# Novel Intuitionistic-Based Interval Type-2 Fuzzy Similarity Measures with Application to Clustering

Sahar Cherif, *Member, IEEE*, Nesrine Baklouti, *Senior Member, IEEE*, Hani Hagras, *Fellow, IEEE*, and Adel M. Alimi, *Senior Member, IEEE*

*Abstract*—Similarity measures have been widely used in applications dealing with reasoning, classification and information retrieval. In this paper, we first propose three new Interval Type-2 Fuzzy Similarity measures (IT-2 FSMs) as a dual concept of some semi-metric distances between Intuitionistic Fuzzy Sets (IFSs). We also prove that the extended IT-2 FSMs satisfy many common properties (i.e. reflexivity, transivity, symmetry and overlapping). Experiments are carried out on a variety of datasets including UCI Learning Machine and real data. Comparative studies between the proposed IT-2 FSMs and the other well-known existing similarity measures (Gorzalczany, Bustince, Mitchell, Zeng and Li as well as VSM and Jaccard) are performed. Obviously, the best results are obtained with the IT-2 FSMs being resilient to the high levels of uncertainty noise. We also prove that our IT-2 FSMs can overcome the drawbacks of some existing similarity measures based on the accuracy rate measure. In addition, the proposed IT-2 FSMs are joined with Fuzzy c-means algorithm as a clustering method and the proposed system is compared against the existing clustering algorithms (Type-1 Fuzzy k-means, Type-1 and Type-2 Fuzzy c-means, Cluster Forest, Bagged Clustering, Evidence Accumulation and Random Projection). Relying on the clustering quality parameters R and C (equivalent to the standard classification accuracy), the advanced IT-2FSMs show higher classification accuracy of about 86% which outperforms nearly the other classifiers.

*Index Terms*—Fuzzy Similarity Measures, Intuitionistic Fuzzy Sets, Interval Type-2 Fuzzy Sets, Fuzzy Distance Measure, Noise, Clustering.

## I. INTRODUCTION

**F**UZZY Logic has been intensively utilized in several applications like pattern recognition, decision making, linguistic knowledge and classification. According to [1], Type-1 Fuzzy systems have the common problem that they cannot fully handle or accommodate for the linguistic and numerical uncertainties associated with changing and dynamic unstructured environments as they use precise Type-1 Fuzzy Sets (T-1 FSs). T-1 FSs handle the uncertainties associated with the inputs and outputs by using precise and crisp membership. Once the Type-1 membership functions have been chosen, all the uncertainty disappears, because Type-1 membership functions are totally precise. Different Type-1 generalizations such as Hesitant Fuzzy Sets [2], Fuzzy Multi

Sets [3], Intuitionistic Fuzzy Sets (IFSs) [4], Interval and General Type-2 Fuzzy Sets (IT-2 FSs and T-2 FSs respectively) [5]–[9] have been adopted. In fact, T-2 FSs have been applied especially to handle high uncertainty levels in some real world data and applications. IFSs have been employed to reveal hesitancy or imprecise information [10], [11]. They have been also applied in decision making [12], edge detection [13], pattern recognition [14] and many other related problems. Besides, IFSs have been used to deal with data hesitancy in describing the real information status.

In some real world applications, data optimization, feature selection [15], [16], classification and similarity factorization [17], [18] are based essentially on the definition of adequate measures. In this context, several types of measures such as distance [4], entropy [19], correlation [20], divergence [21], dissimilarity [22] and similarity measures [10] were introduced in the literature. In order to indicate the degree of closeness between fuzzy sets (FSs) in applications like pattern recognition, computing with words and data mining, Fuzzy Similarity Measures (FSMs) were utilized.

In fact, several FSMs were presented between Type-1, Type-2 and Intuitionistic fuzzy sets satsifying some mathematical properties which are reflexivity, transivity, symmetry and overlapping [23]. Wu and Mendel compared in [24] Type-1 and Type-2 FSMs. They advanced a new ranking IT-2 FSMs by computing the centroids of words membership functions. Indeed, this similarity measure was not appropriate for invertible and non-invertible fuzzy numbers. Motivated by Wu's study, Garcia *et al.* introduced in [25] some fuzzy distances to compare IT-2 fuzzy numbers using membership functions centroids. Nevertheless, the developed approach was applied only on triangular IT-2 Fuzzy numbers.

Heidarzade *et al.* suggested in [26], a new IT-2 FSM for multiple criteria decision-making problems. By this intelligent manner, good results were obtained while handling multiple criteria decision making problems. However, some limits were found in the satisfaction degree for different alternatives in multiple criteria group decision making.

Baccour *et al.* presented in [14] novel Intuitionistic distance measures to compare images, classify shapes and recognize some Arabic sentences. The best results were obtained by the following distance measures $S_8$, d, $D_E$, $d_2$, W, $S_4$ and d [14]. Modest results are given by $S_6$, $S_L$, M and L [14]. The worst results are obtained with the measures T, $S_{GO}$, P and $S_G$ with 0% recognition rate in Image/shape comparisons [14].

Papakostas introduced in [27] a detailed survey for dis-

tances and similarity between IFSs in real pattern recognition application. Besides, several experiments were carried out on some medical cases and benchmarks giving high rate of classification demonstrating the suitability of similarity measures.

Hesamian presented in [28] a general FSM to rank Interval Type-2 Fuzzy Numbers. The proposed measure was tested on some examples in pattern recognition and multi-criteria group decision making. However, it can only be applied for IT-2 Fuzzy numbers with normal fuzzy numbers.

While many FSMs exist for Type-1, Intuitionistic and Type-2 fuzzy sets, only few ones can be applied between general fuzzy sets. McCulloch *et al.* have proposed in [29] and [30] an extension of some IT-2 FSMs to general FSMs for z-slices presentation. The obtained measures preserved all the similarity properties and showed better results compared to some Type-2 FSMs.

In this paper, we present three new Interval Type-2 FSMs (IT-2 FSMs) based on some Intuitionistic distance measures and operations between IT-2 FSs to determine the relation between hesitance (in IFSs) and uncertainty (in IT-2 FSs). This fusion led to Intuitionitic Type-2 Fuzzy similarity measures. We provide mathematical justification in order to demonstrate that the proposed FSMs satisfy proximity properties. Through experiments performed on benchmarks and linguistic data, comparisons are made with other IT-2 FSMs showing consistent and significant results. We finally join the advanced measures with Fuzzy c-means algorithm for Type-2 fuzzy data. The clustering results are evaluated through quality clustering metrics and compared with other clustering methods.

This document is structured as follows: Section II presents an overview on Intuistionistic and IT-2 FSs. Section III lists some existing Intuitionistic and Type -2 FSMs. Section IV depicts the three new Interval Type-2 FSMs followed by proofs and justifications. We illustrate the effectiveness of the proposed measures by numerical examples on some real world data in section V. Finally, section VI provides some conclusions and exposes some future works.

## II. OVERVIEW ON INTUITIONISTIC AND INTERVAL TYPE-2 FUZZY SETS

Intuitionistic Fuzzy Sets (IFS) were introduced by Atanassov in [31] to handle vagueness and uncertainty as cited by Baccour in [10]. An IFS B was illustrated by a truth and a false membership functions. It was expressed in the universe of discourse X by the following equation:

$$B = <x, \mu_B(x), \lambda_B(x)> \ \forall x \in X \quad (1)$$

The truth membership function of $x \in X$ to B is $\mu_B(x)$ while $\lambda_B(x)$ is the false or the non- membership function. To specify the belonging of x to B, an uncertainty function named hesitancy function was also used to measure the belonging of x to B. This function was computed as shown below:

$$\Pi_B(x) = 1 - \mu_B(x) - \lambda_B(x) \quad (2)$$

with $0 \le \Pi_B(x) \le 1$. An Interval Type-2 Fuzzy set $\tilde{A}$ was expressed in the universe of discourse X by a Type-2 membership function $\mu_{\tilde{A}}(x, u)$:

$$\tilde{A} = \int_{x \in X} \int_{u \in J} 1/(x, u) = \int_{x \in X} [\int_{u \in J} 1/u]/x \quad (3)$$

with $J_x \in [0 \ 1]$ and $\int \int$ design the union of all the admissible x in u.

## III. EXISTING FUZZY SIMILARITY MEASURES

### A. Intuitionistic Fuzzy Similarity Measures

In the literature, several similarity and distance measures were proposed between IFSs. We consider two IFSs A and B which have n elements. $\mu_A(x)$ and $\lambda_A(x)$ denote respectively the membership and the non membership functions of the Intuitionistic set A; $\mu_B(x)$ and $\lambda_B(x)$ are the correspondent ones for the IFS B.

*1) Chen's Intuitionistic similarity Measure:* Chen advanced in [32] the first similarity measure between Intuitionistic or vague sets as follows:

$$s_C(A, B) = 1 - \frac{\sum_{i=1}^{n} |s_A(x_i) - s_B(x_i)|}{2n} \quad (4)$$

with $s_A(x_i) = \mu_A(x_i) - \lambda_A(x_i)$, $s_B(x_i) = \mu_B(x_i) - \lambda_B(x_i)$. Several studies (e.g. [33]) showed illogic results of false numerical values in some cases. In fact, high similarity values were obtained even with two disjoint sets.

*2) Hong and Kim's Intuitionistic Similarity Measure:* To overcome Chen's measure limitations, some researchers (as Hong and Kim) introduced, in [33], a new Intuitionistic similarity as illustrated in the following equation:

$$s_H(A, B) = 1 - \frac{\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)| + |\lambda_A(x_i) - \lambda_B(x_i)|}{2n} \quad (5)$$

Although this proposed similarity gave good results mainly for symmetry problems, it does not fit well in some cases. In fact $s_H$ cannot distinguish positive difference from negative difference [34].

*3) Hung's Intuitionistic Similarity Measure:* Motivated by the dual concept context between similarity and distance, Hung developed, in [35], a new Intuitionistic similarity measure as a generalization of the Hamming distance. The equation below illustrates the Hung's similarity:

$$s_{Hu}(A, B) = 1 - \sum_{i=1}^{n} \frac{1 - \frac{1}{2}(|\mu_A(x_i) - \mu_B(x_i)| + |\lambda_A(x_i) - \lambda_B(x_i))|}{n} \quad (6)$$

This Intuitionistic similarity was evaluated on some pattern recognition and classification applications, experimental results show the superiority of the proposed measures for students' evaluation. Some illogic results were found by Szmidt and Kacprzyk in [11].

*4) Baccour's Intuitionistic Distance Measures:* Baccour *et al.* generated, in [14], new intuitionistic distance measures as an extension of the Canberra distance suggested in [36]. The Intuitionistic generated distances $D_{B1}$ and $D_{B2}$ were seen as semi-metric ones:

$$D_{B1}(A, B) = \frac{1}{2n} \sum_{i=1}^{n}$$
$$(\frac{|\mu_A(x_i) - \mu_B(x_i)|}{\mu_A(x_i) + \mu_B(x_i)} + \frac{|\lambda_A(x_i) - \lambda_B(x_i)|}{(\lambda_A(x_i) + \lambda_B(x_i))} \quad (7)$$

$$and \qquad D_{B2}(A, B) = \frac{1}{2n} \sum_{i=1}^{n}$$
$$(\frac{|\mu_A(x_i) - \mu_B(x_i)|}{\mu_A(x_i) + \mu_B(x_i)} + \frac{|\lambda_A(x_i) - \lambda_B(x_i)|}{(2 - \lambda_A(x_i) - \lambda_B(x_i))} \quad (8)$$

Baccour *et al.* applied the obtained measures for shapes classification and Arabic sentences recognition. They obtained some good results and unreasonable ones Authors extended in [14] the Square Chord's distance [37] for IFSs and proposed the following semi-metric distance as:

$$D_{B3}(A, B) = \frac{1}{2n} \sum_{i=1}^{n}$$
$$(\sqrt{|\mu_A(x_i) - \mu_B(x_i)|} + \sqrt{|\lambda_A(x_i) - \lambda_B(x_i)|})^2 \quad (9)$$

Baccour *et al.* also advanced in [14] a new Intuitionistic measure based on some operators like min, max, intersection and union between IFSs . They demonstrated via numerical examples that the distance measure presented in the following equation satisfies the four similarity properties but presents modest results.

$$M_B(A, B) =$$
$$\frac{\sum_{i=1}^{n} min(\mu_A(x_i), \mu_B(x_i)) * min(\lambda_A(x_i), \lambda_B(x_i))}{\sum_{i=1}^{n} max(\mu_A(x_i), \mu_B(x_i)) * max(\lambda_A(x_i), \lambda_B(x_i))} \quad (10)$$

### B. Interval Type-2 Fuzzy Similarity Measures

In the literature, some IT-2 FSMs were advanced to handle uncertainties between vague concepts. We consider 2 Interval Type-2 Fuzzy Sets $\tilde{A}$ and $\tilde{B}$.

*1) Gorzalczany's Type-2 similarity measure:* Gorzalczany defined, in [38], a new proximity measure for interval valued fuzzy sets which provided good computational effectiveness and adequacy. The Gorzalczany's Type-2 similarity measure is :

$$S_G(\tilde{A}, \tilde{B}) =$$
$$[min(\frac{max(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x))}{max \ \underline{\mu}_A(x)}), \frac{max(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x))}{max \ \overline{\mu}_A(x)}),$$
$$max(\frac{min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x))}{max \ \underline{\mu}_A(x)}), \frac{min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x))}{max \ \overline{\mu}_A(x)})] \quad (11)$$

When dealing with linguistic terms, the Gorzalczany's measure does not satisfy reflexivity and gives insignificant results. Some researchers like Tsiporkova and Zimmermann considered in [39] that type-2 Gorzalczany's similarity is not a fuzzy one.

*2) Bustince's Type-2 similarity measure:* Bustince advanced in [40] a new Type-2 fuzzy similarity measure which verifies the inclusion grade indicators for interval-valued fuzzy sets. The Bustince's Type-2 similarity measure is :

$$S_B(\tilde{A}, \tilde{B}) = [s_L(\tilde{A}, \tilde{B}), s_U(\tilde{A}, \tilde{B})] \quad (12)$$

where $s_L$ and $s_U$ designate the lower and upper similarity measures defined respectively by:

$$s_L(\tilde{A}, \tilde{B}) = \Upsilon_L(\tilde{A}, \tilde{B}) * \Upsilon_L(\tilde{B}, \tilde{A})$$
$$s_U(\tilde{A}, \tilde{B})] = \Upsilon_U(\tilde{A}, \tilde{B}) * \Upsilon_U(\tilde{B}, \tilde{A})$$

$$\Upsilon_L(\tilde{A}, \tilde{B}) = inf_{x \in X}$$
$$\left\{ 1, min(1 - \underline{\mu}_{\tilde{A}}(x) + \underline{\mu}_{\tilde{B}}(x), 1 - \overline{\mu}_{\tilde{A}}(x) + \overline{\mu}_{\tilde{B}}(x)) \right\}$$

$$\Upsilon_U(\tilde{A}, \tilde{B}) = inf_{x \in X}$$
$$\left\{ 1, max(1 - \underline{\mu}_{\tilde{A}}(x) + \underline{\mu}_{\tilde{B}}(x), 1 - \overline{\mu}_{\tilde{A}}(x) + \overline{\mu}_{\tilde{B}}(x)) \right\}$$

As explained by Mendel and Wu in [23], the Bustince's similarity does not satisfy the overlapping property.

*3) Mitchell's Type-2 similarity measure:* Mitchell introduced in [41] a general type-2 fuzzy similarity measure for ordering Type-2 fuzzy sets as shown below:

$$S_M(\tilde{A}, \tilde{B}) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} s_{mn} \quad (13)$$

where $s_{mn} = s(A_c^m, A_c^n)$ and s can be any similarity measure for Type-1 fuzzy sets. This measure shows good results for formulating classification problems in pattern recognition. But because embedded Type-1 fuzzy sets can be generated randomly, the overall algorithm is too long and random values for Mitchell's measure can be obtained. Finding obtained by applying Mitchells measure can change from one experiment to another. As proved by Dongrui in [24], Mitchell's similarity presents also other limitations. In fact, it cannot satisfy reflexivity and symmetry properties.

*4) Zeng and Li's Type-2 Similarity Measure:* Zeng and Li introduced in [42] a new interval valued similarity measure for the continuous universe of discourse as demonstrated below:

$$s_Z(\tilde{A}, \tilde{B}) = 1 - \frac{1}{2n} \sum_{i=1}^{n}$$
$$(|\underline{\mu}_{\tilde{A}}(x_i) - \underline{\mu}_{\tilde{B}}(x_i)| + |\overline{\mu}_{\tilde{A}}(x_i) - \overline{\mu}_{\tilde{B}}(x_i)|) \quad (14)$$

They also defined a novel proximity measure for the discrete universe of discourse as follows:

$$s_Z(\tilde{A}, \tilde{B}) = 1 - \frac{1}{2(b - a)}$$
$$\int_a^b ((|\underline{\mu}_{\tilde{A}}(x) - \underline{\mu}_{\tilde{B}}(x)| + |\overline{\mu}_{\tilde{A}}(x) - \overline{\mu}_{\tilde{B}}(x)|)dx \quad (15)$$

Zeng and Li proved that these similarity measures and entropies of interval valued fuzzy sets can be changed by each other. However, according to [23], the Zeng and Li's measure does not preserve the overlapping property. In fact, for disjoint fuzzy sets, when the distance between sets rises, the Zeng and Li's similarity increases proportionally which is irrelevant.

*5) Vector Similarity Measure for Interval Type-2 Fuzzy Sets* : In [23], Mendel and Wu defined a new vector similarity measure (VSM) as an extension of Yager's measure for Type-1 fuzzy sets. The VSM formula is written below:

$$s_v(\tilde{A}, \tilde{B}) = (s_1(\tilde{A}, \tilde{B}), s_2(\tilde{A}, \tilde{B}))^T \qquad (16)$$

$$with \qquad s_1(\tilde{A}, \tilde{B}) =$$

$$\frac{\int_x min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i))dx + \int_x min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i))dx}{\int_x max(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i))dx + \int_x max(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i))dx}$$

$$and \qquad s_2(\tilde{A}, \tilde{B}) = e^{-rd(\tilde{A}, \tilde{B})}$$

with r is a constant $> 0$. Based on some demonstrations and definitions, Mendel and Wu showed that VSM can resolve reflexivity, overlapping and symmetry problems found in the above-described similarities especially when dealing with linguistic data or computing with words applications. In fact, VSM can generate close sets depending on the shape of their membership functions. Some limitations were mentioned in [43] showing that VSM cannot classify crisp numbers and is not appropriate for invertible and non-invertible fuzzy numbers.

*6) Jaccard's Type-2 Similarity Measure:* Wu and Mendel generalized, in [44], the Type-1 Jaccard measure [45] to IT-2 FSs and presented it as:

$$s_J(\tilde{A}, \tilde{B}) =$$

$$\frac{\int_x min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x))dx + \int_x min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x))dx}{\int_x max(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x))dx + \int_x max(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x))dx}$$

$$(17)$$

The main advantage of the Jaccard's Type-2 measure is that it fulfills the four similarity properties and has shown reasonable results in computing fuzzy sets analogy.
A comparison of reflexivity, symmetry, transivity and overlapping properties between the above-described Type-2 FSMs is summarized in Table I.

TABLE I
COMPARISON OF PROXIMITY PROPERTIES BETWEEN MITCHELL (M), GORZALCZANY (G), BUSTINCE (B), ZENG AND LI (Z), VSM (V) AND JACCARD (J) SIMILARITIES

| Properties | IT-2 Fuzzy Similarity Measures | | | | | |
|---|---|---|---|---|---|---|
| | M | G | B | Z | V | J |
| Overlapping | ✓ | ✓ | | | ✓ | ✓ |
| Symmetry | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Reflexivity | | | ✓ | ✓ | ✓ | ✓ |
| Transivity | ✓ | ✓ | ✓ | ✓ | | ✓ |

## IV. THE PROPOSED IT-2 FUZZY SIMILARITY MEASURES: FROM IFS TO IT-2 FSS

In this section, we propose three new fuzzy similarity measures between IT-2 fuzzy sets [46]. The main idea is to extend Baccour's semi-metric distance measures [10] for Intuitionistic Fuzzy Sets (see equations 7, 8 and 10). The genesis of this extension is based on the following motivations: Because similarity and distance are dual proximity concepts, we consider the semi-metric distances presented in [10] as

similarity measures. Originally, Atanassov in [31] advanced the mapping relation from IFS to FS. Some studies (e.g. [10], [47]–[49]) confirmed Atanassov's theory and considered that every classical fuzzy set can be seen as an Intuitionistic one and vice versa. The hesitancy function $\Pi$ is null (see Equation 2) and the truth and false membership functions are hence dual concepts.

$$\Pi_B(x) = 1 - \mu_B(x) - \lambda_B(x) = 0 \qquad (18)$$

$$then \qquad \lambda_B(x) = 1 - \mu_B(x) \qquad (19)$$

However, several works (e.g. [50]–[52]) demonstrated that an IFS is not similar to a FS. Indeed, it is not evident to determine an exact degree of truth of a FS when dealing with IFSs having an hesitancy function. It is also difficult to set membership and non-membership functions when handling FSs (lack of definitions and axioms [49]). To overcome those drawbacks, T-2 FSs have been used to handle uncertainty and hesitance. A new association between IFSs and T-2 FSs was made in [49], [50] demonstrating that IFSs are closer to T-2 FSs. When using IFSs, there is uncertainty on determining the membership and the non-membership functions with a certain hesitancy. However when dealing with T-2 FSs, there is uncertainty only on the selection of the membership function. The switch from IFSs to T-2 FSs is based on the following statement: " *In the uncertainty, there is hesitance and in the hesitance, there is uncertainty*." Zhu *et al.* demonstrated also, in [52], that IFSs are particular cases of T-2 FSs. In [49], authors mentioned that IFSs were organized similarly to Interval Valued Fuzzy Sets considered as special cases of T-2 FSs. Switching from IFSs to T-2 FSs allowed obtaining good results in decision making [51], clustering [50] and medical diagnosis [49] and led to the introduction of the Intuitionistic Type-2 Fuzzy Sets (InT-2 FSs) handling both hesitance and uncertainty. Mathematically, the mapping from an IFS called $A$ to a T-2 FS $\tilde{A}$ is defined as follows:

$$\tilde{A} = \sum [1/(\mu_{\tilde{A}}(x) + p\Pi_A(x)) + 0/(1 - \lambda_A(x) + p\Pi_A(x))]/x \qquad (20)$$

where $p \in [0, 1]$. The true and false membership functions are represented by the secondary grade in Equation 20. When the hesitancy function $\Pi_A(x)$ is null, the $\tilde{A}$ is illustrated by the following formula:

$$\tilde{A} = \sum [1/(\mu_{\tilde{A}}(x)] \qquad (21)$$

The obtained expression (Equation 21) is very compatible to the general definition of an IT-2FS as described in Equation 3. For this reason, we generalize each IFS to an IT-2FS by modifying the false and truth degrees to upper and lower membership functions to handle vague data. We present the three proposed IT-2 FSMs S1, S2 and S3, each of which is justified and proved in order to define its own properties.

### A. Interval Type-2 Fuzzy Similarity Measure S1

The Interval Type-2 fuzzy similarity measure S1 is the dual concept of the semi metric distance described in Equation

TABLE II
COMPARISON OF PROXIMITY PROPERTIES BETWEEN THE PROPOSED IT-2
FSMS: S1, S2 AND S3

| | Proposed IT-2 FSMs | | |
|---|---|---|---|
| **Properties** | **S1** | **S2** | **S3** |
| Overlapping | ✓ | ✓ | ✓ |
| Symmetry | ✓ | | ✓ |
| Reflexivity | ✓ | ✓ | |
| Transivity | ✓ | ✓ | ✓ |

7 between IT-2 FSs. By analogy to Type-2 Fuzzy Logic, the S1 measure is defined through upper and lower values (respectively $\overline{s}1$ and $\underline{s}1$). The false membership function $\lambda(x)$ of any IFS is replaced by $1 - \mu(x)$. For 2 IT-2 FSs $\tilde{A}$ and $\tilde{B}$, Equation 7 will be transformed and written as follows:

$$S1(\tilde{A}, \tilde{B}) = 1 - \frac{(\overline{s}1 + \underline{s}1)}{2} \tag{22}$$

$$with \qquad \overline{s}1 = \frac{1}{2n}\sum_{i=1}^{n}$$

$$(\frac{|\overline{\mu}_{\tilde{A}}(x_i) - \overline{\mu}_{\tilde{B}}(x_i)|}{\overline{\mu}_{\tilde{A}}(x_i) + \overline{\mu}_{\tilde{B}}(x_i)} + \frac{|\overline{\mu}_{\tilde{B}}(x_i) - \overline{\mu}_{\tilde{A}}(x_i)|}{2 - \overline{\mu}_{\tilde{A}}(x_i) - \overline{\mu}_{\tilde{B}}(x_i)}$$

$$and \qquad \underline{s}1 = \frac{1}{2n}\sum_{i=1}^{n}$$

$$(\frac{|\underline{\mu}_{\tilde{A}}(x_i) - \underline{\mu}_{\tilde{B}}(x_i)|}{\underline{\mu}_{\tilde{A}}(x_i) + \overline{\mu}_{\tilde{B}}(x_i)} + \frac{|\underline{\mu}_{\tilde{B}}(x_i) - \underline{\mu}_{\tilde{A}}(x_i)|}{2 - \underline{\mu}_{\tilde{A}}(x_i) - \underline{\mu}_{\tilde{B}}(x_i)}$$

**Proof and Justification:** See Appendix Proof of S1 properties. Through proof and justification, we showed that S1 measure satisfies the four analogy properties.

### B. Interval Type-2 Fuzzy Similarity Measure S2

The Interval Type-2 Fuzzy similarity Measure S2 is defined by computing the mean of lower and upper values ($\overline{s}2$ and $\underline{s}2$). The latter are generalized from the semi-metric distance measure between IFS presented in Equation 8 . The false membership function $\lambda(x)$ written in Equation 8 is modified and computed as $1 - \mu(x)$. For two IT-2 FSs $\tilde{A}$ and $\tilde{B}$, the following formula is hence obtained:

$$S_2(\tilde{A}, \tilde{B}) = 1 - \frac{(\overline{s}2 + \underline{s}2)}{2} \tag{23}$$

$$\overline{s}2 = \frac{1}{2n}\sum_{i=1}^{n}(\frac{|\overline{\mu}_{\tilde{A}}(x_i) - \overline{\mu}_{\tilde{B}}(x_i)|}{\overline{\mu}_{\tilde{A}}(x_i) + \overline{\mu}_{\tilde{B}}(x_i)} + \frac{|\overline{\mu}_{\tilde{B}}(x_i) - \overline{\mu}_{\tilde{A}}(x_i)|}{\overline{\mu}_{\tilde{B}}(x_i) - \overline{\mu}_{\tilde{A}}(x_i)}$$

$$\underline{s}2 = \frac{1}{2n}\sum_{i=1}^{n}(\frac{|\underline{\mu}_{\tilde{A}}(x_i) - \underline{\mu}_{\tilde{B}}(x_i)|}{\underline{\mu}_{\tilde{A}}(x_i) + \overline{\mu}_{\tilde{B}}(x_i)} + \frac{|\underline{\mu}_{\tilde{B}}(x_i) - \underline{\mu}_{\tilde{A}}(x_i)|}{\underline{\mu}_{\tilde{B}}(x_i) - \underline{\mu}_{\tilde{A}}(x_i)}$$

**Proof and justification:** See Proof of S2 properties in Appendix. Through proof and justification, we proved that S2 measure satisfies only reflexivity, transivity and overlapping properties.

### C. Interval Type-2 Fuzzy Similarity Measure S3

The Interval Type-2 Fuzzy Similarity Measure S3 and the distance presented in Equation 10 are dual concepts ( similarity = 1 - Distance). As described for the two previous proposed IT-2 FSMs, S3 can be defined through upper and lower values. The migration from the semi-metric distance between IFS to similarity between IT-2 FSs is illustrated by the following formula:

$$S3(\tilde{A}, \tilde{B}) = 1 - \frac{(\overline{s}3 + \underline{s}3)}{2} \tag{24}$$

$$\overline{s}3 =$$
$$\frac{\sum_{i=1}^{n} min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) * min(1 - \overline{\mu}_{\tilde{A}}(x_i), 1 - \overline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^{n} max(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) * max(1 - \overline{\mu}_{\tilde{A}}(x_i), 1 - \overline{\mu}_{\tilde{B}}(x_i))}$$
$$\underline{s}3 =$$
$$\frac{\sum_{i=1}^{n} min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i)) * min(1 - \underline{\mu}_{\tilde{A}}(x_i), 1 - \underline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^{n} max(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i)) * max(1 - \underline{\mu}_{\tilde{A}}(x_i), 1 - \underline{\mu}_{\tilde{B}}(x_i))}$$

**Proof and justification:** See Proof of S3 properties in Appendix. We demonstrated that S3 does not fulfill the reflexivity property and satisfy the transivity, symmetry and overlapping properties.

The three proposed measures (S1, S2 and S3) are the dual concepts of some semi-metric distances used in pattern recognition, medical diagnosis and decision-making applications. They are effectively Fuzzy Similarity Measures between IT-2 FSs because they satisfy similarity properties (i.e. overlapping, symmetry, reflexivity and transivity) as summarized in Table II.

## V. PERFORMANCE ANALYSIS

### A. Adopted approach

The conformity of the proposed measures was evaluated through some experimentation on examples taken from real data: linguistic, medical, classification, correlation and criminological fields. Comparisons with existing IT-2 FSMs were done and interpretations were given. When dealing with crisp data, each crisp value was substituted by a Type-2 representation having the following form $[leftvalue, rightvalue]$. A certain threshold was used to compute the left and right values in the Type-2 representation. In order to obtain Type-2 membership functions, the Enhanced Interval approach (EIA) [53]–[55], which is a generalization of the Interval Approach (IA) introduced by [56], was applied to model imprecision. The main advantage of the EIA is that it results in thinner and narrower Footprint Of Uncertainty (FOU) than the IA algorithm [57]. The EIA consists mainly in transforming words into IT-2 FSs and generating FOUs for all data attributes. It is composed of two steps: the Data part and the FS part. The former is a pre-processing one in which n data intervals $[a^{(i)}, b^{(i)}]$ are first collected from subjects. Second, the intervals are filtered by eliminating bad data and outliers. However, the latter consists first in computing uncertainty by establishing the FOU's nature (left shoulder, right shoulder or interior). Inadmissible shapes are rejected and both lower and upper membership functions are computed. We describe the different steps of the experimental process followed to

TABLE III
COMPARISON BETWEEN S1, S2 AND S3 AND IT-2 EXISTING SIMILARITY MEASURES FOR MENDEL'S ATTRIBUTES

| | none to very little / extremely low | somewhat high/more or less high | Moderately high/some or less high | high/ very high | very High/ extremely High | Fair/ Very Fair |
|---|---|---|---|---|---|---|
| **Mitchell** | 0.5847 | 0.5082 | 0.4829 | 0.5981 | 0.5033 | 0.5124 |
| **Bustince** | 0.5441 | 0.7433 | 0.6381 | 0.5932 | 0.4572 | 0.8157 |
| **Gorzalczany** | 1 | 0.9577 | **0.9312** | 1 | 0.9937 | 0.9937 |
| **jaccard** | 0.7237 | 0.7689 | 0.6963 | 0.6074 | 0.5270 | 0.8246 |
| **zeng Li** | 0.8929 | 0.9197 | 0.8950 | 0.8123 | 0.7933 | 0.9430 |
| **vsm** | 0.5959 | 0.6689 | 0.5394 | 0.3859 | 0.3796 | 0.7346 |
| **S1** | 0.8718 | 0.9202 | 0.9012 | 0.8655 | 0.8059 | 0.9467 |
| **S2** | 0.5897 | **0.9635** | 0.9225 | **1** | **1** | **1** |
| **S3** | 0.3334 | 0.3057 | 0.2225 | 0.0918 | 0.1107 | 0.4146 |

TABLE IV
COMPARISON BETWEEN S1, S2,S3 AND IT-2 EXISTING SIMILARITY MEASURES FOR WINE'S ATTRIBUTES

| | Ash / Alcanity Of Ash | Flavanoids/ Alcanity of Ash | Ash / Flavanoids | Color intensity / OD280-OD315 of diluted wines | Flavanoids/ Nonflavanoid phenols | Ash/ Nonflavanoid phenols | Alcanity of Ash/ Nonflavanoid phenols |
|---|---|---|---|---|---|---|---|
| **Mitchell** | 0.4184 | 0.5213 | 0.4991 | 0.0858 | 0.4645 | 0.7970 | 0.3987 |
| **Bustince** | 0.5 | 0.8462 | 0.6091 | 0.2231 | 0.5369 | 0.9149 | 0.5 |
| **Gorzalczany** | 0.7074 | 0.8479 | 0.7972 | 0.3226 | **0.9881** | **0.9934** | **0.9985** |
| **jaccard** | 0.4484 | 0.8126 | 0.5575 | 0.1031 | 0.4912 | 0.8813 | 0.4033 |
| **zeng Li** | 0.7599 | **0.9284** | 0.8086 | 0.6111 | 0.7738 | 0.9470 | 0.7350 |
| **vsm** | 0.2408 | 0.6717 | 0.3416 | 0.1029 | 0.2842 | 0.8089 | 0.2014 |
| **S1** | 0.7012 | 0.8908 | 0.7979 | 0.4356 | 0.7625 | 0.9584 | 0.6745 |
| **S2** | **0.7836** | 0.9221 | **0.8634** | **0.7588** | 0.8376 | 0.9908 | 0.7635 |
| **S3** | 0.0954 | 0.3039 | 0.1805 | 0.0178 | 0.1558 | 0.5157 | 0.0892 |

compute similarity measures between two IT-2 FSs based on the proposed IT-2 FSMs. Our IT-2 FSMs were evaluated on both noise free and noisy data. The results obtained are described below.

- **Step 1:** For each Attribute, applying the three steps of the Enhanced Interval Approach: Data pre-processing, establishing FOU Nature and computing the Upper Membership Functions (UMF) and the Lower Membership Functions (LMF). Consequently, each attribute will be presented by an IT-2FS.
- **Step 2:** Computing the similarity between two IT-2FSs. Applying S1 (see Equation 22), S2 (see Equation 23) and S3 (see Equation 24) measures.
- **Step 3:** Comparing between the proposed Similarity measures with other existing IT-2 FSMs.
- **Step 4:** Combining the proposed Similarity Measures with Fuzzy C Means algorithm for Type-2 fuzzy data.
- **Step 5:** Comparing with other clustering algorithms

### B. Noise Free Data

*1) Mendel's Data:* The Mendel's databases [58], [59] are the most known and used data in the linguistic field when the uncertainty rate is very high. The utilized features were collected from the Mendel's databases "Data72" and "Data73". These data are composed of 40 objects and 15 words. Type-2 degrees of truth were obtained and FOUs were generated for the following attributes: "None to very Little", "Extremely Low", "Somewhat High", "Moderately High", "More or less

High", "Some or less high", "High", "Very High", "Extremely High", "Fair" and "Very Fair". The proposed IT-2 FSMs (S1, S2 and S3) were applied on some words and compared to the existing IT-2 ones as depicted in Table III. Generally, attributes are considered close based on their membership functions (MFs) and their FOUs which is confirmed with our IT-2 FSMs. Table III presents the similarity values between the above-mentioned attributes. The proposed IT-2 FSMs S1, S2 and Gorzalczany measures gave the best results compared to other existing IT-2 FSMs especially for approximate degrees of truth. For example, for the attributes "Moderately High" and "Some or Less High", S1 value is equal to 0.9012, S2 is equal to 0.9225 and Gorzalczany measure is about 0.9312. S1 and S2 results were reasonable; However, for similar cases, S3 measure gave lower values even for similar cases which is illogical and invalid. The S2 measure achieved values which are in average 0.17% more efficient than the other existing IT-2FSMs for Mendel Data; whereas the S1 measure achieved a similarity performance which is in average near to 0.14%. The main advantage of S1 measure is that it satisfies all the proximity properties while Gorzalczany and S2 measures do not fulfill the reflexivity and symmetry properties respectively (see Table I).

*2) Wine Data:* The Wine data [60], which are chemical analysis' results of wines, are composed of 178 instances and 13 attributes. The experimental process was applied on Wine attributes in order to generate IT-2 FSs and compute IT-2 FSMs. Table IV reveals that S1 and S2 provided the best
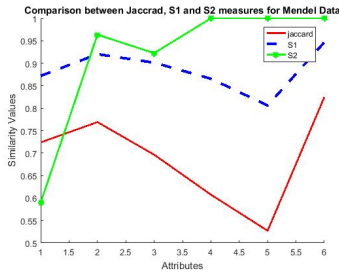
Fig. 1. Comparison Between Jaccard, S1 and S2 measures for Mendel Data
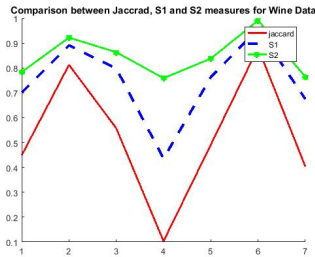


Fig. 2. Comparison between Jaccard, S1 and S2 measures for Wine Data

results compared to other similarities especially S2. For example, for the attributes "Alcanity of Ash" and "Flavanoids", S1 measure is equal to 0.8908; S2 measure is about 0.9221 while Jaccard measure is equal 0.8126. Although, Gorzalczany measure is equal to 0.8479, it did not satisfy all the similarity measures properties. Besides, some researchers considered in [39] that the Gorzalczany similarity measure is not a fuzzy one. The S2 measure allowed reaching higher values which are in average 0.23% more efficient than the other existing similarity measures. The S1 measure achieved, in average, 0.11%. Since Jaccard and S1 measures satisfied all the similarity properties, and S2 gave the best overall results, a comparison between them was done (Figure 1 for Mendel data and Figure 2 for Wine data). Obviously, S1 and Jaccard measures present approximate curves while S2 shows high values for closer sets or approximate shapes.

For the rest of our tests, we focus on Jaccard, S1 and S2 similarity measures (recall that S1 and Jaccard are the only measures that fulfilled the following properties: symmetry, reflexivity, transivity and overlapping).

*3) Mammography Data:* The mammography data presented in [61] is composed of 961 instances and 6 attributes which are "BI-RADS assessment", "Age", "Margin", "Shape", "Density" and "Severity" (benign or malignant). Shape can be round, oval, lobular or irregular. However, Margin can be circumscribed, microlobulated, obscured, ill-defined and spiculated. Density can be high, iso, low and fat-containing. In our experience, we concentrated only on "Margin", "Density" and "Shape" attributes. Data was transformed into left and right values in order to generate IT-2 FSs for each attribute based on the EIA method. The measures computed for S1, S2 and Jaccard are summarized in Table V. Obtained results demonstrate approximate values between all IT-2FMs except for S3 measure which provided lower values even for similar vectors.

TABLE V
COMPARISON BETWEEN S1, S2 AND JACCARD SIMILARITY MEASURES
FOR MAMMOGRAPHY 'S ATTRIBUTES

|  | Jaccard | S1 | S2 |
|---|---|---|---|
| microbulated/circumscribed | 0.5932 | **0.6594** | 0.61 |
| microbulated/obscured | 0.6170 | 0.6778 | **0.6923** |
| microbulated/spiculated | 0.6099 | **0.7149** | 0.5076 |
| obscured/ spiculated | **0.9109** | 0.8552 | 0.8297 |
| spiculated / ill defined | 0.7146 | 0.7333 | **0.9582** |
| oval / lobular | **0.8326** | 0.8241 | 0.8402 |
| Iso / fat containing | 0.7565 | **0.7994** | 0.3435 |

For example, for Microbulated and Spiculated attributes, having closer IT-2 FSs, S3 measure is about 0.1496, S2 is equal to 0.5076, S1 is equal to 0.7149 while Jaccard measure is about 0.6099. S1 measure achieved higher similarity accuracy which is in average 0.2% more efficient than that of Jaccard measure. Figure 3 illustrates the comparison between S1, S2 and Jaccard measures for mammography Data. Jaccard and S1 measures satisfied all the similarity properties. They also presented approximate curves, as described in Figure 3.

*4) Glass Data:* The Glass data, presented in [60], is composed of 214 instances and 10 attributes which are "Id Number", "RI" (Refractive index), "Na" (Sodium), "Mg" (Magnesium), "Al" (Aluminum), "Si" (silicon), "k" (Potassium), "Ca" (Calcium), "Ba" (Barium), "Fe" (Iron) and "Type of glass". After generating an IT-2FS for each attribute, the proximity between features was calculated and compared to the other existing IT-2FSMs. The attributes used in experimentation are : RI , MG, AL and K. Table VI summarizes all the obtained values for S1, S2 and Jaccard for RI, MG, AL and K attributes. For example, between RI and Al attributes, Jaccard measure is equal to 0.4114, S1 is equal to 0.6826 and S2 is about 0.7744. S1 and S2 measures achieved higher performance which are in average 0.3% and 0.79%, respectively more efficient than Jaccard. A comparison between S1, S2 and Jaccard measures is illustrated in Figure 4 demonstrating that S1 and S2 proximity measures always gave better results than Jaccard measure.

TABLE VI
COMPARISON BETWEEN S1, S2 AND JACCARD SIMILARITY MEASURES
FOR GLASS'S ATTRIBUTES

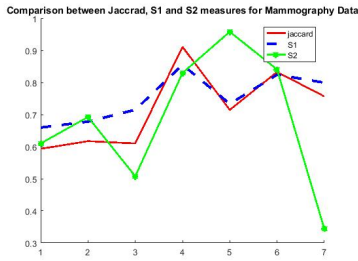|  | Jaccard | S1 | S2 |
|---|---|---|---|
| **Refractive Index / Magnesium** | 0 | **0.4030** | 0.3719 |
| **Refractive Index / Aluminum** | 0.4114 | 0.6826 | **0.7744** |
| **Magnesium / Aluminum** | 0 | 0.4033 | **0.7812** |
| **Aluminum/Potassuim** | 0.0234 | 0.3501 | **0.7296** |
| **Magnesium / Potassuim** | 0 | 0.4291 | **0.7994** |

Fig. 3. Comparison between S1, S2 and Jaccard Measures for Mammography Data

TABLE VII
GENERATED FOUs FROM MENDEL'S DATA UNDER NOISY AND NOISE FREE ENVIRONMENTS WITH STANDARD DEVIATION = 0.1

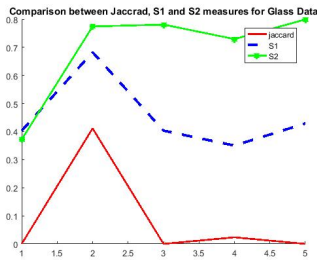| Attribute | noise-Free FOU | noisy FOU |
|---|---|---|
| Very High | | |
| Fair | | |



Fig. 4. Comparison between S1, S2 and Jaccard Measure for Glass data

## C. noisy Data

Compared to the existing IT-2 FSMs, our advanced IT-2 FSMs S1 and S2 provided relevant results when dealing with natural language (Mendel's data) and numerical examples (Glass, Wine and Mammography data). The robustness of the proposed IT-2 FSMs was evaluated when a White Gaussian noise was added to the initial Data with zero mean and Standard Deviations equal to 0.1 and 0.01. We used the same databases under the same conditions for faithful comparison.

*1) noisy Mendel's Data:* The Mendel's Data was affected by a White Gaussian Noise with different standard deviation values (0.01 and 0.1). New IT-2 FSs were also generated (some examples are listed in Table VII: the attributes "Very high" and "Fair") showing that noise can affect the shapes of FOUs leading to new numerical MFs.

With White Gaussian, results are presented in Table VIII (with mean=0 and deviation= 0.1) and Table IX (with mean= 0 and deviation= 0.01). In the case of deviation equal to 0.1 and for the attributes "Moderately High" and "Some or Less High" for example, S1 measure is equal to 0.7074; Jaccard is about 0.7412 and S2 measure is equal to 0.6068. Although

TABLE VIII
COMPARISON BETWEEN S1, S2 AND JACCARD SIMILARITY MEASURES FOR NOISY MENDEL'S ATTRIBUTES WITH STANDARD DEVIATION = 0.1

| | Jaccard | S1 | S2 |
|---|---|---|---|
| None to very little/ Extremly Low | 0.6125 | **0.8875** | 0.942 |
| some what high / more or less high | 0.6512 | **0.8302** | 0.7267 |
| Moderately High / Some or less high | 0.7074 | **0.7412** | 0.6068 |
| high/very high | 0.6970 | 0.7076 | **0.9717** |
| Very High / Extremely High | 0.4681 | 0.6908 | **1** |
| fair /very fair | 0.7133 | 0.8108 | **1** |

TABLE IX
COMPARISON BETWEEN S1, S2 AND JACCARD SIMILARITY MEASURES FOR NOISY MENDEL'S ATTRIBUTES WITH STANDARD DEVIATION = 0.01

| | Jaccard | S1 | S2 |
|---|---|---|---|
| None to very little/ Extremly Low | 0.5198 | 0.8510 | **0.9171** |
| some what high / more or less high | 0.7324 | 0.9023 | **0.9423** |
| Moderately High / Some or less high | 0.6823 | 0.8973 | **0.9444** |
| High/ Very high | 0.6524 | 0.8816 | **1** |
| Very High / Extremly High | 0.7825 | 0.9201 | **1** |
| fair /very fair | 0.8196 | 0.8642 | **1** |

Zeng and Li and Gorzalczany measures gave also high values for similar attributes, they did not satisfy the overlapping and the reflexivity properties which are fundamental in applications like classification, knowledge reduction and decision making. Obtained results demonstrated in Tables VIII and IX show that S1 and S2 measures provided better findings by achieving higher similarity accuracy which is in average 0.3% more accurate than that obtained applying Jaccard measure. The observation confirms the robustness of our measures under noisy environments.

*2) noisy Glass Data:* The previously-described Glass data was also affected by a White Gaussian Noise with zero mean and different standard deviation values (0.1 and 0.01). New IT-2 FSs were hence generated. S1, S2 and Jaccard were applied between the different FOUs and compared to the existing IT-2 FSMs. The results obtained are illustrated in Table X showing the effectiveness of S1 measure even when affected by noise. Zeng and Li measure also gave reasonable and good results compared to the other IT-2 FSMs. However, we should recall that this measure did not satisfy the overlapping property while S1 and Jaccard satisfied all the similarity properties. For example, for RI and Al attributes, S1 measure is equal to 0.7454 while Jaccard measure is about 0.4757 under noise free (see Table VI). When affected by a Gaussian Noise with standard deviation = 0.1, S1 is equal to 0.9259 and Jaccard is about 0.8179 (see Table X). This result confirms our idea that S1 is an appropriate measure between IT-2 FSs even under noisy environments.

TABLE X
COMPARISON BETWEEN S1, S2 AND JACCARD SIMILARITY MEASURES FOR GLASS'S ATTRIBUTES WITH STANDARD DEVIATION = 0.1

| | Jaccard | S1 | S2 |
|---|---|---|---|
| **Refractive Index / Magnesium** | 0.1377 | 0.4627 | **0.6760** |
| **Refractive Index / Aluminum** | 0.8179 | 0.9259 | **0.9414** |
| **Magnesium / Aluminum** | 0.1119 | **0.4085** | 0.3946 |

As illustrated in Figures 1, 2, 3 and 4, S1 and S2 similarity measures provided relevant results even under noisy environ-

TABLE XI
COMPARISON OF SOME CLUSTERING METHODS FOR THE METRIC R

| Dataset | FKM | FCM | FCM2 [62] | CF [63] | BC2 [64] | EA [65] | RP [66] | S1M | S2M |
|---|---|---|---|---|---|---|---|---|---|
| Statlog (Heart) (%) | 51.54 | 52.13 | 52.13 | 56.8 | 51.50 | **53.20** | 52.41 | 52.86 | 53.01 |
| Image Segmentation (%) | 83.95 | 84.39 | 84.49 | 79.71 | 82.19 | 85.75 | **85.88** | 83.78 | 83.52 |
| Robot Execution Failures (%) | 59.83 | 65.28 | 66.09 | 63.42 | 39.76 | 58.31 | 41.52 | 66.1 | **66.21** |
| SPECT Heart (%) | 51.11 | 51.00 | 50.78 | **56.78** | 50.61 | 51.04 | 49.89 | 50.48 | 51.23 |
| Wine (%) | 71.78 | 72.04 | 72.71 | **79.70** | 71.97 | 71.86 | 71.94 | 73.47 | 73.98 |
| Breast Cancer Wisconsin (Diagnostic) (%) | 75.04 | 76.81 | 77.86 | **79.66** | 74.87 | 75.04 | 74.89 | 78.47 | 78.23 |

TABLE XII
COMPARISON OF SOME CLUSTERING METHODS FOR THE METRIC C

| Dataset | FKM | FCM | FCM2 [62] | CF [63] | BC2 [64] | EA [65] | RP [66] | S1M | S2M |
|---|---|---|---|---|---|---|---|---|---|
| Statlog (Heart) (%) | 59.26 | 60.74 | 60.74 | **68.26** | 59.10 | 59.26 | 60.54 | 61.31 | 61.02 |
| Image Segmentation (%) | **58.86** | 57.62 | 57.05 | 48.24 | 49.91 | 51.30 | 47.71 | 56.95 | 57.02 |
| Robot Execution Failures (%) | 42.94 | 46.63 | 47.24 | 41.2 | 35.37 | 37.19 | 35.50 | **47.31** | 47.29 |
| SPECT Heart (%) | 39.7 | 40.45 | 41.2 | **68.02** | 56.28 | 56.55 | 61.11 | 40.78 | 41.09 |
| Wine | 70.22 | 70.79 | 70.79 | **79.19** | 70.22 | 70.22 | 70.79 | 71.01 | 70.86 |
| Breast Cancer Wisconsin (Diagnostic) (%) | 85.41 | 86.64 | 87.35 | **88.7** | 85.38 | 85.41 | 70.79 | 86.45 | 86.43 |

ments. The similarity measure S2 gave better results than S1 (for Mendel, Wine and Glass data); whereas the similarity measure S1 was good for Mammography data. In practical application, we can use either S1 or S2 measure between IT-2 FSs.

### D. Clustering Results

Similarity Measures are often used in applications dealing with clustering, data reduction or knowledge representation. Clustering is an unsupervised methodology applied to group similar objects into the same clusters or classes. Instances in the same class must be as dissimilar as possible to objects in other classes. In general, clustering methods utilize distance or similarity functions like Euclidean, Hamming, Manhattan, Minkowski distances [67], Cosine [68] or Jaccard [24] measures to group items. The performance of any clustering algorithm depends on selecting an appropriate distance or similarity function over the input data set. As cited in [67], distance-based clustering (depending on a distance function) shows some limits in capturing correlations among the input data objects. Similarity-based clustering relies on grouping objects that are not only physically close (distance) but also have similar patterns. Similarity-based clustering (depending on a similarity function) was then adopted for better data classification. In order to prove the effectiveness of the proposed measures (S1 and S2), we combined these measures with Fuzzy c-means algorithm for IT-2FSs. The Euclidean distance was altered with S1 and S2 to determine the belonging of an instance to a specified class. In our clustering experiments, we used 6 datasets from the UCI Machine Learning Repository [63] under noise free environment. These data illustrated in Table XIII are also used to compare our approach with other unsupervised clustering algorithms like Fuzzy k-means (FKM), Cluster Forest (CF) [69], Type-1 and Type-2 Fuzzy C-means (FCM and FCM2 respectively) [62], Bagged Clustering (BC2) [64], Evidence Accumulation (EA) [65] and Random Projection (RP) [66]. As cited in [69], there exist some metrics that can be used for clustering quality comparison. In our experiments, two clustering quality performers were applied

TABLE XIII
DATA COLLECTIONS

| Dataset | Objects | Attributes | Classes |
|---|---|---|---|
| SPECT Heart | 267 | 22 | 2 |
| Image Segmentation | 2100 | 19 | 7 |
| Statlog (Heart) | 270 | 13 | 2 |
| Wine | 178 | 13 | 3 |
| Breast Cancer Wisconsin (Diagnostic) | 596 | 30 | 2 |
| Robot Execution Failures | 164 | 90 | 5 |

in order to evaluate different clustering algorithms. The first metric is called R. It is defined as follows:

$$R = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \sigma(O_i, O_j)\sigma(A_i, A_j)}{N(N-1)}, i \neq j \quad (25)$$

with $\sigma = \begin{cases} 1 & if\ the\ attributes\ are\ the\ same \\ 0 & otherwise \end{cases}$

O is the vector of the original (expected) classes, A represents the vector of the assigned classes and N denotes the number of classified vectors. This metric is based on the assumption that all possible pairs of vectors belonging to the same class in the original class vector, must have the same class in the assigned class vector. The second metric is called C and it is equivalent to the standard classification accuracy. It is represented by the following equation:

$$C = max_{t \subset T} \left\{ \frac{1}{N} \sum_{i=0}^{N} \sigma(t(O_i), A_i) \right\} \quad (26)$$

with T is the set of all class permutations and $\sigma$, $O_i$ and $A_i$ are defined in Equation 25. For the two metrics R and C, the best results correspond to the high value in the clustering experiments. The maximum of threshold of the fitness function was set empirically to 300 for all the used methods for better comparison. In addition, the best-selected method is the approach that can provide the best results with the smallest number of fitness function evaluations. The obtained clustering results are illustrated in Tables XI and XII. Table XI shows the comparison between FKM, FCM, FCM2, CF, BC2, EA, RP, our S1 (S1M) and S2 (S2M) measures for the metric R
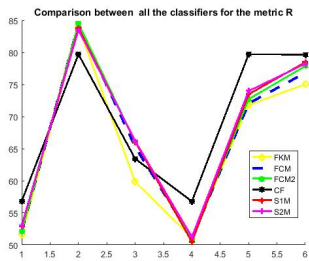
Fig. 5. Comparison between FKM, FCM, FCM2, CF, S1M and S2M for the metric R
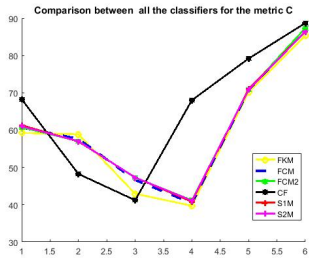


Fig. 6. Comparison between FKM, FCM, FCM2, CF, S1M and S2M for the metric C

while Table XII represents the comparison between the same clustering methods above-cited for the metric C. All results for BC2, EA and RP were taken from [70]. Results for Cluster Forest and those for Type-2 Fuzzy C-means were derived from [69] and [62], respectively. As described in Table XI and XII, the findings, obtained for S1 and S2 measures, show relevant and significant results reaching 83% and 86% for the metric R and C, respectively. For example, for Breast Cancer data and for the metric R, clustering results based on S1 and S2 measures reached higher classification accuracy which is in average 0.6% more accurate than that provided by FCM2, 1% than that given by FCM and 3% than that obtained by FKM, BC2, EA and RP. Our proposed IT-2 FSMs gave better clustering results than all the cited algorithms, except for the CF method. Similarly, for the Statlog (Heart) data and for the metric C, clustering results based on S1 and S2 measures achieved higher classification accuracy which is in average 2% more accurate than FKM; 0.5% than FCM and FCM2; 2% than BC2 and EA and 1% than RP, respectively. From Tables XI and XII, it is clear that S2 measure gave better results than S1 measure for the 2 metrics R and C in term of accuracy. We can then conclude that S1 and S2 measures can perform better in clustering tasks for 6 datasets that differ in instances, features and classes numbers. As depicted by Janousek *et al.* in [69], the CF algorithm outperforms BC2, EA and RP methods. In Figure 5 and 6, only FKM, FCM, CF, FCM2, S1M and S2M classifiers. All these classifiers present similar curves including our S1 and S2 measures which confirms the fact that the proposed IT-2 FSMs can really be used in clustering applications, decision making and knowledge reduction.

In practical application, since S1 measure satisfies the four similarity properties, it may be used in almost all applications especially decision making problems. While, S2 has shown

good results in applications dealing with linguistic terms or computing with words field. Very relevant results were obtained also in classification application or clustering data. The findings obtained by applying S3 similarity measure, have shown that S3 performance may change from one experiment to another. S3 was not good in linguistic data but was efficient in detecting non similar cases.

## VI. CONCLUSIONS AND FUTURE WORK

In this paper, three Interval Type-2 Fuzzy Similarity measures (S1, S2 and S3) were proposed as an extension of some Intuitionistic ones to be used between IT2-FSs in clustering applications. Mathematical Proofs were provided in order to demonstrate that the introduced IT-2 FSMs can satisfy similarity properties (Transivity, reflexivity, overlapping and symmetry). In fact, S1 measure satisfied all the four proximity properties while S2 and S3 did not fulfill the symmetry and the reflexivity property, respectively. These IT-2 FSMs were tested on real data, compared with some existing Type-2 similarity measures (Gorzalczany, Bustince, Mitchell, Zeng and Li, VSM and Jaccard IT-2FSMs) in both noisy and noise free environments showing relevant and reasonable results especially for S1 and S2 measures. The advanced IT-2FSMs achieved higher similarity performance, which is in average, 0.5% more efficient than the existing IT-2FSMs. In order to prove the effectiveness of these IT-2 FSMs, S1 and S2 measures were evaluated in clustering applications. We combined these measures with Fuzzy c-means algorithm by changing the Euclidean distance with the use of the advanced S1 and S2 measures. Then, our algorithm was compared with the existing clustering ones (Type-1 Fuzzy k-means, Type-1 and Type-2 Fuzzy c-means, Cluster Forest, Bagged Clustering, Evidence Accumulation and Random Projection). This comparison demonstrated that our algorithm, based on 2 clustering quality performers: R and C (considered as the standard classification accuracy), reached the highest classification accuracy (86%). The performance of clustering system relying on the advanced IT-2FSMs is almost the same as that of the other exiting classifiers. In our future work, these measures will be tested in the psychology of concepts and knowledge reduction domain by factorizing similar concepts, selecting concepts prototypes and generating the Interval Type-2 Fuzzy Formal Lattice as a knowledge representation form for better decision or subjective judgment.

## REFERENCES

[1] H. A. Hagras, "A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 4, pp. 524–539, 2004.

[2] T. Xie and Z. Gong, "A hesitant soft fuzzy rough set and its applications," *IEEE Access*, vol. 7, pp. 167766–167783, 2019.

[3] X.-W. Liu and S.-L. Han, "Ranking fuzzy numbers with preference weighting function expectations," *Computers & Mathematics with Applications*, vol. 49, no. 11, pp. 1731 – 1753, 2005.

[4] L. Baccour and A. M. Alimi, "Distance measures for intuitionistic fuzzy sets and interval valued intuitionistic fuzzy sets," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2019, pp. 1–6.

[5] G. Ruiz-Garcia, H. Hagras, H. Pomares, and I. R. Ruiz, "Toward a fuzzy logic system based on general forms of interval type-2 fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 12, pp. 2381–2395, 2019.

[6] K. Almohammadi, H. Hagras, D. Alghazzawi, and G. Aldabbagh, "A zslices-based general type-2 fuzzy logic system for users-centric adaptive learning in large-scale e-learning platforms," *Soft Computing*, vol. 21, no. 22, pp. 6859–6880, Nov 2017.

[7] J. M. Mendel, H. Hagras, H. Bustince, and F. Herrera, "Comments on "interval type-2 fuzzy sets are generalization of interval-valued fuzzy sets: Towards a wide view on their relationship"," *Trans. Fuz Sys.*, vol. 24, no. 1, pp. 249–250, Feb. 2016.

[8] D. Wu and J. M. Mendel, "Similarity measures for closed general type-2 fuzzy sets: Overview, comparisons, and a geometric approach," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 3, pp. 515–526, 2019.

[9] M. Zouari, N. Baklouti, J. Sanchez-Medina, H. M. Kammoun, M. B. Ayed, and A. M. Alimi, "Pso-based adaptive hierarchical interval type-2 fuzzy knowledge representation system (pso-ahit2fkrs) for travel route guidance," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2020.

[10] R. J. Leila Baccour, Adel M. Alimi, "Similarity measures for intuitionistic fuzzy sets: State of the art," *Journal of Intelligent & Fuzzy Systems 24 (2013) 3749 DOI:10.3233/IFS-2012-0527 IOS Press*, 2013.

[11] S. E. and K. J., "Analysis of similarity measures for atanassov's intuitionistic fuzzy sets," *In: Proc: IFSA-EUSFLAT, pp. 1416-1421*, 2009.

[12] Z. Pei and L. Zheng, "A novel approach to multi-attribute decision making based on intuitionistic fuzzy sets," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2560–2566, Feb. 2012.

[13] T. Chaira and A. K. Ray, "A new measure using intuitionistic fuzzy set theory and its application to edge detection," *Appl. Soft Comput.*, vol. 8, no. 2, pp. 919–927, Mar. 2008.

[14] L. Baccour, A. M. Alimi, and R. I. John, "Some notes on fuzzy similarity measures and application to classification of shapes, recognition of arabic sentences and mosaic," *IAENG International Journal of Computer Science*, 2014.

[15] S. Cherif, N. Baklouti, A. Alimi, and V. Snasel, "Fuzzy feature selection based on interval type-2 fuzzy sets," in *SPIE Proceedings Vol. 10341*, 03 2017, p. 103412M.

[16] S. Cherif, N. Baklouti, A. M. Alimi, and V. Snasel, "Linguistic representation by fuzzy formal concept and interval type-2 feature selection," in *Intelligent Systems Design and Applications*, A. M. Madureira, A. Abraham, D. Gamboa, and P. Novais, Eds. Cham: Springer International Publishing, 2017, pp. 1071–1081.

[17] R. Belohlavek, J. Dvorak, and J. Outrata, "Fast factorization by similarity in formal concept analysis of data with fuzzy attributes," *Journal of Computer and System Sciences*, vol. 73, no. 6, pp. 1012 – 1022, 2007.

[18] F. Alqadah and R. Bhatnagar, "Similarity measures in formal concept analysis," *Annals of Mathematics and Artificial Intelligence*, vol. 61, no. 3, pp. 245–256, Mar. 2011.

[19] Z. Su, Z. Xu, H. Zhao, Z. Hao, and B. Chen, "Entropy measures for probabilistic hesitant fuzzy information," *IEEE Access*, vol. 7, pp. 65 714–65 727, 2019.

[20] R. D. Garg, H., "A robust correlation coefficient measure of complex intuitionistic fuzzy sets and their applications in decision-making," *Appl Intell 49, 496512*, 2019.

[21] A. Mishra, R. Kumari, and D. Sharma, "Intuitionistic fuzzy divergence measure-based multi-criteria decision-making method," *Neural Computing and Applications*, vol. 31, 09 2017.

[22] G. Smits, O. Pivert, e. J. Duong, Toan Ngoc", M. Ojeda-Aciego, J. L. Verdegay, D. A. Pelta, I. P. Cabrera, B. Bouchon-Meunier, and R. R. Yager, "On dissimilarity measures at the fuzzy partition level," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations*. Cham: Springer International Publishing, pp. 301–312 , isbn=978–3–319–91 476–3, year=2018.

[23] D. Wu and J. M. Mendel, "A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets," *Inf. Sci.*, vol. 178, no. 2, pp. 381–402, Jan. 2008.

[24] W. Dongrui and J. M. Mendel, "A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets," *Inf. Sci.*, vol. 179, no. 8, pp. 1169–1192, Mar. 2009.

[25] J. C. Figueroa-García, Y. Chalco-Cano, and H. Román-Flores, "Distance measures for interval type-2 fuzzy numbers," *Discrete Appl. Math.*, vol. 197, no. C, pp. 93–102, Dec. 2015.

[26] A. Heidarzade and I. Mahdavi, "Multiple attribute group decision making method using a new similarity measure in interval type-2 fuzzy sets: a case study," *Int. J. Mathematics in Operational Research, Vol. 9, No. 2*, 2016.

[27] G. A. Papakostas, A. G. Hatzimichailidis, and V. G. Kaburlasos, "Distance and similarity measures between intuitionistic fuzzy sets: A comparative analysis from a pattern recognition point of view," *Pattern Recogn. Lett.*, vol. 34, no. 14, pp. 1609–1622, Oct. 2013.

[28] G. Hesamian, "Measuring similarity and ordering based on interval type-2 fuzzy numbers," *IEEE Transactions on Fuzzy Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[29] J. McCulloch, C. Wagner, and U. Aickelin, "Extending similarity measures of interval type-2 fuzzy sets to general type-2 fuzzy sets," *CoRR*, vol. abs/1308.5136, 2013.

[30] J. McCulloch and C. Wagner, "Measuring the similarity between zs-lices general type-2 fuzzy sets with non-normal secondary membership functions," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July 2016, pp. 461–468.

[31] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets Syst.*, vol. 20, no. 1, pp. 87–96, Aug. 1986.

[32] S.-M. Chen, "Similarity measures between vague sets and between elements," *Trans. Sys. Man Cyber. Part B*, vol. 27, no. 1, pp. 153–158, Feb. 1997.

[33] D. H. Hong and C. Kim, "A note on similarity measures between vague sets and between elements," *Inf. Sci.*, vol. 115, no. 1-4, pp. 83–96, Apr. 1999.

[34] Y. Li, D. Olson, and Z. Qin, "Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis," *Pattern Recognition Letters*, vol. 28, pp. 278–285, 01 2007.

[35] W.-L. Hung and M.-S. Yang, "On similarity measures between intuitionistic fuzzy sets," *International Journal of Intelligent Systems*, vol. 23, no. 3, pp. 364–383, 2008.

[36] G. N. Lance and W. T. Williams, "Computer programs for hierarchical polythetic classification ('similarity analyses')," *The Computer Journal*, vol. 9, no. 1, pp. 60–64, May 1966.

[37] J. T. Overpeck, T. Webb, and I. C. Prentice, "Quantitative interpretation of fossil pollen spectra: Dissimilarity coefficients and the method of modern analogs," *Quaternary Research*, vol. 23, no. 1, pp. 87–108, 1985.

[38] M. Gorzalczany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets," *Fuzzy Sets and Systems 21 (1987) 1-17.*, 1987.

[39] H.-J. Z. E. Tsiporkova, "Aggregation of compatibility and equality: a new class of similarity measures for fuzzy sets," in *the Seventh International Conference on Information Processing and Management of Uncertainty in KnowledgeBased Systems, Paris, pp. 1769-1776.*, 1998.

[40] H. Bustince, "Indicator of inclusion grade for interval-valued fuzzy sets. application to approximate reasoning based on interval-valued fuzzy sets," *International Journal of Approximate Reasoning*, vol. 23, no. 3, pp. 137 – 209, 2000.

[41] H. B. Mitchell, "Pattern recognition using type-ii fuzzy sets," *Inf. Sci. Inf. Comput. Sci.*, vol. 170, no. 2-4, pp. 409–418, Feb. 2005.

[42] W. Zeng and H. Li, "Relationship between similarity measure and entropy of interval valued fuzzy sets," *Fuzzy Sets Syst.*, vol. 157, no. 11, pp. 1477–1484, Jun. 2006.

[43] P. Singh, "Some new distance measures for type-2 fuzzy sets and distance measure based ranking for group decision making problems," *Front. Comput. Sci.*, vol. 8, no. 5, pp. 741–752, Oct. 2014.

[44] D. Wu and J. Mendel, "Perceptual reasoning for perceptual computing: A similarity-based approach," *Fuzzy Systems, IEEE Transactions on*, vol. 17, no. 6, pp. 1397–1411, Dec 2009.

[45] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bulletin de la Societe de Vaud des Sciences Naturelles 44 (1908) 223.*, 1908.

[46] S. Cherif, N. Baklouti, V. Snasel, and A. M. Alimi, "New fuzzy similarity measures: From intuitionistic to type-2 fuzzy sets," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July 2017, pp. 1–6.

[47] K. Kuratowski and A. Mostowski, *Set theory / [by] K. Kuratowski and A. Mostowski ; Translated from Polish by M. Maczynski.* North-Holland ; Polish Scientific Publishers Amsterdam : Warszawa, 1968.

[48] J. Barwise, *Handbook of Mathematical Logic.* North-Holland, 1977.

[49] C.-M. Own, "Switching between type-2 fuzzy sets and intuitionistic fuzzy sets: an application in medical diagnosis," *Applied Intelligence*, vol. 31, no. 3, p. 283, 2008.

[50] D. Nguyen, L. T. Ngo, and L. Pham, "Interval type-2 fuzzy c-means clustering using intuitionistic fuzzy sets," *2013 3rd World Congress on Information and Communication Technologies, WICT 2013*, pp. 299–304, 05 2015.

[51] S. Singh and H. Garg, "Distance measures between type-2 intuition-istic fuzzy sets and their application to multicriteria decision-making process," *Applied Intelligence*, vol. 46, no. 4, pp. 788–799, Jun. 2017.

[52] Z. X. Bin Zhu and M. Xia, "Dual hesitant fuzzy sets," *Journal of Applied Mathematics, vol. 2012, Article ID 879629, 13 pages, 2012. doi:10.1155/2012/879629*, 2012.

[53] S. Coupland, J. Mendel, and D. Wu, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and convergence of the word fous," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, July 2010, pp. 1–8.

[54] A. Bilgin, H. Hagras, D. Alghazzawi, A. Malibari, and M. J. Alhaddad, "Employing an enhanced interval approach to encode words into linear general type-2 fuzzy sets for computing with words applications," in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Aug 2015, pp. 1–8.

[55] D. Wu, J. M. Mendel, and S. Coupland, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis," *IEEE Trans. Fuzzy Systems*, vol. 20, pp. 499–513, 2012.

[56] F. Liu and J. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," *Fuzzy Systems, IEEE Transactions on*, vol. 16, no. 6, pp. 1503–1521, Dec 2008.

[57] S. Cherif, N. Baklouti, A. M. Alimi, and V. Snasel, *A Type-2 Fuzzy Concepts Lexicalized Representation by Perceptual Reasoning and Linguistic Weighted Average: A Comparative Study*. Cham: Springer International Publishing, 2016, pp. 77–86.

[58] J. Mendel, "Fuzzy sets for words: a new beginning," *Proc. IEEE Int. Conf. Fuzzy Systems, St. Louis, MO, pp. 3742.2003*, 2003.

[59] ——, *Uncertain Rule-based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall PTR, 2001.

[60] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[61] M. Elter, R. Schulz-Wendtland, and T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process," *Medical Physics*, vol. 34, p. 4164, 2007.

[62] A. Ben Ayed, M. Ben Halima, and A. Alimi, "Adaptive fuzzy exponent cluster ensemble system based feature selection and spectral clustering," pp. 1–6, 07 2017.

[63] D. Dheeru and E. Karra Taniskidou, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[64] S. Dudoit and J. Fridlyand, "Bagging to improve the accuracy of a clustering procedure," *Bioinformatics*, vol. 19, no. 9, pp. 1090–1099, 2003.

[65] A. L. N. Fred and A. K. Jain, "Data clustering using evidence accumu-lation," in *ICPR*, 2002.

[66] X. Z. Fern and C. E. Brodley, "Random projection for high dimensional data clustering: A cluster ensemble approach," in *ICML*, 2003.

[67] J. Irani, N. Pise, and M. Phatak, "Article: Clustering techniques and the similarity measures used in clustering: A survey," *International Journal of Computer Applications*, vol. 134, no. 7, pp. 9–14, January 2016, published by Foundation of Computer Science (FCS), NY, USA.

[68] P. Singh, "Similarity measure for type-2 fuzzy sets with an application to students' evaluation," *Computer Applications in Engineering Education*, vol. 23, no. 5, pp. 694–702, 2015.

[69] J. Janoušek, P. Gajdoš, M. Radecký, and V. Snášel, "Application of bio-inspired methods within cluster forest algorithm," in *Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015*, A. Abraham, K. Wegrzyn-Wolska, A. E. Hassanien, V. Snasel, and A. M. Alimi, Eds. Cham: Springer Interna-tional Publishing, 2016, pp. 237–247.

[70] M. I. J. Donghui Yan, Aiyou Chen, "Cluster forests," *arXiv:1104.2930*, 2011.