

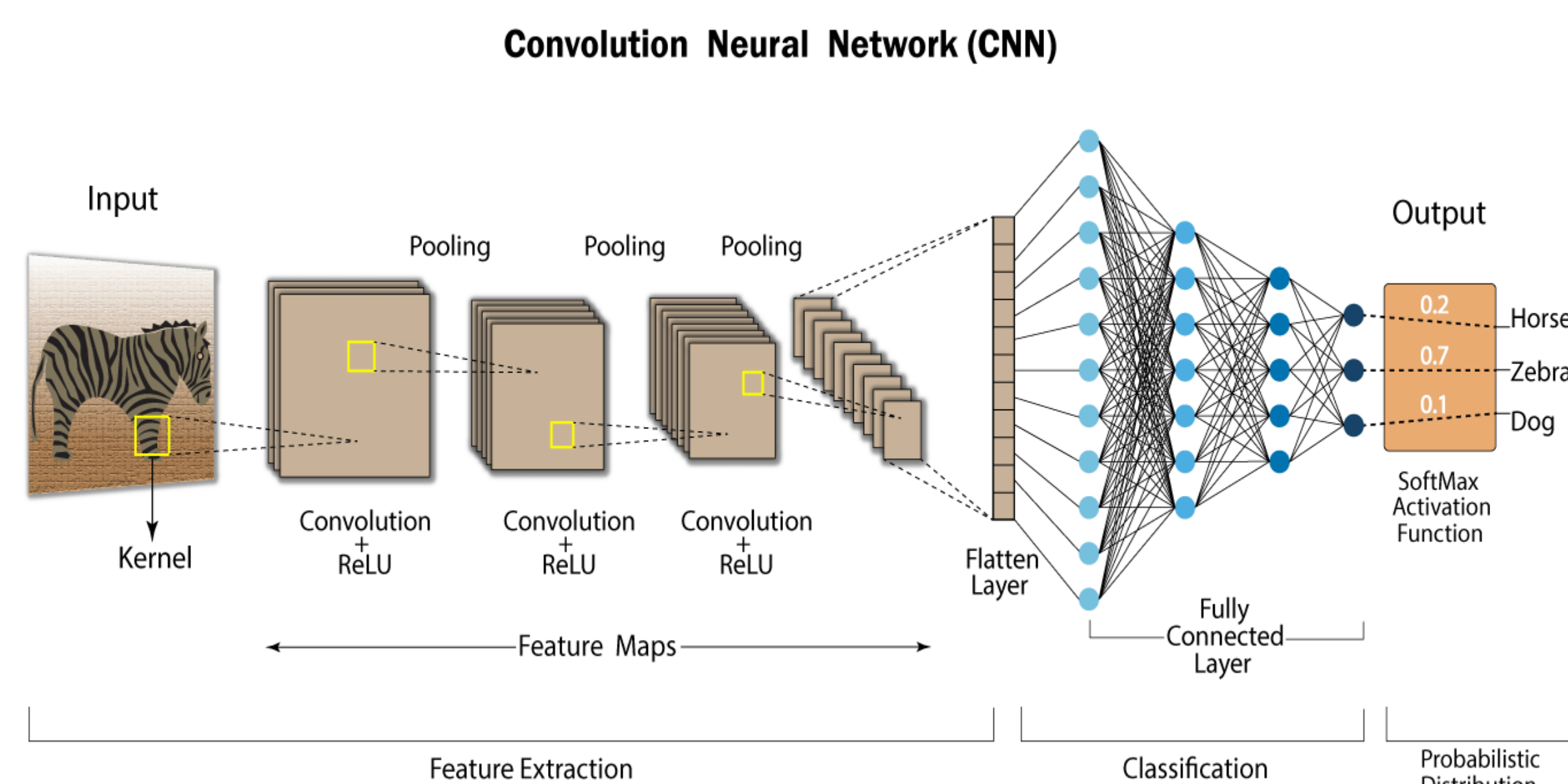
# Comparativa entre K-means y Fuzzy C-means para la agrupación de características en imágenes.

Ángel García Báez, Héctor Acosta Mesa y Sergio Hernández Méndez  
Instituto de Investigaciones en Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, Mé

**Resumen:** Se compararon los algoritmos K-means y Fuzzy C-means para agrupar imágenes de rostros, utilizando vectores de características extraídos con una red neuronal la red convolucional preentrenada VGG16. Se encontró que Fuzzy C-means presenta un mejor desempeño ante datos ruidosos.

## Introducción

La agrupación no supervisada de imágenes es una técnica clave para explorar y descubrir patrones en grandes volúmenes de datos visuales. En este estudio se comparan los algoritmos K-means y Fuzzy C-means aplicados a una base de datos de imágenes de rostros, utilizando una red neuronal convolucional pre-entrenada para extraer representaciones de alto nivel. Se busca evaluar cuál de estos métodos logra una mejor separación de los grupos en el espacio latente generado por la CNN.



## Aplicación

Se extrajeron las características de las 7000 imágenes, en donde cada imagen pasaba a ser un vector de 512 dimensiones. Posteriormente se seleccionaron al azar 800 de esos vectores por limitaciones computacionales. Se determinó el número óptimo de clusters en 3 para K-means y para Fuzzy C-means usando el método del codo de Janbu para graficar la intravarianza. Posteriormente se obtuvieron los clusters y a continuación se muestra el resultado proyectado en 2D con ayuda de PCA.

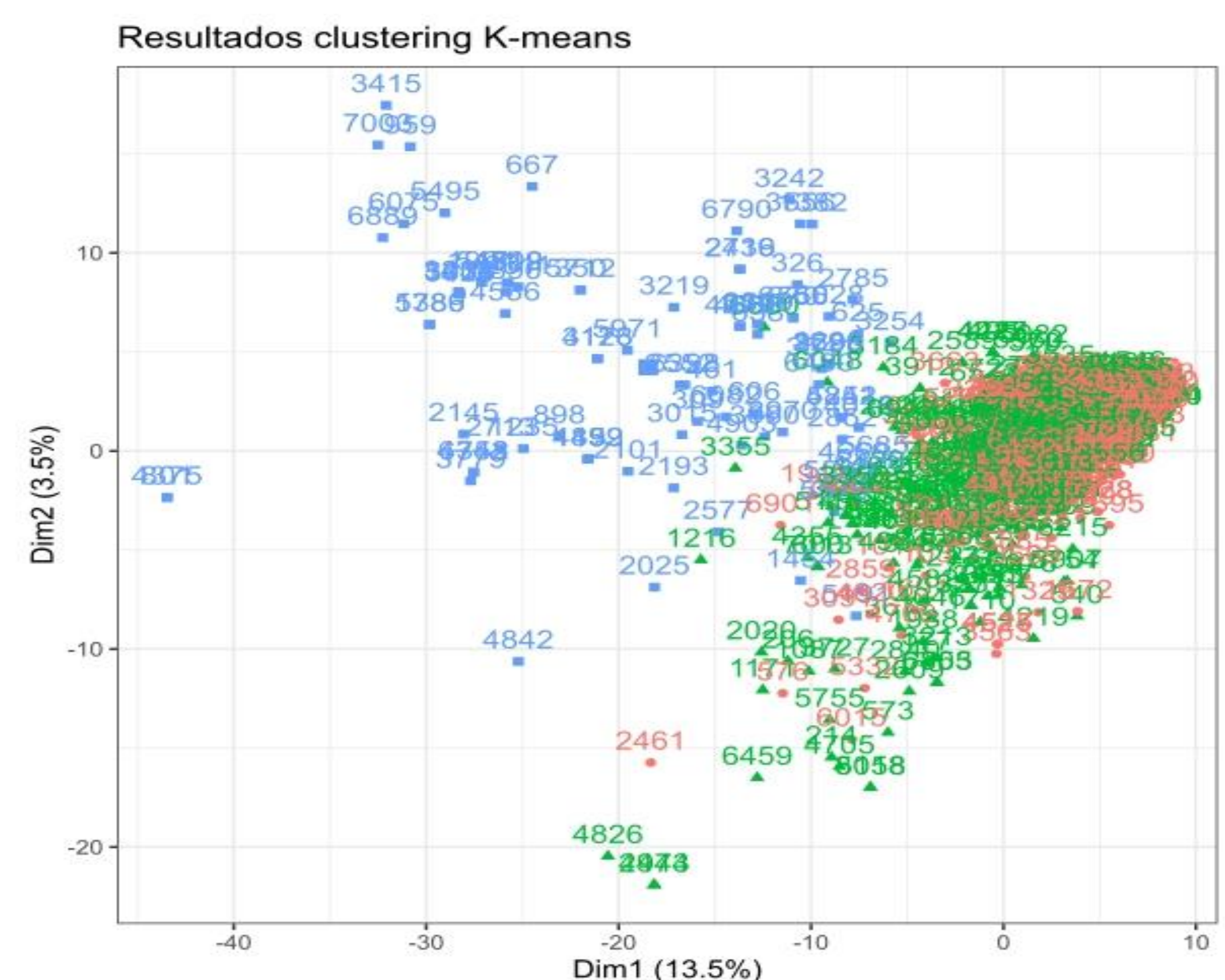


Figura 1: Cluster K-means

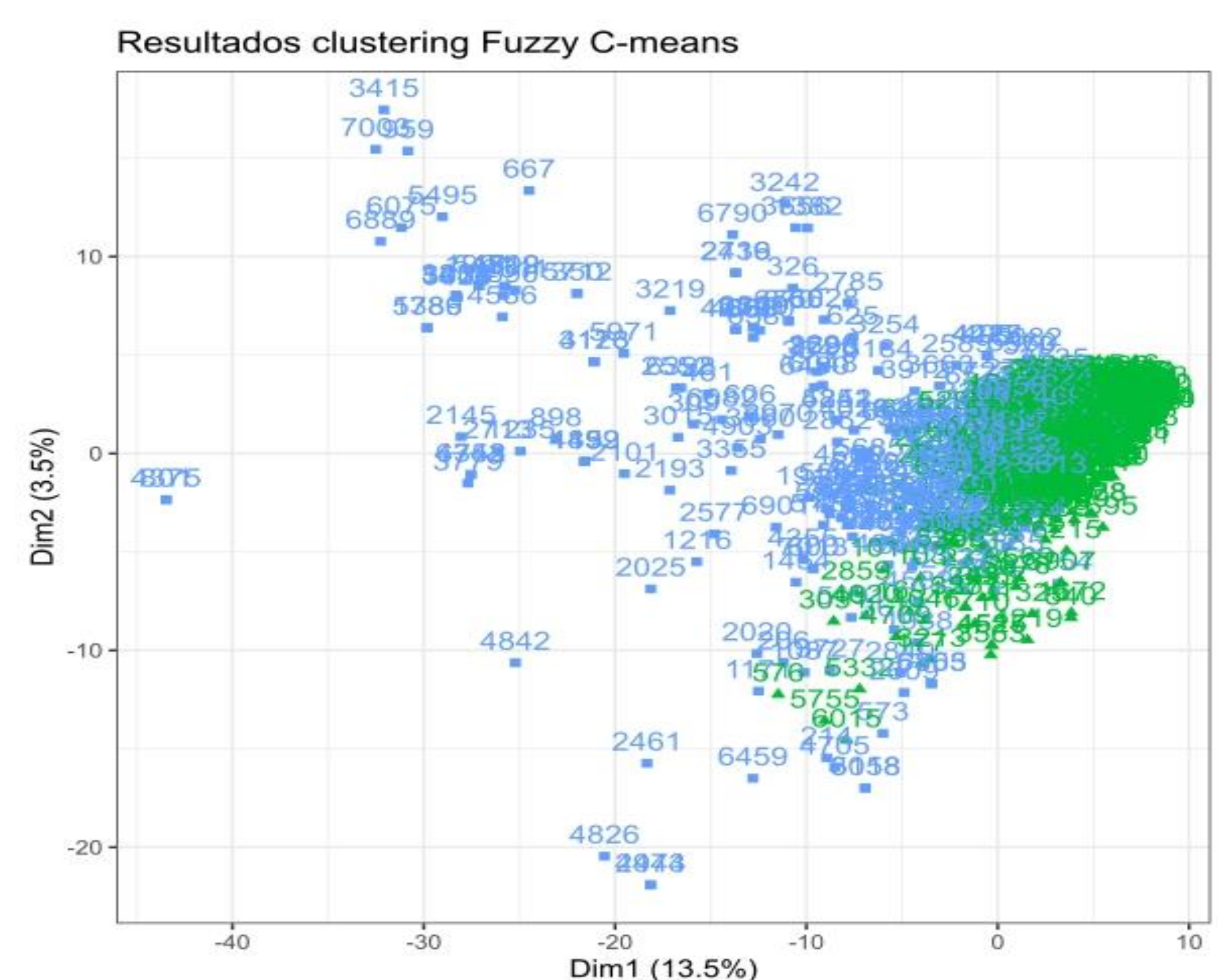


Figura 2: Cluster Fuzzy C-means.

## Objetivo

Evaluar y comparar el desempeño de los algoritmos de agrupamiento no supervisado K-means y Fuzzy C-means

## Metodología

Se obtuvo de Kaggle la base de datos Human Faces compuesta de 7000 imágenes.

Se empleó la red VGG16 pre-entrenada para la extracción de características.

Se seleccionaron al azar 800 imágenes vectorizadas.

Se determinó el número óptimo de clusters para K-means y Fuzzy C-means

Obtención de clusters y métrica de separación intragrupo.

ALGORITMO	G1	G2	G3	INTRAVARIANZA
K-MEANS	294	419	87	2048246
Fuzzy C-means	1	515	284	1149.789

Tabla 1: Resultados comparativos

## Conclusiones

Se encontró que el algoritmo de Fuzzy C-means se desempeña mejor en casos como este donde los datos resultan ser muy ruidosos, esto se puede contrastar mediante la minimización de la intravarianza respecto al algoritmo de K-means.

## Referencias

- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer, New York.  
Edla, D. R., Lone, T., Tapas, N., and Kuppili, V. (2020). Analysis of high dimensional brain data using prototype based fuzzy clustering. Clinical Epidemiology and Global Health, 8(4):1110-1118.  
Johnson, R. A. and Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. Pearson, Upper Saddle River, NJ, 6th edition. Nascimento, S., Mirkin, B., and Moura-Pires, F. (2000). A fuzzy clustering model of data and fuzzy c-means. In Ninth IEEE International Conference on Fuzzy Systems. FUZZ- IEEE 2000 (Cat. No.00CH37063), volume 1, pages 302-307, San Antonio, TX, USA. IEEE.