

Enfoque del Modelo ParFase para el Apoyo a la Toma de Decisiones en el Diagnóstico de Enfermedades

ParFase Model approach to support decision making on diseases diagnosis

García Mejía, José A.*¹; Domínguez Jiménez, Ismael² y Salazar Acosta, J. Javier³

¹Universidad Politécnica Metropolitana de Hidalgo
Tolcayuca, Hidalgo, 43860, México.

²Universidad Autónoma del Estado de Hidalgo
Pachuca, Hidalgo, C.P. 42184, México.

³Área de sistemas del Instituto de Salud (ISEM),
Chalco, Edo. de México., C.P. 56600, México.

*jag.cic.ipn@gmail.com

Resumen

En el presente estudio se desarrolló un modelo denominado ParFase que sirve de base para la construcción de un sistema experto. Este es aplicado como herramienta de apoyo a la toma de decisiones en el diagnóstico de enfermedades. ParFase está conformado por (μ_1, μ_2) , contruidos sobre dos enfoques de razonamiento diferentes: μ_1 está basado en el aprendizaje automático utilizando C4,5; μ_2 utiliza la estadística descriptiva y el simplex inverso; ambos métodos aportan resultados en paralelo para mayor precisión. El caso de estudio se ubica dentro del área veterinaria, en el pre-diagnóstico de las enfermedades caninas diabetes mellitus e hipotiroidismo. Se caracterizaron las enfermedades a partir de sintomatología y expedientes clínicos. La experimentación de ParFase se basa, para μ_1 , en el entrenamiento de una base de casos clínicos para la clasificación de la enfermedad, y validado con k-fold cross validation; para μ_2 , se crea un modelo matemático, que cuenta con una ecuación como función objetivo y la caracterización de las restricciones posibles, para obtener la precisión porcentual de pertenencia. Los resultados experimentales obtenidos están sobre 90 % de precisión.

Palabras clave: Diagnostico asistido por computadora (DAC), estadística descriptiva, pre-diagnóstico de enfermedades, sistema experto.

Abstract

In this work we develop a model called ParFase that is the basis for building a Expert system. This is applied as a tool to support decision making in the diagnosis of diseases. ParFase consists of (μ_1, μ_2) , built approached from two perspectives: μ_1 is based on automatic learning using C4.5; μ_2 uses descriptive statistics and the reverse simplex; Both methods provide results in parallel for greater accuracy. The case study is located within the veterinary area, in the pre-diagnosis of canine diseases; diabetes mellitus and hypothyroidism. Diseases were characterized by symptoms and medical records. ParFase experimentation is based, for μ_1 , base clinical cases training for the classification of the disease, and confirmed with k-fold cross validation; for μ_2 , a mathematical model is created, which has an objective function equation and the characterization of possible restrictions in order to obtain Percentage of ownership accuracy. The experimental results are about 90% of accuracy.

Key words: Descriptive statistics, computer aided diagnosis, pre-diagnosis of diseases, expert system.

1 Introducción

Los sistemas de soporte a la toma de decisiones (SSTD) han evolucionado de forma considerable a lo largo de las últimas décadas. Actualmente para las instituciones es de suma importancia contar con sistemas de este perfil, debido a que proporcionan un contexto de conocimiento más amplio sobre las áreas de estudio. Este tipo de sistemas han mostrado un desarrollo acelerado en su desempeño y los especialistas en el área se apoyan de esta característica para tomar decisiones con mayor precisión. El impacto de adoptar los SSTD en organizaciones ha logrado que este enfoque sea una importante área de investigación (Alavi y Leider, 2001). Debido a su utilidad los SSTD han incursionado en las áreas de gestión clínica y de salud, proporcionando a los usuarios importantes herramientas que son utilizadas para tomar decisiones clínicas sobre un paciente específico (Osheroff, 2009), (Kuper y col., 2007). Los beneficios que se derivan de utilizar este tipo de sistemas permite mejorar la calidad de la atención y reducir el exceso de tratamiento en los pacientes que padecen la enfermedad, así mismo este tipo de sistemas nos permiten entrenar a jóvenes médicos en razonamiento diagnóstico; por lo que en consenso por especialistas del área son de relevante importancia (McGinn y col., 2001), (Berner y col., 2003).

Un área derivada que nos es tan notoria pero no deja de ser importante es el área de medicina veterinaria, en la que, el enfoque es la toma de decisiones basada en evidencias; esta ha sido poco atendida, por lo que, el caso de estudio en esta investigación, se usa para el diagnóstico de enfermedades caninas. En el presente trabajo se muestra el desarrollo de un sistema experto (Russell y Norving, 2002), el cual tiene que alimentar su base de conocimientos de forma a priori para tener un dominio amplio del área, esta información se definen como casos de estudio en el que previamente se diagnosticó la enfermedad con resultado positivo o negativo, a partir de pruebas médicas, análisis clínicos, sintomatología y experiencia de médicos. El módulo ParFase (PFs) representa el núcleo para el procesamiento de datos. El motor de inferencia PFs, por una arista, toma la información de una base de conocimientos de datos históricos con el fin de encontrar similitudes en la base y la sintomatología del paciente que se está analizando, por otra, procesa la información en un modelo matemático propuesto, para complementar la precisión del diagnóstico.

La implementación del sistema experto en cuestión se basa en la arquitectura modelo-vista-controlador (mvc), que permite separar los datos y la lógica de negocio, de la interfaz de usuario. Este sistema presenta módulos de administración clínica de pacientes, consultas, reportes, y el módulo de diagnóstico de la enfermedad. A través de este artículo se realiza una descripción del fundamento teórico de ParFase, la metodología utilizada, los resultados y la experimentación obtenida, así como algunas observaciones del proceso de validación cruzada en los resultados, finalmente en la sección de

conclusiones se describe las ventajas de utilizar este modelo.

2 Sistemas de soporte a la toma de decisiones

A. Sistemas de Soporte a la Toma de Decisiones Clínicas.

Los sistemas de soporte a la toma de decisiones clínicas (SSTDC) han aportado credibilidad a los profesionales de la salud, de acuerdo con Miller y Geissbuhler un sistema de soporte a la toma de decisiones es un algoritmo sistematizado que asiste a profesionales en la salud con uno o más pasos sobre el proceso de diagnóstico (Miller y Randolph, 2007); Musen lo describe como una pieza de software que aplica el enfoque sistémico sobre una situación clínica y produce una salida de inferencias que ayudan a los profesionales a la toma de decisiones, esta pieza de software puede ser interpretada como "inteligente" para los usuarios finales (Musen y col., 2000); recientemente Greenes lo ha definido como el uso de la computadora para brindar conocimiento relevante a ser utilizado en el cuidado de la salud y el bienestar del paciente (Greenes, 2001), (Berner, 2007); el sistema desarrollado en este trabajo se define como un SSTDC que cumple con el enfoque sistémico y que tiene propiedades de razonamiento lógico, precisa una fase de entrenamiento sobre conocimiento clínico histórico y que puede inferir resultados como sugerencias para el pre diagnóstico, este sistema sirve de apoyo a los profesionales de la salud para corroborar su decisión. El área de veterinaria y zootecnia ha sido poco atendida, existen relativamente pocos sistemas reportados en el estado del arte, más aun pocos aplicados a la atención de datos en diagnóstico de enfermedades caninas.

B. Diagnóstico Asistido por Computadora (DAC).

La incursión de las nuevas tecnologías en el ámbito médico ha sido inminente debido al desarrollo de las tecnologías de la información de los últimos años. En el área médica se han incorporado herramientas de software que ayudan a los especialistas a delicada tarea de diagnosticar enfermedades con grandes resultados. A este conjunto de sistemas computacionales se les considera como sistemas para el Diagnóstico Asistido de Computadoras.

C. Sistemas Expertos

La capacidad de procesamiento de un gran conjunto de información histórica alimentada por especialistas en el área, para que una computadora emule el razonamiento de un experto, se considera un sistema experto (SE). Estos sistemas son capaces de analizar múltiples posibilidades y solucionar un conjunto de problemas complejos. Este proceso debe pasar por dos etapas en el motor de inferencia: la parte de entrenamiento, seguido por la parte del motor de inferencia para la clasificación de los datos requeridos. Además el sistema experto debe tener la capacidad para seguir alimentando su base de conocimiento con nuevo conocimiento, generado por el mismo o por nuevos casos clínicos que los especialistas

puedan introducir al sistema. En este artículo se presenta a PFs como un motor de inferencia para clasificar entre las dos enfermedades en estudio.

3 Modulo ParFase.

ParFase está conformado de dos módulos (μ_1, μ_2), contruidos sobre dos enfoques de razonamiento diferentes: μ_1 está basado en el aprendizaje automático utilizando el algoritmo C4,5; μ_2 utiliza la estadística descriptiva y el simplex inverso; ambos métodos aportan resultados en paralelo para obtener mayor precisión. Para realizar el diagnostico el especialista debe ingresar al sistema los signos del paciente y los resultados de las pruebas rápidas; se verifica si existe un historial clínico del paciente, de lo contrario es necesario crearlo, los datos se validan para verificar su veracidad y son almacenados en la base de datos, seguido de esto se inicia el proceso de diagnóstico en el que los datos ingresados por el especialista son procesados por (μ_1, μ_2) de forma paralela; al final se presentan los resultados de ambos módulos para complementar el resultado, este proceso se muestra de forma gráfica en la Fig. 1.

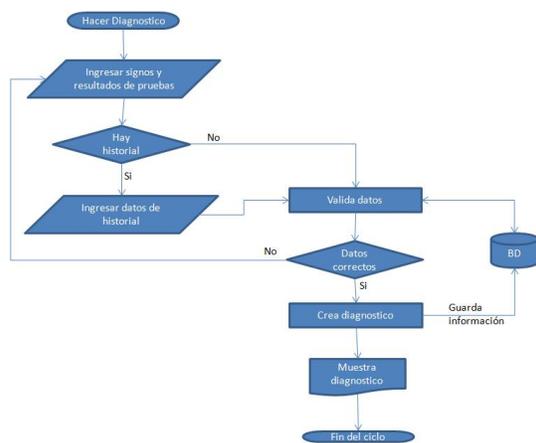


Fig. 1. Diagrama de flujo del proceso de diagnóstico

A. Modulo (μ_1) Algoritmo C4,5

El algoritmo C4,5 (Ramakrishnan, 2009) genera un árbol de decisiones con el conjunto de datos para crear de forma lógica todos los caminos posibles desde las hojas hasta la raíz, utilizando el concepto de entropía de información. Los caminos se van analizando con el fin de medir su eficiencia. Para la etapa de entrenamiento se forma el conjunto indicado en 1 de datos clínicos ya clasificados. Cada caso clínico es un vector donde se representan los atributos o características del ejemplo, como en 1. Los datos de entrenamiento son aumentados con un vector donde se representan la clase a la que pertenece cada muestra o caso clínico. Para PFs el objetivo es clasificar si la enfermedad es diabetes mellitus o

hipotiroidismo. En cada nodo del árbol se elige un atributo de los datos que más eficazmente divide el conjunto de muestras en subconjuntos enriqueciendo una clase u otra. El criterio es normalizado para ganancias de información (diferencia de entropía) que resulta en la elección de un atributo para dividir los datos. El atributo con la mayor ganancia de información normalizada se elige como parámetro de decisión.

$$S = s_1, s_2, s_3, \dots, s_n \tag{1}$$

$$S_i = x_1, x_2, x_3, \dots, x_n$$

$$C = c_1, c_2, c_3, \dots, C_n \tag{2}$$

diabetes aff			
2	xx1	CC	>0.5
3	xx2	Densidad Urinaria	>1.019
4	xx3	Hiperglucemia	>139
5	xx4	Glucosuria	>999
6	xx5	Poliuria	
7	xx6	Polidipsia	
8	xx7	Polifagia	
9	xx8	Perdida de peso repentina	
10	xx9	Problemas Oidos	
11	xx10	Problemas Ópticos	
12	xx11	Proteinuria	
13	xx12	Obesidad	
14			
15	RELATION diabetes		
16			
17	@ATTRIBUTE	x1	REAL
18	@ATTRIBUTE	x2	REAL
19	@ATTRIBUTE	x3	REAL
20	@ATTRIBUTE	x4	REAL
21	@ATTRIBUTE	x5	REAL
22	@ATTRIBUTE	x6	REAL
23	@ATTRIBUTE	x7	REAL
24	@ATTRIBUTE	x8	REAL
25	@ATTRIBUTE	x9	REAL
26	@ATTRIBUTE	x10	REAL
27	@ATTRIBUTE	x11	REAL
28	@ATTRIBUTE	class	{diabetes,sano}
29	@DATA		
30			
31	0.6,1.040,800,2000,1,1,0,0,1,1,0,0,diabetes		
32	0.2,1.030,400,1000,0,0,0,0,1,1,1,0,diabetes		
33	0.4,1.015,140,0,1,1,0,0,0,0,0,0,diabetes		
34	0.4,1.016,140,0,1,1,0,0,1,1,1,0,diabetes		
35	0.8,1.035,400,1000,1,1,1,0,0,1,1,1,diabetes		

Variables que contiene el vector

Etiquetas para la clasificación

Vectores etiquetados

Fig. 2. Estructura de los datos de entrenamiento de casos clínicos

Los atributos que se tomaron en cuenta para cada caso clínico que forma el conjunto de entrenamiento son: Condición Corporal (CC), Densidad Urinaria, Hiperglucemia, Glucosuria, Poliuria, Polidipsia, Polifagia, Perdida de peso repentina, Problemas Oidos, Problemas Ópticos, Proteinuria, Obesidad; estos se puede apreciar en la Fig. 2.

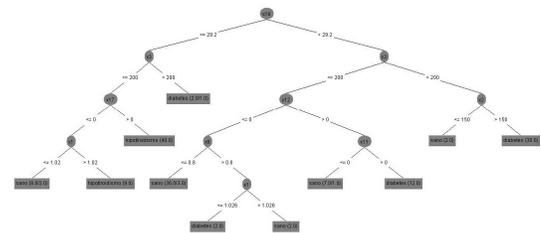


Fig. 3. Árbol de decisión generado por C4,5 para PFs

Las muestras clínicas representadas por cada clase de la enfermedad, así como los casos objetivos que tiene que ser clasificados, son evaluados en el árbol de decisión. En la Fig. 3, se presenta el árbol lógico generado por el algoritmo,

este muestra los datos normalizados para evaluar cada nodo y elegir la clase con mayor ganancia de información.

B. Modulo (μ_2) Modelo matemático.

El modelo matemático, está definido en ParFase como μ_2 , este modelo cuenta con una serie de ecuaciones. Estas ecuaciones son el resultado de utilizar la estadística descriptiva y el método de simplex inverso.

La estadística descriptiva se refiere a la recolección, descripción, análisis e interpretación de una colección de datos. Se utilizó en este trabajo para obtener del conjunto de datos, conclusiones sobre las muestras, después de analizar la información de los expedientes, estas se dividieron en tres rangos: razas chicas, razas medianas y razas grandes.

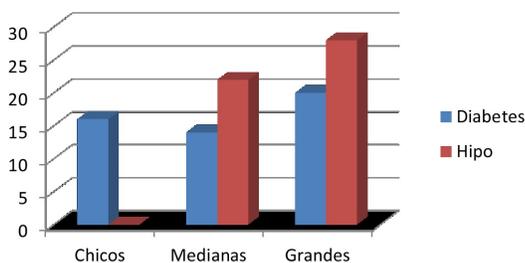


Fig. 4. Muestra de datos de caninos enfermos

Se puede observar en la Fig. 4, que los datos del conjunto muestra la raza con mayor incidencia en ambas enfermedades, es la raza grande", por lo tanto es un factor de gran importancia que se considera en el diagnóstico médico. En μ_2 se define una ecuación como función objetivo y adicionalmente la caracterización de las restricciones posibles, el resultado es obtener la precisión porcentual de pertenencia de la clase.

La **función objetivo**: es la ecuación en la forma ideal que caracteriza todos los factores para determinar cada una de las enfermedades. Se define $f(dm)$ para diabetes mellitus, y $f(ht)$ para hipotiroidismo, la forma general de la ecuación se muestra en 3.

$$F(enfermedad) = C_1X_1 + C_2X_2 + C_3X_3 + \dots + C_nX_n \tag{3}$$

donde C_i es una constante.

Las **restricciones**: son las ecuaciones formuladas con base en la experiencia de los médicos veterinarios y las probabilidades que existen de que aparezca la enfermedad. La forma general de la ecuación de restricción se puede observar en (4).

$$Y_1X_1 + Y_2X_2 + Y_3X_3 + \dots + C_nX_n \tag{4}$$

Dado el espacio muestral, se tiene la necesidad de conocer la probabilidad de encontrar un solo elemento. La ecuación para calcular dicha probabilidad (Martin, y col., 1998), se observa en (5).

$$P(X_i) = \frac{\#(X_i)}{\sum_{i=1}^{\infty} \#(X_i)} \tag{5}$$

Donde:

$\#(X_i)$ Es el numero de ocurrencias del evento.

$\sum_{i=1}^{\infty} \#(X_i)$ Es el total de las muestras.

Esta fórmula es aplicable para conocer los denominados pesos de las funciones de $f(dm)$ y $f(ht)$. Con base a histogramas de cada tamaño de raza, se obtuvieron los valores medios de cada variable (signos) que presenta cada muestra y posteriormente se obtuvieron los pesos principales. Los datos representativos de los datos de los signos de cada una de las enfermedades se pueden observar en la Fig. 5, y la Fig. 6.

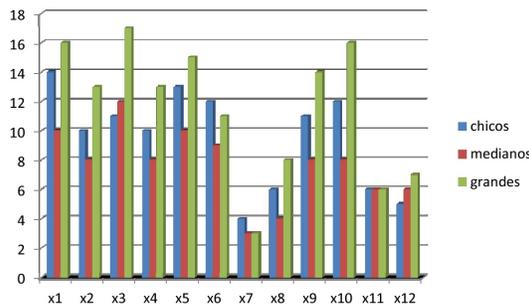


Fig. 5. Histograma de muestras de diabetes mellitus

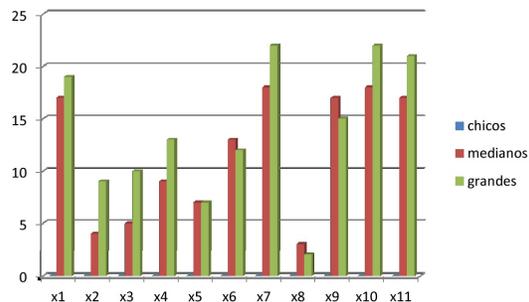


Fig. 6. Histograma de muestras de Hipotiroidismo

Adicionalmente los valores obtenidos de las probabilidades y los valores medios de cada variable de las enfermedades de *diabetes mellitus* e *hipotiroidismo* se muestran en la Tabla 1, y la Tabla 2. Estos representan a los valores de las constantes de la *Función Objetivo*.

Los rasgos representativos de cada variable del conjunto de las X_i , representan las constantes en la Función Objetivo(FO) principal. Para el caso de diabetes

Tabla 1. Probabilidades y valor medio de cada variable para Diabetes Mellitus

Rasgo	Chicos	Medianos	Grandes	Media
X1	12.28	10.99	11.51	11.59
X2	8.77	8.79	9.35	8.97
X3	9.65	13.19	12.23	11.69
X4	8.77	8.79	9.35	8.79
X5	11.4	10.99	10.79	11.06
X6	10.53	9.89	7.91	9.44
X7	3.51	3.3	2.16	2.99
X8	5.26	4.4	5.76	5.14
X9	9.65	8.79	10.07	9.5
X10	10.53	8.79	11.51	10.28
X11	5.26	5.49	4.32	5.02
X12	4.39	6.59	5.04	5.35

Tabla 2. Probabilidades y valor medio de cada variable para Hipotiroidismo

Rasgo	Chicos	Medianos	Grandes	Media
X1	7.89	10.4	9.14	
X2	12.5	13.6	13.05	
X3	14.47	14.4	14.44	
X4	5.92	0.8	3.36	
X5	6.58	4.0	5.29	
X6	1.32	2.4	1.86	
X7	9.87	13.6	11.74	
X8	14.47	14.40	14.43	
X9	4.61	5.6	5.11	
X10	8.55	7.20	7.87	
X11	13.82	13.60	13.71	

mellitus, la forma general de FO se presenta en (6).

$$F(dm) = 11,59X_1 + 8,97X_2 + 11,69X_3 + 8,79X_4 + 11,06X_5 + 9,44X_6 + 2,99X_7 + 5,14X_8 + 9,50X_9 + 10,28X_{10} + 5,02X_{11} + 5,35X_{12} \quad (6)$$

Donde:

- X₁= Condición corporal, X₇= Polifagia
- X₂= Densidad Urinaria, X₈= Perd. de peso repentina
- X₃= Hiperglucemia, X₉= Prob. Óticos
- X₄= Glucosuria, X₁₀= Prob. Ópticos
- X₅= Poliuria, X₁₁= Proteinuria
- X₆= Polidipsia, X₁₂= Obesidad

El siguiente paso es determinar el sistema de ecuaciones de las restricciones, esto es determinar los pesos de las condiciones que los médicos expertos nos proporcionan con base a FO. Tenemos la ecuación general de las restricciones como en (4), de esta manera la ecuación para la obtención de los pesos se presenta en (7).

$$Y_i = C_i \left(\frac{\text{Porcentaje}}{\sum_{i=1}^{12} C_i(X_i)} \right) \quad (7)$$

Donde:

Peso a calcular es=Y_i

Valor de la constante en la FO=C_i

La sum de las constantes de la FO que estan contenidas en las restricciones: $\sum_{i=1}^{12} C_i(X_i)$.

Para finalizar, las ecuaciones de las restricciones resultantes se presentan en (3). Este sistema de ecuaciones se resuelve utilizando el método Simplex, método clásico para la resolución en programación lineal en sistemas que intervienen más de tres variables. Se utilizó el álgebra matricial y el proceso de eliminación por medio Gauss-Jordán para resolver el sistema de ecuaciones lineales.

$$\begin{aligned} &13,3X_1 + 10,29X_2 + 13,42X_3 + 10,29X_4 + 10,9X_9 + 11,08X_{10} \geq 70 \\ &14,51X_1 + 14,64X_3 + 13,85X_5 + 11,82X_6 + 11,9X_9 + 6,28X_{11} \geq 73 \\ &14X_1 + 14,13X_3 + 13,36X_5 + 11,41X_6 + 3,61X_7 + 12,42X_{10} + 6,07X_{11} \geq 75 \\ &16,76X_1 + 16,9X_3 + 15,99X_5 + 13,65X_6 + 7,44X_8 + 7,26x_{11} \geq 78 \\ &12,58X_1 + 12,69X_3 + 12X_5 + 10,24X_6 + 8,58X_8 + 10,31X_1 + 11,16X_{10} + 5,44X_{11} \geq 80 \\ &14,34X_1 + 11,1X_2 + 14,47X_3 + 11,1X_4 + 13,69X_5 + 11,68X_6 + 6,62X_{12} \geq 83 \\ &12,48X_1 + 19,66X_2 + 12,59X_3 + 9,66X_4 + 11,91X_5 + 10,17X_6 + 5,54X_8 + 10,23X_9 + 5,76X_{12} \geq 88 \\ &13,17X_1 + 10,19X_2 + 13,28X_3 + 10,19X_4 + 12,57X_5 + 5,84X_8 + 11,68X_{10} + 6,08X_{12} \geq 83 \\ &13,36X_1 + 13,48X_3 + 12,75X_5 + 10,88X_6 + 5,93X_8 + 10,95X_9 + 11,86X_{10} + 5,79X_{11} \geq 85 \\ &11,13X_2 + 14,5X_3 + 11,13X_4 + 13,72X_5 + 11,77X_9 + 12,75X_{10} \geq 75 \\ &14,31X_2 + 18367X_3 + 14,31X_4 + 17,65X_5 + 15,06X_6 \geq 80 \\ &16,9X_1 + 13,08X_2 + 17,05X_3 + 13,08X_4 + 16,13X_5 + 13,76X_6 \geq 90 \\ &11,91X_1 + 9,22X_2 + 12,01X_3 + 9,22X_4 + 13,37X_5 + 9,7X_8 + 3,07X_{10} + 5,28X_{12} + 10,56X_1 + 5,16X_2 \\ &5,5X_3 \geq 93 \end{aligned} \quad (8)$$

Para fines prácticos se utilizó winQSB para obtener los resultados del sistema de ecuaciones. Después de la resolución del sistema de ecuaciones propuesto en (3), se observó que las variables que tienen más peso para el diagnóstico de la enfermedad de diabetes mellitus son: X₁, X₂, X₄, X₆.

Para el caso de Hipotiroidismo, la forma general de FO se presenta en (9).

$$F(ht) = 9,14X_1 + 13,05X_2 + 14,44X_3 + 3,36X_4 + 5,29X_5 + 1,86X_6 + 11,74X_7 + 14,43X_8 + 5,11X_9 + 7,87X_{10} + 13,71X_{11} \quad (9)$$

Donde:

X_1 = Letargo,	X_7 = Prob. en color de piel
X_2 = Obesidad,	X_8 = Problemas en la piel
X_3 = Alopecia,	X_9 = Problemas óticos
X_4 = Poliuria,	X_{10} = Problemas Ópticos
X_5 = Polidipsia,	X_{11} = T4 Libre
X_6 = Pelo hirsuto	

4 Base de conocimiento

La base de conocimientos se construyó a partir de datos de expedientes o registros de caninos, los cuales ya están diagnosticados con alguna enfermedad en particular. El hospital de medicina veterinaria de la Universidad Nacional Autónoma de México utiliza un sistema proporcionado por la empresa Banfield, el cual contiene registros de toda la población canina que se atiende en el hospital, este sistema es consultado por los médicos veterinarios.

El sistema Banfield contiene más de diez mil registros divididos por enfermedades. Para la búsqueda de expedientes con alguna enfermedad en particular es necesario contar con el número que identifica al paciente, de no tenerlo, se envía una petición a la empresa Banfield, quien se encarga de consultar su base de datos para regresar la respuesta con una lista de números y nombres de los pacientes que tienen dicha enfermedad. Este proceso de solicitud y consulta de información causa un costo monetario adicional, y una inversión de tiempo excesiva, desventajas que se pretenden cubrir con sistemas como el de este trabajo.

Para alimentar a la base de conocimientos se hace uso de 100 expedientes. El sistema analiza dos enfermedades pertenecientes al área médica de Endocrinología: diabetes mellitus, hipotiroidismo. Se analizó la información de acuerdo al contenido de los expedientes y la mayoría de los casos presentan los siguientes datos para determinar un diagnóstico preliminar:

- Resultados de pruebas rápidas. Son los valores de cada una de las pruebas que le hacen al paciente en el momento de la consulta.
- Síntomas que presenta el canino. Son los signos visibles y el resumen de los resultados de las pruebas rápidas.
- Diagnóstico presuntivo. Muestra el listado de las enfermedades relacionadas que posiblemente puede tener el paciente.

- Lista de exámenes médicos. Selección de los exámenes a realizar para los distintos diagnósticos.

Se presentan dos ejemplos de expedientes que fueron analizados y diagnosticados de forma positiva con diabetes mellitus, así como casos diagnosticados con hipotiroidismo. Estos datos se observan en la Tabla 3, y la Tabla 4. Los datos que recopila el especialista en salud de un paciente, debe relacionarse con alguna de estas enfermedades.

Tabla 3. Ejemplos del Analisis de la Enfermedad de Diabetes Mellitus

Paciente	Pruebas Rápidas	Lista de Síntomas	Lista Maestra
1	Glusemia 400 – 800 Multiostix Glucosuria Sin cuerpos cetonicos DU 1.045	3. Poliuria 4. Polidipsia 7. Pérdida de peso 9. Depresión	Diabetes Mellitus (1,2,3,4)
2	Hto 44 Proteínas 82 Glucosa 500 Multiostix 2000 Bilirrubina ++ Cetosnas -- Sangre ++ DU 1,045	1.Hiperglucemia 2.Glucosuria 3.Poliuria 9.Depresion Hipoglucemia Hiporexia Enzimas hepáticas elevadas Convulsiones	Diabetes Mellitus (1,2,3,4,5) Lípidos Hepáticas (7) Inadecuado Insulina (8)

Tabla 4. Ejemplos del Analisis de la Enfermedad de Hipotiroidismo

Paciente	Pruebas Rápidas	Lista de Síntomas	Lista Maestra
1	DU:1,026, CC:4/5 Edad : 5años Peso: 12,70kg Raza: Cocker	Diarrea con moco 1. Alopecia simétrica 2. Aumento de peso 17.Secretión ceruminosa por oídos 13. Seborrea seca 16. Pelo hirsuto	Hipotiroidismo Hiperadrenocortisismo
2	Glucemia 110 Hto: 45 L/L PT: 59 g/L Edad: 11 años Raza: Schnuzer Peso: 7 kg CC: 5/5	Inflamación de glándulas perianales Adenoma de la glándula perianal 2.Obeso 14.Abdomen penduloso 1.Alopecia generalizada 4. Prurito	Abdomen Penduloso Hipotiroidismo

5 Experimentación y Resultados.

5.1 A. Motor de Inferencia ParFase.

Como se plantea al inicio ParFase se compone de dos módulos (μ_1, μ_2), estos forman el motor de inferencia. El modulo (μ_1) tiene dos fases en el proceso del diagnóstico, la fase de entrenamiento descrita en la sección anterior, y

la fase de clasificación. El módulo (μ_2) presenta al modelo matemático con su Función Objetivo (FO). El objetivo principal es que la información se procesa en paralelo sobre estos dos módulos. Los resultados obtenidos al final se complementan para alcanzar una mejor precisión.

El desarrollo de la experimentación para (μ_1) se describe en la fase de clasificación, después de la fase de entrenamiento se presenta un nuevo caso a (μ_1), obtenidos en la interfaz de usuario del sistema de ParFase, se genera un vector normalizado con los datos y este es procesado para obtener con base a su similitud la clasificación de alguna de las dos enfermedades que se proponen, en caso contrario también puede diagnosticar el caso clínico como sano. En la Fig. 6, se observa como son estructurados los datos a clasificar.

```

MuestraNueva.ani
1 $ Rangos
2 %x1 CC >0.5
3 %x2 Densidad Urinaria >1.019
4 %x3 Hiperglucemia >139
5 %x4 Glucosuria >999
6 %x5 Poliuria
7 %x6 Polidipsia
8 %x7 Polifagia
9 %x8 Perdida de peso repentina
10 %x9 Problemas Otricos
11 %x10 Problemas Opticos
12 %x11 Proteinuria
13 %x12 Obesidad
14
15 @RELATION diabetes
16
17 @ATTRIBUTE x1 REAL
18 @ATTRIBUTE x2 REAL
19 @ATTRIBUTE x3 REAL
20 @ATTRIBUTE x4 REAL
21 @ATTRIBUTE x5 REAL
22 @ATTRIBUTE x6 REAL
23 @ATTRIBUTE x7 REAL
24 @ATTRIBUTE x8 REAL
25 @ATTRIBUTE x9 REAL
26 @ATTRIBUTE x10 REAL
27 @ATTRIBUTE x11 REAL
28 @ATTRIBUTE x12 REAL
29 @ATTRIBUTE class {diabetes,sano}
30 @DATA
31
32 0.6,1.040,800,2000,1,1,0,0,1,1,0,0,? Vector a clasificar
    
```

Fig. 7. Estructura de los datos a clasificar

Los resultados de (μ_1), se pueden observar en la Fig. 7, el dato esperado es presentar para el nuevo caso, la sugerencia de pertenencia a la enfermedad.

```

=== Predictions on test data ===

inst#   actual predicted error prediction ()
1       1:1 1:diabetes 1
    
```

Fig. 8. Resultados del algoritmo de clasificación.

La experimentación para (μ_2) está basado en datos reales y la experiencia de los médicos especialistas en el área. El modelo FO aportará como resultado el porcentaje de la enfermedad que presenta el paciente canino. Para obtener esta

información primero se analizan los valores de pruebas médicas rápidas que son introducidas en la interfaz de usuario del sistema experto de ParFase. Esta actividad se puede apreciar en la Fig.9.

Pruebas Medicas Rápidas			
Peso actual	CC actual	Glucosa	Glucosa en la orina
11.5 kg.	205	250 mg/dl	0
Densidad Urinaria	Proteínas	T4 Libre	
1.025	0 g/L	160 pmol/L	

Fig. 9. Datos de las pruebas médicas para el modelo matemático.

El médico especialista también necesita introducir las observaciones pertinentes de acuerdo a las observaciones del médico. Estas observaciones se introducen en una interfaz gráfica, que se observa en la Fig. 10. El sistema experto de ParFase contiene un analizador léxico que identifica las palabras clave de los síntomas de la enfermedad.

Motivo De La Consulta (Propietario)

La propietaria comenta que la perrita no ha comido bien, pero orina mucho y toma mucha agua. También comenta que esta sin animos.

Observaciones (Veterinario)

El paciente presenta cataratas. De igual forma presenta secreción seruminosa en ojos y presenta prurito.

Ingresar manualmente los signos

Fig. 10. Datos de historial médico y observaciones.

Después se genera una lista que relaciona las palabras clave introducidas por el médico, con la lista que identifica la enfermedad en cuestión. La lista de síntomas del tipo de enfermedad se puede apreciar en la Fig. 10. A partir de esta lista y de acuerdo a los síntomas que encontró el especialista se genera un vector de signos identificados, este vector es normalizado y estos datos son tomados en cuenta para que el modelo matemático calcule la pertenencia de la enfermedad. En la Fig. 11, se puede observar los vectores generados con base en las observaciones del médico veterinario.

El vector normalizado es procesado por el modelo matemático, después se calcula la FO. Los resultados son mostrados en una interfaz gráfica para el usuario, como se puede apreciar en la Fig. 13.

5.2 B. Validación de ParFase

Para determinar la probabilidad de acierto del algoritmo de clasificación en el módulo (μ_1) ante nuevos casos, se utiliza el método de validación cruzada conocido como *k – fold cross validation* [13], en el que básicamente consiste



Fig. 11. Vectores de signos encontrados.

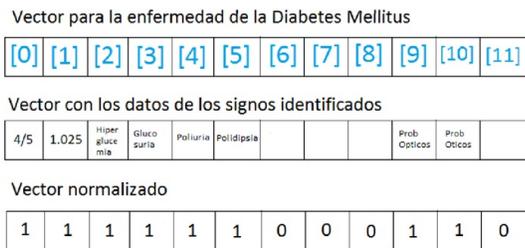


Fig. 12. Vectores generados para el diagnóstico.

en dividir el conjunto total de casos en K subconjuntos disjuntos. De estos k subconjuntos, uno se reserva como datos de validación para probar el motor de inferencia, y los restantes $k - n$ se utilizan como datos de entrenamiento. El proceso se repite K veces (con cada uno de los k , usados solo una vez como datos de validación).

En la Tabla 5, se muestra el análisis y los resultados de la validación cruzada realizados en la base de conocimientos de diabetes mellitus. Podemos observar el comportamiento del resultado del algoritmo al cambiar el tamaño de las muestras (folds). Se muestra que para el $k - fold = 15$, el porcentaje de error al momento de comparar etiquetas, resulta con el porcentaje de error más bajo en comparación con los demás. Por lo tanto, de acuerdo con los resultados que se muestran en la Tabla 5, se indica que con una muestra de fold con $k = 15$ el algoritmo tiene un mejor desempeño en su aprendizaje.

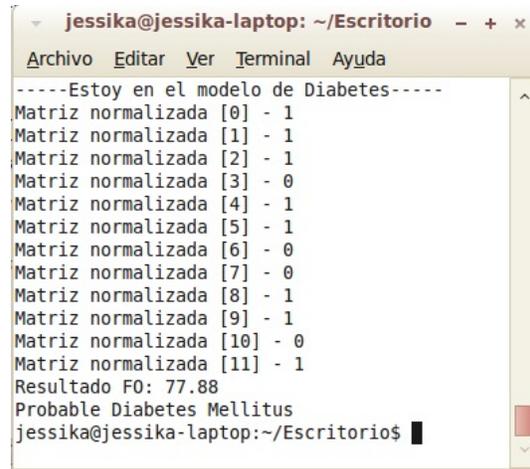


Fig. 13. Vectores normalizados y resultado del modelo matemático.

Tabla 5. Resultados de validación cruzada para la Enfermedad de Diabetes

k-fold	% Aciertos	% Error	# Muestras Clasificadas			
			Correctamente		Incorrectamente	
			Diabetes	Sanos	Diabetes	Sanos
2	83	17	37	46	13	4
5	89	11	42	47	8	3
10	90	10	42	48	8	2
15	93	7	45	48	5	2
25	92	8	45	47	5	3
50	92	8	45	47	5	3
99	92	8	45	47	5	3

5.3 C. Resultados.

Para verificar el funcionamiento del motor de inferencia de ParFase, se hicieron pruebas con 75 muestras: 25 con diabetes mellitus, 25 con hipotiroidismo, y 25 casos clínicos saludables. El enfoque final que se tomo es complementar los resultados aportados por (μ_1) , y complementarlos con los resultados del modelo matemático (μ_2) . Los resultados obtenidos se puede observar en la gráfica de la Fig. 14. La descripción de las etiquetas de la Fig. 3, es: Conjunto, son todas las muestras clasificadas por (μ_1) de forma correcta con el porcentaje de 85,33 % obtenido del modelo; Ayuda FO, son todas aquellas muestras donde la clasificación de (μ_1) fue errónea, pero que el modelo de (μ_2) obtuvo porcentajes altos, en alguna de las clases; Erroneos, son aquellas muestras donde ambos (μ_1) y (μ_2) tuvieron falsos negativos, el porcentaje fue 8 %. Finalmente si complementamos los resultados de (μ_1) y (μ_2) obtenemos una precisión del 91,99 %, al momento de clasificar.

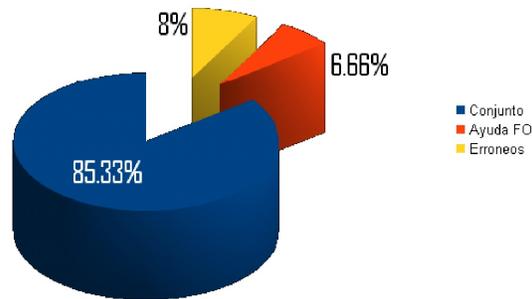


Fig. 14. Resultados finales de (μ_1) con complemento del modelo matemático (μ_2) .

6 Conclusiones

En este trabajo se presenta una alternativa para el análisis, clasificación, y la inferencia del pre diagnóstico de enfermedades caninas. Se presenta a ParFase como un modelo que complementa los resultados de (μ_1) y (μ_2) . Los resultados obtenidos demuestran una mayor precisión. Esta alternativa define un buen aporte para la medicina veterinaria, originada a partir de la experiencia de los médicos que se definen en el modelo matemático. Cabe mencionar que cualquier enfermedad puede ser analizada bajo este esquema; se define su FO y las restricciones para proveer a ParFase como un sistema experto que puede crecer.

Agradecimientos

Se agradece el apoyo recibido para el desarrollo de este proyecto a la Universidad Politécnica Metropolitana de Hidalgo - UPMH, y a PROMEP - SEP-23-03, México.

Referencias

- Alavi M., and Leidner D. E., 2001, Knowledge management and knowledge management systems: Conceptual foundations and research issues, MIS quarterly, pp. 107-136.
- Osheroff J. A., 2009, Improving Medication Use and Outcomes with Clinical Decision Support: A Step by Step Guide. HIMSS.
- Kuperman G. J., Bobb, A., Payne T. H., Avery A. J., Gandhi T. K., Burns G. and Bates D. W., 2007, Medication-related clinical decision support in computerized provider order entry systems: a review, Journal of the American Medical Informatics Association: JAMIA, 14(1), 29-40.
- McGinn T. G., McCullagh L., Kannry J., Knaus M., Sofianou A., Wisnivesky J. P. and Mann D. M., 2013, Efficacy of an evidence-based clinical decision support in primary care practices: a randomized clinical trial. JAMA internal medicine, 173(17), 1584-1591.
- Berner E. S., Maisiak R. S., Heudebert G. R. and Young Jr K. R., 2003, Clinician performance and prominence of

- diagnoses displayed by a clinical diagnostic decision support system. In AMIA Annual Symposium Proceedings, Vol. 2003, p. 76, American Medical Informatics Association.
- Russell S. and Norvig P., 2002, Artificial Intelligence: A Modern Approach (2dn Edition), Prentice Hall.
- Miller, Randolph A. G., 2007, Diagnostic decision support systems, Clinical Decision Support - Theory and Practice, Prentice, Hall Second Edition - Capítulo 5.
- Musen, Mark A., Shahar, Yuval, Shortliffe Edward H., 2000, Clinical decision-support systems, Chapter 16 in Medical Informatics: Computer Applications in Health Care and Biomedicine.
- Greenes, A. Robert, 2006, Clinical Decision Support: The Road Ahead, Academic Press.
- Berner, E. S., 2007, Clinical decision support systems, Vol. 233, New York: Springer Science+ Business Media, LLC.
- Ramakrishnan Naren, 2009, Algorithms in data mining - C4.5, CRC Press, New York.
- Martín, J. A. G., and Callejón, J. M. P., 1998, Sistemas expertos probabilísticos (Vol. 20). Univ de Castilla La Mancha.
- Stone, M., 1974, Cross-validators choice and assessment of statistical predictions. Journal of the royal statistical society. Series B (Methodological), 111-147.

Recibido: 25 de marzo de 2016

Aceptado: 28 de enero de 2017

García Mejía, José Antonio: Es profesor investigador de la Universidad Politécnica Metropolitana de Hidalgo, del P.E. de Ingeniería en Tecnologías de la Información. e-mail: jag.cic.ipn@gmail.com

Domínguez Jiménez, Ismael: Es profesor investigador de la Universidad Autónoma de Estado de Hidalgo. e-mail: ismael@uaeh.edu.mx

Salazar Acosta, J. Javier: Investigador del Instituto de Salud (ISEM). e-mail: phraxoss@gmail.com

