

COMPARACIÓN DE ALGORITMOS EVOLUTIVOS PARA LA OPTIMIZACIÓN EN LA CLASIFICACIÓN DE LA OBESIDAD EN ESCOLARES

COMPARISON OF EVOLUTIVE ALGORITHMS FOR OPTIMIZATION IN THE CLASSIFICATION OF OBESITY IN SCHOOLS

ANA CECILIA ARENAS RODRÍGUEZ
CARLOS ALBERTO TORRES NAIRA
FLOR DE MARÍA VIZCARRA HUYHUA
JOSÉ SULLA TORRES

jsulla@unsa.edu.pe

Departamento de Ingeniería de Sistemas e Informática.

Universidad Nacional de San Agustín, Arequipa, Perú.

JORGE MÉNDEZ CORNEJO

Departamento de Ciencias de la Actividad Física

Universidad Católica del Maule. Talca, Chile.

RESUMEN

El estudio tiene por objetivo comparar tres tipos de algoritmos evolutivos diferentes: Real Encoding Particle Swarm Optimization (REPSO-C), Incremental Learning with Genetic Algorithms (ILGA) y Decision Tree with Genetic Algorithm (DT-GA), para determinar el porcentaje de mejora durante cada iteración que tiene cada uno de los algoritmos sobre una base de datos relacionadas con la obesidad escolar. Se utilizó la herramienta denominada Keel, que permitió cargar los datos, prepararlos, procesarlos y analizar los resultados, tanto en el entrenamiento, como en las pruebas realizadas. La precisión encontrada por la ejecución de cada uno de los algoritmos, permite concluir que el modelo que utiliza árbol de decisión y algoritmo genético DT-GA es el mejor, debido al nivel de precisión que posee.

Palabras clave: algoritmo evolutivo, REPSO-G, ILGA, DT-GA.

ABSTRACT

The aim of the study is to compare three different evolutionary algorithms: Real Encoding Particle Swarm Optimization (REPSO-C), Incremental Learning with Genetic Algorithms (ILGA) and Decision Tree with Genetic Algorithm (DT-GA) to determine the percent improvement during each iteration of each of the algorithms on a database related to school obesity. For this purpose, the Keel tool was used, which allowed data loading, preparation, processing and analysis of results, both in training and in the tests performed. The precision found by the execution of each of the algorithms, allows to conclude that the model using decision tree and genetic algorithm DT-GA is the best one due to the level of precision that it possesses.

Key words: evolutionary algorithm, REPSO-G, ILGA, DT-GA.

1. INTRODUCCIÓN

Los problemas relacionados con la cantidad de grasa en la composición del cuerpo humano actualmente están aumentando significativamente. Esto puede estar asociado a problemas de sedentarismo y hábitos alimenticios. Esta situación afecta a la salud mundial de niños, jóvenes y adultos (OMS, 2016). De hecho, el problema de la obesidad es un asunto de gran preocupación para el individuo y las autoridades gubernamentales (Miyahira & Araujo, 2008), los que son discutidos como temáticas de prioridad.

Actualmente, la minería de datos se aplica en muchos campos de la salud, se alcanza una alta precisión en la predicción de distintos tipos de cáncer utilizando un aprendizaje automático incremental (Nayyeri & Sharifi Noghabi, 2016). Algunas propuestas han sugerido aplicar técnicas de minería de datos para la clasificación de la obesidad infantil (Abdullah et al., 2016). Su aplicación y exactitud depende mucho del conjunto de datos, así como la relación entre los atributos (Zhang et al., 2009).

Actualmente no se cuenta con un sistema que permita predecir la obesidad en las personas en edad escolar de forma precisa, y mucho menos se conoce de un sistema que haga la optimización de un árbol de decisión difuso.

Ante esta situación surge el siguiente cuestionamiento: ¿cuál será el modelo que permite identificar, con mayor precisión, la obesidad en niños y adolescentes?

En ese sentido, el uso de algoritmos evolutivos de inteligencia artificial en un árbol de decisión difuso puede mejorar los datos en la predicción de la obesidad en personas de edad escolar.

Bajo este contexto, el presente estudio tiene como objetivo comparar la aplicación de algoritmos evolutivos para optimizar la clasificación de obesidad escolar, que permitan a los datos ser mejorados de acuerdo a la cantidad de iteraciones realizadas y tomando en cuenta el mejoramiento según los resultados alcanzados.

2. METODOLOGÍA

Tipo de estudio y muestra

Este estudio es descriptivo (transversal). Está basado en datos de niños y adolescentes entre cinco y diecisiete años de la región de Itaipú-Paraná (Brasil). Esta información fue proporcionada por la Red Iberoamericana de Investigación en Desarrollo Biológico Humano. Las variables iniciales (base de datos) se encontraban distribuidas de la siguiente manera: Estatura (m), Peso (kg), Índice de Masa Corporal ($IMC = \text{peso} / \text{estatura}^2$). Producto del uso del IMC, se categorizó a los alumnos en normal, sobrepeso y obesidad. Se utilizaron los puntos de corte de Cole (Cole, Bellizzi, Flegal, & Dietz, 2000).

Para trabajar con la base de datos se utilizó como referencia la Tabla 1. Los datos con los que trabajó para fusificar se encontraban divididos según la Tabla 2:

Tabla 1

Distribución original de los datos.

Categoría	Niños	Niñas
Normal	2331	2429
Sobrepeso	436	459
Obesidad	171	136
Total	2938	3024

Los datos con los que trabajamos para fusificar se encontraban divididos según la siguiente tabla:

Tabla 2

Distribución de los datos seleccionados para el estudio.

Categoría	Niños	Niñas
Normal	980	1078
Sobrepeso	211	206
Obesidad	88	69
Total	1279	1353

Procedimientos

Para realizar la incorporación del algoritmo genético en los datos fusificados que fueron obtenidos anteriormente, se hizo uso de la clasificación de Medsker et al. (Medsker & Bailey, 1992). Se definieron 5 modos de integración: el modelo que se tomó como mejor idea para la realización de nuestro planteamiento es el modelo transformacional.

- Modelos independientes
- Modelos transformacionales
- Modelos ligeramente acoplados
- Modelos fuertemente acoplados
- Modelos totalmente integrados

Modelo transformacional

En este caso se empieza a resolver el problema usando un modelo inteligente y se termina empleando otro algoritmo (Medsker & Bailey, 1992), de modo que era el caso del estudio. Se optó por empezar la resolución de la predicción de la obesidad en personas en edad escolar con un árbol de decisión difuso, para que una vez obtenidos los datos fusificados, los mismos sean nuevamente procesados mediante el algoritmo genético. La Figura 1 muestra el modelo.



Figura 1. Modelo transformacional. Fuente: Elaboración propia.

Herramienta

La herramienta que se usó fue *Keel* (Alcalá-Fdez et al., 2004). Esto permite seleccionar los datos de entrada, procesarlos de acuerdo a los objetivos de estudio. También muestra el porcentaje de precisión del entrenamiento, como de las pruebas.

Para poder hacer uso de la herramienta, se requiere primero realizar la carga de los datos que se quieren procesar. Esto debe estar formateado para que puedan ser reconocidos, descartando los datos erróneos, vacíos, para realizar el experimento.

Luego se selecciona el algoritmo que sea más efectivo de acuerdo a los datos que se tengan. Al final se muestra el porcentaje de precisión.

Algoritmo evolutivo

Este término es empleado para describir sistemas de resolución de problemas de optimización o búsqueda basados en el ordenador, empleando modelos computacionales de algún mecanismo de evolución conocido como elemento clave en su diseño e implementación. Los algoritmos evolutivos trabajan con una población de individuos que representan soluciones candidatas a un problema. Esta población se somete a ciertas transformaciones y después a un proceso de selección. Esto favorece a los mejores.

Cada ciclo de transformación y selección constituye una generación de forma que, después de cierto número de generaciones, se espera que el mejor individuo de la población esté cerca de la solución buscada. Los algoritmos evolutivos combinan la búsqueda aleatoria, dada por las transformaciones de la población, según la selección (Andaluz, 2011).

Principales Componentes:

- Población de individuos, que son una representación (no necesariamente directa) de posibles soluciones.
- Procedimiento de selección basado en la aptitud de los individuos para resolver el problema.
- Procedimiento de transformación para construir nuevos individuos a partir de los anteriores. La Figura 2 muestra la estructura general del algoritmo.

```
t :=0;
Inicializar P(t);
Evaluar P(t)
While no se termine do
begin
    P'(t) := variación [P(t)];
    evaluar [P'(t)];
    P(t+1) := seleccionar [P'(t)∪Q];
    t :=t+1;
end
```

Figura 2. Estructura general del algoritmo evolutivo (Andaluz, 2011).

Características

La característica fundamental de los algoritmos evolutivos radica en los métodos de generación de soluciones: se parte de un conjunto de soluciones iniciales y se van empleando un conjunto de operadores de búsqueda para ir refinando la solución final. Para realizar dicho refinamiento de las soluciones, se pueden utilizar técnicas clásicas, por ejemplo, el seguimiento del gradiente (Hill Climbing), complementadas con mecanismos biológicos de exploración: población de soluciones, operadores genéticos (Andaluz, 2011).

REPSO-C

El descubrimiento de reglas según la optimización del enjambre de partículas, nos dice que tenemos un conjunto de reglas, los cuales vienen conformados por **IF A THEN C**, donde A es un antecedente, la cual es una conjunción de condiciones, y C es una regla consecuente, la cual es una clase de predicción (Liu, Qin, Shi, & Chen, 2004).

La precisión de la clasificación de reglas está definida por la siguiente ecuación:

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (1)$$

Y donde nuestra cobertura estará determinada por la siguiente ecuación:

$$\text{Cobertura} = \frac{TP}{TP + FN} \quad (2)$$

Donde tenemos que:

TP = Número de ejemplos que satisfacen A y C.

FP = Número de ejemplos que satisfacen A pero no C.

FN = Número de ejemplos que no satisfacen A pero satisfacen C.

TN = Número de ejemplos que no satisfacen A ni C.

La comprensión de un conjunto de reglas a menudo se representa en una cantidad menor de condiciones en el antecedente de la regla y menor cantidad de reglas en el conjunto de reglas. En el proceso del descubrimiento de las reglas, usamos la siguiente ecuación:

$$\text{Succinctness} = \frac{\text{countAnt} - 1}{\text{attributedCount}} \quad (3)$$

Donde:

- CountAnt = Número de condiciones en la regla antecedente.
- AttributeCount = significa el número de atributos de decisión en el conjunto de datos.

De las anteriores ecuaciones, entonces se puede definir nuestra ecuación para la función de entrenamiento:

$$Fitness = w_1 * (Precisión = Cobertura) + w_2 * Succinctness + w_3 Interesting \quad (4)$$

Donde:

w_1 , w_2 , y w_3 = son constantes usadas para balancear los pesos de nuestras reglas en el proceso de su descubrimiento.

Interesting = es definido de acuerdo al criterio del usuario, sin embargo, en el presente artículo no será usado.

DT-GA (HYBRID DECISION TREE / GENETIC ALGORITHM)

Este método encuentra las reglas en dos fases de entrenamiento:

- En la primera, ejecuta el árbol de decisión C4.5. El resultado de este será un conjunto de reglas; estas son etiquetadas como una pequeña o gran disyuntiva dependiendo si la cantidad de ejemplos en cada regla es menor a un umbral dado.
- La segunda fase consiste en utilizar el algoritmo genético recorriendo el árbol hasta llegar a un nodo hoja; si pertenece a una de las grandes disyuntivas, este se asigna a una de las reglas de poca disyuntiva descubiertos por el algoritmo genético. El diagrama se puede observar en la Figura 3.

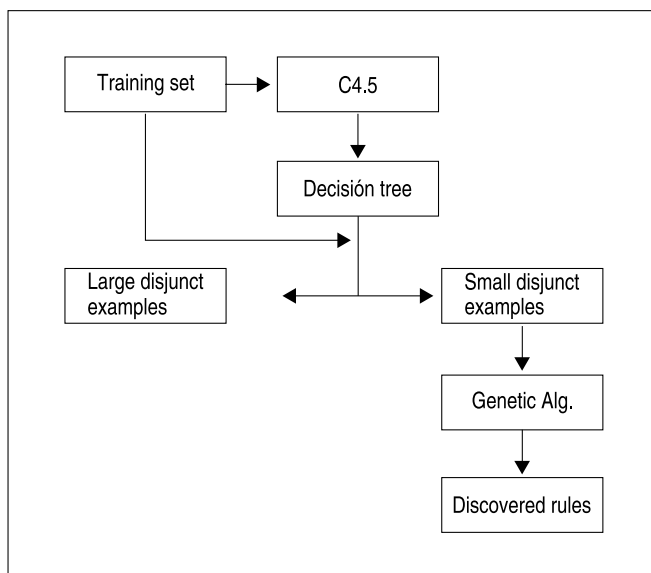


Figura 3. Diagrama del método híbrido DT-GA para el descubrimiento de pequeñas disyuntivas (Nayyeri & Sharifi Noghabi, 2016).

Las reglas del algoritmo genético se inicializan al azar seleccionando un valor para cada atributo de los ejemplos. La representación del cromosoma codifica las condiciones para todos los atributos (nominales y numéricos). La función de entrenamiento se da por una versión cuadrática de la media geométrica de las tasas reales. También utiliza un operador de poda que transforma una condición en un "no me importa", basado en la tasa de precisión asociado a cada atributo.

El criterio para parar está dado por un valor de umbral de los ejemplos que quedan en el "segundo conjunto de entrenamiento". La idea básica de este proceso se muestra en la Figura 4, donde los nodos hoja considerados como pequeñas disyuntivas son denotados con un cuadrado.

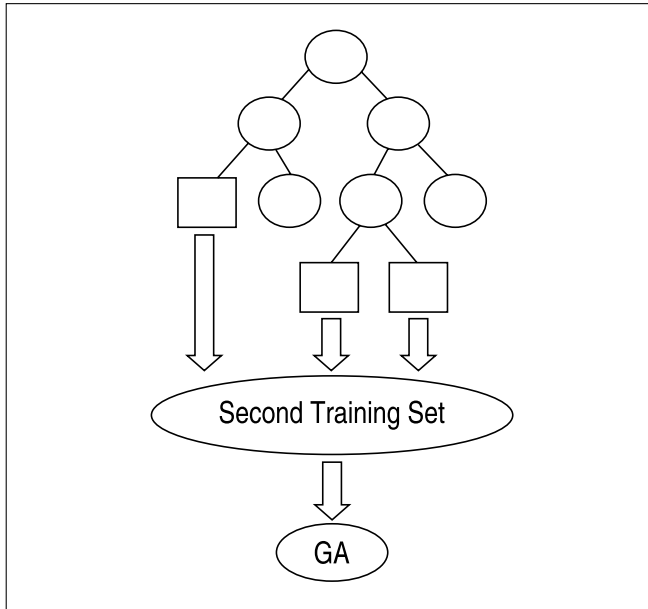


Figura 4. Construcción del segundo conjunto de entrenamiento para tomarlo como entrada al algoritmo genético (Carvalho & Freitas, 2004).

ILGA (INCREMENTAL LEARNING WITH GENETICS ALGORITHMS)

En algunos problemas de clasificación, nuevos datos de entrenamiento, atributos y clases pueden estar disponibles. También algunos elementos existentes pueden ser cambiados. Así, agentes clasificadores deben tener algoritmos para hacer frente a estos cambios. O bien, pueden detectar el entorno y evolucionar sus soluciones por sí mismos, o pueden colaborar para adaptarse al nuevo entorno, como se muestra a continuación en la Figura 5 (Guan & Zhu, 2005).

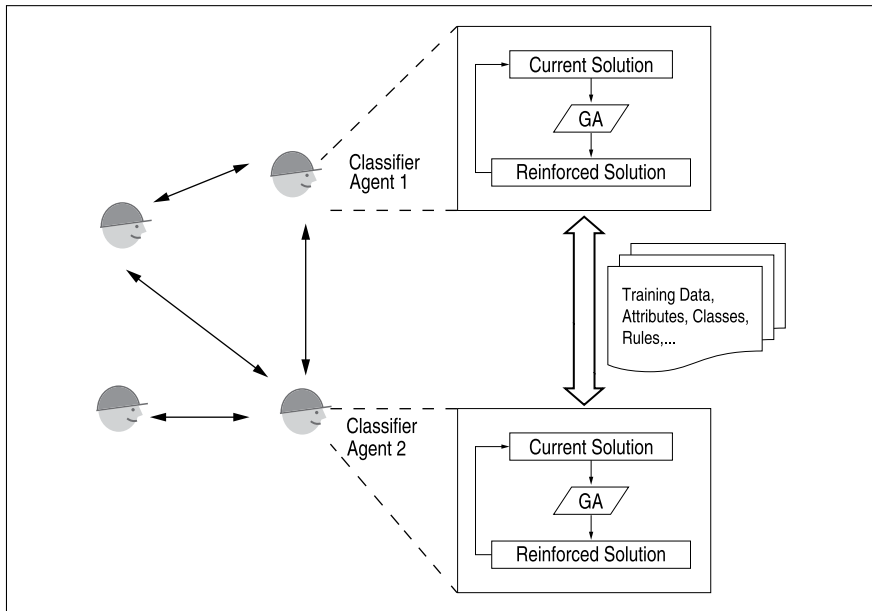


Figura 5. Aprendizaje Incremental de agentes clasificadores con AG (Guan & Zhu, 2005).

Los agentes clasificadores pueden intercambiar información de nuevos atributos y clases. Si está disponible, también pueden intercambiar conjuntos de reglas evolucionadas (cromosomas). Incluso, pueden proporcionar unos a otros nuevos datos de entrenamiento y prueba, o desafiar a otros agentes con problemas no resueltos. También son posibles varias combinaciones de estos modos de funcionamiento (Guan & Zhu, 2005). La Figura 5 muestra también el uso de algoritmos genéticos como el enfoque principal para el aprendizaje gradual, ya sea con el autoaprendizaje o aprendizaje colaborativo.

A continuación, se presenta un modelo para un problema de clasificación simplificados con dos atributos (x_1 , x_2) y dos clases (C1, C2). (Obsérvese Figura 6).

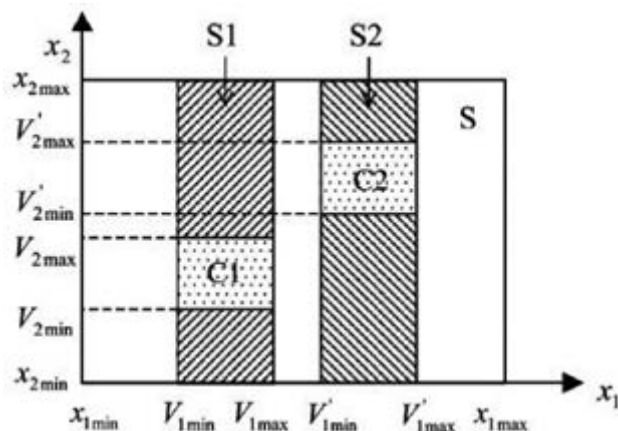


Figura 6. Modelo ILGA para un problema de clasificación simplificado.

Este modelo simplificado se puede ampliar fácilmente para espacios de dimensiones superiores. S indica la posible área confinada por el rango de valores (x_{1min}, x_{1max}) y (x_{2min}, x_{2max}) , que también es el área máxima de búsqueda para el algoritmo genético. $C1$ y $C2$ son las áreas que cubren los casos de entrenamiento pertenecientes a las dos categorías de clase respectivamente. ILGA consiste en dos pasos. El primer paso es una búsqueda unidimensional a lo largo del eje x_1 , tratando de encontrar la información límite para ambas clases, es decir, $V_{1min}, V_{1max}, V'_{1min}, V'_{1max}$. El segundo paso hereda esta información límite, y continúa buscando la información límite para los dos atributos (se busca más $V_{2min}, V_{2max}, V'_{2min}, V'_{2max}$) (Guan & Zhu, 2005).

3. RESULTADOS

REPSO-C

En los resultados obtenidos por el presente algoritmo podemos observar que a medida que las iteraciones aumentan, también aumentan el porcentaje de acierto en el entrenamiento y en la prueba. De modo que en lo que respecta a REPSO-C, tenemos que el total de instancias clasificadas fueron de 2356 escolares. En la Tabla 3 se muestra los resultados encontrados.

Tabla 3*Porcentaje de precisión alcanzado por REPSO-C en la penúltima y última iteración.*

Aspecto calificado	Penúltima iteración Porcentaje (%)	Última iteración Porcentaje (%)
Porcentaje acierto en entrenamiento	87.22%	86.79%
Porcentaje acierto en pruebas	85.82%	88.12%
Número de reglas	16	15
Número de atributos inicial	3	3
Media de atributos/regla	1.0	1.0
Tiempo	0 min. 2 seg. 873 miliseg.	0 min. 2 seg. 826 miliseg.

Como podemos observar en lo que es el aspecto de porcentaje acierto en pruebas, en la última iteración se ve una precisión mayor por aproximadamente un 3%, lo que indicaría que a medida que REPSO – G realiza sus iteraciones, en cada una de ellas, la precisión va mejorando.

Tabla 4*Reglas generadas por el algoritmo REPSO–C.*

```

IF Peso_corporal < 140.0 AND IMC < 23.059019937011502 THEN CLASS = normal
IF Peso_corporal < 140.0 AND IMC < 29.661893877871684 THEN CLASS = sobrepeso
IF Estatura_m < 1.775662659866538 THEN CLASS = obesidad
IF IMC < 27.934647521950634 THEN CLASS = normal
IF      TRUE THEN CLASS =obesidad
IF Estatura_m >= 1.09 AND IMC < 23.65824623043852 THEN CLASS = normal
IF Peso_corporal < 92.729050280177 THEN CLASS = sobrepeso
IF Estatura_m < 1.836009487639064 THEN CLASS = obesidad
IF      TRUE THEN CLASS = normal
IF IMC < 23.06696432823888 THEN CLASS = normal
IF IMC < 29.13370329848766 THEN CLASS = sobrepeso
IF Estatura_m < 1.776233400447717 THEN CLASS = obesidad
IF IMC < 26.440988797260584 THEN CLASS = normal
IF Peso_corporal < 97.821368050496 THEN CLASS = sobrepeso
IF TRUE THEN CLASS = obesidad

```

DT-GA

La Tabla 5 muestra los resultados obtenidos por el modelo híbrido DT-GA comparado con ILGA.

Tabla 5

Porcentaje de precisión alcanzado por DT-GA e ILGA.

Aspecto calificado	DT-GA Porcentaje (%)	ILGA Porcentaje (%)
Porcentaje acierto en entrenamiento	93.76%	86.41%
Porcentaje acierto en pruebas	91.60%	86.41%
Número de reglas	5	12

Se observa un mejor comportamiento del modelo DT-GA respecto al modelo ILGA con porcentajes de acierto en entrenamiento y en pruebas mejores que a los otros para nuestro caso de estudio.

Tabla 6

Reglas generadas por el modelo DT-GA.

Reglas
IF Estatura_m <= 1.66 AND IMC <= 23.14 THEN Class = normal
IF Peso_corporal > 48.4 AND IMC <= 25.5 THEN Class = sobepeso
IF Peso_corporal <= 59.3 AND IMC <= 26.3 THEN Class = sobepeso
IF Peso_corporal > 28.6 AND Estatura <= 1.53 AND IMC <= 25.5 THEN Class = 3
IF Peso_corporal > 28.6 AND Estatura_m <= 1.53 AND IMC <= 25.5 THEN Class = 3

Y las reglas generadas por ILGA son (valores redondeados a dos decimales):

Tabla 7
Reglas generadas por el algoritmo ILGA.

Reglas
IF 127.29 < Peso_corporal < 131.83 AND 1.9 < Estatura_m < 1.32 AND 23.92 < IMC < 43.29 THEN sobrepeso
IF 63.79 < Peso_corporal < 93.27 AND 1.36 < Estatura_m < 1.15 AND 39.34 < IMC < 27.74 THEN normal
IF 35.55 < Peso_corporal < 119.69 AND 1.86 < Estatura_m < 1.12 AND 27.12 < IMC < 14.66 THEN obesidad
IF 24.06 < Peso_corporal < 64.73 AND 1.78 < Estatura_m < 1.17 AND 17.14 < IMC < 33.68 THEN normal
IF 86.8 < Peso_corporal < 100.89 AND 1.55 < Estatura_m < 1.11 AND 11.91 < IMC < 30.75 THEN obesidad
IF 21.14 < Peso_corporal < 56.49 AND 1.19 < Estatura_m < 1.69 AND 15.13 < IMC < 28.25 THEN normal
IF 74.76 < Peso_corporal < 118.41 AND 1.49 < Estatura_m < 1.14 AND 44.25 < IMC < 41.44 THEN obesidad
IF 75.9 < Peso_corporal < 133.81 AND 1.89 < Estatura_m < 1.65 AND 23.59 < IMC < 34.44 THEN sobrepeso
IF 66.64 < Peso_corporal < 82.02 AND 1.59 < Estatura_m < 1.82 AND 12.76 < IMC < 14.81 THEN obesidad
IF 105.3 < Peso_corporal < 108.82 AND 1.81 < Estatura_m < 1.47 AND 42.86 < IMC < 44.47 THEN sobrepeso
IF 56.71 < Peso_corporal < 129.61 AND 1.67 < Estatura_m < 1.71 AND 22.53 < IMC < 13.23 THEN obesidad
IF 82.21 < Peso_corporal < 93.02 AND 1.8 < Estatura_m < 1.73 AND 38.3 < IMC < 16.93 THEN sobrepeso

4. DISCUSIÓN

Los resultados muestran que tras la comparación entre los algoritmos REPSO-C, DT-GA e ILGA. El algoritmo DT-GA evidenció un mejor desempeño, resultado de acuerdo a su precisión de 93.76% en aciertos de entrenamiento y 91.60% en aciertos en pruebas, en relación a los otros algoritmos que presentaron valores relativamente inferiores en los tres modelos analizados. Esto refleja que este modelo DT-GA, puede ser utilizado por los especialistas en ciencias de la salud para identificar con mejor precisión el exceso de peso u obesidad, ya que puede ayudar a tomar decisiones en una institución educativa a través de la determinación de patrones y modelos.

En relación al mayor porcentaje de mejora, los resultados indican que el método DT-GA reflejó mejor porcentaje. Esto se debe a que DT-GA es un modelo híbrido que combina un árbol de decisión mejorado con algoritmos genéticos, donde el componente de evaluación es un árbol de decisión y el componente de búsqueda es un GA.

En consecuencia, se destaca en un estudio anterior, donde toman en cuenta que el REPSO-C es uno de los algoritmos evolutivos más eficientes (Liu et al., 2004), donde gracias a REPSO-C, es posible obtener una clasificación más eficiente. De hecho, en este estudio se comprobó que REPSO-C es bastante óptimo con lo que respecta a la clasificación de un conjunto de datos de obesidad escolar, inclusive, otro estudio reciente mostró similares valores de precisión con el presente estudio, aunque con diferente base de datos y temática de estudio (Liu et al., 2004).

Esta investigación muestra algunos aportes, por ejemplo, el mejoramiento de precisión en los resultados de clasificación por medio de un modelo híbrido, que combina árbol de decisión con algoritmos genéticos. Como aplicación práctica, los resultados sugieren que pueden ser analizados por profesionales de la salud, que puedan interpretar con mayor precisión para determinar de una mejor manera la clasificación de obesidad escolar.

CONCLUSIÓN

A través de los resultados obtenidos por la comparación de los tres algoritmos, se concluye que el modelo DT-GA ofrece un buen porcentaje de precisión (93.76%) con respecto a la clasificación de otros algoritmos evaluados. Estos resultados sugieren que es recomendable utilizar el modelo DT-GA en las diferentes áreas de la salud donde se puedan aplicar las tareas de clasificación y predicción en el exceso de peso y obesidad.

Por lo tanto, los resultados obtenidos constituyen una evidencia significativa con respecto a la precisión, donde el modelo híbrido C4.5/GA puede ser considerada como una mejor solución en cuanto a clasificación en nuestro caso de estudio, superando a REPSO-C y ILGA, de forma rápida y precisa.

AGRADECIMIENTOS

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

REFERENCIAS BIBLIOGRÁFICAS

ABDULLAH, F. S., MANAN, N. S. A., AHMAD, A., Wafa, S. W., SHAHRIL, M. R., ZULAILY, N., ... AHMED, A. (2016). Data Mining Techniques for Classification of Childhood Obesity Among Year 6 School Children. In *International Conference on Soft Computing and Data Mining* (pp. 465–474).

ALCALÁ-FDEZ, J., DEL JESUS, M. J., GARRELL, J. M., HERRERA, F., HERBÁS, C., & SÁNCHEZ, L. (2004). Proyecto KEEL: Desarrollo de una herramienta para el análisis e implementación de algoritmos de extracción de conocimiento evolutivos. *Tendencias de la minería de datos en España, red española de minería de datos y aprendizaje*, 413–424.

ANDALUZ, A. M. (2011). Algoritmos evolutivos y algoritmos genéticos. Ingeniería de Telecomunicaciones, Inteligencia en Redes de Comunicación. Universidad Carlos III de Madrid.

CARVALHO, D. R., & FREITAS, A. A. (2004). New results for a hybrid decision tree/genetic algorithm for data mining. In *Applications and Science in Soft Computing* (pp. 149–154). Springer.

COLE, T. J., BELLIZZI, M. C., FLEGAL, K. M., & DIETZ, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *Bmj*, 320(7244), 1240.

Guan, S. U., & Zhu, F. (2005). An incremental approach to genetic-algorithms-based classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(2), 227–239. <https://doi.org/10.1109/TSMCB.2004.842247>

LIU, Y., QIN, Z., SHI, Z., & CHEN, J. (2004). Rule discovery with particle swarm optimization. In *Content Computing* (pp. 291–296). Springer.

MEDSKER, L. R., & BAILEY, D. L. (1992). Models and guidelines for integrating expert systems and neural networks. In *Hybrid architectures for intelligent systems* (pp. 153–171).

MIYAHIRA, S. A., & ARAUJO, E. (2008). Fuzzy obesity index for obesity treatment and surgical indication. In IEEE International Conference on Fuzzy Systems (pp. 2392–2397). <https://doi.org/10.1109/FUZZY.2008.4630703>

NAYYERI, M., & SHARIFI NOGHABI, H. (2016). Cancer classification by correntropy-based sparse compact incremental learning machine. *Gene Reports*, 3, 31–38. <https://doi.org/10.1016/j.genrep.2016.01.001>

OMS. (2016). Comisión para acabar con la obesidad infantil. *Organización Mundial de La Salud*, 3–5. [https://doi.org/978 92 4 351006 4](https://doi.org/978%204%20351006%204)

ZHANG, S., TJORTJIS, C., ZENG, X., QIAO, H., BUCHAN, I., & KEANE, J. (2009). Comparing data mining methods with logistic regression in childhood obesity prediction. *Information Systems Frontiers*, 11(4), 449–460. <https://doi.org/10.1007/s10796-009-9157-0>