

## **Sistema experto basado en casos para el diagnóstico de la hipertensión arterial**

### **Case based expert system for hypertension diagnosis**

*Santiago Cuadrado Rodríguez<sup>1</sup>, Emilio F. González Rodríguez<sup>2\*</sup>, Haydee Curbelo Hernández<sup>3</sup>, Yaquelin Luna Carvajal<sup>4</sup>, Gladys Casas Cardoso<sup>5</sup>, Iliana Gutiérrez Martínez<sup>6</sup>*

<sup>1</sup>Facultad de Ingeniería Eléctrica, Universidad Central de Las Villas, Carretera Central No. 208 e/Jesús Menéndez y Danielito. Santa Clara, Cuba.

<sup>2\*</sup>Facultad de Ingeniería Eléctrica, Universidad Central de Las Villas, Calle Caridad No.1 (oeste) e/Cuba y Villuendas. Santa Clara, Cuba.

<sup>3</sup>Facultad de Medicina; Universidad Médica de Villa Clara; Santa Clara, Calle Delicias No.2 (altos) e/Síndico y Caridad. Santa Clara, Cuba.

<sup>4</sup>Vicedirección general Programa de Integración y Desarrollo; Salud, Calle D No. 20 e/ 1ra y Carretera de Camajuaní. Santa Clara, Cuba.

<sup>5</sup>Facultad de Matemática-Física-Computación, Universidad Central de Las Villas, Calle Martí No.80 (oeste) e/ Zayas y Esquerra. Santa Clara, Cuba.

<sup>6</sup>Facultad de Matemática-Física-Computación, Universidad Central de Las Villas, Calle Caridad No. 80 e/Colón y Maceo. Santa Clara, Cuba.

(Recibido el 15 de octubre de 2010. Aceptado el 14 de julio de 2011)

#### **Resumen**

El artículo describe un sistema experto basado en casos para el diagnóstico de la hipertensión arterial (HTA) en la ciudad de Santa Clara, Cuba, realizado en el marco de un estudio para conocer la incidencia de la enfermedad en esta población.

La muestra de la población estudiada estaba formada por 455 hombres y 394 mujeres, entre 18 y 78 años de edad. Los individuos fueron clasificados en normotensos (personas con presión arterial normal), prehipertensos (personas en riesgo de padecer HTA) e hipertensos.

Se realizó un procesamiento estadístico en el que se emplearon técnicas multivariadas como el Análisis Discriminante y la Regresión Logística cuyos

---

\* Autor de correspondencia. Teléfono: + 53 + 42 + 281351 ó 209227, correo electrónico: eglez@uclv.edu.cu. (E. F. González)

resultados, junto a los del Método del Triángulo de Füller, fueron utilizados en el sistema para jerarquizar los factores de riesgo de la HTA y obtener el grado de importancia (peso) de estos. Por medio de la técnica de segmentación CHAID se pudo reducir las comparaciones entre los casos haciendo más eficiente este proceso. La obtención de las funciones de comparación por rasgos para las variables continuas se obtuvo de la aplicación conjunta de un análisis de varianza (ANOVA) y el método TwoStep Cluster Analysis.

Todo esto permitió construir la función de semejanza para la comparación entre el nuevo caso a diagnosticar y los casos de la base. La adaptación de la solución de los casos más semejantes se realizó con la aplicación del algoritmo de los k-vecinos más cercanos.

El sistema experto fue validado finalmente y se comprobó una efectividad en el diagnóstico del 96%.

----- **Palabras clave:** sistema experto basado en casos, hipertensión arterial, técnicas estadísticas.

### **Abstract**

This paper describes a case based expert system for hypertension diagnosis in Santa Clara city, Cuba, carried out as part of an investigation to know the incidence of the disease in this population.

The studied population's sample was formed by 455 men and 394 women, with ages between 18 and 78 years. The cases were classified in three groups: normal (people with normal blood pressure), with prehypertension (people that has risk to suffer hypertension) and people with hypertension.

A statistical process was made with this sample in which multivariate techniques as Discriminant Analysis and Logistic Regression were used. The results of these techniques and the application of the Fuller's Triangle Method were used in the system for ranking the hypertension risk factors and for obtain their importance grade (weigh). CHAID technique allowed reduces the comparisons of the cases making more efficient this process. With the application of the analysis of variance (ANOVA) and TwoStep Cluster Analysis the features comparison function for continuous variables was obtained.

With all this information the likeness function was build to compare the new case with the rest of cases of the case base. The adaptation of the solution of the most similar cases was carried out with the application of nearest k-neighbors algorithm.

Finally the system was validated and was checked an effectiveness of 96%.

----- **Keywords:** case based expert system, hypertension, statistics techniques.

## Introducción

La hipertensión arterial (HTA) es una enfermedad crónica caracterizada por un incremento continuo de la presión sanguínea en las arterias, también conocida como presión arterial.

La presión arterial (PA) se clasifica en base a dos tipos de medidas: la sistólica y la diastólica, las

cuales son expresadas, en ese orden, en forma de razón matemática, como por ejemplo, 120/80 y se miden en milímetros de Mercurio (mmHg). Cuando la medida de la presión sistólica o diastólica está por encima de los valores aceptados como normales, el individuo puede clasificarse en la categoría de prehipertenso o hipertenso. La clasificación de la PA para adultos puede verse de forma detallada en la tabla 1. [1, 2]

**Tabla 1** Clasificación de la presión arterial para adultos

<b>Categoría</b>	<b>PA sistólica (mmHg)</b>	<b>PA diastólica (mmHg)</b>
<i>Normotenso (con cifras de PA normales)</i>	Menos de 120	Menos de 80
<i>Prehipertenso (en riesgo de padecer HTA)</i>	120-139	80-89
<i>Hipertenso</i>		
<i>Grado I</i>	140-159	90-99
<i>Grado II</i>	160-179	100-109
<i>Grado III</i>	180 y más	120 y más

Nota: Cuando las cifras de PA sistólica o diastólica caen en diferentes categorías se toma la más elevada para clasificar la lectura.

La HTA se diagnostica fundamentalmente cuando en tres ocasiones diferentes se observan cifras mayores o iguales a 140/90 mmHg [3]. Sin embargo, debido a la naturaleza multicausal de esta enfermedad, su diagnóstico no debe limitarse simplemente a la toma de la PA sistólica y diastólica, sino que deben tomarse en cuenta además, el comportamiento de un grupo de factores de riesgo de los cuales se hará mención más adelante.

La HTA, además de representar por sí misma una enfermedad, constituye un factor de riesgo para enfermedades cardiovasculares, cerebrovasculares y renales. Se considera, por tanto, uno de los problemas de salud pública más importantes y su prevalencia a nivel mundial se estima entre el 21 y el 30% de la población adulta general [4-6].

En encuestas realizadas en Cuba a nivel nacional esta prevalencia se ha estimado en alrededor de un 30% para zonas urbanas [1, 7]. La alta

incidencia de la HTA en la población cubana se ha podido corroborar de manera particular en la ciudad de Santa Clara en la que desde hace varios años se viene realizando una investigación con este propósito por un comité multidisciplinario de expertos.

Tomando como base esta investigación previa, teniendo en cuenta la comprobada alta prevalencia de la HTA en la población y considerando además la diversidad de factores de riesgo a tener en cuenta en esta enfermedad, se comenzó a construir una herramienta de apoyo al médico comunitario con el objetivo de asistirlo en el diagnóstico de esta patología.

### **Sistema experto con razonamiento basado en casos**

El uso de herramientas de Inteligencia Artificial (IA) para la toma de decisiones en el campo de la salud se encuentra muy generalizado

actualmente. Esto se debe a que en este tipo de dominio del conocimiento la información que se maneja resulta difícil de formalizar por el alto grado de subjetividad asociado. Dentro de las herramientas de IA más utilizadas están las tecnologías basadas en el conocimiento y entre estas pueden mencionarse los Sistemas Basados en Reglas, las Redes Neuronales Artificiales, los Sistemas de Inferencia Borrosos y los Sistemas Basados en Casos [8-11]

Debido a las características del problema que se presenta y por el hecho de disponer de una base de casos diagnosticados por expertos médicos, es que se decide aplicar la técnica del razonamiento basado en casos pues esta apoya sus predicciones en ejemplos (casos) y en el que cada nuevo

individuo a diagnosticar se compara con la base de referencia.

En el presente trabajo se muestran los fundamentos de un Sistema Experto Basado en Casos para el diagnóstico de la HTA, el cual se basa en el uso de técnicas estadísticas [12-14].

## Experimentación

La muestra de casos estudiados por el comité de expertos estuvo formada por 455 hombres y 394 mujeres, entre 18 y 78 años de edad, de la ciudad de Santa Clara, Cuba. Dicho comité confeccionó y aplicó un cuestionario (historia clínica) a la muestra estudiada, el cual incluyó un grupo de variables que se describen en la tabla 2

**Tabla 2** Descripción de las variables consideradas por el comité de expertos en el estudio

<i>Variable</i>	<i>Símbolo</i>	<i>Descripción</i>
Edad	<i>edad</i>	-
Sexo	<i>sexo</i>	Femenino - Masculino
Raza	<i>raza</i>	Negra – Mestiza – Blanca
Índice de masa corporal	<i>IMC</i>	$IMC = \text{Peso(kg)}/\text{Talla(m)}^2$
Antecedentes patológicos personales	<i>APP</i>	<i>asma</i> (Sí - No), <i>angina</i> (Sí - No), <i>infarto</i> (Sí - No), <i>insuficiencia cardíaca</i> (Sí - No), <i>enfermedad cerebrovascular</i> (Sí - No), <i>enfermedad renal</i> (Sí - No), <i>diabetes</i> (Sí - No), <i>dislipidemia</i> (Sí - No), <i>insuficiencia arterial</i> (Sí - No).
Antecedentes patológicos familiares de HTA	<i>APF-HTA</i>	<i>madre, padre, abuelos, tíos</i>
Hábito de fumar	<i>fuma</i>	Sí - No
Ingestión de alcohol	<i>alcohol</i>	Sí - No
PA sistólica basal	<i>PAS</i>	-
PA diastólica basal	<i>PAD</i>	-
PA sistólica 1er min	<i>PAS1</i>	-
PA diastólica 1er min	<i>PAD1</i>	-
PA sistólica 2do min	<i>PAS2</i>	-
PA diastólica 2do min	<i>PAD2</i>	-
PA media 2do min	<i>PAM</i>	$PAM = (PAS2 - PAD2)/3 + PAD2$

Continuación Tabla 2

<i>Variable</i>	<i>Símbolo</i>	<i>Descripción</i>
Ácido úrico	<i>acido_urico</i>	-
Creatinina	<i>creatinina</i>	-
Glucemia	<i>glucemia</i>	-
Colesterol Total	<i>ColTot</i>	-
Colesterol HDL	<i>ColHDL</i>	-
Colesterol LDL	<i>ColLDL</i>	-
Triglicéridos	<i>trigliceridos</i>	-
Estrés psicológico	<i>estrés</i>	Bajo – Medio - Alto

Las cifras de PA se tomaron en el marco de una prueba denominada Prueba del Peso Sostenido (PPS). La *PAS* y la *PAD* se obtuvieron en estado de reposo antes de comenzar la prueba.

La PPS es una variante de una prueba denominada “handgrip”, que fundamenta su principio en analizar la reactividad vascular frente a un ejercicio físico isométrico y demostró ser de gran utilidad como método diagnóstico para estudios masivos de HTA debido a su sencillez y alto valor predictivo. Analizando los datos recogidos en la historia clínica y el criterio de los expertos, los individuos fueron clasificados en normotensos, prehipertensos e hipertensos.

En la medida de la PA se utilizó el esfigmomanómetro digital OMRON modelo M6 COMFORT con el brazalete apropiado para cada caso. Los análisis complementarios fueron realizados en el laboratorio clínico del Hospital Docente Celestino Hernández de Villa Clara, con un equipo automatizado marca HITACHI. [15- 17].

### **Análisis estadístico de la muestra**

En el análisis de la muestra se utilizó el paquete estadístico SPSS 13. Entre los métodos de análisis de la estadística multivariada estuvo el Análisis Discriminante, la Regresión Logística y la técnica de segmentación CHAID. Las dos primeras permitieron validar el criterio de los

expertos con respecto al grado de importancia o peso de cada variable para el diagnóstico, lo cual fue utilizado en la construcción del sistema experto; mientras que la técnica CHAID agrupó los sujetos de la base de casos, atendiendo a valores comunes en ciertas variables. En algunos casos esta agrupación clasificó directamente a los individuos, lo cual le aportó eficiencia y rapidez al programa. El criterio de comparación entre las variables continuas se obtuvo de la aplicación conjunta de un análisis de varianza (ANOVA) y el método TwoStep Cluster Analysis.

### **Sistema basado en casos para el diagnóstico de la HTA**

Un Sistema Basado en Casos para la Toma de Decisiones es un Sistema de Decisión  $S$  como el que se presenta en (1), y se define en términos de un par  $(U, X \cup Y)$  donde  $U$  es un conjunto finito no vacío de objetos, eventos llamados casos, mientras  $X$  y  $Y$  son dos conjuntos finitos, no vacíos, de atributos ó propiedades llamados rasgos predictores y objetivos respectivamente.

$$S = (U, X \cup Y) \quad (1)$$

Cada rasgo predictor  $x_i \in X$  puede ser considerado una función, como la que aparece en (2a), que mapea elementos de  $U$  en el conjunto

$M_i = \{x_{i1}, x_{i2}, \dots, x_{i\eta_i}\}$ , el cual se denomina conjunto de valores del rasgo predictor  $X_i$ .

$$x_i : U \rightarrow M_i \quad (2a)$$

De manera análoga cada rasgo objetivo  $y_j \in Y$  puede ser considerado una función que mapea elementos de  $U$  en el conjunto

$N_j = \{y_{j1}, y_{j2}, \dots, y_{j\theta_j}\}$ , que se denomina conjunto de valores del rasgo predictor  $y_j$ . Dicha función puede verse en (2b)

$$y_j : U \rightarrow N_j \quad (2b)$$

Las componentes de un Sistema Basado en Casos son: la base de conocimientos o base de casos, el módulo de recuperación de casos y el módulo de adaptación de las soluciones.

### Base de casos

La base de casos contiene las experiencias o casos con los cuales el sistema realiza sus inferencias y puede ser representada a través de una tabla, cuyas columnas son etiquetadas por variables o atributos que representan los rasgos predictores y objetivos, mientras que sus filas representan los casos. En la tabla 3, la cual representa la estructura de la base de casos de nuestro problema, el conjunto  $U$  está formado por los casos  $O_1, O_2, \dots, O_m$ , el conjunto  $X$  por los atributos  $\{x_1, \dots, x_n\}$ , mientras que el conjunto  $Y$  lo integra el atributo  $y_1$ .

**Tabla 3** Representación de la estructura de la base de casos

	<b>Rasgos predictores</b>			<b>Rasgo objetivo</b>
Casos	$x_1$	...	$x_n$	$y_1$
$O_1$	$x_1(O_1)$	...	$x_n(O_1)$	$y_1(O_1)$
$O_2$	$x_1(O_2)$	...	$x_n(O_2)$	$y_1(O_2)$
...	...	...	...	...
$O_m$	$x_1(O_m)$	...	$x_n(O_m)$	$y_1(O_m)$

En nuestro problema, el conjunto  $U$  estaría integrado por los 849 sujetos de la muestra estudiada, el conjunto  $X$  lo integrarían los factores de riesgo (rasgos predictores) incluidos en la historia clínica y el conjunto  $Y$  estaría formado por el diagnóstico asignado a los casos, el cual, como se ha visto, puede ser normotenso, prehipertenso o hipertenso.

### Módulo de recuperación

El proceso de recuperación consiste en determinar los casos de la base más semejantes a cada nuevo caso. Para escoger una medida de semejanza entre los casos, se han desarrollado varias técnicas. La más sencilla de ellas cuenta el número de rasgos predictores similares entre ambos. Otra técnica utiliza un conjunto de heurísticas que determinan las variables de mayor importancia (peso) en la determinación del rasgo objetivo y se formula una función que teniendo en cuenta esto considera la semejanza o diferencia entre cada uno de los rasgos predictores.

A continuación se describe el algoritmo del proceso de recuperación.

Entrada:  $O_0, O_t$  ( $O_0$ : nuevo problema y  $O_t$ : caso de la base)

Salida:  $\beta(O_0, O_t)$  (Medida de semejanza o distancia entre  $O_0$  y  $O_t$ )

- Para cada rasgo predictor  $x_i$  se requiere:
  - Buscar los valores  $x_i(O_0)$  y  $x_i(O_t)$  (valores del rasgo  $x_i$  en los casos  $O_0$  y  $O_t$ )
  - Calcular una medida de la semejanza  $\delta_i(x_i(O_0), x_i(O_t))$  entre estos valores.
- Tomar en consideración el peso  $W_i$  del rasgo predictor  $x_i$  y  $\delta_i(x_i(O_0), x_i(O_t))$  a través de una función  $f$  como la que se especifica en (3):

$$f\left(\left(w_i, \delta_i(x_i(O_0), x_i(O_t))\right)_{i=1}^n\right) \rightarrow \beta(O_0, O_t) \quad (3)$$

Este algoritmo contiene dos funciones denotadas por  $\delta_i$  y  $f$ , que se describen a continuación:



Para un rasgo predictor  $x_i$ ,  $\delta_i$  establece la semejanza o distancia entre dos valores  $x_i(O_0)$  y  $x_i(O_t)$ , de ese rasgo. La función  $\delta_i$  aparece definida en (4).

$$\delta_i : M_i \times M_i \rightarrow [0,1] \quad (4)$$

La función de semejanza  $f$  tiene la forma que se presenta en (5)

$$\begin{aligned} & \beta(w_i, \delta_i(x_i(O_0), x_i(O_t))) \\ &= \frac{\sum_{i=1}^n w_i \cdot \delta_i(x_i(O_0), x_i(O_t))}{\sum_{i=1}^n w_i} \end{aligned} \quad (5)$$

Las funciones de comparación por rasgos  $\delta_i$  se obtuvieron mediante los siguientes criterios:

- Cuando la variable es discreta, como es el caso de la variable *sexo*, la función de comparación se define como aparece en (6):

$$\begin{aligned} & \delta_i(x_i(O_0), x_i(O_t)) \\ &= \begin{cases} 1 & \text{si } x_i(O_0) = x_i(O_t) \\ 0 & \text{en otro caso} \end{cases} \end{aligned} \quad (6)$$

- Cuando la variable sea continua, se utilizan los resultados del análisis de varianza (ANOVA), que indica la cantidad de grupos homogéneos que se pueden formar por cada una y luego se emplea ese número, para formar clusters mediante el método TwoStep Cluster Analysis.

### Módulo de adaptación

Después de la determinación de los casos más semejantes, las soluciones contenidas en ellos pueden usarse directamente como solución al nuevo problema, pero comúnmente necesitan ser modificadas. Este paso se conoce como proceso de adaptación. Existen diversas maneras

de realizar este proceso. En el sistema se utiliza el algoritmo de los k-vecinos más cercanos considerando un solo rasgo objetivo  $y_1$ , que en nuestro caso sería el diagnóstico. El algoritmo se describe a continuación.

Sean  $K = \{R_1, \dots, R_k\} \subseteq U$  los k-vecinos más cercanos al nuevo problema  $O_0$ , la probabilidad  $P_{O_0, y_{1t}}$  de que  $O_0$  tome el valor  $Y_{1t}$  en el rasgo objetivo  $Y_1$  se define como aparece representado en (7):

$$P_{O_0, y_{1t}} = \frac{\sum_{R \in K} \partial_{R,t} \cdot \beta(O_0, R)^2}{\sum_{R \in K} \beta(O_0, R)^2} \quad (7)$$

donde  $\partial_{R,t}$  es definida de la manera que se presenta en (8)

$$\partial_{R,t} = \begin{cases} 1 & \text{si } y_1(R) = y_{1t} \\ 0 & \text{e.o.c.} \end{cases} \quad (8)$$

Entonces  $Y_1(O_0)$  recibe el valor con más alta probabilidad. [18- 24].

### Resultados y discusión

El Análisis Discriminante jerarquizó los factores de riesgo en el siguiente orden: *PAM*, *PAS2*, *PAD2*, *PAS1*, *PAD1*, *PAS*, *PAD*, *IMC*, *alcohol*, *dislipidemia* y *raza*. Para diferenciar los prehipertensos de los normotensos, la Regresión Logística consideró el siguiente orden: *sexo*, *fuma*, *PAS1*, *PAD1*, *PAS2*, *PAD2* y *PAM*. Sin embargo, para distinguir los prehipertensos de los hipertensos el orden fue: *sexo*, *raza*, *enfermedad renal*, *PAS*, *PAD*, *PAS1*, *PAD1*, *PAS2*, *PAD2*, *PAM*, *IMC*, *acido\_urico* y *ColHDL*. La técnica de segmentación CHAID consideró la *PAM* y el *IMC*, como las variables más importantes para predecir la HTA.

En la tabla 4 aparecen los factores de efectividad (cifras porcentuales) de estas técnicas.

**Tabla 4** Factores de efectividad de las técnicas multivariadas

	<i>Normotenso</i>	<i>Prehipertenso</i>	<i>Hipertenso</i>
A Discriminante	85,7	93,8	85,1
R Logística	98,4	87,4	92,7
CHAID	99,8	92,7	95,5

Con la aplicación del Método del Triángulo de Füller, un método de comparaciones por parejas para tomar decisiones multicriterio a partir de expertos, se pudo obtener la importancia o peso de los rasgos predictores en la determinación del rasgo objetivo [25]. La tabla 5 muestra este resultado en el que las variables han sido ordenadas de forma descendente según el valor de su peso. Estas variables fueron las que finalmente se tomaron en cuenta en la construcción del sistema basado en casos.

**Tabla 5** Variables obtenidas con el Método del Triángulo de Füller y sus pesos

<i>Variable</i>	<i>Peso</i>
PAM	0,69
PAS	0,63
PAD	0,63
PAS2	0,63
PAD2	0,63
PAS1	0,61
PAD1	0,58
IMC	0,53
Diabetes	0,53

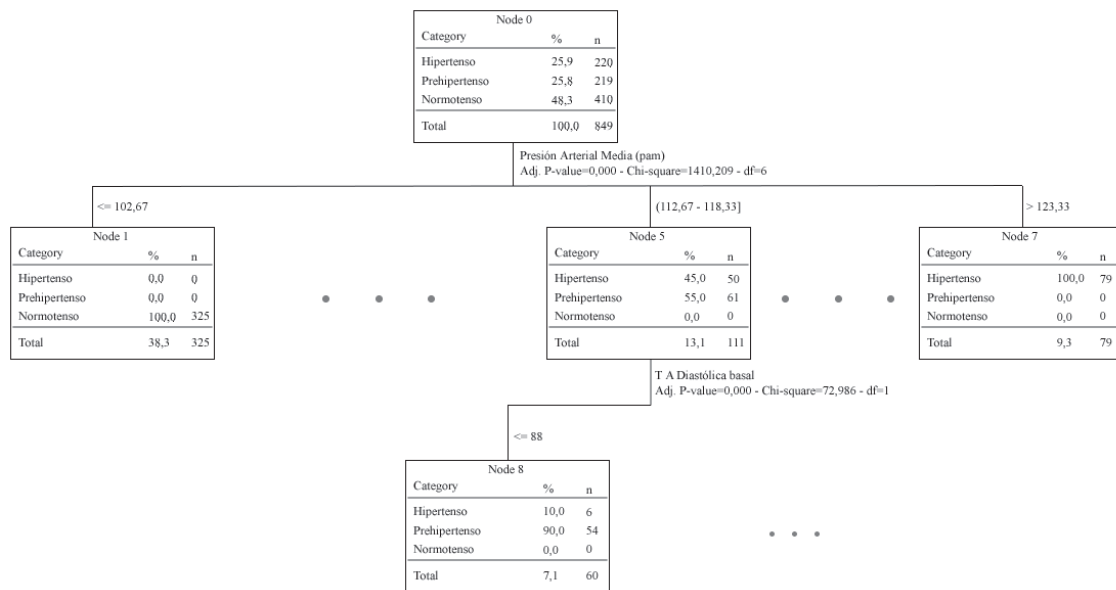
Continuación Tabla 5

<i>Variable</i>	<i>Peso</i>
Padres con HTA	0,53
Abuelos con HTA	0,53
Dislipidemia	0,50
ColTot	0,45
ColLDL	0,42
ColHDL	0,40
Glucemia	0,34
Triglicéridos	0,34
Estrés	0,32
Alcohol	0,29
Sexo	0,24
Raza	0,24
Fuma	0,21

La técnica de CHAID agrupó a los casos de forma que cada nodo, en la estructura jerárquica de árbol, contiene los casos que comparten el mismo rango de valores para una misma variable o grupo de variables. En la figura 1 aparece una versión simplificada de la estructura del árbol obtenida.

Este resultado se utilizó en la implementación del sistema, pues garantizaba que cada nuevo caso a diagnosticar, una vez ubicado en el nodo correspondiente del árbol, solo fuera comparado con los casos de ese nodo. Esto permitió reducir notablemente el número de comparaciones a realizar y aportó gran eficiencia al programa al optimizar el número de comparaciones. En la tabla 6, obtenida del fichero de salida del SPSS, se muestran los resultados de clasificación de la técnica CHAID.





**Figura 1** Versión simplificada de la estructura jerárquica de árbol obtenido por la técnica CHAID

**Tabla 6** Resultados de clasificación de la técnica de CHAID

Observed	Predicted			
	Hipertenso	Prehipertenso	Normotenso	Percent Correct
Hipertenso	210	9	1	95,5%
Prehipertenso	10	203	6	92,7%
Normotenso	0	1	409	99,8%
Overall Percentage	25,9%	25,1%	49,0%	96,8%

Growing Method: CHAID  
Dependent Variable: nuevodiag

Como ya se mencionó anteriormente, las funciones de comparación por rasgos  $\delta_i$  para las variables continuas, se construyeron con ayuda del paquete estadístico. Por ejemplo, en el caso de la variable *PADI* (PA diastólica 1er min) para el sexo femenino, el análisis de varianza (ANOVA) sugirió la formación de 4 grupos homogéneos. Esto aparece reflejado en la tabla 7 (salida del SPSS).

Con este resultado se aplicó el método TwoStep Cluster Analysis, cuyo resultado se puede observar en la tabla 8 (salida del SPSS) en el que se comprueba efectivamente la formación de 4 clusters.

**Tabla 7** Análisis de varianza en el caso de la variable *PADI* para el sexo femenino

		N	% of Combined	% of Total
Cluster	1	137	30,1%	30,1%
	2	87	19,1%	19,1%
	3	144	31,6%	31,6%
	4	87	19,1%	19,1%
	Combined	455	100,0%	100,0%
Total		455		100,0%

**Tabla 8** Aplicación del método TwoStep Cluster Analysis

		<b>PA Diastólica (al 1er minuto)</b>	
		<b>Mean</b>	<b>Std. Deviation</b>
Cluster	1	89,10	1,971
	2	67,80	4,080
	3	79,81	1,044
	4	101,11	5,585
	Combined	84,39	11,412

De esta manera los dos valores a comparar se ubicaron en el clúster correspondiente siguiendo el criterio de la menor distancia de estos a la media de cada clúster. Luego se utilizó la siguiente función de comparación que aparece en (9) [26-30].

$$\delta_i(x_i(O_0), x_i(O_t)) = 1 - \left( \frac{|C_i(x_i(O_0)) - C_i(x_i(O_t))|}{\frac{\sum_{j=1}^k c_{ij}}{k}} \right) \quad (9)$$

Donde  $C_i(x_i(O_0))$  representa la media del cluster al que pertenece el valor  $x_i(O_0)$ ;  $C_i(x_i(O_t))$  representa la media del clúster al que pertenece el valor  $x_i(O_t)$  y  $\frac{\sum_{j=1}^k c_{ij}}{k}$  el promedio de las medias de los clústers obtenidos de esa variable  $x_i$ .

De forma análoga se construyeron las restantes funciones de comparación por rasgos para finalmente aplicar la función de semejanza  $\beta$  representada en 5.

El primer paso para realizar el diagnóstico consistió en ubicar el nuevo caso en el nodo

terminal correspondiente del árbol (ver figura 1). Esta ubicación, en ocasiones, permitió clasificar al caso directamente, dependiendo del nodo del árbol del que se tratase. Por ejemplo, los casos ubicados en el nodo 1 (aquellos con *PAM* menor o igual a 102,67), fueron diagnosticados directamente como normotensos, mientras que los casos ubicados en el nodo 7 (aquellos con *PAM* mayor a 123,33) fueron clasificados como hipertensos. Cuando el caso quedó ubicado en otro nodo, se utilizó el algoritmo de recuperación ya descrito anteriormente pero comparando el nuevo caso sólo con los que se encuentran en el nodo donde fue ubicado. Por ejemplo, si el caso fue ubicado en el nodo 8, la comparación se realizaría con los 60 casos que la técnica de CHAID ubicó en ese nodo.

Una vez determinados los casos más semejantes se aplicó el Algoritmo de los k-vecinos más cercanos para determinar el diagnóstico del nuevo caso.

El sistema, que se construyó bajo el nombre de Tensoft III v2.0, fue validado finalmente y alcanzó una efectividad del 96% la cual puede mejorar a medida que la base de casos crezca y se enriquezca.

## Conclusiones

En este artículo se muestran los resultados de aplicar técnicas de inteligencia artificial, como el razonamiento basado en casos, para el diagnóstico de la HTA. Se comprobó la gran utilidad del empleo de técnicas estadísticas en la construcción del sistema experto pues estas contribuyeron a determinar las variables a considerar, así como las funciones de comparación entre los rasgos.

La construcción de una herramienta dedicada al diagnóstico de la HTA contribuye a hacer más eficaz este proceso y asiste de manera significativa al médico de la atención primaria de salud.

Aunque todos estos resultados han sido aplicados al diagnóstico de la HTA, la metodología seguida aquí puede extenderse a cualquier dominio y emplearse donde sea factible el uso de un sistema experto.

## Referencias

1. [http://es.wikipedia.org/wiki/Hipertensi%C3%B3n\\_arterial](http://es.wikipedia.org/wiki/Hipertensi%C3%B3n_arterial). Consultado el 23 de septiembre de 2010.
2. M. D. P. Caballero, A. D. Herrera, J. A. Guerra, A. V. Vigoa. *Hipertensión Arterial. Guía para la prevención, diagnóstico y tratamiento*. J. Quesada (editor). 1ª ed. Ed. Ciencias Médicas. La Habana, Cuba. 2008. pp. 9-22.
3. <http://www.drscope.com/cardiologia/pac/arterial.htm>. Consultado el 13 de Septiembre de 2010.
4. R. D. Feldman, G. Y. Zou, M. K. Vandervoort. "A simplified approach to the treatment of uncomplicated hypertension. A cluster randomized controlled trial". *Hypertension*. Vol. 53. 2009. pp. 646-653.
5. D. L. Clement, M. L. De Buyzere, D. A. De Bacquer. "Prognostic value of ambulatory blood-pressure recordings in patients with treated hypertension". *J Med*. Vol. 348. 2003. pp. 2407-2415.
6. K. Björklund, L. Lind, B. Zethelius. "Prognostic significance of 24-h ambulatory blood pressure characteristics for cardiovascular morbidity in a population of elderly men". *Hypertension*. Vol. 22. 2004. pp. 1691-1697.
7. [http://scielo.sld.cu/scielo.php?pid=S0864-21251996000200007&script=sci\\_arttext](http://scielo.sld.cu/scielo.php?pid=S0864-21251996000200007&script=sci_arttext). Consultado el 13 de septiembre de 2010.
8. J. L. Kolodner. "An Introduction to Case-Based Reasoning". *Artificial Intelligence Review*. Vol. 6. 1992. pp. 3-34.
9. F. Mariel. "CBR: A categorized bibliography". *The Knowledge Engineering Review*. Vol. 9. 1994. pp. 1-36.
10. A. Aamodt, E. Plaza. "CBR: foundational issues, methodological variations and systems approach". *AI Communications*. Vol. 7. 1996. pp. 1-22.
11. W. Dubitzky. *Knowledge Integration in Case-Based Reasoning: A Concept-Centred Approach*. Thesis submitted in application for the degree of Doctor of Philosophy. Faculty of Informatics, University of Ulster, The United Kingdom. 1997. pp. 226-231.
12. I. G. Martínez, R. E. Bello, Y. C. Bergolla. *URS-HTA: Sistema de Razonamiento en condiciones de Incertidumbre para el diagnóstico de la HTA*. II Simposio Internacional de Hipertensión Arterial. Ed. Feijoo. Villa Clara (Cuba). 2004. pp. 2-4.
13. S. C. Rodríguez, G. C. Cardoso. *Diagnóstico y detección de factores de riesgo de la HTA usando técnicas estadísticas*. III Simposio Internacional de Hipertensión Arterial y I Taller sobre Riesgo Vascular. Ed. Feijoo. Villa Clara, Cuba. 2006. pp. 6-8.
14. A. P. de Armas, O. G. Blanco, H. C. Hernández, E. G. Rodríguez, Y. L. Carvajal, S. C. Rodríguez. *Nuevos métodos para el pesquise y el diagnóstico precoz de la hipertensión arterial esencial*. III Simposio Internacional de Hipertensión Arterial y I Taller sobre Riesgo Vascular. Ed. Feijoo. Villa Clara, Cuba. 2006. pp. 2-4.
15. P. Armario, R. Hernández del Rey, M. Martín. "Estrés, enfermedad cardiovascular e hipertensión arterial". *Medicina Clínica*. Vol. 119. 2002. pp. 23-29.
16. R. S. Vasan, M. G. Larson, E. P. Leip, J. C. Evans, Ch. J. O'Donnell, W. B. Kannel, D. Levy. "Impact of High-Normal blood pressure on the risk of cardiovascular disease". *J Med*. Vol. 345. 2001. pp. 1291-1297.
17. M. Benet, A. J. Yanes, L. J. González, J. J. Apolinaire, J. García del Pozo. "Criterios diagnósticos de la prueba del peso sostenido en la detección de pacientes con hipertensión arterial". *Medicina Clínica*. Vol. 116. 2001. pp. 645-649.
18. W. Wilke, R. Bergmann. *Considering Decision Cost During Learning of Features Weights*. Ed. SpringerLink. University of Kaiserslautern, Germany. 1996. pp. 460-472.
19. S. Cuadrado, G. Casas. *Tensoft: Sistema informativo para el diagnóstico de la HTA sobre bases estadísticas*. Tesis presentada en opción al título de Master of Science. Universidad Central de Las Villas. Santa Clara, Cuba. 2006. pp. 26-44.
20. I. Watson, F. Marir. "Case-Based Reasoning: A review". *The Knowledge Engineering Review*. Vol. 9. 1994. pp. 355-381.
21. J. S. Breese, D. Heckerman. *Decision-theoretic case-based reasoning*. Fifth International Workshop on Artificial Intelligence and Statistics. Fort Lauderdale, Florida. January. 1995. pp. 56-63.
22. I. Gutiérrez, R. Bello. "A Decision Case-Based System, that reasons in Uncertainty Conditions". *Lecture Notes in Artificial Intelligence (LNAI 2504)*. Berlin. Ed. Springer Link. 2002. pp. 54-63.
23. L. Ceccaroni. "Introducción a los sistemas basados en el conocimiento". *Inteligencia Artificial*. Vol. 12. 2008. pp. 25-45.
24. I. Watson, F. Marir. *Case-Based Reasoning: A Review*. University of Auckland. New Zealand. Vol. 3. 2000. pp. 355-381.

25. M.Tabucanon. *Multiple criteria decision making in industry*. Ed. Elsevier. Amsterdam. 1988. pp. 233-299.
26. S. Milton. "Statistical Methods in the Biological and Health Sciences". *Statistics*. 3<sup>rd</sup> ed. Ed. McGraw-Hill. New York. Vol. 1. 1999. pp. 586.
27. J. Jacqueline, Meulman, W. J. Heiser. *Manual de SPSS*. Ed. Oxford: Oxford University Press. 2004. pp. 49-90.
28. S. Glantz. *Primer of biostatistics*. 5<sup>th</sup> ed. Ed. McGraw-Hill. Vol. 1. 2002. pp. 468.
29. A. Field. *Discovering Statistics Using SPSS for Windows. Advanced Techniques for the Beginner*. Ed. U.o.S. Daniel B. Wright. Vol. 1. 2003. pp. 496.
30. A. Aron, E. Aron. *Statistics for the Behavioral and Social Sciences. A Brief Course*. 2<sup>nd</sup> ed. Ed Prentice Hall. Vol. 1. 2002. pp. 330.