



**UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS**

**FACULTAD 7**

**Centro de Informática Médica (CESIM)**

Trabajo final presentado en opción al título de  
Máster en Informática Aplicada

**Método computacional para el diagnóstico de la Hipertensión  
Arterial mediante el procesamiento analítico en línea y el  
razonamiento basado en casos**

**Autor:** Ing. Yovannys Sánchez Corales

**Tutor:** Dr. Pedro López Tamayo

**Co-Tutores:** MSc. Yasel Couce Sardiñas

**La Habana, Diciembre del 2012**

## **AGRADECIMIENTOS**

Al escribir los agradecimientos, pensé que por descuido podía dejar a alguien importante fuera de la mención, pues durante este tiempo son muchas las personas que han participado en este trabajo, por eso desde ya, pido las disculpas correspondientes en caso de que eso suceda.

Antes que nada quiero agradecerle a Dios por hacerme tan humilde y permitirme lograr otra meta más en mi vida.

A Yasel Couce por confiar en mí, y escogermme como su tutelado. GRACIAS HERMANO

Los sabios consejos Mirna, quien ha venido guiando mi formación no solamente académica, sino como persona, sin lugar a dudas me han permitido ver en la naturaleza esa combinación de complejidad y sencillez que a la vez se presenta.

Al tribunal y a los profesores que imparten clases en la maestría Informática Aplicada, que han hecho de mí un mejor profesional, pues me enseñaron más que el saber científico, ser lo que no se aprende en salón de clase y a compartir el conocimiento con los demás.

A Matota y César por su esfuerzo.

Todo esto nunca hubiera sido posible sin el amparo incondicional de mi familia, mis padres y mis hermanos. De ellos es este triunfo y para ellos es todo mi agradecimiento.

A Sailyn, Yoenny, Lucía, Eduardo, Eduardito, Adriana, Geovanis, Ramona. Nieves, Lolita, Deinys, Pipo, Roberto y Jose por la confianza que han brindado.

A los profes del proyecto y del CESIM, por su apoyo en todo momento.

A Anyerman, Rafael, Alfre, Franco, Yami, Expósito, Aly, Joseph, Frank, Cary, las reinas, Daniel, Kenia, Yailin y un largo etcétera.

## DEDICATORIA

Reír a menudo y mucho; ganar el respecto de gente inteligente y el cariño de los niños, conseguir el aprecio de críticos honestos y aguantar la traición de falsos amigos; apreciar la belleza; encontrar lo mejoren los demás; dejar el mundo un poco mejor, sea con un niño saludable, una huerta o una condición social redimida; saber que por lo menos una vida ha respirado mejor porque tú has vivido. Eso es tener éxito.

A mis padres ANGELA CORALES MÁRQUEZ Y JOSÉ ANTONIO SÁNCHEZ LUSSON y padrastro RAMÓN CANTILLO FRÓMETA, porque creyeron en mí y porque me sacaron adelante, dándome ejemplos dignos de superación y entrega, porque en gran parte gracias a ustedes, hoy puedo ser máster, ya que siempre estuvieron impulsándome en los momentos más difíciles y porque el orgullo que sienten por mí, fue lo que me hizo ir hasta el final. Va por ustedes, por lo que valen, porque admiro su fortaleza y por lo que han hecho de mí.

A mis queridos hermanos Liubeth y Leyisbeth, les digo:

*Detrás de cada línea de llegada, hay una de partida.*

*Detrás de cada logro, hay otro desafío.*

*Si extrañan lo que hacían, vuelvan a hacerlo.*

*Sigan aunque todos esperen que abandonen.*

*No dejen que se oxide el hierro que hay en ustedes.*

Al grupo de trabajo Centro de Toma de Decisiones del proyecto APS y demás miembros del proyecto.

A los que nunca dudaron que lograría este triunfo: mis tías Lolita, Nieves, Ramona, Papi Rafael (padre), Deisy y demás tíos, primos, sobrinos y grandes amigos.

A mi ahijado Alain Daniel, gracias por hablarme a menudo por teléfono.

A todos, espero no defraudarlos y contar siempre con su valioso apoyo, sincero e incondicional.

A los presentes y a los ausentes.

A la vida.... Por lo aprendido y aprehendido.

## **DECLARACIÓN DE AUTORÍA**

Declaro por este medio que yo Yovannys Sánchez Corales, con carné de identidad 80022622945, soy el autor principal del trabajo final de maestría Método computacional para el diagnóstico de la Hipertensión Arterial mediante el procesamiento analítico en línea y el razonamiento basado en casos, desarrollado como parte de la Maestría en Informática Aplicada y que autorizo a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Y para que así conste, firmo la presente declaración jurada de autoría en La Habana a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_

Yovannys Sánchez Corales

---

## RESUMEN

La Hipertensión Arterial (HTA) es uno de los problemas de salud complejos frecuentes en Cuba. Gran parte de la complejidad es debido a que se muestra de distintas maneras en los pacientes y constituye el principal factor de riesgo en las enfermedades cardiovasculares y cerebrales. Aunque existen varios enfoques inteligentes para el tratamiento de esta enfermedad, en la Atención Primaria de Salud continúan debutando y asistiendo pacientes hipertensos cuyas manifestaciones no son de fácil detección (pacientes complejos), por lo que no es suficiente abordar el diagnóstico de este complicado problema de salud por un único enfoque de la Inteligencia Artificial.

El presente trabajo plantea como objetivo, desarrollar un método computacional para el diagnóstico de la Hipertensión Arterial (HTA) en la Atención Primaria de Salud, de manera que facilite el apoyo en la toma de decisiones clínicas integrando el Procesamiento Analítico en Línea para el análisis de los factores de riesgo de la HTA y el Razonamiento Basado en Casos para la clasificación y proceder de la enfermedad.

El método es validado mediante la construcción de una Herramienta Clínica con Soporte a la Toma de Decisiones (*Clinical Decision Support System - CDSS*) como parte del Sistema Integral para la Atención Primaria de Salud (alás SIAPS). Se utilizó además un clasificador que arroja un 91% de aciertos, así como el análisis de las opiniones de un conjunto de expertos, los cuales concuerdan en un 99% de confianza.

**Palabras Claves:** Hipertensión Arterial, Procesamiento Analítico en Línea, Razonamiento Basado en Casos, Sistema Clínico para la Toma de Decisiones.

## ÍNDICE

<b>INTRODUCCIÓN</b> .....	2
<b>CAPÍTULO I. FUNDAMENTOS TEÓRICOS METODOLÓGICOS DE TÉCNICAS COMPUTACIONES PARA EL DIAGNÓSTICO DE LA HIPERTENSIÓN ARTERIAL.</b> .....	8
1.1 Evolución histórica de los métodos computacionales para el diagnóstico médico .....	8
1.2 Métodos de referencia para el diagnóstico de la HTA.....	9
1.3 Principales métodos utilizados para el diagnóstico médico .....	11
1.4 Relación de razonamiento basado en casos y el diagnóstico médico .....	13
1.4.1 Proceso de Recuperación .....	14
1.4.2 Medidas de similitud .....	15
1.4.3 Tipos de adaptación.....	16
1.4.4 Retención y mantenimiento de la base de casos .....	16
1.4.5 Análisis del RBC para el diagnóstico médico .....	17
1.4.6 Valoración del RBC sobre otras tecnologías empleadas en el diagnóstico médico ..	18
1.5 Relación del OLAP y el diagnóstico médico .....	19
1.5.1 Mercado de Datos.....	20
1.5.2 Modelado de un mercado de datos. Almacenamiento .....	22
1.5.3 Procesos de integración .....	23
1.6 Integración de RBC y OLAP para el diagnóstico médico .....	23
1.7 Consideraciones del capítulo.....	24
<b>CAPÍTULO II. MÉTODO TEÓRICO PARA EL DIAGNÓSTICO DE LA HIPERTENSIÓN ARTERIAL MEDIANTE LA INTEGRACIÓN DEL PROCESAMIENTO ANALÍTICO EN LÍNEA Y EL RAZONAMIENTO BASADO EN CASOS</b> .....	25
2.1 Precondiciones del método .....	25
2.2 Diagnóstico de la hipertensión arterial.....	25
2.2.1 Razonamiento basado en Casos y sus etapas .....	25
2.2.2 Base de casos de pacientes hipertensos. ....	26
2.2.3 Indexación de los casos de pacientes hipertensos .....	29
2.2.4 Recuperación de pacientes semejantes.....	29
2.2.5 Adaptación de pacientes hipertensos semejantes .....	33
2.2.6 Retención de pacientes hipertensos.....	34
2.2.7 Explicación de la inferencia realizada.....	35
2.3 Elementos causales de la hipertensión arterial .....	35
2.3.1 Procesamiento analítico en línea y sus peculiaridades .....	35
2.3.2 Almacenamiento de los factores de riesgo de la hipertensión arterial .....	36
2.3.3 Integración de los factores de riesgo de la hipertensión arterial .....	38
2.3.4 Visualización de los factores de riesgo de la hipertensión arterial .....	39

2.4	Integración y método final.....	40
2.5	Política de respaldo y recuperación de los datos .....	41
2.6	Consideraciones del capítulo.....	41
<b>CAPÍTULO III. SISTEMA CLÍNICO PARA LA TOMA DE DECISIONES.....</b>		<b>43</b>
3.1	Herramienta desarrollada .....	43
3.1.1	Estructura del repositorio clínico del sistema alas SIAPS .....	44
3.1.2	Diagnóstico de la HTA. Toma de decisiones.....	44
3.1.3	Análisis de los elementos causales. Toma de decisiones.....	46
3.2	Validación cruzada estratificada .....	50
3.1.1	Bases de datos utilizadas por el clasificador .....	50
3.1.2	Análisis de resultados del clasificador .....	51
3.2	Validación por comité de expertos.....	54
3.2.1	Aplicación del método.....	56
3.2.2	Análisis de los resultados .....	56
3.2.3	Evaluación de concordancia .....	60
3.3	Consideraciones del capítulo.....	62
<b>CONCLUSIONES .....</b>		<b>63</b>
<b>TRABAJO FUTURO .....</b>		<b>64</b>
<b>BIBLIOGRAFÍA .....</b>		<b>64</b>
<b>ANEXOS .....</b>		<b>73</b>

## INTRODUCCIÓN

La informática en la medicina (1; 2) es una de las aplicaciones más comunes (3; 4; 5) e importantes desde hace varias décadas, lo que ha permitido al sector de la salud, no solo poseer métodos novedosos, sencillos y eficaces de gestión administrativa en consultas, hospitales y centros de investigación biomédica, sino también disponer de complejas aplicaciones que reducen la posibilidad de error en el diagnóstico de las enfermedades y la gestión de las Historias Clínicas Electrónicas (HCE) (1; 6). El diagnóstico en el campo médico es, sin dudas, una de las áreas de la Inteligencia Artificial que supone todavía de un gran desafío.

En el contexto de la Informática Médica surgen los Sistemas Clínicos para la Toma de Decisiones (*CDSS - Clinical Decision Support System*) (7; 8; 9; 1) como componentes disponibles en aplicaciones informáticas sanitarias. Los *CDSS* ofrecen soporte en la toma de decisiones mediante la recolección de información, su análisis y la generación de mensajes de alertas o recordatorios ante situaciones alarmantes. Aportan recomendaciones necesarias para soportar la selección de estudios mediante reglas lógicas o análisis de decisión, la elección del tratamiento, ya sea para ajuste de dosis o seguimiento en situaciones especiales o evaluar el pronóstico de un paciente.

Con la evolución de las Tecnologías para la Información y las Comunicaciones (TIC) gran cantidad de datos recogidos en la HCE han sido estudiados y clasificados a partir técnicas de análisis de datos como el Procesamiento Analítico en Línea (*OLAP - On Line Analytical Processing*) (10). Una de las principales ventajas de la utilización de *OLAP* en el desarrollo de los Sistemas Clínicos de Soporte para la Toma de Decisiones es su rapidez al generar conocimiento ante los posibles sucesos vinculados a los problemas de salud y útiles al orientar algún diagnóstico basado en información que no puede apreciarse a simple vista.

Las tecnologías de ayuda a la toma de decisiones están cada vez más disponibles para los clínicos de medicina general. En los últimos años, se han desarrollado una gran variedad de programas (11; 12; 13; 14), diseñados para ayudar en la dosificación de la medicación, mantenimiento de la salud, diagnóstico y otras decisiones clínicas relevantes.

Como un tipo simplificado de *CDSS* se encuentran los sistemas expertos (15; 16), los cuales brindan asesoría al tomar decisiones altamente especializadas sobre un área en particular. A diferencia de los sistemas expertos, los *CDSS* poseen un enfoque más integral ya que incluyen tareas que requieren de integración con servidores

terminológicos (17), descubrimiento de información y manipulación o interpretación de las HCE para facilitar buenas prácticas basada en la evidencia médica.

Algunos de los sistemas expertos utilizan como método de solución a los problemas la técnica de razonamiento basado en casos (RBC) (18; 19; 20), cuyos pilares son dos suposiciones fundamentales acerca de la naturaleza del mundo:

- Su regularidad. De situaciones similares se extraen conclusiones y se aprenden lecciones similares. En consecuencia, las conclusiones y lecciones que acompañan experiencias previas pueden ser la base de las que correspondan a situaciones nuevas.
- La recurrencia de las experiencias. Es altamente probable que las situaciones futuras sean variantes de las actuales.

Partiendo de la suposición básica de que “problemas parecidos tienen soluciones parecidas” y que “los nuevos problemas son similares a problemas previamente resueltos” (21), la resolución de problemas basada en casos saca partido de las relaciones entre dos tipos de similitud. Esta teoría ha sido ampliamente utilizada para el diagnóstico (22; 23; 24), pues apelar a la experiencia basada en soluciones previas, acerca este tipo de razonamiento al realizado por los humanos.

La salud en Cuba se encuentra al alcance de todos, es gratuita y de acuerdo con Arocha y Castillo (25) se desarrolla a partir de un concepto social que va más allá del diagnóstico de las enfermedades, pues pone en primer lugar la atención de las necesidades de la población en el medio en que se desarrolla. El Ministerio de Salud Pública de la República de Cuba (MINSAP) es el órgano rector del Sistema Nacional de Salud (SNS). El SNS se estructura en tres niveles de atención: nivel primario, secundario y terciario. El nivel primario es el más importante pues es donde asisten la mayor cantidad de pacientes a las consultas (26; 27).

En el seguimiento integral de los problemas de salud en el nivel primario, se detecta con mayor incidencia la Hipertensión Arterial (HTA) (28; 29), la cual constituye el principal riesgo de las enfermedades cardiovasculares. Según el reporte de la Organización Mundial de la Salud (OMS) del 2012 (30), 1 de cada 3 personas en el mundo padece de Hipertensión Arterial; además agrega que 1 de cada 10 personas es diabética y se reflejan en estudios (31; 32) la alta frecuencia con que las personas diabéticas desarrollan previamente una Hipertensión Arterial. El MINSAP por su parte desarrolla un Programa de Control de la Hipertensión Arterial mediante acciones (33; 34; 35) como respuesta a este importante problema de salud.

Varios CDSS han sido propuestos con la finalidad de apoyar el diagnóstico temprano de la HTA, entre los que se pueden destacar: HiperWeb (36) basado en reglas configurables, URS-HTA (*Uncertainty Reasoning System for HTA*) (37) que utiliza un método basado en probabilidades para contemplar la incertidumbre en la base de casos, la red bayesiana a partir de factores de riesgo de la Hipertensión Arterial (38), la cual utiliza la técnica del CHAID (*Chi-squared Automatic Interaction Detector*) para emitir el diagnóstico, Mathcad (39) el cual utiliza una red neuronal artificial para su funcionamiento y Tensoft III (29), el cual emplea un método de diagnóstico basado en técnicas regresión logística y análisis de discriminante.

Las anteriores aplicaciones tienen como única finalidad la orientación del diagnóstico hacia la clasificación de pacientes en prehipertensos, normotensos e hipertensos y no contemplan el tipo de hipertensión que posee el paciente o si existe probabilidad de que algún órgano diana (28) sea dañado. Además no permiten realizar un análisis del comportamiento a lo largo del tiempo de los factores de riesgo que utilizan para su funcionamiento, así como la determinación de lugares donde la enfermedad se presenta como foco de atención. Debido a su concepción de CDSS monolítico, se dificulta la integración con otros sistemas sanitarios. Limitan el diagnóstico solo a pacientes puntuales, restringiendo realizar un diagnóstico con características grupales, aunque establecen un posible diagnóstico de la enfermedad, no proponen el posible tratamiento correspondiente.

Existen aplicaciones informáticas como el Sistema Integral para la Atención Primaria de Salud (alás SIAPS) y el Sistema de Información Hospitalaria (alás HIS), ambas desarrolladas por el Centro de Informática Médica (CESIM) perteneciente a la Universidad de las Ciencias Informáticas (UCI). Estos sistemas comparten un repositorio de HCE que almacena el historial de consultas realizadas por los pacientes (incluyendo los hipertensos) como parte de su funcionamiento, lo cual constituye una motivación para explorar la posibilidad de integrar el diagnóstico desde una perspectiva puntual debido a los numerosos casos de pacientes hipertensos existentes y a la vez permita identificar los brotes y causas que provocan la HTA para posteriormente sugerir acciones preventivas que eviten el inicio de este problema de salud en las personas.

Por lo anteriormente planteado se define como **problema de investigación:** Insuficiente pertinencia de los sistemas informáticos para la toma de decisiones clínicas relacionadas con el diagnóstico de la Hipertensión Arterial en la Atención Primaria de Salud.

Se plantea como **objeto de estudio** el proceso de diagnóstico en la Atención Primaria de Salud.

Se define como **objetivo general** de la investigación: desarrollar un método computacional para el diagnóstico de la Hipertensión Arterial en la Atención Primaria de Salud, de manera que proporcione apoyo en la toma de decisiones clínicas integrando el Procesamiento Analítico en Línea para el análisis de los factores de riesgo y el Razonamiento Basado en Casos para la clasificación y proceder de la enfermedad.

El **campo de acción** se enfoca en los métodos computacionales para el diagnóstico clínico relacionados con la Hipertensión Arterial en la Atención Primaria de Salud.

Los **objetivos específicos** son:

- Elaborar el marco teórico de la investigación identificando las principales tendencias, limitaciones y ventajas de diferentes métodos aplicados en el diagnóstico médico y en particular la Hipertensión Arterial.
- Desarrollar un método que implemente las técnicas del procesamiento analítico en línea y el razonamiento basado en casos orientado al diagnóstico de la Hipertensión Arterial.
- Validar el método propuesto a partir de una herramienta piloto y del criterio de expertos.

La **hipótesis** planteada es la siguiente: La aplicación de un método computacional científicamente fundamentado, que combine el procesamiento analítico en línea y el razonamiento basado en casos orientado al diagnóstico integral de la HTA en la Atención Primaria de Salud, posibilitará una correcta clasificación de la enfermedad y la identificación de los principales factores de riesgos, apoyando la toma de decisiones clínicas.

Como parte del diseño metodológico fue seguida una **estrategia explicativa** pues los conocimientos precedentes acerca del problema han sido suficientes para plantear una hipótesis explicativa y la representación del problema es clara con respecto a la caracterización del fenómeno en sus aspectos externos.

Los **métodos de investigación** utilizados se exponen a continuación:

### **Métodos Teóricos**

- Histórico-lógico para investigar y analizar los orígenes y evolución de los CDSS, con el objetivo de lograr una mejor comprensión del objeto y campo de estudio proponiendo la solución al problema científico planteado.
- Análisis-síntesis en el estudio de las principales teorías, metodologías, técnicas y estándares existentes para la creación de sistemas con soporte para la toma de decisiones clínicas; para el procesamiento de los datos empíricos y en la elaboración de la memoria escrita.
- Inducción-deducción para el análisis y la definición de la hipótesis de la investigación que será verificada o probada en función del estudio de elementos más particulares, de menor nivel de generalidad.

### **Métodos Empíricos**

- Entrevistas a médicos especialistas que trabajan en la Atención Primaria de Salud para conocer su estado de opinión acerca de la población y frecuencia que asisten a sus consultas enfermos de Hipertensión Arterial, así como sus valoraciones sobre la ayuda e importancia de los sistemas computarizados que tienen a su alcance.
- Observación directa del uso, aplicación y resultados de los sistemas computarizados para diagnóstico de la Hipertensión Arterial en los policlínicos UCI.

### **Métodos Estadísticos**

- Estadística descriptiva para el procesamiento de los datos obtenidos como resultado de la aplicación de los métodos empíricos y la presentación de tablas y gráficos.
- Delphi para la selección, aplicación y análisis de encuestas aplicadas a los expertos.

El presente trabajo presenta la integración de un núcleo OLAP y otro RBC de modo que se pueda establecer un diagnóstico general sobre la HTA. Esto permite definir un método computacional de manera que apoye a la toma de decisiones clínicas relacionadas con el seguimiento y control de la HTA a través de distintas acciones de salud en la Atención Primaria de Salud.

La investigación se presenta en un documento que se estructura como sigue:

- **Capítulo I.** *FUNDAMENTOS TEÓRICOS METODOLÓGICOS DE TÉCNICAS COMPUTACIONES PARA EL DIAGNÓSTICO DE LA HIPERTENSIÓN ARTERIAL*, incluye antecedentes y trabajos relacionados así como algunos aspectos esenciales para entender el entorno del problema a resolver; conceptos fundamentales y temas relacionados con la HTA, RBC y OLAP.
- **Capítulo II.** *MÉTODO TEÓRICO PARA EL DIAGNÓSTICO DE LA HIPERTENSIÓN ARTERIAL MEDIANTE LA INTEGRACIÓN DEL PROCESAMIENTO ANALÍTICO EN LÍNEA Y EL RAZONAMIENTO BASADO EN CASOS*. Se exponen cada uno de los procesos que conforman el método propuesto y se detallan cada uno de los elementos que lo conforman.
- **Capítulo III.** *SISTEMA CLÍNICO PARA LA TOMA DE DECISIONES*. Se realiza una certificación de la propuesta de solución basada en una herramienta informática. Es aplicado un clasificador que utiliza la Validación Cruzada Estratificada y el método Delphi para la validación de los expertos. También es realizado un análisis comparativo tomando una parte de la base de conocimiento de CDSS con cuatro bases de datos internacionales.

## **CAPÍTULO I. FUNDAMENTOS TEÓRICOS METODOLÓGICOS DE TÉCNICAS COMPUTACIONES PARA EL DIAGNÓSTICO DE LA HIPERTENSIÓN ARTERIAL.**

### **1.1 Evolución histórica de los métodos computacionales para el diagnóstico médico**

Las tecnologías de ayuda a la decisión están cada vez más disponibles para los clínicos de medicina general. Los problemas relacionados con el diagnóstico médico no pueden resolverse de forma algorítmica, por lo que en muchos casos hay que recurrir a las técnicas de Inteligencia Artificial (40; 41; 42) para la resolución de los mismos. En los últimos años, se han desarrollado para el mercado médico una gran variedad de sistemas informáticos basados en este enfoque y diseñados para ayudar en la dosificación de la medicación, mantenimiento de la salud, diagnóstico y otras decisiones clínicas relevantes. En (43) se describen algunos de ellos.

Los primeros años de la década de los 70 se produjo un gran auge para los investigadores en el campo de la Inteligencia Artificial en la medicina. El éxito inicial de sistemas informáticos como CASNET (44), DENDRAL (45) y MYCIN (46) constituyen ejemplos de herramientas basadas en métodos inteligentes orientados al diagnóstico médico.

#### **Toma de decisiones relacionadas con el diagnóstico médico**

La toma de decisiones es la actividad esencial del clínico. En teoría, tomar una decisión implica crear una lista de posibles estrategias y/o acciones (47), determinar la consecuencia de cada decisión, y seleccionar la solución más apropiada para el contexto. Un problema típico de toma de decisiones es el problema del diagnóstico (19), el cual se define como el proceso de determinar mediante un examen la naturaleza y las circunstancias de una situación de enfermedad (8). Supone una secuencia de tareas interdependientes (48), a menudo altamente individualizadas: hacer referencia a la historia inicial del paciente mediante un análisis de la información asociada reflejada en su historia clínica y a los resultados de un examen físico; evaluar y refinar hipótesis de diagnóstico a través de la información adicional sobre el paciente; indicar un tratamiento acorde al diagnóstico identificado; sugerir acciones preventivas asociadas, entre otras.

## 1.2 Métodos de referencia para el diagnóstico de la HTA

### HiperWeb

**HiperWeb** (36) es un sistema para la toma de decisiones con respecto a la HTA en el ámbito nacional, fue desarrollado en la UCI en el año 2006 y ha sido implementado sobre tecnología Web. Contiene información sobre el Séptimo Reporte del *Joint National Committé* (JNC-7) (49) y la Guía Cubana para el Tratamiento de la Hipertensión Arterial del año 2004. Se estructura de dos módulos: indicación y control del tratamiento no farmacológico de los pacientes hipertensos, así como, un módulo orientado a mejorar el estilo de vida de los pacientes hipertensos. La inferencia de HiperWeb sigue un método basado en reglas de producción. Fue construido utilizando la plataforma .NET.

*Hiperweb* aunque es un sistema basado en reglas configurables, éstas tienden a crecer según evolución de la enfermedad debido a la tendencia en el crecimiento de pacientes complejos en el área de la atención primaria; por lo que la complejidad de la enfermedad dificulta la determinación de nuevas reglas.

### URS-HTA: Sistema de Razonamiento en condiciones de Incertidumbre para el diagnóstico de la HTA

URS-HTA (*Uncertainty Reasoning System for HTA*) (37) es un sistema desarrollado en Universidad Central de las Villas, Cuba, el cual implementa un procedimiento para Razonamiento Basado en Casos basado en técnicas estadísticas para el manejo de la incertidumbre presente en la base de conocimientos. Permite la representación eficiente del conocimiento garantizando eficiencia y fiabilidad en el proceso de diagnóstico. El diagnóstico que brinda sobre un paciente puede ser de dos tipos: hipertenso y normotenso. Fue desarrollado con la plataforma *Borland Delphi*.

El anterior procedimiento aunque ofrece robustez ante la incertidumbre en la base de casos y el acceso de forma jerárquica, es difícil realizar operaciones como la adición y modificación de casos, teniendo en cuenta que en este tipo de representación se desencadenan tareas no triviales que en ocasiones, constituyen en la actualidad un problema de investigación. No incluye soporte para el análisis en línea de los factores de riesgos ni brinda algún tratamiento asociado. El tipo de plataforma utilizada no permite ser incorporado en esta investigación.

### **Tensoft III**

Este método (29) es una propuesta desarrollada en Universidad Central de las Villas, Cuba, en el marco de un estudio para conocer la incidencia de la enfermedad en esa provincia. Se basa en el uso de técnicas estadísticas y tiene como objetivo el diagnóstico de HTA en los pacientes (hipertenso, no hipertenso o prehipertenso). La base de casos es creada a partir de la Historia Clínica PROCDEC (50) creada por un grupo multidisciplinario de esa provincia. Incluye procesamiento estadístico en el que se emplean las técnicas multivariadas como el Análisis Discriminante y la Regresión Logística. La herramienta escogida para su desarrollo fue *Borland Delphi*.

El método que implementa el sistema Tensoft III es dependiente del conjunto de HCE PROCDEC, la cual no sigue estándares médicos internacionales como HL7/-CDA (51), lo que dificulta el uso del método para este trabajo.

### **Red Bayesiana a partir de factores de riesgo de la Hipertensión Arterial**

Esta red (38) es creada en la Universidad Central de las Villas y realiza el análisis a partir de distintos factores de riesgos y atributos que permiten caracterizar los tipos de HTA. A partir de casos con su diagnóstico se obtiene una red bayesiana, la cual utiliza la técnica del CHAID (*Chi-squared Automatic Interaction Detector*) (52) para su funcionamiento. La red bayesiana obtenida es útil para realizar tanto el diagnóstico de la HTA, como para analizar cuál sería el comportamiento de los atributos predictores más significativos. La base de datos está formada por un conjunto de atributos predictivos y un atributo objetivo (diagnóstico) cuyo valor (hipertenso, normotenso o hiperactivo) se infiere a partir del valor de los atributos predictivos.

Este método no garantiza que exista autoaprendizaje, es decir, aumentar la experiencia debido a problemas similares tratados anteriormente; elemento imprescindible a tener en cuenta debido a la complejidad de la enfermedad.

### **Mathcad**

Mathcad (39) implementa una Red Neuronal Artificial (RNA) que utiliza métodos de aprendizaje para modelar una memoria auto asociativa bidireccional. La simulación y modelización del arreglo en la RNA apoya a los diagnósticos generando la matriz de conocimientos de la RNA para determinar la posible predicción de los pacientes hipertensos, partiendo desde una matriz nula tomando como entrada pacientes sanos. El sistema fue implementado con la herramienta *Borland Delphi*.

El método que sigue Mathcad es costoso computacionalmente, tomando como criterio principal las bajas prestaciones de las computadoras que poseen las distintas unidades que prestan servicios en la atención primaria.

### **Valoración general de los métodos de referencia**

Un diagnóstico general de HTA ofrece los tipos de hipertensión según sus características, el nivel a que pertenece y si ha provocado algún problema secundario o ha afectado o no algún órgano diana como corazón, riñones, ojos, cerebro, entre otros; también la clasificación según el grado de afectación de órganos diana (28; 53). Teniendo en cuenta estos elementos, se puede decir que la mayoría de los sistemas anteriores no contemplan en su concepción elementos de lo antes descrito, tampoco el tratamiento dictaminado por Junta Internacional de HTA ante cada situación (JNC-7) (49). Tienen como único propósito la clasificación de pacientes en prehipertensos, normotensos e hipertensos y el problema de salud abordado no es codificado por algún estándar médico internacional de manera que se dificulta el análisis estadístico de la HTA en planos internacionales.

Las aplicaciones desarrolladas con los métodos descritos anteriormente, la mayoría resultaron ser monolíticas e implementadas con tecnologías dependientes del sistema operativo Windows. Algunas están desarrolladas con software privativo, de modo que no cumplen con la política de la soberanía tecnológica que se lleva a cabo el MINSAP, lo que trae consigo que se dificulte la modificación de sus funciones para adaptarlo a la problemática planteada.

### **1.3 Principales métodos utilizados para el diagnóstico médico**

#### **Razonamiento basado en reglas**

El razonamiento basado en reglas (54; 55), proceso complejo y prolongado pues la extracción se refiere a la formalización de reglas y el pensamiento humano no siempre está regido conscientemente por las reglas de la lógica; en ocasiones es básicamente un procesamiento de información recuperada con el tiempo. Las reglas de producción se utilizan como forma de representar el conocimiento y usualmente la búsqueda en profundidad con dirección hacia delante o dirigida por datos y hacia atrás o dirigida por objetivo como método de solución de problemas. Auxilian a médicos generales en el diagnóstico de enfermedades (56; 57) donde los especialistas no se encuentran disponibles. En (58) se combina esta técnica con el RBC.

### **Redes Neuronales Artificiales**

Una red neuronal artificial (RNA), según Freman y Skapura (59), es un sistema de procesadores paralelos conectados entre sí en forma de grafo dirigido. Esquemáticamente cada elemento de procesamiento de la red (neuronas) se representa como un nodo. Estas conexiones establecen una estructura jerárquica que, tratando de emular la fisiología del cerebro, busca nuevos modelos de procesamiento para solucionar problemas concretos del mundo real. La utilidad de las RNA radica en la capacidad de aprender, reconocer y aplicar relaciones entre objetos y tramas de objetos del mundo real. Su comportamiento depende de los pesos de las funciones de entrada-salida (Funciones de transferencia). Generalmente las funciones se clasifican en: función lineal, gatillo y sigmoide (14).

En medicina la aplicación más común de estos procedimientos, es la clasificación de patrones con el propósito de apoyar al médico en el diagnóstico y tratamiento del paciente. En (60) se refleja el éxito de las redes neuronales combinadas con el RBC como apoyo al diagnóstico médico. A pesar de los buenos resultados, el uso de redes neuronales puede no ser práctico, pues recuperar un caso de cientos posibles alternativas haría demasiado grande la red y por consiguiente el proceso de recuperación se torna lento. Además podría ser complicado aprender acerca de los casos dentro de una estructura de conexiones.

### **Arboles de Decisión**

Son propuestos por Howard Raiffa (61) y constituyen una forma de representar el conocimiento obtenido en el proceso de aprendizaje inductivo. Cada registro está formado por el conjunto de valores de las variables predictoras y el valor de la variable resultado que corresponde a cada caso. Esta partición binaria recursiva se plasma en una organización jerárquica del espacio de representación que puede modelarse mediante una estructura tipo árbol. Cada nodo interior contiene una pregunta sobre una variable predictora concreta (con un hijo para cada una de las dos posibles respuestas) y cada nodo hoja se refiere a un resultado (o clasificación). Su utilización se ha extendido a áreas de la biología y la medicina.

### **Algoritmos Genéticos**

Los algoritmos genéticos (AG) (62; 63) se basan en el funcionamiento de la evolución natural para resolver determinadas situaciones, es decir, imitan a la evolución biológica como estrategia para resolver problemas computacionales. Dado un problema

específico a resolver, la entrada del AG es un conjunto de soluciones potenciales a ese problema, codificadas de alguna manera y una métrica llamada función de aptitud que permite evaluar cuantitativamente a cada candidata. Las aplicaciones de los AG en Medicina e Ingeniería biomédica son múltiples, pues han tenido grandes resultados en la identificación del cáncer (64), manipulación de imágenes médicas (65), evaluación de desempeño de fármacos (66), entre otros. Algunos estudios (67; 68) vinculan esta técnica con el RBC.

### **Algoritmos Clínicos**

Los algoritmos clínicos (69; 70), también conocidos como mapas u diagramas de flujo, son representaciones esquemáticas a partir de las cuales se pueden seguir distintos caminos y así llegar a los distintos nodos. Existen los nodos de información (de forma cuadrada) y los nodos de decisión (rombos) donde se toman las decisiones relacionadas con el sendero a seguir. Los algoritmos clínicos han utilizados desde hace varias décadas y de diversas maneras, permitiendo representar el conocimiento en forma gráfica llevando a los usuarios paso a paso. Debido a su sencillez, son relativamente fáciles de codificar.

#### **1.4 Relación de razonamiento basado en casos y el diagnóstico médico**

El Razonamiento Basado en Casos (RBC) es uno de los métodos actuales para construir Sistemas Basados en el Conocimiento (19) ampliamente utilizado en el diagnóstico de enfermedades (71). Un nuevo problema se compara con los casos almacenados previamente en la base de casos (Memoria de Casos) y se recuperan uno o varios casos. Posteriormente se evalúa una solución sugerida, por los casos que han sido seleccionados con anterioridad para tratar de aplicarlos al problema actual (23). A los sistemas informáticos que solucionan nuevos problemas por analogía con los antiguos se les denomina, sistemas de Razonamiento Basados en Casos (18). De acuerdo con Aamodt y Plaza, el ciclo de razonamiento basado en casos consta de cuatro etapas o pasos como se muestra en la figura 1.

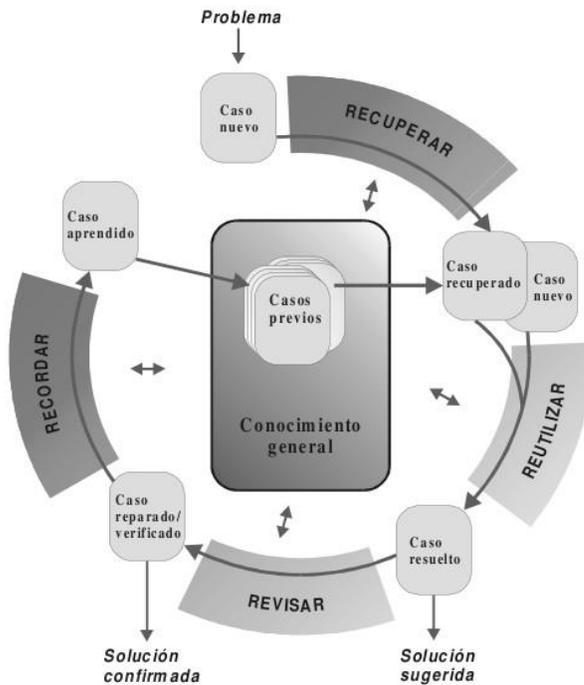


Fig. 1. El ciclo de razonamiento basado en casos (72).

**Recuperar** el caso o casos pasados más similares a la nueva situación presentada.

**Reutilizar** la información y conocimiento de este caso o casos recuperados para resolver el nuevo problema.

**Revisar** cuando una solución para un caso generado en la fase de reutilización no es correcta.

**Retener** las partes de esta experiencia que puedan ser útiles para la resolución de futuros problemas.

El principal activo de un sistema RBC es la base de casos. Según Kolodner (73), las formas de representar la solución, la justificación y el resultado varían según el dominio de aplicación del sistema RBC y la aproximación utilizada. En (74) se distinguen dos grandes tipos de representaciones para los casos: representaciones planas y representaciones estructuradas aunque se pueden encontrar distintos criterios y soluciones (75; 76; 77) en la literatura consultada.

#### 1.4.1 Proceso de Recuperación

En esta etapa se busca en la base de casos, aquellos que sean lo más parecidos al caso actual. Frecuentemente, los casos a recuperar se buscan por casos enteros, comparando sus características con las del nuevo caso (problema), aunque hay ocasiones en que es suficiente comparar solo una porción del caso. Este tipo de situaciones normalmente surgen cuando en la base de casos no existe un caso que coincida completamente con el caso que se está estudiando. El proceso de recuperar un caso es altamente dependiente de la estructura de la base de casos y el método de indexación.

Es importante en la etapa de recuperación la selección del peso de cada atributo o característica, debido a que influye directamente sobre la eficiencia y la precisión del

RBC. La literatura consultada muestra algunos métodos para determinar el peso (78; 79; 80; 81; 82), los cuales han tenido resultados positivos en la recuperación de casos.

Durante la etapa de recuperación se determina el grado de proximidad entre los casos, lo que se conoce como medida de semejanza o similitud (83; 84). La similitud de un paciente hipertenso con respecto a otros se obtiene calculando la similitud de cada rasgo predictivo para llegar a una medida global mediante el método del vecino más cercano (85; 86), k-vecinos cercanos (*knn* - *K nearest neighbors*). Pese a su sencillez, el *knn* arroja buenos resultados experimentales (86; 87; 88; 89) en materia de precisión o coste computacional. Una vez recuperado el caso, se realiza un análisis para determinar si ese caso es suficientemente parecido al problema o si la búsqueda debe ser modificada.

#### 1.4.2 Medidas de similitud

En términos matemáticos, la medida de similitud (*matching*) significa definir un espacio métrico entre los casos  $x$  e  $y$ . Debe cumplirse la existencia de una función  $d(x, y)$  real no negativa y un conjunto  $U$ , tal que se cumplan las siguientes condiciones (83):

Siendo  $d(x, y)$ ,  $U \in \mathbb{R}^+$ .  $\forall x, y \in U$  se cumple:

$$d(x, y) = 0 \Leftrightarrow x = y \quad (\text{Axioma de la separación})$$

$$d(x, y) = d(y, x) \quad (\text{Axioma de la simetría})$$

$$d(x, y) \geq 0 \Leftrightarrow x \neq y$$

$$d(x, y) \leq d(x, z) + d(z, y) \text{ para } z \in U \quad (\text{Desigualdad triangular})$$

Si  $d(x, y)$  es una función de distancia, cuantifica el valor de semejanza de forma inversa (73), es decir, mientras menor sea la distancia, mayor será la semejanza entre los casos.

Las medidas de similitud frecuentemente utilizadas en el RBC son: la distancia Euclidiana (90) normalizada por la importancia (pesos) de los rasgos, la distancia de Hamming (91), la medida de Tversky (92), la de Hunt (93), la medida del coseno, el índice de Jaccard (94), Minkowski (95), Mahalanobis (96), entre otras. En (90) aparecen todas las medidas más utilizadas en este tipo de razonamiento.

En la literatura consultada se han desarrollado varios tipos de medidas. Algunas utilizan lógica difusa (97), enmarcadas en algunos tipos de contextos (98) como la búsqueda de documentos (99). Otras medidas de similitud entre conjuntos difusos se

encuentran en (100; 101; 102). Existen además las medidas híbridas empleadas cuando el caso posee atributos de distintas naturaleza, estudios (103) sobre atributos continuos, discretos o ambos reflejan lo antes planteado.

### 1.4.3 Tipos de adaptación

La adaptación es el proceso de transformar la solución recuperada, en una nueva solución apropiada para resolver el problema. Esta etapa del ciclo de RBC es la encargada de añadir la inteligencia al método. Puede realizarse por el usuario o de manera automática por el sistema. Si la adaptación la realiza el usuario, el sistema sólo realizaría la búsqueda de los casos similares. Si la adaptación la ejecuta el sistema, entonces debe contener algún conocimiento, tal como fórmulas o reglas. Los tipos de adaptación pueden ser (19):

- *Reinstanciación*: este tipo de técnica reutiliza los algoritmos, métodos o reglas usados para generar la solución, pero con nuevos argumentos.
- *Búsqueda local*. Se realizan búsquedas dentro de jerarquías semánticas y se soluciona el problema por analogía.
- *Ajuste de parámetros*: se ajustan los parámetros de la solución de casos anteriores de acuerdo con las diferencias entre las descripciones de los casos en cuestión.

Los siguientes trabajos (104; 105; 106; 107) muestran variantes de realizar el proceso de adaptación. Luego de elegida la estrategia de adaptación hay que comprobar si la solución adaptada tiene en cuenta diferencias entre el caso recuperado y el problema actual, es decir, si realmente resolverá el problema. En este punto también hay que decidir si la solución propuesta no resuelve el problema con éxito. Finalmente se realiza la evaluación del caso, puede ser por un experto o semiautomática para ser reparado en caso de ser necesario.

Es necesario destacar que durante el proceso de recuperación se pueden obtener un conjunto de casos candidatos recuperados. En este caso, o bien se puede seleccionar uno de ellos basado en algún criterio específico o hacer utilizar soluciones conjuntas de todos ellos en la búsqueda de una solución apropiada al caso actual.

### 1.4.4 Retención y mantenimiento de la base de casos

Una vez obtenida la el caso apropiado, se comprueban los resultados obtenidos con los esperados. Para ello se necesita algún método (108) capaz de valorar el

rendimiento de la prueba. Luego es adicionado el caso a la base; de esta forma aumenta la probabilidad de encontrar un caso similar, así como la solución que el sistema es capaz de crear. Se evidencia que el aprendizaje puede ocurrir cuando el caso ha sido solucionado con éxito o cuando interesa tener en cuenta las causas del error cometido. El nivel de éxito o fracaso de la solución y la relevancia del caso son situaciones en que se debe agregar o descartar un caso.

En la labor de mantenimiento se agrega, actualiza, revisa y elimina casos. Se tiene en cuenta el crecimiento de la base de casos, pues se pueden resolver problemas de forma errónea, almacenar datos conflictivos o información redundante, entre otros. Un buen perfilado de los datos como parte del proceso de calidad, facilita mejores razonamiento en el futuro. Algunos estudios (109; 110; 111) enfocados en el mantenimiento, profundizan como llevar esta labor.

Existe cierta relación entre la cantidad de casos retenidos y la eficiencia de la recuperación: mientras mayor es el número de casos, mayor será el número de problemas cubiertos. También es posible que impacte en la calidad del razonamiento el volumen de la base de casos, es por ello que la calidad de los casos cobra importancia durante el mantenimiento de los mismos.

#### **1.4.5 Análisis del RBC para el diagnóstico médico**

La toma de decisiones consiste, básicamente, en elegir una alternativa entre las disponibles. En general, la segunda vez que se intenta resolver un problema es más fácil que la primera, pues su forma de solución es recordada y repetida. Se es más inteligente la segunda vez porque se recuerdan los errores pasados y se intentan evitarlos. Los expertos médicos y en las leyes se sigue este proceder (18).

El proceso diagnóstico médico es similar al funcionamiento del RBC: el médico posee una serie de experiencias almacenadas que se corresponden con las enfermedades de los distintos pacientes a los que ha tratado a lo largo de su carrera. Suponiendo que el recuerdo fue causado por la semejanza de los síntomas importantes entre los pacientes, el médico utiliza el diagnóstico del paciente previo para determinar la patología y el tratamiento del paciente actual.

En el caso específico de la HTA, el razonamiento basado en casos se puede relacionar con el diagnóstico médico debido a que se pueden establecer las analogías: caso - paciente hipertenso, rasgos predictivos - cuadro clínico y rasgos objetivos - elementos

que componen el diagnóstico. Existe además una alta frecuencia de repetición al analizar el cuadro clínico de estos pacientes que, en muchas ocasiones poseen características similares. De la información de los pacientes hipertensos que radica en las HCE, se pueden crear casos concretos y ser empleados como base para prevenir la ocurrencia de la enfermedad de otros pacientes con características similares. Además resulta muy difícil establecer reglas generalizables que permitan inferir las distintas características que posee un paciente hipertenso debido a la complejidad del problema de salud. Por las razones antes expuestas, se decide aplicar como parte del presente trabajo el Razonamiento Basado en Casos.

#### **1.4.6 Valoración del RBC sobre otras tecnologías empleadas en el diagnóstico médico**

El RBC ante otras técnicas utilizadas en el diagnóstico médico presenta las siguientes características (112):

- La rapidez de respuesta del RBC está dada, entre otros factores, por el hecho de que generalmente no se realizan retrocesos (*backtracking*), sino que se recuperan soluciones previas completas en un sólo paso; lo cual acerca este tipo de razonamiento al realizado por los humanos.
- No es necesario entrenamiento previo como las redes neuronales para su funcionamiento y el costo de implementación es pequeño.
- En los sistemas basados en reglas el antecedente de la regla tiene que corresponderse exactamente con la información contenida en la memoria de trabajo mientras que en los sistemas basados en casos los casos pueden igualar la entrada de forma parcial.

Entre las ventajas se encuentran (113):

- La experiencia y conocimientos obtenidos y desarrollados por los expertos, pueden perdurar y crecer en el tiempo de forma indefinida.
- Son útiles en la pedagogía médica, ya que pueden ser consultados por personas u otros sistemas informáticos teniendo la posibilidad de proporcionar una explicación sobre el procedimiento que han seguido para llegar a una solución.
- Están disponibles ininterrumpidamente, ofreciendo siempre su máximo desempeño.

- Su experiencia puede utilizarse como banco de conocimiento en una unidad de salud, lo que permite difundir el conocimiento experto por todo el equipo de salud y entrenar al nuevo personal en materia de diagnóstico.

Algunas de las desventajas que sufre esta técnica son (19; 20):

- Dependiendo del algoritmo implementado, pueden tener una noción muy limitada acerca del contexto del problema, por lo que no pueden percibir todas las cosas que un experto humano puede apreciar de una situación de diagnóstico.
- Puede ser que no recupere el caso más apropiado para la solución del nuevo caso.
- El Razonamiento Basado en Casos depende de una adecuada función de semejanza la cual no es fácil de encontrar para cada aplicación.

### **1.5 Relación del OLAP y el diagnóstico médico**

El Procesamiento Analítico en Línea (*OLAP - On Line Analytical Processing*) (10; 114), es un tipo de procesamiento de datos que se caracteriza, entre otras cosas, por permitir el análisis multidimensional y es empleado principalmente en el campo de la Inteligencia de Negocios. Su principal objetivo es el de brindar rápidas respuestas a complejas preguntas interpretando la situación de un negocio sobre grandes bases de datos (básicamente un almacén o mercado de datos) o sistemas transaccionales y de esa forma poder tomar decisiones (115) al respecto.

Este tipo de procesamiento puede analizar la información desde diferentes escenarios históricos mediante informes estadísticos y proyectar cómo se ha venido comportando y evolucionando en un ambiente multidimensional, o sea, mediante la combinación de diferentes criterios, temas de interés o dimensiones lo que permite deducir tendencias por medio del descubrimiento de relaciones entre los criterios.

#### **Análisis del procesamiento analítico en línea en el diagnóstico médico**

El éxito que se obtiene al emplear esta técnica de análisis e información radica en el objetivo que se persigue. OLAP se utiliza esencialmente en el análisis de la información provocada por los distintos departamentos de las empresas, pero puede ser aplicado para el diagnóstico mediante el análisis de la información contenida en un repositorio (banco de datos) de Historias Clínicas y de esta forma, obtener información acerca del comportamiento histórico o tendencia de enfermedades como la HTA y sus

factores de riesgo, mediante informes generados, así como su proceder médico. OLAP puede utilizar el modo de almacenamiento ROLAP (116) que, ante otros métodos como MOLAP (116) y HOLAP (116), se caracteriza por utilizar las ventajas del modelo relacional arrojando rápidos resultados en las consultas complejas durante la creación de los informes generados.

Al realizar este tipo de procesamiento en línea, de manera rápida se puede obtener información valiosa sobre los factores de riesgos de la HTA, monitorear su comportamiento y detectar alguna anomalía en ellos. De esta forma se realizan acciones médicas de carácter preventivas, las cuales reducen el número de pacientes hipertensos en las consultas y por consiguiente facilitan el proceso de diagnóstico.

### **1.5.1 Mercado de Datos**

Según W. H. Inmon, un Almacén de Datos (AD) (117) tiene como base el modelo multidimensional correspondiente. En dicho modelo los datos se organizan en torno a los hechos (118) sobre un contexto determinado refiriéndose a los datos históricos. Tienen atributos o medidas (119) que pueden verse en mayor o menor detalle según ciertas dimensiones (120). Las medidas responden generalmente a la pregunta “cuánto”, mientras que las dimensiones responderán al “cuándo”, “qué”, “dónde”, entre otros.

De acuerdo con Kimball (118), “...un mercado de datos es un conjunto flexible de datos, idealmente basado en el dato más atómico posible (granular) para ser extraído de las fuentes operacionales y presentado en un modelo simétrico (dimensional), que es más resistente cuando se enfrentan con las más inesperadas consultas de los usuarios...”.

Los Mercados de Datos (*DM - Data Mart*) poseen las características de un AD, incluyendo su arquitectura, pero con un alcance mucho más pequeño; están restringidos a un área de negocio específica, lo que implica que se obtengan de un menor número de fuentes de datos. Se caracterizan por disponer de una estructura adecuada de los datos, para analizar la información al detalle, condición que facilita su análisis desde todas las perspectivas que afecten a los procesos de dicha área. Los datos que contienen pueden ser explorados, agrupados y propagados de diversas formas, y así los usuarios pueden realizar su explotación de acuerdo con sus

necesidades.

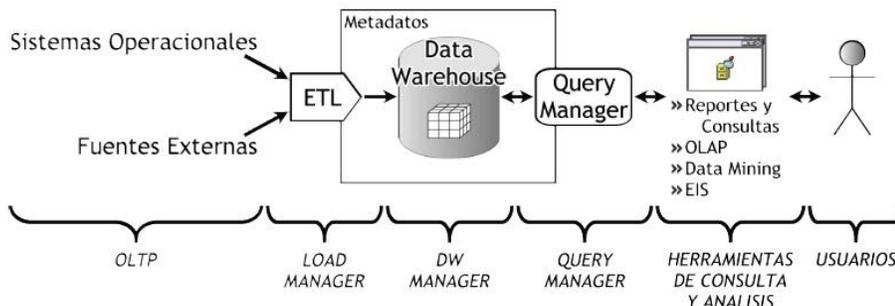


Fig. 2. Arquitectura de un mercado o almacén de datos, tomado de Michael Silver (121)

Existen varias metodologías para la construcción de un DM (117; 122) las cuales guían el proceso de desarrollo e identificación de los subsistemas que conformarán la solución final. Algunas de las metodologías fueron diseñadas incluyendo elementos existentes en otras metodologías (*híbridos*) como es el caso de la propuesta por González (123) la cual combina los enfoques de Kimball (124) y Zepeda (125) para modelar el DM.

### Vinculación de un mercado de datos con el diagnóstico de la HTA

Los DM tienen un grupo de características que pueden ser aprovechadas en el proceso de diagnóstico de la HTA. Entre ellas se encuentran:

- Al tener como principal fuente de datos las historias clínicas de los pacientes, el DM alcanza un enfoque clínico, de tal forma que la información obtenida se relaciona con los problemas de salud asociados a la HTA, factores de riesgos, evolución del paciente, entre otros elementos de interés.
- En la complejidad de la HTA se involucran muchas variables, por lo que la aceleración al ejecutar las acciones de tipo SELECT sobre las historias clínicas tributarán a la rapidez de los informes generados sobre dicha enfermedad.
- Se puede mantener el historial de las consultas (*sin ser alteradas*) de HTA de los pacientes atendidos en las unidades de salud ya que los datos no sufrirán variaciones en el tiempo.
- Al existir un número elevado de pacientes hipertensos atendidos en el nivel primario, el uso de un DM se ajusta mucho a su finalidad, es decir, almacenar grandes cantidades de información que existe en las historias clínicas para luego ser procesada.

- El DM almacena datos pre-procesados en gran medida, por tanto enfermedades como la HTA puedan ser analizadas fácilmente en profundidad y realizar acciones diagnósticas precoces.
- Por su carácter *Integrador*, el DM puede contener información centrada en el paciente e información basada en el conocimiento del mismo; recopilando de esta forma datos que se encuentren en otros sistemas legados, hojas de cálculo como Excel, entre otros que traten la HTA.

### 1.5.2 Modelado de un mercado de datos. Almacenamiento

Las bases de datos multidimensionales implican tres variantes posibles de modelamiento según su complejidad:

- *Esquema en estrella*: consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves.
- *Esquema copo de nieve*: representa una extensión del modelo en estrella cuando las dimensiones se organizan en jerarquías de dimensiones, posibilitando flexibilidad y segregación de los datos relacionados con las dimensiones. Aunque posee mayor complejidad en su estructura, es útil en tablas de dimensiones de muchos registros.
- *Esquema constelación o copo de estrellas*: compuesto por una serie de esquemas en estrellas. Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves con un mínimo esfuerzo adicional de diseño, ya que se pueden reutilizar las dimensiones. Tiene como limitante que no es soportado por algunas herramientas.

Los DM utilizan el modelo multidimensional para el almacenamiento de la información de forma tal que sea de fácil entendimiento por parte del usuario y que a su vez, garantice la rápida ejecución de las consultas. Esta estructura de almacenamiento tiene forma de cubo multidimensional (124), donde las aristas representan las dimensiones y las intersecciones entre estas aristas son los hechos.

- **hechos**: son los elementos primarios del modelo multidimensional y contienen los valores del negocio. Una fila de la tabla de hechos corresponde a una medida que constituye la intersección entre todas las dimensiones;

- **dimensiones:** contienen las descripciones textuales del negocio, juegan un papel vital dentro de los DM ya que son la base de los reportes y consultas ejecutadas;
- **medidas:** son características cualitativas o cuantitativas de los objetos que se desean analizar, que proporcionan valores numéricos importantes para el usuario.

### 1.5.3 Procesos de integración

Los procesos de Extracción, Transformación y Carga de datos (*Extract, Transform and Load - ETL*) (126) se utilizan para integrar datos heterogéneos a un mismo tipo o formato. También se utilizan para la integración con sistemas heredados (aplicaciones antiguas). Kimball identificó además 34 subprocesos (126) que mejoran el funcionamiento de los procesos ETL. Dichos subprocesos constituyen buenas prácticas y son incluidos en el método que se propone en la presente investigación. A continuación se detallan brevemente estos procesos:

**Extracción:** constituye el primer paso para la obtención de la información de las diferentes fuentes hacia el MD. Generalmente los datos provienen de fuentes distintas. La extracción prepara los datos en un formato listo para la transformación.

**Transformación:** una vez que la información es extraída, el proceso de transformación se encarga de preparar los datos de la manera adecuada para integrarlos en el DM. Dicho proceso se compone de actividades como: limpieza de datos, integración de formato e integración de datos.

**Carga:** al extraer los datos de las diferentes fuentes y luego transformados, se procede a cargar el MD, después de la carga inicial, el siguiente paso es mantener el mercado de forma periódica.

### 1.6 Integración de RBC y OLAP para el diagnóstico médico

Durante el cuidado de la salud de un paciente, se utilizan acciones de salud fundamentales como la prevención, el diagnóstico y el tratamiento. A continuación se aborda como un método de integración puede contribuir tener una valoración integral de dichos pacientes.

Las acciones preventivas constituyen una de los enfoques primordiales en el nivel primario de atención. En este sentido la toma de decisiones relacionada con el

diagnóstico clínico puede ser respaldada al utilizar análisis personalizado (RBC) para dar solución a problemas similares y grupales (OLAP) e inferir nuevo conocimiento al respecto. Las acciones relacionadas con la prevención pueden reducir la ocurrencia de nuevos casos de HTA con la identificación del foco de atención y las causas que provocan el problema. Esta fusión permite identificar tareas preventivas como parte del diagnóstico médico, así como proveer criterios para elegir una buena decisión.

El diagnóstico médico constituye uno de los escenarios en que el riesgo es un factor importante. Dicho riesgo puede estar condicionado por el déficit de información, por lo que es importante tener presente diversos criterios mediante distintas vías para poder emitir un diagnóstico más completo. Es por ello que imprecisiones ocurridas durante la prescripción médica como las cascadas diagnósticas y la polimedicación (127) pueden ser reducidas al emplear RBC y OLAP de manera conjunta.

Al atender un paciente hipertenso determinado utilizando RBC, y apreciar que han asistido varios casos de pacientes parecidos de los cuales se han aplicado soluciones similares o que posean algún tipo de relación entre ellos; es posible revisar mediante OLAP el comportamiento en el tiempo del problema de salud así como la asociación con otros problemas comunes presentes en área. Se pueden analizar indicadores como la cantidad de pacientes con la misma enfermedad o identificar las causas que la provocan; de esta forma se reducen consecuencias no deseadas.

### **1.7 Consideraciones del capítulo**

El método que sustenta la aplicación informática *URS-HTA*, aunque es robusto debido a introducción de incertidumbre en la base de casos, es más eficiente cuando las memorias que almacenan los casos son grandes.

El resto de los métodos computacionales estudiados que tributan al diagnóstico de la HTA no satisfacen completamente las necesidades de la investigación, pues solo diagnostican al paciente en prehipertensos, normotenso e hipertenso, además no incluyen análisis causales en los factores de riesgo utilizados.

Se realizó un análisis del RBC y de OLAP quedando justificado el motivo de su elección.

La combinación RBC y OLAP demostraron ser métodos útiles para el diagnóstico integral de la HTA.

## **CAPÍTULO II. MÉTODO TEÓRICO PARA EL DIAGNÓSTICO DE LA HIPERTENSIÓN ARTERIAL MEDIANTE LA INTEGRACIÓN DEL PROCESAMIENTO ANALÍTICO EN LÍNEA Y EL RAZONAMIENTO BASADO EN CASOS**

### **2.1 Precondiciones del método**

Para obtener resultados satisfactorios a través del presente método es necesario comprender las siguientes precondiciones:

- *Estandarización del conjunto de Historias Clínicas Electrónicas:* Es conveniente que las HCE sigan el estándar médico HL7/CDA (51), avalado por el Instituto Nacional Estadounidense de Estándares (*American National Standards Institute - ANSI*) y presente en la mayoría de los sistemas sanitarios, debido a que recoge la mayor cantidad de información; de ahí que quede representado mayor volumen de conocimiento sobre el problema de salud a tratar, sus factores de riesgo así como su relación con otros problemas de salud que con cualquier otro estándar. La razón está en que en dependencia de la información en la HCE, será la calidad del procesamiento analítico y los casos de pacientes hipertensos manejados por el método.
- *Calidad de los datos iniciales:* Se refiere a la no existencia de ruido (126) en los mismos. Para garantizar la calidad de los datos y reducir el ruido existente, se llevan a cabo las tareas de perfilado, limpieza y auditoría (126). Como resultado de este proceso se pueden establecer reglas de negocio o transformación, realizar tratamiento de excepciones o aplicar algún tipo de imputación (128). Kimball, propone clasificar la calidad de los datos en correcta, inequívoca, coherente y completa (126) para un mejor entendimiento de los mismos.

Es necesario destacar que la calidad de los datos no es abordada con profundidad en la presente investigación por cuestiones de alcance a pesar de su importancia para este método.

### **2.2 Diagnóstico de la hipertensión arterial**

#### **2.2.1 Razonamiento basado en Casos y sus etapas**

A continuación se describen las etapas del RBC utilizado en el método para determinar el diagnóstico de HTA basado en el nivel, los tipos de HTA identificados, así

como el tratamiento asociado (general o particular) a dicho problema de salud con la ayuda de especialistas en el tema.

### 2.2.2 Base de casos de pacientes hipertensos.

La base de casos está formada por un conjunto de características (pacientes hipertensos) representados en un vector de características o rasgos predictivos que describen las particularidades del caso y rasgos objetivos los cuales representan el diagnóstico de la HTA basado el tipo de HTA, el nivel, los tratamientos así como las complicaciones asociadas como se muestra en la Tabla 1. Los rasgos objetivos identificados se mencionan a continuación:

- **Tipo de HTA (Tipo):** primaria (P), secundaria acelerada maligna (SAM) y secundaria renovascular (SR). Aunque existen otros tipos de HTA, no son incluidos en esta investigación por las complicaciones que generan.
- **Nivel o grado (Nivel):** niveles en que se ubica la HTA en dependencia del tipo y daño ocasionado. Pueden ser: I (*sin daños*), II (*análisis alterados*) y III (*daño en algún órgano diana*).
- **Tratamiento base (TB):** tratamiento sugerido por la Junta Internacional de HTA (JNC-7) (49) para cada tipo de HTA.
- **Tratamiento personalizado (TP):** tratamiento particular sugerido por el médico para cada tipo de HTA de acuerdo a las características de cada paciente.
- **Complicaciones (Cs):** Complicaciones médicas producidas por tipo de HTA presente en cada paciente.

Tabla 1. Estructura del caso de paciente hipertenso

CASO PACIENTE HIPERTENSO					
Rasgos Predictivos	Rasgos Objetivos				
Conjunto de rasgos predictores	Tipo	Nivel	TB	TP	Cs

Los rasgos predictores se describen a continuación:

- **Sexo/Edad (sexo\_edad):** Son incluidas solo las personas *adultas maduras* y los *ancianos* según la clasificación emitida por la OMS en la presente investigación, debido a que son más propenso a sufrir HTA que las personas de menor edad. La falta de ovarios en las mujeres adultas maduras contribuye al padecimiento de la HTA acelerada. El rasgo *sexo\_edad* es de tipo nominal y toma el valor de “riesgo” si:

- Masculinos mayores de 39 años.
- Femenino entre 20 y 59 años sin ovarios.
- Femenino mayor de 60 años.

En caso contrario el rasgo sexo\_edad toma valor “no riesgo”.

- **Etnia** (*etnia*): Rasgo nominal y se refiere a la etnia del paciente hipertenso. La etnia negra es más vulnerable a padecer de HTA. El rasgo toma los valores negra, blanca o mestiza.
- **Etapas Menopáusicas** (*etap\_men*): Es un rasgo binario que se determina en correspondencia con la edad. Toma valor uno si la edad es mayor que 45 y lleva 12 meses de amenorrea (sin ver la menstruación) que es cuando se considera válida la etapa menopáusica. En caso contrario el rasgo toma valor falso. La amenorrea tributa a la HTA secundaria en muchas de sus variantes y los tipos y causas que la condicionan que, aunque son importantes tratarlas, no son consideradas en este trabajo. Tampoco es considerado la menopausia presente en los hombres.
- **Hábito fumar** (*hab\_fum*): constituye un rasgo numérico y para ello se propone dividir el mismo de acuerdo con la OMS. Para facilitar el trabajo se realiza una clasificación según los tipos de fumadores (activos y pasivos).

En las personas fumadoras pasivas se considera válido el valor del rasgo, entendiéndose si están expuestos al humo en el trabajo, hogar u otra área de gran frecuencia. Mientras que para las activas se clasifican en dependencia de la cantidad de cigarrillos diarios en:

- *No fumadores* (1).
- *Fumadores pasivos* (2)
- *Ligero* (3): Menos de 10 cigarrillos
- *Moderado* (4): Entre 11 y 20 cigarrillos
- *Severo* (5): Entre de 21 y 30 cigarrillos
- *Muy Severo* (6): Mas de 31 cigarrillos

El rasgo toma valores en la escala de 1 – 6 según la clasificación realizada. Las personas ex fumadoras no se consideran en el método.

- **Hipertrofia Ventricular Izquierda** (HVI): Rasgo textual condicionado por la descripción del examen de electrocardiograma (ECG), pues este tipo de

examen contribuye a la estratificación del riesgo cardiovascular del paciente hipertenso.

- **Antecedentes** (*antec*): Constituye un rasgo binario y toma valor verdadero si existe presencia de la HTA en los familiares más cercanos (padres, hermanos, abuelos), en caso contrario toma valor falso.
- **Ingesta de Sal** (*ing\_sal*): Rasgo nominal clasificado por la OMS y la experiencia empírica. Para ello se crea una variable *consumo* la cual toma valor “ingesta normal” si:
  - No se le agrega sal adicional a la dieta común
  - Ingesta menor de 6 gramos diarios

En cualquier otro caso la variable toma valor “ingesta adicional” lo cual constituye un riesgo.

- **centellograma** (*cent\_renal*): Rasgo textual que recoge la información de las estenosis (*hipertensión sistémica resultante de una afectación de la arteria renal*) mediante un centellograma renal, el cual constituyen un método útil en el proceder de las arterias renales. Clásicamente, relacionar directamente la HTA con una lesión arterial depende de la resolución de la HTA después de solventar la obstrucción (129). Uno de los resultados de esta investigación es revertir el modo clásico de dicha relación.

La organización de los casos de pacientes hipertensos responde a un modelo secuencial, útiles para memorias pequeñas. El presente método incluye una memoria de este tipo debido a que no concibe el crecimiento acelerado de la base de casos, pues al constituir el RBC un complemento de OLAP, solo son diagnosticados los pacientes que ya presentan la enfermedad y así evitar que nuevos pacientes se inicien como hipertensos. A continuación se muestra el modelo de datos propuesto para el almacenamiento de casos de pacientes hipertensos:

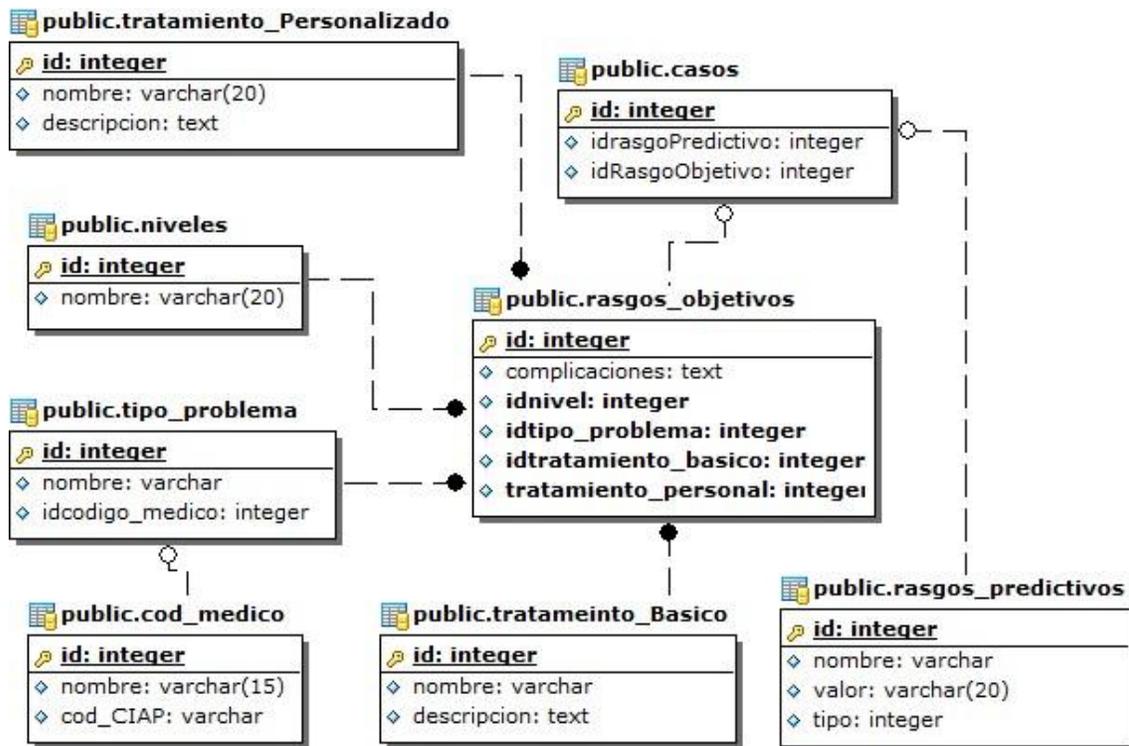


Fig. 3. Esquema de la base de casos de pacientes hipertensos

### 2.2.3 Indexación de los casos de pacientes hipertensos

La indexación propuesta consiste en asignar uno o varios índices (identificación de rasgos predictivos significativos) a cada caso de paciente hipertenso para que sea más fácil la recuperación del mismo. En los estudios (130; 131; 132; 133; 134) existen métodos de elección del índice más adecuado.

En este trabajo es escogido para la base de casos de pacientes hipertensos el tipo de índice *B-tree* sobre el atributo identificador del caso para agilizar las operaciones de unión entre las tablas que almacenan la información del caso. Todos los rasgos son importantes para la recuperación. De esta manera se utiliza como estrategia todos los rasgos del caso único con un rasgo abstracto *id*, garantizando el acceso rápido a la información relacionada con los casos de pacientes hipertensos que residen en las distintas tablas debido a la estructura del índice en forma de árbol balanceado.

### 2.2.4 Recuperación de pacientes semejantes

En esta etapa del razonamiento se definen las métricas de similitud locales (*SimL*) por tipos de rasgos y luego se establece una similitud global (*SimG*) por casos de pacientes

hipertenso escogidas en el método. Las métricas propuestas en este trabajo se exponen a continuación:

### Similitud local entre rasgos de pacientes hipertensos

Entre las medidas mencionadas en el capítulo I, se escogieron las que a continuación se detallan por tipo de rasgos presente en los pacientes hipertensos:

- **Para rasgos numéricos:** Manhattan ajustada (135)

La distancia absoluta o Manhattan se representa como la diferencia absoluta sobre todas las dimensiones. En la mayoría de las aplicaciones, esta genera resultados similares a la distancia euclidiana, pero con más bajo nivel computacional. La expresión que se muestra a continuación corresponde a la distancia Manhattan:

$$d(x, y) = |x - y|$$

La anterior expresión devuelve los valores en dependencia del valor de los parámetros, lo que trae consigo un reajuste para lograr resultados en un rango deseado como se muestra a continuación:

$$SimL(x, y) = 1 - \left| \frac{x - y}{r_{max} - r_{min}} \right| \quad (1)$$

Donde  $r_{max}$  y  $r_{min}$  son los valores de rango máximo y mínimo respectivamente del conjunto de valores existentes.

- **Para rasgos nominales y binarios:** Hamming (91)

Para rasgos o atributos textuales esta medida cuenta el número de palabras diferentes. Como paso previo, dicho texto debe ser depurado, es decir, se realiza una normalización del texto (136; 137) eliminando aquellas palabras vacías que no resulten de interés como artículos, preposiciones, signos de puntuación, entre otros. Por ejemplo sean 2 textos de 4 palabras:

$$T_1 = (p_1, p_2, p_1, p_4) \text{ y } T_2 = (p_1, p_3, p_1, p_4)$$

La distancia entre los textos  $T_1$  y  $T_2$  es 1.

Esta distancia también puede ser utilizada cuando todos los atributos son binarios. La cantidad de atributos que tienen en común los ejemplos, indican los valores de semejanza. Los casos con mayor cantidad de atributos iguales, serán los más similares.

La medida de Hamming para dos casos  $C_x$  y  $C_y$  viene dada por la siguiente expresión:

$$d(C_x, C_y) = \sqrt[r]{\sum_{i=1}^n |x_i - y_i|^r} \quad (2)$$

Donde  $n$  es la cantidad de atributos;  $r=2$  para la distancia de Hamming y los valores  $X_i$  e  $Y_i$  representan el valor del atributo  $i$  para cada caso respectivamente. La expresión queda simplificada para determinar la distancia entre dos atributos de la siguiente manera:

$$d(x, y) = |x - y| \quad (3)$$

$$Si \begin{cases} 0 & si \ x = y \\ 1 & si \ x \neq y \end{cases}$$

Para cuantificar el valor de semejanza de forma inversa en este trabajo, es necesario realizar un ajuste de la expresión quedando de la siguiente manera:

$$SimL(x, y) = 1 - |x - y| \quad (4)$$

$$Si \begin{cases} 0 & si \ x = y \\ 1 & si \ x \neq y \end{cases}$$

- **Para rasgos textuales:** Jaccard (94)

Se define como el tamaño de la intersección dividido entre el tamaño de la unión entre dos conjuntos de datos y su valor está en  $[0, 1]$ . Compara dos conjuntos de palabras en cuanto a la cantidad presente en cada uno por sobre la cantidad total de ellos en ambos conjuntos. Para los casos  $C_j$  y  $C_k$  su fórmula se define como sigue.

$$IJ(C_j, C_k) = \frac{\sum_{l=1}^n i_{jl} i_{kl}}{\sum_{l=1}^n i_{jl}^2 + \sum_{l=1}^n i_{kl}^2 - \sum_{l=1}^n i_{jl} i_{kl}} \quad (5)$$

Donde:

$IJ$ : Índice de Jaccard.

$\sum_{l=1}^n i_{jl} i_{kl}$  : Número de rasgos predictivos comunes de  $C_j$  y  $C_k$

$\sum_{l=1}^n i_{jl}^2$  : Número de rasgos predictivos de  $C_j$

$\sum_{l=1}^n i_{kl}^2$  : Número de rasgos predictivos de  $C_k$

$$Si \begin{cases} 0 & si \ j = k \\ 1 & si \ j \neq k \end{cases}$$

Una forma simplificada de la expresión anterior es la siguiente:

$$SimL(x, y) = 1 - \frac{|x \cap y|}{|x \cup y|} \quad (6)$$

Siendo  $x$  e  $y$ , los conjuntos de palabras mencionados.

En caso que existen valores ausentes en los atributos se procede según Moreno (84): cuando faltan los dos atributos no se lleva a cabo el cálculo de la distancia, pero cuando falta un atributo la distancia se hace 1, también llamada distancia trivial.

### Similitud global entre casos de pacientes hipertensos

La similitud global o función de semejanza para dos pacientes hipertensos ( $P_o$  y  $P_t$ ) se calcula utilizando medidas de similitudes locales y la importancia de cada rasgo predictivo. La expresión para calcular la similitud global se muestra a continuación:

$$SimG(P_o, P_t) = \frac{\sum_{i=1}^n w_i \cdot SimL(P_{oi}, P_{ti})}{\sum_{i=1}^n w_i} \quad (7)$$

Donde  $P_o$  es un nuevo paciente hipertenso,  $P_t$  es un paciente registrado en la base de casos,  $w_i$  constituye el peso de cada rasgo predictivo otorgado por el médico especialista, mientras que  $d(P_{oi}, P_{ti})$  corresponden a las medidas de similitud locales (1), (4) y (6) según el tipo de rasgo predictivo presente en  $P_o$  y  $P_t$  respectivamente. Otro ejemplo de similitud global se encuentra en (19), de acuerdo a la literatura consultada.

### Pasos para la recuperación de pacientes hipertensos

Considerando un valor umbral  $\mu$  tomado como patrón para medir la semejanza y predefinido por el especialista, el algoritmo de recuperación consta de los siguientes pasos:

#### Algoritmo 1. Acceso a los pacientes semejantes

##### Entradas:

- $\mathbf{X}=(x_1, x_2, \dots, x_n)$ : vector que contiene los rasgos predictivos correspondientes a un nuevo paciente hipertenso.
- $\mathbf{Y}=(y_1, y_2, \dots, y_m)$ : lista de los pacientes hipertensos de la base de casos, siendo  $Y_t=(y_{t,1}, y_{t,2}, \dots, y_{t,n})$ , donde  $m$  es la cantidad de casos,  $y_{t,n}$  corresponde a un caso  $t$  con  $n$  atributos pertenecientes a dicho caso.
- $\mathbf{W}=(w_1, w_2, \dots, w_n)$ : vector con los pesos correspondientes a cada rasgo predictivo.

### **Salidas:**

- $\mathbf{S}=(S_1, S_2, \dots, S_k)$  lista de los  $k$  pacientes más similares, con  $\mathbf{S}_k=(S_{k,1}, S_{k,2}, \dots, S_{k,n})$ .

### **Pasos:**

Para cada rasgo predictivo de  $\mathbf{X}_i$  y  $\mathbf{Y}_{j,i}$  perteneciente a la base de casos de pacientes hipertensos.

1. Calcular la semejanza entre los casos,  $SimG(X(x_i), Y_j(y_{j,i})) \geq \mu$  descrita en la ecuación (7) según el tipo de rasgo predictivo y el peso  $w_i$  asociado al mismo.
2. Guardar en una lista  $\mathbf{S}$  los valores de semejanzas obtenidos, así como los rasgos objetivos.
3. Ordenar la lista  $\mathbf{S}$  de forma creciente según el valor de semejanza calculado y obtener solo los primeros  $k$  elementos.

#### **2.2.5 Adaptación de pacientes hipertensos semejantes**

Luego de obtener los  $k$  casos de pacientes hipertensos más semejantes, se procede a la verificación del valor de semejanza de dichos casos con respecto al caso inicial.

El tipo de adaptación utilizada es la nula, en el que se reutiliza el valor del atributo objetivo de los casos semejantes recuperados respecto al nuevo caso.

Si la función de semejanza alcanza un valor de 1 es que los casos son semejantes completamente y el proceso de adaptación se realiza automáticamente a partir del valor máximo de semejanza, sino se realiza el proceso de forma semiautomática debido a que el especialista debe escoger de los casos más semejantes recuperados el que, por su criterio, mejor se ajuste con el nuevo caso.

A continuación se expone el algoritmo de adaptación utilizado por el método:

#### **Algoritmo 2. Adaptación de pacientes semejantes**

##### **Entradas:**

- $\mathbf{X}=(x_1, x_2, \dots, x_n)$ : Vector que contiene los  $n$  rasgos predictivos correspondientes a un nuevo paciente hipertenso.
- $\mathbf{S}=(S_1, S_2, \dots, S_k)$ : Matriz que contiene los vectores con los valores de los rasgos objetivos de los  $k$  pacientes más similares.

**Salida:**

- $Y = (x_1, x_2, \dots, x_n, S_i)$ : Vector que contiene los rasgos correspondientes a un paciente hipertenso adaptado, donde la componente  $S_i$ , se corresponde con el vector de valores de los rasgos objetivos de uno de los  $k$  pacientes más similares.

**Pasos:**

1. Recuperar los  $k$  casos más semejantes al nuevo caso.
2. Determinar si existe al menos un caso  $C_i$  que tenga una semejanza de un 100 por ciento con el nuevo caso ( $SimG(X, C_i)=1$ ):
3. El vector  $S_i$  toma los valores del vector de rasgos objetivos del caso recuperado ( $C_i$ ).

**2.2.6 Retención de pacientes hipertensos**

Luego de aplicar los algoritmos antes descritos comienza la etapa de almacenamiento o retención. Los casos adaptados no son retenidos directamente en la base de casos de pacientes hipertensos, sino que son almacenados temporalmente como casos pendientes para una posterior revisión por los expertos, los cuales tienen la responsabilidad de decidir si tiene sentido o no incorporarlos a la base de casos. El proceso de retención final se muestra a continuación:



**Fig. 4.** Almacenamiento de pacientes hipertensos

A partir del almacenamiento temporal de los casos adaptados, es mostrada la lista de dichos casos a las personas que pueden realizar el almacenamiento en la base de casos hipertensos. Los casos rechazados pasan a formar parte de un registro o *log* de casos rechazados, de modo que se pueda analizar las causas que provocaron el

rechazo. Los casos aceptados se almacenan finalmente en la base de casos de paciente hipertensos.

### **2.2.7 Explicación de la inferencia realizada**

A continuación se procede a describir como se realiza la inferencia en el proceso de recuperación de casos de pacientes semejantes con respecto a un nuevo caso presentado.

El proceso de inferencia consiste en establecer una comparación entre los rasgos predictivos de un caso obtenido de la base de casos con el nuevo caso. Esta operación se realiza rasgo a rasgo teniendo en cuenta su tipo, por lo que la función de similitud local utilizada es diferente para cada par. Basándose en los resultados de las comparaciones anteriormente referidas se calcula el valor de la similitud global, para lo cual se tiene en cuenta la influencia de los pesos de cada atributo. Este proceso se realiza tantas veces como cantidad de casos existan registrados en la base de casos.

Para que un caso sea recuperado como posible solución, el valor de semejanza obtenido debe ser mayor al umbral definido. De esta manera se evidencia como se alcanza al listado de casos semejantes a partir del caso inicial.

## **2.3 Elementos causales de la hipertensión arterial**

Luego de obtener la posible clasificación de HTA y sus elementos asociados en pacientes atendidos, se puede realizar un análisis a través del procesamiento analítico en línea con el objetivo de detectar las posibles causas que provocan la enfermedad en los pacientes asistidos. Si los pacientes clasificados por el RBC siguen patrones determinados como: aumento de la frecuencia de los pacientes en la consulta, igual área de convivencia, iguales apellidos, entre otros, se puede utilizar este tipo de análisis para esclarecer las causas asociadas. En las siguientes secciones se describe como se realiza este tipo de procesamiento analítico.

### **2.3.1 Procesamiento analítico en línea y sus peculiaridades**

Este enfoque se compone de procesos utilizados en la arquitectura de un Data Mart o Mercado de Datos (124) propuesta por Kimball, así como su componente principal: el procesamiento analítico en línea. Los procesos corresponden al almacenamiento, integración y visualización de los datos que, mediante su vinculación, es obtenida la información que permite entender el comportamiento de los principales factores de riesgos de la HTA.

### 2.3.2 Almacenamiento de los factores de riesgo de la hipertensión arterial

En este proceso se modela el ambiente multidimensional, el cual contiene información de la HTA, problemas asociados y factores de riesgo a través de los hechos, las dimensiones (*conceptos*) o temas de interés y las medidas de análisis a partir de una topología Constelación. Los hechos están constituidos por la unión de las dimensiones que responden a factores de riesgos de la HTA en torno al paciente (*h\_paciente*); así como las consultas realizadas por el mismo (*h\_consulta*) como se muestran en las tablas 2 y 3. La tabla 4 describe cada una de las dimensiones propuestas por el método.

Tabla 2. Especificación del hecho *h\_paciente*

Nombre hecho	Descripción		
<i>h_paciente</i>	Almacena las mediciones referentes a los pacientes asistidos en consulta.		
Medidas	Unidad de Medida	Aditividad	Función de agregación
<i>cantidad_problemas</i>	Entero	Aditiva	Suma
<i>cantidad_antecedentes</i>	Entero	Aditiva	Suma
<i>cantidad_habitos</i>	Entero	Aditiva	Suma

Tabla 3. Especificación del hecho *h\_consulta*

Nombre hecho	Descripción		
<i>h_consulta</i>	Almacena las mediciones a las consultas realizadas en la unidad de salud.		
Medidas	Unidad de Medida	Aditividad	Función de agregación
<i>cantidad_consultas</i>	Entero	Aditiva	Suma
<i>cantidad_problemas</i>	Entero	Aditiva	Suma

Tabla 4. Relación de dimensiones identificadas

Dimensión 1: <i>datos_paciente</i>		
<b>Descripción:</b> Esta dimensión almacena los datos relacionados con la etnia y el género que presenta cada paciente atendido en una unidad de salud determinada.		
<b>Frecuencia de actualización:</b> Siempre que se actualicen la etnia y género en el repositorio de historias clínicas electrónicas.		
<b>Jerarquía:</b> No		
Atributo	Tipo	Descripción
<i>id_paciente</i>	numérico	Identificador
<i>grupo_etnico</i>	cadena	Etnia del paciente

genero_paciente	cadena	Género del paciente
<b>Dimensión 2: hábitos personales</b>		
<b>Descripción:</b> Esta dimensión almacena los datos relacionados con los distintos hábitos personales que presenta cada paciente como el hábito de fumar, ingestión de bebidas alcohólicas y el consumo de café. Asimismo se incluyen la fecha en que se adquiere los hábitos y la frecuencia de aplicación de los mismos.		
<b>Frecuencia de actualización:</b> Siempre que se actualicen los hábitos personales identificados en el repositorio de historias clínicas electrónicas		
<b>Jerarquía:</b> No		
<b>Atributo</b>	<b>Tipo</b>	<b>Descripción</b>
id_habito	numérico	Identificador
nombre_habito	cadena	Nombre del hábito del paciente
fecha_comienzo	fecha	Fecha del comienzo del hábito
frecuencia_habito	cadena	Frecuencia del hábito del paciente
descripción	cadena	Descripción del hábito del paciente
<b>Dimensión 3: antecedentes</b>		
<b>Descripción:</b> Esta dimensión almacena los datos relacionados con los antecedentes personales de cada paciente. Se recogen los antecedentes por parentesco de nivel y por tipo de antecedentes ( <i>familiares y personales</i> ). Además se incluye la fecha de detección de dichos antecedentes.		
<b>Frecuencia de actualización:</b> Siempre que se actualicen los antecedentes personales identificados en el repositorio de historias clínicas electrónicas		
<b>Jerarquía:</b> No		
<b>Atributo</b>	<b>Tipo</b>	<b>Descripción</b>
id_antecedente	numérico	Identificador
nombre	cadena	Nombre del antecedente
tipo_antecedente	cadena	Tipo de antecedente
parentesco_antecedente	cadena	Parentesco del antecedente
fecha_deteccion	fecha	Fecha de detección
descripción	cadena	Descripción del antecedente.
<b>Dimensión 4: problema salud</b>		
<b>Descripción:</b> Recoge los problemas de salud detectados al paciente incluyendo la hipertensión arterial. También el código de la enfermedad correspondiente al Clasificador Internacional de Enfermedades (CIE-10).		
<b>Frecuencia de actualización:</b> Siempre que se actualicen los problemas de salud presente en los pacientes los pacientes asistidos en el repositorio de historias clínicas electrónicas		
<b>Jerarquía:</b> No		
<b>Atributo</b>	<b>Tipo</b>	<b>Descripción</b>
id_problema_salud	numérico	Identificador
codigo_cie	cadena	Código de la enfermedad según el codificador internacional de enfermedades
problema_salud	cadena	Nombre del problema
descripción	cadena	Descripción del problema
<b>Dimensión 5: control tiempo</b>		
<b>Descripción:</b> Esta dimensión agrupa información relacionada con la periodicidad y el comportamiento histórico con que se recoge la información relacionada con los factores de riesgos previamente identificados.		
<b>Frecuencia de actualización:</b> Siempre que se actualicen los datos relacionados con el tiempo de los sucesos en el repositorio de historias clínicas electrónicas		

Jerarquía: año- semestre- trimestre- mes- día		
Atributo	Tipo	Descripción
id_tiempo	numérico	Identificador
fecha_hecho_medico	fecha	Fecha en que se produce el evento médico
año_hecho_medico	cadena	Se deriva de la fecha
semestre_hecho_medico	cadena	Se deriva de la fecha
trimestre_hecho_medico	cadena	Se deriva de la fecha
mes_hecho_medico	cadena	Se deriva de la fecha
día_hecho_medico	cadena	Se deriva de la fecha

Se determina además la relación que existe entre dimensiones y hechos a través de una matriz bus o dimensional como se muestra en la tabla 5, y de esta forma evitar solapamientos entre las dimensiones y los hechos. Se puede apreciar como existen tres dimensiones compartidas entre los hechos.

Tabla 5. Matriz bus

Dimensiones \ Hechos	Dimensiones				
	D1	D2	D3	D4	D5
h_paciente	X	X	X	X	X
h_consulta	X			X	X

El resultado final es la conformación física del DM con la estructura correcta para almacenar datos relacionados con los factores de riesgos la HTA, sus afectaciones y otros problemas de salud asociados a la HTA. A continuación se ilustra el flujo de ejecución del proceso de almacenamiento:

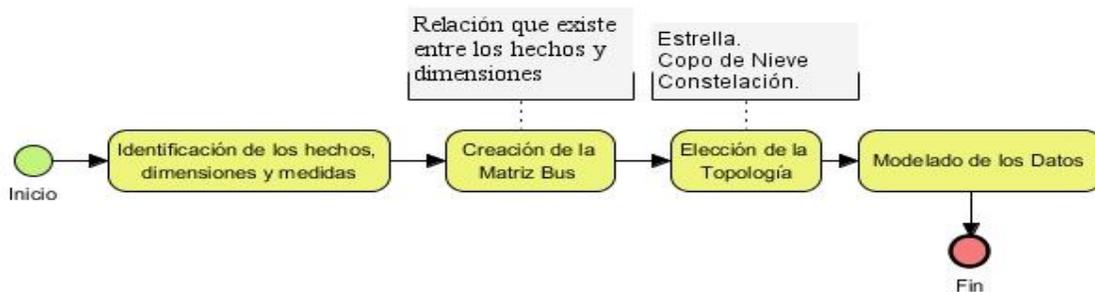


Fig. 5. Proceso de almacenamiento OLAP

### 2.3.3 Integración de los factores de riesgo de la hipertensión arterial

En esta etapa se siguen una serie de pasos que definen el flujo a seguir de cada proceso de integración mediante transformaciones se encargan de extraer y procesar los datos correspondiente a la HTA y sus factores de riesgos del conjunto de HCE. Estas transformaciones finalizan con la carga de los datos en el DM y proceder al análisis. El flujo de integración es descrito como sigue:

En la etapa de estandarización se seleccionan los datos que se les realizarán la transformación. Luego se obtiene la fecha correspondiente a los datos. Los términos del diccionario de datos establecen una correspondencia entre los valores de la fuente y los que maneja el DM, de esta manera se describen cada una de las dimensiones y sus variables, especificando el significado que tienen éstas y los posibles valores que pueden tomar. Se procede a buscar el identificador en la dimensión tiempo, de igual forma en el resto de las dimensiones definidas y se validan seleccionando solo aquellos datos que serán cargados al hecho correspondiente. Los datos que no cumplan con esta validación serán enviados a un archivo de errores. La figura 6 muestra el proceso general de integración:

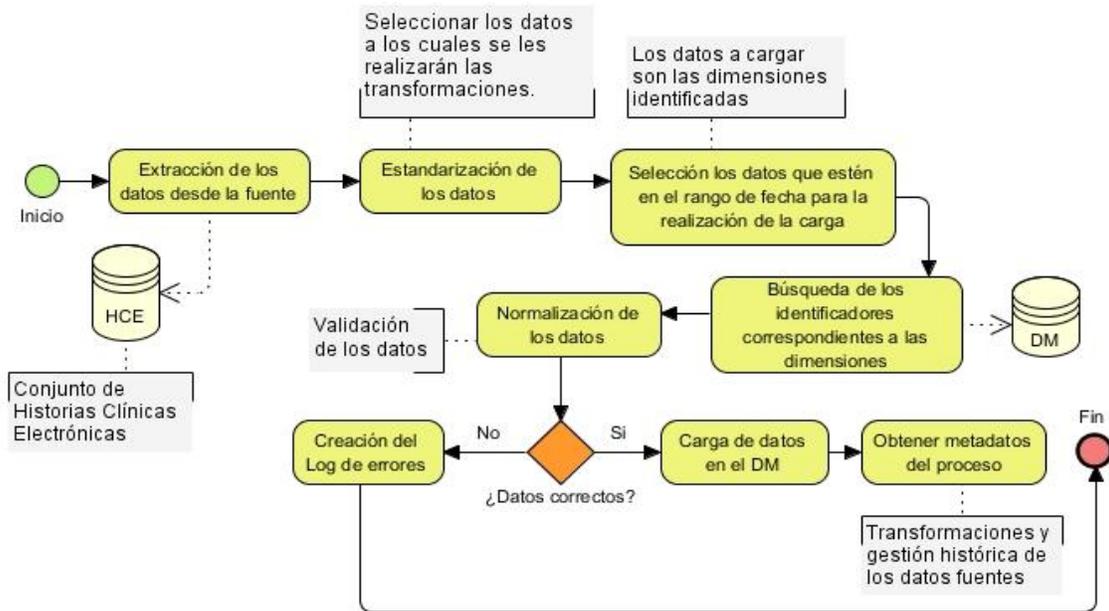


Fig. 6. Proceso de integración OLAP

### Estrategia de carga de los datos

La carga de los datos hacia el DM se hará a través de trabajos (*jobs*), los cuales agrupan las transformaciones y se ejecutan periódicamente. Para las dimensiones lentamente cambiantes (*Slowly Changing Dimensions - SCD*) (126), se utilizan las de tipo 2 para almacenar los cambios producidos con el paso del tiempo y brindar una perspectiva histórica de los factores de riesgos de la HTA.

#### 2.3.4 Visualización de los factores de riesgo de la hipertensión arterial

En este proceso se define la arquitectura de información correspondiente al área de análisis. También se diseñan los informes o reportes que serán visualizados con la

información correspondiente a los factores de riesgo de HTA. Finalmente se construyen los cubos OLAP asociados a los hechos identificados y se especifican las jerarquías de cada dimensión. La figura 7 muestra el flujo correspondiente al proceso de visualización:

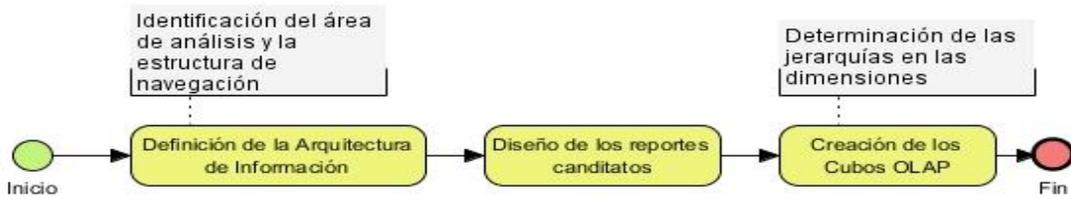


Fig. 7. Proceso de visualización

## 2.4 Integración y método final

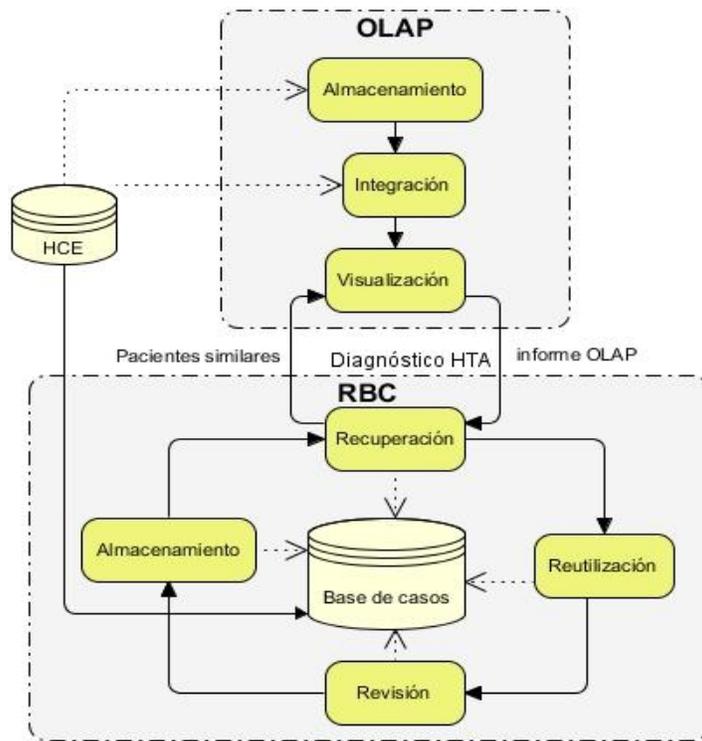


Fig. 8. Integración de RBC y OLAP

Se puede apreciar como a partir de RBC se puede utilizar OLAP para descubrir las causas y otros temas de interés relacionado con la enfermedad. De esta manera el diagnóstico de la HTA es aglomerado con acciones preventivas y de tratamiento.

Una vez determinado el resultado de un paciente mediante el RBC se puede revisar mediante OLAP si el área de convivencia de ese paciente constituye un foco de atención y tomar las acciones de salud al respecto.

## **2.5 Política de respaldo y recuperación de los datos**

El método propone llevar a cabo una política de respaldo y recuperación basada en la periodicidad de las salvadas realizadas. Lo anterior es evidenciado después de cada carga al DM, así como de la base de casos de pacientes hipertensos; de las cuales se verificará que exista una copia escrita de la estructura de la base de casos, las tablas dimensiones y las tablas hechos almacenada en el servidor. Las salvadas realizadas no serán reemplazadas o sobrescritas, debido a que se mantendrá el historial realizado hasta el momento. El estado de cada salva es chequeado mensualmente mediante pruebas de rendimiento, las cuales son condicionadas por su ejecución y puesta en marcha. Cada salva será un archivo compactado y el acceso al mismo será mediante una contraseña como medida de seguridad.

## **2.6 Consideraciones del capítulo**

Es necesario realizar un análisis de calidad de los datos sobre las HCE para garantizar buenos resultados durante el proceso de integración OLAP y la confección de los casos de pacientes hipertensos.

Fue elegida la estrategia de indexación *B-tree*, utilizando como atributo el identificador de caso tributa al acceso rápido y por ende optimiza el proceso de recuperación del RBC.

Para el cálculo de similitud local se evaluaron las medidas de distancias usadas para tipos de atributos numéricos, nominales y textuales.

Se realizó un ajuste de una función de Manhattan de modo que el resultado oscile en un rango de valores determinado.

La retención de los casos de pacientes hipertensos es realizada mediante un flujo de trabajo, permitiendo identificar los casos exitosos de los rechazados.

El presente método incluye en el núcleo RBC una representación de los casos de manera plana, lo que puede atentar contra proceso de recuperación a medidas que crece la base de casos.

La matriz bus permite describir la trazabilidad de las dimensiones y los hechos, lo que evita que existan dos o más hechos que compartan exactamente las mismas dimensiones.

Fueron diseñados los procesos fundamentales del método lo que permitió tener la concepción general del sistema y con ello facilitar los procesos para su posterior implementación.

## CAPÍTULO III. SISTEMA CLÍNICO PARA LA TOMA DE DECISIONES

### 3.1 Herramienta desarrollada

Se desarrolló un Sistema Clínico para la Toma de Decisiones (*CDSS - Clinical Decision Support System*) que implementa el método propuesto como parte del sistema alas SIAPS aunque también puede ser incorporado al sistema alas-HIS como se muestra en la figura 9. Es utilizado el repositorio de HCE de dicho sistema además de otras historias clínicas escogidas de los consultorios pertenecientes al policlínico “Ernesto Che Guevara”, ubicado en la Universidad de las Ciencias Informáticas (UCI), con la participación de médicos enfermeras que laboran en esas unidades de salud.

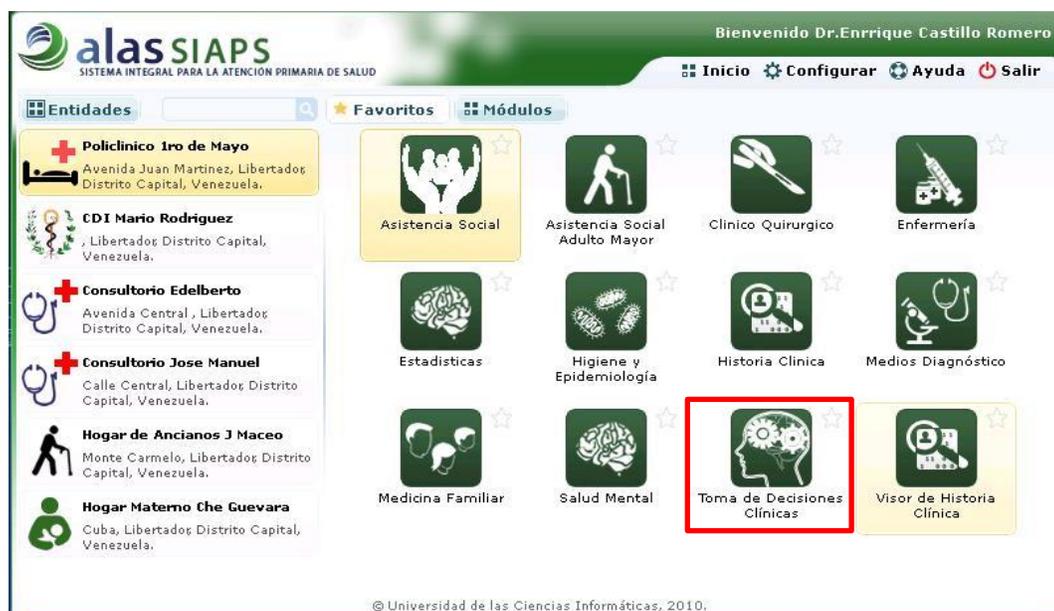
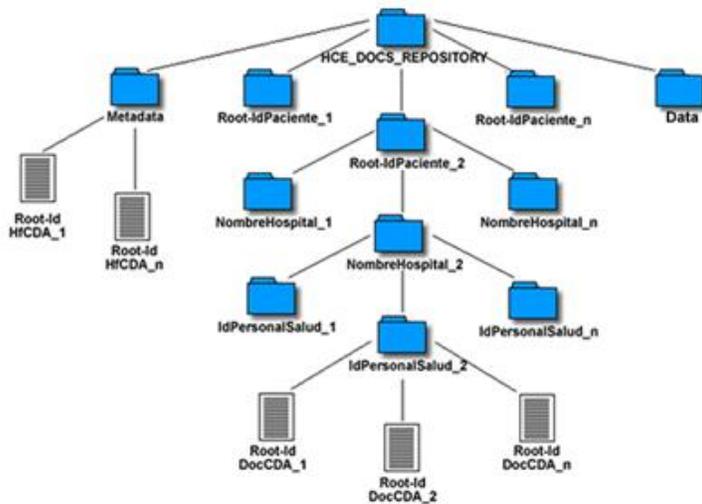


Fig. 9. CDSS en el sistema informático alas SIAPS

Todo el desarrollo es sustentado con la plataforma libre *J5EE (Java 5 Enterprise Edition)* con ayuda del marco de trabajo *JBPM (Java Business Process Management)* para el flujo de trabajo incluido en la retención de casos de pacientes hipertensos. Las reglas de negocio fueron manipuladas mediante *Jboss Drools*. También se utiliza el gestor de bases de datos *PostgreSQL* para el almacenamiento de los casos de pacientes hipertensos y el DM creado. La seguridad del *CDSS* es heredada del módulo de configuración del sistema alas SIAPS. Para optimizar la fase de recuperación de pacientes hipertensos, fue escogida la estrategia de indexado *B-tree* incluida en el gestor de bases de datos. El algoritmo de recuperación es implementado haciendo uso de la tecnología de hilos. Para el análisis *OLAP* se utilizó la herramienta *Pentaho Mondrian*.

A continuación se aborda como el CDSS desarrollado ayuda a establecer un diagnóstico integral de la HTA mediante distintas acciones de salud.

### 3.1.1 Estructura del repositorio clínico del sistema alas SIAPS



El repositorio de HCE está compuesto por una carpeta *Metadata* que almacena las hojas frontales (información demográfica) de los pacientes, una carpeta *Data* que contiene el historial clínico de los pacientes basados en los documentos clínicos tal y como se muestra en la figura siguiente:

Fig. 10. Estructura del repositorio HCE del sistema alas SIAPS

Los documentos son archivos XML normados por el estándar médico HL7/CDA. Dicho repositorio almacena varios tipos de documentos clínicos, de ellos son utilizados por el método los que se mencionan a continuación:

*Hoja Frontal:* Contiene los datos personales del paciente y algunos de los factores de riesgos de la HTA.

*Hoja General de Consulta:* Almacena los problemas del paciente identificados en la consulta de Medicina Interna para extraer solo los datos relacionados con la HTA, así como otros problemas asociados a la misma.

*Evolución Médica:* Almacena los datos de la evolución médica del paciente así como el proceder durante cada consulta.

### 3.1.2 Diagnóstico de la HTA. Toma de decisiones

Los casos referentes a pacientes hipertensos fueron conformados a partir de la información contenida en el repositorio de HCE de sistema alas SIAPS. El umbral de semejanza y valor de la importancia de los rasgos predictivos es condicionado por la media de los criterios otorgados por los especialistas.

El CDSS muestra el código internacional CIAP correspondiente a la enfermedad y la lista de casos correspondientes a los pacientes hipertensos semejantes ordenada (por valor de semejanza) de forma descendente como se muestra en la figura siguiente:

Casos semejantes					
ID caso	Codigo CIAP	Tipo de hipertención	Diagnóstico	Detalles	
7	A18	Secundaria	Hipertenso Grado 1		
14	A17	Primaria	Hipertenso Grado 2		
15	A17	Primaria	Hipertenso Grado 2		
16	A18	Secundaria	Hipertenso Grado 1		
17	A18	Secundaria	Hipertenso Grado 1		
18	A18	Secundaria	Hipertenso Grado 1		
19	A18	Secundaria	Hipertenso Grado 1		
20	A18	Secundaria	Hipertenso Grado 2		
21	A18	Secundaria	Hipertenso Grado 2		
22	A18	Secundaria	Hipertenso Grado 2		

Fig. 11. Recuperación de pacientes hipertensos

El umbral de semejanza y la cantidad de casos a mostrar también son configurados como muestra la siguiente figura:

Configuraciones Globales	
Cantidad de Casos Semejantes :	<input type="text" value="10"/>
Umbral de Semejanza :	<input type="text" value="65"/>
<input type="button" value="Aceptar"/> <input type="button" value="Volver"/>	

Fig. 12. Configuraciones globales

De cada paciente hipertenso recuperado se muestra además el tratamiento asociado al tipo de hipertensión, que puede ser farmacológico, no farmacológico o la combinación de ambos. Dicho tratamiento es creado por la unión del tratamiento previo que tiene el caso recuperado, el cual puede ser modificado; y el que sugiere el estándar JNC-7 como se muestra a continuación:

**Tratamiento**

**Tratamiento JNC:** Se recomienda iniciar el tratamiento con un solo fármaco. si el paciente hipertenso tiene síndrome hiperkinético, la mejor opción es un beta bloqueador; es conveniente el uso de betabloqueadores como el calcio antagonista o bloqueadores alfa ya que exacerban la hipercinesia circulatoria

**Tratamiento Personalizado:** Propanolol (25-100 mg). Aumento de ejercicios físicos y control periódico de la tensión arterial

Aceptar Volver

Fig. 13. Tratamiento de pacientes hipertensos

Además se muestran las complicaciones sufridas producto del tipo de hipertensión padecida, así como la opción de revisión del caso como paso previo al proceso de adaptación.

**Información general**

Detalles del Diagnóstico	
<b>Nivel:</b>	III
<b>Tipo:</b>	Secundaria acelerada-maligna
<b>Semejanza:</b>	79 %
<b>Utilizado:</b>	7 veces

Revisar

Complicaciones
Glaucoma. Aneurisma Órtica. Diabetes Mellitus. Perturbaciones del ritmo cardíaco

Aceptar Volver

Fig. 14. Detalles y complicaciones de un paciente hipertenso

Durante la adaptación son mostrados los rasgos predictivos del paciente nuevo así como un listado de los pacientes semejantes de los cuales se muestran el valor de semejanza y sus rasgos predictivos como detalle. Los rasgos objetivos son adicionados al nuevo caso directamente.

### 3.1.3 Análisis de los elementos causales. Toma de decisiones

Al seleccionar las dimensiones problemas de salud y tiempo, el CDSS muestra a través de OLAP la incidencia de la HTA en el área atendida por la unidad de salud. Al realizar un desglose por años y trimestre, como muestra la figura 16 se puede apreciar en que época del año se hace más frecuente la aparición de dicha enfermedad en los pacientes (en este caso los primeros trimestres de cada año).



Fig. 15. Análisis con respecto al tiempo

La gráfica correspondiente al anterior informe se ilustra a continuación:

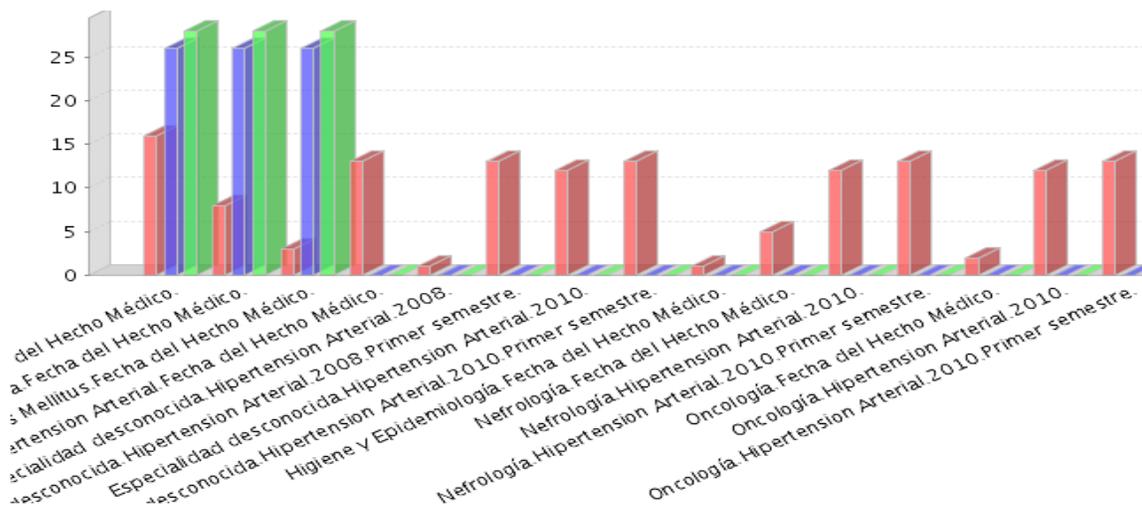


Fig. 16. Gráfica de la vista de análisis problemas de salud-tiempo

Otro de los reportes generados es el de la figura 18, el cual muestra la relación entre la etnia de los pacientes y los hábitos personales. Se aprecian como el consumo excesivo de sal en la comida y el hábito de fumar constituye factores esenciales en la aparición de HTA en los pacientes. También como existe gran influencia por los antecedentes.

Etnia del paciente	Hábitos personales	Medidas	
		↻ Cantidad de problemas	↻ Cantidad de antecedentes
	+ Hábitos personales	16	26
	+ Hábitos personales	7	10
Blanco	- Hábitos personales	2	3
	Fumar	0	0
	Sal	2	3
Negra	+ Hábitos personales	1	0
Negra	- Hábitos personales	2	3
	Fumar	0	0
	Sal	2	3
Mestiza	+ Hábitos personales	1	2
Negra	+ Hábitos personales	1	2
	+ Hábitos personales	9	16
Blanco	+ Hábitos personales	1	2
Blanco	+ Hábitos personales	1	2
Negra	+ Hábitos personales	1	2
Negra	- Hábitos personales	2	2
	Beber	0	0
	Fumar	0	0
	Sal	2	2
Blanco	+ Hábitos personales	1	2
Negra	+ Hábitos personales	1	2
Negra	+ Hábitos personales	1	2
Negra	+ Hábitos personales	1	2

Fig. 17. Análisis de la etnia con respecto a los hábitos personales

Finalmente al analizar la combinación de los criterios antecedentes y problemas de

Problemas de Salud	Antecedentes	Medidas	
		↻ Cantidad de problemas	↻ Cantidad de antecedentes
- Problemas_Salud	+ Antecedentes	16	26
- Especialidad desconocida	+ Antecedentes	8	26
Diabetes Mellitus	- Antecedentes	3	26
	- Antecedente Familiar	0	12
	+ Abuela	0	1
	+ Abuelo	0	1
	+ Madre	3	4
	+ Padre	0	7
	+ Antecedente Personal	3	14
Hipertension Arterial	- Antecedentes	13	0
	- Antecedente Personal	13	0
	+ Madre	13	0
	Hipertension Arterial	13	0
+ Higiene y Epidemiología	+ Antecedentes	1	0
Diabetes Mellitus	+ Antecedente Personal	3	14
- Nefrología	+ Antecedentes	5	0
Hipertension Arterial	- Antecedente Personal	13	0
	- Madre	13	0
	Hipertension Arterial	13	0
- Oncología	+ Antecedentes	2	0
Hipertension Arterial	- Antecedente Personal	13	0
	+ Madre	13	0
	Hipertension Arterial	13	0

salud mostrada en la figura 19, se observa como de las madres se puede heredar más fácilmente la HTA que de otros familiares.

Al unir el resultado de los tres análisis anteriores se puede inferir que en los primeros trimestres de cada año aumentan los casos de HTA en los pacientes cuyas madres

Fig. 18. Análisis de la HTA con respecto a los antecedentes del paciente

también padecen dicho problema de salud. Además, la enfermedad está condicionada principalmente por el consumo de sal y el hábito de fumar.

La gráfica correspondiente a la anterior vista de análisis se muestra a continuación

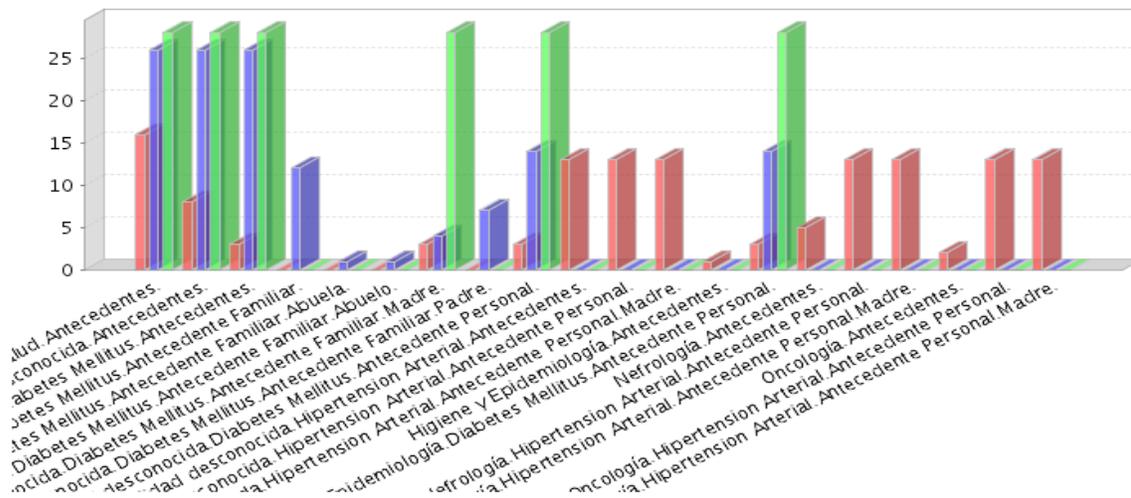


Fig. 19. Gráfica de la vista de análisis problemas de salud-antecedentes personales

Teniendo en cuenta esta información obtenida, son detectadas las causas principales de la HTA presente en los pacientes. Con este conocimiento adquirido sobre la enfermedad, se puede entonces realizar acciones de salud preventivas y/o de seguimiento y control respecto a la HTA.

### Principales causas de HTA en la UCI

En un estudio realizado en el año 2008 (138), la UCI contaba con 112 personas que sufren Hipertensión Arterial. El estudio basado en el Análisis de la Situación de Salud (ASS) de ese año, análisis de entrevistas y encuestas realizadas al personal médico, así como las estadísticas relacionadas con la HTA muestra entre los principales factores de riesgo: el sedentarismo, los antecedentes familiares o personales, exceso del peso corporal, dieta y el embarazo.

En entrevistas realizadas a las enfermeras de cuatro consultorios ubicados en la UCI, se determinó que entre los principales factores de riesgo que afectan a los pacientes hipertensos de dicha universidad están náuseas 30 %, calambre 50%, consumo de bebidas alcohólicas 70%, dieta (exceso de sal) 25 %, lesión OD, 75 %, consumo de café 65% y colesterol 67%.

De las anteriores causas de HTA, algunas son arrojadas como resultado por la herramienta desarrollada, lo que demuestra una correlación entre la información proporcionada por el personal médico y los resultados de la herramienta.

### 3.2 Validación cruzada estratificada

Un clasificador es un modelo general que representa el conocimiento intrínseco existente en una base de datos (139). Una vez construido el modelo, es utilizado para “clasificar” casos nuevos de los que no se conozca el valor de su clase. En la presente investigación se crea un modelo clasificador (MC) basado en el método de validación cruzada de  $k$  hojas (*k - fold cross - validation*) (139) utilizando la función de semejanza descrita en la expresión (7) del capítulo II. El procedimiento es aplicado para 7 repeticiones ( $k - fold = 7$ ).

La validación cruzada consiste en: dado un número  $n$  de casos, se dividen los datos en partes (*folds*), con se construye el MC con las  $n-1$  partes y con el resto se valida. Cada partición del conjunto de datos es mutuamente excluyente y de igual tamaño. La tasa de error se calcula con la media de los errores asociados a las  $n$  partes.

La base de casos de pacientes hipertensos contiene 100 instancias, por lo que fue utilizada la herramienta WEKA (*Waikato Environment for Knowledge Analysis*) (140) para facilitar el análisis de los resultados obtenidos.

Es utilizada la validación cruzada debido a que ofrece gran precisión en los resultados que brinda, evitando así el solapamiento de los conjuntos de prueba, es decir, el problema de sobre ajuste de los datos (*overfitting*) de los clasificadores. Es necesario destacar que la validación cruzada constituye una técnica costosa computacionalmente.

#### 3.1.1 Bases de datos utilizadas por el clasificador

También es empleado el MC con cuatro bases de datos internacionales publicadas en el *UCI Repository Machine Learning Repository* de la Universidad de California (*Center for Machine Learning and Intelligent Systems*) (141). Estas bases de datos han sido ampliamente utilizadas dada su disponibilidad de forma pública por lo que distintos clasificadores se han probado sobre dichos conjuntos de datos. La selección de las bases de datos para este trabajo se apoya en el ámbito de aplicación (*diagnóstico*) por las que fueron publicadas. Las mismas se detallan a continuación:

- *Breast Cancer Wisconsin (Diagnostic)*, suministrada por William H. Wolberg, W. Nick Street y Olvi L. Mangasarian 1995. Instancias: 569. Tipos de datos: reales.
- *Dermatology*, suministrada por Nilsellter, H. Altay Guvenir, 1998. Presenta datos ausentes. Instancias: 366. Tipos de datos: reales y nominales.

- *Echocardiogram*, suministrada por Steven Salzberg, 1989. Instancias: 132. Tipos de datos: categóricos, enteros y reales. Presenta datos ausentes.
- *Heart Disease*, suministrada por Andras Janosi, William Steinbrunn, Matthias Pfisterer, Robert Detrano, 1988. Presenta datos ausentes. Instancias: 303. Tipos de datos: enteros, nominales y reales.

### 3.1.2 Análisis de resultados del clasificador

Para medir el correcto funcionamiento (eficiencia) del MC se emplean algunos indicadores como: sensibilidad y especificidad. En (142) se encuentran otras medidas para medir la eficiencia de los clasificadores. A continuación se analizan cada uno de los indicadores arrojados por el modelo.

#### Tasas de acierto y error del clasificador.

Para calcular el error que se comete al intentar clasificar todos los registros, se debe tener en cuenta las medidas anteriormente mencionadas. El interés, categoría sensible o *clase positiva* suscita a cada tipo de HTA (*primaria (P)*, *secundaria acelerada maligna (SAM)* y *secundaria renovascular (SR)*). Como la clase positiva es multiclásica (*alcanza 3 valores correspondientes a cada tipo de HTA*), para facilitar el trabajo se decidió utilizar un enfoque biclásico dónde se tiene la categoría sensible y el resto de los valores, también llamado *clase negativa*; de este modo es aplicado el clasificador para cada clase positiva. A continuación se describen cada una de las medidas utilizadas:

- **Verdaderos Positivos (VP) (Sensibilidad)**: Cantidad de veces que el clasificador predice que un paciente tiene un tipo de hipertensión y es verdad.
- **Verdaderos Negativos (VN) (Especificidad)**: Cantidad de veces que el clasificador predice el tipo de hipertensión que no pertenece a un paciente y es verdad.
- **Falsos Positivos (FP)**: Los pacientes que no presentan un tipo de hipertensión y el algoritmo predijo que si lo presentan.
- **Falsos Negativos (FN)**: Los pacientes que presentan un tipo de hipertensión y el algoritmo no lo predijo.

La matriz base de contingencia para cada clase positiva se muestra a continuación:

**Tabla 6.** Tabla de contingencia base para cada tipo de clase positiva

		<b>Clase predicha</b>	
		Clase positiva	Otros
<b>Clase verdadera</b>	Clase positiva	VP	FP
	Otros	FN	VN

Por tanto se tiene para cada clase positiva VP+VN, que son los casos donde la clasificación es correcta, mientras que en FP+FN son los casos se han cometido errores en la clasificación. Las medidas anteriormente mencionadas así como las tasas de acierto (TA) y error (TE) se determinan de la siguiente forma (142):

$$VP = \frac{VP}{VP + FN} \quad FN = \frac{FN}{VP + FN} \quad TE = \frac{FP + FN}{VP + FP + FN + VN}$$

$$VN = \frac{VN}{VN + FP} \quad FP = \frac{FP}{VN + FP} \quad TA = \frac{VP + VN}{VP + FP + FN + VN}$$

Para cada clase positiva se obtuvieron los siguientes resultados:

**Tabla 7.** Tabla de contingencia HTA Primaria

	P	Otros
P	58	1
Otros	5	36

**Tabla 8.** Tabla de contingencia HTA Secundaria Maligna

	SAM	Otros
SAM	19	4
Otros	8	69

**Tabla 9.** Tabla de contingencia HTA Renovascular

	SR	Otros
SR	23	3
Otros	7	67

La TA promedio para cada clase positiva es 94.4%, 88% y 90 % respectivamente. La TE se corresponden con los valores que siguen 5,6%, 12% y 10%. Se obtiene además un 90, 7 ≈ 91 % y 9,3% de TA y TE general respectivamente. Como TA > TE para cada clase positiva entonces se puede plantear que es correcto el funcionamiento del MC.

Finalmente se compara la eficiencia del MC creado con la de otros clasificadores como ID3 (143), C4.5 (144) y la Clasificación por Regresión o Regresión Lineal (RL) (145) utilizados con la misma base de casos de pacientes hipertensos. La siguiente tabla muestra los resultados alcanzados:

**Tabla 10.** Por ciento de casos bien clasificados de MC

ID3	C4.5	RL	MC
86 %	79%	82%	91%

Los resultados del MC arrojaron un comportamiento superior al de los otros clasificadores quedando evidenciado la eficiencia del mismo.

**Análisis de las bases de datos internacionales**

También es aplicado el MC sobre las bases de datos internacionales citadas anteriormente, las cuales han sido utilizadas por clasificadores encontrados en la literatura consultada. A continuación se muestran las tablas de comparaciones realizadas.

En (142; 139) se utilizan los algoritmos C 4.5 e ID3 con las bases de datos *Breast Cancer Wisconsin* y *Heart Disease*. La siguiente tabla muestra una comparación de dichos algoritmos con el MC.

**Tabla 11.** Resultados de aplicar C 4.5, ID3 y el MC

Clasificadores	Bases de Datos	Tasa de Aciertos (%)
C 4.5	<i>Breast Cancer Wisconsin</i>	93 %
ID3	Heart Disease	91 %
MC	<i>Breast Cancer Wisconsin</i>	95 %
MC	Heart Disease	90 %

Adicionalmente en (146) se utiliza la RL con la base de datos *Dermatology*. Es mostrado en la tabla 7 los resultados comparativos.

**Tabla 12.** Resultados de aplicar RL y el MC

Clasificadores	Bases de Datos	Tasa de Aciertos (%)
RL	<i>Dermatology</i>	94 %
MC	<i>Dermatology</i>	91 %

Los casos con valores perdidos (ausentes) fueron omitidos en las bases de datos utilizadas. El MC creado proporciona una tasa de acierto por encima del 90 %, al igual que los resultados obtenidos por los algoritmos aplicados a las bases de datos anteriormente mencionadas. De manera general son buenos los resultados obtenidos por el MC de modo que es correcto su funcionamiento en bases de datos independientes.

### 3.2 Validación por comité de expertos

Para llevar a cabo la validación del Componente para la Toma de Decisiones Clínicas se utilizó el método Delphi (147). Se contó con la participación de nueve especialistas a los cuales se autoevaluaron sus conocimientos y habilidades relacionados con la prevención, diagnóstico y tratamiento de la HTA, así como las herramientas y/o técnicas para el apoyo a la toma de decisiones. De ellos, 6 son doctores con más 5 años de experiencia en la Atención Primaria de Salud. En relación al grado académico 3 ostentan la categoría de máster.

Luego de haber seleccionado a los posibles expertos se les explicó en qué consistía el método que se aplicaría y una vez dado su consentimiento para la evaluación, se les entregó un cuestionario impreso (Ver Anexo 1) que debían responder con el objetivo de valorar su nivel de competencia y conocimiento a fin de determinar su experticia en el tema.

En esta metodología la competencia de los expertos se determina por el coeficiente  $k$ , el cual se calcula de acuerdo con la opinión del experto sobre su nivel de conocimiento acerca del problema que se está resolviendo y con las fuentes que le permiten argumentar sus criterios.

El coeficiente de competencia se calcula por la siguiente fórmula:

$$K = \frac{1}{2} (k_c + k_a)$$

**$k_c$** : es el coeficiente de conocimiento o información que tiene el experto acerca del problema, calculado sobre la valoración del propio experto en una escala del 1 al 10, de esta forma, la evaluación "1" indica que el experto tiene muy poco conocimiento de la problemática correspondiente, mientras que la evaluación "10" significa todo lo contrario.

**ka:** es el coeficiente de argumentación o fundamentación de los criterios del experto, obtenido como resultado de la suma de los puntos alcanzados a partir de una tabla patrón como la siguiente:

**Tabla 13.** Criterios de los expertos

FUENTES DE ARGUMENTACION	Grado de influencia de cada una de las fuentes en sus criterios.		
	A (alto)	M (medio)	B (bajo)
Análisis teóricos realizados por usted (AT)	0.3	0.2	0.1
Su experiencia obtenida (EO)	0.5	0.4	0.2
Trabajos de autores nacionales (AN)	0.05	0.05	0.05
Trabajos de autores extranjeros (AE)	0.05	0.05	0.05
Su propio conocimiento del estado del problema en el extranjero (PC)	0.05	0.05	0.05
Su intuición (I)	0.05	0.05	0.05

Los resultados mostrados del cálculo del índice de experticia se muestran a continuación:

**Tabla 15.** Resultados del índice de experticia

Expertos	Kc	AT	EO	AN	AE	PC	I	Ka	K	Competencia
1	0,8	0,2	0,4	0,05	0,05	0,05	0,05	0,8	0,8	Alta
2	0,9	0,2	0,5	0,05	0,05	0,05	0,05	0,9	0,9	Alta
3	0,9	0,2	0,2	0,05	0,05	0,05	0,05	0,6	0,75	Media
4	0,8	0,2	0,5	0,05	0,05	0,05	0,05	0,9	0,85	Alta
5	0,7	0,2	0,2	0,05	0,05	0,05	0,05	0,6	0,65	Media
6	0,7	0,2	0,4	0,05	0,05	0,05	0,05	0,8	0,75	Media
7	0,9	0,2	0,4	0,05	0,05	0,05	0,05	0,8	0,85	Alta
8	0,8	0,2	0,2	0,05	0,05	0,05	0,05	0,6	0,65	Media
9	0,9	0,3	0,4	0,05	0,05	0,05	0,05	0,9	0,85	Alta

Después del cálculo de k, se determina el nivel de competencia de la siguiente manera (148):

- **Alto:** si K es mayor o igual que 0,8

- **Medio:** si K es mayor o igual que 0,5 y menor que 0.8
- **Bajo:** si K es menor que 0,5.

Como el coeficiente de conocimientos de todos los expertos es mayor que 0,7, se decidió seleccionar todos los expertos que se autoevaluaron en la primera ronda de preguntas.

### 3.2.1 Aplicación del método

Se les aplicó a los expertos un segundo cuestionario (Ver Anexo 2) que permitió comprobar el nivel de coincidencia entre la opinión de los expertos y los elementos que se tuvieron en cuenta para diseñar el CDSS desarrollado. Ofreció además, una medida de la utilidad del uso de dicho sistema como ayuda en la detección los factores de riesgo de HTA. También proporcionó un valor origen para decidir cuándo es válida la clasificación realizada por el CDSS en relación al tipo de HTA presente en los pacientes.

El cuestionario fue entregado en copia dura a los expertos, y se les anexó un documento con la descripción del trabajo, características y métodos propuestos. El mismo consta de siete preguntas categorizadas en:

- Completamente(C1)
- En gran medida (C2)
- Poco (C3)
- Muy poco (C4)
- Nada (C5)

### 3.2.2 Análisis de los resultados

Para llevar a cabo el análisis de los resultados se utilizaron métodos estadísticos tal y como indica el método Delphi. La siguiente tabla muestra los resultados obtenidos una vez aplicada la segunda encuesta a través de la frecuencia absoluta:

Tabla 16. Resultados de la encuesta

No	Aspectos	C1	C2	C3	C4	C5	Total
1	Utilización de una herramienta informática que ayude a la correcta clasificación de los tipos de HTA.	8	1				
2	Elección correcta de los tipos de HTA utilizados	7	2				

	por la herramienta informática.						
3	Identificación de los rasgos predictivos y objetivos para la creación de los casos de pacientes hipertensos.	4	5				
4	Validez de los tratamientos sugeridos por la herramienta informática	7	2				
5	Valor de umbral de semejanza utilizado por la herramienta informática.	1	7	1			
6	Dimensiones utilizadas en la herramienta informática durante la generación de reportes OLAP.	4	3	1	1		
7	Detección de causas de la HTA por parte de la herramienta informática	8	1				

A partir de la frecuencia absoluta se puede calcular la frecuencia absoluta acumulada, la misma se obtiene sumando los valores de cada fila con el valor anterior, quedando de la siguiente manera:

**Tabla 17.** Frecuencia absoluta respecto a los resultados obtenidos

No	Aspectos	C1	C2	C3	C4	C5
1	Utilización de una herramienta informática que ayude a la correcta clasificación de los tipos de HTA.	8	9	9	9	9
2	Elección correcta de los tipos de HTA utilizados por la herramienta informática.	7	9	9	9	9
3	Identificación de los rasgos predictivos y objetivos para la creación de los casos de pacientes hipertensos.	4	9	9	9	9
4	Validez de los tratamientos sugeridos por la herramienta informática	7	9	9	9	9
5	Valor de umbral de semejanza utilizado por la herramienta informática.	1	8	9	9	9
6	Dimensiones utilizadas en la herramienta informática	4	7	8	9	9

	durante la generación de reportes OLAP.					
7	Detección de causas de la HTA por parte de la herramienta informática	8	9	9	9	9

La frecuencia relativa acumulada se obtiene al dividir la frecuencia absoluta acumulada entre la cantidad de expertos como se muestra a continuación:

**Tabla 18.** Frecuencia absoluta acumulada respecto a los resultados obtenidos

No	Aspectos	C1	C2	C3	C4	C5
1	Utilización de una herramienta informática que ayude a la correcta clasificación de los tipos de HTA.	0.88	0.99	0.99	0.99	0.99
2	Elección correcta de los tipos de HTA utilizados por la herramienta informática.	0.77	0.99	0.99	0.99	0.99
3	Identificación de los rasgos predictivos y objetivos para la creación de los casos de pacientes hipertensos.	0.44	0.99	0.99	0.99	0.99
4	Validez de los tratamientos sugeridos por la herramienta informática	0.77	0.99	0.99	0.99	0.99
5	Valor de umbral de semejanza utilizado por la herramienta informática.	0.11	0.88	0.99	0.99	0.99
6	Dimensiones utilizadas en la herramienta informática durante la generación de reportes OLAP.	0.44	0.77	0.88	0.99	0.99
7	Detección de causas de la HTA por parte de la herramienta informática	0.88	0.99	0.99	0.99	0.99

Por último se buscan las imágenes de las frecuencias relativas acumuladas por medio de la Distribución Normal Estándar Invertida y se adicionan las siguientes columnas:

- Sumatoria (S): sumatoria de cada fila y columna
- Promedio (P): promedio de los criterios de los expertos

- N: es el resultado de dividir la sumatoria de las sumas entre el producto del número de categorías por el número de pasos.
- N-P: es entonces el valor promedio que le otorgan los expertos consultados a cada paso de la metodología
- Punto de corte: es el promedio de cada columna.

La siguiente tabla resume estos resultados:

**Tabla 16.** Frecuencias relativas mediante la Distribución Normal Estándar Invertida

No	Aspectos	C1	C2	C3	S	NP		Competencia
						P	N-P	
1	Utilización de una herramienta informática que ayude a la correcta clasificación de los tipos de HTA.	1.22	3.71	4.26	13.46	3.3	-0.60	Completamente
2	Clasificación correcta de HTA.	0.76	3.71	4.26	13.01	3.2	-0.49	Completamente
3	Identificación del cuadro clínico para la confección de los casos de pacientes hipertensos.	-0.13	3.71	4.26	12.10	3.0	-0.26	Completamente
4	Validez de los tratamientos sugeridos	0.76	3.71	4.26	13.01	3.2	-0.49	Completamente
5	Valor mínimo adecuado para la clasificación de HTA	-1.22	1.22	4.26	8.52	2.1	0.62	En gran medida
6	Criterios ofrecidos durante la confección de informes	-0.13	0.76	1.22	6.10	1.5	1.23	En gran medida
7	Detección de causas y eventos anormales de la HTA	1.22	3.71	4.26	13.46	3.3	-0.60	Completamente
Suma		1.25	16.9	22.5				
Punto de Corte		0.21	2.81	3.75				

Los puntos de corte sirven para determinar la categoría o grado de adecuación de cada criterio según la opinión de los expertos consultados. La determinación del grado de adecuación de cada aspecto a validar se realiza de la siguiente forma:

- Si N-P es menor o igual que 0,21 entonces es Completamente

- Si N-P es mayor que 0,21 y menor o igual que 2,81 es En gran medida
- Si N-P es mayor que 2,81 y menor o igual que 3,75 es Poco
- Si N-P es mayor que 3,75 es Muy Poco.

A partir de este análisis todos los aspectos del método propuesto fueron catalogados de Completamente a excepción del valor mínimo que toma la función de semejanza para emitir resultados y los criterios condicionados por los factores de riesgo de la HTA que se tienen al hacer procesamiento analítico en línea. Los señalamientos van encaminados a que el valor mínimo resulta complejo de determinar debido a la escasa información sobre los rasgos que puede tener un paciente con hipertensión primaria, mientras que los criterios utilizados en el procesamiento analítico en línea son esenciales en el diagnóstico de la HTA.

### 3.2.3 Evaluación de concordancia

La evaluación de la concordancia se realizó mediante el cálculo del coeficiente de Kendall ( $W$ ). Para comprobar el grado de significación de Kendall se aplica la prueba de Chi-cuadrado, planteándose la hipótesis nula y la alternativa respectivamente. Cada experto midió mediante una escala del 1 al 5 cada uno de los indicadores, donde: 5: Manifestación COMPLETAMENTE; 4: Manifestación EN GRAN MEDIDA; 3: Manifestación POCO; 2: Manifestación MUY POCO; 1: Manifestación NADA. Los resultados se muestran a continuación:

Tabla 19. Evaluación de concordancia

	E1	E2	E3	E4	E5	E6	E7	E8	E9	Ep	$\Sigma E$	$\Delta C$	$\Delta C^2$
P1	5	5	5	5	5	5	4	5	5	4.9	44	-3.8	14.69
P2	5	5	5	4	4	5	5	5	5	4.8	43	-4.8	23.36
P3	5	5	5	4	5	4	4	4	4	4.4	40	-7.8	61.36
P4	5	5	5	5	5	5	5	5	5	5.0	45	3.7	13.79
P5	4	4	4	4	4	5	3	4	4	4	36	-11.8	140.03
P6	5	4	3	5	4	5	4	5	2	4.1	37	-10.8	117.4
P7	5	5	5	5	5	4	5	5	5	4.9	44	-3.8	14.69

Donde  $C$  es la cantidad de criterios presentes,  $E_p$  es el promedio de los criterios dados (filas),  $\Sigma E$  es la sumatoria de los criterios dados (filas),  $\Delta C$  es la diferencia entre el total de  $\Sigma E$  y  $\Sigma E$  entre  $C$ , finalmente  $\Delta C^2$  es el cuadrado de  $\Delta C$ .

Con estos datos se calcula  $S$  según la expresión que a continuación se muestra:

$$S = \left| \sum \left( \sum E - \sum \frac{\sum E}{C} \right)^2 \right|$$

Y a continuación el coeficiente de Kendall (W) y Chi-cuadrado ( $X^2$ ) según las siguientes expresiones:

$$W = \frac{S}{\frac{E^2}{12}(C^3 - C)}$$

$$X^2 = E(C - 1)W$$

Obteniendo como resultado un coeficiente de Kendall (W) de 0,463 y un valor de Chi-cuadrado de 19,6.

La hipótesis nula y la alternativa se plantean de la siguiente forma (148):

- $H_0$ : no existe concordancia en los criterios emitidos por los expertos.
- $H_1$ : existe concordancia en los criterios emitidos por los expertos.

Por otra parte, se busca el Chi-cuadrado tabulado en la tabla del percentil de la distribución Chi-cuadrado con un nivel de significación  $\alpha$  y  $C - 1$  grados de libertad, representado por  $X^2_{\text{tab}} = X^2_{\alpha; C-1}$ .

Al ser comparado el valor de Chi-cuadrado calculado con el valor de Chi-cuadrado tabulado con un nivel de significación de 0,01, se obtiene que  $19.6 > 16,81$ , por tanto se rechaza  $H_0$ , es decir, se acepta  $H_1$ . Se puede afirmar entonces que los expertos concuerdan al 99% de confianza en los criterios emitidos.

La media de los criterios expresados es superior a 4, medido en una escala de 0 a 5, por lo que la emisión de criterios de evaluación es alta, demostrando de esta forma la factibilidad del CDSS desarrollado.

Los comentarios aportados por los expertos para mejorar la herramienta estuvieron enfocados a conjugar o integrar la propuesta presentada con otros estudios científicos relacionados con el apoyo médico a para así darle una mayor connotación a los resultados finales que se podrían obtener con esta investigación.

### 3.3 Consideraciones del capítulo

En el anterior capítulo se llevó a cabo la validación del sistema teniendo en cuenta las valoraciones estadísticas del modelo clasificador elaborado y el criterio de los expertos.

Se arribaron a las siguientes conclusiones:

- Se construyó un *CDSS* que sustenta el método descrito como un módulo del sistema alas SIAPS.
- Se conocen las causas de la HTA debido a los informes resultantes del procesamiento analítico en línea.
- Es posible conocer el tipo de hipertensión, nivel y las complicaciones asociadas así como aplicar un proceder correspondiente en pacientes hipertensos.
- Los resultados arrojados por el clasificador mostraron un correcto comportamiento debido a las tasas de acierto y error respectivamente.
- El *CDSS* fue presentado a un conjunto de expertos que valoraron las características que se proponen en el presente trabajo. Los resultados del análisis fueron los que a continuación se muestran:
  - Los criterios ofrecidos durante la confección de informes y la detección de los factores de riesgos de la HTA fueron valorados como *En gran medida* puesto a que los expertos optaron por la conjugación de estas características con estudios de similar relación. El resto de las características fueron valoradas de *Completamente*.
  - Se determinó la concordancia de criterios preferenciales entre los expertos encuestados.

## CONCLUSIONES

- El estudio del estado del arte realizado evidenció una tendencia internacional de la aplicación de sistemas para la toma de decisiones clínicas en el diagnóstico de la Hipertensión Arterial en el sector de la salud.
- El método propuesto a través de la integración de Razonamiento Basado en Casos y el Procesamiento Analítico en Línea muestra potencialidades para la prevención y diagnóstico temprano de la HTA.
- La utilización de Procesamiento Analítico en Línea además de ser novedoso en los Sistemas Clínicos para la Toma de Decisiones, apoya el proceder de la Hipertensión Arterial en las áreas más comunes mediante la generación de nuevo conocimiento.
- A partir del método propuesto se implementó un Sistema Clínico para la Toma de Decisiones, que además de constituir un pilar para el diagnóstico integral de la Hipertensión Arterial, también contribuye a reducir imprecisiones que ocurren durante el diagnóstico médico como las cascadas diagnósticas y la polimedicación.
- El Sistema Clínico para la Toma de Decisiones fue validado utilizando un modelo clasificador obteniendo una precisión superior al 90 por ciento, arrojando resultados satisfactorios al ser comparado con clasificadores del estado del arte.
- Los resultados de las encuestas aplicadas a los expertos demuestran un alto grado de aceptación de la herramienta y la efectividad de la misma para el diagnóstico de la enfermedad.

## TRABAJO FUTURO

Se recomienda para futuras versiones de la presente investigación:

- Incluir el tratamiento de rasgos difusos en el RBC para incrementar la precisión en la recuperación de casos semejantes, así como el tratamiento de la incertidumbre de los rasgos.
- Diseñar un sistema de alertas (Monitor de Eventos Clínicos) que sea integrado al CDSS usando el sistema de notificaciones del sistema alas SIAPS de modo que los médicos puedan ser prevenidos automáticamente antes posibles sucesos de la HTA.

## BIBLIOGRAFÍA

1. **Hersh, W.** *Medical informatics: improving health care through information*. 16, USA : JAMA, 2002, Vol. 288, pp. 1955-8.
2. **Anderson, J. and Gremy, F.** *Education in Informatics of Health Personnel*. Europe : North-Holland Publ. Co, 1974. Vol. 1.
3. **Van Bemmel, J.H.** *Medical Informatics. Art or Science?* Netherlands : Meth Inform Me, 1996. pp. 157-172. Vol. 35.
4. **Bertalanffy, L.V.** *General systems theory: foundation, development, applications*. New York : Brazilliez, 1978. p. 45.
5. **Navarro, H. and Pastor, V.** *Manual de gestión hospitalaria*. [book auth.] H Navarro. *Sistemas de Información*. Madrid : Interamericana-Mc Graw-Hill, 1992, pp. 189-208.
6. **Cosials, D.** *Información para la gestión clínica. Contrato de servicio*. Madrid : ELSEVIER ESPAÑA, S.A, 2000. Vol. 2. 9788481744828. pp. 19-26.
7. **Wright, A.** *ANDS: An Architecture for Clinical Decision Support in a National Health Information Network*. Toronto : AMIA Annu Symp Proc, 2007. pp. 816-820.
8. **Berner, Eta S.** *Clinical Cecision Support Systems: Theory and Practice*. US : Health Information Series, 2007. 0387-38319-0. pp. 77-103
9. **Haynes, R.** *Of studies, syntheses, synopses, and systems: the "4S" evolution of services for finding current best evidence*. 2, US : ACP J Club, 2002, Vol. 134, pp. A11-3.
10. **Codd, E.F. and Salley, C.T.** *Providing OLAP (on-line Analytical Processing) to User-analysts: An IT Mandate*. US : Codd & Date, Inc, 1993. Vol. 32. 0718-3305. p. 31
11. **Bichindaritz, I.** *A case-based assistant for clinical psychiatry expertise*. Paris : Prog Annu Symp Comput Appl Med Care, 1994. pp. 673-7.
12. **Macura, R.** *Computerized case-based instructional system for computed tomography an magnetic resonance imaging of brain tumors*. Augusta : Invest Radiol, 1994. Vol. 4. 497-506. pp. 497-506.
13. **Bartels, P. H.** *Automated reasoning system in histopathologic diagnosis and prognosis of prostate cancer and its precursors*. US : Eur Urol, 1996. Vol. 2. pp. 222-33.
14. **García, M. M. and Bello, R. E.** *A model and its different applications to case-based reasoning*. 7, Cuba : Knowledge-Based Systems, November 1996, Vol. 9. 0950-7051.
15. **ROLSTON, D. W.** *Principios de Inteligencia Artificial y Sistemas Expertos*. Universidad Catolica San Pablo : DSpaces, 2002. 970-10-0074-9.
16. **GARCÍA, G. L.** Los sistemas automatizados de acceso a la información bibliográfica: Evaluación y tendencias en la era de Internet. *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*. Salamanca : Universidad de Salamanca, 2010, Vol. 11, pp. 436-437.
17. **Navas, H. et al.** *Creation and evaluation of a terminology server for the interactive coding of discharge summaries*. Buenos Aires : Stud Health Technol Inform, 2007. pp. 650-4. 16.
18. **Kolodner, J. and Lake, D.** *A Tutorial Introduction to Case-Based Reasoning*. [book auth.] J Kolodner. *Case-Based Reasoning: Experiences, Lessons, and Future Directions*. United States of America : AAAI Press / The MIT Press, 1996.

19. **Gutiérrez, I and Bello, R.** *Un Sistema Basado en Casos para la Toma de Decisiones en Condiciones de Incertidumbre.* Villa Clara : Revista Investigación Operacional, 2002, Vol. 23.
20. **Martínez, N. and García, M. et al.** *El Paradigma del Razonamiento Basado en Casos en el Ámbito de los Sistemas de Enseñanza/Aprendizaje Inteligentes.* Villa Clara, Cuba : Revista Electrónica de Tecnología Educativa, 2009. 1135-9250.
21. **Rivera, S. M.** *Modelo de un Sistema de Razonamiento basado en casos para el análisis en la Gestión de Riesgos.* La Habana : Tesis para optar por el grado de Máster en Informática Aplicada, 2010.
22. **Bello, Rafael.** *El empleo del razonamiento basado en casos en el desarrollo de sistemas basados en el conocimiento para el diagnóstico.* Villa Clara, Cuba : Trabajo de Tesis en opción al grado científico de Doctor en Ciencias Técnicas, Universidad de las Villas, 1997.
23. **Febles, J. P. and Estrada, V.** *Uso del razonamiento basado en casos para la enseñanza de temas médicos.* 1, La Habana, Cuba : Revista Ingeniería Industrial, 2002, Vol. 23. 0258-5960.
24. **Watson, I.** *Applying Case-Based Reasoning: Techniques for Enterprise Systems.* Morgan Kaufmann., 1997.
25. **Arocha, C. and Castillo, A.** *Experiencia cubana en la organización de los servicios de salud, una alternativa ante la privatización.*1, La Habana, Cuba : Revista Cubana de Salud Pública, 2007, Vol. 33. ISSN 0864-3466.
26. **Stusser, R. J. and Rodríguez, A.** *La informatización de la atención primaria de salud.* 4, La Habana, Cuba : Revista Cubana de Medicina General Integral, 2006, Vol. 22.
27. **Narey, B. and Aldereguía, G.** *Medicina social y salud pública en Cuba.* La Habana : Editorial Pueblo y Educación, 1990. p. 151. Vols. La Habana: Editorial Pueblo y Educación, 1990:151.
28. **Guerra, J. and Alfonso, P.** *Hipertensión Arterial en la Atención Primaria de Salud.* La Habana Cuba : Ciencias Médicas, 2009. 978-959-212-530-8.
29. **Cuadrado, S. and Casas, G.** *Tensoft: Sistema informativo para el diagnóstico de la HTA sobre bases estadísticas.* Universidad Central de Las Villas. Santa Clara : Tesis presentada en opción al título de Máster en Ciencias, 2006. pp. 26-44.
30. **OMS.** *Estadísticas Sanitarias Mundiales 2012.* US : World Health Organization, 2012. 978 92 4 356444 9.
31. **Cordero, A. et al.** *Novedades en hipertensión arterial y diabetes mellitus.* Madrid : Revista española de cardiología, 2010. Vol. 63. 0300-8932. pp. 101-115.
32. **Aguilar, C.** *Riesgo cardiovascular en diabetes mellitus e hipertensión arterial en España. Estudio CORONARIA.* 4, Madrid : Medicina Clínica, 2006, Vol. 127. 13090381. pp. 126–132.
33. **Soca, M. ; Enrique, P. and Sarmiento, Y.** *Hipertensión arterial, un enemigo peligroso.* 5, Habana : ACIMED, 2009, Vol. 2. 1561-2880.
34. **Barreto, D. G.** *La hipertensión arterial en la tercera edad.* 2, Habana : Revista Médica, 2009, Vol. 48.
35. **Agramonte, M. et al.** *Prevalencia del síndrome metabólico en pacientes diabéticos tipo 2.* 1, Habana : Revista Cubana de Medicina, 2010, Vol. 48. SSN 1561-302X.
36. **Oliva, A. M.** *Hiper web Sistema para la ayuda a la toma de decisiones en el diagnóstico, evaluación, control y tratamiento de los pacientes con hipertensión arterial.* La Habana, Cuba :

Universidad de las Ciencias Informáticas– ICID– Hospital Hnos. Ameijeiras. XVI Forum Nacional de Ciencia y Técnica, 2007.

37. **Gutiérrez, I. et al.** *URS-HTA, sistema para el diagnóstico de la Hipertensión Arterial*. Villa Clara : II Simposio Internacional de Hipertensión Arterial. Ed. Feijoo, 2003. pp. 2-4.

38. **Chávez, M. C. ; et al.** *Red Bayesiana a partir de factores de riesgo de la Hipertensión Arterial*. Santa Clara, Villa Clara, CUBA : Universidad Central de Las Villas (UCLV), 2008.

39. **Mateos, J. G.** *Modelo con RNA para Diagnosticar Hipertensión Arterial*. Guadalajara : Universidad de Guadalajara. Centro Universitario de Ciencias Exactas e Ingenierías. Departamento de Electrónica, 2004. 959-250-161-0.

40. **Turing, A.** *Computing machinery and intelligence*. Inglaterra : Mind, 1950. pp. 433-460. 59.

41. **Ginsberg, M.** *Essentials of artificial intelligence*. San Francisco : Morgan Kaufmann Publishers, 1993. 1-55860-221-6.

42. **Blum, A. L.** Selection of relevant features and examples in machine learning. *Artificial Intelligence*. USA : Carnegie Mellon University Pittsburgh, 1997, Vol. 97.

43. **Berner, Eta S.** *Clinical Decision Support Systems: State of the Art*. US : Agency for Healthcare Research and Quality, 2009. 09-0069-EF.

44. **Weiss, S. M. et al.** *A model based method for computer-aided medical decision making*. *Artificial Intelligence*, 1978. Vol. 11. pp. 145-172.

45. **Buchanan, B. G. et al.** *Heuristic DENDRAL: a program for generating explanatory hypothesis in organic chemistry*. [ed.] Edinburgh University Press. Edinburgh : Machine Intelligence, 1969. Vol. 4. pp. 209-254.

46. **Shortliffe, E. H.** *MYCIN: A rule-based computer program for advising physicians regarding antimicrobial therapy selection*. California : Dissertation, Stanford University, 1974.

47. **Engle, E.L.** *Attempts to use computers as diagnostic aids in medical decision making: a thirty-year*. New York : Perspect. Biol. Med, 2008. Vol. 35. 1472-6947-5-1. pp. 207-219.

48. **Miller, R.** *Why the standard view is standard: people, not machines, understand patients problems*. *J. Med. Philos*, 1990. Vol. 15. 15-559-579. pp. 581-591.

49. **Chobanian A.V. et al.** *The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure: The JNC 7 Report*. *JAMA*, 2003. pp. 2560-72. 289.

50. **Gonzalez, E. and Pérez, A.** *Diagnostico de referencia PROCDEC*. la Habana : Sitio INFOMED/HTA, 2004. 959-250-161-0.

51. **López, D. and Blobel, B.** *Arquitecturas para la Implementación de Sistemas de Información en Salud Basadas en el Estándar HL7. 2*, Alemania : EUROPEAN JOURNAL FOR BIOMEDICAL INFORMATICS, 2010, Vol. 8. ISSN 1801-5603.

52. **Kass, G.** *An Exploratory Technique for Investigating Large Quantities of Categorical Data*. *Applied Statistics*. 2, South African : South African Medical Journal, 1980, Vol. 29, pp. 19–127.

53. **Olaz, F. and Berjón, J.** *Valoración cardiológica del paciente hipertenso*. 1, Madrid : ANALES, 2009, Vol. 21.

54. **Hanney, W. and Keane, M.T.** *Learning Adaptation Rules from a Case-Base In Proc. Advances in Case-Based Reasoning*. Lausanne : 3rd European Workshop, EWCBR-96, November 1996. pp. 179-192.
55. **Hand, D.J.** *Construction and Assessment of Classification Rules*. Chichester : John Wiley & Sons, 1997.
56. **Galvez L. D.** *Sistemas basados en el conocimiento*. Villa Clara : Universidad Central de las Villas, 2006. pp. 4-37.
57. **Rích, E.** *nteligencia artificial*. España : 2da ed. McGraw-Hill, Interamericana de España, 1994. pp. 135-160.
58. **Rissland, E.L. and Skalak, D.B.** *Cabaret: rule interpretation in a hybrid architecture..* 6, International Journal of Man-Machine Studies, 1991, Vol. 34, pp. 839-887.
59. **Freeman, J. A.** *Redes Neuronales. Algoritmos, aplicaciones y técnicas de propagación*. Mexico : Addison-Wesley , 1993.
60. **Raphael, B. and Domer, B.** *Incremental develoment of cbr strategies for computing projects cost probabilities*. Advanced Engineering Informatics, 2008. 21(3). pp. 311-321.
61. **Howard, R.** *Decision Analysis: Introductory Readings on Choices Under Uncertainty*. US : McGraw Hill, 1997. 0-07-052579-X.
62. **Watson, W and Pollack, L.** *How symbiosis can guide evolution*. Springer-Verlag : Proceedings of the Fifth European Conference on Artificial Life, 2000. pp. 29-38.
63. **Srinivas, M.** *Genetic Algorithms: A Survey*. US : IEEE Computer, 2000. pp. 17-26.
64. **Ahmad S, Bergen.** *A genetic algorithm approach to the inverse problem of treatment planning for intensity-modulated radiotherapy*. 3, US : Biomedical Signal Processing and Control, 2010, Vol. 5, pp. 189-95.
65. **Yeh, J. and Fu, J.** *A hierarchical genetic algorithm for segmentation of multi-espectral human-brain MRI*. 34, Shangai : Expert Systems with Applications, 2008, Vol. 2, pp. 1285-95.
66. **Qiang, F. and Ming. L.** *Analysis of cefalexin with NIR spectrometry coupled to artificial neural networks with modified genetic algorithm for wavelength selection*. 97, Pekin : Chemometrics and Intelligent Laboratory Systems, 2009, Vol. 2, pp. 127-31.
67. **Huang, M. J. and Che, M. Y.** *Constructing a personalized e-learning system based on genetic algorithm and cased-based reasoning approach*. Expert System with Application : s.n., 2009. 33 Vol 3. pp. 551-554.
68. **Chang, P.C. and Lai, C.** *A hybrid system by envolving case-based reasoning with genetic algorithm in wholesaler returning book forecasting*. Decision Support System, 2008. 42(3). pp. 1715-1729.
69. **Martín, L. M.** *Guías y algoritmos clínicos de tratamiento de la depresión en atención primaria*. 37, Madrid : Semergen, 2011, Vol. 2, pp. 54-60.
70. **Yusta, A.** *Algoritmos Clínicos en Medicina*. México : Hospital Universitario Guadalajara. Facultad de Medicina. Universidad de Alcalá, 2009.
71. **Hsu, C. and Ho, S.** *A new hybrid case-based architecture for medical diagnosis*. Pekin : Information Sciences, 2004. Vol. 166. pp. 1-4
72. **Aamodt, A. and Plaza, R.** *Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches*. US : AI Communications. IOS Press, 1994. pp. 39-59.

73. **Kolodner, J. L.** *An Introduction to Case-Based Reasoning*. 1992.
74. **Aamodt, A.** *Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches*. US : AI Communications. IOS Press, 1994. pp. 39-59.
75. **Bergmann, R and Stahl, A.** *Similarity measures for object-oriented case representations. En Advances in Case-Based Reasoning, Lecture Notes in Computer Science*. London : Elsevier Science Publishers, 1998. Vol. 1448. pp. 25-36.
76. **Piñero-Pérez, P. Y.** *Un modelo para el aprendizaje y la clasificación automática basado en técnicas de softcomputing*. Villa Clara, Cuba : Tesis Doctoral, Universidad Central de Las Villas, 2005.
77. **Juarez, j. M. et al.** *Case representation ontology for case retrieval system in medical domains*. s.l. : Proceedings of the 25th International Multi-Conference artificial intelligence and application (IASTED), 2007. pp. 168-173.
78. **Yan, W. et al.** *A novel discovery technique on attribute weight of engine cbr desingn system*. Advanced Materials Research, 2010. 97(101).
79. **Reauda, J. et al.** *Weights determination of owa operators by parametric indentification*. Mathematics and Computers Simulation, 2008. 77(6). pp. 499-511.
80. **Ha, S.** *A personalized counseling system using case-based reasoning with neural symbolic feature weighting*. Applied Intelligence, 2008. 29(3). pp. 279-288.
81. **Gancarski, P. et al.** *Comparison betwen two coevolutionary feature weigting algorithms in clustering*. Pattern Recongnition, 2008. 41(3). pp. 983-994.
82. **Tabucanon, M.** *Multiple criteria decision making in industry*. Tokio : Elsevier Amsterdam - Oxford - New York, 1988.
83. **Michavila, F.** *Fundamentos del calculo numerico I: Topología numérica. Editorial*. Barcelona, España : Reverte, SA, 1986
84. **Moreno, J. M. D.** *Introducción a la topología de los espacios métricos*. Universidad de Cadiz, España : Editorial Servicio Publicaciones, 1998.
85. **Liu, T. et al.** *An investigation of practical approximate nearest neighbor algorithms*. Pittsburg : School of Computer Science, Carnegie-Mellon University, 1998.
86. **Wang, Y.** *Approximating nearest neighbor among triangles in convex position*. Information Processing Letter, 2008. 108(6). pp. 379-385.
87. **Viswanath, P. and Narasimha, M.** *Fusion of multiple apppximate nearest neighbor classifiers for fast and efficient classification*. Information Fusion, 2004. 5(4). pp. 239-250.
88. **Lai, J. and Liaw, C.** *Fast k-nearest-neighbor search based on projection and triangular inequality*. Pattern Recongnition, 2007. 40(2). pp. 351-359.
89. **Finley, A. and McRoberts, R.** *Efficient k-nearest neighbor search for multi-source forest attribute mapping*. Remote Sensing of Enviroment, 2008.112(5). pp. 2203-2211.
90. **Warren, L. and Zhang, L.** *Similarity measures for retrieval in case-based reasoning systems*. Applied Artificial Intelligence, 1998. (12)4. pp. 267-288.
91. **Hamming, R.** *Error detecting and error correcting codes*. 2, Bell System Technical Journal, 1950, Vol. 29, pp. 147-160.
92. **Tversky, A.** *Featuring of similarity*. Psychological Review, 1977. Vol. 84. pp. 2427-2438.

93. **Hung, J. E.** *Case memory and retrieval based on the immune system*. Proceedings of the first international Conference on Case-base Reasoning (ICCBR), 1995. pp. 205-216.
94. **Jaccard, P.** *Etude comparative de la distribution florale dans une portion de alpes et du jura*. Bulletin de Société vandoise des Sciences Naturelles, 1901. pp. 547-579. 37.
95. **Bachelor, B.** *Pattern recognition: Ideas in practice*. New York : Plenum Press, 1978. pp. 71-72.
96. **Morton, N. and Smith, E.** *Pattern recognition engineering*. New York : Wiley, 1993. pp. 293-294.
97. **Zadeh, L.A.** *Fuzzy Set*. Information and Control, 1965. 8(3). pp. 338-353.
98. **Wu, D and Mendel, J. M.** *A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets*. Information Sciences, 2008. 22(1). pp. 381-402.
99. **Armengol, E. and Plaza, E.** *Relational case-based reasoning for carcinogenic activity prediction*. Artificial Intelligence Review, 2003. 20(1). pp. 121-141.
100. **Dengfenga, L. and Chuntianb, C.** *New simillarity measures of intuionistics fuzzy sets and application to pattern recongnition*. Pattern Recognitions Letters, 2002. 23(3). pp. 221-225.
101. **Szmidt, E. and Kacprzyk, J.** *A new concept of similarity measure for intuitionistics fuzzy sets and its use groups decision making*. Lecture Notes in Computers Sciences, 2005. pp. 272-282.
102. **Wang, D and Meng, Y.** *A fuzzy similarity inference method for fuzzy reasoning*. Computer & Mathematics with Applications, 2008. pp. 2445-2454.
103. **Yu, Z and Zhou, L.** *A hybrid similarity measure of contens for tv personalization*. Multimedia Systems, 2010.16(4). pp. 231-141.
104. **Huang, B.W. and Shih, N.** *Price information evaluation and prediction for broiler using adapted case-based reasoning approach*. 5, Expert Systems with Applications, 2009, Vol. 36, pp. 1014-1019.
105. **Leake, D. B. and Kendall, J.** *Four heads are better than one: Combinings suggetions for case adaptation*. Lectures Notes in Computers Sciences, 2009. pp. 165-179.
106. **Neagu, N. and Faltings, B.** *Exploiting interchangeabilities for case adaptation*. Lecture Notes in Computer Science, 2001. pp. 422-436.
107. **Passone, S. and Chung, P.H.** *New knowledge extration technique using probability for case-based reasoning: application to medical diagnosis*. 3, Expert Systems, 2006, Vol. 19, pp. 192-201.
108. **Santiago, S. and Plaza, E.** *Collaborative case retention strategies for cbr*. Lecture Notes in Computers Sciences, 2003. pp. 1063-1063.
109. **Liu, C. H. and Chen, L.** *An association based case case reduction technique for case-based reasoning*. 17. Information Sciences, 2008, Vol. 178, pp. 3337-3355.
110. **Perner, P.** *Case-base maintenace by concept clustering of graphs*. 4. Engineering Applications of Artificial Intelligence, 2006, Vol. 19, pp. 381-393.
111. **Yang, C. and Farley, B.** *Automated case creation and management for diagnostic cbr system*. 1. Applied Intelligence, 2008, Vol. 28, pp. 17-28.

112. **Lio, D.** *Sistemas Basados en Casos. Sistemas Basados en el Conocimiento.* Las Villas : Grupo de Investigación en Inteligencia Artificial. Departamento de Ciencia de la Computación Facultad de Matemática, Física y Computación Universidad Central “Martha Abreu”, 1998.
113. **García, D. L. et al.** *Sistema Experto Basado en Casos para el Diagnóstico de la Diabetes Pediátrica.* La Habana, Cuba : Memorias del Evento INFORMÁTICA, 2011. 978-959-7213-01-7.
114. **Gray, J. et al.** *A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals.* Redmon : Institute of Electrical and Electronics Engineers, 1995. p. 21.
115. **Sellappan, CH. S. L.** *Clinical Decision Support Using OLAP With Data Mining.* 9, Selangor, Malaysia : International Journal of Computer Science and Network Security, 2011, Vol. 8.
116. **Tamayo, M. and Moreno, F.** *Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP.* 3, Bogotá : Ingeniería e Investigación, 2010, Vol. 26. 0120-5609.
117. **Inmon, W. H.** *Building the Data Warehouse.* Indianapolis, Indiana : John Wiley & Son Inc, 2002. 0-471-08130-2.
118. **Kimball, R. and Ross, M.** *The Data Warehouse Toolkit.* Canberra, Australia : John Wiley & Sons Incorporated, 2006. 0-471-15337-0..
119. **Bernabeu, R. D.** *Hefesto: Metodología propia para la Construcción de un Data Warehouse.* Córdoba, Argentina : Dataprix, 2010.
120. **Hernández, L.** *Data Warehouse.* La Habana, Cuba : Italian Cuban Seminar on Informatics Methods and Applications, Departamento de Ciencia de la Computación, Universidad de La Habana, 2002.
121. **Silver, M. et al.** *Case Study: How to Apply Data Mining Techniques in a Healthcare Data Warehouse.* 2, USA : Healthcare Information Management Systems Society, A Publishing Unit of John Wiley & Sons, Inc, 2002, Vol. 15.
122. **Bernabeu, R. D.** *Hefesto: Metodología propia para la construcción de un Data Warehouse.* Córdoba, Argentina, 2007.
123. **González, Y.** *PROPUESTA DE METODOLOGIA PARA EL DAsARROLLO DE ALMACENES DE DATOS EN DATEC.* La Habana : Centro de Tecnología de Gestión Datos, UCI. 2012. T51.
124. **Kimball, R. and Ross, M.** *The Data Warehouse Toolkit.* New York : Wiley Computer Publishing, 2002. 0-471-20024-7. pp. 95-101.
125. **Sánchez, L. Z.** *Metodología para el Diseño Conceptual de Almacenes de Datos.* Universidad Politécnica de Valencia , 2008..
126. **Kimball, R. and Caserta, J.** *The Data Warehouse ETL Toolkit.* Indianapolis : Wiley Publishing, Inc, 2004. 0-7645-7923 -1.
127. **Zúccaro, G.** *El error médico.* Buenos Aires, Argentina : Revista argentina de neurocirugía, 2011, Vol. 24. 1850-1532.
128. **Sánchez, J. J.** *Manual de Análisis de Datos.* Madrid : Alianza Editorial, 1995. pp. 208-211. 9788420681504.
129. **Oliveras, A.** *Hipertensión arterial vascularrenal de origen aterosclerótico. Diagnóstico y manejo terapéutico.* 3, Barcelona, España : NefroPlus, 2009, Vol. 2, pp. 34-44.

130. **Bonzano, A. et al.** *Using introspective learning to improve retrieval in cbr: a case study in air traffic control*. Proceedings of the second International Conference in Case Based Reasoning (ICCBR), 1997. pp. 291-302.
131. **Chen, F. et al.** *Design and realization of case-indexing model based on ahp*. 8. Journal of Software, 2010, Vol. 5, pp. 851-857.
132. **Chen, M. et al.** *A parallelized indexing method for large-scale case reasoning*. 2. Expert system with applications, 2002, Vol. 23, pp. 95-102.
133. **Kim, S.** *The cluster-indexing method for case-based reasoning using self-organization maps and learning vector quantization for bond rating cases*. 3. Applied Intelligence, 2004, Vol. 21, pp. 147-156.
134. **Wang, D. et al.** *Research on ontology-based case indexing in cbr*. Proceedings of International Conference on Artificial Intelligence and Computational Intelligence, 2009. pp. 238-241.
135. **Eibe, F. and Witten, L.** *Data Mining: Practical Machine Learning Tools and Techniques*. Amazon, 2008.
136. **Beall, J.** *The Weaknesses of Full-Text Searching*. 5, Denver : The Journal of Academic, 2008, Vol. 34. 0099-1333. pp. 438-444.
137. **Lefever, E. et al.** *Clustering web people search results using fuzzy ants*. [book auth.] E Lefever. Information Sciences. NY : Elsevier Science Inc, 2010, Vol. 180, pp. 3192-3209.
138. **Fernández, E.** *Propuesta didáctica para la promoción de salud en el caso de la HTA en los pacientes de la Universidad de las Ciencias Informáticas*. La Habana : Tesis para optar por el grado de Máster en Tecnología de los Procesos Educativos, 2008, pp 49-52.
139. **Castellano, F. and García, J.** *Modelos bayesianos para la clasificación supervisada. Aplicaciones al análisis de datos de expresión genética*. Granadas, España : Tesis doctoral, Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granadas, 2009. 978-84-692-8389-9.
140. **Witten, I. H. and Frank, E.** *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann Series in Data Management Systems, 2005.
141. **Blake, C. and Merz, C.** UCI repository of machine learning databases. [Online] University of California, Irvine, Dept. of Information and Computer Sciences, 1998. [Cited: January 21, 2012.] <http://www.ics.uci.edu/mllearn/MLRepository.htm>.
142. **Castellano, F. J.** *Modelos Bayesianos para la clasificación supervisada*. Tesis Doctoral. Departamento de Ciencias de la Computación e Inteligencia Artificial. Granada, 2010.
143. **Quinlan, J.R.** *Discovering rules by induction from large collection of examples*. Knowledge-base system in the Micro Electronics Age, 1979. pp. 168-271.
144. **Quinlan, J. R.** *Programs for machines learning*. Morgan Kaufmann Publishers Inc, 1993.
145. **Cárdenas, C. and Olesia, C.** *Biplot con información externa basado en modelos bilineales generalizados*. Caracas. Venezuela : Universidad Central de Venezuela (Caracas, Venezuela). Consejo de Desarrollo Científico y Humanístico, 2008.
146. **Fernández, A.** *Modelos Probabilísticos para la Inteligencia Artificial y la Minería de Datos*. Almería : Tesis para optar por el grado de Doctor en Ciencias, 2007.

147. **Bravo, M. L. and Arrieta, J. J.** *El Método Delphi. Su implementación en una estrategia didáctica para la enseñanza de las demostraciones geométricas.* Cienfuegos : Revista Iberoamericana de educación, 2006. 1681-5653.

**ANEXOS**

Encuesta No.1																																								
<u>Nombre:</u> <u>Fecha:</u> <u>Cargo o Responsabilidad:</u> <u>Entidad:</u>																																								
<p><b>Nota:</b>  <i>El objetivo de esta encuesta es solamente investigativa. El responsable de esta encuesta se compromete a mantener total privacidad de la información recopilada.</i></p>																																								
<p><b>Pregunta 1</b>            Marque con una X, en una escala del 1 al 10 el valor que se corresponde con el grado de conocimiento o información que usted considera que tiene respecto a <b>HTA y los sistemas informáticos con soporte a la toma de decisiones clínicas</b>. 1 indica que no tiene ningún conocimiento sobre el tema y 10 indica que tiene pleno conocimiento sobre él.</p>																																								
<table border="1" style="margin: auto; border-collapse: collapse;"> <tr> <td style="width: 20px; text-align: center;">1</td> <td style="width: 20px; text-align: center;">2</td> <td style="width: 20px; text-align: center;">3</td> <td style="width: 20px; text-align: center;">4</td> <td style="width: 20px; text-align: center;">5</td> <td style="width: 20px; text-align: center;">6</td> <td style="width: 20px; text-align: center;">7</td> <td style="width: 20px; text-align: center;">8</td> <td style="width: 20px; text-align: center;">9</td> <td style="width: 20px; text-align: center;">10</td> </tr> <tr> <td style="height: 20px;"></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>										1	2	3	4	5	6	7	8	9	10																					
1	2	3	4	5	6	7	8	9	10																															
<p><b>Pregunta 2</b>            Señale con una X el nivel de influencia que ha tenido cada una de las fuentes indicadas en su conocimiento sobre la <b>HTA y los sistemas informáticos con soporte a la toma de decisiones clínicas</b>.</p>																																								
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th rowspan="2" style="width: 50%; text-align: center; vertical-align: middle;">FUENTES DE ARGUMENTACION</th> <th colspan="3" style="text-align: center;">Grado de influencia de cada una de las fuentes en sus criterios.</th> </tr> <tr> <th style="width: 15%; text-align: center;">A (alto)</th> <th style="width: 15%; text-align: center;">M (medio)</th> <th style="width: 15%; text-align: center;">B (bajo)</th> </tr> </thead> <tbody> <tr> <td>Análisis teóricos realizados por usted</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Su experiencia obtenida</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Trabajos de autores nacionales</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Trabajos de autores extranjeros</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Su propio conocimiento del estado del problema en el extranjero</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Su intuición</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>										FUENTES DE ARGUMENTACION	Grado de influencia de cada una de las fuentes en sus criterios.			A (alto)	M (medio)	B (bajo)	Análisis teóricos realizados por usted				Su experiencia obtenida				Trabajos de autores nacionales				Trabajos de autores extranjeros				Su propio conocimiento del estado del problema en el extranjero				Su intuición			
FUENTES DE ARGUMENTACION	Grado de influencia de cada una de las fuentes en sus criterios.																																							
	A (alto)	M (medio)	B (bajo)																																					
Análisis teóricos realizados por usted																																								
Su experiencia obtenida																																								
Trabajos de autores nacionales																																								
Trabajos de autores extranjeros																																								
Su propio conocimiento del estado del problema en el extranjero																																								
Su intuición																																								

**Anexo 1.** Encuesta para determinar el nivel de competencia.

Encuesta No.2
<u>Nombre:</u> <u>Fecha:</u> <u>Cargo o Responsabilidad:</u> <u>Entidad:</u>
<p><b>Nota:</b>            El objetivo de esta encuesta es solamente investigativa. El responsable de esta encuesta se</p>

compromete a mantener total privacidad de la información recopilada.

1) ¿Cree usted que es necesaria la utilización de una herramienta informática para que ayude a la correcta clasificación de los tipos de HTA así como la detección de las causas que la provocan?

Completamente \_\_\_ En gran medida \_\_\_ Poco \_\_\_ Muy poco \_\_\_ Nada \_\_\_

2) ¿Considera usted correcta la elección de los tipos de HTA utilizados por la herramienta informática?

Completamente \_\_\_ En gran medida \_\_\_ Poco \_\_\_ Muy poco \_\_\_ Nada \_\_\_

3) ¿Cree usted que los rasgos predictivos y objetivos identificados son correctos?

Completamente \_\_\_ En gran medida \_\_\_ Poco \_\_\_ Muy poco \_\_\_ Nada \_\_\_

4) ¿Considera usted que es válido el tratamiento sugerido por la herramienta informática?

Completamente \_\_\_ En gran medida \_\_\_ Poco \_\_\_ Muy poco \_\_\_ Nada \_\_\_

5) ¿Considera usted que el umbral de semejanza utilizado por la herramienta informática es adecuado?

Completamente \_\_\_ En gran medida \_\_\_ Poco \_\_\_ Muy poco \_\_\_ Nada \_\_\_

6) ¿Considera usted que las dimensiones o criterios utilizados en la herramienta informática durante la generación de reportes OLAP son adecuados?

Completamente \_\_\_ En gran medida \_\_\_ Poco \_\_\_ Muy poco \_\_\_ Nada \_\_\_

7) ¿Cree usted que a partir de los reportes OLAP generados por la herramienta informática se pueden detectar causas que provocan la HTA?

Completamente \_\_\_ En gran medida \_\_\_ Poco \_\_\_ Muy poco \_\_\_ Nada \_\_\_

**Anexo 2.** Encuesta para determinar el nivel de coincidencia entre la opinión de los expertos.