

Modelo difuso para la predicción de casos de obesidad empleando el árbol GFID3 generalizado

Christian Suca, André Córdova, Abel Condori, Jordy Cayra

Universidad Nacional de San Agustín,
Escuela Profesional de Ingeniería de Sistemas, Arequipa,
Perú

{andrepacord, csucavelando, perez}@gmail.com, acondoricas@unsa.edu.pe

Resumen. La obesidad es la base de muchas enfermedades crónicas importantes, y por lo tanto es uno de los mayores problemas de salud en el mundo, siendo el IMC el principal indicador empleado para la predicción, sin embargo, el porcentaje de grasa corporal es uno de los factores asociados con problemas de obesidad, siendo estos porcentajes distintos para hombres y mujeres. El objetivo de este artículo es desarrollar un modelo novel difuso clasificado para hombres y mujeres para la predicción de casos de obesidad en personas de 6-17 años. El presente trabajo estudió los factores e indicadores empleados como medidas de predicción de casos de obesidad y desarrolló un modelo difuso que puede ser utilizado como un índice de predicción de obesidad en el campo médico utilizando percentiles de peso, altura y el porcentaje de grasa corporal (BF). Los pasos involucrados en este estudio son: una revisión de los factores de obesidad infantil, colección Pre-procesamiento y clasificación de los datos, fuzzificación de las entradas y la generación de reglas óptimas, basadas en el árbol GFID3. Los resultados obtenidos demuestran que el modelo difuso desarrollado obtiene una exactitud 83.65% para hombres y 76.13% en mujeres para la predicción de casos de obesidad.

Palabras clave: Obesidad, árbol GFID3, lógica difusa, IMC.

Fuzzy Model for Prediction Obesity Cases Using the Generalized GFID3 Tree

Abstract. Obesity is the cause of many major chronic diseases, and therefore is one of the biggest health problems in the world, being the IMC the primary indicator used for prediction, however, the percentage of body fat is one of the factors associated with obesity problems, these percentages being different for men and women. The aim of this paper is to develop a new classified novel fuzzy model for men and women for the prediction of cases of obesity in people aged 6-17 years. This paper studied the factors and indicators used as measures of prediction cases of obesity and developed a fuzzy model which can be used as a predictor of obesity in the medical field using percentiles weight, height and percentage of body fat (BF). The steps involved in this study are: a review of the

factors of childhood obesity, Pre-processing collection and classification of data fuzzification of inputs and generating optimal rules, based on the tree GFID3. The results show that the developed fuzzy model accuracy gets 83.65% and 76.13% for men and women for predicting cases of obesity.

Keywords: Obesity, GFID3 tree, fuzzy logic, IMC.

1. Introducción

La condición clínica conocida como obesidad actualmente asume la característica de epidemia universal, esta enfermedad es causa importante de muerte y enfermedades relacionadas (también denominadas co-morbilidades), por otra parte se relaciona directamente y de forma proporcional a la magnitud de la condición clínica de la Obesidad. La OMS [6] indica que el número de niños y lactantes que padecen sobrepeso aumento de 32 millones a 42 millones en el 2013, la OMS establece que de acuerdo a la tendencia posiblemente este número crezca hasta 70 millones para el 2025. La definición de obesidad, medición, clasificación y tratamiento no se determinan fácilmente, encontrar un sistema para evaluar y clasificar el nivel de Obesidad, sobre todo si se refiere a los riesgos o la prescripción de los tratamientos es de gran interés clínico [1]. El exceso de peso puede provenir de músculo, hueso, grasa y / o agua en el cuerpo, la presencia de exceso de grasa en el tronco o en el abdomen fuera de proporción con la grasa corporal total es un predictor independiente de los factores de riesgo y morbilidad [19]. Hoy en día se realizan múltiples investigaciones para clasificar a una persona en condición de sobrepeso, obesidad, etc. utilizando diversos parámetros antropométricos, cuando el interés es determinar la mejor indicación terapéutica y el tratamiento, así como el mejor mecanismo apropiado para el análisis de datos, la búsqueda de un método más preciso para evaluar la obesidad y en consecuencia el mejor tratamiento [1]; por este motivo surge la pregunta, es posible la aplicación y utilización de un modelo difuso basado en árboles de decisión que pueda ser utilizado como un índice de predicción de casos de obesidad empleando un conjunto de registros médicos de crecimiento de personas entre 6-17.

En esta investigación, se desarrolló un modelo difuso para la predicción de casos de obesidad. Este enfoque novel combina las técnicas de árboles de decisión difuso ID3 y la lógica difusa para la generación de un índice basado en un conjunto de reglas así como un modelo para la predicción de casos de obesidad tomando como referencia los percentiles de edad, peso y BF y que pueden ser tomados como ayuda en el campo médico. Este modelo está clasificado tanto para hombres como para mujeres, tomando como referencia el hecho de que las medidas y rangos, como el porcentaje de grasa tienen rangos distintos en hombres y mujeres [1], por lo tanto el modelo clasificado es capaz de predecir si una persona es propensa a casos de obesidad con mayor precisión y ofrecen una alternativa de ayuda en el campo médico. Para relacionar las diferentes variables de nuestra base de datos se utiliza la Clasificación asociativa que es un campo integrado de la Minería de Reglas de Asociación (ARM) y Clasificación. El ARM tradicional fue diseñado teniendo en cuenta que los elementos tienen la misma importancia y en la base de datos simplemente se menciona su presencia o ausencia [18].

El presente artículo está estructurado de la siguiente forma: en la sección 2: Trabajos Previos, donde se verá las investigaciones realizadas anteriormente. Posteriormente en la sección 3, se mostrará los Métodos y Herramientas utilizados en el modelo difuso. En la Sección 4 se mostrará los Experimentos y Resultados que se obtuvieron luego, además de la comparación con otros artículos que analizaron la misma problemática. Finalmente, en la sección 5 Conclusiones y Trabajo a Futuro, se establecerá cuál fue el resultado obtenido mediante el modelo propuesto.

2. Trabajos relacionados

En el 2008 se propuso un nuevo índice de medida, como una alternativa en la toma de decisiones en cirugía bariátrica con respecto a su grado de obesidad, tomando como referencia el IMC y el porcentaje de grasa corporal en personas adultas, tomando como referencia los lineamientos de la OMS y la experticia de los autores para la generación de reglas [1]. En el contexto mundial de la salud el índice de masa corporal (IMC) es considerado el principal criterio para determinar si una persona tiene problemas de obesidad., sin embargo, el exceso de grasa en relación al porcentaje de grasa corporal (% BF), es en realidad el principal factor perjudicial en la enfermedad de la obesidad que por lo general se pasa por alto. El estudio propone un modelo difuso mediante la asociación de IMC y %BF.

En el 2013 se presenta un modelo difuso para la evaluación de la obesidad abdominal (AO) y el riesgo Cardio-Metabólico (CMR) [2]. Sobre la base de los indicadores más conocidos de mediciones antropométricas; como el índice de masa corporal (IMC), así como la circunferencia de la cintura (CC) y la relación cintura-altura (RCEst); se hizo la construcción de una función de pertenencia (OF), así también se mapeo los conjuntos difusos de IMC, AO.

En el 2014 se realizó el sistema de toma de diagnóstico médico difusos [3] para ayudar y apoyar a evaluar los pacientes con dolor torácico de acuerdo al grado de obesidad. Para este artículo se fuzzificó el índice de masa corporal (IMC). Ese mismo año [4] realizó un nuevo árbol de decisión difusa. Se propuso un algoritmo ID3 fuzzy generalizado en el uso generalizado de entropía de la información (abreviado como GFID3), que mejoró las deficiencias de algoritmos de árboles de decisión fuzzy existentes, ya que no toman en cuenta el impacto de la no linealidad; por lo cual son incapaces de integrar la selección de los atributos extendidos. En el 2015 se hizo el estudio en los niños de 4 a 6 años [5], en el estudio se tienen en cuenta el índice de masa corporal y la actividad física que realizan para el análisis de la obesidad en los niños.

3. Métodos y herramientas

La fuente de datos que se utilizó son un conjunto de 5962 registros médicos de personas entre los 6-17 años de edad, proporcionadas por instituciones educativas de Brasil. Los parámetros seleccionados y que fueron empleados para la predicción se muestran en la Tabla 1.

Posteriormente, se procedió al cálculo del IMC de acuerdo a las pautas clínicas dada por [6], de acuerdo al análisis realizado sobre los datos recogidos, este cálculo se realizó

con respecto a los percentiles de edad, altura y peso y sexo. Después de calcularse el IMC en los niños y adolescentes, los valores percentiles del IMC y edad se muestran en la figura 1, proporcionada por los Centros para el Control y la Prevención de Enfermedades (CDC) para el IMC por edad para niñas para obtener la categoría del percentil. El IMC por edad para niños y niñas para obtener la categoría del percentil se muestran en la Fig. 1 y 2.

Tabla 1. Parámetros de entrada.

| Parámetros | Descripción |
|--------------------|--|
| Percentil Estatura | Índice de clasificación de la estatura por edad basado en percentiles (0-100). |
| Percentil Peso | Índice de clasificación del peso por edad basado en percentiles (0-100). |
| BF | Porcentaje de grasa |

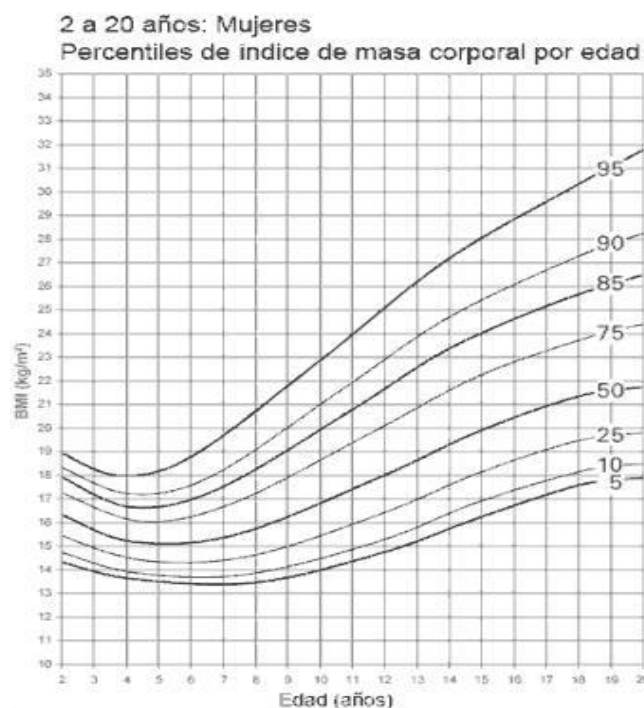


Fig.1. Índice de masa corporal en niñas basado en percentiles [6].

Los percentiles son el indicador que se utiliza con más frecuencia para evaluar el tamaño. El percentil indica la posición relativa del número del IMC del niño entre niños del mismo sexo y edad. Las categorías del nivel de peso del IMC por edad y sus percentiles correspondientes se muestran en la siguiente Tabla 1.

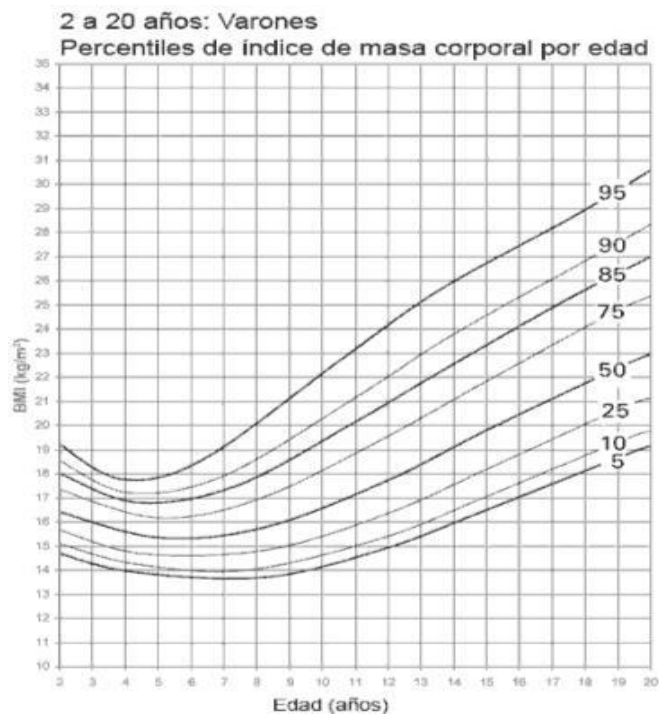


Fig.2. Índice de masa corporal en niños basada en percentiles [6].

Tabla 2. Pautas clínicas en la identificación de la obesidad en los niños [6].

| Clasificación | Percentil IMC |
|---------------|---------------|
| Bajo peso | <5 |
| Normal | >=5 y <85 |
| Sobrepeso | >=85 y <95 |
| Obesidad | >=95 |

Tabla 3. Pautas clínicas en la valoración del porcentaje de grasa en los niños [7].

| Clasificación | Porcentaje de grasa |
|--------------------|---------------------|
| Muy Bajo | >0 y <5.5 |
| Bajo | >=5.5 y <11.5 |
| Rango Optimo | >=11.5 y <21 |
| Moderadamente Alto | >=21 y <26 |
| Alto | >=26 y <31 |
| Muy Alto | >=31 y <=100 |

El %BF se considera un mejor índice para evaluar la obesidad que el IMC ya que este último sólo funciona como un indicativo del exceso de peso y el primero es un indicativo de exceso de la masa grasa y el riesgo de seguir siendo obesos [1][7]. Por lo tanto, %BF ha sido suficientemente aprovechado en este artículo. Los valores para niñas y niños se muestran en las Tablas 2 y 3.

Tabla 4. Pautas clínicas en la valoración del porcentaje de grasa las niñas [7].

| Clasificación | Porcentaje de grasa |
|--------------------|---------------------|
| Muy Bajo | >0 y <11 |
| Bajo | >=11 y <15 |
| Rango Optimo | >=15 y <23.5 |
| Moderadamente Alto | >=23.5 y <31.5 |
| Alto | >=31.5 y <34.5 |
| Muy Alto | >=31 y <=100 |

Para introducir un mecanismo coherente y hacer frente a la obesidad en este artículo se consideró tanto el percentil de peso, el percentil de estatura y %BF para presentar un nuevo índice de la obesidad en niños mediante la teoría de la lógica difusa.

4. Experimentos y resultados

Se consideró 4 etapas para desarrollar el trabajo, estos son: Pre-procesamiento de datos, Fuzificación, Generación de Reglas, Aplicación del Sistema Mandani, Clasificación de Resultados. Además de la comparación del Sistema Difuso con un clasificador usando un algoritmo *Machine Learning*, que en este caso es un árbol de decisión C4.5.

4.1 Pre-Procesamiento de datos

En la etapa inicial, se disponía de una base de datos que poseía un total de 14 atributos entre edad, sexo, índices de masa corporal, variables de esfuerzo físico, pliegues cutáneos, etc. De acuerdo a [1] se establece los valores Crisp para poder delimitar los límites y rangos de los atributos IMC y BF, sin embargo estas medidas son adecuadas para personas adultas, además se debe establecer de acuerdo a los expertos cuales son los atributos de medida que son factores críticos en la predicción de obesidad [7, 1]. En su trabajo de investigación [1] propone un nuevo índice de clasificación de obesidad para adultos tomando como referencia IMC y el porcentaje de grasa, tomando como referencia los resultados dados por [1] se tomó como entradas al sistema, el porcentaje de grasa corporal, así como los percentiles de peso y edad, ya que el modelo propuesto está orientado a personas de 6-17 años. Los valores de porcentaje de grasa (BF) tienen valores Crisp distintos para hombres y mujeres, por lo que fue necesaria una clasificación de los registros de datos en hombres y mujeres. Otro punto que se considero es la conversión de los atributos de BMI, peso y estatura a un percentil, debido a que como lo menciona la OMS [8] una medida básica para el control

de la obesidad es el percentil del BMI, ya que esto permite poder establecer un porcentaje y compararlo con otras evaluaciones hechas a otros niños y ver si está en un rango aceptable. Se tomó una plantilla base para la conversión de BMI a Percentiles de Luis F. Romero, investigador de la universidad de Málaga [9]. El siguiente paso fue una clasificación de los datos en hombres y mujeres así como la clasificación de sus registros médicos en cada una de las clases Crisp para el cálculo del promedio de cada uno de los atributos, así como la desviación estándar que serán utilizados en la etapa de fuzificación.

4.2 Fuzzificación

Una vez con los datos preparados se necesitaba tener los rangos y clases para cada uno de estos atributos cada uno de estos se basaron en una guía de expertos como es la OMS y otros artículos de investigación sobre Obesidad Infantil. El primero que es para el caso de Percentiles de BMI, peso y estatura se pudo conseguir gracias a la OMS en [10] que dispone de una serie de gráficas y tablas de percentiles tanto para niños como para niñas siendo las clases: bajo peso, peso normal, sobrepeso, y obesidad. En cuanto al Porcentaje de Grasa de acuerdo a un artículo visto en [7] donde se establecía de acuerdo a los pliegues cutáneos el resultado de la masa grasa, masa magra y el porcentaje de grasa corporal igualmente para niños y niñas siendo las clases: muy bajo, bajo, normal, moderado, alto y muy alto. Hay que considerar también la salida de clasificación, esto es importante porque se necesita datos pre-iniciales o de entrenamiento que servirá en la etapa de generación de reglas. Con los rangos y clases ya tomados fueron utilizados como entrada al sistema Mandani para lógica difusa.

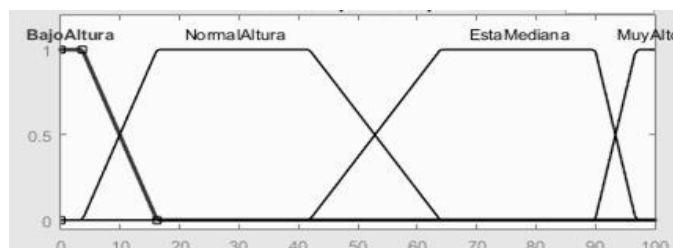


Fig.3. Índice Percentil Estatura en niños.

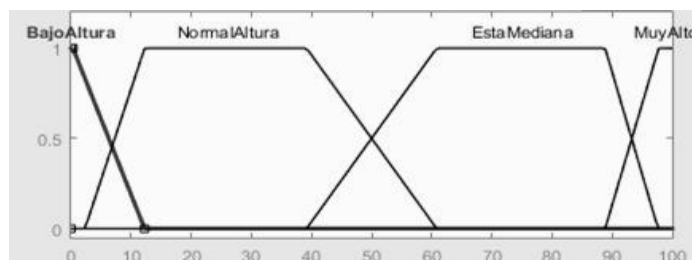


Fig.4. Índice Percentil Estatura en niñas.

En este caso se seleccionó la función de pertenencia y la que más se adapta al contexto del problema es la función de pertenencia de forma trapezoidal de acuerdo a

[1], estas funciones de pertenencia fueron establecidas mediante la función trapezoidal en base al promedio y las desviaciones estándares de cada rango para poder conseguir los cuatro parámetros de cada trapecio. La función de pertenencia trapezoidal se muestra en la ecuación 1, donde x es el atributo de entrada a ser fuzzificado, a , b , c , d son los rangos del trapecio.

$$f(x, a, b, c) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases} \quad (1)$$

Finalmente se obtuvo los resultados de cada atributo con sus valores fuzzificados o graduados pudiendo pertenecer a más de una clase. En las siguientes Figuras 3, 4, 5, 6, 7, 8, se muestran las funciones de pertenencias de cada atributo de entrada.

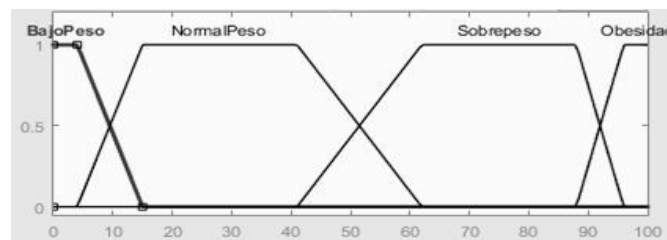


Fig.5. Índice Percentil Peso en niños.

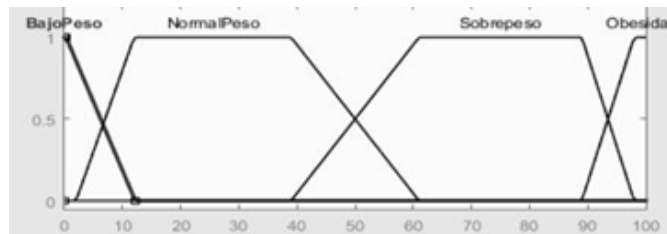


Fig.6. Índice Percentil Peso en niñas.

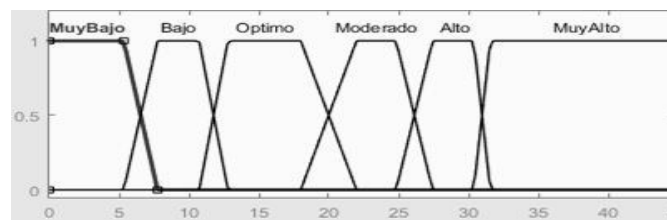


Fig.7. Índice BF en niños.

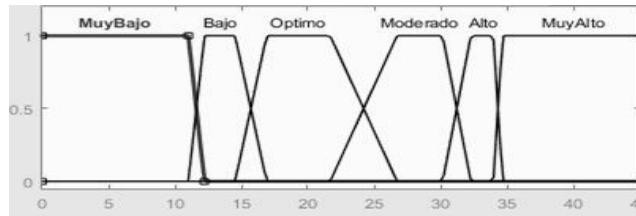


Fig.8. Índice BF en niña.

4.3 Generación de reglas

Para esta etapa, existen diversas herramientas con base en algoritmos de máquinas de aprendizaje que son generalizadas para trabajar con datos difusos, de acuerdo al estado del arte existen muchas opciones que depende mucho sobre los datos que se estén trabajando, así como la cantidad que representa, algunas técnicas poseen ventajas en cuanto a la robustez, pero también cuentan con desventajas como caer en óptimos locales. Otras técnicas permiten conseguir reglas muy óptimas, pero requieren de gran esfuerzo computacional para procesarlo. Y algunos carecen de métodos de simplificación, para sí obtener reglas con alta precisión, existen desde redes neuronales [11], mapas cognitivos [12], arboles de decisión [13, 14, 15] o algoritmos genéticos [16], todos generalizados para trabajar con datos fuzzificados.

Árbol ID3 Fuzzificado

El algoritmo ID3 fuzzificado, es una extensión del algoritmo clásico ID3, como un método para construir un árbol de decisión. En la construcción recursiva del árbol, se adopta una estrategia ‘divide y vencerás’. El algoritmo Fuzzy ID3 calcula la entropía de partición difusa basada en el atributo y luego selecciona el atributo con la mínima partición de entropía difusa con una extensión del atributo, la entropía de partición es una extensión de la entropía de partición Crisp del algoritmo ID3 [17].

Sea $\Omega = \{1,2, \dots, n\}$, un conjunto de entrenamiento, $A = \{A_1, A_2, \dots, A_m\}$ un conjunto de atributos, donde $\{A_{i1}, A_{i2}, \dots, A_{iki}\}$ es el rango del atributo A_i , en el árbol ID3 con valores fuzzificados, los nodos de decisión se consideran como conjuntos difusos en Ω , la entropía de partición difusa de un nodo no hoja D está dado por $FE(D, A_i)$ [4]:

$$FE(D, A_i) = \sum_{j=1}^{ki} \frac{m_{ij}}{m_i} E(D, A_{ij}), \quad (2)$$

$$E(D, A_{ij}) = - \sum_{k=1}^{ni} \frac{m_{ijk}}{\bar{m}_{ij}} \log_2 \frac{m_{ijk}}{\bar{m}_{ij}}, \quad (3)$$

donde $m_{ij} = M(A \cap A_{ij}), m_i = \sum_{j=1}^k m_{ij}, m_{ijk} = M(D \cap A_{ij} \cap C_k), \bar{m}_{ij} = \sum_{k=1}^n m_{ijk}$.

Una de las características del árbol ID3 difuso basado en la generalización de entropía (FID3) propuesto por [17], es que puede ser aplicado tanto a conjunto de datos con valores de atributos fuzzificados como a atributos numéricos y continuos.

El método elegido fue un árbol de decisión fuzzificado con el algoritmo ID3 generalizado para este propósito [17]. Aparte de considerar los buenos resultados que proporciona esta técnica, es necesario considerar sobre todo la rapidez con la cual puede proporcionar las reglas. Las reglas obtenidas, así como la tasa de precisión para cada clase de salida se muestran en las Tablas 4 y 5.

Tabla 5. Reglas difusas para Niñas.

| Nro | Reglas | Precisión |
|------------|---|------------------|
| 1 | IF Peso IS Normal AND Estatura IS MuyAlta AND BF IS MuyAlto THEN BajoPeso | 1.00 |
| 2 | IF Peso IS Sobrepeso AND BF IS Optimo AND Estatura IS Bajo THEN Obesidad | 1.00 |
| 3 | IF Peso IS Sobrepeso AND BF IS Alto AND Estatura IS Bajo THEN Sobrepeso | 1.00 |
| 4 | IF Peso IS Sobrepeso AND BF IS MuyAlto THEN Obesidad | 1.00 |
| 5 | IF Peso IS Normal AND Estatura IS Bajo THEN Sobrepeso | 0.98 |
| 6 | IF Peso IS Normal AND Estatura IS Normal AND BF IS MuyBajo THEN BajoPeso | 0.90 |
| 7 | IF Peso IS Sobrepeso AND BF IS Bajo AND Estatura IS Alta THEN Normal | 0.90 |
| 8 | IF Peso IS Normal AND Estatura IS MuyAta AND BFIS Moderado THEN BajoPeso | 0.88 |
| 9 | IF Peso IS Sobrepeso AND BF IS Bajo AND Estatura IS Normal THEN Normal | 0.86 |
| 10 | IF Peso IS Sobrepeso AND BF IS Moderado THEN Sobrepeso | 0.84 |

Tabla 6. Reglas difusas para Niños.

| Nro | Reglas | Precisión |
|------------|---|------------------|
| 1 | IF Peso IS BajoPeso AND Estatura IS Normal AND BF IS Moderado THEN BajoPeso | 1.00 |
| 2 | IF Peso IS Normal AND Estatura IS Bajo AND BF IS Optimo THEN Normal | 1.00 |
| 3 | IF Peso IS Normal AND Estatura IS Bajo AND BF IS Alto THEN Sobrepeso | 1.00 |
| 4 | IF Peso IS Normal AND Estatura IS Normal AND BF IS Moderado THEN Sobrepeso | 1.00 |
| 5 | IF Peso IS Normal AND Estatura IS Normal AND BF IS Alto THEN Sobrepeso | 1.00 |

| Nro | Reglas | Precisión |
|-----|--|-----------|
| 6 | IF Peso IS Sobrepeso AND BF IS MuyBajo AND Estatura IS Normal THEN Normal | 1.00 |
| 7 | IF Peso IS Sobrepeso AND BF IS MuyBajo AND Estatura IS SobreAlto THEN Normal | 1.00 |
| 8 | IF Peso IS Sobrepeso AND BF IS MuyBajo AND Estatura IS Alto THEN BajoPeso | 1.00 |
| 9 | IF Peso IS Sobrepeso AND BF IS MuyAlto AND Estatura IS Bajo THEN Obesidad | 1.00 |
| 10 | IF Peso IS Sobrepeso AND BF IS MuyAlto AND Estatura IS Alto THEN Sobrepeso | 1.00 |

Con esto finalmente se pudo insertar dentro del requerimiento de reglas del sistema difuso para la clasificación de peso infantil, siendo dichas reglas las que pasaran a formar parte del resultado o salida del sistema difuso.

4.4 Clasificación y resultados

Después de proporcionar el conjunto de entrenamiento con un total de 5962 registros; 20% de los datos fueron utilizados para entrenar el árbol GFID3, mientras que el conjunto de datos total fueron empleados para test. El modelo difuso generado se muestra en la Fig. 9.

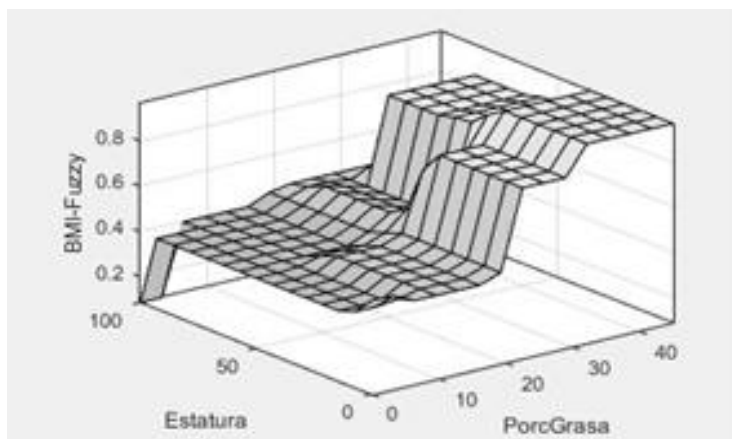


Fig.9. Relación Porcentaje de Grasa vs Estatura del Sistema de Clasificación Difusa para Niños.

En cuanto a los resultados de comparación se evaluó la categoría de precisión de las reglas, así como de las salidas. En la Tabla 6 y 7, se muestran algunos resultados de los primeros 10 registros para hombres y mujeres respectivamente.

Los resultados de clasificación en 2,935 niños y 3,021 en niñas estos resultados se muestran en porcentajes en las Tablas 8 y 9.

Tabla 7. Comparación de Salidas para Niños.

| Estatura Percentil | Peso Percentil | %Grasa | Salida | Clasificación del Modelo | Salida Experto | Clasificación Experto |
|---------------------------|-----------------------|---------------|---------------|---------------------------------|-----------------------|------------------------------|
| 56.56 | 37.03 | 11.9 | 38.3 | Normal | 27.273 | Normal |
| 83.45 | 54.57 | 10.1 | 43.40 | Normal | 24.03 | Normal |
| 52.75 | 55.81 | 11.5 | 43.85 | Normal | 60.84 | Normal |
| 86.76 | 96.28 | 27.8 | 97.54 | Obesidad | 96.21 | Obesidad |
| 99.93 | 99.97 | 33.5 | 97.54 | Obesidad | 99.71 | Obesidad |
| 97.37 | 99.41 | 28.6 | 97.54 | Obesidad | 98.70 | Obesidad |
| 87.30 | 99.28 | 30.2 | 97.541 | Obesidad | 99.13 | Obesidad |
| 98.58 | 97.00 | 18.5 | 85.19 | Sobrepeso | 91.24 | Sobrepeso |
| 96.40 | 96.66 | 17.8 | 85.31 | Sobrepeso | 93.15 | Sobrepeso |
| 97.98 | 97.14 | 16.7 | 85.43 | Sobrepeso | 92.35 | Sobrepeso |

Tabla 8. Comparación de Salidas para Niñas.

| Estatura Percentil | Peso Percentil | %Grasa | Salida | Clasificación del Modelo | Salida Experto | Clasificación Experto |
|---------------------------|-----------------------|---------------|---------------|---------------------------------|-----------------------|------------------------------|
| 82.40 | 30.15 | 13.7 | 7.38 | Normal | 1.69 | Bajo Peso |
| 78.08 | 58.84 | 11.9 | 45.29 | Normal | 37.25 | Normal |
| 84.90 | 54.29 | 15.4 | 43.82 | Normal | 19.87 | Normal |
| 100.00 | 100.00 | 36.2 | 97.53 | Obesidad | 99.82 | Obesidad |
| 86.53 | 99.28 | 30.7 | 97.53 | Obesidad | 99.38 | Obesidad |
| 98.73 | 99.60 | 33.8 | 97.53 | Obesidad | 98.92 | Obesidad |
| 98.39 | 99.57 | 24.6 | 97.53 | Obesidad | 98.92 | Obesidad |
| 30.76 | 73.25 | 25.9 | 85.56 | Sobrepeso | 91.87 | Sobrepeso |
| 96.62 | 96.80 | 29.7 | 91.88 | Sobrepeso | 93.32 | Sobrepeso |
| 70.34 | 90.57 | 30.7 | 86.08 | Sobrepeso | 94.02 | Sobrepeso |

Tabla 9. Resultados de Clasificación para Niños.

| Niños | Precisión |
|------------------------------|------------------|
| Correctamente Clasificados | 83.65 |
| Incorrectamente Clasificados | 16.35 |

Tabla 10. Resultados de Clasificación para Niñas.

| Niñas | Precisión |
|------------------------------|------------------|
| Correctamente Clasificados | 76.13 |
| Incorrectamente Clasificados | 23.87 |

4.5 Discusiones

De acuerdo al sistema difuso y los resultados de clasificación con el algoritmo difuso basado en árboles de decisión GFID3, se observa claramente que el sistema difuso es mejor en cuanto a la clasificación dando un promedio de precisión de reglas 82.79% para niños y un porcentaje de 76.37% de precisión para niñas. A diferencia del estudio realizado por Miyahira y su índice difuso basado en IMC y BF propuesto para la predicción de casos de obesidad en personas adultas, el presente estudio tomo como entrada al sistema difuso los valores percentiles de IMC, peso, estatura y %BF, considerando la distribución de estos en el rango de edad de 2-20 años y generando las reglas más óptimas mediante un conjunto de 5643 registros médicos y la aplicación del GFID3 generalizado que ha demostrado ser más robusto que las variantes propuestas del árbol ID3 y por tanto adecuado en el campo médico.

5. Conclusiones y trabajo a futuro

Este trabajo presenta un índice difuso para la predicción de casos de obesidad tanto en hombres como mujeres, tomando como entradas los Percentiles de peso, estatura y el %BF, basado en la lógica y los conjuntos difusos. Los resultados obtenidos son un conjunto de reglas y un modelo difuso que permite determinar si una persona es propensa o no a problemas de obesidad partir de sus registros médicos de peso, estatura y %BF, estas reglas fueron generadas mediante la fuzzificación de las entradas y como resultado del árbol GFID3 que ha demostrado tener gran exactitud frente a las variantes propuestas del ID3. El índice difuso desarrollado puede ser utilizado en el campo médico para la predicción de casos de obesidad, sin embargo, no está destinado a reemplazar o sustituir las métricas de evaluación empleados por médicos con experiencia; por el contrario, se debe ver como una herramienta de apoyo para la predicción de casos de obesidad en personas de 6-17 años. Como trabajos futuros se pretende extender este estudio para personas adultas, optimizando el conjunto de reglas generadas mediante un algoritmo evolutivo como los algoritmos genéticos y probar y validar sus resultados en el campo médico.

Referencias

1. Miyahira, S.A., de Azevedo, J.L.M.C., Araújo, E.: Fuzzy obesity index (MAFO I) for obesity evaluation and bariatric surgery indication. *J. Transl. Med.*, Vol. 9, No. 1, p. 134 (2011)
2. Nawarycz, T., Pytel, K., Drygas, W., Gazicki-Lipman, M., Ostrowska-Nawarycz, L.: A fuzzy logic approach to the evaluation of health risks associated with obesity. *Computer Science and Information Systems (FedCSIS)* (2013)
3. Orsi, T., Araujo, E., Simoes, R.: Fuzzy chest pain assessment for unstable angina based on Braunwald symptomatic and obesity clinical conditions. *Fuzzy Systems (FUZZ-IEEE), IEEE International Conference on*, pp. 1076–1082 (2014)
4. Jin, C., Li, F., Li, Y.: A generalized fuzzy ID3 algorithm using generalized information entropy. *Knowledge-Based Syst.*, Vol. 64, pp. 13–21 (2014)
5. Khanna, M., Srinath, D.N.K., Mendiratta, D.J.K.: The Study Of Obesity In Children Using Fuzzy Logic. *IJITR*, Vol. 3, No. 1 (2015)

6. OMS, (2007-2009). Comisión para acabar con la obesidad infantil: Datos y cifras sobre obesidad infantil Publishing. Recuperado de <http://www.who.int/end-childhood-obesity/facts/es/>.
7. Hall López, J.A., Monreal Ortiz, L.R., Ochoa Martínez, P.Y. Alarcón Mesa, E.: Porcentaje de grasa corporal en niños de edad escolar. Vol. 2, No. 1, pp. 13–17 (1988)
8. OMS, Centro para el Control y Prevención de Enfermedades CDC, Disponible en: http://www.cdc.gov/healthyweight/spanish/assessing/bmi/childrens_bmi/acerca_indice_masa_corporal_ninos_adolescentes.html (2015)
9. Romero, L.F.: Tabla de Crecimiento (Control de Percentiles), Disponible en: <http://www.ac.uma.es/~felipe/personal.html> (2009)
10. OMS, CDC growth charts to monitor growth for children age 2 years and older in the U.S. Disponible en: http://www.cdc.gov/growthcharts/clinical_charts.htm
11. Jang, J.-S.R.: ANFIS adaptive-network-based fuzzy inference system. Systems, Man and Cybernetics, IEEE Transactions on, Vol. 23, No.3, pp.665–685 (1993)
12. Kosko, B.: Fuzzy cognitive maps. Int. J. Man. Mach. Stud., Vol. 24, No. 1, pp. 65–75, Jan. (1986)
13. Lee, K.M., Lee, K.M. Lee, J.H., Lee-Kwang, H.: A fuzzy decision tree induction method for fuzzy data. Fuzzy Systems Conference Proceedings (1999)
14. Yuan, Y., Shaw, M.J.: Induction of fuzzy decision trees. Fuzzy Sets Syst., Vol. 69, No.2, pp. 125–139 (1995)
15. Janikow, C.Z.: Fuzzy decision trees: issues and methods. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, Vol.28, No.1, pp. 1–14 (1998)
16. Yuan, Y., Zhuang, H.: A genetic algorithm for generating fuzzy classification rules. Fuzzy Sets Syst., Vol. 84, No. 1, pp. 1–19 (1996)
17. Jin, C., Li, F., Li, Y.: A generalized fuzzy ID3 algorithm using generalized information entropy. Knowledge-Based Syst., Vol. 64, pp. 13–21 (2014)
18. Soni, S., Vyas, O.P.: Fuzzy Weighted Associative Classifier: A Predictive Technique for Health Care DataMining. Int. J. Comput. Sci. Eng. Inf. Technol., Vol. 2, No. 1, pp. 11–22 (2012)
19. Bouharati, S., Bounechada, M., Djoudi, A., Harzallah, D., Alleg, F., Benamrani, H.: Prevention of Obesity using Artificial Intelligence Techniques. International Journal of Science and Engineering Investigations, Vol. 1, No. 9, pp. 146–150 (2012)