

# Clasificación binaria de neuropatía diabética implementando el operador *PS-Merge*

América Velasco-Cétera<sup>1</sup>, Eddy Sánchez-DelaCruz<sup>1</sup>,  
Pilar Pozos-Parra<sup>2</sup>

<sup>1</sup>Tecnológico Nacional de México,  
Departamento de Posgrado e Investigación,  
Mexico

<sup>2</sup>Universidad Autónoma de Baja California,  
México

{amecetera, eddsacx}@gmail.com, maria.pozos@uabc.edu.mx

**Resumen.** En el campo de la inteligencia artificial, algunos algoritmos se diseñan para que, a partir de datos, se desempeñen métodos predictivos o descriptivos con la finalidad de ayudar a los expertos en la toma de decisiones. Este tipo de método de la inteligencia artificial es conocido como aprendizaje automático, existen métodos clásicos, los cuales son utilizados en la estadística con procesos complejos y sofisticados, principalmente para la clasificación y la regresión como: los algoritmos bayesianos, la regresión, la agrupación, los árboles de decisión y las redes neuronales, siendo esta última la más utilizada, para su implementación se requiere de un experto que programe los algoritmos considerando que se necesita ajustar gran variedad de parámetros para que un sistema aprenda correctamente. En esta investigación se realizó una clasificación binaria utilizando *PS-Merge*, un operador empleado para la fusión de creencias (conocimiento) y poco empleado en el aprendizaje automático. Los resultados mostraron una precisión superior al 80 % de instancias correctamente clasificadas de casos de neuropatía diabética, empleando tres nociones basadas en lógica proposicional: fusión de creencias, revisión de creencias y la noción de satisfacción parcial, lo que permitió combinar información resultante de diferentes fuentes contradictorias, durante el proceso no se requiere ajustar parámetros, además, se puede trazar las causas de los resultados, logrando así un aprendizaje automático explicable.

**Palabras clave:** Aprendizaje automático, clasificación, *PS-Merge*.

## Binary Classification of Diabetic Neuropathy implementing the *PS-Merge* operator

**Abstract.** In the field of artificial intelligence, some algorithms are designed so that, from data, predictive or descriptive methods are performed

in order to help experts in decision making. This type of method of artificial intelligence is known as automatic learning, there are classic methods, which are used in statistics with complex and sophisticated processes, mainly for classification and regression as: Bayesian algorithms, regression, grouping, decision trees and neural networks, the latter being the most used, for its implementation requires an expert to program the algorithms considering that you need to adjust a variety of parameters for a system to learn correctly. In this research a binary classification was made using PS-Merge, an operator used for the fusion of beliefs (knowledge) and little used in automatic learning. The results showed an accuracy higher than 80 % of correctly classified instances of diabetic neuropathy cases, using three notions based on propositional logic: belief fusion, belief review and the notion of partial satisfaction, which allowed combining information resulting from different contradictory sources, during the process it is not required to adjust parameters, in addition, it is possible to trace the causes of the results, thus achieving an explainable automatic learning.

**Keywords:** Machine learning, classification, *PS-Merge*.

## 1. Introducción

Los trastornos que afectan al sistema nervioso se llaman neuropatías y generalmente afectan a las extremidades inferiores (piernas y pies), pero varían desde dolor y entumecimiento en los pies hasta problemas con la función de los órganos internos, como el corazón y la vejiga. Las personas diabéticas que no cuidan adecuadamente su control glucémico, corren el riesgo de desarrollar una neuropatía diabética, que es el resultado de la diabetes y afecta la forma en que las personas caminan, que conduce a una marcha alterada y a un deterioro progresivo [17]. El número de diabéticos en el mundo se estima en 285 millones, y se espera que alcance los 438 millones para 2030 [22].

Según las estimaciones de la Organización Mundial de la Salud, 422 millones de adultos en todo el mundo tenían diabetes en 2014, en comparación con 108 millones en 1980. En la población adulta, esta proporción aumentó de 4.7 % a 8.5 % [4]. La neuropatía diabética es el tipo más común de complicaciones microvasculares diabéticas, siendo una de las causas más común de amputación, falla autonómica, morbilidad y mortalidad [13].

Por otro lado, la Inteligencia Artificial (IA) constituye uno de los campos interdisciplinarios y transfronterizos donde convergen muchas ciencias. Se han introducido programas basados en IA que ayudan a los médicos a tomar decisiones [14], la mayoría basados en aprendizaje automático. Para procesar datos, construir e investigar sistemas que pueden aprender de los conjuntos de datos de entrenamiento y mejorar el proceso de clasificación y predicción [8]. Esta subdisciplina de la IA cada vez está más presente en tecnologías del entorno cotidiano, mediante técnicas que permiten reducir la intervención humana en la toma de decisiones.

Dado que la clasificación se usa más en el aprendizaje automático [12], se han propuesto diferentes algoritmos, tales como: funciones lineales, árboles de decisión, redes neuronales, clasificadores bayesianos, por mencionar algunos. Sin embargo, estos enfoques clásicos involucran funciones que los expertos deben calibrar manualmente, por ejemplo, el concertar gran variedad de parámetros y pesos, los cuales deben estar bien configurados para hacer que funcione correctamente el sistema.

Ahora, en lógica, se emplean técnicas de fusión que utilizan conocimiento, creencias o metas de un grupo de sujetos, así como creencias o metas de cada miembro del grupo. La *fusión de creencias* es una técnica basada en la lógica proposicional, que permite combinar información potencialmente conflictiva de diferentes fuentes definidas en un conjunto de bases de creencias inconsistentes, para generar una base de conocimiento única y consistente [6]. En cambio, la *revisión de creencias* se basa en el principio de separar el conocimiento de la información de diferentes fuentes para cambiar el estado del conocimiento en un marco lógico [19], o como una agregación de nuevo conocimiento en una base de conocimientos (KB) ya establecida [10], al cambiar el conocimiento original lo menos posible y manteniendo la consistencia de la KB. La *satisfacción parcial* es otra forma de medir la satisfacción de una fórmula.

Este concepto permite tener un valor de satisfacción en  $[0,1]$ . La explicación intuitiva para la satisfacción parcial indica que: es natural pensar que, si tiene una combinación de dos literales y sólo una está satisfecha, el agente estará satisfecho al 50%. Al resumir esta idea, la satisfacción de la conexión de uno o más literales se puede medir en función de la suma de la interpretación de conjunciones y la suma de literales [1]. Este trabajo se inspira en [9], donde se usó una variante del operador PS-Merge para diagnosticar el cáncer oral. Los resultados muestran una precisión de más del 75%, sin tener que ajustar parámetros en el conjunto de datos, determinando que el operador es adecuado para aplicaciones y tecnologías de aprendizaje automático en conjuntos de datos médicos.

Actualmente se carece de implementaciones con operadores que no utilicen la estadística y optimización sin la necesidad de hacer configuraciones en parámetros o iteraciones. En esta investigación, se realizó una implementación del PS-Merge [1], un operador de fusión relativamente nuevo y poco conocido, basado en la noción de satisfacción parcial, que combina la fusión de creencias y la revisión de creencias, aplicado a un conjunto de datos de la marcha de pacientes que padecen neuropatía diabética, para una clasificación binaria {sano, enfermo}. En la literatura las técnicas de *fusión de creencias*, *revisión de creencias* y la *noción de satisfacción parcial* se han utilizado por separado, éste es el primer intento utilizando PS-Merge en casos de neuropatía diabética, por lo que los resultados de esta investigación se plantean originales en el área.

La distribución del documento es la siguiente: la sección 2 presenta trabajos relacionados, en la sección 3 se presentan los materiales y métodos, la sección 4 presenta los resultados y la discusión. Por último, la sección 5 presenta conclusiones y trabajos futuros.

## 2. Trabajos relacionados

El uso del aprendizaje automático basado en algoritmos de IA para sistemas de diagnóstico y análisis ha guiado importantes mejoras en la predicción de enfermedades, en donde los sistemas pueden identificar si un paciente sufre una enfermedad de acuerdo a condiciones presentes, al proporcionar suficientes datos no sólo se puede detectar una enfermedad con la misma precisión que un patólogo humano sino que se puede anticipar a detectar en el paciente signos que puedan conducir a una enfermedad. No obstante, las tres nociones que utiliza *PS-Merge*, mantienen avances significativos como:

En [18], formalizaron el concepto de juego de revisión de creencias (BRG) para modelar la dinámica de las creencias de un grupo de agentes. Señalando un conjunto de propiedades que abordan varios temas de preservación, así como la capacidad de respuesta, la monotonicidad y la convergencia. Agregaron varias clases de políticas de revisión que se basan en la fusión de creencias de los operadores, considerando un caso en que todos los agentes utilizan la misma política de revisión. Desarrollaron un software que consiste en una interfaz gráfica que permite jugar con los BRGs considerando cualquiera de las 18 políticas de revisión de  $\{ R_{\Delta}^k \mid k \in \{ 1, \dots, 6 \}, \Delta \in \{ \Delta^{d_D, \Sigma}, \Delta^{d_H, \Sigma}, \Delta^{d_H, Gmin} \} \}$ .

En [5] presentaron una solución genérica al problema de la revisión de los marcos de argumentación que se aplica a muchas argumentaciones prominentes de I-máximo semántica, con el fin de probar un teorema de representación para dos tipos de revisión. Combinando avances de argumentación y el cambio de creencias, es decir, tomaron conceptos de realizabilidad en la argumentación y la noción de cumplimiento como se utiliza en la revisión de Horn, los resultados son genéricos en el sentido que tienen una amplia gama de semántica.

En [7], fusionaron  $\delta$ -ontologías usando la revisión de creencias y lógica inviable programación, extendiendo el marco de  $\delta$ -ontologías al combinar la revisión de creencias, ontologías lógicas de argumentación y descripción, de dos ontologías tales que la unión de éstas podría ser inconsistente. Mientras que, en [3] analizaron el marco de los fragmentos de lógica propositiva considerando que el resultado del cambio de operación esté en el mismo fragmento que el cambio de creencia, así como, mantener el comportamiento del operador original sin cambios en caso de que entregue un resultado que ya encaja en el fragmento. Como resultado obtuvieron un enfoque que no se limita a un solo fragmento de lógica propositiva, sino que es aplicable a tres fragmentos de lógica propositiva.

En [2], describen un método de revisión que considera las bases de creencias "planas" sin información extra, adaptando la fusión de creencias bajo restricciones para llevar a cabo la revisión de las mismas, al comparar el  $\Delta_{ps_{\mu}}$  contra dos operadores de revisión de creencias: Dalal y BHQ, los resultados mostraron el potencial del enfoque de los *PS*, en donde las operaciones eran más simples así como más apropiadas en algunos escenarios y desarrollaron un prototipo de software de revisión de creencias.

En [21], ampliaron la noción de satisfacción, para incluir una "medida" de satisfacción y resolver problemas de la Primera Competencia Internacional de Horarios, donde fue necesario encontrar algunas heurísticas que permitieron reducir

el espacio de memoria. Integraron esta expansión en una fase de preprocesamiento que definió estados mentales consistentes, deseos (Belief-Desires-Intentions) de los agentes. El resultado fueron los estados que maximizan la Suma de la Satisfactibilidad Parcial del perfil, sin ser necesario definir un preordenamiento parcial.

En [17], se usó un conjunto de datos de una red de sensores, para lograr biomarcadores de marcha para pacientes con neuropatía diabética e individuos sanos, crearon un dataset que incluyó características de personas sanas o al contrario, así como el estado de salud conocido de cada caso, pero sin etiquetar. Los datos de prueba confirmaron la eficiencia de los modelos tras la aplicación de una búsqueda exhaustiva que combinó diversos algoritmos (*clasificadores ensamblados+aprendizaje profundo*) y la selección del que tuvo el máximo porcentaje de instancias correctamente clasificadas, en donde las instancias mostraron un alto grado de certeza con la existencia de atrofia en los músculos que conducen a una marcha anormal debido a la neuropatía diabética, logrando una precisión del 86.46 %.

### 3. Materiales y métodos

#### 3.1. Conjunto de datos de neuropatía diabética

El conjunto de datos se utilizó en [17], en donde los autores realizaron una clasificación mediante la combinación de algoritmos ensamblados y aprendizaje profundo. Los datos se obtuvieron del departamento de servicios médicos de la Universidad Autónoma de Tabasco. Los registros se tomaron de la marcha de pacientes con neuropatía diabética y sanos, utilizando cinco sensores acelerómetros ADXL-335 en tres ejes (x, y, z), conectados a una tarjeta Arduino MEGA.2560; se colocó un sensor en la espalda baja, otro en la rodilla derecha, otro en la rodilla izquierda, otro en el tobillo derecho, y el último en el tobillo izquierdo del paciente. Asimismo, la tarjeta Arduino MEGA se configuró para el registro de 400 puntos en 5 minutos, con un total de diez pacientes con neuropatía y cinco pacientes sanos, y la lectura interna de las posiciones contiene 400 registros por paciente.

Los nombres y representaciones de las tres posiciones que ocupó el sensor son los siguientes: espalda baja (col-x, col-y, col-z), rodilla derecha (rodDer-x, rodDer-y, rodDer-z), rodilla izquierda (rodIzq-x, rodIzq-y, rodIzq-z), tobillo derecho (tobDer-x, tobDer-y, tobDer-z), y tobillo izquierdo (tobIzq-x, tobIzq-y, tobIzq-z). Para determinar si el registro pertenece a pacientes con neuropatía diabética, se agregó un valor positivo (TRUE) y falso (FALSE) de pacientes sanos, lo que generó un tipo de conjunto binario (enfermo, sano) con un total de 16 atributos. La Tabla 1 muestra las características de los pacientes: género, edad, peso, altura, años de sufrimiento y etiología. La Tabla 2 muestra un segmento del conjunto de datos, que representa las tres posiciones de cada sensor ocupado (col-z, col-y, col-x, rodDer-z, rodDer-y, rodDer-x, ...), y la columna 16 indica si el paciente tiene la enfermedad o es sano, el dataset contiene un total de 6000 registros y 16 columnas.

**Tabla 1.** Información de pacientes.

Género	Edad (años)	Peso (Kg)	Estatura (m)	Tiempo de padecimiento (años)	Causa
Masculino	54	89	1.70	5	Hereditario
Masculino	60	108	1.65	10	Alimentación
Femenino	56	99	1.60	4	Hereditario
Masculino	56	81.5	1.62	6	Hereditario
Masculino	62	73	1.57	15	Alimentación
Femenino	50	70	1.59	8	Hereditario
Masculino	58	102	1.61	6	Alimentación
Masculino	57	87.7	1.58	8	Alimentación
Femenino	61	90	1.65	3	Hereditario
Masculino	50	83.2	1.63	5	Hereditario
Femenino	35	72	1.61	0	Sano
Masculino	38	82	1.65	0	Sano
Masculino	45	95	1.67	0	Sano
Masculino	40	75	1.59	0	Sano
Femenino	29	59	1.55	0	Sano

**Tabla 2.** Dataset de neuropatía diabética (fragmento).

col-z	col-y	col-x	rodIzq-z	rodIzq-y	... caso
1.65	1.74	1.99	1.60	1.69	... NeuroDiabe
1.65	1.75	2.01	1.61	1.70	... NeuroDiabe
1.66	1.75	2.01	1.61	1.70	... NeuroDiabe
1.67	1.86	1.97	1.64	1.74	... Sano

En el método, los datos se distribuyen de la siguiente manera:  $l_1$  representa el sensor en la posición z (col-z),  $l_2$  en la posición y (col-y),  $l_3$  en la posición x (col-x),  $l_4$  en la posición z (rodIzq-z),  $l_5$  en la posición y (rodIzq-y),  $l_6$  en la posición x (rodIzq-x),  $l_7$  en la posición z (rodDer-z),  $l_8$  en la posición y (rodDer-y),  $l_9$  en la posición x (rodDer-x),  $l_{10}$  en la posición z (tobIzq-z),  $l_{11}$  en la posición y (tobIzq-y),  $l_{12}$  en la posición x (tobIzq-x),  $l_{13}$  en la posición z (tobDer-z),  $l_{14}$  en la posición y (tobDer-y),  $l_{15}$  en la posición x (tobDer-x), y  $l_{16}$  el caso del diagnóstico del paciente {enfermo o sano}.

### 3.2. Binarización del conjunto de datos

La binarización incluye la reducción de información, donde el único valor posible es verdadero (1) o falso (0) [20]. Se han propuesto diferentes métodos de binarización, pueden reducirse seleccionando un umbral, un promedio y otras operaciones clásicas para obtener resultados satisfactorios [15]. El método más simple de binarización es la determinación de umbrales, también llamado umbralización, y puede ser de manera global, local o adaptativa [11,16].

Considerando que el operador PS-Merge trabaja con datos de entrada y salida en formato binario, se binarizaron todas las variables del conjunto de datos, que consisten en 16 columnas y 6000 registros (como se describe en la Sección 3.1). Se aplicaron 8 algoritmos de agrupación robustos (*Spectral Clustering*, *Affinity Propagation Clustering*, *MiniBatchKMeans*, *MeanShift*, *Agglomerative Clustering*, *DBSCAN*, *OPTICS*, *KMeans*) para binarizar, e identificar cuál es la mejor separación en dígitos binarios. Así como tres métodos: primero, al tomar el valor de umbral. En seguida, al tomar el promedio global de cada columna. Finalmente, al tomar el valor promedio de 4 algoritmos de agrupación más adecuados para las dos categorías, es decir, de los 8 algoritmos de agrupación, cuatro pueden separar mejor los datos entre clases.

### 3.3. Operador PS-Merge

Antes de definir este paradigma proposicional, es necesario describir los conceptos utilizados.

Sea  $\mathcal{L}$  un lenguaje proposicional que consiste en: un conjunto de  $n$  variables proposicionales  $P = \{p_1, p_2, p_3, \dots, p_n\}$ , y conectores lógicos ( $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ ) para formar proposiciones compuestas. Hay dos formas normales, las disyuntivas (FND) y las conjuntivas (FNC). En el paradigma disyuntivo son disyunciones de conjunciones de literales, y el paradigma conjuntivo son conjunciones de disyunciones de literales, donde un literal ( $l$ ) es una variable proposicional o su negación ( $p, \neg p, q, \neg q, \dots$ ).

Una cláusula es una disyunción de literales finita de la forma  $l_1 \vee \dots \vee l_n$ , donde cada  $l_i$  con  $i = (1, \dots, n)$  es un literal.

Un término es una conjunción de literales finita de la forma  $l_1 \wedge \dots \wedge l_n$ , donde cada  $l_i$  con  $i = (1, \dots, n)$  es un literal.

Una fórmula  $v$ , está en forma normal disyuntiva (FND) si es de la forma  $v_1 \vee v_2 \vee \dots \vee v_n$  donde cada  $v_i$  con  $i = 1, \dots, n$  es un término.

Una fórmula  $\psi$ , está en forma normal conjuntiva (FNC) si es de la forma  $\psi_1 \wedge \psi_2 \wedge \dots \wedge \psi_n$  donde cada  $\psi_i$  con  $i = 1, \dots, n$  es una cláusula.

Si  $K$  es una proposición representando una base de conocimientos o creencias, entonces  $Q_K$  representa la base de conocimiento equivalente en su FND.

Una interpretación, estado, o mundo es una función  $\omega$  de  $P$  a 0, 1, estos dos valores son identificados con los valores clásicos de Verdadero (V) y Falso (F) respectivamente.

El conjunto de todas las posibles interpretaciones se denotará como  $\mathcal{W}$  y sus elementos se denotarán como vectores de la forma  $(\omega(p_1), \dots, \omega(p_n))$ . La cardinalidad del conjunto  $\mathcal{W}$  es  $|\mathcal{W}| = 2^n$ .

Si  $\omega(v) = 1$  para una determinada interpretación  $\omega$ , una vez que  $\omega$  se extiende a la conjunción y el resto de los conectores de manera convencional, la fórmula  $v$  se denomina satisfecha o consistente. En este caso,  $\omega$  es un modelo de  $v$ .

Si tanto  $l$  como  $\neg l$  están en un *multiset*, entonces contienen pares  $\langle l, \neg l \rangle$  inconsistentes.

Cuando cualquier interpretación que hace que la premisa sea verdadera (operando izquierdo) también hace que la conclusión sea verdadera (operando dere-

cho), se produce una relación semántica ( $\models$ ) de resultados. Cuando cada modelo de la premisa es también el modelo de conclusión. La conclusión es el resultado lógico de la premisa.

El método empleado utiliza la *fusión de creencias*, *revisión de creencias* y *noción de satisfacción parcial*. Para la fusión de creencias se emplea *PS-Merge*, el que considera bases de creencias inconsistentes, extrayendo información que no cause inconsistencias de las bases de creencias y sólo examina el mejor valor de los casos, está basado en la satisfacción parcial, la cual se define directamente desde la distancia del mundo  $w$  hasta la base  $K$  sin recurrir a la distancia entre los mundos. Para calcular la distancia del mundo a un perfil se emplea la suma de la *satisfacción parcial* de cada base de creencias. En [1] se define el operador como se muestra a continuación.

**Definición 1.** La *satisfacción parcial* se define a partir de una base de conocimiento normalizada en formato FND:  $Q_K$ ,  $\omega \in \mathcal{W}$  una interpretación cualquiera y  $|P|=n$ . La satisfacción parcial de  $Q_k$  para  $\omega$  denotada por  $\omega_{ps}(Q_k)$ , se define como:

1. Si  $Q_K$  es una conjunción  $C_1 \wedge \dots \wedge C_s$  donde cada  $C_i$  es una literal, entonces:

$$\omega_{ps}(Q_K) = \max \left\{ \sum_{i=1}^s \omega_{ps} \left( \frac{\omega(C_i)}{s} \right), \left( \frac{n - |P(\bigwedge_{i=1}^s C_i)|}{2n} \right) \right\}, \quad (1)$$

donde:  $n$ : es el número de variables proposicionales del lenguaje considerado y  $s$ : es el número de literales que aparecen en la conjunción.

2. Si  $Q_K$  es una disyunción  $D_1 \vee \dots \vee D_r$  donde cada  $D_i$  es una conjunción de literales como:

$$\omega_{ps}(Q_K) = \max\{\omega_{ps}(D_1), \dots, \omega_{ps}(D_r)\}. \quad (2)$$

Cada caso de la base de datos es una fuente de información, es decir, cada caso (base de conocimiento)  $K$  es una conjunción de literales.

**Definición 2.** El operador  $\Delta_{ps}$  se define a partir de los valores que maximizan la suma de la satisfacción parcial de las bases del perfil. Sea  $E$  un perfil de conocimiento obtenido de las bases de conocimiento  $\{Q_{k_1}, \dots, Q_{k_m}\}$ , la fusión de  $E$  empleando satisfacción parcial, denotada por  $\Delta_{ps}(E)$  es una función que va del conjunto de perfiles de creencias al conjunto de bases de conocimiento tal que:

$$mod(\Delta_{ps}(E)) = \left\{ \omega \in \mathcal{W} \mid \sum_{i=1}^m \omega_{ps}(Q_{K_i}) \geq \sum_{i=1}^m \omega'_{ps}(Q_{K_i}) \text{ para todo } \omega' \in \mathcal{W} \right\}. \quad (3)$$

### 3.4. Implementación del operador PS-Merge

El método trata cada caso del conjunto de datos como una fuente de información, y cada base de conocimiento tiene la siguiente forma:  $K = l_1 \wedge l_2 \wedge l_3 \wedge \dots \wedge l_{15} \rightarrow l_{16}$ . Por ejemplo, cuando ocurre el caso 10, se representa como  $K_{10} =$



(0,0,1,1,1,1,0,0,0,1,1,1,1,1,0,0), dado que los datos de entrada y salida están en formato binario.

Se identifican 16 variables en el conjunto de datos, y la fórmula general para obtener el número de combinación es:  $2^n$ , es decir, las posibles combinaciones son  $2^{16} = 65,536$ :

- 1) 0000000000000000
- 2) 0000000011000001
- 3) 0001100010000010
- ...
- 65,536) 1111111111111111

Si los casos están en FNC, deben transformarse a FND.

De acuerdo a la expresión  $a \rightarrow b \equiv \neg a \vee b$ , se puede determinar una forma normal disyuntiva (FND) o forma normal conjuntiva (FNC) correspondiente a cada K que niega cada variable y conecta todos los elementos con disyunciones,  $\neg l_1 \vee \neg l_2 \vee \neg l_3 \vee \neg l_4 \vee \neg l_5 \vee \neg l_6 \vee \neg l_7 \vee \neg l_8 \vee \neg l_9 \vee \neg l_{10} \vee \neg l_{11} \vee \neg l_{12} \vee \neg l_{13} \vee \neg l_{14} \vee \neg l_{15} \vee l_{16}$ , después de la transformación de los casos en sus formas normales, se aplica el algoritmo, el objetivo que es el valor que debe ser y el resultado que es la evaluación obtenida por el algoritmo.

Al realizar la fusión se obtienen combinaciones, en donde a veces se puede obtener dos resultados 0, 1 con el mismo valor de atributos, dado que las características de los datos siempre vienen en paquetes considerando todas las variables, se pueden obtener ambas, pero no son contradicciones, es decir, el algoritmo puede encontrar una fórmula como  $\{\neg p_1 \wedge \neg p_2 \wedge p_3 \wedge p_4 \wedge p_5 \wedge p_6 \wedge \neg p_7 \wedge \neg p_8 \wedge \neg p_9 \wedge p_{10} \wedge p_{11} \wedge p_{12} \wedge p_{13} \wedge p_{14} \wedge p_{15} \wedge p_{16}\}$  y  $\{\neg p_1 \wedge \neg p_2 \wedge p_3 \wedge p_4 \wedge p_5 \wedge p_6 \wedge \neg p_7 \wedge \neg p_8 \wedge \neg p_9 \wedge p_{10} \wedge p_{11} \wedge p_{12} \wedge p_{13} \wedge p_{14} \wedge p_{15} \wedge \neg p_{16}\}$ , lo cual es un diagnóstico ambiguo: el paciente tiene neuropatía diabética, pero en el conjunto de prueba es sano.

Se dividió la base de conocimiento en dos conjuntos, el conjunto de entrenamiento y el conjunto de pruebas, asignando el 70% del conjunto de datos original al entrenamiento y el 30% para el conjunto de pruebas, esto de manera aleatoria, la asignación quedó con 4200 casos de entrenamiento y 1800 de casos para prueba.

Ahora bien, en la primer fase se fusionaron 4200 casos del conjunto de entrenamiento. Para obtener el resultado de solución; se compararon los casos de la nueva base de conocimiento generada por *la fusión de creencias* con los casos del conjunto de prueba; si existían modelos con el mismo valor que la base obtenida por la fusión, podría ser 0, 1 o ambos.

Si encontró ambos, significa que en la base de fusión hay tanto positivo como negativo con las mismas variables y se representa como BB (Both-Both), en otras palabras, si en la fusión existen ambos casos 0 o 1, quiere decir que no sabe cómo dar solución, porque no hay ningún caso que pueda decir que sea 0 o 1, incluso, si una de las soluciones obtenida de la fusión coincide con un caso del conjunto de prueba, entonces es 0 y se interpreta con B, mientras que, al

segundo dígito para que sea 0, 1, se le asigna al primer dígito una B, para que no tenga 0 o ambos, porque puede que en el siguiente caso coincidan.

Para el tercer método binario, el algoritmo no encontró una respuesta en 2 casos, con 16 variables; otro aspecto del algoritmo fue que obtuvo 88 resultados de falsos negativos, en donde el resultado fue B, es decir, el algoritmo no encontró el valor correspondiente, para determinar si es un caso con neuropatía diabética o sano.

Para reducir los casos confusos se realizaron más pruebas. Mientras que, para el segundo método binario, el algoritmo encontró 65 resultados de falsos negativos, 254 resultados falsos positivos, dicho de otra manera, encontró casos de neuropatía diabética en donde los casos deberían ser sano. Asimismo, el algoritmo no encontró respuesta en 1 caso. Y finalmente para el primer método binario, el algoritmo encontró 56 casos falsos positivos, asignando un valor que se asimilara al valor que debería ser, en los casos en donde ambos tienen que es igual 0,1 entonces sí pudo definir un solo valor, ya sea que esté correcto o incorrecto, pero determinó un valor o se quedó como BB.

En otras palabras, se transformó lo que se inició como 0,1 a BB, en el caso que sea BB, ya sea que entró como BB desde el inicio, o que entró como 0,1, hace el proceso para que encuentre un resultado, si no lo encuentra, queda con los dígitos que inicio. Si son iguales ambos dígitos, y se asignó BB o encontró una solución, al segundo dígito se asigna una B, en el caso que ambas sean diferentes, sólo 1 se satisface y no ambas, sólo se le asigna B a una y no serán iguales, en el caso que ambas se satisfacen se asigna BB.

En la segunda fase, el algoritmo analizó los resultados obtenidos de la primer fase aplicando *la fusión de creencias*, los cuales son resultados de reducción de los B, en un momento dado con cada caso se recorre toda la base de datos para distinguir que tanto satisface a toda la base de datos, inicialmente nadie está satisfecho, aunque son 4200 registros, realmente en las combinaciones no se utilizan todas, se pueden encontrar algunas que se repiten, los resultados deben satisfacer las 16 variables para todos los casos.

Cuando en la primera entrada el máximo en satisfabilidad es 0, muy probablemente cambiará, por ejemplo, tenemos  $K_{10} = 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$  y la base de datos en cuestión esté como  $K = 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1$ , pero si la suma parcial al menos tiene un valor que satisface parcialmente una pequeña porción de la base de datos en cuestión, es una solución, significa que satisface toda la base de datos de datos.

Analiza los 4200 casos hasta obtener una mejor solución, de las 65536 combinaciones sólo unas cuantas maximizan toda la satisfacción de la base de datos en cuestión, en efecto, cuando el algoritmo obtiene resultados que hacen que la mayoría de los casos estén satisfechos al máximo, en los casos en donde ambos tienen 0,1 sí pudo definir un solo valor para el conjunto de prueba, ya sea que este correcto o incorrecto, pero determinó un solo valor o quedó como BB, dando una solución correcta o incorrecta.

#### 4. Resultados y discusiones

En esta investigación se aplicó el operador *PS-Merge* a un conjunto de datos con biomarcadores de la marcha de neuropatía diabética, con registros de personas con la enfermedad y personas sanas.

Una persona con neuropatía diabética que tiene un resultado positivo en la prueba y se denomina Verdadero Positivo (TP), mientras que una persona con neuropatía diabética, pero con una prueba negativa se denomina Falso Negativo (FN), una persona sin neuropatía diabética que tiene un resultado positivo se denomina Falso Positivo (FP), mientras que una persona sin neuropatía diabética, pero con una prueba negativa se denomina Verdadero Negativo (TN) [9]. La precisión se calculó a partir de la ecuación  $((TP+TN)/(TP+FP+FN+TN))$ . En la Tabla 3, se observa la matriz de confusión base.

**Tabla 3.** Matriz de confusión.

	Predijo sí	Predijo no
Real sí	Verdadero positivo	Falso positivo
Real no	Falso negativo	Verdadero negativo

Deriva una matriz de confusión para los tres métodos binarios, se utilizaron 4200 registros para cada método. Para el primer método el operador clasificó 4198, en donde, clasificó correctamente 1142 casos verdaderos positivos, 340 casos verdaderos negativos, por lo contrario, 88 casos falsos positivos y 228 falso negativo, y en 2 registros no considero correcto o incorrecto un resultado. La precisión se calculó a partir de la ecuación  $((TP+TN)/(TP+FP+FN+TN))$ , tal como,  $(1142+340) / (1142+88+228+340) = 82.42\%$ , (Véase en la Tabla 4).

**Tabla 4.** Matriz de confusión, primer método binario.

Clase	Sano en control	Enfermo
Sano en control	1142	88
Enfermo	228	340

En el segundo método binario, el operador clasificó correctamente 1124 casos verdaderos positivos, 356 casos verdaderos negativos, no obstante, el operador obtuvo 65 casos falsos positivos y 254 casos falsos negativos, así como 1 registro en el cual no encontró respuesta (Véase en la Tabla 5). La precisión fue  $(1124+356) / (1124+65+254+356) = 82.26\%$ .

Finalmente, para el tercer método binario, el operador clasificó correctamente 1132 casos verdaderos positivos, 340 casos verdaderos negativos, en cambio, 56 casos falsos positivos y 272 falso negativo (Véase en la Tabla 6). Asimismo, la precisión fue  $(1132+340) / (1132+56+272+340) = 81.77\%$ .

**Tabla 5.** Matriz de confusión caso 2.

Clase	Sano en control	Enfermo
Sano en control	1124	65
Enfermo	254	356

**Tabla 6.** Matriz de confusión caso 3.

Clase	Sano en control	Enfermo
Sano en control	1132	56
Enfermo	272	340

*PS-Merge* es un operador basado en lógica proposicional, que no se había implementado para clasificar datos de neuropatía diabética, el proceso utilizó de tres técnicas: *fusión de creencias*, *revisión de creencias* y la *noción de satisfacción parcial*. En esta investigación se logró clasificar correctamente datos de neuropatía diabética utilizando *la fusión de creencias*, obteniendo resultados de precisión de más del 80 %; resultados aceptables, en comparación a lo obtenido con técnicas de aprendizaje automático utilizando el mismo conjunto de datos [17], y al aplicar clasificadores ensamblados y aprendizaje profundo, con un resultado de precisión de más del 85 %. Ahora bien en [9], aplicaron el operador *PS-Merge* para casos de diagnóstico de cáncer oral, los resultados permitieron plantear la posibilidad de fusionar un conjunto de datos sin afinar parámetros, en donde los mejores resultados dan una precisión superior al 75 % de instancias correctamente clasificadas.

## 5. Conclusiones y trabajos futuros

En esta investigación hemos logrado clasificar clases binarias, implementando el operador *PS-Merge*, utilizando la noción de *fusión de creencias*, *revisión de creencias* y la *noción de satisfacción parcial*, en casos de neuropatía diabética, en donde los resultados mostraron un valor de precisión superior al 80 % de instancias correctamente clasificadas en los tres métodos binarios, por lo que es comparable a lo que se logró con técnicas de aprendizaje automático empleando el mismo conjunto de datos. En la Sección 3.2 presentamos una idea para binarizar el conjunto de datos. Asimismo, en la Sección 4 señalamos los resultados aplicando el operador *PS-Merge*, así como la matriz de confusión para cada uno de los métodos binarios. Se puede observar que el operador puede ser utilizado en cualquier conjunto de datos que esté representado en vectores binarios. Puesto que, el operador no utiliza la estadística y configuración alguna de parámetros, como lo realizan las técnicas de aprendizaje automático con algoritmos clásicos para datos del área médica, tampoco se requieren técnicas de preprocesamiento de datos, sólo se llevó a cabo una selección de variables para identificar el número de variables a fusionar.

Como trabajo a futuro se pretende:

- Implementar el operador *PS-Merge* en otros conjuntos de datos y comparar los resultados del operador con resultados de algoritmos clásicos.
- Ampliar el operador para clases multiclase ya que sólo considera clases binarias.

## Referencias

1. Chávez-Bosquez, O., Pozos-Parra, P.: Modelando la toma de decisiones mediante fusión de creencias. *Recta* 15(2) (2014)
2. Chávez-Bosquez, O., Pozos-Parra, P., Ma, J.: Implementing  $\Delta$  ps (ps-merge) belief merging operator for belief revision. *Computación y Sistemas* 21(3), 419–434 (2017)
3. Creignou, N., Papini, O., Rümmele, S., Woltran, S.: Belief merging within fragments of propositional logic. *ACM Transactions on Computational Logic (TOCL)* 17(3), 1–28 (2016)
4. Cuevas, I.J.F., Núñez, Z.A.C., Ascencio, R.L., Vásquez, C.: Detección de neuropatía diabética periférica en adultos mayores de 60 años en el centro de salud “méxico bid” de colima, méxico. *Archivos de medicina* 14(4), 1 (2018)
5. Diller, M., Haret, A., Linsbichler, T., Rümmele, S., Woltran, S.: An extension-based approach to belief revision in abstract argumentation. *International Journal of Approximate Reasoning* 93, 395–423 (2018)
6. Everaere, P., Konieczny, S., Marquis, P.: On egalitarian belief merging. In: Fourteenth International Conference on the Principles of Knowledge Representation and Reasoning (2014)
7. Gómez, S.A., Simari, G.R.: Merging of ontologies using belief revision and defeasible logic programming. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial* 16(52), 16–28 (2013)
8. Ing, V.A., Salazar, J., Garicano, C., MgS, M.C.G.M., Añez, R., et al.: Una introducción a las aplicaciones de la inteligencia artificial en medicina: Aspectos históricos. *Revista Latinoamericana de Hipertension* 14(5), 590–600 (2019)
9. Kareem, S.A., Pozos-Parra, P., Wilson, N.: An application of belief merging for the diagnosis of oral cancer. *Applied Soft Computing* 61, 1105–1112 (2017)
10. Luna, I., De, G., Zacarias Flores, F., García García, A.D.: Algoritmo basado en modelos para la revisión de creencias entre formas normales conjuntivas. *Computación y Sistemas* 21(3), 435–448 (2017)
11. Molina, E., Diaz, J., Hidalgo-Silva, H., Chávez, E.: Algoritmos de binarización robusta de imágenes con iluminación no uniforme. *Revista Iberoamericana de Automática e Informática industrial* 15(3), 252–261 (2018)
12. Osisanwo, F., Akinsola, J., Awodele, O., Hinmikaiye, J., Olakanmi, O., Akinjobi, J.: Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)* 48(3), 128–138 (2017)
13. Pop-Busui, R., Boulton, A.J., Feldman, E.L., Bril, V., Freeman, R., Malik, R.A., Sosenko, J.M., Ziegler, D.: Diabetic neuropathy: a position statement by the american diabetes association. *Diabetes care* 40(1), 136–154 (2017)
14. Ramírez, R.B., Arellano, R.M., Arévalo, J.J.: la tecnología y la inteligencia artificial como futuro en el área médica. *Universitas, Revista de Ciencias Sociales y Humanas* (21), 185–190 (2014)

15. Rodríguez, R., Sosa, J.: Importance of the use of the parametric logarithm in a non-supervised strategy for the image binarization. *IEEE Latin America Transactions* 14(3), 1434–1439 (2016)
16. Sáenz, K.d.J.B.: Clasificación de los frutos de café según su estado de maduración y detección de la broca mediante técnicas de procesamiento de imágenes. *Prospectiva* 14(1), 15–22 (2016)
17. Sánchez-DelaCruz, E., Weber, R., Biswal, R., Mejía, J., Hernández-Chan, G., Gómez-Pozos, H.: Gait biomarkers classification by combining assembled algorithms and deep learning: Results of a local study. *Computational and Mathematical Methods in Medicine* (2019)
18. Schwind, N., Inoue, K., Bourgne, G., Konieczny, S., Marquis, P.: Belief revision games. In: *Twenty-Ninth AAAI Conference on Artificial Intelligence* (2015)
19. Shear, T., Fitelson, B.: Two approaches to belief revision. *Erkenntnis* 84(3), 487–518 (2019)
20. Triana, N., Jaramillo, A.E., Gutiérrez, R.M., Rodríguez, C.A.: Técnicas de umbralización para el procesamiento digital de imágenes de gem-foils. *Scientia et Technica* 21(4), 352–359 (2016)
21. Walulya, I., Palyvos-Giannas, D., Nikolakopoulos, Y., Gulisano, V., Papatriantafilou, M., Tsigas, P.: Viper: A module for communication-layer determinism and scaling in low-latency stream processing. *Future Generation Computer Systems* 88, 297–308 (2018)
22. Zhang, S.S., Wu, Z., Zhang, L.C., Zhang, Z., Chen, R.P., Huang, Y.H., Chen, H.: Efficacy and safety of pregabalin for treating painful diabetic peripheral neuropathy: a meta-analysis. *Acta Anaesthesiologica Scandinavica* 59(2), 147–159 (2015)