

XVI Fórum de Ciencia y Técnica

ConFuCiuS, Sistema Basado en Casos Borroso Conexionista y sus aplicaciones.

Autores:

Yanet Rodríguez Sarabia¹

María Matilde García Lorenzo¹

Carlos Morell Pérez¹

Rafael Bello Pérez¹

Ricardo Grau Avalo¹

Rafael Falcón Martínez¹

Mabel González Castellanos¹

Mayelis Espinosa García¹

Adonis Aguirre Roque²

Haydee Leyva Castellano¹

Maybel Martín Pérez¹

Eduviel Ramos Cárdenas³

Paulino Hernández Hernández³

¹Centro de Estudios de Informática, Facultad de Matemática-Física y Computación

Universidad Central “Marta Abreu” de Las Villas

²Hospital Pediátrico Universitario “José Luis Miranda”

³Hospital Universitario “Arnaldo Milián Castro

Síntesis del trabajo

En este trabajo se presentan un modelo híbrido para resolver problemas de clasificación usando técnicas de Inteligencia Artificial (IA), así como las aplicaciones hasta ahora desarrolladas y que muestran su valor práctico en la solución de problemas concretos, así como impacto social y económico reportado por la aplicación de estos productos de software desarrollados. Este sistema se referencia como ConFuCiuS, y su modelo computacional combina de una forma coherente una red neuronal artificial de tipo asociativa y el razonamiento basado en casos con elementos de la Teoría de los Conjuntos Borrosos (*Fuzzy Set Theory*).

Los sistemas computacionales desarrollados con este modelo garantizan disminuir la ingeniería del conocimiento requerida en este tipo de sistemas, y al mismo tiempo permiten:

- Una representación más natural y efectiva de los rasgos de tipo numérico que describen un problema.
- Un desempeño al menos comparable con otros clasificadores reportados en la bibliografía, y mostrando los mejores resultados en dominios de aplicación cuyas muestras se caracterizan por atributos mezclados y con ausencia de información.
- Interpretar y explicar la solución dada a un problema concreto en el contexto de experiencias pasadas

La novedad científica de la propuesta esta dada por:

- La definición de dos familias de funciones de distancia que generalizan la Métrica Basada en Valores (VDM), concebida originalmente para tratar atributos nominales, al procesamiento de atributos numéricos. Dichas generalizaciones se fundamentan en un criterio de comparación local uniforme para cualquier tipo de atributo. El concepto básico con el cual se trabaja en esta generalización es la noción de variable lingüística, juntamente con la partición borrosa que describe sus valores. Ambas familias de funciones reflejan un enfoque unipolar más tradicional basado en cardinalidades (frecuencias relativas) y un enfoque bipolar contemporáneo basado en uninormas y coeficientes de correlación. Tal enfoque bipolar está en línea con los últimos avances en la toma de decisiones multi-criterio, en particular con los modelos basados en la integral de Choquet, pero el mismo ha sido desarrollado aquí de manera muy independiente y original. Este primer aporte por sí solo pudiera ser de interés para una variedad de disciplinas, dada la popularidad que ha ganado el hecho de medir la similitud entre los objetos.
- La concepción del nuevo modelo híbrido para resolver problemas de clasificación con aprendizaje basado en instancias, y los resultados obtenidos de su validación utilizando conjuntos de datos internacionalmente reconocidos con esta finalidad.

La aplicabilidad, impacto y pertenencia de los resultados se propicia a partir de:

- Su implementación computacional en un moderno ambiente de aprendizaje automatizado (WEKA, *Waikato Environment for Knowledge Analysis*), lo cual propicia además la visibilidad científica de estos resultados.
- La implementación de varios productos de software útiles a terceros: ConFuCiuS 1.0, NeuroEvaluator 1.0, SAPRIM 1.0 y NeproSystem 1.0, ya registrados en CENDA (Aval 1 del Anexo I) o en proceso.

Los productos de software desarrollados han tenido dos alternativas de aplicación:

- Como medios de enseñanza para el desarrollo de habilidades prácticas en la construcción de sistemas inteligentes, tanto en cursos de pregrado como postgrado (Aval 2, 3, 4 del Anexo I); lo cual está muy acorde con la política educacional actual del estado cubano, en el cual se potencia el uso de las TIC como medios de enseñanza.
- Para desarrollar sistemas basados en el conocimiento para resolver problemas concretos, permitiendo simplificar el proceso de ingeniería del conocimiento así como el desarrollo de productos de software a la medida de los usuarios de estas aplicaciones.
- Para asistir a los especialistas en la toma de decisiones en dominios específicos de aplicación. Se presentan avales de los clientes en el anexo I: sobre aplicación de SAPRIM (Aval 5,6 y7), sobre NephroSystem (Aval 8) y sobre la factibilidad de aplicación de ConFuCiuS en aplicaciones medio ambientales (Aval 9).

La pertinencia de la investigación desarrollada se muestra por:

- Estar enmarcada en un proyecto financiado por el CITMA dentro del Programa Nacional Científico-Técnico de Tecnologías de la Información, y con resultados premiados CITMA provincial en el 2006 (Aval 1 del anexo II) y Tesis de doctorado defendida en el tribunal de Ciencias técnicas en Computación y Automática (Aval 2 del anexo II).
- Proyecto contratado a nivel ramal por el MINSAP (aval 3 del anexo II).
- Premio Provincial en la 12 Exposición Forjadores del Futuro de las BTJ (aval 4 del anexo II).
- Valoración del impacto social y económico del sistema SAPRIM (aval 5 del anexo II).

Lo antes expuesto se avala por la defensa de una tesis de doctorado, por la presentación de los resultados en eventos científicos (anexo III); así como su publicación en revistas del *web of science*, referenciadas o como capítulo de libros, algunas de las cuales serán citadas a continuación en la descripción más detallada que se presenta del trabajo.

Descripción detallada

Los problemas de clasificación aparecen prácticamente en todos los dominios de aplicación, por eso la solución a esta clase de problemas se ha enfrentado desde diferentes disciplinas. Un factor importante para decidir desde cuál perspectiva abordarlo es el tipo de conocimiento necesario para realizar la clasificación, el cual puede ser estructurado y no estructurado. En el primer caso contamos con un algoritmo que nos permite realizar la clasificación del evento u objeto a partir de su descripción, mientras que en la clasificación no estructurada no contamos con esa vía de clasificación. En esta última alternativa es donde las técnicas de Inteligencia Artificial (IA) son aplicables y necesarias para poder construir un procedimiento o modelo de clasificación.

Se han considerado diferentes técnicas de la IA para resolver un problema de clasificación, entre ellas: los Sistemas Basados en Reglas, los Sistemas Basados en Casos y las Redes Neuronales Artificiales. Pero la IA ofrece otra ventaja en la solución de los problemas de clasificación, y es que además de aportar estos modelos para la construcción de clasificadores, ofrece técnicas para ayudar en la construcción de los mismos mediante el aprendizaje automático. Por otro lado, en muchos campos de aplicación como el diagnóstico médico¹, generalmente los especialistas relacionan síntomas (valores de los rasgos predictores) con un diagnóstico (clasificación), relación dependiente del contexto. Al mismo tiempo es poco usual que la relación predictiva entre atributos de naturaleza numérica y la solución de un problema detalle cada valor específico del dominio de un rasgo, sino a valores que puedan ser representativos de un grupo de valores próximos a él, a los cuales usualmente se refieren mediante términos (ejemplo, “fiebre alta”, “bajo peso”). Luego, aunque el algoritmo admita atributos numéricos, el tratamiento de esta información en el modelo computacional no se correspondería con la del especialista. Adicionalmente, mucho de estos términos no tienen siempre límites bien definidos, lo que constituye una fuente de imprecisión.

Utilizar los conjuntos borrosos para representar este tipo de información, y contar con un modelo computacional “que razone” considerando estas reflexiones, garantizaría acercarnos a la forma en que los humanos resuelven estas tareas. Utilizar conjuntos borrosos, además de propiciar mayor naturalidad en la modelación computacional de un problema, a su vez permite manejar el grado de pertenencia del valor real a cada término lingüístico como una medida de similitud o cercanía con los elementos representativos.

A continuación se describe brevemente el modelo computacional, las características generales de sus implementaciones computacionales y las aplicaciones hasta ahora desarrolladas.

¹ Kuncheva, L. and F. Steimann (1999). "Fuzzy diagnosis." *Artificial Intelligence in Medicine* 16(2): 121-128.

1 Descripción del modelo computacional

Se propone un nuevo modelo híbrido para resolver problemas de clasificación con aprendizaje basado en instancias, el cual se referenciará como ConFuCiuS; y que a diferencia de los modelos puramente conexionistas esencialmente en posponer la generalización hasta recibir una solicitud q (aprendizaje perezoso). Los sistemas desarrollados a partir de éste muestran un desempeño al menos comparable con los clasificadores existentes, y los mejores resultados se obtienen cuando el conjunto de instancias contiene valores ausentes y mezclados así como con rasgos irrelevantes [ROD07c]² [ROD07b]³.

ConFuCiuS sigue un enfoque híbrido conexionista borroso, y se define como una instancia del clasificador basado en los vecinos más cercanos; utilizando un nuevo criterio de comparación derivado de la Métrica Basada en la Diferencia de Valores (VDM) para seleccionar los casos más similares. La implementación del concepto de similitud (distancia) global, se define como la suma de criterios de comparación a nivel de rasgo siguiendo el principio global-local. La definición del criterio de comparación local (unidimensional) se basa en la relación dinámica de cada valor posible para ese rasgo con la clase correspondiente (factor diagnóstico de un rasgo) basado en conocimiento del dominio (un conjunto de instancias).

La propuesta consiste en extender sobre variables lingüísticas [ROD05]⁴ [ROD06a]⁵ y generalizar la métrica VDM que ha mostrado buenos resultados en problemas de clasificación y en presencia de rasgos irrelevantes, pero solo aplicable en atributos simbólicos. Como resultado resulta novedoso la definición de dos nuevas familias de funciones de distancia para comparar dos instancias, que se basa en un criterio de

² **Generalización de la Métrica Basada en la Diferencia de Valores para Variables Lingüísticas y su Aplicación en Sistemas Basados en el Conocimiento.** Tesis Doctoral Defendida por Yanet Rodríguez Sarabia en el Tribunal Nacional de Automática y Computación, Año 2007. Tutor Dra. María M. García Lorenzo de la UCLV.

³ **Integrating ANN and CBR using Fuzzy Sets to develop Hybrid Systems. Hybrid Artificial Intelligence Systems. HAIS 2006. ISBN: 84-934181-9-6. Javier Bajo, Emilio S. Corchado, Álvaro Herrero and Juan M. Corchado (Eds.). Universidad de Salamanca. Salamanca. España. 2007. pp 101-108.** Yanet Rodríguez, María M. García, Bernard De Baets, Rafael Bello y Carlos Morell.

⁴ **Extending CBR-ANN Hybrid Models Using Fuzzy Sets.** Proceedings of International Conference on Neural Networks and Brain, 2005 (ICNN&B'05). Vol. 3, pp. 1755-1760. ISBN 0-7803-9422-4. Oct. 2005, China. Printed by IEEE Computer Society. Yanet Rodríguez, María M. García, Rafael Bello and Rafael Falcón.

⁵ **Extending a Hybrid CBR-ANN Model by Modeling Predictive Attributes using Fuzzy Sets. Lecture Notes on Artificial Intelligence 4140, Springer Verlag (2006), pp 238-248.** Yanet Rodríguez, María M. García, Bernard de Baets, Rafael Bello and Carlos Morell

comparación local conceptualmente uniforme para todo tipo de rasgo, a partir de extender la métrica VDM sobre variables lingüísticas y generalizar el concepto de asociación entre valores.

1.1 Caracterización general de ConFuCiuS

El modelo ConFuCiuS se concibe como una variante del k -NN utilizando como criterio de comparación a nivel de rasgo un ejemplo particular de las familias de distancia que se proponen en este trabajo. A continuación se describe atendiendo a cada uno de los aspectos que caracteriza el aprendizaje basado en instancias:

- Se tienen almacenadas instancias de entrenamiento que se representan como pares de atributo-valor.
- El valor para la clase del problema a resolver se corresponde con la clase que más aparece (más probable) en los k casos más similares recuperados (*majority vote*⁶), luego de calcular la probabilidad de cada clase atendiendo a las siguientes expresiones:

$z = ArgMax_{l=1}^c P_r(C_l/q)$	(1)
$Pr^k(C/q) = \frac{\sum_{i \in K_q} C(z_i)}{ K_q }$	(2)

donde K_q es el conjunto de los k vecinos más cercanos, subconjunto del conjunto de instancias de entrenamiento; y que se seleccionan como resultado de comparar cada instancia de entrenamiento x con la solicitud q . Para cada conjunto de instancias se determina un valor de k óptimo o ideal aplicando el método de validación cruzada dejando uno fuera, que se refiere al valor de $k= 1..15$ que minimice el LOOCE (leave-one-out classification error⁷).

- Para comparar dos instancias (en el espacio m -dimensional) se utiliza la distancia Euclideana que emplea el k -NN estándar; considerando como criterio de comparación a nivel de rasgo una de las funciones de distancia propuestas también en el trabajo. Si para este criterio utilizamos la expresión general FADM, resulta la expresión siguiente:

⁶ Wettschereck, D., D. W. Aha, et al. (1997). "A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms." Artif. Intell. Rev. **11**(1-5): 273-314.

⁷ Wilson, D., R. Randall, et al. (2000). "An integrated instance-based learning algorithm." Computacional Intelligence **16**(1): 1-28.

$$d_E(q, e) = \sqrt{\sum_{j=1}^m (d_{FADM}(q_j, d_j))^2} \quad (3)$$

A este aspecto del modelo se debe el alias “ConFuCiuS” (Connectionist Fuzzy Case-based System), por su estrecha relación con el modelo conexionista antes explicado [ROD06b]⁸. Nótese que el criterio de comparación a nivel de rasgo (FADM en general) se define a partir de la función $I(x, C)$; la cual se basa en la asociación entre los conjuntos borrosos $A_j \in F(X)$ y $C \in F(Y)$ y se corresponde en el modelo FSIAC con el valor del peso $w_{A,C}$ asociado al enlace entre las neuronas correspondientes a sus términos lingüísticos. En particular, en [ROD07a]⁹ se presenta la variante de este modelo utilizando FCDM.

1.2 Dos nuevas familias de funciones de distancia basadas en la asociación entre conjuntos borrosos

Se considera que dos valores para un rasgo predictor¹⁰ son similares en la medida que, en un conjunto de instancias conocidas, la asociación (o relación) de estos valores sea similar con cada uno de los valores posibles del rasgo objetivo¹¹. La propuesta se basa en considerar para cada atributo numérico una partición borrosa sobre su dominio; y definir un criterio de comparación a nivel de rasgo (local) mediante la combinación, utilizando operadores de agregación de la lógica borrosa, del grado de pertenencia (o compatibilidad) de un valor numérico con cada término lingüístico definido sobre su dominio y una medida que cuantifique la asociación entre dos conjuntos borrosos.

Definición 1: Métrica basada en la diferencia de asociaciones borrosas (FADM, Fuzzy Association Difference Metric).

Sea X el dominio de un atributo predictor, Y el dominio del atributo objetivo. Sean $x, y \in X$ y $\Pi(Y) = \{C_1, C_2, \dots, C_c\}$ una partición borrosa sobre Y . La distancia entre dos elementos $x, y \in X$ $d_{FADM}: X \times X \rightarrow \mathbb{R}^+$ se define por la siguiente expresión:

⁸ **A Connectionist Fuzzy Case-Based Reasoning Model.** A. Gelbukh and C.A. Reyes-Garcia (Eds.): **Lecture Notes on Artificial Intelligence 4293.** Springer-Verlag Berlin Heidelberg (2006), pp 176–185. Yanet Rodríguez, María M. García, Bernard De Baets, Carlos Morell, and Rafael Bello.

⁹ **Improving a Fuzzy ANN Model using Correlation Coefficients.** *Advances in Soft Computing.* Springer-Verlag . Berlin/Heidelberg 2007. ISSN 1615-3871. ISBN 978-3-540-72433-9. Yanet Rodríguez, Bernard De Baets, María M. García, Ricardo Grau, Carlos Morell, and Rafael Bello.

¹⁰ Un rasgos predictor se refiere a un rasgo que describe una instancia

¹¹ Es una rasgo de salida para el cual se desea inferir un valor

$$d_{FADM}(x, y) = \sqrt{\sum_{l=1}^c (I(x, C_l) - I(y, C_l))^2} \quad (4)$$

donde $I(x, C)$ es una medida de la influencia de un elemento $x \in X$ en la $C \in \mathcal{I}(Y)$ según definición 2.1.

Otro aspecto se refiere al tratamiento de los valores ausentes en la comparación, el cual queda implícito en la expresión anterior por la relación entre la teoría de los conjuntos borrosos y la teoría de la posibilidad. Si el valor de x se desconoce para la i -ésima instancia ($x_i = ?$), se interpreta que este evento tiene máxima compatibilidad con cada uno de los términos lingüísticos $A_j \in F(X)$; es decir, $A_j(x_i) = 1, j = 1, \dots, N$.

En la bibliografía consultada no se reportan funciones que se basen en conjuntos borrosos, y a su vez en la relación del valor del rasgo a comparar con el rasgo objetivo basado en conocimiento del dominio (conjunto de instancias). En este trabajo se definen dos nuevas familias de funciones de distancia con esta característica, en correspondencia con las dos variantes para calcular la asociación entre dos conjuntos borrosos.

I) Extensión de VDM sobre variables lingüísticas

Definición 2: Familia de métricas basadas en la diferencia de valores borrosos (FVDM, Fuzzy Value Difference Metric).

Sea X el dominio de un atributo predictor, Y el dominio del atributo objetivo. Sean $A, B \in \mathcal{I}(X)$ y $\mathcal{I}(Y) = \{C, C_2, \dots, C_c\}$ una partición borrosa sobre Y . La distancia entre los conjuntos borrosos A, B se define como: $d_{FVDM}: \mathcal{I}(X) \times \mathcal{I}(X) \rightarrow \mathbb{R}^+$, cuya expresión es:

$$d_{FVDM}(A, B) = \sqrt{\sum_{l=1}^c (S^T(A, C_l) - S^T(B, C_l))^2} \quad (5)$$

donde $S(A, C)$ es la medida de asociación entre dos conjuntos borrosos basado en frecuencia relativa siendo la t-norma T el parámetro que define esta familia.

Nótese que independientemente de la t-norma utilizada en la función $S(\cdot)$, en un problema de clasificación con clases duras este criterio de comparación para un rasgo simbólico se reduce a la métrica VDM propuesta por Stanfill y Waltz como caso particular.

Definición 3: Familia FVDM entre elementos.

Sea X un atributo predictor y $\mathcal{I}(X) = \{A_1, A_2, \dots, A_N\}$ una partición borrosa definida sobre X . Sea Y el dominio del atributo objetivo y $\mathcal{I}(Y) = \{C_1, C_2, \dots, C_c\}$ una partición borrosa sobre Y , $C \in \mathcal{I}(Y)$. La distancia entre los elementos $x, y \in X$ se define como $d_{FVDM}: X \times X \rightarrow \mathbb{R}^+$, cuya expresión es:

$$d_{FVDM}(x, y) = \sqrt{\sum_{l=1}^c \left(\sum_{j=1}^N A_j(x) S^T(A_j, C_l) - \sum_{j=1}^N A_j(y) S^T(A_j, C_l) \right)^2} \quad (6)$$

II) Familia de funciones basada en coeficientes de correlación

Definición 4 Métrica basada en la diferencia de correlaciones borrosas (FCDM, Fuzzy Correlation Difference Metric).

Sea X el dominio de un atributo predictor, Y el dominio del atributo objetivo. Sean $A, B \in \mathcal{I}(X)$ y $\mathcal{I}(Y) = \{C_1, C_2, \dots, C_c\}$ una partición borrosa sobre Y . La familia de distancia FCDM para comparar conjuntos borrosos se define como $d_{FCDM}: \mathcal{I}(X) \times \mathcal{I}(X) \rightarrow \mathbb{R}^+$, cuya expresión es:

$$d_{FCDM}(A, B) = \sqrt{\sum_{l=1}^c (r_{A_j, C_l} - r_{B_j, C_l})^2} \quad (7)$$

donde $r_{A,C}$ es la medida de asociación entre dos conjuntos borrosos basado en el coeficiente de correlación de Spearman.

Definición 5 Métrica basada en la diferencia de correlaciones (CDM, Correlation Difference Metric).

Cuando los conjuntos A y C son conjuntos ordinarios como caso particular, resulta una nueva función de distancia para comparar atributos simbólicos y que referenciaremos como CDM (Correlation-Difference Metric).

Definición 6 Familia FCDM entre elementos.

Sea X un atributo predictor y $\mathcal{I}(X) = \{A_1, A_2, \dots, A_N\}$ una partición borrosa definida sobre X . Sea Y el dominio del atributo objetivo y $\mathcal{I}(Y) = \{C_1, C_2, \dots, C_c\}$ una partición borrosa sobre Y , $C \in \mathcal{I}(Y)$. La distancia entre los elementos $x, y \in X$ se define como $d_{FCDM}: X \times X \rightarrow \mathbb{R}$, mediante la expresión:

$$d_{FCDM}(x, y) = \sqrt{\sum_{l=1}^c \left(\sum_{j=1}^N (A_j(x) * r_{A_j, C_l}) - \sum_{j=1}^N (A_j(y) * r_{A_j, C_l}) \right)^2} \quad (8)$$

donde la uninorma U es el parámetro que define esta familia.

1.3 ConFuCiuS en el contexto de los modelos perezosos

Las características de la función de distancia que utiliza ConFuCiuS compromete el enfoque puramente perezoso de la variante del k -NN propuesta, debido a que la función que se utiliza para comparar dos valores de un rasgo se basa en una función $I(\cdot)$ que generaliza la relación entre estos valores en el conjunto de instancias conocidas. Las diferencias entre ConFuCiuS y el k -NN estándar son esencialmente basadas en el criterio de proximidad que cada uno implementa, debido a que:

- Ambos almacenan las instancias como pares atributo-valor, pero ConFuCiuS adicionalmente requiere una modelación borrosa de los atributos numéricos.

- Al comparar dos instancias ConFuCiuS tiene en cuenta dos tipos de conocimiento: los valores que en éstas toma cada rasgo y el conocimiento generalizado del conjunto de instancias; mientras que el k -NN se basa en la HEOM que es una métrica superficial.
- Como consecuencia, de lo anterior, ConFuCiuS es más costoso computacionalmente que el k -NN al requerir adicionalmente generalizar información contenida en el conjunto de entrenamiento; la cual permanece invariable mientras no se modifique el conjunto de instancias de entrenamiento.
- ConFuCiuS utiliza un criterio uniforme para comparar a nivel de rasgo, que no requiere normalización, a diferencia del k -NN que utiliza una distancia heterogénea.
- ConFuCiuS garantiza medidas de proximidad graduales para aplicar la función de predicción.
- En la comparación de atributos simbólicos o en presencia de valores ausentes ConFuCiuS garantiza una mayor diferenciación entre las instancias porque:
 - o En la comparación con un valor ausente (?) la distancia HEOM retorna valor 1 (máxima distancia), mientras que la función FADM retorna un valor atendiendo al valor conocido en la solicitud y el conocimiento generalizado sobre este rasgo en el conjunto de instancias. Solamente en la variante que ambos valores sean desconocidos, se considera entonces un valor constante (contrario al anterior que asume valor cero y distancia mínima).
 - o Si el atributo a comparar es simbólico, la distancia HEOM retorna un valor binario (función overlap), mientras que el criterio de comparación local FADM en el que se basa ConFuCiuS permite una mayor spectrum de valores
- ConFuCiuS modela los atributos numéricos como variables lingüísticas, permitiendo una modelación del problema más cercana a la realidad al garantizar manejar la imprecisión causada por el uso de conceptos tales como “fiebre alta”, “precio elevado”, “rendimiento normal”, entre otros; los cuales se pueden utilizar tanto en la descripción de una instancia como al especificar una solicitud.
- ConFuCiuS utiliza un criterio de proximidad que depende del contexto y que se define automáticamente a partir del conjunto de entrenamiento; garantizando considerar un conocimiento más específico para un dominio de aplicación y al mismo tiempo disminuir la ingeniería del conocimiento requerida.
- Utilizar la modelación borrosa de los atributos numéricos permite manejar el grado de pertenencia del valor a cada uno de los términos lingüísticos definidos como una medida de similitud o compatibilidad con estos; propiciando ConFuCiuS soluciones efectivas y más naturales al problema que se resuelve.

ConFuCiuS, a diferencia del k -NN, no muestra signos de degradación ante rasgos irrelevantes; a partir de utilizar como criterio de proximidad ejemplos de distancias basadas en la extensión y generalización de VDM. Por otro lado, muestra un desempeño al menos comparable con los clasificadores existentes, y los mejores resultados cuando la muestra se describe por rasgos numéricos y simbólicos y en ella aparecen valores ausentes o rasgos irrelevantes.

2 Implementación computacional

Sobre la implementación de NeuroEvaluator

Esta herramienta [FAL05]¹² se presenta como una plataforma para desarrollar sistemas neuro-borrosos asociativos. La herramienta brinda facilidades para: la visualización y preprocesamiento de los datos, la configuración del modelo de RNA, así como un módulo para la evaluación de los resultados. Luego, permite la modelación de las variables lingüísticas, facilitando la obtención de la mejor variante de matriz de pesos a utilizar seguidamente por ConFuCiuS.

Para un conjunto de entrenamiento se automatiza la selección de la mejor configuración del modelo, generando las configuraciones posibles con los discretizadores y alternativas para definir las funciones de pertenencia antes mencionadas y evaluando con cada una de ellas el desempeño del modelo; de lo cual resulta la formulación de un problema de configuración, pues se requiere encontrar la configuración óptima variando los aspectos no predefinidos a priori por el ingeniero en conocimiento y donde la función a optimizar es el desempeño del modelo. De aquí que este software se conciba como una plataforma evaluadora de modelos neuro-borrosos.

En aras de ganar eficiencia computacional, la implementación de este módulo experimentador explota el procesamiento distribuido de tareas, donde en cada cliente se pone a evaluar una configuración posible. Adicionalmente este problema de configuración se resuelve utilizando el método de ascensión de colinas como estrategia de búsqueda heurística [FAL06]¹³. Dichas características diferencian esta herramienta de

¹² **Plataforma Evaluadora de Modelos Asociativos Neuroborrosos.** Memorias del XII Simposio de Ingeniería Eléctrica (SIE), 18-23 de junio 2005. ISBN 959-250-201-3. Rafael J. Falcón Martínez, Yanet Rodríguez Sarabia, María Matilde García Lorenzo, Rafael Bello Pérez, Alain Jose Varela Martín, Yasel Couce Sardiñas.

¹³ **Módulo de Experimentación Combinatoria para la Evaluación de Modelos Neuro-Borrosos Asociativos.** VII Conferencia Científica Internacional UNICA. III Computer Science Workshop CSW'06. Ciego de Avila, Cuba. Octubre, 2006. ISBN: 959-16-0473-4. Rafael J. Falcón Martínez, Aylin Labrador Morales, Angel Miguel Navarro Moya, Yanet Rodríguez Sarabia, María Matilde García Lorenzo.

otras existentes. Su factibilidad se constató en el desarrollo del sistema para el tratamiento de aguas residuales [ROD05b]¹⁴

Sobre la implementación de ConFuCiuS

La implementación de ConFuCiuS se facilitó al concebirse como una extensión a la herramienta de aprendizaje automatizado WEKA, de código abierto y orientada a la extensibilidad. Para ello se requiere la adición de un nuevo tipo de dato (LINGUISTIC) y los filtros correspondientes para manejar un atributo numérico como variable lingüística [MOR06]¹⁵. Este nuevo tipo de dato permite asociar a cada atributo numérico un conjunto de términos lingüísticos con las funciones de pertenencia correspondientes; cuyo procesamiento conllevó a modificaciones del intérprete del Weka por la introducción de un analizador descendente recursivo. Luego en la modelación de los rasgos, el ingeniero del conocimiento puede adoptar dos conductas: especificar para un atributo numérico en el fichero de definición de datos los conjuntos borrosos si ya se tiene una modelación a priori (por ejemplo, siguiendo el criterio de los expertos humanos) o que el propio Weka sea el encargado de hallarlos semi-automáticamente mediante un algoritmo de pre-procesamiento. El pre-procesamiento en Weka se implementa mediante filtros, lo cual adicionalmente tiene la ventaja de independizar el pre-procesamiento de los datos de la implementación de los algoritmos.

El k-NN es un clasificador de la familia de los algoritmos basados en instancias, cuya implementación en Weka ya aparece (IBk) formando parte del paquete “weka.classifiers.lazy”. Este tipo de clasificador requiere como parámetro de entrada la función de distancia a utilizar para calcular la similitud entre dos instancias, entre otros. Luego, la implementación de ConFuCiuS en Weka se redujo a implementar las nuevas funciones de comparación a nivel de rasgos que utiliza.

¹⁴ **Artificial Intelligence hybrid model applied to wastewater treatment.** Proceedings of the International e-Conference on Computer Science 2005 (IeCCS 2005). Lecture Series on Computer and Computational Sciences, 2. ISBN 90-6764-425-0. Yanet Rodríguez Sarabia, Xiomara Cabrera Bermúdez , Rafael Jesús Falcón Martínez, Zenaida Herrera Rodríguez, Ana M. Contreras Moya, Maria Matilde García Lorenzo.

¹⁵ **Una metodología para extender el ambiente de aprendizaje automatizado weka.** Informe de Investigación terminada, Universidad Central de Las Villas 2006. Carlos Morell, Yanet Rodríguez, Héctor Matías y Liana I. Araujo.

Además en [ROD06d]¹⁶ se resumen las ventajas de validar nuevos algoritmos de la IA utilizando la plataforma Weka, y de realizar la implementación del modelo en esta plataforma. Entre ellas se destaca que facilita la comparación de los nuevos algoritmos con otros modelos ya existentes e implementados en la herramienta, y propicia la visibilidad científica de los resultados en la comunidad de científicos de este campo. Con esta implementación de ConFuCiuS fueron desarrolladas las aplicaciones que se relacionan a continuación

3 Descripción de algunas aplicaciones

Sistema Automático Predictor de Riesgo Infantil de Mortalidad (SAPRIM¹⁷)

Esta aplicación se desarrolla en coordinación con especialistas del Hospital Pediátrico “José Luis Miranda” de Santa Clara. Esta nueva herramienta permite dar una primera evaluación del estado del paciente sirviéndole de guía al especialista (no experto necesariamente) para dar recomendaciones al ingreso del paciente, mientras se esperan por resultados del laboratorio para complementar su diagnóstico.

A partir de la muestra recopilada, y con la ayuda de los especialistas, se define una base de casos con 1079 ejemplos (538 “de cuidado” y 541 “grave-crítico”). Cada caso se caracteriza por 32 (3 de tipo numérico) rasgos, uno de los cuales se refiere al rasgo objetivo definiendo como sus valores posibles: “de cuidado” y “Grave-crítico”. Se utiliza la herramienta Weka para definir la base de casos; y las variables lingüísticas correspondientes a los rasgos “Edad”, “Tensión arteria; sistólica” y “Tensión arterial diastólica” se modelan por funciones Trapezoidales a partir del filtro “MDLFuzzifier”.

En [ROD05c]¹⁸ se presenta la solución computacional que se da inicialmente a este problema utilizando una red neuronal de tipo MLP como clasificador (eliminando previamente la ausencia de información), mostrando mejor desempeño con respecto a la variante FSIAC. Sin embargo, aplicando ConFuCiuS (3 casos similares para pronosticar la solución y FCDM como criterio de comparación a nivel de rasgo), se superan estos resultados al mismo tiempo que se garantiza justificar la solución encontrada.

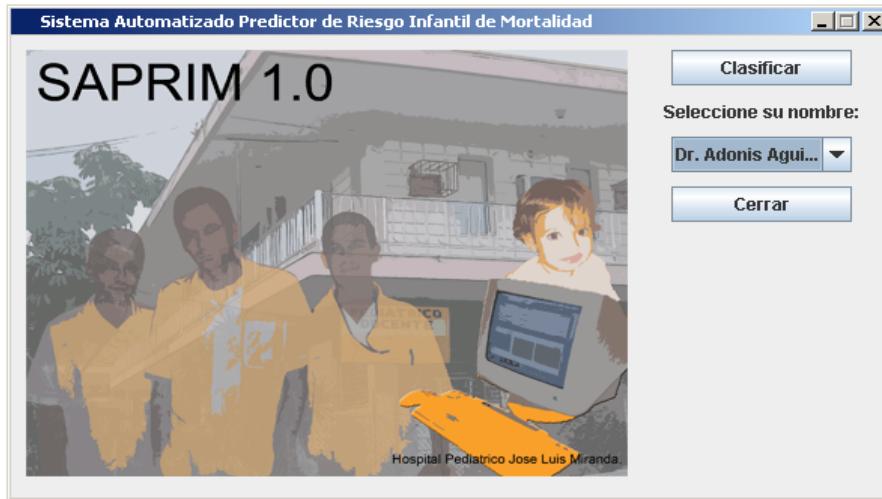
A continuación se muestran ventanas que se refieren a la aplicación de este sistema:

¹⁶ **Validación de nuevos algoritmos de la Inteligencia Artificial que utilizan conjuntos borrosos utilizando la herramienta para el aprendizaje automatizado WEKA.** VII Conferencia Científica Internacional UNICA. III Computer Science Workshop CSW’06. Ciego de Avila, Cuba. Octubre, 2006. ISBN: 959-16-0473-4. Yanet Rodríguez, Carlos Morell, Héctor Matías, Liana I. Araujo

¹⁷ Rodríguez, Y. y otros. SAPRIM: Sistema Automatizado Predictor de Riesgo Infantil de Mortalidad.

¹⁸ Rodríguez, Y. and e. al. (2005). Sistema Experto Híbrido para el Pronóstico del nivel de gravedad del paciente pediátrico grave Paciente pediátrico grave. COMPUMAT.

Esta es la ventana principal de SAPRIM, como el sistema está en etapa de prueba se le da al especialista la posibilidad de identificarse antes de interactuar con el sistema, ya que luego se le va a pedir su propia opinión con respecto al caso que se está analizando.



Luego se elige clasificar.

La anterior ventana permite captar todos los rasgos clínicos y epidemiológicos del paciente. Para valorizar las opciones que aparecen en gris, que son los rasgos con valores múltiples, se elige primero la opción Evaluado.

Ahora se introduce como historia clínica 345678, se elige modo de ingreso directo, y se selecciona que no es un reingreso. La edad es de 15 meses, el estado nutricional y de hidratación son normales y en la temperatura se escoge Hipertermia. El color de la piel es normal, el estado de salud previo, sano y el diagnóstico es adenoiditis.

Nuevo Lazo

Datos Generales | Signos respiratorios | Signos Vitales | Signos circulatorios | Signos Neurológicos

Datos

Historia Clínica: 345678

Modo Ingreso: Directo

Reingreso: ☐ Si ☒ No

Edad y otros

Edad (Meses): 15

Estado nutricional: Normal

Temperatura: Hipertermia

Estado de hidratación: Normal

Color de la piel

☒ Evaluado

- ☒ Normal
- ☐ Palida
- ☐ Rubrecundez
- ☐ Cianosis
- ☐ Terrosa
- ☐ Ictero
- ☐ Livedo-reticular

Estado de salud previo

☒ Evaluado

- ☒ Sano
- ☐ Acidosis_tubular_renal
- ☐ Adenoiditis_crónica
- ☐ Alergia
- ☐ Anemia_hemolítica
- ☐ Aplasia_medular
- ☐ Asma_bronquial
- ☐ Cardiopatía

Diagnostico

☒ Evaluado

- ☒ Adenoiditis
- ☐ Ahogamiento_incompleto
- ☐ Anemia
- ☐ Apendicitis_aguda
- ☐ Arritmia
- ☐ Artritis
- ☐ Broncoaspiracion

Clasificar

En la sección de Signos Respiratorios la frecuencia respiratoria y el murmullo vesicular se seleccionan como normales. En estridor, quejido y aleteo se elige "no" al igual que en tiraje y estertores.

Nuevo Caso

Datos Generales Signos respiratorios Signos Vitales Signos circulatorios Signos Neurológicos

Frecuencia respiratoria: Normal

Murmullo vesicular: Normal

Tiraje

☒ Evaluado

☒ No

☐ Subcostal

☐ Intercostal

☐ Supraclavícula

☐ Supraesternal

☐ Generalizado

Estridor inspiratorio

☐ Si ☒ No

Quejido Espiratorio

☐ Si ☒ No

Aleteo nasal

☐ Si ☒ No

Estertores

☒ Evaluado

☒ No

☐ Roncos

☐ Silbantes

☐ Crepitantes

☐ Subcrepitantes

☐ Ruidos

Clasificar

La TAS es de 110 y la TAD de 70. La permisión tisular buena y el pulso y la frecuencia cardiaca normales.

Nuevo Caso

Datos Generales Signos respiratorios Signos Vitales Signos circulatorios Signos Neurológicos

Tension Arterial

TAS: 110

TAD: 70

Perfusion tisular: Buena

Pulso: Normal

Frecuencia Cardiaca: Normal

Clasificar

Nuevo Caso

Datos Generales Signos respiratorios Signos Vitales **Signos circulatorios** Signos Neurológicos

Hematurias

☐ Si ☒ No

Sangramientos

☐ Si ☒ No

Purpuras

☒ Evaluado

☒ No

☐ Petequias

☐ Esquimosis

☐ Hemetomas

☐ Vasculitis

Diuresis Normal

En la sección Signos circulatorios el paciente no presenta hematurias ni sangramiento, la diuresis es normal y no tiene púrpuras.

Clasificar

En la sección Signos Neurológicos el sensorio se encuentra despejado, las pupilas isocóricas reactivas, el tono muscular normal y la fontanela normotensa. No presenta convulsiones ni defecto motor.

Nuevo Caso

Datos Generales Signos respiratorios Signos Vitales Signos circulatorios **Signos Neurológicos**

Sensorio Despejado

Pupilas Isocóricas reactivas

Tono Muscular Normal

Fontanela Normotensa

Convulsiones No

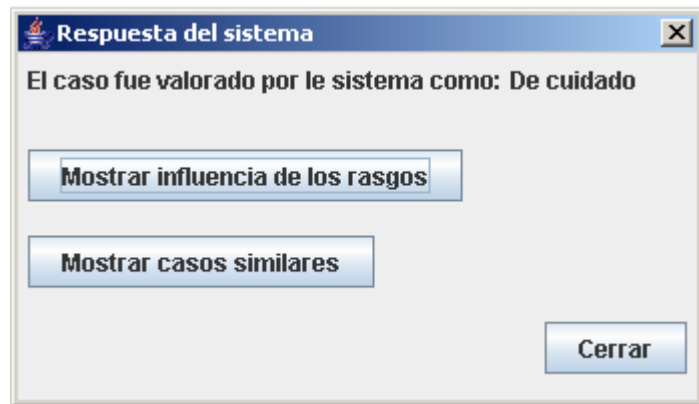
Defecto Motor

☐ Si ☒ No

Luego se elige Clasificar.

Clasificar

Si se selecciona Mostrar Casos Similares se visualizan los casos más semejantes al caso que se está analizando



A screenshot of a software window titled "Respuesta del sistema". It contains the text "El caso fue valorado por le sistema como: De cuidado". Below this text are two buttons: "Mostrar influencia de los rasgos" and "Mostrar casos similares". At the bottom right is a "Cerrar" button.

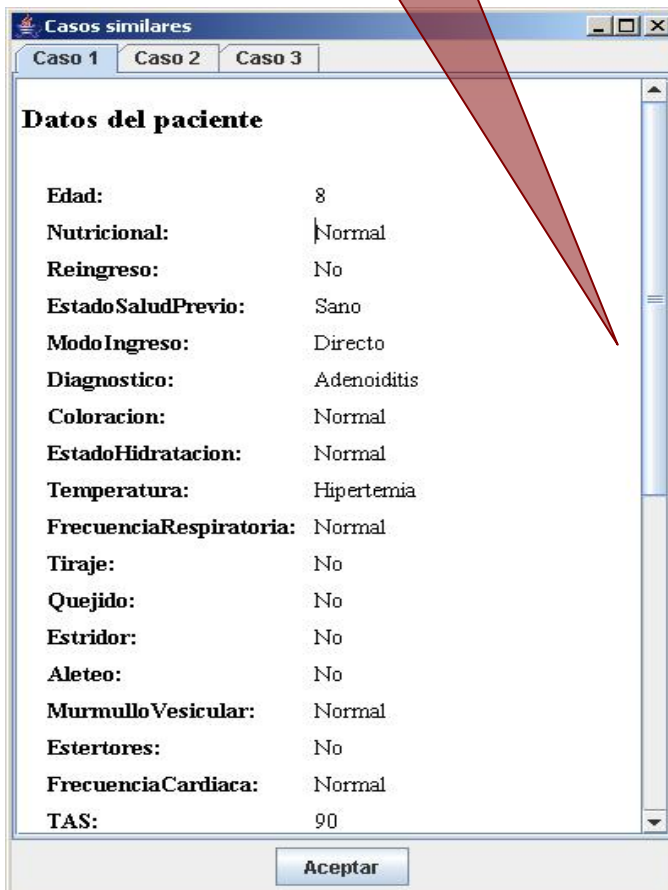
Respuesta del sistema

El caso fue valorado por le sistema como: De cuidado

Mostrar influencia de los rasgos

Mostrar casos similares

Cerrar



A screenshot of a software window titled "Casos similares". It has three tabs: "Caso 1", "Caso 2", and "Caso 3". The "Caso 1" tab is selected. Below the tabs is a section titled "Datos del paciente" containing a list of patient data. At the bottom is an "Aceptar" button.

Casos similares

Caso 1 Caso 2 Caso 3

Datos del paciente

Edad:	8
Nutricional:	Normal
Reingreso:	No
EstadoSaludPrevio:	Sano
Modo Ingreso:	Directo
Diagnostico:	Adenoiditis
Coloracion:	Normal
EstadoHidratacion:	Normal
Temperatura:	Hipertemia
FrecuenciaRespiratoria:	Normal
Tiraje:	No
Quejido:	No
Estridor:	No
Aleteo:	No
Murmullo Vesicular:	Normal
Estertores:	No
FrecuenciaCardiaca:	Normal
TAS:	90

Aceptar

Influencia de los rasgos en el pronóstico obtenido

Influencia de los rasgos

EstadoSaludPrevio:	0.25326702830025144
Nutricional:	0.2450103086969927
Reingreso:	0.23757523563818636
Tiraje:	0.2097228639372366
Pupilas:	0.1493788118966967
DefectoMotor:	0.12080490044973696
FrecuenciaRespiratoria:	-0.10942200982581313
MurmulloVesicular:	0.10229912071426164
Aleteo:	0.0837685537319943
Convulsiones:	-0.06102781976973618
Perfusión:	0.05322331929053619
Diuresis:	0.0532233192905359

Aceptar

Si se selecciona **Mostrar influencia de los rasgos** aparece en que medida influyo cada rasgo en la respuesta que dio el sistema. Influencias positivas favorecen la respuesta

NephroSystem: Sistema Automatizado para el servicio de Nefrología

Esta aplicación se desarrolla en coordinación con especialistas del Hospital Universitario "Arnaldo Milián Castro" de Santa Clara. Esta tiene como propósito brindar ayuda al personal especializado en la toma de decisiones, específicamente a la hora de determinar la esperanza de vida de un paciente y si el mismo es transplantable. Para ello es necesaria la entrada de datos referentes al estado clínico del paciente. Debido a que este software será manipulado por personal médico y no por especialistas de la computación, nos propusimos que todas las operaciones que se realizaran fueran de forma sencilla, cómoda y manejable para todo tipo de usuario, sin la necesidad de poseer conocimientos computacionales profundos.

Para la aplicación se utilizó como lenguaje de programación Java y se uso como plataforma de desarrollo el NetBeans 5.0. A continuación se muestran las principales facilidades y componentes del sistema.



Ventana de presentacion

Panel esperanza de vida

En el panel “Esperanza de vida” que se muestra en la figura siguiente se observan los componentes que sirven para construir el nuevo caso que será clasificado.

nephroSystem

Esperanza de vida **Transplante Renal**

Edad Estado Nutricional **Bien Nutrido** ▼

Etiologia






- Diabetes mellitus
- Nefritis intersticial
- Nefropatia diabetica
- Destransplante
- Nefropatia vascular
- Nefroangioesclerosis

Patologias asociadas

- Fallo cardiaco congestivo
- Diabetes mellitus
- Enf. vascular periferica
- Enf. vascular
- HTA severa
- Enf. obstructiva cronica
- Enf. isquemica
- IMA

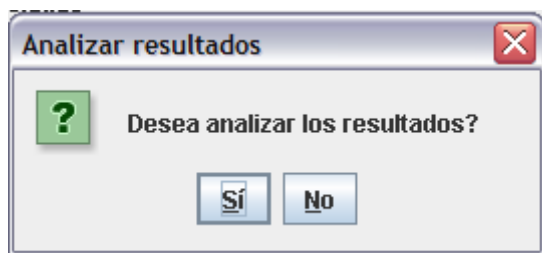
Para posibilitar lo descrito anteriormente hay un conjunto de acciones que se deben ejecutar sobre distintos componentes y que se explicarán a continuación.

El Spinner “Edad” se acciona mediante los botones por defecto de este componente aunque también es posible introducir el dato directamente. El cuadro combinado “estado nutricional” se acciona solamente presionando el botón por defecto del componente y seleccionando la opción deseada. El cuadro de selección “preparado” se marca o desmarca según sean los datos conocidos. El botón >> se acciona cuando es necesario pasar un elemento de la lista “etiología” a la lista destino, vale aclarar que se puede seleccionar uno o varios elementos de la lista “etiología” (selección múltiple con intervalos continuos o

discontinuos), en el caso de los intervalos continuos es manteniendo la tecla Shift (derecha o izquierda) presionada, realizar clic en el primer y el último elemento del intervalo y en el caso discontinuo la tecla ctrl. Y realizar un clic sobre los elementos deseados. El botón  es usado para eliminar de la lista etiología (destino) uno o varios elementos y devolverlos a la lista etiología. Así mismo  se acciona cuando es necesario pasar un elemento de la lista “patologías asociadas” a la lista destino, y  es usado para eliminar de la lista “patologías asociadas (destino)” uno o varios elementos y devolverlos a la lista “patologías asociadas”, con las mismas especificaciones anteriormente descritas para  y . También se pueden enviar los elementos de la lista “etiología” a la lista “etiología (destino)” y de la lista “patología asociada” a la lista “patologías asociadas (destino)” y viceversa realizando doble clic sobre cada elemento (esto solo permite efectuar selección simple).

Una vez entrados todos los datos a recopilar del paciente se acciona el botón “Calcular”, e inmediatamente se muestra en el cuadro de edición los resultados, en el botón “calcular” se realiza un chequeo de la validez de los datos y en el caso de que se hayan dejado datos vacíos se muestra la ventana de la figura 3.3 donde se tiene una sola opción:

Aceptar: se pasa nuevamente al panel activo para entrar los datos restantes.



La ventana anterior brinda la posibilidad de visualizar los casos similares al analizado, muestra dos botones de acción. Al seleccionar “No” se cierra esta ventana y al seleccionar “Sí” se accede a la ventana que se observa en la figura 3.5, en esta ventana se da la posibilidad de entrar la cantidad de casos similares a mostrar (una cantidad menor que 10, por defecto se muestran tres) una vez entrada la cantidad de casos se presiona el botón procesar y se visualizan los casos. Después de haber accedido a esta ventana solo se tendrá la opción de ver los datos mostrados o cerrar la ventana mediante el botón “Cerrar”, una vez se cierra la ventana se vuelve al panel “Esperanza de vida” donde se muestra en el botón “Introducir nuevo caso” para analizar otro paciente.

Analisis de resultados

Casos a consultar

3

+

-

Procesar

Rasgos	Caso No. 1	Caso No. 2	Caso No. 3
Edad								
CI Diabetes mellitus								
CI Nefritis Intersticial								
CI Nefropatía diabética								
CI Destransplante								
CI Nefropatia Vascular								
CI ENF. R. Poliq. AD								
CI NVHipertensiva								
CI Nefroangioesclerosis								
CI Valva Uretra Posterior								
CI GLOM. S y Vasculitis								
CI No Filiada								
CI ENF. Quísticas								
CI Nefropatia Isquemica								
CI HTA								
CI Glom. Primaria								

La otra funcionalidad principal de esta aplicación se accede a través del **panel “transplante”** que se muestra en la figura siguiente. Para introducir los datos de un paciente para analizar si es transplantable hay un conjunto de acciones que se deben ejecutar sobre distintos componentes y que se explicarán a continuación.

nephroSystem

Esperanza de vida **Transplante Renal**

Edad Sexo ☐ Foco septico

Estudios imaginologicos

☐ Ultrasonido abdominal y renal ☐ Cistografia miccional
☐ Alta estomatologica ☐ Ecocardiograma
☐ Rx de senos perinasales ☐ Ultrasonido renal
☐ Rx de estomago y duodeno ☐ TUS

☐ Antigeno prostatico ☐ Prueba citologica

Estudios virologicos

☐ Hepatitis C
☐ Hepatitis B
☐ VIH
☐ Serologia
☐ Citomegalovirus


Estudios inmunologicos

☐ Cross match
☐ HLA

Creatinina mg% ☐ Cuagulograma

T.G.P. U/L Peso Kg

Panel Transplante renal

Causas por las que el paciente no esta apto para el tra... 

Resultados elevados


TGP : 78.0

Resultados de otras pruebas

Cross match : Presenta problemas

HLA : Presenta problemas

Prueba Citologica : Presenta problemas



Cerrar