

# Análisis comparativo de la calidad de agrupamiento del algoritmo Intuitionistic Fuzzy C-Means

Celia Ramos-Palencia, Dante Mújica-Vargas, Manuel Mejía-Lavalle

Tecnológico Nacional de México,  
CENIDET,  
México

{celia.ramos17ce, dante.mv}@cenidet.tecnm.mx

**Resumen.** En este artículo se presenta un estudio comparativo de la calidad del agrupamiento obtenido por el algoritmo Intuitionistic Fuzzy C-Means respecto a los algoritmos Fuzzy C-Means, Gath-Geva, Gustafson Kesel, K-Means y K-Medoid. La evaluación se hace en tareas de reconocimiento de patrones, utilizando las métricas: Xie-Beni, de Dunn, de Partición, Coeficiente de Partición, Entropía de Clasificación, de Separación y alternativo de Dunn. Los resultados indican que el algoritmo Intuitionistic Fuzzy C-Means es más eficiente en la calidad de agrupamiento y convergencia en comparación con los otros algoritmos utilizados para este trabajo.

**Palabras clave:** Agrupamiento, Intuitionistic Fuzzy C-means, algoritmos clásicos de agrupamiento, calidad de agrupamiento.

## Comparative Analysis of the Clustering Quality of the Intuitionistic Fuzzy C-Means Algorithm

**Abstract.** In this paper we proposed a comparative study of the clustering quality obtained by the Intuitionistic Fuzzy C-Means algorithm about the Fuzzy C-Means, Gath-Geva, Gustafson Kesel, K-Means and K-Medoid algorithms. The evaluation is done in pattern recognition tasks, using the metrics: Xie-Beni, Dunn's, Index, Partition Coefficient, Classification Entropy, Index Separation, and Dunn's Alternative Index. The experimental results show that the Intuitionistic Fuzzy C-Means algorithm is more efficient in the quality of clustering and convergence compared to the other algorithms used in this work.

**Keywords:** Clustering, Intuitionistic Fuzzy C-means, classic clustering algorithms, clustering quality.

## 1. Introducción

*Clustering* o *Agrupamiento* puede ser definido como un método que se utiliza para crear un determinado número de grupos de objetos, de tal forma que los objetos pertenecientes a un grupo son muy similares entre sí y, al mismo tiempo, dichos objetos son distintos respecto a los objetos de otros grupos [1].

La investigación en el área de agrupamiento de datos ha producido una gran cantidad de diferentes métodos que se centran en medir la efectividad de la calidad del agrupamiento y su impacto en el rendimiento de este tipo de algoritmos. En [2], se comparó la efectividad de algunos índices difusos internos ampliamente utilizados, usando el algoritmo Fuzzy C-Means, estando dirigido a investigar los cambios de efectividad de los índices de validez cuando los datos difusos están a diferentes distancias de los centroides. En [3], se mostró un trabajo experimental que compara 30 índices de validez en diversos entornos con características diferentes. Los resultados pueden servir como una guía para seleccionar el índice más adecuado para cada aplicación posible y proporcionar una visión detallada de las diferencias de rendimiento. En [4], se presentó una revisión de los índices de validación difusos disponibles en la literatura, haciendo una comparación exhaustiva junto con el algoritmo Fuzzy C-Means en varias bases de datos ampliamente utilizadas.

En [5], se realizó un estudio del comportamiento de algunos índices en algunas bases de datos con ruido o valores atípicos. Los algoritmos utilizados fueron K-Means, agrupamiento jerárquico, usando el método de Ward y agrupamiento basado en modelos usando mezclas Gaussianas, el algoritmo de maximización de expectativas (EM). El objetivo fue conocer con cual método se obtenían un mejor comportamiento en cuanto al ruido o la presencia de valores atípicos. En [6], se presentó un índice de acuerdo de agrupamiento (CAI) para comparar la disyunción y superposición de la agrupación. Se utilizaron dos índices: ARI y NMI y el propósito fue extender a casos superpuestos, generalizar medidas existentes e introducir medidas novedosas.

En [7], se mencionó que la principal desventaja de K-Means es que el número de los grupos debe ser predifinido, por lo tanto, en este estudio se compararon diferentes índices de validación para la determinación automatizada y reproducible del número óptimo de grupos.

En [8], se evaluaron las características de un conjunto representativo de índices de validación de agrupamiento interno con diferentes bases de datos. El prototipo basado en el marco del agrupamiento incluye estimaciones estadísticas múltiples, clásicas y robustas de la ubicación del grupo, lo que hace este trabajo novedoso. Se utilizaron los algoritmos K-Means, K-medians y K-spatialmedians. En [9], se presentó un análisis comparativo entre las metodologías basadas en máquinas de vectores de soporte y Fuzzy C-Means para agrupar señales de ECG normales y patológicas en un espacio de complejidad de Lempel-Ziv y entropía de Shannon. En [10], se presentó un estudio comparativo para el reconocimiento de patrones con algoritmos de agrupamiento en la clasificación en señales de comunicación analógicas moduladas.

Los algoritmos utilizados fueron K-Means y Fuzzy C-Means. En [11], se hizo una evaluación de las métricas disponibles para comparar algoritmos de agrupamiento. También se definieron algunas restricciones formales intuitivas sobre tales métricas, además de proponer un conjunto de restricciones que una buena métrica de evaluación debe satisfacer en un problema de agrupamiento genérico.

Se pudo observar que la comparación de algunos índices se ha realizado con un solo algoritmo de agrupamiento, en este caso Fuzzy C-Means, o se han hecho estudios comparativos de algunas métricas para su mejora, sin embargo no se ha realizado un estudio para comprobar la calidad de agrupamiento de un algoritmo de agrupamiento difuso intuicionista, por ello se propone realizar un análisis comparativo para medir la calidad de agrupamiento de un algoritmo difuso intuicionista y demostrar que es superior a otros algoritmos clásicos en términos de convergencia y calidad de agrupamiento, para posteriormente poder mejorarlo a través de alguna fusión con otras metodologías, por ejemplo, dotarlo de capacidad de aprendizaje con una red neuronal.

La estructura de este documento es de la siguiente manera: en la Sección 2 se presenta una breve descripción de los conceptos utilizados en este documento. En la Sección 3 se presenta la evaluación comparativa de los algoritmos y las métricas en las diferentes bases de datos seleccionadas para este propósito. En la Sección 4 se muestra la descripción de los resultados. Por último se presentan las conclusiones y posibles trabajos futuros.

## 2. Marco teórico

Dentro de la clasificación de algoritmos de agrupamiento, se encuentran los de agrupamiento difuso, que forman una clase en la cual cada uno de los elementos tiene un cierto grado de pertenencia a los grupos y no solo a uno, como por ejemplo, el clásico K-Means. Este tipo de algoritmos considera que cada elemento se puede agrupar inequívocamente con los elementos de su grupo y que, por lo tanto, no se asemeja al resto de los elementos [12].

### 2.1. Algoritmo K-Means

El algoritmo K-Means es un método de partición más conocido y utilizado; agrupa los datos en función de su cercanía entre sí de acuerdo con una medida de distancia, por ejemplo, la euclidiana. Minimiza la siguiente función objetivo [1]:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|. \quad (1)$$

En la Figura 1 se muestran los pasos del algoritmo y sus requerimientos.

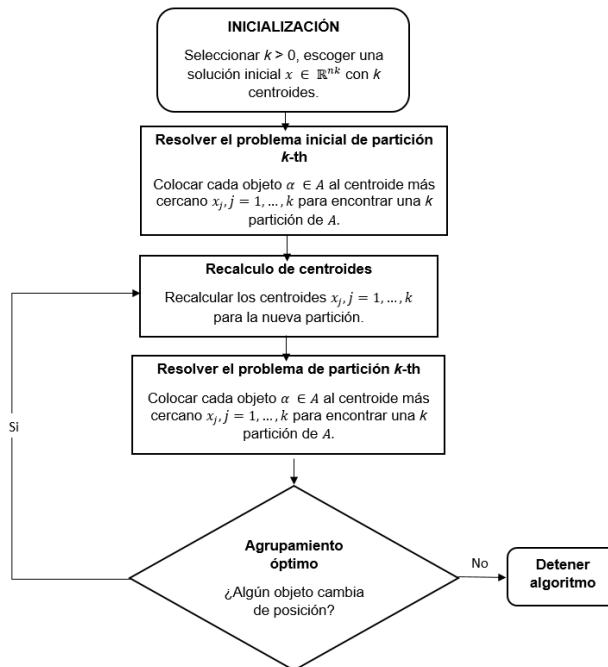


Fig. 1. Algoritmo K-Means.

### 2.2. Algoritmo Fuzzy C-Means

En Fuzzy C-Means, los datos pueden pertenecer a más de un grupo y también tener diferentes grados de pertenencia en los diferentes grupos. Minimiza la siguiente función objetivo [12]:

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2. \tag{2}$$

En la Figura 2 se observa la descripción de los pasos de este algoritmo.

### 2.3. Algoritmo Gath Geva

Permite detectar conjuntos de objetos de distintas formas, tamaños y densidades [12,13]. A causa del tipo de distancia que se utiliza (Euclidiana, Manhattan, Mahalanobis, entre otras), los conjuntos no están restringidos a un solo grupo como en el caso de K-Means. Sin embargo, es muy sensible a la inicialización, ya que el algoritmo tiende a converger en óptimos locales. La Figura 3 muestra una descripción detallada del funcionamiento del algoritmo.

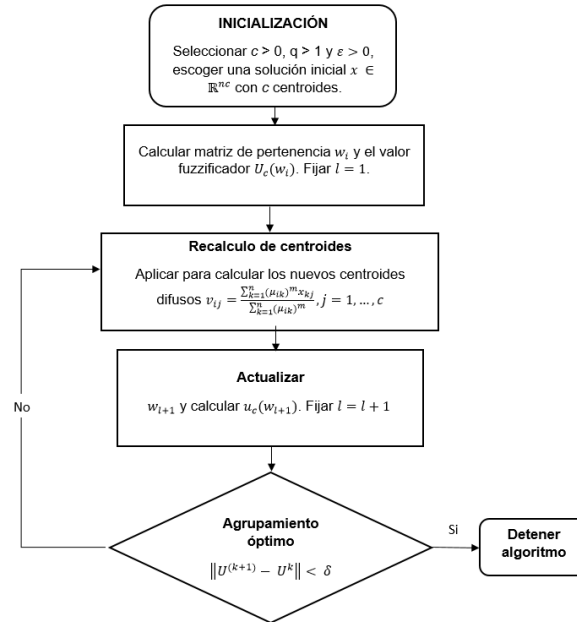


Fig. 2. Algoritmo Fuzzy C-Means.

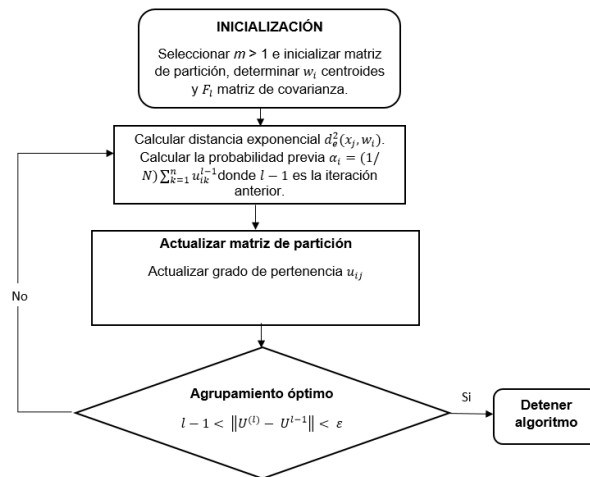


Fig. 3. Algoritmo Gath Geva.

#### 2.4. Algoritmo Gustafson Kessel

Muy similar a FCM, los parámetros  $m, \varepsilon, \mu$  y  $j$  son los mismos en los dos algoritmos [14]. Un medoide es un objeto ubicado más al centro del grupo, con

una suma mínima de distancias a otros puntos. En la Figura 4 se presentan los pasos y especificaciones de este método.

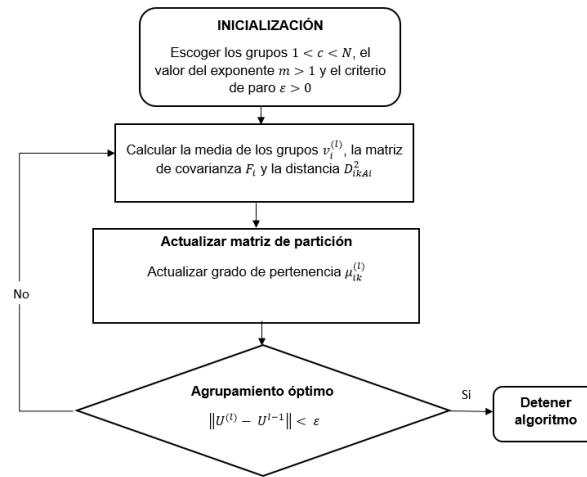


Fig. 4. Algoritmo Gustafson Kessel.

## 2.5. Algoritmo K-Medoids

Es un algoritmo usado para encontrar los medoides, el cual es un punto situado en el centro de un grupo [15,16]. Además de ser más robusto al ruido y valores atípicos. La Figura 5 muestra los pasos del algoritmo.

## 2.6. Agrupamiento difuso intuicionista

Gracias a la teoría de conjuntos difusos intuicionistas surge Intuitionistic Fuzzy C-Means (IFCM) o Fuzzy C-Means intuicionista, que además de agregar un grado de pertenencia, asignan a cada elemento un grado de no pertenencia y uno de indecisión, permitiendo que este método sea más poderoso que el tradicional FCM [17].

Los IFS asignan a cada elemento un grado de pertenencia  $\mu_F(x)$ , un grado de no pertenencia  $\nu_F(x)$  y un grado de indecisión  $\pi_F(x)$ , permitiendo que este método sea más poderoso que el algoritmo difuso tradicional.

*Intuitionistic Fuzzy Sets:* Dado  $X$  como el conjunto de datos, dado  $A \subset X$  como un conjunto fijo. Entonces un IFS  $A$  en  $X$  es expresado de la siguiente manera:

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \}, \quad (3)$$

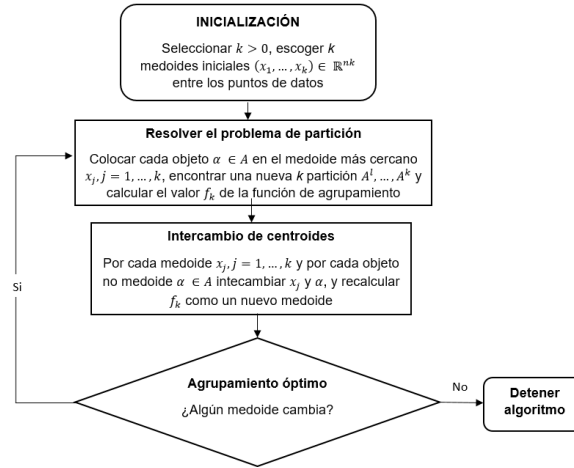


Fig. 5. Algoritmo K-Medoids.

donde las funciones  $\mu_A : A \mapsto [0, 1]$  y  $\nu_A : A \mapsto [0, 1]$  son los grados de pertenencia y no pertenencia del elemento  $x \in X$  al conjunto  $A$ , respectivamente, y para cada  $x \in X$ :

$$0 \leq \mu_A + \nu_A \leq 1. \quad (4)$$

Cada conjunto difuso intuicionista debe ser:

$$A^* = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \}. \quad (5)$$

El siguiente parámetro es el grado de indecisión  $\pi_A(x)$ , al ser incluido en la expresión anterior, debe tomar la siguiente forma:

$$A^* = \{ \langle x, \mu_A(x), \nu_A(x), \pi_A(x) \rangle \mid x \in X \}, \quad (6)$$

donde el grado de indecisión puede ser expresado como  $\pi_A = 1 - \mu_A(x) - \nu_A(x)$ .

IFCM requiere menor tiempo de ejecución debido a que la cantidad de iteraciones necesarias para lograr un ejercicio de agrupamiento específico corresponde a la precisión requerida. El algoritmo IFCM encuentra una partición difusa de los datos minimizando la siguiente función objetivo:

$$\min J_m (X^{IFS}; U, V^{IFS}) = \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m d^2 (x_j^{IFS} - v_i^{IFS}). \quad (7)$$

Sujeto a la condición:

$$\sum_{i=1}^c u_{ij} = 1, 1 \leq j \leq N, \quad (8)$$

donde  $X^{FIS} = (x_1^{FIS}, x_2^{FIS}, \dots, x_n^{FIS})$  son  $N$  datos difusos intuicionistas y cada  $x_j^{FIS}$  está representado en términos de conjuntos difusos intuicionistas como  $x_j^{FIS} = \{(\mu(x_j), \nu(x_j), \pi(x_j))\}$ ,  $m(1 < m < \infty)$  es el factor difuso,  $c(1 < c < N)$  es el número de grupos, y  $V^{FIS} = \{v_1^{FIS}, v_2^{FIS}, \dots, v_c^{FIS}\}$  son los centroides de los grupos y cada  $v_i^{FIS}$  está representado en términos de conjuntos difusos intuicionistas como  $v_i^{FIS} = \{\mu(v_i), \nu(v_i), \pi(v_i)\}$ ,  $u_{ij} (0 \leq u_{ij} \leq 1)$  es el grado de pertenencia y  $U = (u_{ij})_{c \times N}$  es la matriz de pertenencia. La matriz de pertenencia  $U$  también cumple la condición  $0 < \sum_{i=1}^c u_{ij} < N, \forall i$ .

La pertenencia y los centroides  $v_i^{FIS} = \mu(v_i, \nu(v_i), \pi(v_i))$  se calculan de la siguiente manera:

$$u_{ij} = \frac{1}{\sum_{r=1}^c \left( \frac{d^2(x_j^{FIS}, v_i^{FIS})}{d^2(x_j^{FIS}, v_r^{FIS})} \right)^{\frac{1}{m-1}}}, 1 \leq i \leq c, 1 \leq j \leq N. \quad (9)$$

El vector de prototipos está dado por:

$$v_i = \{\mu(v_i), \nu(v_i), \pi(v_i)\}, \quad (10)$$

donde cada uno de los componentes se da de la siguiente manera:

$$\mu(v_i) = \frac{\sum_{j=1}^N u_{ij}^m \mu(x_j)}{\sum_{j=1}^N u_{ij}^m}, 1 \leq i \leq c, \quad (11)$$

$$\nu(v_i) = \frac{\sum_{j=1}^N u_{ij}^m \nu(x_j)}{\sum_{j=1}^N u_{ij}^m}, 1 \leq i \leq c, \quad (12)$$

$$\pi(v_i) = \frac{\sum_{j=1}^N u_{ij}^m \pi(x_j)}{\sum_{j=1}^N u_{ij}^m}, 1 \leq i \leq c. \quad (13)$$

### 3. Evaluación comparativa

En esta sección se muestra el proceso de experimentación, los datos y las métricas de evaluación utilizados para verificar que el algoritmo IFCM es superior al resto de los otros métodos mencionados, en cuanto a calidad de agrupamiento y convergencia

#### 3.1. Evaluación

Se utilizaron bases de datos reales obtenidas del repositorio de la UCI [18]. En la Tabla 1 se muestran la bases de datos utilizadas y el total de objetos de cada una. Para verificar la calidad del agrupamiento de los algoritmos y hacer el estudio comparativo, se utilizaron diferentes métricas: Coeficiente de Partición (PC), Clasificación de la Entropía (CE), Índice de Partición (SC), Índice de Separación (S), Xie-Beni (XB), de Dunn (DI) y Alternativo de Dunn (ADI).



**Tabla 1.** Bases de Datos.

Nombre	Número de objetos
Iris	150
Wine	175
Ecoli	336
Breast Cancer	696
RSSI	1420
Skin	9568
Carbon	10721
Bank	45210
Power Plant	245057
Buzz	260059

1. *Coficiente de Partición*(PC): Mide la cantidad de traslape entre grupos. Fue definido por Bezdek. [19,20]:

$$PC(c) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^2, \quad (14)$$

donde  $\mu_{ij}$  es la pertenencia del  $i$ -ésimo grupo. El número óptimo de grupos es indicado por un valor máximo de este índice.

2. *Entropía de Clasificación* (CE): Este índice mide la falta de difusividad de los grupos [9]:

$$CE(c) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c \mu_{ij} \log(\mu_{ij}). \quad (15)$$

3. *Índice de partición* (SC): Es la división de la suma de separación y compactibilidad de los grupos; a su vez, es una suma de medidas individuales de la validez de los grupos que es normalizada a través de la división de la cardinalidad difusa de cada grupo. El número óptimo de grupos se logra al minimizar este índice [9]:

$$SC(c) = \sum_{j=1}^c \frac{\sum_{i=1}^n (\mu_{ij})^m \|x_i - c_j\|^2}{n_j \sum_{k=1}^c \|c_k - c_j\|^2}. \quad (16)$$

4. *Índice de separación* (S): A diferencia del anterior, este índice utiliza una separación de distancia mínima para validar el agrupamiento [9]:

$$S(c) = \frac{\sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^2 \|x_i - c_j\|^2}{n \min_{k,j} \|c_k - c_j\|^2}. \quad (17)$$

5. *Índice de Xie y Beni* (XB): Ayuda a medir la relación de la variación total dentro de los grupos y la separación de los mismos:[21]:

$$XB(c) = \frac{\sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m \|x_i - c_j\|^2}{n \min_{i,j} \|x_i - c_j\|^2}. \quad (18)$$

6. *Índice de Dunn* (DI): Este índice es originalmente propuesto para usarse en la identificación de “grupos compactos y bien separados” [21]:

$$DI(c) = \min_{j \in c} \left\{ \min_{i \in c, i \neq j} \left\{ \frac{\min_{x \in C_j, y \in C_i} d(x, y)}{\max_{k \in c} \{ \max_{x, y \in C} d(x, y) \}} \right\} \right\}, \quad (19)$$

el principal inconveniente de este índice es que su carga computacional se vuelve muy expansiva en función del incremento de  $n$  y  $c$ .

7. *Índice Alternativo de Dunn* (ADI): En este caso, el cálculo se vuelve más simple que en el índice de Dunn tradicional, cuando la función de disimilaridad entre dos grupos ( $\min_{x \in C_i, y \in C_j} d(x, y)$ ) [21]:

$$d(x, y) \geq |d(y, v_j) - d(x, v_j)|, \quad (20)$$

donde  $v_j$  es el centroide del grupo del  $j$ -ésimo centroide:

$$ADI(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x_i \in C_i, x_j \in C_j} |d(y, v_j) - d(x_i, v_j)|}{\max_{k \in c} \{ \max_{x, y \in C} d(x, y) \}} \right\} \right\}. \quad (21)$$

## 4. Resultados

En esta sección se presenta la evaluación comparativa de algoritmos de agrupamiento clásicos y el algoritmo Intuitionistic Fuzzy C-Means, a través de gráficas comparativas utilizando las bases de datos mencionadas anteriormente, con los siguientes parámetros: en el caso de los difusos  $m=2$ , y  $\epsilon = 1e - 6$ . Como nota, este criterio de paro no es aplicable a todos los algoritmos ya que durante los pasos del mismo, algunos de los valores obtenidos son *infinitos* o *NaN*, que significa no un número, por ejemplo, hacer una operación inusual, como dividir entre cero.

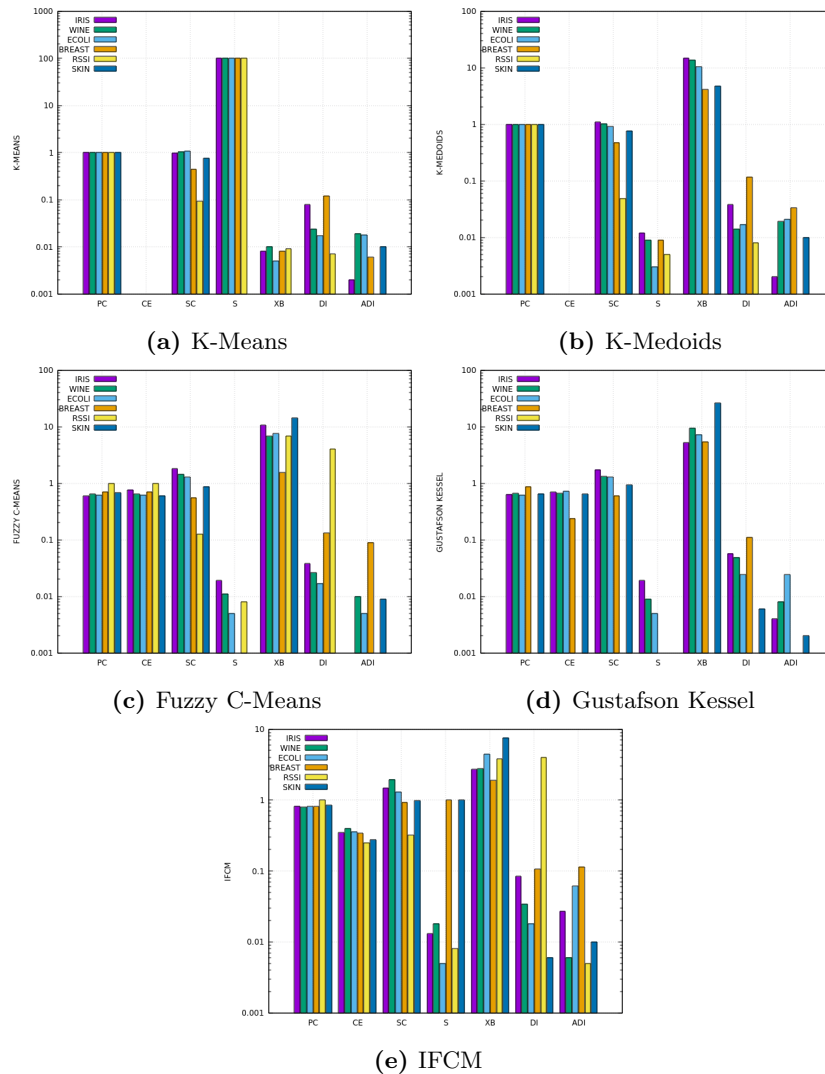
La experimentación se llevó a cabo en MATLAB R2018, usando el *Toolbox Clustering* [22] para los algoritmos clásicos, IFCM fue programado por separado; una computadora con 1 Tb de disco duro de almacenamiento, 8Gb de RAM, procesador iCore i7 de 7ma generación, y Linux Xubuntu 18.04.

### 4.1. Métricas e iteraciones

En la Tabla 2 se observan los resultados de cada métrica con la ejecución de cada uno de los algoritmos, para cada base de datos; y en las gráficas de la Figura 6 se muestra cómo con el algoritmo IFCM, los resultados obtenidos en cada métrica tienden a alcanzar los valores óptimos. Anteriormente se mencionó

**Tabla 2.** Resultados cuantitativos.

Base de datos	Metrica	Kmeans	Kmedoids	FCM	GG	GK	IFCM
Iris	PC	1	1	0.59	0.98	0.63	0.82
	CE	NaN	NaN	0.76	NaN	0.69	0.34
	SC	1.27	0.87	1.80	NaN	1.74	1.47
	S	0.009	0.006	0.01	NaN	0.01	0.01
	XB	8.38	Inf	10.58	NaN	5.20	2.70
	D	0.04	0.04	0.03	0.05	0.05	0.08
	ADI	0.03	0.001	0.001	0.003	0.004	0.02
	<hr/>						
Wine	PC	1	1	0.80	-	0.89	0.79
	CE	NaN	NaN	0.38	-	0.17	0.39
	SC	0.21	0.22	0.49	-	0.000006	1.95
	S	0.001	0.001	0.004	-	0.000006	0.01
	XB	47.6	Inf	36.44	-	48	2.77
	D	0.03	0.01	0.01	-	0.004	0.03
	ADI	0.00001	0.000007	0.000001	-	0	0.006
	<hr/>						
Ecoli	PC	1	1	0.61	-	0.62	0.80
	CE	NaN	NaN	0.73	-	0.72	0.35
	SC	1.07	0.88	1.27	-	1.27	1.30
	S	0.005	0.003	0.005	-	0.005	0.005
	XB	11.9	Inf	7.52	-	7.20	4.46
	D	0.01	0.02	0.01	-	0.02	0.01
	ADI	0.01	0.001	0.005	-	0.02	0.06
	<hr/>						
Breast Cancer	PC	1	1	0.70	-	0.86	0.81
	CE	NaN	NaN	0.56	-	0.23	0.33
	SC	0.47	0.52	0.55	-	0.60	0.93
	S	0.001	0.001	0.001	-	0.001	0.001
	XB	3.75	Inf	1.54	-	5.37	1.89
	D	0.10	0.11	0.13	-	0.11	0.10
	ADI	0.03	0.11	0.09	-	0.00001	0.11
	<hr/>						
RSSI	PC	1	1	0.99	-	-	0.99
	CE	NaN	NaN	0.01	-	-	0.001
	SC	0.04	0.08	0.12	-	-	0.32
	S	0.000005	0.000008	0.000008	-	-	0.00004
	XB	NaN	Inf	6.92	-	-	3.79
	D	0.082	0.007	4.03	-	-	4.03
	ADI	0	0	0	-	-	0
	<hr/>						
Skin	PC	1	1	0.67	-	-	0.87
	CE	NaN	NaN	0.60	-	-	0.27
	SC	0.76	0.66	0.87	-	-	0.98
	S	0.00001	0.000009	0.00001	-	-	0.00001
	XB	47.4	Inf	142.16	-	-	7.61
	D	0.001	0.001	0.001	-	-	0.00006
	ADI	0.01	0.005	0.009	-	-	0.001
	<hr/>						
Carbon	PC	1	1	0.64	0.90	0.64	0.82
	CE	NaN	NaN	0.70	0.15	0.65	0.34
	SC	0.61	0.61	0.77	1.55	1.11	0.99
	S	0.00005	0.000005	0.000007	0.00001	0.00001	0.000009
	XB	44.8	Inf	18.42	1.58	28.09	5.85
	D	0.003	0.002	0.001	0.001	0.001	0.001
	ADI	0.001	0.001	0.004	0.02	0.004	0.02
	<hr/>						
Bank	PC	1	1	0.74	-	-	0.88
	CE	NaN	NaN	0.47	-	-	0.20
	SC	0.99	1.05	1.44	-	-	0.90
	S	0.000003	0.000003	0.000004	-	-	0.000003
	XB	9.23	Inf	10.76	-	-	7.79
	D	0.001	0.001	0.01	-	-	0.00001
	ADI	0	0	0	-	-	0
	<hr/>						
Power Plant	PC	1	1	0.99	-	-	0.88
	CE	NaN	NaN	0.01	-	-	0.20
	SC	0.49	0.50	0.12	-	-	0.90
	S	0.0000002	0.0000002	0.000008	-	-	0.000003
	XB	73.02	Inf	6.92	-	-	7.79
	D	0.002	0.028	4.03	-	-	0.00001
	ADI	0.003	0.003	0	-	-	0
	<hr/>						
Buzz	PC	1	1	0.95	-	-	0.34
	CE	NaN	NaN	0.08	-	-	1.18
	SC	0.43	0.50	0.59	-	-	3.43
	S	0.000002	0.000002	0.000003	-	-	0.12
	XB	159.42	Inf	367.27	-	-	8.78
	D	0.00004	0.0028	0.00005	-	-	0.000001
	ADI	0.0000001	0.0037	0.0000001	-	-	0.0000002



**Fig. 6.** Representación gráfica de los resultados cuantitativos.

que debido al criterio de paro, algunos algoritmos se ciclaban y no terminaban de iterar, es por ello, que en la Tabla 2 y en las subfiguras de la Figura 6, se observan lugares vacíos.

Esta situación no se presentó con IFCM; con este método cada una de las bases de datos pudo ser ejecutada sin problema, con un menor tiempo de ejecución (número de iteraciones), las cuales se redujeron en promedio entre un 93% y un 99% respecto al número de iteraciones realizadas por los otros algoritmos, y/o con una mayor calidad de agrupamiento.

## 5. Conclusiones y trabajos futuros

En este artículo se presentó un estudio comparativo de la calidad del agrupamiento obtenido por el algoritmo Intuitionistic Fuzzy C-Means y se realizó una comparación con algunos métodos de agrupamiento utilizados en la literatura. La evaluación fue hecha en tareas de reconocimiento de patrones, utilizando índices de validación. Los resultados indican que el algoritmo IFCM reduce hasta en un 93% y 99% la cantidad de iteraciones necesarias para converger, en cuanto a la calidad de agrupamiento, en la mayoría de las métricas se alcanzan los valores cercanos a los óptimos para una buena calidad de agrupamiento. Con estos resultados se pretende, como trabajo futuro, dotar de aprendizaje neuronal a este algoritmo ya que es eficiente en las tareas de agrupamiento y reconocimiento de patrones.

**Agradecimientos.** Los autores de este trabajo expresan su gratitud al CONACYT, así como al Tecnológico Nacional de México/CENIDET por el financiamiento a través del proyecto “Controlador Difuso para ajuste de coeficientes de rigidez de un modelo deformable para simulación en tiempo real de los tejidos del hígado humano”.

## Referencias

1. Mújica-Vargas, D.: Segmentación de imágenes utilizando algoritmos robustos de agrupamiento difuso. Tesis Doctoral, Instituto Politécnico Nacional (2013)
2. Cebeci, Z.: Comparison of internal validity indices for fuzzy clustering. *Journal of Agricultural Informatics*, 10(2) (2019)
3. Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I.: An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1), pp. 243–256 (2013)
4. Wang, W., Zhang, Y.: On fuzzy cluster validity indices. *Fuzzy Sets and Systems*, 158(19), pp. 2095–2117 (2007)
5. Guerra, L., Robles, V., Bielza, C., Larrañaga, P.: A comparison of clustering quality indices using outliers and noise. *Intelligent Data Analysis*, 16(4), pp. 703–715 (2012)
6. Rabbany, R., Zaiane, O.R.: Generalization of clustering agreements and distances for overlapping clusters and network communities. *Data Mining and Knowledge Discovery*, 29(5), pp. 1458–1485 (2015)
7. Abualhaj, B., Weng, G., Ong, M., Attarwala, A.A., Molina, F., Büsing, K., Glatting, G.: Comparison of five cluster validity indices performance in brain [18F]FET-PET image segmentation using k-means. *Medical Physics*, 44(1), pp. 209–220 (2017)
8. Hämmäläinen, J., Jauhiainen, S., Kärkkäinen, T.: Comparison of internal clustering validation indices for prototype-based clustering. *Algorithms*, 10(3), pp. 105 (2017)
9. Villazana, S., Arteaga, F., Seijas, C., Rodríguez, O.: Estudio comparativo entre algoritmos de agrupamiento basado en svm y c-medios difuso aplicados a señales electrocardiográficas arritmicas. *Revista Ingeniería*, pp. 16–24 (2012)

10. Guldemir, H., Sengur, A.: Comparison of clustering algorithms for analog modulation classification. *Expert Systems with Applications*, 30(4), pp. 642–649 (2006)
11. Amigó, E., Gonzalo, J., Artiles, J., Verdejo, F.: A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval*, 12(5), pp. 613–613 (2009)
12. Ramos-Palencia, C.: Mejora de un algoritmo de agrupamiento del estado del arte mediante heurísticas aplicadas a la mejora del algoritmo K-Means. Tesis de Maestría, Tecnológico Nacional de México (2019)
13. Lange, O., Meyer-Baese, A., Hurdal, M., Foo, S.: A comparison between neural and fuzzy cluster analysis techniques for functional MRI. *Biomedical Signal Processing and Control*, 1(3), pp. 243–252 (2006)
14. Miller, D.J., Nelson, C.A., Cannon, M.B., Cannon, K.P.: Comparison of fuzzy clustering methods and their applications to geophysics data. *Applied Computational Intelligence and Soft Computing*, pp. 1–16 (2009)
15. Arora, P., Deepali, D., Varshney, S.: Analysis of k-means and k-medoids algorithm for big data. *Procedia Computer Science*, 78, pp. 507–512 (2016)
16. Park, H.S., Jun, C.H.: A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36(2), pp. 3336–3341 (2009)
17. Xu, Z.: Intuitionistic fuzzy clustering algorithms. In: *Intuitionistic Fuzzy Aggregation and Clustering Studies in Fuzziness and Soft Computing*, pp. 159–267 (2012)
18. Lichman, M.: UCI Machine learning repository. (2013)
19. Bezdek, J.C.: Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1(1), pp. 57–71 (1974)
20. Bezdek, J.C.: Cluster validity with fuzzy sets. *Journal of Cybernetics*, 3(3), pp. 58–73 (1973)
21. Desgraupes, B.: Clustering indices. University of Paris Ovest-Lab Modal’X, 1 (2013)
22. Clustering Toolbox: MathWorks. MATLAB (2012)
23. Gustafson, D., Kessel, W.: Fuzzy clustering with a fuzzy covariance matrix. In: *IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes* (1978)