantea). 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. La estructura de la red, por si misma aporta gran cantidad de información cualitativa. En efecto, un arco XY indica, ya antes de conocer el valor concreto de probabilidad condicional, que hay una correlación entre ambas variables: el valor que toma X influye sobre la probabilidad de Y, y viceversa.
Dado que se conoce el valor de alguna(s) variable(s) se puede actualizar las probabilidades del resto de las variables; esto comúnmente se llama propagación de probabilidades, propagación de evidencia o inferencia. O sea: La inferencia es el proceso de introducir nuevas observaciones y calcular las nuevas probabilidades que tendrán el resto de las variables, por lo tanto, dicho proceso consiste en calcular las probabilidades a posteriori $P(X|Y)=y_i$ de un conjunto de variables X, después de obtener un conjunto de observaciones $Y= y_i$ donde Y es la lista de variables observadas e y_i es la lista correspondiente a los valores observados.
Las probabilidades a posteriori $P(X|Y= y_i)$, se pueden obtener a partir de la probabilidad marginal $P(X|Y)$, que a su vez puede obtenerse de la probabilidad conjunta $P(x_1, x_2, xi)$ sumando los valores para todas las variables que no pertenezcan al conjunto X e Y. La red tiene información cualitativa (estructura de grafos) y cuantitativa (las probabilidades a priori y condicionales). Esta red constituye nuestro modelo causal y, salvo que se introduzca algún mecanismo de aprendizaje, es invariable. Este modelo permite tanto un razonamiento diagnóstico (cual es la causa más probable) como predictivo (que valor de cierta variable aparecerá con mayor probabilidad). Una ventaja de las Redes Bayesianas (RRBB) es que un mismo nodo puede ser fuente de información u objeto de predicción dependiendo de cuál sea la evidencia disponible

4 Metodología

Se investigaron las metodologías para la construcción de un sistema experto, decidiendo para este trabajo la propuesta por Blanqué y García Martínez [García Martínez y Blanqué, 1997; 1998] BGM [8], según se detalla en la figura Nº 1.Se

incorpora en la etapa de planteo de causalidades el aprendizaje paramétrico a través de la aplicación de redes bayesianas.

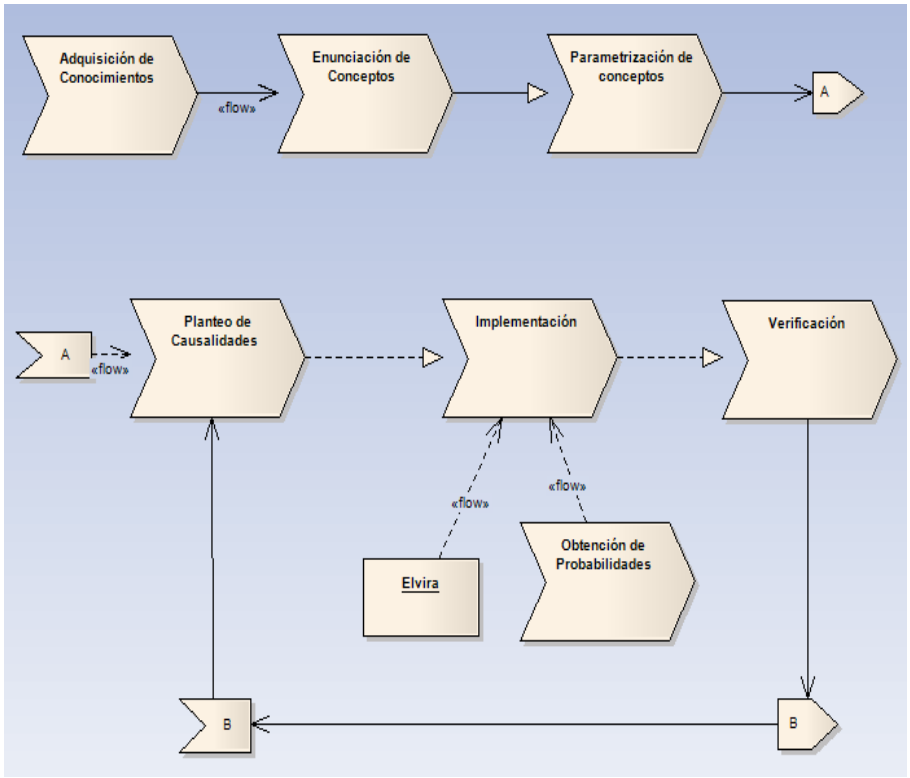

Figura Nº 1"Etapas de la metodología BGM"

Etapa 1: Adquisición de conocimientos: en esta fase se recogió información sobre el tipo de problema a resolver, el razonamiento que siguen los especialistas en el tema y los conceptos que aparecen en el razonamiento. Se consultó bibliografía específica del dominio para determinar algunos valores de probabilidad. Pero la participación del experto en este trabajo fue fundamental para la construcción de esta red causal.

Etapa 2 Enunciación de Conceptos: en esta etapa se definieron los conceptos primarios y secundarios y la relación existente entre los distintos conceptos a nivel médico.

Etapa 3: Parametrización de conceptos: se definieron los parámetros de los factores de riesgo tales como fumador/ exfumador,/no fumador, dieta equilibrada/ no equilibrada ,valores probabilísticos, valores bioquímicos.

Etapa 4: Planteo de Causalidades: en esta fase se establecen las relaciones de causalidad entre los conceptos mencionados armando la red causal.

Etapa 5 Implementación: se empleo la herramienta ELVIRA [10] para implementar la red causal y para asignar a cada nodo los factores asociados a las probabilidades determinadas con el experto.

Etapa 6: Verificación: se efectúo la aceptabilidad de la red con el experto en el campo. Esto se realizó usando casos de testeo que se consideraron típicos, se compararon los resultados con los dados para esos mismos casos por el experto, y en base a esa comparación, se modificó, tanto en forma cualitativa como cuantitativa la red.

5. Solución Propuesta

Como resultado de la educción de conocimiento al experto, se presenta a continuación el modelo causal, con sus arcos correspondientes. La figura 2 muestra la red causal propuesta en este trabajo, que plantea una nueva manera de tratar el tema de riesgo cardiovascular, a través de la evaluación clínica de estratificación de riesgo inferido cardiovascular mediante redes bayesianas.

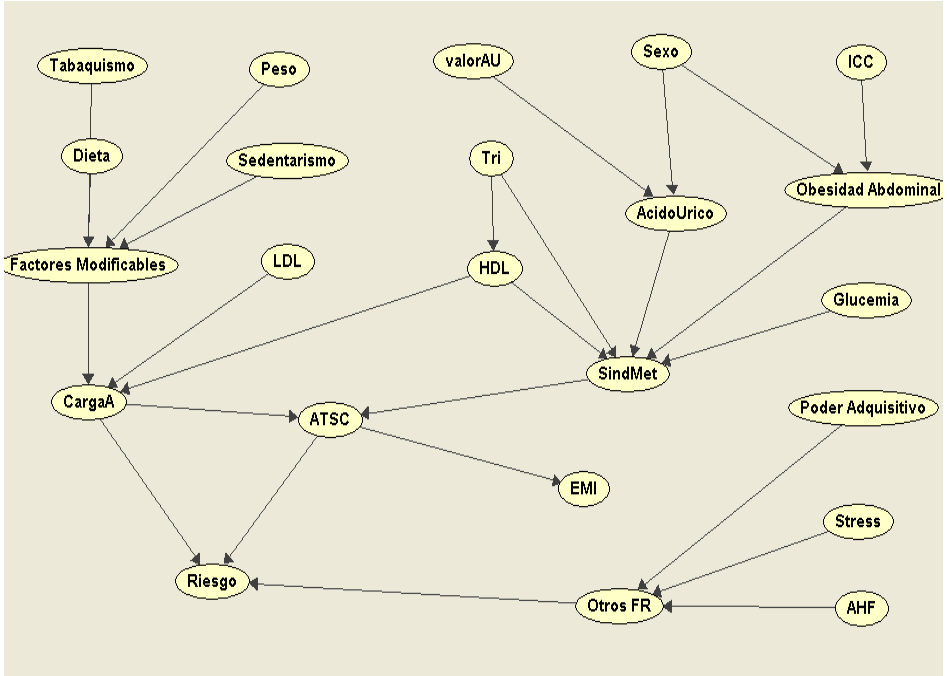

Figura 2 "Red Causal de estratificación de riesgo inferido cardiovascular"

6. Caso de prueba

Se desarrollaron casos de pruebas correspondiente a cada nodo, en este orden se presenta en la tabla Nº 2 los resultados obtenidos de un caso de prueba correspondiente al nodo riesgo, que contempla como resultados: (1) el comportamiento con los valores en la red sin el Valor EMI (Espesor Mio Intimal) y (2) considerando el valor de EMI en la red

Elemento a probar: **Nodo: Riesgo** Característica: Persona con 4 factores de riesgo con valor de Eco Doppler .					
Entradas			Resultado		
Variable	Valor Asignado sin EMI	Valor de EMI asignado	Valor ERICA sin EMI (1)	Valor ERICA con EMI(2)	Valor esperado
Sexo	Masculino				
LDL	normal				
HDH	normal				
Tabaquismo *	presente				
Sedentarismo*	presente				
Peso*	presente				
Dieta*	presente				
EMI Espesor Mio Intimal		<1,5 mm	0.8- 0,9mm 45 %		
C Aterogénica			Media 51 %	Media 91 %	Medio
ATSC Aterosclerosis sub-clínica			Inicial 49 % inter. 41 %	Alta 87%	Alto
Síndrome M	ausente				
Otros F R	ausente				
Riesgo			Moder.49% bajo 40 %	Alto 79 %	Alto

Tabla Nº 2 "Resultados del nodo riesgo con y sin el valor de EMI, aplicando ERICA"

Sobre la base de caso de prueba presentado, se muestra en la figura 3 los valores en el modo inferencia de la red Causal propuesta en ERICA, 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.

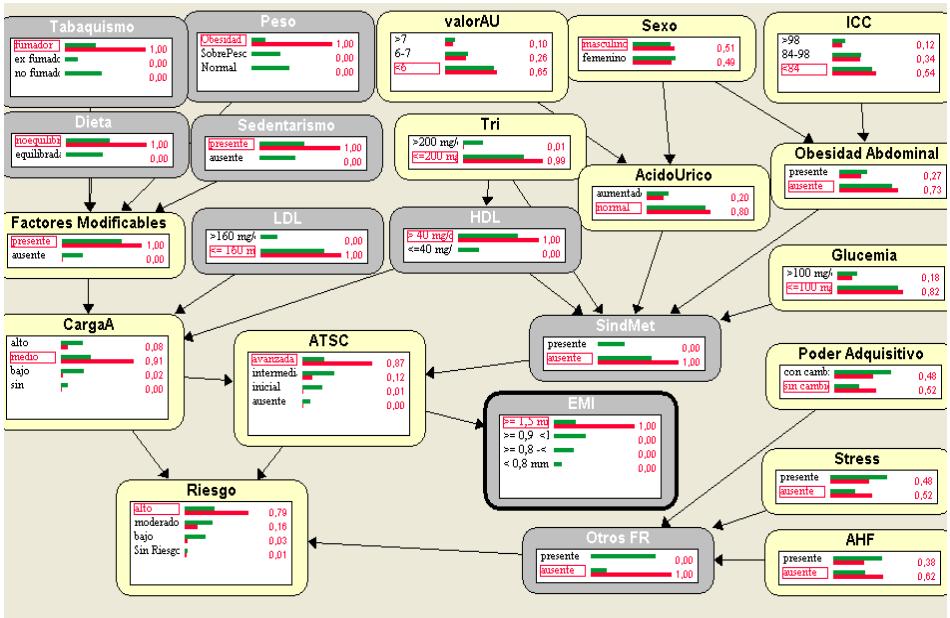

Figura 3 “Valores en el modo inferencia empleando ELVIRA”

7. Resultados

Se realizó una comparación de la estratificación hecha a través de la tabla propuesta por el American Heart Association (A.H.A) y los resultados arrojados por ERICA de 50 casos clínicos sin introducir el valor del Eco Doppler Carotideo, el cual se utiliza como parámetro para certificar los resultados. En la tabla Nº 3 se muestran los resultados obtenidos.

Cantidad de personas con Eco Doppler		American Heart Association			ERICA		
		Riesgo			Riesgo		
Resultado	Cant	Bajo	Moderado	Alto	Bajo	Moderado	Alto
Normal	15	6	9	0	5	5	5
Aumentado	8	5	3	0	1	4	3
Placa	27	14	10	3	0	4	23
Total Pacientes	50	25	22	3	6	13	31

Tabla Nº 3 “Comparación de resultados A.H.A y ERICA”

Se observa que en 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.

Se muestran en la figura Nº 4, los resultados que se alcanzaron al comparar la cantidad de pacientes con valores de EMI (obtenidos por el estudio Doppler), y los resultados obtenidos para dichos pacientes considerando la tabla propuesta por el A.H.A. En la figura 5 se muestran los resultados al comparar la cantidad de pacientes con valores de EMI obtenidos por el estudio Doppler correspondiente y los resultados obtenidos para dichos pacientes con la tabla propuesta por ERICA

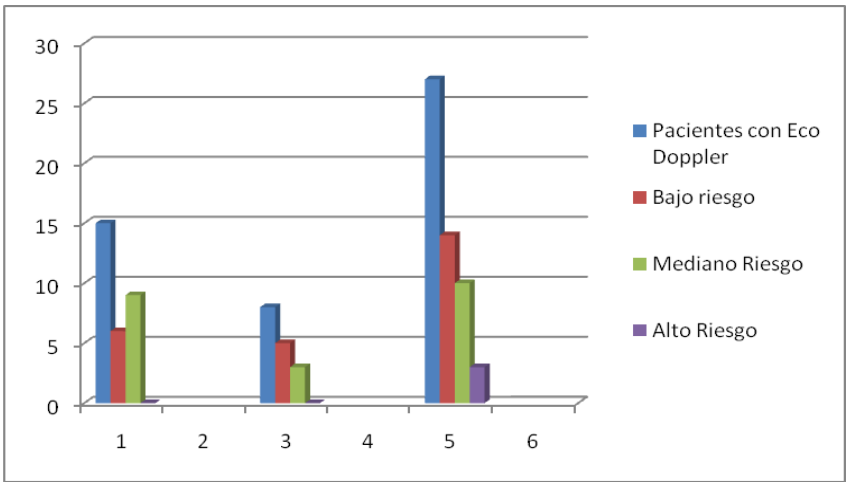

Figura 4 "Comparación EMI con los valores propuestos por A.H.A"

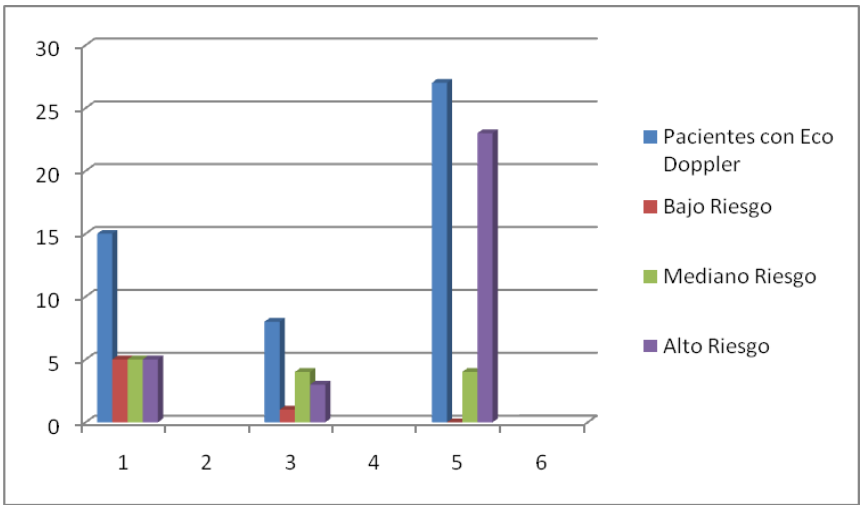

Figura 5 "Comparación EMI con la propuesta de ERICA"

8. Conclusiones y futuras líneas de investigación

El marco metodológico ofrecido por BGM resultó ser adaptable en el desarrollo de la red propuesta para el presente trabajo. Las redes bayesianas están estratégicamente diseñadas para abordar los problemas de diagnóstico e inferencia y pueden considerarse efectivas para el caso específico de la estimación de riesgo cardiovascular. Para la ciencia médica constituye un aporte original al permitir evaluar el riesgo en poblaciones abiertas es decir, son extrapolables con solo cambiar la prevalencia de los valores de la población en estudio. Posibilita introducir nodos a la red que ofrezcan información diagnóstica, tal es el caso del Espesor Mio Intimal (EMI), que hace posible certificar el riesgo, caso contrario, permite re-estratificarlo. Se utilizan factores de riesgo no convencionales representativos para la población en estudio. Los escores por puntuaciones subestiman el riesgo, ERICA permite una evaluación conservadora de las poblaciones en eventual riesgo cardiovascular. El mayor grado de coherencia entre la estimación del riesgo mediante redes bayesianas y la presencia de EMI y placas fibrolipídicas, sugeriría que la guía ERICA es la más adecuada para predecir eventos coronarios en nuestra población. De la experiencia adquirida durante la realización del presente trabajo quedan abiertas las siguientes líneas de investigación en orden a comparar ERICA con la red propuesta por los distintos clasificadores que tenemos disponibles (Naives Bayes, Tan, KDB), ampliar la capacidad de la red ajustando probabilidades contenidas en la misma, refinar la base de conocimientos, 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.

9. Referencias

[1] Khut y col. Prevalence of Convencional Risk Factors in Patients with Coronary Heart Decease. JAMA 2003; 290: 898-904

[2]. http://www.framinghamheartstudy.org/about/spanish.html

[3] Silva LC, Suárez P. ¿Qué es la inferencia bayesiana? JANO 2000; 58: 65-66.

[4] Silva LC, Muñoz A. Debate sobre métodos frecuentistas vs bayesianos. Gac Sanit 2000; 14: 482-494

[5] F. J. Díez, J. Mira, E. Iturralde y S. Zubillaga. DIAVAL, a Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, 10:59-73, 1997.

[6] Mar Abad-Grau, Jorge Ierache, Claudio Cervino, Paola Sebastiani "Evolution and challenges in the design of computational systems for triage assistance in the Emergency Department" Special Issue of the Journal of Biomedical Informatics on Computerized Decision Journal of Biomedical Informatics, Editorial Elsevier JBI-07-75R2 , 2008.

[7] M Abad Graud, C Cervino, J Ierache "Modelado de Sistema Experto para el Triaje en servicios de urgencias medicas CACIC 2007, Universidad Nacional del Noreste-Facultad de Ciencias Exactas y Agrimensura Corrientes y Facultad Regional-Resistencia Universidad Tecnológica Nacional: Páginas 1734-1744, ISBN:978-950-656-109-3

[8] García Martínez, R. 1992. Construcción de Sistemas Expertos CEI-UBA. Argentina

[9] Castillo, Gutiérrez y Hadi. Sistemas Expertos y Modelos de Redes Probabilísticas

[10] http://www.ia.uned.es/~elvira.