Chapter 5

Performance Analysis of Type-1 and Type-2 Fuzzy Clustering Algorithms for Digital Capability-Based Synthetic Data

Mükerrem Bahar Başkır[1]

Abstract

In this study, type-1 fuzzy c-means and its interval and general type-2 versions were applied to a digital capability-based synthetic data. The synthetic data were generated via a Monte Carlo method under multivariate normality parameterized by an empirical mean vector and a shrunk covariance matrix. The performances of type-1 and type-2 fuzzy clustering algorithms were evaluated using a cluster validity index and discriminant analysis. Kim and Ramakrishna's validity index was utilized to determine the optimal fuzzy partitions of the synthetic data. The synthetic data was partitioned into three fuzzy clusters. Discriminant analysis was conducted to assess the separability of the obtained fuzzy clusters. Consequently, the type-1, interval type-2, and general type-2 fuzzy clustering algorithms achieved the highest group case proportions under "Small", "Medium", and "High" fuzzifier settings, respectively.

1. Introduction

Digitalization is a key strategic indicator for organizational decision-making. Organizations focus on measuring and evaluating their digital growth and maturity using objective and data-driven approaches. There are various data-driven methods to improve the strategic decision-making processes. A system considered within a decision problem contains epistemic uncertainties due to both the system–components' crisp structures and the interactions/relations among these components. Zadeh's (1965) fuzzy set theory–based approaches analyze the epistemic uncertainties addressed in decision problems.

1 Assoc.Prof.Dr., Bartın University, mbaskir@bartin.edu.tr, https://orcid.org/0000-0002-1107-0659

https://doi.org/10.58830/ozgur.pub1148.c4769

Fuzzy clustering is an unsupervised machine learning task under fuzzy environment. This unsupervised clustering task is commonly used in modeling and eliminating the uncertainties in complicated decision systems. As the well-known fuzzy clustering algorithm, type-1 fuzzy c-means (T1-FCM) (Dunn, 1973; Bezdek, 1981) assigns observations (data points) to multiple clusters based on their membership degrees. Numerous type-1 fuzzy clustering algorithms were developed for different purposes (e.g., Hathaway & Bezdek, 1993; Höppner & Klawonn, 2003; Pedrycz, 2004; Pal et al., 2005; Chaira et al., 2007; Celikyilmaz & Turksen, 2008b; Baskir & Turksen, 2013). There were enhanced various type-2 fuzzy clustering algorithms to overcome the uncertainties arising from the crisp primary memberships produced by T1-FCM (e.g., Hwang & Rhee, 2007; Linda & Manic, 2012; Raza & Rhee, 2012; Turksen, 2014, Nguyena et al., 2014; Rubio et al., 2017; Zhao et al., 2019; Yang et al., 2021; Baskir 2022; Wu & Peng, 2023; Huang et al., 2024; Kchaou et al., 2025). The fuzzifier parameter m in T1-FCM has a significant role in shaping an appropriate fuzzy partition structure. In this context, Hwang and Rhee (2007) introduced an interval type-2 fuzzy c-means (IT2-FCM) algorithm that utilizes an interval-valued fuzzifier $m \in [m_L, m_R]$ to capture uncertainty in T1-FCM-based fuzzification process. As an extension of IT2-FCM, a general type-2 fuzzy c-means (GT2-FCM) was proposed by Linda and Manic (2012). GT2-FCM was structured using linguistic fuzzifier M obtained by type-1 fuzzy sets (T1FSs) and α-planes description of general type-2 fuzzy sets to calculate its secondary memberships. Type-reduction and center-updating procedures can be performed by using the enhanced Karnik-Mendel algorithm (Wu & Mendel, 2009). Type-1 and type-2 fuzzy clustering algorithms discover the intrinsic data structure without label information. Due to the unsupervised nature of these algorithms, the suitable fuzzy partition of the data is determined using cluster validity indices (CVIs). CVIs (e.g., Bezdek's Partition Coefficient and Partition Entropy (1973, 1974, 1981), Xie & Beni (1991), Kim & Ramakrishna (2005), Çelikyılmaz & Türksen (2008a, 2008b), Baskir & Turksen, 2013, among others) quantitatively evaluate intra-cluster compactness and inter-cluster separation of the fuzzy partition.

In this study, a comparative performance analysis of T1-, IT2-, and GT2-FCM algorithms was conducted on digital capability-based synthetic data. The performance of these algorithms was tracked using the synthetic data. The digital capability-based synthetic data were generated according to the empirical mean and shrinkage covariance of Aramburu et al.'s (2021) original data. The clustering structure-qualities of these algorithms were assessed using Kim-Ramakrishna's validity index. Additionally, the accuracy of cluster assignments and the overall classification performance was assessed through discriminant analysis to validate the fuzzy clustering results. The highlights of this study are as follows:

- The synthetic data were generated via Monte Carlo approach under the assumption of multivariate normality parameterized empirical mean and covariance of the digital capability-based original data.
- T1-, IT2-, and GT2-FCM were comparatively evaluated using the internal validity index and discriminant analysis.
- The results revealed similar and/or distinct behaviors of type-1 and type-2 fuzzy clustering algorithms in terms of their membership patterns.

The remainder of this study is organized as follows: Section 2 presents the theoretical frameworks of type-1 and type-2 fuzzy-based clustering algorithms, and the mathematical structure of the cluster validity index. The performance results of the relevant algorithms are presented in Section 3. The conclusion is given in Section 4.

2. Materials And Methods

This section presents the analytical structures/procedures of the type-1 and type-2 fuzzy clustering algorithms and the cluster validity index.

2.1. From Classical To Type-1 Fuzzy Clustering

Clustering is a frequently used machine learning task that partitions any given input data into clusters based on their similarity and/or dissimilarity patterns. The main goal is to maximize similarity within clusters and maximize dissimilarity between clusters. Classical clustering recognizes data patterns by assigning each data point to a specific cluster based on distance measures/similarity metrics. The well-known classical clustering task is the *k*-means algorithm. The *k*-means was introduced by MacQueen (1967) to partition a given data into *k* clusters, such that each data point belongs to only one cluster. Under this algorithm, the boundaries of the clusters are distinctly defined. The *k*-means algorithm aims to obtain homogeneous and well-separated cluster structures by minimizing the sum of squared errors within each cluster. However, data points located in boundary regions may exhibit similarities to multiple clusters. This situation leads to limitations in classical clustering algorithms due to their rigid assignment mechanisms. To eliminate the limitations of classical clustering, fuzzy clustering was developed to construct flexible and soft cluster structures.

The conventional type-1 fuzzy clustering was first introduced by Dunn (1973), and then developed by Bezdek (1981) as an optimization problem in Eq. (1), where J is the minimized objective function, u_{ij} is the membership degree, v_i is the cluster center, c and m are cluster number and fuzzifier, respectively, and ||.|| is the Euclidean norm.

$$\min J(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} \mu_{ik}^{m}\left(\left\|x_k - v_i\right\|\right)$$
$$\text{sbj to:} 0 \le \mu_{ik} \le 1,\ \forall i,k;\ \sum_{i=1}^{c}\mu_{ik} = 1,\ \forall k;\ 0 \le \sum_{k=1}^{n}\mu_{ik} \le n,\ \forall i \quad (1)$$

The solutions (center and membership function) of Eq. (1) are given in Eqs. (2)-(3):

$$\text{Center: } v_i = \frac{\sum_{k=1}^{n}(\mu_{ik})^m x_k}{\sum_{k=1}^{n}(\mu_{ik})^m} \quad (2)$$

$$\text{Membership function: } \mu_{ik} = \left[\sum_{j=1}^{c}\left(\frac{\left\|x_k - v_i\right\|}{\left\|x_k - v_j\right\|}\right)^{2/(m-1)}\right]^{-1}, i \ne j \quad (3)$$

The pseudo-code of T1-FCM is given in Algorithm I.

Algorithm I. T1-FCM algorithm.

Input: Input data ($X = \{x_1, \ldots x_n\}$), cluster number (c), fuzzifier (m), error (ε), iteration number ($iter$)
Output: Membership matrix (U^*), Center vector (V^*)
1. Initial cluster centers are randomly determined.
2. Repeat $t < iter$:
3. For $i = 1,2, \ldots, n$
4. For $k = 1,2, \ldots, c$
5. Memberships are calculated using Eq. (3).
6. Centers are computed using Eq. (2).
7. If $\parallel v^{(t)} - v^{(t-1)} \parallel < \varepsilon$ then Stop.
8. End
9. End

2.2. Interval Type-2 Fuzzy Clustering Algorithm

Hwang and Rhee (2007) proposed the interval type-2 fuzzy c-means (IT2-FCM) algorithm by defining the fuzzifier m in T1-FCM as an interval-valued parameter $[m_L, m_R]$. The objective function of T1-FCM is restructured using m_L and m_R, separately, to obtain the objective functions of IT2-FCM. The solutions of IT2-FCM objective functions are given in Eqs. (4)-(6):

$$\text{Center: } \bar{v}_i = \left[v_i^L, v_i^R\right] = \sum_{u \in J_{x_1}} \ldots \sum_{u \in J_{xn}} 1 / \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_k}{\sum_{k=1}^{n} (\mu_{ik})^m} \tag{4}$$

$$\text{Memberships: } \bar{\mu}_{ik,t} = \max(\left[\sum_{j=1}^{c}\left(\frac{\|x_k - v_{i,t-1}\|}{\|x_k - v_{j,t-1}\|}\right)^{2/(m_L-1)}\right]^{-1}, \left[\sum_{j=1}^{c}\left(\frac{\|x_k - v_{i,t-1}\|}{\|x_k - v_{j,t-1}\|}\right)^{2/(m_R-1)}\right]^{-1}) \tag{5}$$

$$\underline{\mu}_{ik,t} = \min(\left[\sum_{j=1}^{c}\left(\frac{\|x_k - v_{i,t-1}\|}{\|x_k - v_{j,t-1}\|}\right)^{2/(m_L-1)}\right]^{-1}, \left[\sum_{j=1}^{c}\left(\frac{\|x_k - v_{i,t-1}\|}{\|x_k - v_{j,t-1}\|}\right)^{2/(m_R-1)}\right]^{-1}) \tag{6}$$

The crisp representative centers and memberships are obtained as in Eqs. (7)-(8), respectively.

$$v_i = \frac{v^L + v^R}{2} \quad (7)$$

$$\mu_{ik} = \frac{\mu_{ik}^L + \mu_{ik}^R}{2} \quad (8)$$

Type-reduction and cluster center updating procedures can be examined using the enhanced Karnik-Mendel (E-KM) algorithm (Wu & Mendel, 2009).

The pseudo-code of IT2-FCM is given in Algorithm II.

Algorithm II. IT2-FCM algorithm.

Input: Input data ($X = \{x_1, \dots x_n\}$), cluster number (c), fuzzifier (m=[m_L,m_R]), error (ε), iterarion number ($iter$)

Output: Lower and Upper Membership matrices ($\underline{U}$ and $\overline{U}$), Interval Centers ($\underline{V}$, $\overline{V}$)

1. Initial interval centers are randomly generated.
2. Repeat $t < iter$:
3. For $i = 1,2, \dots, n$
4. For $k = 1,2, \dots, c$
5. Lower and upper memberships are updated using Eqs. (5)-(6).
6. Interval centers are calculated using Eq. (4).
7. Type-reduction procedure can be performed using the E-KM algorithm.
8. If $\| v^{(t)} - v^{(t-1)} \| < \varepsilon$ then Stop.
9. End
10. End

2.3. General Type-2 Fuzzy Clustering Algorithm

Linda and Manic (2012) enhanced a general type-2 fuzzy c-means (GT2-FCM) algorithm by defining the linguistic fuzzifier M. GT2-FCM comprises α-planes of general type-2 memberships, which can be generated by α-cuts of M. The secondary memberships in GT2-FCM can be expressed as in Eq. (9):

$$\tilde{\mu}_{ji} = \bigcup_{\alpha\in[0,1]} \alpha / S_{\tilde{\mu}_j}(x_i|\alpha) = \bigcup_{\alpha\in[0,1]} \alpha / \left[s^L_{\tilde{\mu}_j}(x_i|\alpha), s^R_{\tilde{\mu}_j}(x_i|\alpha) \right] \tag{9}$$

where $s^L_{\tilde{\mu}_j}(x_i|\alpha)$ and $s^R_{\tilde{\mu}_j}(x_i|\alpha)$ are left and right boundaries for α-planes of GT2 memberships in Eq. (10):

$$\tilde{\mu}_j(\alpha) = \sum_{x_i\in X} S_{\tilde{\mu}_j}(x_i|\alpha) \tag{10}$$

where

$$s^R_{\tilde{\mu}_j}(x_i|\alpha) = \max\left(\left[\sum_{j=1}^{c} \left(\frac{\|x_k - v_{i,t-1}\|}{\|x_k - v_{j,t-1}\|} \right)^{2/(s^L_M(\alpha)-1)} \right]^{-1}, \left[\sum_{j=1}^{c} \left(\frac{\|x_k - v_{i,t-1}\|}{\|x_k - v_{j,t-1}\|} \right)^{2/(s^R_M(\alpha)-1)} \right]^{-1} \right)$$

$$s^L_{\tilde{\mu}_j}(x_i|\alpha) = \min\left(\left[\sum_{j=1}^{c} \left(\frac{\|x_k - v_{i,t-1}\|}{\|x_k - v_{j,t-1}\|} \right)^{2/(s^L_M(\alpha)-1)} \right]^{-1}, \left[\sum_{j=1}^{c} \left(\frac{\|x_k - v_{i,t-1}\|}{\|x_k - v_{j,t-1}\|} \right)^{2/(s^R_M(\alpha)-1)} \right]^{-1} \right)$$

The fuzzy cluster-centre can be obtained as in Eq. (11):

$$\tilde{v}_j = C_{\tilde{\mu}_j} = \sum_{\tilde{\mu}\in J_{x_1}} \dots \sum_{\tilde{\mu}\in J_{xn}} (f_{\tilde{\mu}_j}(x_1) * \dots * f_{\tilde{\mu}_j}(x_N)) / \frac{\sum_{k=1}^{n} \left(\tilde{\mu}_{jk}\right)^m x_k}{\sum_{k=1}^{n} \left(\tilde{\mu}_{jk}\right)^m}, \tag{11}$$

$$where\ C_{\tilde{\mu}_j} = \bigcup_{\alpha\in[0,1]} \alpha / \left[c^L_{\tilde{\mu}_j}(\alpha), c^R_{\tilde{\mu}_j}(\alpha) \right]$$

The centroid $C_{\tilde{\mu}_j}$ is calculated using Liu's theorem (2008), where $\left[c^L_{\tilde{\mu}_j}(\alpha), c^R_{\tilde{\mu}_j}(\alpha) \right]$ is the individual interval center.

The precise cluster-centre is computed as in Eq. (12), where K is the number of discretized type-2 fuzzy sets:

$$v_j = \frac{\sum_{i=1}^{K} C_{\tilde{\mu}_j}(y_i)\, y_i}{\sum_{i=1}^{K} C_{\tilde{\mu}_j}(y_i)} \tag{12}$$

Type-reduction and center-updating procedures can be performed using the E-KM algorithm. Defuzzification approach of IT2-FCM is extended for GT2-FCM using Liu's theorem.

The pseudo-code of GT2-FCM is given in Algorithm III.

Algorithm III. GT2-FCM algorithm.

Input: Input data ($X = \{x_1, \dots x_n\}$), cluster number (c), linguistic fuzzifier (M), zSlice number (S) or α-planes set, error (ε), iteration number ($iter$)

Output: GT2-memberships matrix $(\tilde{U}^*)$, type-reduced cluster centers (V^*)

1. Initial cluster prototypes are randomly determined.
2. Repeat $t < iter$:
3. For $s = 1,2, \dots, S$
4. For $i = 1,2, \dots, n$
5. For $k = 1,2, \dots, c$
6. Update GT2 memberships via zSlices / α-planes.
7. Update GT2 cluster prototypes (centroids) per slice.
8. Type-reduction across slices (e.g., the E-KM algorithm)
9. If $\| v^{(t)} - v^{(t-1)} \| < \varepsilon$ then Stop.
10. End
11. End
12. End

2.4. Fuzzy Cluster Validity Indices

The optimal fuzzy partition of an input data can be obtained by selecting the appropriate cluster number and fuzzifier parameters in fuzzy clustering algorithms. The cluster validity indices (CVIs) were enhanced to select the optimal fuzzy clustering parameters. There were enhanced numerous CVIs (e.g., Bezdek's indices (1973, 1974, 1981), Xie & Beni (1991), Kim & Ramakrishna (2005), Çelikyılmaz & Türksen (2008a, 2008b), Baskir & Turksen, 2013, among others). These CVIs vary depending on the geometric structure of the data. The Kim-Ramakrishna's (KR) validity index in Eq. (13) was used in this study.

$$v_{KR}(u) = \left\{ \frac{\max\limits_{i=1,\ldots,c} \left\{ \frac{1}{n} \sum_{k=1}^{n} u_{ik}^{2} \left\| x_k - v_i \right\|^2 \right\}}{\left(\min\limits_{i \neq j} \left\{ \left\| v_i - v_j \right\|^2 \right\} \right)} \right\} \tag{13}$$

3. Results And Discussion

This section presents discriminant-analysis based performance evaluations of the type-1 and type-2 fuzzy clustering algorithms.

3.1. Digital Capability-Based Synthetic Data

Digitalization is a crucial indicator for making strategic organizational decisions. In this context, the DIGROW framework for digital maturity (North et al., 2020) guides firms through two key assessment processes (Aramburu et al., 2021): i) Evaluating their digital maturity level, ii) assessing the capabilities associated with each maturity level to support digitally enabled growth. The DIGROW-stages and their capacities are given in Figure 1:

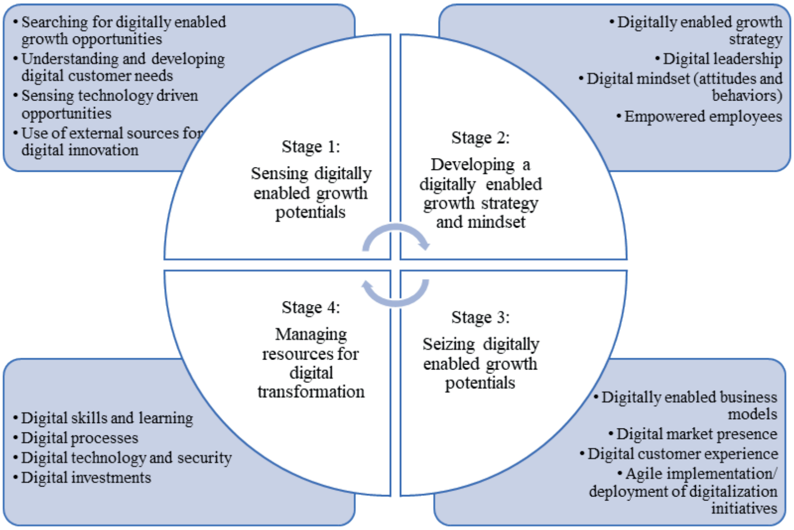


Fig. 1. The DIGROW stages (Aramburu et al., 2021; Di Felice et al., 2022).

In this study, T1-, IT2-, and GT2-FCM algorithms were performed on digital capability-based synthetic (S_DC) data. The original data released by Aramburu et al. in 2021 for researchers was utilized to create the S_DC data. First, the overall scores for each DIGROW stage in the original data were calculated by averaging the scores of the four relevant capacities. Then, the S_DC data were generated using a parametric Monte Carlo approach assuming multivariate normality. The mean vector and covariance structure were estimated from the overall score-based original data. To mitigate the effects of multicollinearity in discriminant analysis, a shrinkage-based covariance adjustment was employed. This procedure preserves the empirical mean structure while systematically reducing interdimensional correlations. The S_DC data size was set to 400 to ensure clustering stability and maintain consistency with the original data size.

3.2. Type-1 and Type-2 Fuzzy Clustering Results

The digital capability-based synthetic overall scores (400×4-dimensional S_DC data) were classified using T1-, IT2-, and GT2-FCM. Fuzzifier settings of T1-, IT2-, and GT2-FCM are given in Table 1:

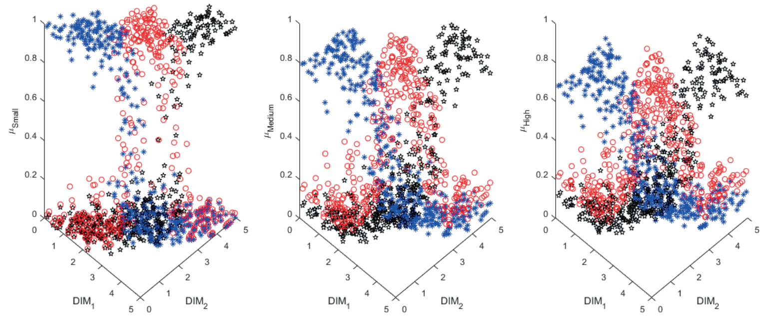
Table 1. Fuzzifier settings of type-1 and type-2 fuzzy c-means.

Algorithm	Fuzzifier settings
T1-FCM	Small (m=1.4), Medium (m=2.0), High (m=2.6)
IT2-FCM	Small ($[m_L,m_R]=[1.4,2.2]$), Medium ($[m_L,m_R]=[1.6,2.4]$), High ($[m_L,m_R]=[1.8,2.6]$)
GT2-FCM	Small (*M*: Gaussian T1FS with the parameters (1.6, 0.06)), Medium (*M*: Gaussian T1FS with the parameters (2.0, 0.1)), High (*M*: Gaussian T1FS with the parameters (2.4, 0.3))

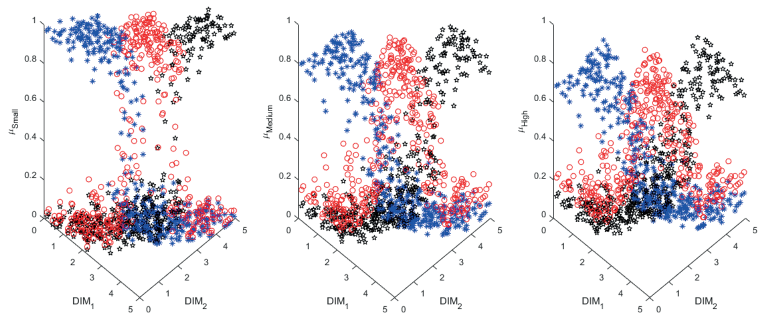
The KR validity index (as shown in Eq. (13)) results for the T1-, IT2-, and GT2-FCM algorithms are presented in Table 2. For all fuzzifiers considered for the type-1 and type-2 fuzzy clustering algorithms, the KR validity index consistently identified three as the optimal cluster number. T1-, IT2-, and GT2-FCM-based clustering structures are illustrated in Figure 2.

Table 2. KR-index values for the synthetic data.

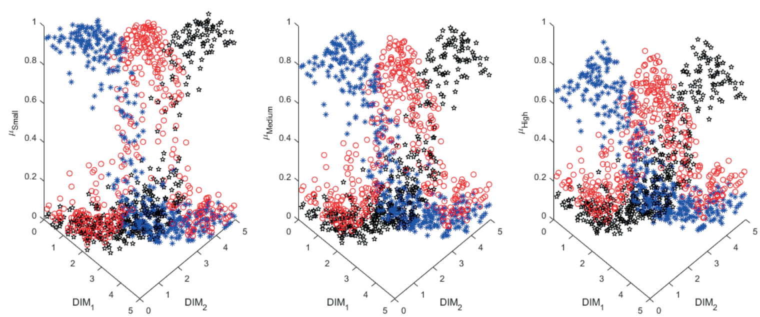
KR-index values		c=2	c=3	c=4	c=5	c=6	c=7	c=8
T1-FCM	Small (m=1.4)	0.08	**0.08**	0.09	0.088	0.098	0.089	0.058
	Medium (m=2.0)	0.066	**0.059**	0.06	0.074	0.065	0.091	0.075
	High (m=2.6)	0.072	**0.068**	0.068	0.101	0.109	0.186	0.237
IT2-FCM	Small ($[m_L,m_R]=[1.4,2.2]$)	0.079	**0.079**	0.089	0.087	0.098	0.078	0.054
	Medium ($[m_L,m_R]=[1.6,2.4]$)	0.066	**0.059**	0.06	0.075	0.064	0.088	0.073
	High ($[m_L,m_R]=[1.8,2.6]$)	0.072	**0.067**	0.068	0.104	0.108	0.179	0.313
GT2-FCM	Small (M: Asym. Gaussian T1FS)	0.018	**0.018**	0.019	0.019	0.02	0.019	0.014
	Medium (M: Sym. Gaussian T1FS)	0.017	**0.015**	0.015	0.019	0.016	0.022	0.019
	High (M: Asym. Gaussian T1FS)	0.017	**0.016**	0.016	0.023	0.023	0.034	0.043



(a)



(b)



(c)

Fig. 2. (a) T1-, (b) IT2-, (c) GT2-FCM representations (*: Cluster-1, o: Cluster-2, ♦: Cluster-3)

The hard partitioning procedures were executed for the relevant algorithms by selecting the highest membership degrees.

3.3. Discriminant-Based Performance Analysis

The discriminant analysis was used to assess the separability of each fuzzy clustering structure. The multivariate normality assumption of the digital capability-based synthetic (S_DC) data was provided according to the Mardia (1970) test results (p-values > 0.01 for multivariate skewness and kurtosis tests). The results of Box's M tests in Table 3 indicate no statistically significant difference among the covariance matrices of the relevant clusters. Therefore, linear discriminant analysis (LDA) was conducted for T1-, IT2-, and GT2-FCM models of the S_DC data. The first discriminant functions obtained for fuzzy clustering algorithms were statistically significant (p-values=0.000).

Table 3. LDA results.

S_DC data		Box's M	F	p-value
T1-FCM	SMALL	26.831	1.321	0.153
	MEDIUM	23.377	1.151	0.288
	HIGH	27.450	1.352	0.134
IT2-FCM	SMALL	26.831	1.321	0.153
	MEDIUM	24,623	1.212	0.232
	HIGH	27.450	1.352	0.134
GT2-FCM	SMALL	28.345	1.395	0.112
	MEDIUM	22.527	1.109	0.331
	HIGH	25.041	1.233	0.215

The equality tests of group (cluster) means for type-1 and type-2 fuzzy clustering models are given in Table 4. According to Table 4, the S_DC stages (independent variables) were found to have statistically significant mean differences across the groups (clusters) of the dependent variable (p-values=0.000). Discriminant analysis and variance analysis revealed similar results.

Table 4. The group (cluster) means – equality tests.

Model	S_DC Dimension	Wilks' λ	F	p-value
T1-FCM (SMALL)	DIM1_score	0.329	405.173	0.000
	DIM2_score	0.230	663.346	0.000
	DIM3_score	0.214	730.772	0.000
	DIM4_score	0.294	477.533	0.000
T1-FCM (MEDIUM)	DIM1_score	0.325	412.884	0.000
	DIM2_score	0.221	701.113	0.000
	DIM3_score	0.224	685.971	0.000
	DIM4_score	0.301	460.156	0.000
T1-FCM (HIGH)	DIM1_score	0.318	425.030	0.000
	DIM2_score	0.222	697.414	0.000
	DIM3_score	0.230	666.201	0.000
	DIM4_score	0.306	450.010	0.000
IT2-FCM (SMALL)	DIM1_score	0.329	405.173	0.000
	DIM2_score	0.230	663.346	0.000
	DIM3_score	0.214	730.772	0.000
	DIM4_score	0.294	477.533	0.000
IT2-FCM (MEDIUM)	DIM1_score	0.328	406.237	0.000
	DIM2_score	0.222	694.671	0.000
	DIM3_score	0.218	710.902	0.000
	DIM4_score	0.301	460.168	0.000
IT2-FCM (HIGH)	DIM1_score	0.318	425.030	0.000
	DIM2_score	0.222	697.414	0.000
	DIM3_score	0.230	666.201	0.000
	DIM4_score	0.306	450.010	0.000
GT2-FCM (SMALL)	DIM1_score	0.327	407.651	0.000
	DIM2_score	0.231	661.458	0.000
	DIM3_score	0.215	725.677	0.000
	DIM4_score	0.293	478.475	0.000
GT2-FCM (MEDIUM)	DIM1_score	0.318	426.121	0.000
	DIM2_score	0.224	689.325	0.000
	DIM3_score	0.225	694.964	0.000
	DIM4_score	0.305	453.309	0.000
GT2-FCM (HIGH)	DIM1_score	0.316	429.092	0.000
	DIM2_score	0.221	699.482	0.000
	DIM3_score	0.228	673.094	0.000
	DIM4_score	0.312	438.246	0.000

The cross-validation results obtained from discriminant analysis of T1-, IT2-, and GT2-fuzzy clustering models are given in Tables 5-7, respectively. According to Tables 5-7, the high cross-validation rates (proportions) indicate that the clusters are highly separable and statistically consistent in the S_DC feature spaces.

Table 5. Cross-validation rates of the T1-FCM-based cluster-assignments

T1-FCM (SMALL)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3
Original Group	Cluster-1	149	1	0
	Cluster-2	3	154	1
	Cluster-3	0	1	91
Assignment-rates for the clusters		99.3%	97.5%	98.9%
n=400, n (Correct assignment)=394, Cross-validated group cases-rate=98.5%				
T1-FCM (MEDIUM)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3
Original Group	Cluster-1	139	4	0
	Cluster-2	3	158	0
	Cluster-3	0	3	93
Assignment-rates for the clusters		97.2%	98.1%	96.9%
n=400, n (Correct assignment)=390, Cross-validated group cases-rate=97.5%				
T1-FCM (HIGH)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3
Original Group	Cluster-1	136	3	0
	Cluster-2	2	158	1
	Cluster-3	0	5	95
Assignment-rates for the clusters		97.8%	98.1%	95.0%
n=400, n (Correct assignment)=389, Cross-validated group cases-rate=97.3%				

Table 6. Cross validation rates of the IT2-FCM-based cluster-assignments

IT2-FCM (SMALL)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3
Original Group	Cluster-1	149	1	0
	Cluster-2	3	154	1
	Cluster-3	0	1	91
Assignment-rates for the clusters		99.3%	97.5%	98.9%
n=400, n (Correct assignment)=394, Cross-validated group cases-rate=98.5%				
IT2-FCM (MEDIUM)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3
Original Group	Cluster-1	144	1	0
	Cluster-2	5	155	0
	Cluster-3	0	3	92
Assignment-rates for the clusters		99.3%	96.9%	96.8%
n=400, n (Correct assignment)=391, Cross-validated group cases-rate=97.8%				
IT2-FCM (HIGH)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3

Original Group	Cluster-1	136	3	0
	Cluster-2	2	158	1
	Cluster-3	0	5	95
Assignment-rates for the clusters		97.8%	98.1%	95.0%
n=400, n (Correct assignment)=389, Cross-validated group cases-rate=97.3%				

Table 7. Cross validation rates of the GT2-FCM-based cluster-assignments

GT2-FCM (SMALL)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3
Original Group	Cluster-1	149	1	0
	Cluster-2	3	153	1
	Cluster-3	0	2	91
Assignment-rates for the clusters		99.3%	97.5%	97.8%
n=400, n (Correct assignment)=393, Cross-validated group cases-rate=98.3%				
GT2-FCM (MEDIUM)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3
Original Group	Cluster-1	137	5	0
	Cluster-2	4	159	0
	Cluster-3	0	3	92
Assignment-rates for the clusters		96.5%	97.5%	96.8%
n=400, n (Correct assignment)=388, Cross-validated group cases-rate=97.0%				
GT2-FCM (HIGH)		Predicted Group		
		Cluster-1	Cluster-2	Cluster-3
Original Group	Cluster-1	135	1	0
	Cluster-2	2	161	1
	Cluster-3	0	5	95
Assignment-rates for the clusters		99.3%	98.2%	95.0%
n=400, n (Correct assignment)=391, Cross-validated group cases-rate=97.8%				

4. Conclusion

This study presents a comparative performance analysis of type-1, interval type-2, and general type-2 fuzzy c-means algorithms for digital capability-based synthetic (S_DC) data. A Monte Carlo approach was employed to generate the S_DC data under the assumption of multivariate normality. The empirical mean vector and a shrinkage-adjusted covariance matrix derived from the original data were utilized as distributional parameters. The appropriate fuzzy partitions of the S_DC data were determined using Kim-Ramakrishna's validity index. The optimal cluster number was determined to be three. Under the fuzzifier settings (Table 1), the behaviors of the T1-, IT2-, and GT2-FCM algorithms for three clusters were assessed by their cross-validated cluster assignment-based and overall group case-based proportions calculated through discriminant analysis. Under the "Small" fuzzifier settings, the three algorithms yielded highly similar group cases-proportions. Additionally, the three algorithms achieved the similar/close cluster assignment proportions in Cluster-1/Cluster-2 (fuzzifier settings "Small"), in Cluster-3 (fuzzifier settings "Medium"/"High"). This may indicate that the assignment changes occurred at boundary data points, which exhibit sensitive assignment behaviors to small perturbations based on the relevant membership structure. T1- and IT2-FCM algorithms defined under the "Small" fuzzifier settings attained the highest assignment proportions for Cluster-3. Under the "Medium" setting, IT2-FCM achieved the highest proportions in Cluster-1, while T1-FCM exhibited its highest values in Cluster-2. GT2-FCM under the "High" fuzzifier settings demonstrated the highest cluster assignment proportions across Clusters 1 and 2. T1-, IT2-, and GT2-FCM algorithms achieved the highest group case proportions under "Small", "Medium", and "High" fuzzifier settings, respectively.

In future studies, clustering algorithms based on type-1 fuzzy sets and their several extensions (e.g., intuitionistic, picture, hesitant fuzzy sets) will be comparatively investigated. The effects of these algorithms on fuzzy system modeling will be analysed for big data analytics-based applications such as traffic congestion prediction, energy efficiency analysis, climate change and environmental risk modeling, disease diagnosis, and genomic analysis.

References

Baskir, M.B. & Turksen, I.B. (2013). Enhanced fuzzy clustering algorithm and cluster validity index for human perception. *Expert Systems with Applications, 40(3),* 929-937. https://doi.org/10.1016/j.eswa.2012.05.049

Baskir, M.B. (2022). An adaptive self-reduction type-2 fuzzy clustering algorithm for pattern recognition. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 30(06)*, 991-1017. https://doi.org/10.1142/S0218488522500301

Bezdek, J.C. (1973). Cluster validity with fuzzy sets. *J. Cybernetic, 3*, 58-72. https://doi.org/10.1080/01969727308546047

Bezdek, J.C. (1974). Numerical taxonomy with fuzzy sets. *J.Math. Biol., 1*, 57-71. https://doi.org/10.1007/BF02339490

Bezdek, J.C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

Çelikyılmaz, A. & Türkşen, I.B. (2008a). Validation criteria for enhanced fuzzy clustering. *Pattern Recognition Letters, 29 (2),* 97-108. https://doi.org/10.1016/j.patrec.2007.08.017

Celikyilmaz, A. & Turksen, I.B. (2008b). Enhanced fuzzy system models with improved fuzzy clustering algorithm. *IEEE Transaction on Fuzzy Systems, 16,* 779-794. doi: 10.1109/TFUZZ.2007.905919

Chaira, T., Ray, A.K. & Salvetti, O. (2007). Intuitionistic fuzzy c means clustering in medical image segmentation. *Advances in Pattern Recognition, 1*, 226–230. https://doi.org/10.1142/9789812772381_0037

Di Felice, P., Paolone, G., Di Valerio, D., Pilotti, F., & Sciamanna, M. (2022). Transforming DIGROW into a multi-attribute digital maturity model: Formalization and implementation of the proposal. In O. Gervasi, B. Murgante, S. Misra, A. M. A. C. Rocha, & C. Garau (Eds.), *Computational science and its applications – ICCSA 2022 workshops* (Lecture Notes in Computer Science, Vol. 13378). Springer. https://doi.org/10.1007/978-3-031-10562-3_38

Dunn, J.C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters. *Journal of Cybernetics, 3*, 32-57. https://doi.org/10.1080/01969727308546046

Hathaway, R.J. & Bezdek, J.C. (1993). Switching regression models and fuzzy clustering. *IEEE Transactions on Fuzzy Systems, 1,* 195-203.

Höppner, F. & Klawonn, F. (2003). Improved fuzzy partitions for fuzzy regression models. *International Journal of Approximate Reasoning,* 32, 85-102. https://doi.org/10.1016/S0888-613X(02)00078-6

Huang, C., Lei, H., Chen, Y., Cai, J., Qin, X., Peng, J., Zhou, L., & Zheng, L. (2024). Interval type-2 enhanced possibilistic fuzzy C-means noisy image segmentation algorithm amalgamating weighted local information. *Engineering Applications of Artificial Intelligence, 137(Part A),* 109135. https://doi.org/10.1016/j.engappai.2024.109135

Hwang, C. & Rhee, F.C.H. (2007). Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means. *IEEE Trans. Fuzzy Syst., 15,* 107–120. doi: 10.1109/TFUZZ.2006.889763

Kchaou, H., Abbes, W., Kechaou, Z., & Alimi, A. M. (2025). Interval type-2 fuzzy c-means collaborative clustering approach for scientific cloud workflows. *International Journal of Computers and Applications, 47(4),* 388–397. https://doi.org/10.1080/1206212X.2025.2471885

Kim, M. & Ramakrishna, R.S. (2005). New indices for cluster validity assessment. *Pattern Recognition Letters, 26*, 2353-2363. https://doi.org/10.1016/j.patrec.2005.04.007

Linda, O. & Manic, M. (2012). General type-2 fuzzy c-means algorithm for uncertain fuzzy clustering. *IEEE Trans. Fuzzy Syst.,* 20, 883–897. doi: 10.1109/TFUZZ.2012.2187453

Liu, F. (2008). An efficient centroid type-reduction strategy for general type-2 fuzzy logic system. *Information Sciences, 178,* 2224–2236. https://doi.org/10.1016/j.ins.2007.11.014

MacQueen, J.B. (1967). Some methods for classification and analysis of multivariate observations. *Proc. Symp. Math. Statist. and Probability (5th)*, 281–297.

Mardia, K.V., (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519-530. https://doi.org/10.2307/2334770

Nguyena, D.D., Ngoa, L.T., & Watadab, J. (2014). A genetic type-2 fuzzy c-means clustering approach to M-FISH segmentation. *Journal of Intelligent & Fuzzy Systems, 27*, 3111–3122. https://dl.acm.org/doi/10.5555/2699159.2699194

North, K., Aramburu, N., and Lorenzo, O. (2020). Promoting digitally enabled growth in SMEs: a framework proposal. *J. Enterpr. Inf. Manag., 33*, 238–262. doi: https://doi.org/10.1108/JEIM-04-2019-0103

Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C. (2005). A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems, 13(4),* 517–530. https://doi.org/10.1109/TFUZZ.2004.840099

Pedrycz, W. (2004). Fuzzy clustering with knowledge-based guidance. *Pattern Recognition Letters, 25*, 469–480. https://doi.org/10.1016/j.patrec.2003.12.010

Raza, M.A. & Rhee, F.C.H. (2012). Interval type-2 approach to kernel possibilistic c-means clustering, in *Proc. IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE*, (Brisbane, Australia; June 2012), pp. 1–7. doi: 10.1109/FUZZ-IEEE.2012.6251233

Rubio, E., Castillo, O., Valdez, F., Melin, P., Gonzalez, C.I., & Martinez, G. (2017). An extension of the fuzzy possibilistic clustering algorithm using type-2 fuzzy logic techniques. *Advances in Fuzzy Systems*, 7094046. https://doi.org/10.1155/2017/7094046

Turksen, I.B. (2014). From type 1 to full type N fuzzy system models. *Journal of Multiple-Valued Logic and Soft Computing*, *22(4-6),* 543-560.

Xie, X.L. & Beni, G.A. (1991). Validity measure for fuzzy clustering. *IEEE Trans. Pattern and Machine Intelligence, 3(8)*, 841-846. doi: 10.1109/34.85677

Wu, D. & Mendel, J. M. (2009). Enhanced Karnik–Mendel algorithms. *IEEE Transactions on Fuzzy Systems, 17(4)*, 923–934. doi:10.1109/TFUZZ.2008.924329

Wu, C. & Peng, S. (2023). Robust interval type-2 kernel-based possibilistic fuzzy clustering algorithm incorporating local and non-local information. *Advances in Engineering Software, 176*, 103377. https://doi.org/10.1016/j.advengsoft.2022.103377

Yang, X., Yu, F., & Pedrycz, W. (2021). Typical characteristic-based type-2 fuzzy c-means algorithm. *IEEE Transactions on Fuzzy Systems, 29(5),* 1173–1187. https://doi.org/10.1109/TFUZZ.2020.2969907

Zadeh, L.A. (1965). Fuzzy sets. *Information Control, 8(3)*, 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhao, F., Chen, Y., Liu, H., & Fan, J. (2019). Alternate PSO-based adaptive interval type-2 intuitionistic fuzzy c-means clustering algorithm for color image segmentation. *IEEE Access,* 7, 64028 – 64039. doi:10.1109/ACCESS.2019.2916894