

Neuro-Diffuse System with Optimization via Particulate Swarms for the Classification of Obesity in Children and Adolescents

José Sulla-Torres^{1,2}, Christian Soto-Paredes¹, Rocío Cárdenas-Soria¹, Lucía Huancco-Coila¹
fatima_113@hotmail.com

¹ Universidad Nacional de San Agustín, Arequipa-Perú, josullato@gmail.com, cjsotopa@gmail.com, cardenassoria.r@gmail.com,

² Universidad Católica de Santa María, Arequipa-Perú, jsullato@ucsm.edu.pe

Abstract- The following article aims to classify obesity in children and male adolescents in a range from six to seventeen, using neural networks and fuzzy logic, for which the neuro-fuzzy model ANFIS (used artificial Neural Network Fuzzy Inference System) which is optimized using PSO (particle swarm optimization). Experimental tests show an error RMSE 8.31, after making 500 iterations of the algorithm PSO. This result is considered acceptable within the characteristics of this research conditions.

Keywords-- Classification, ANFIS, PSO, IMC, RMSE, Obesity.

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2016.1.1.247>

ISBN: 978-0-9822896-9-3

ISSN: 2414-6390

Sistema Neurodifuso con Optimización por Enjambre de Partículas para la Clasificación de la Obesidad en Niños y Adolescentes

José Sulla-Torres^{1,2}, Christian Soto-Paredes¹, Rocío Cárdenas-Soria¹, Lucia Huancco-Coila¹

¹ Universidad Nacional de San Agustín, Arequipa-Perú, josullato@gmail.com, cjsotopa@gmail.com, cardenassoria.r@gmail.com, fatima_113@hotmail.com

² Universidad Católica de Santa María, Arequipa-Perú, jsullato@ucsm.edu.pe

Resumen— *El siguiente artículo tiene como objetivo clasificar la obesidad en niños y adolescentes del sexo masculino, en un rango que va desde los seis hasta los diecisiete años, utilizando Redes Neuronales y Lógica difusa, para ello se utiliza el modelo neuro-difuso ANFIS (Artificial Neural Network Fuzzy Inference System) el cual se optimizará utilizando PSO (Optimización por Enjambre de Partículas). Las pruebas experimentales realizadas muestran un error RMSE de 8.31, después de realizar 500 iteraciones del algoritmo PSO. Este resultado se considera aceptable dentro de las condiciones propias de esta investigación.*

Abstract— *The following article aims to classify obesity in children and male adolescents in a range from six to seventeen, using neural networks and fuzzy logic, for which the neuro-fuzzy model ANFIS (used artificial Neural Network Fuzzy Inference System) which is optimized using PSO (particle swarm optimization). Experimental tests show an error RMSE 8.31, after making 500 iterations of the algorithm PSO. This result is considered acceptable within the characteristics of this research conditions.*

Palabras clave: *Clasificación, ANFIS, PSO, IMC, RMSE, Obesidad.*

I. INTRODUCCIÓN

La identificación y tratamiento de la obesidad y el sobrepeso en la niñez, constituye un aspecto importante de la atención preventiva y la salud pública, que afectan a la mejoría de la salud, física, social y emocional de los niños y que puede repercutir en la edad adulta [13], para realizar el diagnóstico de sobrepeso y obesidad es necesario contar con indicadores antropométricos como son: edad, peso, estatura, IMC (índice de masa corporal), BF (grasa corporal), entre otros.

En la literatura revisada se ha encontrado sistemas de clasificación de la obesidad desarrollados con sistemas neurodifusos que combinan lógica difusa y redes neuronales; como son los trabajos desarrollados por (Uduak y Umoh) [11] y (khanna, Srinath y Mendiratta) [12], en estos trabajos se realiza un proceso de clasificación mediante el modelo ANFIS, los autores plantean conjuntos difusos y reglas difusas a partir de indicadores antropométricos para clasificación de la obesidad en adultos.

Pero hasta la fecha no se ha encontrado trabajos que busquen optimizar estos modelos neurodifusos con algún algoritmo evolutivo, que permita optimizar el error al clasificar la obesidad en niños y adolescentes; estos

algoritmos evolutivos como: algoritmos genéticos, colonia de hormigas, enjambres de partículas permitirían optimizar los resultados obtenidos en un sistema de clasificación de la obesidad.

En el presente trabajo se plantea la pregunta de que si la implementación de un sistema híbrido neurodifuso (ANFIS) con PSO (optimización por enjambre de partículas) optimizará el error de clasificación de la obesidad en niños y adolescentes.

Para responder la pregunta planteada se diseñó el sistema de clasificación de la obesidad en niños y adolescentes del sexo masculino entre seis a diecisiete años de edad de la región del Paraná Brasil; el sistema tiene como atributos de entrada a: edad, peso, estatura e IMC (índice de masa corporal); y como salida la clasificación de: normal, sobrepeso y obesidad. El entrenamiento y pruebas de la clasificación se llevan a cabo con el modelo neurodifuso ANFIS (Artificial Neural Network Fuzzy Inference System) de Matlab r2015a, mediante la función `genfis3`, con un modelo difuso tipo sugeno, una técnica de clusterización FCM (fuzzy c-means clustering), con un aprendizaje no supervisado; el FIS, creado a partir de ANFIS es evaluado con la finalidad de optimizar el error cuadrático medio obtenido en la clasificación mediante el uso del algoritmo PSO.

El artículo está estructurado de la siguiente manera: un breve resumen del artículo, en la sección uno está conformado por la introducción, la sección dos describe el estado de arte consultado, la sección tres detalla la metodología seguida en el trabajo, en la sección cuatro se muestran los resultados obtenidos, en la sección cinco una evaluación y discusión y la sección seis conclusiones a las que se llegaron.

II. ESTADO DEL ARTE

Dentro del estado del arte consultado se tiene una descripción de artículos consultados sobre teoría e investigaciones relacionados con el tema propuesto en este artículo que son descritos en los siguientes párrafos.

En [1] se menciona que los sistemas neurodifusos forman parte de la computación flexible (soft-computing) que engloban un conjunto de técnicas que tienen en común la robustez en el manejo de información imprecisa e incierta que existe en los problemas relacionados con el mundo real, las técnicas de computación flexible pueden ser combinadas

para aprovechar sus ventajas individuales, ANFIS (Adaptive Neuro Fuzzy Inference System) es un método que permite crear la base de reglas de un sistema difuso, utilizando el algoritmo de entrenamiento de retropropagación a partir de la recopilación de datos de un proceso. Su arquitectura es funcionalmente equivalente a una base de reglas tipo Sugeno.

En [2] se menciona que la optimización por enjambre de partículas (PSO) es una metaheurística poblacional que ha sido exitosamente aplicada para resolver problemas de optimización. Está inspirado en el comportamiento social del vuelo de bandadas de aves o el movimiento de los bancos de peces. El algoritmo PSO fue desarrollado por Kennedy y Eberhart basado en un enfoque de la metáfora social [3], y se fundamenta en los factores que influyen en la toma de decisión de una partícula que forma parte de un conjunto de partículas similares. La decisión de cada partícula se realiza conforme a un componente social y un componente individual, mediante los que se determina el movimiento de esta partícula para alcanzar una nueva posición en el espacio de soluciones. La metaheurística intenta simular este comportamiento para resolver problemas de optimización.

La clusterización puede definirse como el proceso de agrupar un conjunto de objetos abstractos o físicos en clases similares [2]. La clusterización es una técnica de aprendizaje no supervisado [19], un buen método de clusterización debe identificar cluster que sean tan compactos como separados entre sí, es decir que tengan alta similaridad intra-cluster y baja similaridad inter-cluster [18]. Los métodos de clusterización utilizadas en ANFIS son: *Subtractive Clustering* y *fuzzy c-means clustering* (FCM). *Subtractive Clustering* es un algoritmo rápido, de un solo paso para estimar el número de grupos y de los centros de los conglomerados en un conjunto de datos, se implementa mediante la función *subclust*, la función *genfis2* utiliza este método para la generación un sistema de inferencia difusa (FIS) [13]. FCM es un método de agrupación desarrollado por Dunn en 1973 y mejorado por Bezdek en 1981, este método permite determinar la pertenencia de un dato a un cluster, a partir de su grado de pertenencia a cada uno de los clusters predefinidos y de la distancia del dato a cada uno de los centros de los clusters, mediante una función de optimización; este método de clusterización es usado por *genfis3* para generar un sistema de inferencia difusa (FIS) [14][4].

En el trabajo presentado por (Franco) [5] nos presenta una optimización por enjambre de partículas (PSO) fue utilizado para resolver el problema de asignación axial 3-dimensional. El algoritmo PSO se utiliza por lo general para optimizar problemas de naturaleza continua, el problema axial 3-d es un problema de naturaleza discreta de tipo NP-duro, los operadores clásicos de PSO fueron redefinidos y un algoritmo en paralelo fue implementado. El algoritmo propuesto se aplicó a un conjunto de problemas de referencia y los resultados fueron comparados con los existentes de otros algoritmos que solucionan 3AP. La evaluación muestra que el algoritmo propuesto es competitivo.

En el artículo presentado por (Fazel et al.) [6] nos presenta un sistema adaptativo neurodifuso (ANFIS) entrenado por el algoritmo de optimización con enjambre de partículas (PSO) para la predicción del precio de las acciones, se implementan cuatro estrategias diferentes de PSO: *gbest*, *lbest-a*, *lbest-b* y euclidiana, Los resultados muestran la superioridad del modelo propuesto, en comparación con los modelos disponibles en términos de minimización de errores, robustez y flexibilidad.

En el trabajo de (Salazar) [7] propone una optimización estructural con un nuevo algoritmo híbrido auto-configurado a partir de dos algoritmos como son: optimización por enjambre de partículas (PSO) y un algoritmo genético (GA). Con esta estrategia se pretende mejorar la confiabilidad del procedimiento. Este nuevo algoritmo combina las características de búsqueda del PSO para mejorar su capacidad de localizar puntos óptimos.

En el artículo presentado por (Seydi, Aliyary y Teshnehlad) [8], propone un entrenamiento de ANFIS con un algoritmo PSO modificado. El modelo propuesto es aplicado para la identificación de un sistema dinámico no lineal. ANFIS utiliza el método de mínimos cuadrados para calcular el error, El algoritmo PSO modificado, remueve la peor partícula del enjambre y lo reemplaza por dos partículas que se generan por un operador de cruce (crossover) a partir de dos partículas una seleccionada de manera aleatoria y la otra se selecciona por la característica de ser la peor mejor local de su generación. El Algoritmo PSO modificado, logro mejorar el error del FIS (sistema de inferencia difusa) original. De este artículo se tomó la idea de mejorar el error encontrado en un FIS, utilizando el algoritmo PSO para mejorar el error de clasificación tanto para los datos de entrenamiento como de prueba.

En el artículo presentado por (Kothandaraman y Ponnusamy) en [9], presenta un estudio de afinamiento de un controlador neurodifuso de tipo ANFIS mediante el algoritmo PSO, para un sistema de suspensión vehicular. Se utiliza el método de control LQR (regulador lineal-cuadrático) es utilizado para obtener los datos de entrenamiento, se utiliza la técnica *Subtractive clustering* para formular el ANFIS, con aproximaciones de la fuerza de salida del actuador como una función de estados. PSO busca optimizar los radios para *subtractive clustering*; El entrenamiento se realiza fuera de línea y la función de coste se basa en la minimización del error real y salida aproximada. Del artículo mencionado se toma la idea de optimizar el error del FIS para la clasificación de la obesidad mediante PSO, la función de coste a optimizar será el error cuadrático medio.

Según (Vanegas et. al.) [20] encontró que para las funciones analizadas el algoritmo PSO fue significativamente más veloz, en comparación con otra técnica de optimización. Por eso motivo y según la revisión de los artículos anteriores, se eligió PSO como algoritmo de optimización ya que se ajusta a la naturaleza de estudio.

III. METODOLOGIA

El estudio busca clasificar la obesidad en 2938 escolares del sexo masculino entre 6 a 17 años de edad de la región del Paraná Brasil.

Mediante un sistema neurodifuso (ANFIS), cuyo error cuadrático medio (MSE) será optimizado mediante el algoritmo de optimización por enjambre de partículas (PSO).

El sistema será realizado en Matlab R2015a.

La metodología seguida para el siguiente trabajo está conformada por las etapas mostradas en la Figura 1.

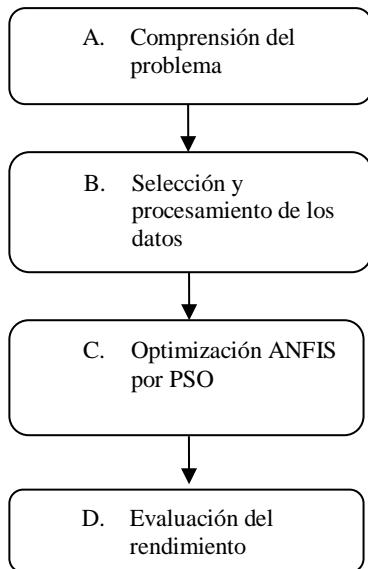


Fig. 1. Metodología propuesta para la clasificación de obesidad

A. Comprensión del problema

La clasificación de la obesidad se realizara sobre una base de datos de 2938 registros, cada registro posee cuatro atributos de entrada al sistema y un atributo de salida que es Obesidad, como se muestra en la Tabla I.

TABLA I. ATRIBUTOS DE ENTRADAY DE SALIDA

Edad (años)	peso (kg)	Estatura (cm)	IMC	Obesidad
-------------	-----------	---------------	-----	----------

La etiqueta Edad va desde los 6 años hasta los 17 años y Obesidad toma tres valores de salida: normal (valor 0), sobrepeso (50) y obesidad (100).

El trabajo de investigación fue implementado en Matlab r2015a utilizando la herramienta genfis3, se tomó como base el algoritmo de PSO de (Mostapha Kalami Heris) [10], al cual se realizó modificaciones para adaptarlo a la naturaleza del problema tratado en este artículo.

B. Selección y procesamiento de los datos

Para facilitar el trabajo los 2938 registros se almacenaron en un archivo (.mat), por medio de un script (.m) encargado de convertir datos a este formato, este archivo (.mat) es una estructura propia de matlab que sirve

para almacenar datos, la cual internamente consta de 2 Matrices: Matriz de Entradas y Matriz de Salidas.

La Matriz de Entradas consta de Edad, Peso, Estatura e IMC y la Matriz de Salidas consta de Obesidad.

Para el balanceo primero los registros se desordenaron aleatoriamente, luego se escogió de entre estos el 70% para el entrenamiento y el 30% para pruebas.

Este 70% se dividió en: “Entradas para entrenamiento” y “Salidas para entrenamiento”; y el 30% en: “Entradas para prueba” y “Salidas para prueba”.

C. Optimización ANFIS por PSO

La red ANFIS [14] posee para su implementación la aplicación *Neuro-Fuzzy Designer* de Matlab, esta se muestra en la Figura 2.

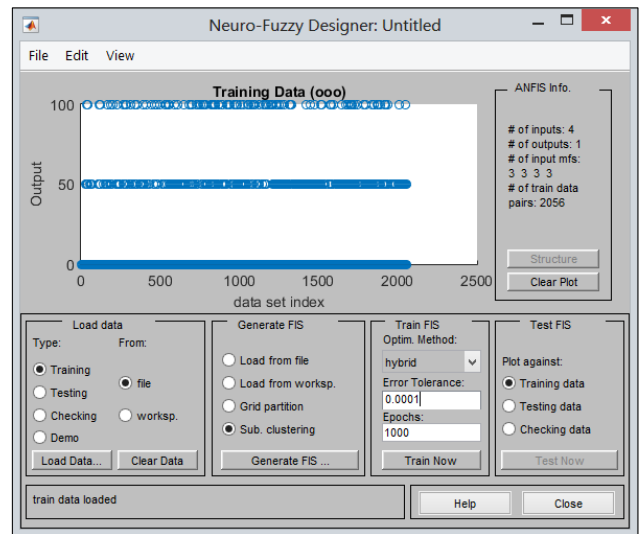


Fig. 2. Red ANFIS implementada con Neuro-Fuzzy Designer

Se observa que consta de 4 partes bien definidas como son:

- Load Data (Carga de Datos)
- Generate FIS (Generación del FIS)
- Train FIS (Entrenamiento del FIS)
- Test FIS (Pruebas y Resultados del FIS)

Teniendo como base estos pasos o etapas se desarrolló un Sistema Híbrido optimizado por PSO que trata de mejorar este proceso, el cual se observa en la Figura 3. Se pasa a detallar en el siguiente apartado.



Fig. 3. Sistema Híbrido optimizado por PSO

1) Carga de Datos

A diferencia de la herramienta *Neuro-fuzzy Designer* este proceso se mejora al incorporarle características de balanceo de los datos.

En la Figura 4. Se observa el proceso de Balanceo de los datos descrito en el apartado B. Además los resultados globales de este proceso se observan en la Tabla II.

Cabe acotar que la separación entre Entradas y Salidas se debe a que facilita el proceso de fusificación o Generación del Fis del apartado siguiente.

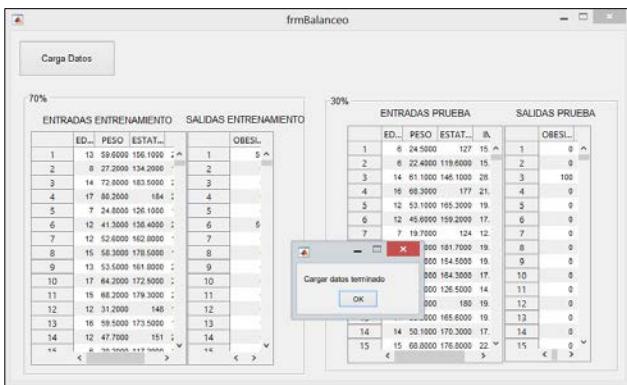


Fig. 4. Balanceo de los datos

TABLA II. RESULTADOS GLOBALES DEL BALANCEO DE LOS DATOS

	Entrenamiento	Prueba
Entradas	2057	2057
Salidas	881	881
TOTAL	2938	2938

2) Generación del FIS

Para la generación del FIS se optó por usar el *fuzzy c-means clustering* (FCM), ya que este permite generar un modelo más adaptable al problema de la clasificación de la obesidad, como se sabe este problema presenta distintos rangos para cada edad específica, es decir un niño de 6 años no se le puede tratar de igual manera que un adolescente de 17 años, debido a la diferencia en los percentiles de los indicadores antropométricos (peso, estatura e IMC) con la edad.

Una alternativa fue crear 12 modelos FIS de tipo sugeno con funciones de pertenencia trapezoidal y salida lineal, para registros agrupados por edad.

Se vio la manera de mejorar esta creación de 12 modelos FIS, para ello se optó por una clasificación difusa no supervisada que con un solo modelo FIS que abarque todas las edades consideradas, es decir una clusterización difusa que infiriera automáticamente las funciones de pertenencia necesarias que puedan regir tanto para niños como para adolescentes.

La solución fue utilizar el modelo de fusificación *Fuzzy c-means* por medio de la función *Genfis3* propia de Matlab.

Genfis3 genera un sistema de inferencia difusa (FIS) a partir de datos proporcionados previamente, utiliza la clusterización fuzzy c-means (FCM), la cual mediante la extracción de un conjunto de reglas modela el comportamiento de los datos. La función requiere conjuntos separados de datos de entrada y de salida.

Cuando sólo hay una salida, se puede utilizar *genfis3* para generar una FIS inicial para la formación del ANFIS. El método de extracción utiliza primero la función FCM para determinar el número de reglas y funciones de pertenencia de los antecedentes y consecuentes.

En la ecuación (1) se observa los parámetros que toma el *genfis3*, *Xin* constituye la matriz de los datos de entrada, *Xout* constituye la matriz de los datos de salida, *type* especifica el tipo (sugeno o mamdani), *cluster_n* especifica el número de grupos y en *fcmoptions* se detalla especificaciones de agrupamiento para el algoritmo FCM.

$$fismat = genfis3(Xin, Xout, type, cluster_n, fcmoptions) \quad (1)$$

El tipo de función de pertenencia de entrada es 'gaussmf'. Por defecto, el tipo de función de pertenencia de salida es 'linear'. Sin embargo, si se especifica *type* como 'mamdani', entonces el tipo de función de pertenencia de salida es 'gaussmf'.

Los parámetros para el *genfis3* se muestran en la Tabla III.

TABLA III. PARÁMETROS PARA EL GENFIS3

Parámetros	Valor
Numero Clusters	10
Numero de Reglas	10
Exponente U	2
Max. Número de Iteraciones	100
Cantidad mínima de mejora	1e-5

En la Figura 5. Se muestra la generación del modelo difuso FCM para 10 clusters por medio del llamado a la función *genfis3*.

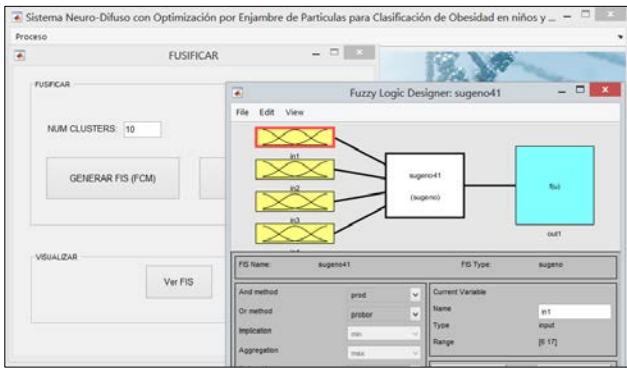


Fig. 5. Fuzzy c-means para 10 clusters.

En la Figura 6 se observa la generación de la función de pertenencia sugeno de tipo gaussiano para la variable de entrada Edad.

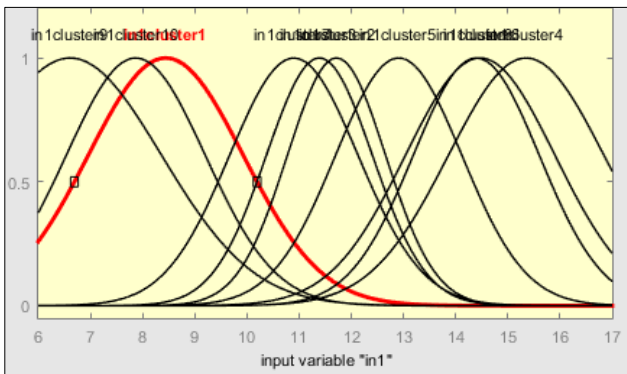


Fig. 6. Función de pertenencia para la variable Edad.

En la Figura 7 se observa la generación de la función de pertenencia sugeno de tipo gaussiano para la variable de entrada Peso.

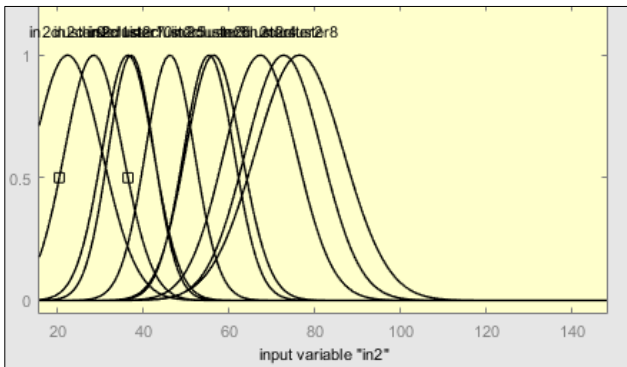


Fig. 7. Función de pertenencia para la variable Peso.

En la Figura 8 se observa la generación de la función de pertenencia sugeno de tipo gaussiano para la variable de entrada Estatura.

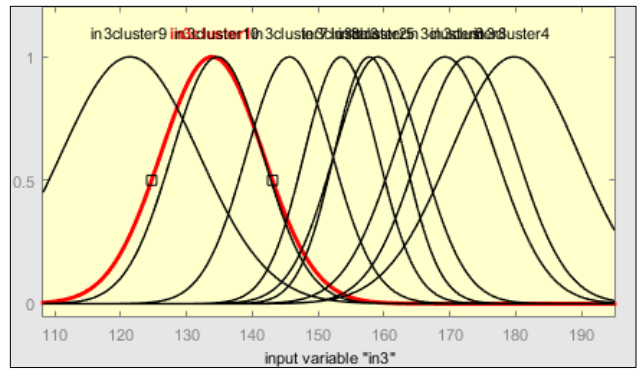


Fig. 8. Función de pertenencia para la variable Estatura.

En la Figura 9 se observa la generación de la función de pertenencia sugeno de tipo gaussiano para la variable de entrada IMC.

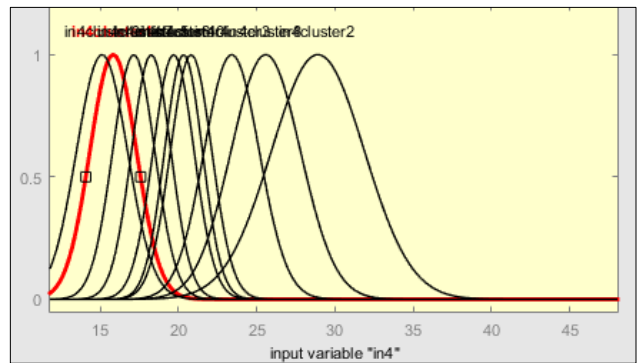


Fig. 9. Función de pertenencia para la variable IMC

Las 10 reglas de inferencia obtenidas se muestran en la Figura 10.

```

1. If (in1 is in1cluster1) and (in2 is in2cluster1) and (in3 is in3cluster1) and (in4 is in4cluster1) then (out1 is out1cluster1) (1)
2. If (in1 is in1cluster2) and (in2 is in2cluster2) and (in3 is in3cluster2) and (in4 is in4cluster2) then (out1 is out1cluster2) (1)
3. If (in1 is in1cluster3) and (in2 is in2cluster3) and (in3 is in3cluster3) and (in4 is in4cluster3) then (out1 is out1cluster3) (1)
4. If (in1 is in1cluster4) and (in2 is in2cluster4) and (in3 is in3cluster4) and (in4 is in4cluster4) then (out1 is out1cluster4) (1)
5. If (in1 is in1cluster5) and (in2 is in2cluster5) and (in3 is in3cluster5) and (in4 is in4cluster5) then (out1 is out1cluster5) (1)
6. If (in1 is in1cluster6) and (in2 is in2cluster6) and (in3 is in3cluster6) and (in4 is in4cluster6) then (out1 is out1cluster6) (1)
7. If (in1 is in1cluster7) and (in2 is in2cluster7) and (in3 is in3cluster7) and (in4 is in4cluster7) then (out1 is out1cluster7) (1)
8. If (in1 is in1cluster8) and (in2 is in2cluster8) and (in3 is in3cluster8) and (in4 is in4cluster8) then (out1 is out1cluster8) (1)
9. If (in1 is in1cluster9) and (in2 is in2cluster9) and (in3 is in3cluster9) and (in4 is in4cluster9) then (out1 is out1cluster9) (1)
10. If (in1 is in1cluster10) and (in2 is in2cluster10) and (in3 is in3cluster10) and (in4 is in4cluster10) then (out1 is out1cluster10) (1)

```

Fig. 10. Reglas de inferencia.

3) Optimización del entrenamiento ANFIS con PSO.

Para la implementación del PSO se hizo uso del algoritmo implementado en [10].

El algoritmo PSO se describe como: los individuos que conviven en una sociedad tienen una opinión que es parte de un conjunto de creencias (el espacio de búsqueda) compartido por todos los posibles individuos. Cada individuo puede modificar su propia opinión basándose en tres factores:

- Su conocimiento sobre el entorno (su valor del *fitness*).
- Su conocimiento histórico o experiencias anteriores (su memoria).
- El conocimiento histórico o experiencias de los individuos situados en su vecindario.

Un algoritmo de PSO mantiene un cúmulo de partículas, donde cada partícula representa una solución al problema. Las partículas vuelan a través de un espacio de búsqueda multidimensional, donde la posición de cada partícula se ajusta de acuerdo a su propia experiencia y a la experiencia de sus vecinos.

PSO se inicializa con un grupo de partículas al azar (soluciones) y después busca óptimas mediante la actualización de iteraciones. En cada iteración, cada partícula es actualizado por los siguientes dos valores "mejores". La primera de ellas es la mejor solución (fitness) que se ha logrado hasta ahora y está representado como $Pbest$. El Otro mejor valor es la mejor solución obtenido hasta el momento por cualquier partícula en la población, esto es representado como $Gbest$. Cada partícula conoce el mejor valor hasta la fecha ($Pbest$) y el mejor valor en el grupo ($Gbest$). Las partículas intentan modificar su posición usando la velocidad actual y la distancia, desde $Pbest$ a $Gbest$. La velocidad ajustada y la posición ajustada de cada partícula pueden ser calculadas utilizando las fórmulas de las ecuaciones (2) y (3).

$$V_i^{t+1} = w^t V_i^t + c_1 r_1^t (Pbest_i - X_i^t) + c_2 r_2^t (Gbest_i - X_i^t) \quad (2)$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \quad (3)$$

Donde:

V_i^{t+1} Velocidad ajustada

W^t Inercia del propio movimiento

c_1 : Coeficiente de confianza en la experiencia

c_2 : Coeficiente de confianza en la experiencia del grupo

$Pbest_i$: Mejor posición previa de i

X_i^t : Posición actual de i

$Gbest_i$: Mejor posición previa encontrada por el grupo

r_1^t y r_2^t : Operadores aleatorios entre 0 y 1

X_i^{t+1} : Posición de la partícula i después del ajuste

De manera más detallada, el método de PSO se describe como sigue:

- Inicializar la población. La posición de cada una de las partículas es determinada de manera aleatoria.
- La mejor posición previa es igualada a la posición actual.
- Cada posición es evaluada en la función de aptitud para determinar la calidad de la solución.
- Se compara la aptitud de la posición actual con la mejor previa.
- Asignar informantes (vecindario) de tamaño k a la partícula.
- Determinar la mejor partícula del vecindario.
- Ajustar la velocidad.
- Ajustar la posición.
- Verificar si se cumple el criterio de parada.
- Si no se cumple, regresar al paso e.

Los valores que toma PSO para una prueba inicial de optimización del entrenamiento se detallan en la Tabla IV. Cabe destacar que más pruebas serán realizadas en la sección "Resultados" para llegar a dilucidar conclusiones finales.

TABLA IV. PARÁMETROS PARA EL PSO

Parámetros	Valor
Numero Iteraciones	100
Población	25
Peso de Inercia	1
Factor de amortiguamiento	0.99
Coefficiente entrenamiento personal C1	1
Coefficiente entrenamiento general C2	2

En la Figura 11. Se observa la optimización de entrenamiento para ANFIS utilizando PSO.

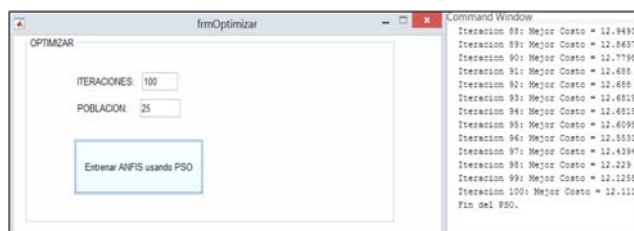


Fig. 11. Optimización del entrenamiento

D. Evaluación del rendimiento

Para la evaluación del rendimiento se utilizó el error cuadrático medio (MSE). El error cuadrático medio (MSE) de un estimador mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. El MSE es una función de riesgo, correspondiente al valor esperado de la pérdida del error al cuadrado o pérdida cuadrática. La diferencia se produce debido a la aleatoriedad o porque el estimador no tiene en cuenta la información que podría producir una estimación más precisa [15].

El MSE es el segundo momento (sobre el origen) del error, y por lo tanto incorpora tanto la varianza del estimador así como su sesgo. Para un estimador insesgado, el MSE es la varianza del estimador. Al igual que la varianza, el MSE tiene las mismas unidades de medida que el cuadrado de la cantidad que se estima. En una analogía con la desviación estándar, tomando la raíz cuadrada del MSE produce el error de la raíz cuadrada de la media o la desviación de la raíz cuadrada media (RMSE o RMSD), que tiene las mismas unidades que la cantidad que se estima; para un estimador insesgado, el RMSE es la raíz cuadrada de la varianza, conocida como la desviación estándar.

IV. RESULTADOS

En la Tabla V. Se muestran los resultados para el entrenamiento considerando una población de 25 y a diferentes niveles de iteraciones:

TABLA V. RESULTADOS DEL ENTRENAMIENTO

	ITERACIONES			
	100	200	500	1000
Costo	12.42	10.06	8.31	8.6
MSE	154.37	101.21	69.03	73.96
RMSE	12.42	10.06	8.31	8.6
Error Medio	1.43	-0.03	0.02	-0.17
Error St. D.	12.34	10.06	8.31	8.60

En la Figura 12. Se muestra la comparación entre la salida (color negro) y los resultados esperados (color rojo) para el entrenamiento con el mejor resultado obtenido, que fue 500 iteraciones.

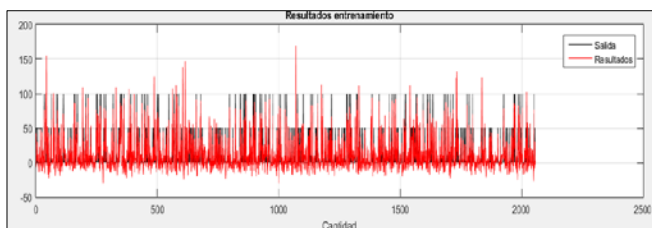


Fig. 12. Resultado de entrenamiento tras 500 iteraciones

En la Tabla VI. Se muestran los resultados para las pruebas considerando una población de 25 y a diferentes niveles de iteraciones:

TABLA VI. RESULTADOS DE LAS PRUEBAS

	ITERACIONES			
	100	200	500	1000
Costo	12.42	10.06	8.31	8.6
MSE	155.49	101.68	69.23	71.25
RMSE	12.47	10.08	8.32	8.44
Error Medio	1.18	0.16	0.06	0.076
Error St. D.	12.42	10.09	8.32	8.45

En la Figura 13. Se muestra la comparación entre la salida (color negro) y los resultados esperados (color rojo) para las prueba con el mejor resultado obtenido, que fue 500 iteraciones.

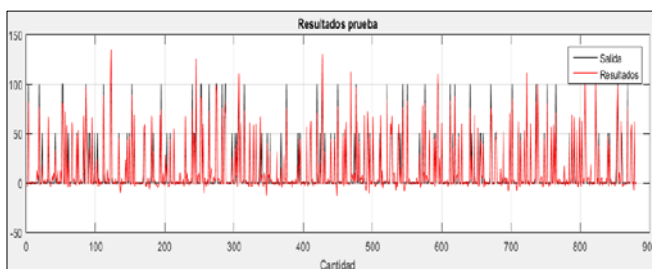


Fig. 13. Resultado de las Pruebas tras 500 iteraciones

V. EVALUACIÓN Y DISCUSIÓN

En base a los resultados obtenidos en las Tablas V y VI, se observa que el error así como el costo disminuye a un mayor número de iteraciones, más específicamente cuando llega a 500 iteraciones, en cambio sí se sigue aumentando el número de iteraciones se ve que los resultados decaen, porque se considera que ha ocurrido un sobre-entrenamiento

quedando ajustado a unas características muy específicas de los datos de entrenamiento que no tienen relación causal con la función objetivo.

Se propone mejorar el entrenamiento PSO aumentando el tamaño del enjambre así como paralelizando el algoritmo.

VI. CONCLUSIONES

La implementación del Fuzzy c-means hizo más manejable abordar el problema de la clasificación de la Obesidad.

La optimización de la Red ANFIS por medio de Optimización por Enjambre de Partículas PSO dio resultados aceptables dentro de las condiciones propias de esta investigación.

AGRADECIMIENTOS

Se agradece a la Red Iberoamericana de investigación en el Desarrollo Biológico Humano por los datos proporcionados.

REFERENCIAS

- [1] J. C. Chahuara Quispe, «Control Neuro-Difuso Aplicado a una Grúa Torre.,» Oficina General del Sistema de Bibliotecas y Biblioteca Central de la UNMSM, Lima, 2005.
- [2] A. Villagra, «Metaheurísticas aplicadas a Clustering,» Argentina.
- [3] J. Kennedy , R. C. Eberhart y Y. Shi, «Swarm Intelligence,» Academic Press, San Diego, 2001 .
- [4] «A Tutorial on Clustering Algorithms,» 2016. [En línea]. Available: http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/
- [5] J. A. Franco Gómez, «Un Algoritmo basado en la Optimización por Enjambre de Partículas para el problema de Asignación Axial 3-Dimensional,» La Paz, Baja California Sur, 2011.
- [6] M. H. Fazel Zarandi, M. Avazbeigi y M. Alizadeh, «A Neuro-Fuzzy Expert System Trained by Particle Swarm Optimization for Stock Price Prediction,» IGI Global, 2012.
- [7] P. Y. Salazar Pinto, «Algoritmo Híbrido Auto Configurado para Optimización Estructural,» Bucaramanga, 2009.
- [8] Ghomsheh, V. Seydi, M. Aliyari Shoorehdeli, and M. Teshnehlab. "Training ANFIS structure with modified PSO algorithm." *Control & Automation, 2007. MED'07. Mediterranean Conference on. IEEE*, 2007.
- [9] R. Kothandaraman y L. Ponnusamy , «PSO tuned Adaptive Neuro-fuzzy Controller for Vehicle Suspension Systems,» *JOURNAL OF ADVANCES IN INFORMATION TECHNOLOGY*, vol. 3, nº 1, 2012.
- [10] M. Kalami Heris , «Evolutionary ANFIS Training in MATLAB,» yarpiz, 2015. [En línea]. Available: <http://yarpiz.com/319/ypfz104-evolutionary-anfis-training>. [Último acceso: 2016].
- [11] U. A. Umoh y E. Isong, «Design Methodology of Fuzzy Expert System for the Diagnosis and Control of Obesity,» *Computer Engineering and Intelligent Systems*, vol. 6, nº 9, pp. 12 - 32, 2015.

- [12] M. KHANNA, N. K. SRINATH y J. K. MENDIRATTA, «The Study of Obesity in Children Using Fuzzy Logic,» *INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH*, vol. 3, n° 1, p. 1833 – 1836, December-January 2015.
- [13] C. Ureña Trigueros, «CONCEPTOS BÁSICOS DE OBESIDAD EN LA NIÑEZ Y LA ADOLESCENCIA: (Definición, Diagnóstico, Epidemiología, Factores de Riesgo, Complicaciones),» *REVISTA MEDICA DE COSTA RICA Y CENTROAMERICA*, vol. LXV , n° 582, pp. 31-35, 2008.
- [14] O. Lutfy y E. T. Khduir, «A PID-like ANFIS Controller Trained by PSO Technique to Control Nonlinear MIMO Systems,» *Australian Journal of Basic and Applied Sciences*, 2013.
- [15] E. Lehmann y G. Casella, *Theory of Point Estimation*, New York: Springer, 1998.
- [16] MathWorks, «Subclust - Documentación Matlab R2015a,» 2016. [En línea]. Available: http://www.mathworks.com/help/fuzzy/subclust.html?s_tid=gn_loc_drop.
- [17] MathWorks, «Fcm - Documentación de Matlab R2015a,» MathWorks, 2016. [En línea]. Available: <http://www.mathworks.com/help/fuzzy/fcm.html>.
- [18] S. Schiaffino, «Clustering,» 2015. [En línea]. Available: http://www.exa.unicen.edu.ar/catedras/optia/public_html/clustering.pdf.
- [19] CIOS, Krzysztof J., et al. *Unsupervised learning: Clustering*. En *Data Mining*. Springer US, 2007. p. 257-288.
- [20] Vanegas-Ardila, Diego, Karol Sebastián Barragán-Niño, and Rodrigo Correa-Cely. "Comparación de las técnicas de optimización por análisis de intervalos y la de enjambre de partículas para funciones con restricciones." *Ingeniería y Universidad* 15.1 (2011): 47.