



Clasificación de pacientes a partir de factores de riesgo cardiovascular mediante mapas auto-organizados de Kohonen.

Clasificación de pacientes a partir de factores de riesgo cardiovascular mediante mapas auto-organizados de Kohonen.¹

Patients classification based on cardiovascular risk factors by means of Kohonen's self-organized maps.

Juan Carlos Maldonado Beltrán²

¹Universidad Industrial de Santander, Colombia.

Artículo recibido en el mes de Octubre de 2013; artículo aceptado en el mes de Noviembre de 2013

Citación del artículo: Maldonado, J. (2013). Clasificación de pacientes a partir de factores de riesgo cardiovascular mediante mapas auto-organizados de Kohonen. *I+D Revista de Investigaciones*, 2(2), 61-75.

Resumen

Se presenta una alternativa para generar una estructura de clasificación de pacientes a partir de sus factores de riesgo cardiovascular, mediante la implementación de mapas auto-organizados de Kohonen. Se desarrolló una aplicación en Matlab® que automatiza la búsqueda de los parámetros para el diseño y entrenamiento de un mapa de Kohonen, a partir de un conjunto de datos de pacientes, y que posteriormente los clasifica a partir de la similitud de sus factores de riesgo. Se clasificaron 190 pacientes colombianos y se

comparó la clasificación resultante respecto a la propuesta por el estudio de Framingham en los Estados Unidos, de predicción de riesgo cardiovascular, encontrando características similares entre los métodos de clasificación; no obstante, se observó que es posible identificar grupos de pacientes en la muestra clasificada utilizando mapas auto-organizados, que no se identifican mediante la clasificación del estudio de Framingham.

Palabras clave: Mapa auto-organizado de Kohonen, factores de riesgo cardiovascular, modelo de predicción de riesgo cardiovascular.

¹Artículo de investigación científica, de enfoque cuantitativo, resultado de un proyecto de investigación denominado: Mapas auto-organizados de Kohonen y riesgo cardiovascular, desarrollado en el grupo de investigación de la Universidad Industrial de Santander de la ciudad de Bucaramanga (Colombia). Carrera 27 Calle 9. PBX: (57) (7) 6344000. La investigación fue financiada por la misma institución universitaria. Carrera 27 calle 9. PBX: (57) (7) 6344000.

²Ingeniero de Sistemas, Magíster en Ingeniería de Sistemas e Informática de la Universidad Industrial de Santander. Docente-Investigador.

Abstract

This research shows a new choice in the patient's classification structure based on their cardiovascular risk factors, by means of the application of Kohonen's self-organized maps. It was developed an application in Matlab® which automatizes the search of the parameters for designing and training a Kohonen's map with a given patients' data set, in order to classify them under risk factor similarity criterion. They were classified 190 Colombian patients and this classification was compared against that set out by Framingham's cardiovascular risk prediction study in the United States, finding similar characteristics between the two classification methods; nevertheless it was observed that, it is possible to identify groups of patients in the classified sample using self-organized maps, not identified in the Framingham study classification method.

Keywords: Kohonen's self-organized maps, cardiovascular risk factors, cardiovascular risk prediction model.

Introducción

En el entorno clínico, la palabra riesgo suele definirse como la probabilidad de que ocurra un evento adverso. El riesgo se considera generalmente durante un período determinado y finito de tiempo (normalmente 10 años), aunque también hay interés en predecir el riesgo de toda la vida (Jackson, Kerr & Wells, 2010). Actualmente existe un gran número de trabajos relativos al estudio del riesgo de enfermedad cardiovascular (Conroy, Pyorala, Fitzgerald, 2003; Paynter, et

al., 2010) como el estudio de Framingham (National Heart, Lung, and Blood Institute, 1923–2011) que es ampliamente utilizado para la toma de decisiones terapéuticas con base en la estimación de riesgo proporcionada por el modelo, al introducir las características de riesgo del paciente concreto.

Los modelos actuales de predicción de riesgo cardiovascular utilizan variables denominadas factores de riesgo cardiovascular, con base en las cuales se estima la probabilidad de ocurrencia de una enfermedad cardiovascular (ECV) en un periodo de tiempo determinado, es decir, asignan un valor o probabilidad de ocurrencia de ECV a un individuo con base en los valores de sus factores de riesgo. Entre los factores de riesgo tradicionales se encuentran: edad, género, presión arterial sistólica (PAS), colesterol total (LDL-C, por sus siglas en inglés, *low-density lipoprotein*), colesterol HDL (HDL-C, por sus siglas en inglés, *high-density lipoprotein*) y el tabaquismo. Adicionalmente, se ha encontrado una fuerte asociación entre los índices de variabilidad de la frecuencia cardíaca (VFC) en el dominio de la frecuencia y el riesgo de presentar un evento cardíaco. De igual manera, según algunos estudios (Sloan, Shapiro, Bagiella, Myers, Gorman, 1999), los índices de variabilidad de la presión arterial (VPA) pudieran ser factores de riesgo cardiovascular.

Cuando se analiza el comportamiento de estos modelos de predicción de riesgo cardiovascular, se observa que para individuos con valores de factores de riesgo similares, se obtienen porcentajes de riesgo similares, es decir, que en una población (p.e. la población colombiana), existen grupos de individuos que tienen

características similares de acuerdo con los valores de sus factores de riesgo, y por lo tanto, se puede plantear una hipótesis, en la cual se supone que es posible encontrar una estructura de clasificación de los individuos de una población en categorías de riesgo, donde cada una de estas categorías corresponde a un subconjunto de individuos con características similares, y consecuentemente, con porcentaje de riesgo de ECV similar. Aunque los modelos actuales de predicción de riesgo cardiovascular permiten realizar dicha clasificación, el poder predictivo de estos modelos es bastante pobre (Menotti & Puddu, 2000; Payne, 2012) en términos de calibración, discriminación y generalización. En el presente trabajo se resalta el problema de la baja discriminación de los actuales modelos de predicción de riesgo cardiovascular, es decir, qué tan bien la puntuación de riesgo calculada usando un modelo, puede con precisión estratificar o clasificar a los pacientes.

Se propone una solución por medio de la elaboración de un modelo preliminar de clasificación de pacientes, que permite la identificación de las clases presentes en una muestra, a partir de la similitud de los factores de riesgo de sus individuos. La detección de las categorías o clases en la muestra se realiza de manera no supervisada, mediante la implementación de mapas auto-organizados (SOM, por sus siglas en inglés, *self-organizing map*) de Kohonen, cuyos parámetros de diseño y entrenamiento pasan por un proceso previo de optimización, con el objetivo de encontrar la configuración ideal para la muestra que va a ser clasificada.

Las redes neuronales artificiales (RNA) han emergido como una potente herramienta para el modelado estadístico, orientadas principalmente al reconocimiento de patrones, tanto en la vertiente de predicción como de clasificación (Martínez, s.f). Entre los paradigmas de RNA que son usados en diversos campos de aplicación, se destacan los SOM de Kohonen (Kohonen, 2001). En la corteza cerebral existen zonas especializadas en ciertas tareas, donde neuronas asociadas a características similares, son vecinas. Producto de este comportamiento, se generan mapas donde se ordenan las características de los estímulos recibidos. Kohonen diseñó un modelo de red competitiva que emula esta idea, mediante el algoritmo SOM (Kohonen, 1981, 1982), que se introdujo en 1981 y cuyas primeras aplicaciones se realizaron principalmente en ingeniería. Posteriormente, el algoritmo se convirtió progresivamente en un método estándar para el análisis de datos, en una gran variedad de campos donde se utiliza el aprendizaje no supervisado: agrupamiento, visualización, organización de datos, caracterización y exploración, aplicado a la solución de problemas en múltiples áreas del conocimiento (Kaski, Kangas & Kohonen, 1998; Oja, Kaski & Kohonen, 2003; Pöllä, Honkela & Kohonen, 2009).

Con base en lo anterior, se realiza una comparación con el modelo de Framingham por medio del análisis gráfico exploratorio de un conjunto de 190 pacientes, entre ellos 156 mujeres entre los 15 y 75 años, y 34 hombres entre los 18 y 45 años, para los cuales se registraron seis variables asociadas en la bibliografía como factores de riesgo cardiovascular tradicionales.

Método

Tipo de estudio

Se siguió un tipo de investigación no experimental-descriptiva-transversal.

Participantes

Se utilizó una muestra no probabilística o dirigida, correspondiente a 190 pacientes de nacionalidad colombiana, entre ellos 156 mujeres de 15 a 75 años y 34 hombres de 18 a 45 años, para los que se registraron las variables correspondientes a los factores de riesgo cardiovascular tradicionales: género, edad, LDL-C, HDL-C, PAS y tabaquismo.

La decisión de usar este tipo de muestra, se debe a la disponibilidad del acceso a bases de datos creadas en investigaciones anteriores por la Fundación Cardiovascular de Colombia (Guzmán et al., 2006; García et al., 2012), donde se han captado las variables deseadas (factores de riesgo cardiovascular tradicionales) para la elaboración y prueba de la aplicación presentada en este trabajo, además de los elevados costos que implica la realización de este tipo de procesos de recolección de datos.

Las medidas de tendencia central y de dispersión para la muestra de 190 pacientes se presentan en tabla 1.

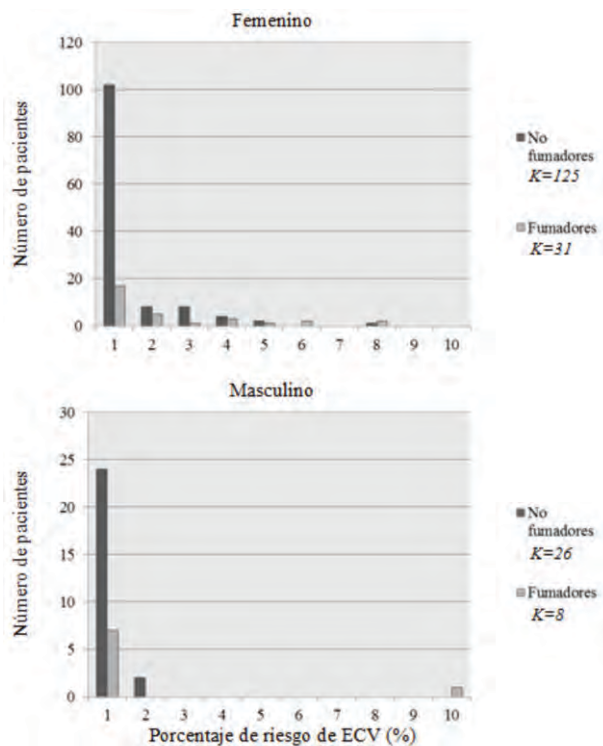
Tabla 1. *Medidas de tendencia central y dispersión de la muestra*

Género	Femenino (n = 156)			
Fumadores	31			
	Edad (años)	LDL-C (mg/dl)	HDL-C (mg/dl)	PAS (mmHg)
Mediana	34.0	186.8	48.5	110.0
Media	36.1	186.0	48.4	110.0
Desviación estándar	14.2	39.4	12.4	16.4
Valor máximo	75.0	285.0	81.4	198.0
Valor mínimo	15.0	90.0	22.0	75.0
Rango	60.0	195.0	59.4	123.0
Género	Masculino (n = 34)			
No fumadores	8			
	Edad (años)	LDL-C (mg/dl)	HDL-C (mg/dl)	PAS (mmHg)
Mediana	26.5	177.0	42.6	110.5
Media	27.2	185.9	41.0	111.6
Desviación estándar	5.6	45.6	11.3	10.8
Valor máximo	45.0	301.0	69.0	137.0
Valor mínimo	18.0	115.5	18.2	94.0
Rango	27.0	185.5	50.8	43.0

Fuente: Autor.

La muestra se dividió en cuatro grupos por género (masculino/femenino) y tabaquismo (fumadores/no fumadores), para no incluir estas variables categóricas en el diseño del SOM y se calculó el porcentaje de riesgo según el modelo de Framingham para cada paciente.

Figura 1. Porcentajes de riesgo mediante el modelo de Framingham para la muestra..



Fuente: Autor.

Materiales e instrumentos

Matlab® (Laboratory of Computer and Information Science) y **Somvis®** (Vienna University of Technology). La estandarización de datos, diseño y entrenamiento del SOM se automatizó por medio del desarrollo de una aplicación, usando el entorno de desarrollo integrado de *Matlab® R2010a* de MathWorks®, que

implementa la *SOMToolbox* para Matlab® y el paquete de visualización *SOMVIS®* (Vienna University of Technology). La aplicación permuta los parámetros de estructura y entrenamiento de un SOM (topología, forma del mapa, función de vecindad, algoritmo de entrenamiento) con el objetivo de encontrar la mejor configuración de parámetros y permite la visualización de los resultados. Para un conjunto de datos de entrada

$$E| = \sum_{j=1}^M \sum_{k=1}^K h_{jc(k)} \|w_j - X_k\|^2 \quad (1)$$

Donde W_j corresponde al vector de referencia de la neurona de la capa de salida j para $(j=1,2,3,\dots, M)$.

Procedimiento

Para la realización del estudio se llevaron a cabo los siguientes pasos:

El proceso de estandarización de datos no es indispensable, pero es importante cuando las variables que serán presentadas a la capa de entrada del SOM, manejan diferentes unidades de medida y escala, debido a que algunas técnicas como la distancia euclídea requieren que todos sus parámetros manejen la misma escala, para poder obtener una comparación ecuánime. En tal caso, este proceso resulta como una fase previa al entrenamiento del SOM, dependiendo de las características de las variables que se han medido del fenómeno en estudio.

Sea $X_k \in \mathbb{R}^N$ una muestra (correspondiente a los factores de riesgo del paciente k) dispuesta para ser presentada a la capa de entrada del SOM, la muestra puede ser descrita como:

$X_k = [Y_{1k}, Y_{2k}, Y_{3k}, \dots, Y_{Nk}]$ donde $Y_{nk} \in \mathbb{R}^K$ corresponde a la k -ésima medición de la variable

n (p.e. edad), para el paciente k , para ($k=1,2,3, \dots, K$). De manera que cada medición puede ser estandarizada con el objetivo de tipificar la media y la varianza de las Y_n variables medidas del fenómeno, mediante transformaciones que, aunque modifican sus rangos, hacen permanecer invariantes sus distribuciones; luego, el comportamiento de dichas variables puede ser analizado mediante la implementación del SOM, garantizando que la magnitud de las mediciones, que depende de las escalas y unidades de medida usadas para muestrear el fenómeno, no influyen la clasificación obtenida. Un método para estandarizar las muestras Y_{nk} sin modificar la distribución de las Y_n variables, se describe mediante la transformación:

$$Y_{nk} = \frac{(Y_{nk} - \bar{Y}_n)}{\sigma_{Y_n}}$$

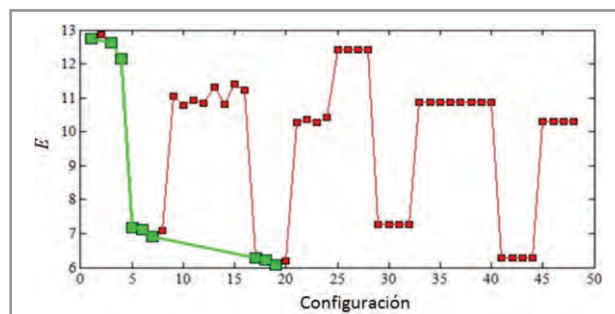
Donde \bar{Y}_n es la media de la variable Y_n y σ_{Y_n} es su desviación estándar. Finalizada la estandarización de datos, la fase siguiente corresponde al diseño y entrenamiento del SOM.

Para el entrenamiento secuencial se utilizaron los valores $\alpha_0 = 1$ y $\beta_0 = \sqrt{M}$ (Kohonen, 1982). El número de neuronas de la capa de entrada, corresponde a la cantidad de factores de riesgo medidos para cada paciente ($N = 4$) El número de neuronas de la capa de salida se determinó mediante una regla heurística que calcula dicho valor a partir del tamaño de la muestra dispuesta a ser clasificada ($M = 5\sqrt{K}$) [37].

En la figura 2 se presenta el proceso de búsqueda de los parámetros óptimos para la

creación y entrenamiento de SOM automatizado, a partir del grupo de género femenino y no fumadores ($K=125$). Se observa que se probaron 48 configuraciones posibles, es decir, que se crearon y entrenaron igual cantidad de SOM para cada grupo de la muestra, evaluando la calidad de cada uno de éstos por medio del cálculo de la función de optimización que minimiza (1). En color rojo se observan los valores calculados para cada configuración de parámetros, mientras que en verde las configuraciones que minimizaron el valor de E en alguna iteración. Finalmente, se observa que en la configuración 19 se obtiene el valor mínimo para E .

Figura 2. Proceso de búsqueda de parámetros óptimos para un SOM



Fuente: Autor.

El proceso de búsqueda de parámetros se realizó para los cuatro grupos en que se dividió la muestra, resultando en el diseño y entrenamiento de igual número de SOM, donde cada uno de estos representa las características del grupo de la muestra de origen.

Resultados

Se presentan los SOM obtenidos a partir de los grupos de la muestra, producto del proceso de búsqueda de parámetros. Para su visualización se

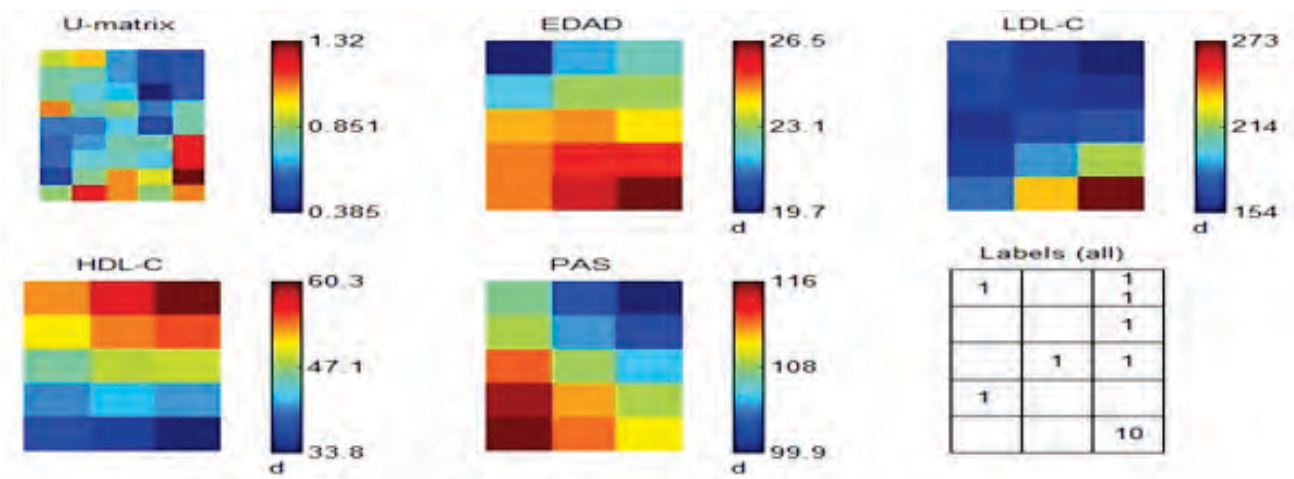
utiliza inicialmente la matriz de distancias unificadas (U-matrix, por su nombre en inglés, *unified distance matrix*), los planos de cada variable y el histograma de mapeo de casos en la tabla 2 (Ultsch & Siemon, 1990). La tabla 2 muestra la configuración del SOM para el grupo de 8 pacientes de género masculino y fumadores.

Tabla 2. Configuración del SOM para el grupo de pacientes de género masculino-fumadores

Parámetro	Tipo
Topología	Rectangular
Forma	Laminar
Función de vecindad	Gaussiana
Algoritmo de entrenamiento	Secuencial

Fuente: Autor. ($K=8, M=15$)

Figura 3. U-matrix, planos de factores de riesgo e histograma de mapeo grupo masculino y fumadores.



Fuente: Autor

La U-matrix en la figura 3, muestra las distancias calculadas entre los pesos de los vectores de referencia de neuronas adyacentes, donde se pueden identificar regiones densas denominadas *clusters*, en las cuales las distancias entre los vectores de referencia de las neuronas adyacentes son pequeñas. En esta matriz no todas las celdas corresponden a una neurona de la capa de salida, sino que se han incluido celdas alrededor de cada neurona, que corresponden a las distancias calculadas entre los pesos de las neuronas y sus adyacentes. En la posición de la neurona se localiza el promedio de las distancias calculadas respecto a todas sus

vecinas. En la figura 3, se muestra en la parte superior derecha de la U-matrix un sector en color azul que puede corresponder a un *cluster*.

Por otra parte, un plano de una variable en la figura 3, muestra los valores de los pesos de los vectores de referencia de cada neurona para una variable particular en todo el mapa, y permite identificar el tipo de influencia que tiene una variable en determinado sector del mapa. En la figura 3, se presentan los planos para los factores de riesgo y se observa que en el plano correspondiente a la edad, los pacientes de menor edad tendrán la tendencia a quedar mapeados en la

parte superior izquierda, mientras que los de mayor edad en la parte inferior derecha.

Finalmente, el histograma de mapeo de casos permite identificar a cuál neurona de la capa de salida se asigna un caso presentado a la capa de entrada, cuando el SOM se encuentra en modo de operación normal. Se conoce como se revela en la figura 3, que el grupo de pacientes clasificados contiene un único caso correspondiente a un paciente con 10% de riesgo de ECV según el modelo de Framingham, y siete casos correspondientes a pacientes con 1% de riesgo de ECV. Por otra parte, se observa en la figura 3, que en la celda de la esquina inferior derecha de la U-matrix, que la distancia entre el vector de referencia de la neurona representada por dicha celda y los vectores de referencia de las neuronas adyacentes a ésta, presentan valores elevados, indicando esto, que la neurona corresponde a una característica aislada del mapa. Adicionalmente, en el histograma de mapeo de casos, se observa que el paciente con 10% de riesgo de ECV se encuentra asignado a la neurona de la celda de la esquina inferior derecha del mapa, lo que se puede interpretar como un comportamiento coherente del SOM que logra aislar un caso particular y acercar los casos similares.

Tabla3. Configuración del SOM para el grupo de pacientes de género masculino-no fumadores

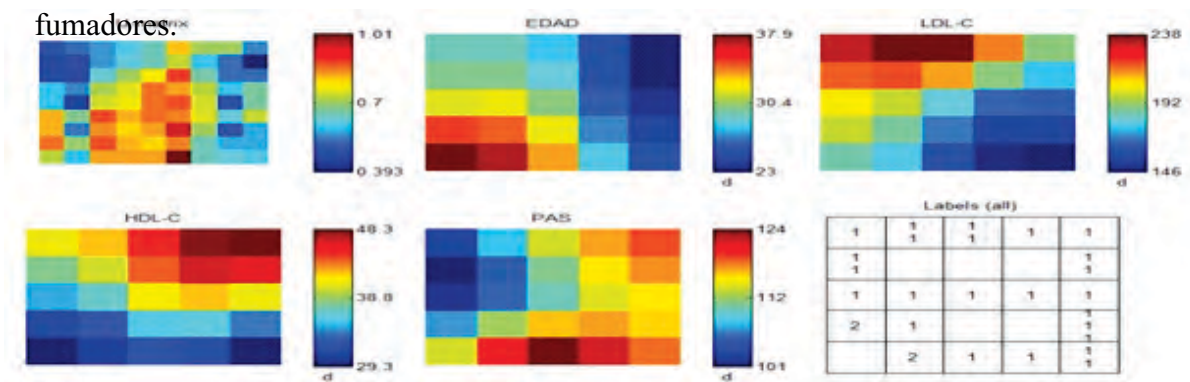
Parámetro	Tipo
Topología	Rectangular
Forma	Laminar
Función de vecindad	Gaussiana recortada
Algoritmo de entrenamiento	Secuencial

Fuente: Autor ($K=23, M=25$)

La tabla 3 presenta la configuración del SOM para el grupo de 26 pacientes de género masculino y no fumadores. Se conoce de este grupo que contiene 2 pacientes con porcentaje de ECV del 2% y que los pacientes restantes tienen un 1% de riesgo de ECV según el modelo de Framingham.

Se observa en el histograma de mapeo de casos, figura 4, que los dos pacientes cuyo porcentaje de riesgo es 2%, se encuentran clasificados en las neuronas de la parte inferior izquierda del mapa y separados de los demás pacientes. No obstante, este comportamiento no implica que los dos pacientes correspondan a la misma característica, es decir, que pertenezcan a una misma clase; si bien se encuentran en neuronas topológicamente cercanas, dichas neuronas no pertenecen a un mismo *cluster* como se observa en la U-matrix, por lo tanto, cada neurona representa una característica particular del grupo y si se observan en los planos de las variables, los tonos en las neuronas a las que se encuentran asignados dichos pacientes, son similares en la mayoría de las variables, excepto en el plano de la PAS. A medida que aumenta el número de casos de $K=8$ a $K=26$, se observa para la muestra, que la proporción de neuronas que clasifican más de un paciente, también aumenta.

Figura 4. U-matrix, planos factores de riesgo e histograma de mapeo masculino y No fumadores.



Fuente: Autor.

La tabla 4 presenta la configuración del SOM para el grupo de 31 pacientes de género femenino y fumadores. Este grupo presenta una mayor variabilidad en el porcentaje de riesgo de ECV según el modelo de Framingham de sus pacientes con un rango entre 1% y 8%

De manera similar a los grupos anteriores, la clasificación generada por el SOM presenta

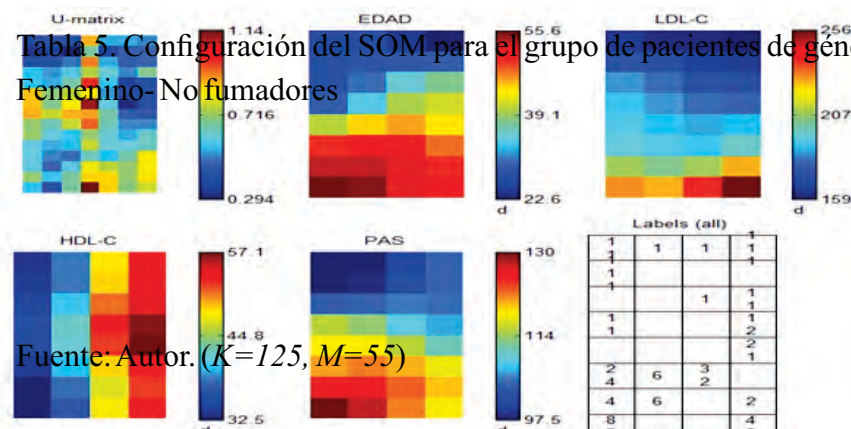
Tabla 4. Configuración del SOM para el grupo de pacientes de género Femenino- fumadores

Parámetro	Tipo
Topología	Rectangular
Forma	Laminar
Función de vecindad	Gaussiana recortada
Algoritmo de entrenamiento	Por lotes

Fuente: Autor (K=31, M=25)

desde la parte superior del mapa, hacia la parte inferior del mismo.

Tabla 5. Configuración del SOM para el grupo de pacientes de género Femenino- No fumadores



Fuente: Autor. (K=125, M=55)

La tabla 5 presenta la configuración del SOM para el grupo de 125

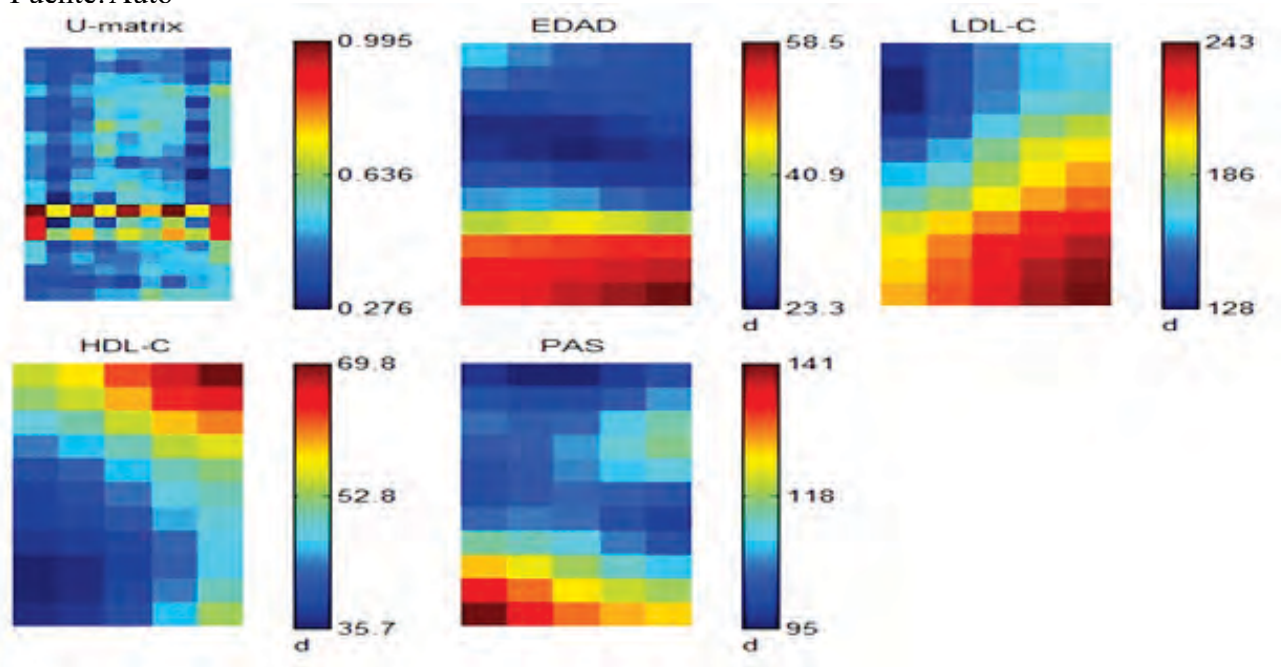
pacientes de género femenino y no fumadores. Este grupo presenta una distribución de porcentajes de riesgo de ECV de sus pacientes entre 1% y 8%, con una cantidad de pacientes (102 pacientes) con 1% de riesgo de ECV considerablemente mayor que la cantidad (23 pacientes) cuyo porcentaje de riesgo de ECV se encuentra entre 2% y 8%.

pacientes de la parte inferior, que tienden a presentar valores mayores de dicha variable.

Figura 6. U-matrix, planos de los factores de riesgo grupo femenino y No fumadores.

Para este grupo de pacientes, se observa en la figura 6, que la U-matrix indica la presencia de dos *clusters* definidos. El primero de estos lo compone la sección superior del mapa, mientras que el segundo lo compone la parte inferior. Al observar el plano de la variable edad, se nota una coincidencia de dichos *clusters* con el comportamiento de esta variable, donde la sección superior del mapa corresponde a los pacientes de menor edad en comparación con los

Fuente: Auto

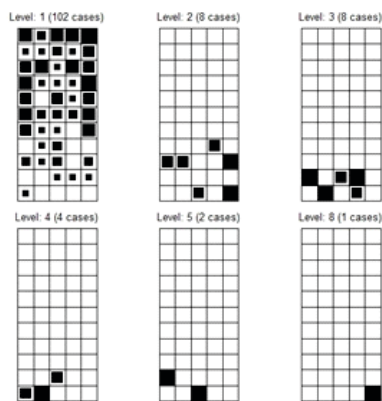


Fuente: Autor

El gráfico de vecindad es una técnica de visualización reciente basada en el método de los k -vecinos cercanos (Polzlbauer, Rauber & Dittenbach, 2005), que tiene en cuenta la densidad de los datos. Este método genera un gráfico basado en la medición de la distancia al vecino más cercano basado en un radio k .

Una vez aplicada la técnica, se comprueba la existencia de dos *clusters* en el mapa, es decir, que los pacientes clasificados en cada *cluster*, comparten características similares detectadas por el SOM, que no son discriminadas por el modelo de Framingham para la muestra.

Figura 8 Gráfico de vecindad mediante el método de los k -vecinos cercanos ($k=2$).

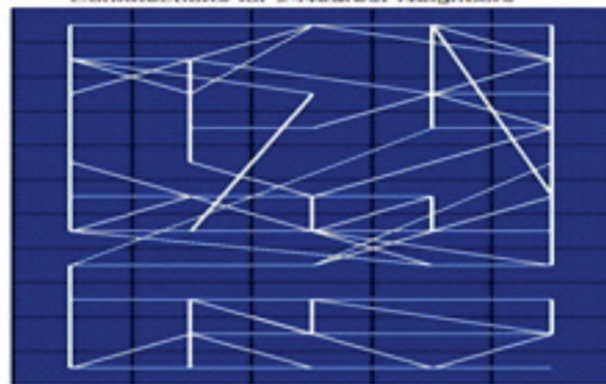


Fuente

Discusión

La implementación de mapas auto-organizados, permitió la creación una estructura alternativa para la clasificación de los pacientes de la muestra tomada de la población colombiana y la realización de un análisis exploratorio de los datos asociados a dicha muestra, que de manera coherente con la hipótesis planteada en la introducción del presente trabajo, generó diferentes estructuras a partir de los datos, aumentando la comprensión de la muestra analizada.

Así mismo, teniendo en cuenta que los porcentajes de riesgo de ECV se utilizaron únicamente con fines comparativos y no para el diseño ni para el entrenamiento del SOM, los resultados obtenidos mediante la aplicación de



ción del modelo de Framingham, debido a que no garantiza la detección de subclases contenidas dentro de las clases enmarcadas en cada intervalo de 1% de riesgo de ECV, ni es sensible a la identificación de características similares entre pacientes con porcentajes de riesgo de ECV distintos, que fueron detectadas en la muestra mediante la aplicación de mapas auto-organizados de Kohonen.

Adicionalmente, debido a que la muestra usada para generar la clasificación, presenta una

proporción considerablemente mayor de pacientes con porcentaje de riesgo de ECV de 1%, comparada con la cantidad de pacientes con porcentaje de riesgo mayor que 1%, el método de clasificación demuestra una inferencia deseable, al permitir la detección de estas características propias de un subgrupo pequeño de la muestra, con lo cual se esperaría que en una muestra con valores de porcentajes de riesgo de ECV con una mayor distribución de frecuencias, se genere una clasificación, con clases mejor definidas y la posible aparición de nuevos clusters dispuestos para ser analizados desde el punto de vista médico.

Se propone la adición de nuevos factores de riesgo descubiertos recientemente y la actualización constante de la clasificación de manera adaptativa, por medio de la realimentación con historia clínica electrónica, teniendo en cuenta las ocurrencias de ECV en la muestra como condiciones preliminares para la elaboración de un modelo de predicción de riesgo de ECV, y de esta manera generar una alternativa complementaria al modelo de Framingham.

Finalmente, se requiere de una muestra de tamaño considerablemente mayor y que sea representativa, para que las observaciones realizadas sobre la muestra en el presente trabajo puedan ser validadas y los resultados generalizados a una población.

Agradecimientos

El autor expresa su agradecimiento a García R. G. por sus aportes durante la elaboración de este trabajo y a la Fundación Cardiovascular de Colombia por el soporte clínico.

Referencias

Conroy, R. M., Pyorala, K., & Fitzgerald, A. P. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project; SCORE project group. *European Heart Journal*, 24(11):987-1003.

García, R. G., Zarruk, J. G., Guzmán, J. C., Barrera, C., Pinzón, A., Trillos, E., López-Jaramillo, P., Morillo, C.A., & Tomaz, C. (2012). Sex differences in cardiac autonomic function of depressed Young adults during orthostatic stress. *Biological Psychology*, 90, 179-185.

Guzmán, J. C., García, R. G., Dillemburg, R., Silva, F.A., López-Jaramillo, P. & Morillo, C. A. (2006). Central serotonergic response to orthostatic challenge in patients with neurocardiogenic syncope. *Europace*. 8, 306-311.

Guzmán, J. C., García, R. G., Silva, F. A., Jaimes, H., López-Jaramillo, P. & Morillo, C.A. (2007). Efectos de la terapia de reemplazo hormonal sobre la función autonómica en mujeres postmenopáusicas. *Revista Colombiana de Cardiología*, 14, 9-17.

Jackson, R., Kerr, A., & Wells, S. (2010). Is estimating lifetime cardiovascular risk useful?

British Medicine Journal, 341,73-79.

Kohonen, T. (1981). Automatic formation of topological maps of patterns in a self-organizing system. En E. Oja & O. Simula (eds.), *Proceedings of 2SCIA*. Simposio llevado a cabo en el Scandinavian conference on image analysis, Helsinki, Finland.

Kohonen, T. (1982). Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59-69.

Kohonen, T. (2001). *Self Organizing Maps*. New York: Springer.

Kohonen, T. (2001). *Self-Organizing Map*. Springer, Berlin, Heidelberg. Third, Extended Edition. 1995.

National Heart, Lung, and Blood Institute. *The Framingham Heart Study*. 1923-2011.

Oja, M., Kaski, S. & Kohonen, T.(2003). Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum, *Neural Computing Surveys*, 3, 1-156.

Paynter, N. P., Chasman, D. I., Pare, G., Buring, J. E., Cook, N. R., Miletich, J. P. & Ridker P. M. (2010) Association between a literature-based genetic risk score and cardiovascular events in women. *Journal of American Medicine Association*, 303, 631-637.

Paynter, N. P., Chasman, D. I., Pare, G., Buring, J.E., Cook, N.R., Miletich, J.P., Ridker, P.M. (2010). Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 2010; 303:631-7.

Pöllä, M., Honkela, T. & Kohonen, T. (2009), Bibliography of Self-Organizing Map (SOM) Papers: 2002-2005 Addendum. *TKK Reports in Information and Computer Science*, Helsinki University of Technology, Report TKK-ICS-R23.

Pözlbauer,G., Rauber, A. & Dittenbach, M. (2005). Graph projection techniques for self-organizing maps. En *Proceedings of XIIIth European Symposium on Artificial Neural Networks*, Bruges (Belgium).

Sloan, R. P., Shapiro, P. A., Bagiella, E., Myers, M. M., & Gorman, J. M. (1999). Cardiac autonomic control buffers blood pressure variability responses to challenge: a psychophysiological model of coronary artery disease. *Psychosomatic Medicine*, 61:58-68.

Ultsch,K., Siemon,H.& Kohonen, T. (1990). *Self-organizing feature maps for exploratory data analysis*. International Neural Network Conference.

Vesanto, J., Himberg, J., Alhoniemi, E. & Parhankas, J. (2000). *Organizing Map in Matlab: the SOM Toolbox*. Laboratory of Computer and Information Science Helsinki University of Technology. Finland.

