IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 26, 2019, accepted April 3, 2019, date of publication April 18, 2019, date of current version May 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2910195

Fuzzy K-Means Using Non-Linear S-Distance

ADITYA KARLEKAR[1], AYAN SEAL[2,3], (Senior Member, IEEE), ONDREJ KREJCAR[3], AND CONSUELO GONZALO-MARTIN[4]

[1]Hitkarini College of Engineering and Technology, Jabalpur 482005, India
[2]PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India
[3]Faculty of Informatics and Management, Center for Basic and Applied Research, University of Hradec Kralove, 50003 Hradec Kralove, Czech Republic
[4]Center for Biomedical Technology, Universidad Politecnica de Madrid, 28223 Madrid, Spain

Corresponding author: Ayan Seal (ayan.seal@gmail.com)

This work was supported by the project Smart Solutions in Ubiquitous Computing Environments, Grant Agency of Excellence, University of Hradec Kralove, Faculty of Informatics and Management, Czech Republic under Grant UHK-FIM-GE-2019. The work of A. Seal was supported by the Media Lab Asia, Ministry of Electronics and Information Technology, Government of India, for providing him young faculty research fellowship.

ABSTRACT A considerable amount of research has been done since long to select an appropriate similarity or dissimilarity measure in cluster analysis for exposing the natural grouping in an input dataset. Still, it is an open problem. In recent years, the research community is focusing on divergence-based non-Euclidean similarity measure in partitional clustering for grouping. In this paper, the Euclidean distance of traditional Fuzzy k-means (FKM) algorithm is replaced by the S-distance, which is derived from the newly introduced S-divergence. Few imperative properties of S-distance and modified FKM are presented in this study. The performance of the proposed FKM is compared with the conventional FKM with Euclidean distance and its variants with the help of several synthetic and real-world datasets. This study focuses on how the proposed clustering algorithm performs on the adopted datasets empirically. The comparative study illustrates that the obtained results are convincing. Moreover, the achieved results denote that the modified FKM outperforms some state-of-the-art FKM algorithms.

INDEX TERMS Fuzzy K-means clustering, S-distance, S-divergence.

I. INTRODUCTION

Clustering is an imperative unsupervised machine learning approach employed in identifying some inherent structure exists in a set of patterns or objects. The aim of cluster analysis is to split set of objects, commonly vectors in a multi-dimensional space, are grouped into subsets so that the objects in the same subset are similar in some perception and objects in different clusters are dissimilar in the same perception.

Different selection of measured data or features, proximity measures, clustering criteria, and clustering algorithms may lead to totally different clustering results. In this study, Fuzzy k-means (FKM) algorithm applies on some real and synthetic datasets, where clustering criteria is same throughout the study. It means the selection of proximity measure plays an imperative job to find the cluster structure in data [1]. Even though the Euclidean distance has been the standard of squared error distortion, a considerable amount of research has been done to introduce non-linear distance measures [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Manuel Rosa-Zurera.

Recently, researchers have been substituting the Euclidean distance in k-means by various non-linear distance metrics, some of them even non-metric i.e. do not follow the triangle inequality property [2]–[4]. One motivation is to introduce non-linearity in clustering may result in identifying more accurate cluster boundaries. One such attempt was made to introduce general Bregman divergence as the similarity measure in the k-means algorithm to increase its efficacy [2]. Some noteworthy divergence based similarity measures for clustering include [5]–[9].

II. CLUSTERING

In this section, we first provide a formal definition of the clustering problem. A brief introduction of the traditional FKM is also presented since we have compared the performance of the proposed FKM with traditional one.

A. BASIC PRINCIPLE

Clustering is the process of splitting $d-$dimensional m data-points or the observations, $A[= u_1, u_2, \ldots, u_m]$, in R^n_+ into ‘k’ groups of homogeneous data-points, $C[= (C_1, C_2, \ldots, C_k)]$ in such a way that the degree of strong association within group and weak association between different groups. Then

$$\begin{aligned} C_i &\neq \phi \quad for\ i = 1, \ldots, k, \\ C_i \cap C_j &= \phi \quad for\ i = 1, \ldots, k; j = 1, \ldots, k\ and\ i \neq j, \\ \cup_{i=1}^{k} C_i &= C \end{aligned}$$

B. FUZZY K-MEANS

From a machine learning perspective, clustering analysis will give us an accurate and deep understanding into hidden patterns of different groups. After clustering, the partition matrix would be expressed as $W(A)_{(k\times m)}$. It can further be represented as $W = [w_{ab}]$, where w_{ab} is the membership of a data-point, u_b, from cluster center, c_a, of cluster, C, and 'a' and 'b' vary from 1 to k and 1 to m respectively. The value of w_{ab} is 1 only when u_b belongs to c_a otherwise, w_{ab} would be 0 in case of crisp partitioning. Clustering is the subject of active research in several fields such as pattern recognition [10], image processing [11], [12] especially in satellite image analysis [13]–[17] and data mining [18], [19] such as scientific data exploration, information retrieval and text mining. In [20], Dalal and Harale presented clustering analysis techniques and divided them into three categories: partition based clustering [21], density based [22] and hierarchical clustering [23], which are further sub-divided and discussed in [19]. Interested readers are referred to explore [24]–[26] for studying clustering algorithms in detail. Undoubtedly, Fuzzy k-means (FKM) is one of the frequently used partition-based clustering algorithms in pattern recognition. This algorithm was introduced by Dunn in 1973 [27] and enhanced by Peizhuang in 1981 [28]. FKM performs grouping by exploring a set of fuzzy groups, W, and the associated group centers, C, that denote the structure of the data-point as best as possible iteratively. The FKM algorithm depends on the input provide by the user, which represents the number of clusters present in the data-points to be fuzzy clustered by minimizing the within group sum of squared error objective function, $E_s(W, C)$, which is shown in equation 1.

$$E_s(C, W; A) = \sum_{b=1}^{m}\sum_{a=1}^{k}(w_{ab})^s||u_b - c_a||^2, \quad 1 \leq s < \infty, \tag{1}$$

where the real number 's' is familiar as fuzziness coefficient. Moreover, it manages the influence of membership grades in the performance index. The division becomes fuzzier (not crisp) with the increasing of 's' and researchers proved that the FKM algorithm converges for any $s \in (1, \infty)$. The w_{ab} indicates the degree of membership of u_b in the cluster a, u_b is the b^{th} of d-dimensional features, c_a is the d-dimensional center of the cluster and $|| \cdot ||$ is some inner product induced norm expressing the similarity between any features and the center. The error function relies on W and C, subject to two constraints, which are shown in equations 2 and 3.

$$\sum_{a=1}^{k} w_{ab} = 1, \quad b = 1, 2, \ldots, m, \tag{2}$$

where $w_{ab} \in [0, 1]$, $a = 1, 2, \ldots, k$ and $b = 1, 2, \ldots, m$.

$$0 < \sum_{b=1}^{m} w_{ab} < k, \quad a = 1, 2, \ldots, k \tag{3}$$

Fuzzy division is done through an iterative approach by minimizing an error function $E_s(W, C)$ as depicted in equation 1, with the update of membership w_{ab} and the cluster centers c_a using equations 4 and 5 respectively.

$$w_{ab}^{(e+1)} = \frac{1}{\sum_{l=1}^{k}\left(\frac{u_b - c_a^{(e)}}{u_b - c_l^{(e)}}\right)^{\frac{2}{s-1}}} \tag{4}$$

$$c_a^{(e+1)} = \frac{\sum_{b=1}^{m}\left[w_{ab}^{(e+1)}\right]^s \cdot u_b}{\sum_{b=1}^{m}\left[w_{ab}^{(e+1)}\right]^s} \tag{5}$$

This iteration will terminate while $max_{ab}\{|w_{ab}^{(e+1)} - w_{ab}^{(e)}|\} < \epsilon$, where ϵ is a stopping criterion between 0 and 1, whereas 'e' depicts iteration or epoch. This process converges to a local minimum or a saddle point of $E_s(W, C)$.

In this work, S-distance is used, which is derived from the notion of S-divergence [3], [4], [29]. Various properties of this distance have been studied. All the experiments have been performed on some real and synthetic datasets. All the simulation results depict that the FKM using S-distance outperforms the traditional FKM algorithm with Euclidean-distance and along with different variant of FKM such as Weighting in FKM, Minkowski metric weighted FKM. Our claim has been validated by performing statistical analysis on the obtained results.

III. S-DISTANCE AND ITS PROPERTIES

The definition followed by different properties of S-distance is discussed in this section.

Definition 1: S-divergence is a metric stated over a set of all positive definite matrices of size $n \times n$, S_n, which could be computed by equation 6.

$$\partial_s(U, V) = log(|\frac{U+V}{2}|) - \frac{log(|U|) + log(|V|)}{2}, \tag{6}$$

where $|U|$ = determinant of U. An injective function is stated as $\psi : \mathbb{R}_+^n \to S_n$ such that $\psi(u) = \text{diag}(u_1, u_2, \ldots, u_m)$, where $u = (u_1, u_2, \ldots, u_m) \in \mathbb{R}_+^n$ is a real positive vector and $\mathbb{R}_+^n$ is $(0, \infty) \times (0, \infty) \times \ldots \times (0, \infty)(n \quad times)$. The definition of S-distance is as follows:

Definition 2: The S-distance between any two data-points, $u, v \in \mathbb{R}_+^n$, could be stated as a function $SD : \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R}_+ \cup \{0\}$, which can also be expressed by equation 7.

$$SD(u, v) = \partial_s(\psi(u), \psi(v)) \tag{7}$$

The S-distance metric satisfies the following properties:

Proposition 1: Non-negativity: $SD(u, v) \geq 0$.

Proposition 2: Symmetry: $SD(u, v) = SD(v, u)$.

Proof: $SD(u, v) = \partial_s(\psi(u), \psi(v)) = \partial_s(\psi(v), \psi(u)) = SD(v, u)$.

Proposition 3: $SD(u, v) \geq 0$ and $SD(u, v) = 0$ iff $u = v$.

Proof: $SD(u, v) = \partial_s(\psi(u), \psi(v)) \geq 0$ and $SD(u, v) = 0$ iff $\partial_s(\psi(u), \psi(v)) = 0$ iff $\psi(u) = \psi(v)$ iff $u = v$.

Proposition 4: Triangle inequality: $SD(u, v) \leq SD(u, r) + SD(r, v)$.

Proof: $SD(u, v) = \partial_s(\psi(u), \psi(v)) \leq \partial_s(\psi(u), \psi(r)) + \partial_s(\psi(r), \psi(v)) = SD(u, r) + SD(r, v)$.

Thus, elements of A are called points of the metric space, and SD is called a metric or distance function on $\mathbb{R}^n_+$, which could be thought as $SD(u, v) = \sum_{i=1}^{n} \partial_s(u_i, v_i)$.

Now, its time to investigate some of the properties of S-distance measure in the form of following theorems.

Theorem 1: S-distance is not a Bregman divergence.

Proof: We will prove by contradiction. Let us assume that the original statement i.e. S-distance were a Bregman divergence $SD(u, v)$ is false. The original statement is false if $SD(u, v)$ is convex in $\mathbf{u}$. Now, SD, could also be expressed by equation 8.

$$SD(u, v) = \sum_{i=1}^{n} [log((u_i + v_i)/2) - (log(u_i) + log(v_i)/2] \quad (8)$$

The following expression could be obtained by taking double-derivative both sides of equation 8 w.r.t. u_i. $\frac{\partial^2 SD}{\partial_s u_i^2} = \frac{1}{u_i + v_i} - \frac{1}{2u_i}$ $\frac{\partial^2 SD}{du_i du_j} = 0$ when $i \neq j$ otherwise, $\frac{\partial^2 SD}{\partial_s u_i^2} = -\frac{1}{(u_i + v_i)^2} + \frac{1}{2u_i^2}$ $\frac{\partial^2 SD}{\partial u_i^2} < 0$ could be obtained for $i \in \{1, 2, \ldots, n\}$ when $v_i < (\sqrt{2} - 1)u_i \forall i \in \{1, 2, \ldots, n\}$. So, a diagonal matrix with negative diagonal entries can be obtained, which is known as Hessian matrix. So, $SD(u, v)$ is not convex in u. Therefore, it is proved that S-distance is not a Bregman divergence.

Theorem 2: $SD(x \circ u, x \circ v) = xSD(u, v)$ for $x \in \mathbb{R}^n_+$, where $x \circ u$ represents the Hadamord product between x and u.

Proof: We know, $(x \circ u) = (x_1u_1, x_2u_2, \ldots, x_nu_n)$. So, $\delta_s(x_iu_i, x_iv_i) = log(\frac{(x_iu_i + x_iv_i)}{2}) - \frac{log(x_iu_i) + log(x_iv_i)}{2} = log(x_i) + log(\frac{(u_i + v_i)}{2}) - \frac{(log(u_i) + log(v_i) + 2log(x_i))}{2} = log(\frac{(u_i + v_i)}{2}) - \frac{(log(u_i) + log(v_i) +)}{2} = \delta_s(u_i, v_i)$ $\sum_{i=1}^{n} \delta_s(x_iu_i, x_iv_i) = \sum_{i=1}^{n} x_i\delta_s(u_i, v_i)$ implying So, $SD(x \circ u, x \circ v) = xSD(u, v)$

Theorem 3: S-distance is f-divergence.

Proof: A divergence is called as f-divergence when that divergence can be expressed in the following form $\phi(t) = u\phi(\frac{v}{u})$, where $t = \frac{v}{u}$ The S-distance between $u \in \mathbb{R}^n_+$ and $v \in \mathbb{R}^n_+$ is given by $SD(u, v) = \sum_{i=1}^{n} [log((u_i + v_i)/2) - (log(u_i) + log(v_i)/2]$ putting $t_i = \frac{v_i}{u_i}$ $SD(u, v) = \sum_{i=1}^{n} [log((u_i + u_it_i)/2) - (log(u_i) + log(u_it_i)/2]$ $SD(u, v) = \sum_{i=1}^{n} [log(\frac{(1+t_i)}{2}) - \frac{log(t_i)}{2})] = \sum_{i=1}^{n} \phi(t) = \sum_{i=1}^{n} x_i\phi(\frac{v_i}{u_i})$ Since, $SD(u, v)$ can be expressed as $\sum_{i=1}^{n} x_i\phi(\frac{v_i}{u_i})$. Thus, S-distance is f-divergence.

IV. PROPOSED FUZZY K-MEANS WITH S-DISTANCE

The FKM with S-distance achieves grouping by solving equation 9.

$$\min_{\substack{C=(c_1,c_2,\ldots,c_k)\in R^{n\times k}\\ W\in M}} E_s(C, W; A) = \sum_{b=1}^{m}\sum_{a=1}^{k} (w_{ab})^s SD(u_b, c_a), \quad 1 \leq s < \infty, \quad (9)$$

where

$$M = \left\{ W = [w_{ab}]_{\substack{a=1,2,\ldots,k\\ b=1,2,\ldots,m}} \middle| w_{ab} \in [0, 1], \sum_{a=1}^{k} w_{ab} = 1, \sum_{b=1}^{m} w_{ab} > 0 \right\} \quad (10)$$

Generally, closed-form solution of equation 10 does not exist [30]. According to the below mentioned theorem, an alternating optimization method with modified equations of 4 and 5 are available in literature to find a solution.

Theorem 4 (Alternating Optimization of FKM): Let $\tau_b = \left\{a | a \in \{1, \ldots, k\}, u_b = c_a^{(e)}\right\}$, where e represents the e^{th} epoch. Equation 11 is the modified form of equation 4 whereas equation 5 would be same, which are required in alternating optimization algorithm for proof of convergence globally to a minimizer or a saddle point of E_s. Theorem 4.1 is identical to that in [31].

$$w_{ab}^{(e+1)} = \begin{cases} \left(\sum_{l=1}^{k} \left[\frac{SD(u_b, c_a^{(e)})}{SD(u_b, c_l^{(e)})} \right]^{\frac{2}{s-1}} \right)^{-1}, & \text{if } \tau_b = 0 \\ \frac{1}{|\tau_b|}, & \text{if } \tau_b \neq 0 \text{ and } a \in \tau_b \\ 0, & \text{if } \tau_b \neq 0 \text{ and } a \notin \tau_b \end{cases} \quad (11)$$

The FKM criterion in equation 9 can be expressed with the help of reduced unconstrained FCM criterion in theorem 4.2.

Theorem 5 (Reduced FKM Criterion [30], [32], [33]): The reduced FKM criterion is reported in equation 12, which is identical to

$$\min_{C\in R^{n\times k}} E_s'(C; A) = \sum_{b=1}^{m} \left[\sum_{a=1}^{k} SD(u_b, c_a)^{\frac{2}{1-s}} \right]^{1-s} \quad (12)$$

C^* is a local or global minimizer or a saddle point of E_s' if (C^*, W^*) is a local or global minimizer or a saddle point of E_s. $(C^*, F(C^*))$ is a local or global minimizer or a saddle point of E_s if C^* is a local or global minimizer or a saddle point of E_s', where $F : R^{n\times k} \to M; F(C) = W$ with each w_{ab} estimated using equation 11. The detail proof of theorem is explained in [32], [33].

V. CONVERGENCE ANALYSIS

Let us consider $f(\gamma_1, \gamma_2, \ldots, \gamma_k) = (\sum_{a=1}^{k} \gamma_a^q)^{1/q}$, where $q = \frac{1}{1-s} < 0$. Then equation 12 can be expressed as

$$E'_s(c_1, c_2, \ldots, c_k; A) = \sum_{b=1}^{m} f(\gamma_{k1}, \gamma_{k2}, \ldots, \gamma_{kb})|_{\gamma_{ab}=d(u_b,c_a)^2} \tag{13}$$

Lemma 5.1. can be stated with the help of equation 13, where the right side of equation 14 is known as a majorant of E'_s.

Lemma 1 (Majorant of E'_s):

$$\begin{aligned} &E'_s(c_1, c_2, \ldots, c_k; A) \\ &\quad \le maj^e E'_s = E'_s(c_1^{(e)}, c_2^{(e)}, \ldots, c_k^{(e)}; A) \\ &\quad + \sum_{b=1}^{m}\sum_{a=1}^{k} \frac{df}{d\gamma_{ab}}|_{(e)}\Big(SD(u_b, c_a)^2 - SD(u_b, c_a^e)^2\Big), \end{aligned} \tag{14}$$

where the derivative is taken at $c_1^{(e)}, c_2^{(e)}, \ldots, c_k^{(e)}$.

Proof: In [30], Gröll and Jakel proved that $f(\gamma_1, \gamma_2, \ldots, \gamma_k)$ is concave. So,

$$f(\gamma_1, \gamma_2, \ldots, \gamma_k) \le f(\rho_1, \rho_2, \ldots, \rho_k) + \sum_{a=1}^{k} \frac{df}{d\gamma_{a_{\rho_a}}}(\gamma_a - \rho_a) \tag{15}$$

Equation 16 can be obtained by replacing γ_a and ρ_a using γ_{ab} and ρ_{ab} respectively and taking the sum over all b.

$$\begin{aligned} \sum_{b=1}^{m} f(\gamma_{1b}, \gamma_{2b}, \ldots, \gamma_{kb}) &\le \sum_{b=1}^{m} f(\rho_{1b}, \rho_{2b}, \ldots, \rho_{kb}) \\ &\quad + \sum_{b=1}^{m}\sum_{a=1}^{k} \frac{df}{d\gamma_{ab_{\rho_{ab}}}}(\gamma_{ab} - \rho_{ab}) \end{aligned} \tag{16}$$

Equation 17 can be derived after putting the value of $\gamma_{ab} = d(u_b, c_a)^2$ and $\rho_{ab} = d(u_b, c_a^{(e)})^2$ as well as equation 14.

$$\begin{aligned} &E'_s(c_1, c_2, \ldots, c_k; A) \\ &\quad \le E'_s(c_1^{(e)}, c_2^{(e)}, \ldots, c_k^{(e)}; A) \\ &\quad + \sum_{b=1}^{m}\sum_{a=1}^{k} s\frac{df}{d\gamma_{ab}}|_{(e)}\Big(SD(u_b, c_a)^2 - SD(u_b, c_a^e)^2\Big) \end{aligned} \tag{17}$$

Every majorant obtained from equation 14 is a global majorant. It could be said clearly that a majorant along an random search direction is a global majorant. If and only if the search direction passes the global minimizer of the global majorant then the minimizers of global and directional majorants are identical.

Theorem 6 (Alternating optimization as a Steepest Descent Algorithm): If the steplength is tuned using the majorization principle by the majorant based on equation 14 then the sequences $c_a^{(e+1)}$ produced by the alternating optimization algorithm in equations 11 and 5 and the sequences of a steepest descent algorithm applied to equation 12 are equivalent.

Proof: All coefficients $\frac{df}{d\gamma_{ab}}$ of the strictly convex terms $d(u_b, c_a)^2$ are non-negative.

$$\begin{aligned} \frac{df}{d\gamma_{ab}} &= \frac{d}{d\gamma_{ab}}\left[\sum_{a=1}^{k}\gamma_{ab}^q\right]^{\frac{1}{q}} = \left[\sum_{a=1}^{k}\gamma_{ab}^q\right]^{\frac{1}{q-1}} \gamma_{ab}^{q-1} \\ &= \left[\sum_{l=1}^{k}\left(\frac{1}{\gamma_{lb}}\right)^{-q}\right]^{\frac{1}{q-1}} (\gamma_{ab}^{-q})^{\frac{1}{q-1}} = \left[\sum_{l=1}^{k}\frac{\gamma_{ab}^{-q}}{\gamma_{lb}^{-q}}\right]^{\frac{1}{q-1}} \\ &= \left[\left(\sum_{l=1}^{k}\left[\frac{SD(u_b, c_a)^2}{SD(u_b, c_l)^2}\right]^{\frac{1}{s-1}}\right)^{-1}\right]^{s} = (w_{ab})^s \ge 0 \end{aligned} \tag{18}$$

Hence, the majorant is convex w.r.t all c_a. Moreover, majorant is convex because at least one of the coefficients corresponding to any c_a is non-negative. So, the unique minimizer, c_a^o, of the majorant using the first-order both necessary and sufficient condition is

$$\begin{aligned} \nabla_{c_a} maj^{(e)} &= E'_s(c_1, c_2, \ldots, c_k; A)|_{c_a=c_a^o} \\ &= -2\sum_{b=1}^{m}\frac{df}{d\gamma_{ab}}|_{(e)}(u_b - c_a^o) = 0, a = 1, 2, \ldots, k \end{aligned} \tag{19}$$

So, the value of $c_a^{(e+1)}$ is

$$c_a^{(e+1)} = c_a^o = \frac{\sum_{b=1}^{m}\frac{df}{d\gamma_{ab}}|_{(e)}u_b}{\sum_{b=1}^{m}\frac{df}{d\gamma_{ab}}|_{(e)}} \tag{20}$$

Equation 20 would be same as equation 5 after replacing $\frac{df}{d\gamma_{ab}}|_{(e)}$ by $(w_{ab}^{(e+1)})^s$. The steepest descent can be computed using equation 21.

$$\begin{aligned} c_a^{(e+1)} &= c_a^{(e)} \\ &\quad - \underbrace{\frac{1}{2\sum_{b=1}^{m}\frac{df}{d\gamma_{ab}}|_{(e)}}}_{\text{steplength } \alpha_a^{(e)}} \cdot \underbrace{\left(-2\sum_{b=1}^{m}\frac{df}{d\gamma_{ab}}|_{(e)}(u_b - c_a^{(e)})\right)}_{\nabla_{c_a} E'_s(c_1,c_2,\ldots,c_k;A)|_{c_l=c_l^{(e)},\ l=1,2,\ldots,k}} \end{aligned} \tag{21}$$

Lastly, the global minimizer majorants are majorants along the direction with steepest descent.

$$\begin{aligned} &\nabla_{c_a} maj^{(e)} E'_s(c_1, c_2, \ldots, c_k; A)|_{c_l=c_l^{(e)}} \\ &\quad = \underbrace{\nabla_{c_a} E'_s(c_1, c_2, \ldots, c_k; A)|_{c_l=c_l^{(e)},\ l=1,2,\ldots,k}}_{\nabla_{c_a} E_s'^{(e)}} \end{aligned} \tag{22}$$

Now, it is easy to define the convergence properties as following three corollaries using above mentioned optimization theory.

Corollary 1 (Global Convergence of Reduced FKM): The reduced FKM state in equation 9 converges globally to a local minimizer or saddle point.

Proof: using lemma 5.1. $E_s^{'(e)} - E_s^{\prime(e+1)} \geq E_s^{\prime(e)} - maj^{(e)} E_s'\Big(c_1^{(e+1)}, c_2^{(e+1)}, \ldots, c_k^{(e+1)}; A\Big)$
$= \sum_{b=1}^{m} \sum_{a=1}^{k} \frac{df}{d\gamma_{ab}}|_{(e)} \Big(SD\big(u_b, c_a^{(e)}\big)^2 - SD\big(u_b, c_a^{(e)} + \alpha_a^{(e)} \cdot \nabla_{c_a} E_s^{\prime(e)}\big)^2\Big)$
$= \sum_{b=1}^{m} \sum_{a=1}^{k} \frac{df}{d\gamma_{ab}}|_{(e)} \Big(- 2\alpha_a^{(e)} \big(u_b - c_a^{(e)}\big)^T \nabla_{c_a} E_s^{\prime(e)} - \big(\alpha_a^{(e)}\big)^2 \big|\big|\nabla_{c_a} E_s^{\prime(e)}\big|\big|_2^2\Big)$
$= \sum_{a=1}^{k} \Big(- 2\alpha_a^{(e)} \sum_{b=1}^{m} \frac{df}{d\gamma_{ab}}|_{(e)} \big(u_b - c_a^{(e)}\big)^T \nabla_{c_a} E_s^{\prime(e)} - \sum_{b=1}^{m} \frac{df}{d\gamma_{ab}}|_{(e)} \big(\alpha_a^{(e)}\big)^2 \big|\big|\nabla_{c_a} E_s^{\prime(e)}\big|\big|_2^2\Big)$
$= \sum_{a=1}^{k} \Big(\alpha_a^{(e)} \big|\big|\nabla_{c_a} E_s^{\prime(e)}\big|\big|_2^2 - 2\frac{1}{\alpha_a^{(e)}} \big(\alpha_a^{(e)}\big)^2 \big|\big|\nabla_{c_a} E_s^{\prime(e)}\big|\big|_2^2\Big)$,
where $\nabla_{c_a} E_s^{\prime(e)} = -2 \sum_{b=1}^{m} \frac{df}{d\gamma_{ab}}|_{(e)} \big(u_b - c_a^{(e)}\big)$ and $\alpha_a^{(e)} = \frac{1}{2\sum_{b=1}^{m} \frac{df}{d\gamma_{ab}}|_{(e)}}$
$= \sum_{a=1}^{k} \frac{\alpha_a^{(e)}}{2} \big|\big|\nabla_{c_a} E_s^{\prime(e)}\big|\big|_2^2 \geq \sum_{a=1}^{k} \frac{1}{4m} \big|\big|\nabla_{c_a} E_s^{\prime(e)}\big|\big|_2^2$ since $\sum_{b=1}^{m} \frac{df}{d\gamma_{ab}}|_{(e)} \leq m$ and hence $\alpha_a^{(e)} \geq \frac{1}{2m}$
So, $E_s^{'(e)} \geq E_s^{\prime(e+1)}$. Together with the boundedness of E_s' obeys $\lim_{e\to\infty} \Big(E_s^{\prime(e)} - E_s^{\prime(e+1)}\Big) = 0$. The right-hand side of the inequality would be zero if the lest-hand side tends to zero since it is bounded by zero. Convergence to a stationary point can be deduced from $\lim_{e\to\infty} \big|\big|\nabla_{c_a} E_s^{\prime(e)}\big|\big|_2^2 = 0 \, \forall a$, and so, $\lim_{e\to\infty} \nabla_{c_a} E_s^{\prime(e)} = 0$. Only local minimizers or saddle points appear as limit points since $\{E_s^{\prime(e)}\}$ is a non-decreasing sequence.

Corollary 2 (Local Convergence of Reduced FKM): There exists some neighborhood $Z(c_1^*, c_2^*, \ldots, c_k^*)$ if $C = (c_1^*, c_2^*, \ldots, c_k^*)$ is a strict local minimizer of E_s' such that if a starting point $C^{(0)} = (c_1^{(0)}, c_2^{(0)}, \ldots, c_k^{(0)})$ is chosen in this neighborhood, the FKM algorithm converges to $C^* = (c_1^*, c_2^*, \ldots, c_k^*)$.

Proof: Since the FKM algorithm is a globally convergent gradient method by corollary 5.1., the well-known Capture Theorem can be applied [34].

Corollary 3 (Convergence Rate of FKM): FCM converges linearly near a nonsingular local minimum (a minimum with a positive definite Hessian matrix).

Proof: The proof of this statement uses a Taylor expansion of E_s' and the well-known convergence rate theorem for quadratic functions [34].

VI. EXPERIMENTS

A. DATASET DESCRIPTION

Some real and synthetic datasets are used to validate the proposed FKM in this study. Multi-class synthetic datasets have been generated by assigning each class one or more normally distributed clusters of points. The synthetic datasets consist of 2_blobs (DB_1), 3_blobs (DB_2), 5_blobs (DB_3), and 10_blobs (DB_4) in this work. The first row of figure 1 shows the data-points of DB_1, DB_2, DB_3 and DB_4 respectively. On the other hand, the real datasets are collected from the UCI Machine Learning Repository [35] and Keel Repository [36]. The real datasets include Iris (DB_5), Glass (DB_6), Cleveland (DB_7), Bank Note Authentication (DB_8), Appendicitis (DB_9), Breast Cancer Wisconsin (DB_{10}) and Mammography (DB_{11}). Two sets of multi-spectral and panchromatic images along with their ground truths are also used in this work. Both the datasets acquired using Worldview-2 sensor at 1:25,000 scale. Satellite image data and spectral measurements were correlated and classified to identify the location in the studied region. The first set of images named as DB_{12}, consists of one panchromatic band of high resolution 0.46m and 8 multi-spectral bands having 1.8m resolution. The multi-spectral band includes red, blue, green, near infrared, red-edge, coastal, yellow and near infrared 2. The first four are the standard bands while the rest are new bands [37]. The size of each image is 2048 × 2048 pixels. These images were captured on 11th September, 2011. The images enclosed an area of 10.485 ha and the coordinates of upper left corner are $S32°51'7.91''$ and $W70°39'5.10''$ respectively. The area corresponds to a rural zone located at Valparaiso region in Comuna de Los Andes, Chile. Seven land covers were found in the studied area in DB_{12}. These land covers are generic agricultural land, water bodies, and four different types of crops (nectarine, grapevine, alfalfa and maize) in different phenological stages and buildings and urban construction. The nectarine crops area could be separated further in two different crop areas: nectarin_1 and nectarin_2. The panchromatic image of DB_{12} is shown in figure 2a.

The second set of images was captured on 19^{th} January, 2012 using the same sensor. The images are covered in an area of 157 ha of croplands at Coihueco district, in Nuble province, Biobio region, Chile (S36°37′15.7″ and W71°53′57.7″). The area is a good representation of diverse vegetation, forests, rural constructions, and crops. The satellite images consist of one panchromatic image of 0.59m resolution and 4 multi-spectral of resolution 2.36m. Four spectral bands are as follows: blue band (450 to 510nm), green band (510 to 580nm), red band (630 to 690nm), and near infrared (NIR, 770 to 895nm) band data. The size of each image is 2006 × 2172 pixels. Five land covers were found in DB_{13}. These land covers are forest, soil, crop, fruit, and urban construction. The panchromatic image of DB_{13} is shown in figure 3a. The labeled polygons included in the vector file are overlaid on figure 2a and figure 3a, shown in figure 2b and figure 3b respectively. Beside these *in-situ* data, pixel-wise ground truth of land cover types are also provided with both the datasets.

B. CLUSTER VALIDITY INDEX

For cluster analysis, the analogous question is how to quantify the "goodness" of the resulting clusters? The notion of "goodness" is evaluated using validity indexes. The concept of a validity index can be expressed mathematically. Let us consider m data-points, A. A clustering algorithm divides A into k-partitions namely, $A_1, A_2, \ldots, A_k$. The values of their corresponding validity indexes are $Z_1, Z_2, \ldots, Z_k$. The $Z_{h1} \geq Z_{h2} \geq \ldots \geq Z_{hk}$ will depict that $A_{h1} \uparrow A_{h2} \uparrow \ldots \uparrow A_{hk}$,

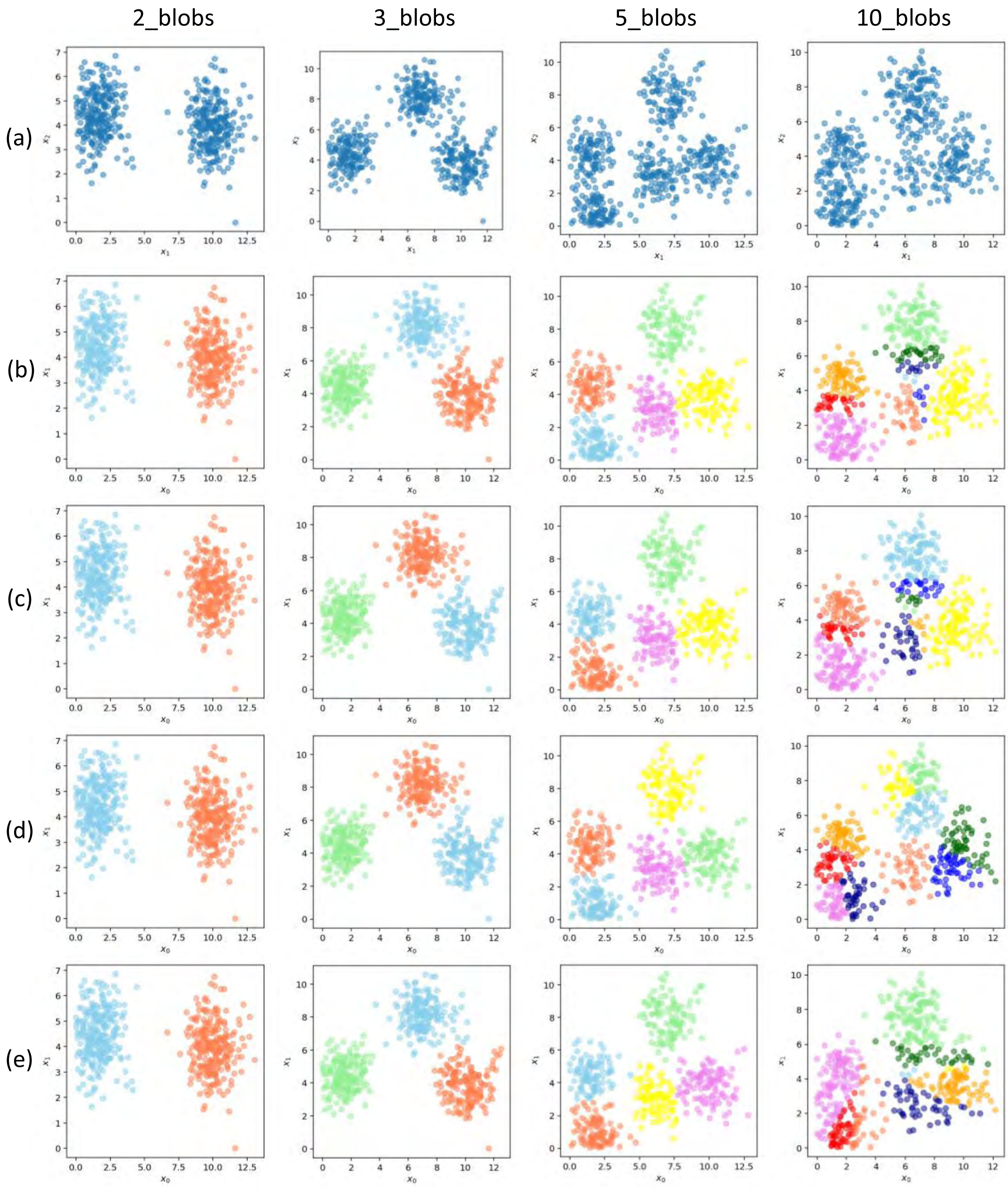

FIGURE 1. Clustering Results: (a) Original Structure (b) Grouping using FKM with Euclidean distance (M_1) (c) Grouping using Minkowski weighted FKM (M_2) (d) Grouping using Weighting in FKM (M_3) and (e) Grouping using the proposed FKM (M_4).

for some permutation $h1, h2, \ldots, hk$ of $\{1, 2, \ldots, k\}$, where $A_i \uparrow A_j$ denotes that the partition A_i is a better cluster than A_j [38]. Numerical measures that are applied to conclude different aspects of cluster validity, are classified into the following two types mainly. External index is used to measure the extent to which cluster labels match externally supplied class labels. Normalized Mutual Information (NMI) [39] and Adjusted Rand Index (ARI) [40] are two external validity indexes in this work. On the other hand, internal index is used to measure the goodness of a clustering structure without respect to external information. For this study, three internal evaluation schemes namely, Silhouette index (SI) [41], Dunn index (DI) [38] and Davies Boulden Index (DBI) [38] have also been considered to determine the cohesiveness of the obtained clusters. NMI is used as an index to compare performance between two groups of data-points. On the other hand, ARI has been considered as a cluster validation index. Both of these metrics depict the mismatch between two data clustering of a given set of data-points. The highest value i.e. 1 indicates no mismatch whereas the lowest value i.e. 0 represents complete mismatch. Both of the metrics are used to compare the partition achieved by the algorithms and the

FIGURE 2. **(a) Panchromatic image of DB_{12} and (b) corresponding *in-situ* data showing selected regions of different types of land covers with color labels.**

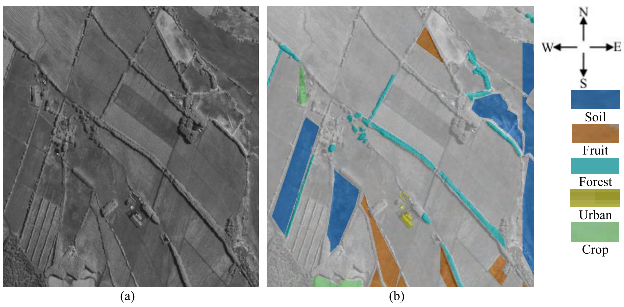

FIGURE 3. **(a) Panchromatic image of DB_{13} and (b) corresponding *in-situ* data showing selected regions of different types of land covers with color labels.**

TABLE 1. **The values of ARI, NMI, SI, DI, and DBI for synthetic and real datasets.**

	Dataset	M_1	M_2	M_3	M_4
ARI	DB_1	**1.0**	**1.0**	**1.0**	**1.0**
	DB_2	0.9880180	0.982036	0.982036	**0.990181**
	DB_3	0.875980	0.875845	0.898623	**0.91747**
	DB_4	0.463736	0.448327	0.432681	**0.475745**
	DB_5	0.921409	0.902619	0.979932	**0.989932**
	DB_6	0.446451	0.431981	0.5476595	**0.556994**
	DB_7	0.577500	0.78239	0.484217	**0.889659**
	DB_8	0.550272	0.550272	0.532513	**0.567329**
	DB_9	0.027349	0.0317099	0.020098	**0.032149**
	DB_{10}	0.528617	0.5391607	0.491424	**0.578617**
	DB_{11}	0.107842	0.107842	0.106418	**0.126852**
	DB_{12}	0.406837	0.307993	0.407870	**0.507011**
	DB_{13}	0.320471	0.320471	0.226208	**0.378029**
NMI	DB_1	**1.0**	**1.0**	**1.0**	**1.0**
	DB_2	0.977740	0.966614	0.982036	**0.990181**
	DB_3	0.866303	0.868204	0.898623	**0.90747**
	DB_4	0.654322	0.643144	0.432681	**0.675745**
	DB_5	0.897779	0.879632	0.979932	**0.989932**
	DB_6	0.616544	0.593026	0.5476595	**0.656994**
	DB_7	0.396832	0.53360	0.484217	**0.589659**
	DB_8	0.451882	0.4510854	0.532513	**0.567329**
	DB_9	0.027349	0.020370	0.020098	**0.029349**
	DB_{10}	0.486361	0.5040837	0.491424	**0.578617**
	DB_{11}	0.086286	0.086286	0.106418	**0.126852**
	DB_{12}	0.328378	0.295578	0.512618	**0.534485**
	DB_{13}	0.019634	0.440703	0.418665	**0.474168**
SI	DB_1	0.794916	0.794873	0.798578	**0.799972**
	DB_2	0.6931593	0.693228	**0.695169**	**0.695169**
	DB_3	0.572960	0.571924	0.571924	**0.583560**
	DB_4	0.256123	0.248279	0.248279	**0.257125**
	DB_5	0.577102	0.574844	0.577102	**0.589541**
	DB_6	0.315921	0.246860	0.315921	**0.320874**
	DB_7	0.399785	0.394392	0.399785	**0.401259**
	DB_8	0.410091	0.407459	0.410091	**0.412235**
	DB_9	0.027349	0.369059	0.027349	**0.373494**
	DB_{10}	0.691067	0.690935	0.691067	**0.692024**
	DB_{11}	0.530125	0.530125	0.530125	**0.535214**
	DB_{12}	0.314455	0.390665	0.297673	**0.408613**
	DB_{13}	0.021033	0.025401	0.021698	**0.112395**
DI	DB_1	1.904015	1.914016	1.904015	**1.924316**
	DB_2	1.667249	1.704798	1.667249	**1.990181**
	DB_3	1.25921	1.307801	1.307801	**1.90747**
	DB_4	0.463736	0.220063	0.463736	**1.475745**
	DB_5	1.897049	1.840235	1.840235	**1.989932**
	DB_6	0.444175	0.065613	0.476595	**0.556994**
	DB_7	1.423736	1.364377	1.484217	**1.889659**
	DB_8	1.00112	1.001708	1.532513	**1.567329**
	DB_9	1.257200	**1.297200**	1.020098	1.029349
	DB_{10}	1.300558	1.300383	1.491424	**1.578617**
	DB_{11}	1.397413	1.397412	1.106418	**1.406852**
	DB_{12}	0.392247	0.474642	0.509411	**0.512247**
	DB_{13}	0.391262	0.391262	0.377334	**0.404804**
DBI	DB_1	0.144642	**0.144604**	0.144634	**0.144604**
	DB_2	**0.156546**	0.156589	0.165546	**0.156546**
	DB_3	0.122504	0.121585	0.121585	**0.120504**
	DB_4	0.187655	0.196655	**0.141915**	**0.141915**
	DB_5	**0.167452**	0.167492	0.185452	**0.167452**
	DB_6	0.507572	0.574572	**0.253338**	**0.253338**
	DB_7	0.485210	0.472901	0.472901	**0.468776**
	DB_8	0.523546	0.521241	**0.518281**	**0.518281**
	DB_9	0.498974	0.499391	0.499391	**0.027349**
	DB_{10}	0.258860	0.257766	0.257766	**0.257680**
	DB_{11}	**0.319149**	**0.319149**	0.324503	**0.319149**
	DB_{12}	0.272598	0.448134	0.5574520	**0.136888**
	DB_{13}	0.228156	0.228156	0.387455	**0.202353**

ground truth to estimate the performance of these algorithms. These internal indexes measure how similar a data point is to its own group (cohesion) compared to other groups (separation). The ranges of SI varies from -1 to $+1$, where a high value depicts that the data point is well matched to its own group and poorly matched to neighbor groups. A higher DI and lower DBI depict better clustering.

C. COMPUTATIONAL PROTOCOLS

Four experimental studies have been conducted on the above said datasets using FKM with Euclidean distance (M_1), Minkowski weighted FKM (M_2), Weighting in FKM (M_3) and the proposed FKM (M_4).

PERFORMANCE COMPARISON: It is made sure that same randomly chosen centroids were considered for each of the algorithm and for estimating ARI, NMI, SI, DI, and DBI values to maintain the consistency in results. The performance of each clustering techniques do not rely on the selection of initial set of centroids. However, the performance relies on the clustering algorithm itself. The procedure is repeated 10 times on each dataset. Then Wilcoxon signed, ranksum and signtest have been conducted to know whether two dependent data-points from populations having same distribution on the obtained values of ARI, NMI, SI, DI and DBI using M_i, where $1 \leq i \leq 4$.

VII. RESULTS AND DISCUSSION

Table 1 reports the mean ARI, NMI, SI, DI and DBI values computed by M_i, where $1 \leq i \leq 4$ on synthetic and real datasets. It is clear from table 1 that the proposed M_4 outperforms M_i, where $1 \leq i \leq 3$ since most of the ARI and NMI values are closer to the highest value i.e. 1. In most of

TABLE 2. P-Values obtained from ARI, NMI, SI, DI, and DBI for Wilcoxon signed rank test to compare M_4 with M_i, where $1 \leq i \leq 3$.

	Dataset	M_1	M_2	M_3
	DB_1	1.0	1.0	1.0
	DB_2	0.0025	0.0025	0.0025
	DB_3	0.0020	0.002	0.0020
	DB_4	0.0020	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0371	0.0440	0.0195
ARI	DB_7	0.0020	0.0020	0.0020
	DB_8	0.0022	0.0020	0.0015
	DB_9	0.0020	0.0020	0.0015
	DB_{10}	0.002	0.002	0.002
	DB_{11}	0.002	0.002	0.002
	DB_{12}	0.0015	0.0025	0.0020
	DB_{13}	0.0020	0.0025	0.0015
	DB_1	1.0	1.0	1.0
	DB_2	0.0025	0.0025	0.0025
	DB_3	0.0020	0.002	0.0020
	DB_4	0.0020	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0022	0.0020	0.0020
NMI	DB_7	0.0020	0.0020	0.0020
	DB_8	0.0020	0.0020	0.0020
	DB_9	0.0025	0.0020	0.0025
	DB_{10}	0.002	0.002	0.002
	DB_{11}	0.002	0.002	0.002
	DB_{12}	0.0020	0.0020	0.0020
	DB_{13}	0.0025	0.0015	0.0030
	DB_1	0.0020	0.0020	0.0020
	DB_2	0.0020	0.0020	0.0020
	DB_3	0.0020	0.0020	0.0020
	DB_4	0.0025	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0371	0.0840	0.0195
SI	DB_7	0.0020	0.0022	0.0020
	DB_8	0.0022	0.0020	0.0020
	DB_9	0.0020	0.0020	0.0022
	DB_{10}	0.002	0.0025	0.002
	DB_{11}	0.002	0.002	0.002
	DB_{12}	0.0015	0.0015	0.0025
	DB_{13}	0.0020	0.0020	0.0020
	DB_1	0.0020	0.0020	0.0020
	DB_2	0.0020	0.0020	0.0020
	DB_3	0.0020	0.0020	0.0020
	DB_4	0.0025	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0025	0.0020	0.0020
DI	DB_7	0.0020	0.0022	0.0020
	DB_8	0.0022	0.0025	0.0020
	DB_9	0.0020	0.0020	0.0022
	DB_{10}	0.0022	0.0025	0.0022
	DB_{11}	0.0020	0.0022	0.0020
	DB_{12}	0.0025	0.0002	0.0005
	DB_{13}	0.0025	0.0020	0.0015
	DB_1	0.0020	0.0020	0.0020
	DB_2	0.0020	0.0020	0.0020
	DB_3	0.0020	0.0020	0.0020
	DB_4	0.0025	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0025	0.002	0.0025
DBI	DB_7	0.0020	0.0022	0.0020
	DB_8	0.0022	0.0020	0.0020
	DB_9	0.0020	0.0020	0.0022
	DB_{10}	0.0025	0.0025	0.0025
	DB_{11}	0.0020	0.0025	0.0020
	DB_{12}	0.0001	0.0005	0.0002
	DB_{13}	0.0025	0.0025	0.0015

TABLE 3. P-Values obtained from ARI, NMI, SI, DI, and DBI for Wilcoxon ranksum test to compare M_4 with M_i, where $1 \leq i \leq 3$.

	Dataset	M_1	M_2	M_3
	DB_1	1.0	1.0	1.0
	DB_2	0.0079	0.0079	0.0079
	DB_3	0.002	0.002	0.002
	DB_4	1.5938e-05	1.5938e-05	1.5938e-05
	DB_5	1.1758e-05	1.5938e-05	1.3659e-05
	DB_6	1.5938e-05	1.1758e-05	1.5938e-05
ARI	DB_7	1.4523-05	1.5938e-05	1.5938e-05
	DB_8	1.5938e-05	1.4523-05	1.5938e-05
	DB_9	1.5938e-05	1.3659e-05	1.1758e-05
	DB_{10}	1.5938e-05	1.5938e-05	1.5938e-05
	DB_{11}	1.5938e-05	1.5938e-05	1.5938e-05
	DB_{12}	5.6265e-05	0.0025	0.0015
	DB_{13}	0.0025	1.5254e-05	5.8927e-05
	DB_1	1.0	1.0	1.0
	DB_2	0.0079	0.0079	0.0079
	DB_3	1.5938e-05	1.5938e-05	1.5938e-05
	DB_4	1.5938e-05	1.5938e-05	1.5938e-05
	DB_5	1.6874e-05	1.5938e-05	1.5938e-05
	DB_6	1.5938e-05	1.6874e-05	1.5938e-05
NMI	DB_7	1.4523-05	1.4528e-05	1.5938e-05
	DB_8	1.5938e-05	1.5938e-05	1.6874e-05
	DB_9	1.5938e-05	1.5938e-05	1.5938e-05
	DB_{10}	1.5938e-05	1.4528e-05	1.5938e-05
	DB_{11}	1.5938e-05	1.5938e-05	1.5938e-05
	DB_{12}	1.5938e-05	0.0015	4.5938e-05
	DB_{13}	0.0025	6.5938e-05	4.5938e-05
	DB_1	1.5938e-05	1.5938e-05	1.5938e-05
	DB_2	1.6874e-05	1.5938e-05	1.3254e-05
	DB_3	1.5938e-05	1.5938e-05	1.5938e-05
	DB_4	1.5938e-05	1.6874e-05	1.5938e-05
	DB_5	1.5938e-05	1.5938e-05	1.5938e-05
	DB_6	1.5938e-05	1.5938e-05	1.3254e-05
SI	DB_7	1.5938e-05	1.6874e-05	1.5938e-05
	DB_8	1.5938e-05	1.4528e-05	1.5938e-05
	DB_9	1.6874e-05	1.5938e-05	1.3254e-05
	DB_{10}	1.5938e-05	1.5938e-05	1.4528e-05
	DB_{11}	1.5938e-05	1.4528e-05	1.5938e-05
	DB_{12}	0.0035	1.4587e-05	1.4523e-05
	DB_{13}	4.5938e-05	0.0025	1.9854e-05
	DB_1	1.5938e-05	1.5938e-05	1.5938e-05
	DB_2	1.6874e-05	1.5938e-05	1.3524e-05
	DB_3	1.5938e-05	1.5938e-05	1.5938e-05
	DB_4	1.5938e-05	1.6874e-05	1.5938e-05
	DB_5	1.5938e-05	1.3524e-05	1.5938e-05
	DB_6	1.5938e-05	1.5938e-05	1.5938e-05
DI	DB_7	1.5938e-05	1.6874e-05	1.5938e-05
	DB_8	1.5938e-05	1.4528e-05	1.3524e-05
	DB_9	1.6874e-05	1.3524e-05	1.5938e-05
	DB_{10}	1.5938e-05	1.5938e-05	1.4528e-05
	DB_{11}	1.5938e-05	1.4528e-05	1.5938e-05
	DB_{12}	1.5938e-05	4.1236e-05	2.4888e-05
	DB_{13}	2.5938e-05	1.5938e-05	0.00002
	DB_1	1.5938e-05	1.5938e-05	1.5938e-05
	DB_2	1.6874e-05	1.5938e-05	1.5938e-05
	DB_3	1.5938e-05	1.5938e-05	1.5938e-05
	DB_4	1.5938e-05	1.6874e-05	1.5938e-05
	DB_5	1.5938e-05	1.5938e-05	1.5938e-05
	DB_6	1.5938e-05	1.5938e-05	1.5938e-05
DBI	DB_7	1.5938e-05	1.6874e-05	1.5938e-05
	DB_8	1.5938e-05	1.4528e-05	1.5938e-05
	DB_9	1.6874e-05	1.5938e-05	1.5938e-05
	DB_{10}	1.5938e-05	1.5938e-05	1.4528e-05
	DB_{11}	1.5938e-05	1.4528e-05	1.5938e-05
	DB_{12}	2.1689e-05	0.0002	1.5688e-05
	DB_{13}	5.5938e-05	1.3625e-05	2.4582e-05

the cases, M_4 produces greater mean ARI and NMI values over other algorithms, which denotes the effectiveness of M_4. The values of SI, DI and DBI for the same datasets are showed in table 1. The obtained results again prove the effectiveness of the M_4 over M_i, where $1 \leq i \leq 3$ because the values generated by M_4 are more closer to idea values

TABLE 4. P-Values obtained from ARI, NMI, SI, DI, and DBI for Wilcoxon sign test to compare M_4 with M_i, where $1 \le i \le 3$.

	Dataset	M_1	M_2	M_3
ARI	DB_1	1.0	1.0	1.0
	DB_2	0.0025	0.0025	0.0025
	DB_3	0.0020	0.0020	0.0020
	DB_4	0.0020	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0371	0.0440	0.0195
	DB_7	0.0020	0.0020	0.0020
	DB_8	0.0022	0.0020	0.0015
	DB_9	0.0020	0.0020	0.0015
	DB_{10}	0.0020	0.0020	0.0020
	DB_{11}	0.0020	0.0020	0.0020
	DB_{12}	0.0025	0.0015	0.0020
	DB_{13}	0.0020	0.0020	0.0020
NMI	DB_1	0.0020	0.0020	0.0020
	DB_2	0.0025	0.0025	0.0025
	DB_3	0.0020	0.0020	0.0020
	DB_4	0.0020	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0022	0.0020	0.0020
	DB_7	0.0020	0.0020	0.0020
	DB_8	0.0020	0.0020	0.0020
	DB_9	0.0025	0.0020	0.0025
	DB_{10}	0.0020	0.0020	0.0020
	DB_{11}	0.0020	0.0020	0.0020
	DB_{12}	0.0020	0.0020	0.0020
	DB_{13}	0.0025	0.0025	0.0020
SI	DB_1	0.0020	0.0020	0.0020
	DB_2	0.0020	0.0020	0.0020
	DB_3	0.0020	0.0020	0.0020
	DB_4	0.0025	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0371	0.0840	0.0195
	DB_7	0.0020	0.0022	0.0020
	DB_8	0.0022	0.0020	0.0020
	DB_9	0.0020	0.0020	0.0022
	DB_{10}	0.0020	0.0025	0.0020
	DB_{11}	0.0020	0.0020	0.0020
	DB_{12}	0.0015	0.0020	0.0025
	DB_{13}	0.0020	0.0025	0.0020
DI	DB_1	0.0020	0.0020	0.0020
	DB_2	0.0020	0.0020	0.0020
	DB_3	0.0020	0.0020	0.0020
	DB_4	0.0025	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0025	0.0020	0.0020
	DB_7	0.0020	0.0022	0.0020
	DB_8	0.0022	0.0025	0.0020
	DB_9	0.0020	0.0020	0.0022
	DB_{10}	0.0022	0.0025	0.0022
	DB_{11}	0.0020	0.0022	0.0020
	DB_{12}	0.0002	0.0010	0.0025
	DB_{13}	0.0025	0.0005	0.0020
DBI	DB_1	0.0020	0.0020	0.0020
	DB_2	0.0020	0.0020	0.0020
	DB_3	0.0020	0.0020	0.0020
	DB_4	0.0025	0.0020	0.0020
	DB_5	0.0020	0.0020	0.0020
	DB_6	0.0025	0.0020	0.0025
	DB_7	0.0020	0.0022	0.0020
	DB_8	0.0022	0.0020	0.0020
	DB_9	0.0020	0.0020	0.0022
	DB_{10}	0.0025	0.0025	0.0025
	DB_{11}	0.0020	0.0020	0.0020
	DB_{12}	0.0020	0.0005	0.0015
	DB_{13}	0.0002	0.0001	0.0020

compared to values obtained by M_i, where $1 \le i \le 3$. The non-parametric Wilcoxon signed, ranksum and signtest have also been conducted for comparison of M_4 with M_i,

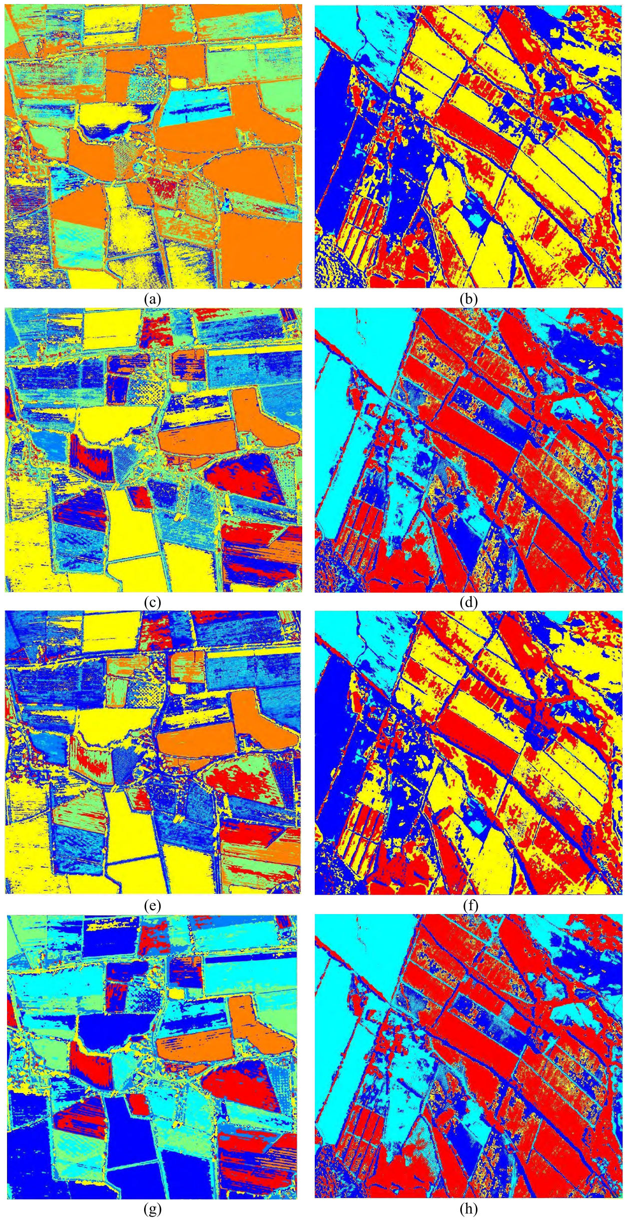

FIGURE 4. Clustering Results: (a) The output of DB_{12} using M_1 (b) The output of DB_{13} using M_1 (c) The output of DB_{12} using M_2 (d) The output of DB_{13} using M_2 (e) The output of DB_{12} using M_3 (f) The output of DB_{13} using M_3 (g) The output of DB_{12} using M_4 (h) The output of DB_{13} using M_4.

where $1 \le i \le 3$ based on the p-values estimated by ARI, NMI, SI, DI and DBI. The second, third, fourth and fifth rows of figure 1 show the output of clustering algorithms namely, M_1, M_2, M_3 and M_4 respectively on synthesis datasets. Figures 4a, 4c, 4e and 4g show the clustering outputs of $Data_{12}$ using M_1, M_2, M_3 and M_4 respectively. On the other hand, the clustering results of $Data_{13}$ using M_1, M_2, M_3 and M_4 are shown in figures 4b, 4d, 4f and 4h respectively. Table 2 to table 4 show the calculated p-values. All most all the achieved results suggest that we can reject the null hypothesis for 5 % level of significance. It means significant evidence is available based on data with us in order to say that M_4 algorithm outperforms M_i, where $1 \leq i \leq 3$ discussed in this study. It is also clear from table 2 to table 4 that the statistical results for DB_1 is not significant in some cases since p-values are greater than 0.05. However, ARI, NMI, SI, DI and DBI values of table 1 are good enough. So, we can conclude that most of the cases the proposed method with S-distance outperforms state-of-the-art methods.

VIII. CONCLUSION

In this study, a new distance metric on $\mathbb{R}_+^n$ has been addressed using S-divergence. Different properties of the S-distance has also been discussed. Classical FKM algorithm has been revised, where Euclidean distance has been replaced by the proposed distance. A theoretical analysis of the FKM with S-distance has also been studied by providing the proof of convergence. The study of data complexity metrics is an promising area of research in the field of clustering. It deserves further study. So, we would focus on the analysis of several dataset characteristics to retrieve information from them and this could further be considered to design the proper clustering algorithm.

REFERENCES

[1] L. Bottou and Y. Bengio, "Convergence properties of the k-means algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 1995, pp. 585–592.

[2] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with bregman divergences," *J. Mach. Learn. Res.*, vol. 6, pp. 1705–1749, Oct. 2005.

[3] S. Chakraborty and S. Das, "k-Means clustering with a new divergence-based distance metric: Convergence and performance analysis," *Pattern Recognit. Lett.*, vol. 100, pp. 67–73, Dec. 2017.

[4] L. Legrand and E. Grivel, "Jeffrey's divergence between moving-average models that are real or complex, noise-free or disturbed by additive white noises," *Signal Process.*, vol. 131, pp. 350–363, Feb. 2017.

[5] F. Nielsen and R. Nock, "Total jensen divergences: Definition, properties and clustering," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2015, pp. 2016–2020.

[6] F. Nielsen, R. Nock, and S. I. Amari, "On clustering histograms with k-means by using mixed a-divergences," *Entropy*, vol. 16, pp. 3273–3301, May 2014.

[7] R. Nock, F. Nielsen, and S.-I. Amari, "On conformal divergences and their population minimizers," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 527–538, Jan. 2016.

[8] M. D. Gupta, S. Srinivasa, J. Madhukara, and M. Antony, "Kl divergence based agglomerative clustering for automated vitiligo grading," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2700–2709.

[9] A. Notsu, O. Komori, and S. Eguchi, "Spontaneous clustering via minimum gamma-divergence," *Neural Comput.*, vol. 26, no. 2, pp. 421–448, Feb. 2014.

[10] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York, NY, USA: Wiley, 1973.

[11] H. Yao, Q. Duan, D. Li, and J. Wang, "An improved K-means clustering algorithm for fish image segmentation," *Math. Comput. Model.*, vol. 58, nos. 3–4, pp. 790–798, Aug. 2013.

[12] L.-H. Juang and M.-N. Wu, "Psoriasis image identification using k-means clustering with morphological processing," *Measurement*, vol. 44, no. 5, pp. 895–905, Jun. 2011.

[13] A. Paoli, F. Melgani, and E. Pasolli, "Clustering of hyperspectral images based on multiobjective particle swarm optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4175–4188, Dec. 2009.

[14] M. Volpi, D. Tuia, and M. Kanevski, "Memory-based cluster sampling for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3096–3106, Aug. 2012.

[15] B. Banerjee, F. Bovolo, A. Bhattacharya, L. Bruzzone, S. Chaudhuri, and B. K. Mohan, "A new self-training-based unsupervised satellite image classification technique using cluster ensemble strategy," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 741–745, Apr. 2015.

[16] N. Gillis, D. Kuang, and H. Park, "Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2066–2078, Apr. 2015.

[17] A. Meka and S. Chaudhuri, "A technique for simultaneous visualization and segmentation of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1707–1717, Apr. 2015.

[18] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, "Advances in knowledge discovery and data mining," Tech. Rep., 1996.

[19] Garima, H. Gulati, and P. K. Singh, "Clustering techniques in data mining: A comparison," in *Proc. 2nd Int. Conf. Comput. Sustain. Global Develop.*, Mar. 2015, pp. 410–415.

[20] M. A. Dalal and N. D. Harale, "A survey on clustering in data mining," in *Proc. Int. Conf. Workshop Emerg. Trends Technol.*, Feb. 2011, pp. 559–562.

[21] C.-R. Lin and M.-S. Chen, "Combining partitional and hierarchical algorithms for robust and efficient data clustering with cohesion self-merging," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 2, pp. 145–159, Feb. 2005.

[22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, Aug. 1996, pp. 226–231.

[23] I. Gurrutxaga *et al.*, "SEP/COP: An efficient method to find the best partition in hierarchical clustering based on a new cluster validity index," *Pattern Recognit.*, vol. 43, no. 10, pp. 3364–3373, Oct. 2010.

[24] A. K. Jain and R. C. Dubes, "Algorithms for clustering data," Tech. Rep., 1988.

[25] K. Fukunaga, "Statistical pattern recognition," in *Handbook Of Pattern Recognition And Computer Vision*. Singapore: World Scientific, 1999, pp. 33–60.

[26] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

[27] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," Tech. Rep., 1973.

[28] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," *Adv. Appl. Pattern Recognit.*, vol. 25, no. 3, p. 442, 1983.

[29] S. Sra, "Positive definite matrices and the s-divergence," *Proc. Amer. Math. Soc.*, vol. 144, pp. 2787–2797, Oct. 2015.

[30] L. Groll and J. Jakel, "A new convergence proof of fuzzy c-means," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 5, pp. 717–720, Oct. 2005.

[31] F. Hoppner and F. Klawonn, "A contribution to convergence theory of fuzzy c-means and derivatives," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 5, pp. 682–694, Oct. 2003.

[32] N. R. Pal, J. C. Bezdek, and R. J. Hathaway, "Sequential competitive learning and the fuzzy c-means clustering algorithms," *Neural Netw.*, vol. 9, no. 5, pp. 787–796, 1996.

[33] W. Wei and J. M. Mendel, "Optimality tests for the fuzzy c-means algorithm," *Pattern Recognit.*, vol. 27, no. 11, pp. 1567–1573, 1994.

[34] D. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena scientific, 1995.

[35] D. Dheeru and E. Karra. (2017). *Uci Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml

[36] J. Alcala-Fdez *et al.*, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *J. Multiple-Valued Logic Soft Comput.*, vol. 17, no. 2, pp. 255–287, Jan. 2011.

[37] *Digital Globe*. [Online]. Available: http://worldview2.digitalglobe.com/docs/WorldView-2_8-Band_Applications_Whitepaper.pdf./

[38] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.

[39] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, Oct. 2010.

[40] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, Dec. 1985.

[41] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. math.*, vol. 20, pp. 53–65, Nov. 1987.

ADITYA KARLEKAR received the B.Tech. degree in engineering from the Hitkarini College of Engineering and Technology, Jabalpur. He is currently an Intern with the Computer Science and Engineering Department, PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, Madhya Pradesh, India. His current research interests include image processing and pattern recognition.

AYAN SEAL received the Ph.D. degree in engineering from Jadavpur University, West Bengal, India, in 2014. He is currently an Assistant Professor with the Computer Science and Engineering Department, PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Madhya Pradesh, India. He is also with the Faculty of Informatics and Management, Center for Basic and Applied Research, University of Hradec Kralove, Hradec Kralove, Czech Republic. He has visited the Universidad Politecnica de Madrid, Spain, as a Visiting Research Scholar. He has authored or coauthored several journals, conferences, and book chapters in the area of biometric and medical image processing. His current research interests include image processing and pattern recognition. He was a recipient of several awards. He has received Sir Visvesvaraya Young Faculty Research Fellowship from Media Lab Asia, Ministry of Electronics and Information Technology, Government of India.

ONDREJ KREJCAR is currently a Full Professor in systems engineering and informatics with the University of Hradec Kralove (UHK), Czech Republic. He is also a Vice-Dean of science and research with the Faculty of Informatics and Management, UHK. He is also a Director of the Center for Basic and Applied Research, UHK. His h-index is 16 with more than 700 citations received in Web of Science. He has published more than 40 IF research papers. He has a number of cooperations throughout the world (e.g. Malaysia, Spain, UK, Ireland, Etiopia, Latvia, and Brazil). His research interests include control systems, smart sensors, ubiquitous computing, manufacturing, wireless technology, portable devices, biomedicine, image segmentation and recognition, biometrics, technical cybernetics, and ubiquitous computing. His second areas of interest include biomedicine (image analysis), biotelemetric system architecture (portable device architecture and wireless biosensors), and development of applications for mobile devices with the use of remote or embedded biomedical sensors. He is currently on the editorial board of *MDPI Sensors journal* JCR index, and several others ESCI indexed journals. He has been a Vice-leader and Management Committee member at WG4 at project COST CA17136, since 2018. He has also been a Management Committee member substitute at the project COST CA16226, since 2017. In 2018, he was awarded as 14th Top peer reviewer in Multidisciplinary in the World by Publons. Since 2019, he has been the Chairman of Program Committee of the KAPPA Program, Technological Agency of the Czech Republic, and will be a regulator of EEA/Norwegian Financial Mechanism in Czech Republic from 2019 to 2024. Since 2014, he has been the Deputy Chairman of the Panel 7 (Processing Industry, Robotics and Electrical Engineering) of the Epsilon Program, Technological Agency of the Czech Republic. At the University of Hradec Kralove, he is a guarantee of doctoral study program Applied Informatics, where he is focusing on lecturing of Smart Approaches to Development of Information Systems and Applications in Ubiquitous Computing Environments.

CONSUELO GONZALO-MARTIN received the B.A. degree in physics from Salamanca University, in 1986, and the Ph.D. degree in physics from the Complutense University of Madrid, in 1989. Since 1993, she has been an Assistant Professor with the Department of Architecture and Technology of Computer systems, Faculty of Computer Science, Universidad Politecnica de Madrid (UPM). She has coauthored 25 international journal publications, most of them with high impact factor in the remote sensing area as the *International Journal of Remote Sensing* or the *Canadian Journal of Remote Sensing* between others and has participated in 40 international congresses. Her research interests include image processing and artificial neural networks. In particular, she was involved in the development of different algorithms for satellite optical image fusion, and ART and SOM artificial neutral networks for applications in remote sensing. In 2012, she joined the Center for Biomedical Technology of the UPM, where she was involved in research and development projects for text and image mining in the health care domain. As a result of her research, she has directed more than 10 financed research projects and participated in around 15 more.