2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

# Enhanced Cluster Validity Index for the Evaluation of Optimal Number of Clusters for Fuzzy C-Means Algorithm

Neha Bharill
Department of Computer Science and Engineering
Indian Institute of Technology
Indore, India 453331
Email: phd12120103@iiti.ac.in

Aruna Tiwari
Department of Computer Science and Engineering
Indian Institute of Technology
Indore, India 453331
Email: artiwari@iiti.ac.in

***Abstract*—Cluster validity index is a measure to determine the optimal number of clusters denoted by $(\mathcal{C})$ and an optimal fuzzy partition for clustering algorithms. In this paper, we proposed a new cluster validity index to determine an optimal number of hyper-ellipsoid or hyper-spherical shape clusters generated by Fuzzy C-Means (FCM) algorithm called as $VI_{DSO}$ index. The proposed validity index jointly exploits all the three measures named as intra-cluster compactness, an inter-cluster separation and overlap between the clusters. The proposed intra-cluster compactness is based on relative variability concept which is a statistical measure of relative dispersion or scattering of data in various dimensions within the clusters. The proposed inter-cluster separation measure indicates the isolation or distance between the fuzzy clusters. The proposed inter-cluster overlap measure determines the degree of overlap between the fuzzy clusters. The best fuzzy partition produced by the $VI_{DSO}$ index is expected to have low degree of intra-cluster compactness, higher degree of inter-cluster separation and low degree of inter-cluster overlap. The efficacy of $VI_{DSO}$ index is evaluated on six benchmark data sets and compared with a number of known validity indices. The experimental results and the comparative study demonstrate that, the proposed index is highly effective and reliable in estimating the optimal value of $\mathcal{C}$ and an optimal fuzzy partition for each data set because, it is insensitive with change in values of fuzzification parameter denoted by *m*. In contrast, the other indices [2], [3], [6], [7] fails to achieve the optimal value of $\mathcal{C}$ due to it is susceptibility with change in *m*.**

## I. Introduction

In pattern recognition, one of the most widely used technique is Clustering [1]. It is an unsupervised learning approach in which collection of unlabeled samples are grouped into a meaningful clusters so that, the similarity between samples within the cluster is maximized whereas the similarity between the clusters is minimized. The major objective of clustering algorithm is to partition the data set into $\mathcal{C}$ homogeneous clusters [2] and derive insightful information based on the similarity exhibits within each cluster. The partitions generated by the clustering algorithm determines the belongingness of these samples to the clusters. The produced partitions may define the hard boundary for samples called as hard clustering. In hard clustering, each sample belongs to only one cluster with degree of membership equal to one or zero. In contrast, the clustering algorithm may produce the fuzzy partitions where the data points are given partial degree of memberships in multiple nearby clusters.

The data points belong to multiple clusters with a degree of membership between 0 and 1. One of the most widely used clustering algorithm based on the fuzzy sets is defined as Fuzzy C-Means (FCM) algorithm proposed by Bezdek [4], [11]. The FCM algorithm detects clusters having centroid prototypes of a roughly similar size i.e. distribution of data is in form of hyper-ellipsoid or hyper-spherical shape [14]. It is required to pre-specified number of clusters denoted by $\mathcal{C}$ for the computation of fuzzy partitions in FCM algorithm. However, the partitions produced by the clustering algorithm (hard or fuzzy) depend on the choice of $\mathcal{C}$. The problem of estimating the correct value of $\mathcal{C}$ and finding the best partition from the partitions produced by the clustering algorithm is done by calculating cluster validity index [1]. In this paper, we propose a modified cluster validity index designed for validating the fuzzy partitions produced by standard FCM algorithm. However, many clustering algorithms [4], [5] has been proposed by the researchers to produce the fuzzy partitions. These algorithms generates the fuzzy partitions, that are validated by the cluster validity index. As stated above, FCM algorithm works well for the data sets where the distribution of data is of hyper-ellipsoid or hyper-spherical shape. Therefore, many cluster validity indices [1], [2], [3] have been proposed by the researchers for validating the fuzzy partitions produced by FCM algorithm on different value of $\mathcal{C}$ where the generated clusters is of hyper-ellipsoid or hyper-spherical shape. All these indices are mainly based on the optimization of measures known as compactness and separation, to determine best fuzzy partition. Compactness measures the scattering or dispersion of samples within each cluster. The small value of compactness measure indicates that, the samples within each cluster are less scattered and tightly bounded with each other. In contrast, separation determines the segregation between the clusters from one another [2].

The first validity index proposed by Bezdek known as partition coefficient $(V_{PC})$ [7] and partition entropy $(V_{PE})$ [6] based on optimization of compactness measurement. The $(V_{XB})$ index proposed by Xie and Beni [8] is based on the optimization of both the compactness and separation measure to find the optimal fuzzy partitions. Another index proposed by Fukuyama and Sugeno known as $V_{FS}$ index [9] also determines the best fuzzy partition by optimizing compactness and separation measure. Similar to the $V_{FS}$ index, the

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

Rezaee proposed the $V_{CWB}$ index [3] which measures the compactness by computing the variance of samples within the cluster with respect to average scattering and then, it combined the compactness measure with the $Dist(c)$ which measures the separation between the clusters. The value of $\mathcal{C}$ which minimize the $V_{CWB}$ index, corresponds to the optimal value of $\mathcal{C}$. The above stated indices based on optimization of intra-cluster compactness and inter-cluster separation but, it fails to emphasize on the inter-cluster overlap measure which has significant impact over fuzzy partitions. The next index proposed by the Dae-Won Kim called as $v_{os}$ index [2], determines the optimal value of $\mathcal{C}$ by measuring the inter-cluster separation and inter-cluster overlap. The optimal value of $v_{os}$ index maximizes the separation and minimizes the overlapping to determine the best fuzzy partition. However, the existing validity indices, determines the true number of clusters by jointly exploiting the compactness and separation measure or, in other way, we can say that overlap and separation measures are considered jointly. But, all these indices fail to exploit all three measures jointly i.e. the intra-cluster compactness within the cluster, inter-cluster separation and overlap measure between the clusters.

In this paper, we proposed a new cluster validity index ($VI_{DSO}$) to find the optimal clusters form the number of clusters generated by FCM algorithm where the generated clusters is of hyper-ellipsoid or hyper-spherical shape [14]. It also validate the fuzzy partitions produced by the FCM algorithm by jointly exploits all the three measures i.e., intra-cluster cohesion or compactness based on relative variability, the inter-cluster separation and overlap to determine the optimal number of clusters ($\mathcal{C}$) and an optimal fuzzy partition. The proposed intra-cluster compactness or cohesion, measure the relative scattering or dispersion of data points in all the dimensions. Thus, the dimension in which data points have maximum dispersion is minimized to increase the overall compactness within the clusters. The inter-cluster separation computes the distance between fuzzy clusters, where the larger distance indicates greater separation. Therefore, the distance between the clusters that are less separated from each other is maximized to increase the overall separation among all the clusters. The inter-cluster overlap measure indicates the degree of overlap between the fuzzy clusters. This degree of overlap among the highly overlapped clusters is need to be minimized, to reduce the overall overlapping among the fuzzy partitions produced by FCM algorithm. Hence, the best fuzzy partition is expected to have smaller value of cohesion or compactness measure, low degree of overlap measure and larger distance or separation measure.

The remainder of this paper is organized as follows: Section II, is presented with the brief overview of the FCM algorithm; The formulation of the proposed cluster validity index for FCM algorithm is described in Section III; Section IV, illustrate the experimental results and comparison of proposed validity index with various popular validity index on variety of benchmark data sets. Finally, Section V, is presented with the concluding remarks.

## II. Fuzzy C-Means Clustering

Clustering is a mechanism of grouping the collection of unlabeled data points into $\mathcal{C}$ homogeneous clusters such that, the data points within the cluster are similar to each other. One of the most widely used unsupervised clustering algorithm that produces the fuzzy partitions called as FCM algorithm [10]. Let a set $X$=$[x_1, x_2, ..., x_n]$ denotes the $n$ data points in $d$-dimensional Euclidean space $\mathbf{R^d}$. The FCM algorithm generates the fuzzy clusters denoted as $\overline{FP}$=$\{\overline{FP}_1, ..., \overline{FP}_c\}$. It allows the data points to belong to multiple clusters with varying degree of membership denoted as $\mu_{\overline{FP}_i}(x_j)\in[0,1]$, such that $\sum_{i=1}^{c}\mu_{\overline{FP}_i}(x_j)$=1; $\forall x_j \in X$. Therefore, $U_c = \{\mu_{11}, ..., \mu_{1c}\}$ represents the set of fuzzy membership degree of all the data points present in set $\boldsymbol{X}$ corresponding to $\mathcal{C}$ clusters.

The FCM algorithm iteratively minimizes the objective function. The formulation of objective function is defined as follows:

$$J_m(U,V,X) = \sum_{j=1}^{n}\sum_{i=1}^{c}(\mu_{\overline{FP}_i}(x_j))^m \|x_j - v_i\|^2, 1 < m < \infty \tag{1}$$

where, $V$=$(v_1, ..., v_c)$ denotes the set of cluster prototype of fuzzy clusters $\overline{FP}$=$\{\overline{FP}_1, ..., \overline{FP}_c\}$; $V_i \in R^d$

$m$, the weighting exponent also called as fuzzification parameter. It controls the degree of fuzziness between clusters and also has significant impact over the performance of FCM clustering algorithm. The steps of FCM algorithm are presented in Algorithm 1, subsequently.

**Algorithm 1** Fuzzy C-Means Clustering FCM(U,V)

**Input:** $X = \{x_1, x_2, ..., x_n\}; \epsilon = 0.001;$
$U = \{\mu_{\overline{FP}_1}(x_j), .., \mu_{\overline{FP}_c}(x_j)\}; 1 < m < \infty$
**Output:** $U, V$

1: Given a pre-decided number of clusters $\mathcal{C}$ where $c_{\min} \leq \mathcal{C} \leq c_{\max}$; $c_{\min}$=2, $c_{\max}$=$\sqrt{N}$; $N$ represents the number of training samples, initialize the fuzzy partition matrix $U$ corresponding to $\forall x_j \in \overline{FP}$ where $\overline{FP}$=$\{\overline{FP}_1, ..., \overline{FP}_c\}$, such that

$$\sum_{i=1}^{c}\mu_{\overline{FP}_i}(x_j) = 1 \tag{2}$$

2: **while** $\| U^{l+1} - U^l \| > \epsilon$ **do**
3: Compute the fuzzy cluster centers for all $i = 1, ..., c$

$$v_i = \frac{\sum_{i=1}^{c}[\mu_{\overline{FP}_i}(x_j)]^m x_j}{\sum_{i=1}^{c}[\mu_{\overline{FP}_i}(x_j)]^m}, \forall i \tag{3}$$

4: Update the fuzzy cluster membership

$$\mu_{\overline{FP}_i}(x_j) = \frac{\| x_j - v_i \|^{\frac{-2}{m-1}}}{\sum_{k=1}^{c}\| x_j - v_k \|^{\frac{-2}{m-1}}}, \forall i, j \tag{4}$$

5: Check fuzzy membership matrix obtained in Eq (4) such that summation of degree of belongingness of each data point $x_j$ to all the clusters should be 1 i.e.

$$\sum_{i=1}^{c}\mu_{\overline{FP}_i}(x_j) = 1 \tag{5}$$

6: **end while**

The FCM algorithm iteratively minimizes the objective function by randomly initializing the membership matrix $U_c$.

It improves the set of cluster prototype $V$=$(v_1, ..., v_c)$ and the fuzzy membership set $U_c$ in subsequent iterations. The algorithm terminates when the change in the membership values between two successive iterations is less than the predefined threshold value $\epsilon$.

The FCM algorithm produces the fuzzy partitions on the pre-specified value of $\mathcal{C}$. It is not always possible to predict the correct value of $\mathcal{C}$ in advance and also it is not sure that the pre-specified value of $\mathcal{C}$ will always produced an optimal fuzzy partition. Different specifications of $\mathcal{C}$ values will produce different fuzzy partitions. Therefore, the fuzzy partitions produced by FCM algorithm on each value of $\mathcal{C}$ require a validation methodology. The validation methodology is used to find the correct value of $\mathcal{C}$ is referred as cluster validity index [1]. The cluster validity index [2] is a mathematical formula evaluated for each fuzzy partition generated by FCM algorithm on pre-assumed value of $\mathcal{C}$ where $\mathcal{C} \in [c_{\min}, .., c_{\max}]$. The value of $\mathcal{C}$ on which the validity indices achieves its optimal value will indicate the true number of clusters ($\mathcal{C}$). Thus, the cluster validity index searches the true value of $\mathcal{C}$ which leads to an optimal fuzzy partition.

## III. The Proposed Validity Index For FCM

In this paper, we proposed a new validity index named as $VI_{DSO}$. The $VI_{DSO}$ index jointly exploits intra-cluster compactness, inter-cluster separation and inter-cluster overlap measure to evaluate the quality of fuzzy partitions produced by FCM algorithm. The intra-cluster compactness, indicates the density of the data points present within the cluster. The small value of this term, indicates that data points are more tightly coupled within the cluster thus, it results in higher compactness. The proposed definitions and the formulation for the same is briefly discussed in Section-A. Another important measure used to validate the fuzzy partitions is defined as inter-cluster separation, it indicates that how far apart the clusters are located from each other. Higher value of this term indicates the larger separation between the clusters. Related basics and the proposed definitions involve in the computation of inter-cluster separation is briefly discussed in Section-B. Next, one more important factor used to evaluate the quality of fuzzy partitions is the inter-cluster overlap, it indicates the degree of overlap of data points between fuzzy clusters. The small value of this term, indicates that the data points are more clearly classified to one cluster. The proposed definitions for the computation of inter-cluster overlap is presented in Section-C. Thus, value of $\mathcal{C}$ which minimizes the $VI_{DSO}$ index is consider as optimal number of clusters (or an optimal fuzzy partition). Hence, optimal value of $\mathcal{C}$ is expected to minimize the intra-cluster compactness within the cluster, maximize the inter-cluster separation between the clusters and minimize the inter-cluster overlap between the clusters.

### A. Proposed Intra-cluster Compactness

As discussed earlier, the existing validity index [3] measure the intra-cluster compactness based on the average variation of data points within the cluster. Here, we proposed the new intra-cluster compactness measure based on relative variability concept known as coefficient of variation. It measures the relative dispersion of data points in all the dimensions. The dimension in which data points have maximum dispersion is minimized by the proposed compactness measure denoted as $Disp(\mathcal{C}, U)$ which in turn indicate that the overall dispersion of data points in various dimensions within the cluster is minimized. The proposed definitions for the computation of intra-cluster compactness ($Disp(\mathcal{C}, U)$) is presented as follows:

**Definition 1**: Standard deviation of data points present in set $X$ in $n^{th}$ dimension is denoted by $\sigma(X)^n$ and is defined as:

$$\sigma(X)^n = \sqrt{(\sum_{i=1}^{p}(x_i^n)^2 - (\mu(X)^n \times p)^2/p)/p} \tag{6}$$

where, $X^n = \{x_1^n, .., x_i^n, .., x_p^n\}$; $\forall x_i \in R^n$ such that $\sigma(X) = [\sigma(X)^1, ..., \sigma(X)^n]$; $X \in R^n$ denote the data points present in set $X$ in $n$ dimensions; $\mu(X)^n = \sum_{i=1}^{p} x_i^n/n$ indicates the mean of data points in $n^{th}$ dimensions; $\forall x_i \in X$.

**Definition 2**: Coefficient of Variation of all the data points present in set $X$ in $n^{th}$ dimensions is denoted by $Coff\_var(X)^n$ and is defined as:

$$Coff\_var(X)^n = \frac{\sigma(X)^n}{\mu(X)^n} \tag{7}$$

where, $Coff\_var(X)^n \in Coff\_var(X)$ such that $Coff\_var(X) = [Coff\_var(X)^1, ..., Coff\_var(X)^n]$

**Definition 3**: Standard Deviation of $c^{th}$ cluster in $n^{th}$ dimensions is denoted as $\sigma_{v_c}^n$ and given by:

$$\sigma_{v_c}^n = \sqrt{\frac{1}{p}[\sum_{i=1}^{p}(x_i^n - v_c^n)^2]} \tag{8}$$

where $\sigma_{v_c} = [\sigma_{v_c}^1, ..., \sigma_{v_c}^n]$ ; $v_c \in R^n$ with $p$ data points such that $X = [x_1, .., x_i, .., x_p]; \forall x_i \in R^n$

**Definition 4**: Coefficient of Variation of $c^{th}$ clusters in $n^{th}$ dimension is denoted as $Coff\_var_{v_c}^n$ and is defined as:

$$Coff\_var_{v_c}^n = \frac{\sigma_{v_c}^n}{v_c^n} \tag{9}$$

where, $Coff\_var_{v_c}^n \in Coff\_var_{v_c}$ such that $Coff\_var_{v_c}$ $= [Coff\_var_{v_c}^1, ..., Coff\_var_{v_c}^n]$

**Definition 5**: The overall Dispersion within $\mathcal{C}$ number of clusters is defined as:

$$Disp(\mathcal{C}, U) = \frac{\max\limits_{1\le i\le c}\max\limits_{1\le j\le n}[Coff\_var_{v_i}^j]}{\max\limits_{1\le j\le n}[Coff\_var(X)^j]} \tag{10}$$

In Eq (6), we proposed a formulation to compute the dispersion of all the data points in each dimensions. The small value of this terms indicates that the data points tends to be close to each other in respective dimensions. Next, we proposed the formulation in Eq (7) to compute the relative dispersion of each data point in $n$ dimensions. In Eq (8), we proposed the formulation to compute the dispersion of $\mathcal{C}$ number of clusters in each dimension. The small value of

this term, indicates that the dispersion within $\mathcal{C}$ number of clusters in each dimension is minimum. Next in Eq (9), we proposed a formulation to compute the relative dispersion of each cluster in $n$ dimensions. Small value of this term indicates the dimension in which each cluster has less dispersion with respect to other dimensions.

Finally, the proposed definition in Eq (10), is a ratio of $\max_{1\leq i\leq c}\max_{1\leq j\leq n}[Coff\_var_{v_i}^{j}]$ and $\max_{1\leq j\leq n}[Coff\_var(X)^{j}]$. The numerator compute the dispersion for all the clusters in each dimension and finally consider the cluster which has maximum variation in particular dimension. The denominator consider the dimension in which all the data points have maximum dispersion with respect to other dimensions. Therefore, evaluation of mathematical expression $Disp(\mathcal{C},U)$ considers the dimension in which the data points and clusters have maximum dispersion relatively with other dimensions. Thus, it reflect the overall dispersion corresponding to all the data points and $\mathcal{C}$ clusters. Hence, small value of $Disp(\mathcal{C},U)$ indicates the higher compactness within the cluster.

### *B. Proposed Inter-cluster Separation based on Fuzzy Set*

Inter-cluster separation is also an important measure for estimating the quality of fuzzy partitions produced by FCM algorithm. The proposed separation measure evaluates the distance between the clusters by using the distance measure in the fuzzy sets. Therefore, we utilize the similarity measure suggested by Lee et al. [10]. The similarity between two fuzzy cluster $F_l$ and $F_r$ at data point $x_j$ is defined as follows:

$$S(F_l,F_r)=\max_{1\leq j\leq p}\min(\mu_{F_l}(x_j),\mu_{F_r}(x_j));\forall x_j\in X \quad (11)$$

The proposed separation measure represented as $Sep(\mathcal{C},U)$ considers the clusters with maximum similarity which conversely, results in consideration of the clusters with minimum separation. Therefore, the distance between the clusters with minimum separation is maximized. Thus, it indicates that the overall separation between the $\mathcal{C}$ clusters is maximized. Hence, the value of $\mathcal{C}$ for which the distance between the minimum separated cluster increases, indicates the overall well separated fuzzy partitions. Various definitions of proposed inter-cluster separation measure is presented as follows:

**Definition 1**: The separation between the two fuzzy clusters $F_l$ and $F_r$ is defined as follows:

$$Dist(F_l,F_r)=1-S(F_l,F_r) \quad (12)$$

**Definition 2**: The overall separation among $(\mathcal{C})$ number of clusters is defined as:

$$Sep(\mathcal{C},U)=min(Dist(F_l,F_r)) \quad (13)$$

The proposed definition in Eq (12), compute the separation among all the pairs of fuzzy clusters $F_l$ and $F_r$. The proposed definition for $Sep(\mathcal{C},U)$ in Eq (13), consider a pair of fuzzy clusters with minimum separation. Therefore, large value of $Sep(\mathcal{C},U)$ indicates that, the clusters which are separated by minimum distance are far apart from each other which in turns indicate the larger separation between other pairs of fuzzy clusters. Thus, it results in generation of well separated fuzzy partitions.

### *C. Proposed Inter-cluster Overlap Measure*

In fuzzy clustering, overlapping is an important factor need to be quantified by computing an inter-cluster overlap between fuzzy clusters. The proposed inter-cluster overlap measure compute overlap of each data point $x_j$ between two fuzzy clusters is represented by $R(x_j,c_p,c_q)$ in Eq (14). Each data point $x_j$ is assigned a degree of overlap depending on the belongingness of that data point with respect to the clusters and is denoted by $\delta(x_j)$. The vague data is assigned a higher degree of overlap than a clearly classified data point. Thus, total overlap between two fuzzy clusters is defined as $O(F_l,F_r)$ in Eq (17). It is obtained by summing the overlap of all the data points present in these clusters. Finally, the overlap between all pairs of fuzzy clusters is defined as $Overlap(\mathcal{C},U)$ in Eq (18) which is computed by considering the pair of fuzzy clusters having maximum overlap.

The fuzzy partitions and value of $\mathcal{C}$ on which $Overlap(\mathcal{C},U)$ achieve its minimum value will indicate that, the overlap between the highly overlapped pair of clusters is minimum. Thus, it reflects that, the other pairs of fuzzy clusters will have lesser overlap. Therefore, it achieves the best fuzzy partitions and results in well classified data points within the clusters. Various formal definitions for computing inter-cluster overlap is defined as follows:

**Definition 1**: The overlap of each data point $x_j$ between two fuzzy clusters $F_l$ and $F_r$ is defined as follows:

$$R(x_j,F_l,F_r)=\begin{cases}\delta(x_j), & if\Big(Dom_{\min}(x_j)>0\ \&\\ & \quad Dom_{\max}(x_j)<1\Big)\\ 0.0, & Otherwise\end{cases} \quad (14)$$

Where,

$$Dom_{\min}(x_j)=\min(\mu_{F_l}(x_j),\mu_{F_r}(x_j)) \quad (15)$$

$$Dom_{\max}(x_j)=\max(\mu_{F_l}(x_j),\mu_{F_r}(x_j)) \quad (16)$$

If data point $x_j$ is highly vague i.e. maximum degree of membership $Dom_{\max}(x_j)\leq 0.5$, then degree of overlap $\delta(x_j)$=1.0. Conversely, if the data point $x_j$ is not vague i.e. maximum degree of membership $Dom_{\max}(x_j)>0.5$ & $Dom_{\max}(x_j)<1$, then degree of overlap $\delta(x_j)=[0.9,0.1]$. Otherwise, if data point $x_j$ is clearly classified to particular cluster i.e. maximum degree of membership $Dom_{\max}(x_j)$=1, then degree of overlap $\delta(x_j)$=0.0.

**Definition 2**: The total overlap between two pairs of fuzzy clusters $F_l$ and $F_r$ is defined as follows:

$$O(F_l,F_r)=\sum_{j=1}^{n}R(x_j,F_l,F_r) \quad (17)$$

**Definition 3**: The total overlap between all pairs of fuzzy clusters is denoted by $Overlap(\mathcal{C},U)$ and is defined as follows:

$$Overlap(\mathcal{C},U)=\max_{l\neq r}(O(F_l,F_r)) \quad (18)$$

### D. Formulation of Proposed Validity Index

The proposed three measures i.e., $Disp(\mathcal{C}, U)$ in Eq (10), $Sep(\mathcal{C}, U)$ in Eq (13) and $Overlap(\mathcal{C}, U)$ in Eq (18), are jointly utilized to propose a new cluster validity index, $VI_{DSO}$. These three measures are of varying scales therefore, it needs to conciliate through normalization approach. Thus, all the three measures over varying $\mathcal{C}$ is defined as follows:

where, $\mathcal{C}$=$[c_{\min}, ..., c_{\max}]$; $c_{\min}$=2, $c_{\max}$=$\sqrt{N}$; $N$ denotes the number of samples.

$$Disp(\mathcal{C}, U) = [Disp(2, U), ..., Disp(c_{\max}, U)] \tag{19}$$

$$Sep(\mathcal{C}, U) = [Sep(2, U), ..., Sep(c_{\max}, U)] \tag{20}$$

$$Overlap(\mathcal{C}, U) = [Overalp(2, U), ..., Overlap(c_{\max}, U)] \tag{21}$$

The maximum value corresponding to each measure is computed as:

$$Disp_{\max} = \max_{c_{\min} \leq \mathcal{C} \leq c_{\max}} [Disp(\mathcal{C}, U)] \tag{22}$$

$$Sep_{\max} = \max_{c_{\min} \leq \mathcal{C} \leq c_{\max}} [Sep(\mathcal{C}, U)] \tag{23}$$

$$Overlap_{\max} = \max_{c_{\min} \leq \mathcal{C} \leq c_{\max}} [Overlap(\mathcal{C}, U)] \tag{24}$$

We normalize $Disp(\mathcal{C}, U)$, $Sep(\mathcal{C}, U)$ and $Overlap(\mathcal{C}, U)$ for each value of $\mathcal{C}$ with respect to their maximum values $Disp_{\max}$, $Sep_{\max}$ and $Overlap_{\max}$. Thus, normalized value of these measures are represented as:

$$Disp^N(\mathcal{C}, U) = \frac{Disp(\mathcal{C}, U)}{Disp_{\max}} \tag{25}$$

$$Sep^N(\mathcal{C}, U) = \frac{Sep(\mathcal{C}, U)}{Sep_{\max}} \tag{26}$$

$$Overlap^N(\mathcal{C}, U) = \frac{Overlap(\mathcal{C}, U)}{Overlap_{\max}} \tag{27}$$

Where, $Disp^N(\mathcal{C}, U)$ represents the normalized value of dispersion of data points within the cluster, $Sep^N(\mathcal{C}, U)$ represents the normalized value of separation of data points between fuzzy clusters and $Overlap^N(\mathcal{C}, U)$ represents the normalized value of overlap of data points between fuzzy clusters for a fuzzy partition with a particular $U$ and $C$; $U$ denotes partition matrix; $\mathcal{C}$ denotes the number of clusters.

The formulation of proposed validity index $VI_{DSO}$ is defined as:

$$VI_{DSO}(\mathcal{C}, U) = \frac{Disp^N(\mathcal{C}, U) + Overlap^N(\mathcal{C}, U)}{Sep^N(\mathcal{C}, U)} \tag{28}$$

The $VI_{DSO}$ index is calculated over varying values of $\mathcal{C} = [c_{\min}, ..., c_{\max}]$. The value of $\mathcal{C}$ and fuzzy partitions are optimal on which the value of $VI_{DSO}(C, U)$ is minimum. Thus, minimum value of $VI_{DSO}(\mathcal{C}, U)$ indicates that, the data points present within the clusters are more compact, clusters are overlapped with a lesser degree and well separated from each other. In Algorithm 2, we discuss the steps involved in the validation of fuzzy partitions produced by FCM algorithm which is validated using $VI_{DSO}$ index.

**Algorithm 2** Evaluation of Proposed $VI_{DSO}(\mathcal{C}, U)$ index

**Input:** $X = \{x_1, x_2, ..., x_n\}; U_c = \{\mu_{11}, .., \mu_{1c}\}; c_{\min} = 2;$ $\mathcal{C} = [c_{\min}, ..., c_{\max}]; m = [1.5, 2.5]; \epsilon = 0.001;$ $c_{\max} = \sqrt{N}$; N denotes the number of samples

**Output:** $VI^{\min}_{DSO}(c, U)$

1: Initialize $\mathcal{C} = c_{\min}$; $c_{\min} = 2$.
2: **if** $\mathcal{C} \leq c_{\max}$ **then**
3: Iteratively call $FCM(U, V)$ in Algorithm 1 for specified value of $\mathcal{C}$.
4: Compute and store the value of compactness measure $Disp(\mathcal{C}, U)$ in Eq (10), Separation measure $Sep(\mathcal{C}, U)$ in Eq (13) and Overlap measure $Overlap(\mathcal{C}, U)$ in Eq (18) for a fuzzy partitions obtained by $FCM(U, V)$ in Algorithm 1 for the specified value of $\mathcal{C}$.
5: $\mathcal{C} \longleftarrow \mathcal{C} + 1$
6: goto step 2
7: **else**
8: goto step 10
9: **end if**
10: Compute the normalized compactness measure $Disp^N(\mathcal{C}, U)$ in Eq (25), normalized separation measure $Sep^N(\mathcal{C}, U)$ in Eq (26) and normalized overlap measure $Overlap^N(\mathcal{C}, U)$ in Eq (27) for all values of $\mathcal{C}$ where, $\mathcal{C} = [c_{\min}, ..., c_{\max}]$.
11: Compute the proposed validity index $VI_{DSO}(\mathcal{C}, U)$ in Eq (28) for all values of $\mathcal{C}$ where, $\mathcal{C} = [c_{\min}, ..., c_{\max}]$.
12: Find the correct value of $\mathcal{C}$ or an optimal fuzzy partition and store the value of $\mathcal{C}$ that minimizes $VI_{DSO}$ represented as:

$$VI^{\min}_{DSO}(\mathcal{C}, U) = \min_{c_{\min} \leq \mathcal{C} \leq c_{\max}} [VI_{DSO}(\mathcal{C}, U)] \tag{29}$$

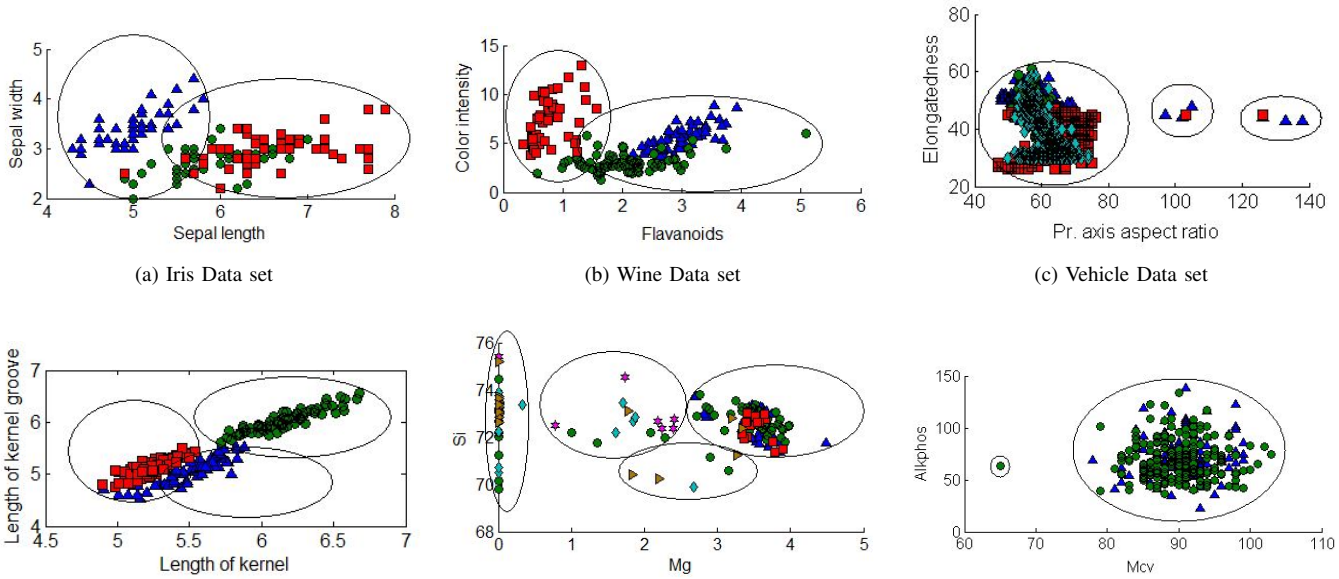
## IV. Experimental Results

In this section, experimentation is carried out to demonstrate the effectiveness of the proposed validity index $VI_{DSO}$ which measures intra-cluster compactness or dispersion based on relative variability, inter-cluster separation based on similarity measure and inter-cluster overlap. The effectiveness of $VI_{DSO}$ index in comparison with other four indices is tested over six benchmark data sets [12]. TABLE I, illustrates the description of these data sets.

For each data set, we investigate the performance of validity indices as discussed in Section I and Section III-D on standard FCM algorithm for each value of $\mathcal{C} = [c_{\min}, ..., c_{\max}]$.

TABLE I. Basic information of data sets

| Data sets | Samples | Features | Classes |
|---|---|---|---|
| Iris | 150 | 4 | 3 |
| Wine | 178 | 13 | 3 |
| Vehicle | 946 | 18 | 4 |
| Seeds | 210 | 7 | 3 |
| Glass | 214 | 10 | 6 |
| Bupa | 345 | 7 | 2 |




(a) Iris Data set (b) Wine Data set (c) Vehicle Data set

(d) Seeds Data set (e) Glass Data set (f) Bupa Data set

Fig. 1. Scatter plot in two dimensional space indicating the optimal number of clusters ($\mathcal{C}$) with circle for (a) Iris (b) Wine (c) Vehicle (d) Seeds (e) Glass (f) Bupa datasets according to data distribution.

TABLE II. PARAMETER SPECIFICATION

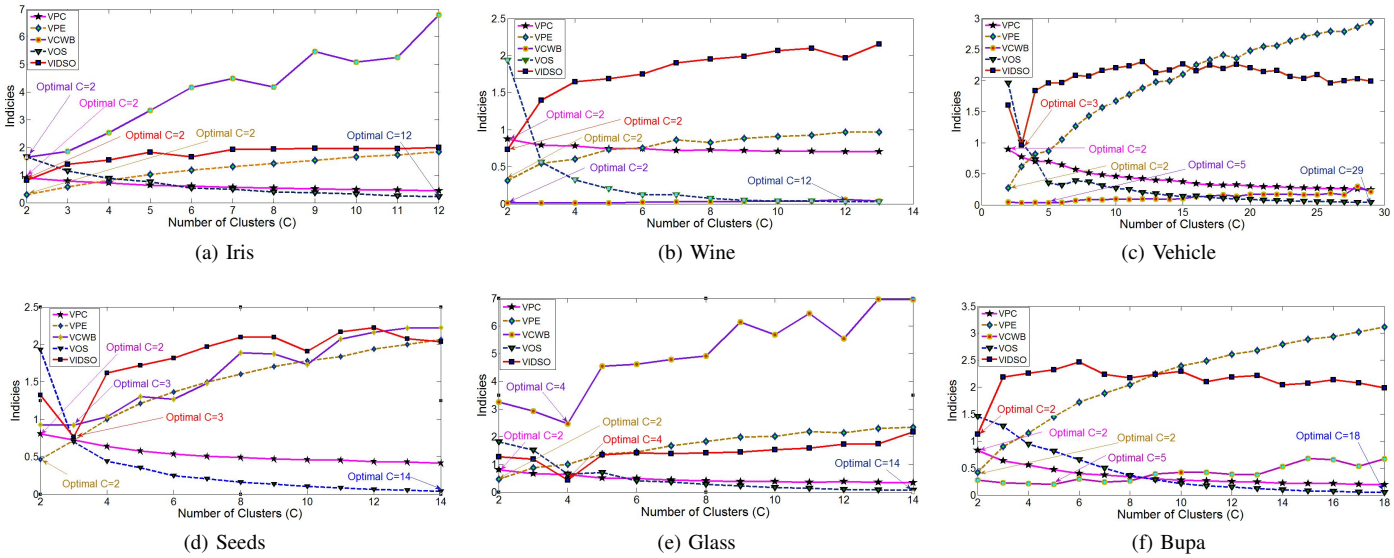
| Parameters | Description | Values |
|---|---|---|
| $\epsilon$ | Termination criteria | 0.001 |
| $m$ | Weighting exponent | $1 < m < \infty$ |
| $c_{\min}$ | Minimum number of cluster ($\mathcal{C}$) | 2 |
| $c_{\max}$ | Maximum number of cluster ($\mathcal{C}$) | $\sqrt{N}$ |
| $N$ | Number of input samples | Size of data set |

TABLE II, represents the specification of parameters required in computation of all the indices. The scatter plot of Iris, Wine, Vehicle, Seeds, Glass, Bupa data sets in two dimensional space is presented in Fig. 1(a) – 1(f) respectively. Fig. 2(a) – 2(f) display the results corresponding to optimal number of clusters determined by each index for $m = 2$ on six data sets.

Fig. 1(a), represents the plot of Iris data set contain 150 samples distributed in two dimensions i.e. sepal width and sepal length. The distribution of data indicates that out of three classes, two classes have substantial overlap while the third class is well separated from the other two. One can argue in favor of choosing $\mathcal{C}$=2 or 3 but rather than considering three separate clusters for three classes, if overlapped classes are grouped in one single cluster then, it will reduce the number of overlapped clusters. Therefore, $\mathcal{C}$=2 is considered as optimal number of clusters according to the geometric structure of data as mentioned by Pal and Bezdek [13]. Fig. 2(a), presents the optimal value of $\mathcal{C}$ determined by various validity indices over varying $\mathcal{C}$=$[c_{\min}, ..., c_{\max}]$; $c_{\min} = 2$; $c_{\max} = \sqrt{N} \approx 12$. The proposed validity index $VI_{DSO}$ and the existing indices $V_{PC}$ [7], $V_{PE}$ [6], $V_{CWB}$ [3] achieve optimal value of $\mathcal{C}$ at 2. Thus, it can be verified from above discussion that, $\mathcal{C}$=2 is the correct value of $\mathcal{C}$ [13] and it also represent the true number of cluster according to the distribution of data. In contrast, $v_{os}$ [2] index indicates $\mathcal{C}$=12 as optimal number of clusters, thus it fails to determine the correct value of $\mathcal{C}$.

Fig. 1(b), represents the plot of Wine data set contain 178 samples distributed in two dimensions i.e. color intensity and Flavonoids. The distribution of data shows that two classes have substantial overlap while the third class is well separated from the other two. So, $\mathcal{C}$=2 or 3 can be chosen as optimal clusters but rather than considering three separate clusters for three classes, if overlapped classes are grouped in one single cluster then, it will reduce the number of overlapped clusters. Therefore, $\mathcal{C}$=2 can be chosen as the optimal number of clusters according to the geometric structure of data as mentioned by Pal and Bezdek [13]. Fig. 2(b), highlight the results computed where, the optimal value of $\mathcal{C}$ determined by various validity indices over varying $\mathcal{C}$=$[c_{\min}, ..., c_{\max}]$; $c_{\min} = 2$; $c_{\max} = \sqrt{N} \approx 13$. The proposed validity index $VI_{DSO}$ and the existing indices $V_{PC}$ [7], $V_{PE}$ [6], $V_{CWB}$ [3] achieves its minimum value at $\mathcal{C}$=2; thus, it indicate $\mathcal{C}$=2 as the optimal number of clusters which can also be inferred from the above discussion and distribution of data that, true value of $\mathcal{C}$ is 2. Thus, $V_{PC}$, $V_{PE}$, $V_{CWB}$ and the proposed $VI_{DSO}$ index correctly recognize the presence of two clusters. In contrast, $v_{os}$ [2] index indicates $\mathcal{C}$=12 as optimal number of clusters. This, index is sensitive towards the large value of $\mathcal{C}$ therefore, it always achieves its optimal value at maximum value of $\mathcal{C}$. Thus, it fails to correctly determine the correct value of $\mathcal{C}$.

Fig. 1(c), represents the plot of Vehicle data set contain 946 samples distributed in two dimensions i.e. Elongatedness and PR. axis aspect ratio. Thus, spread of data in two dimensions, indicates the presence of 3 as true number of clusters ($\mathcal{C}$). Fig. 2(c), demonstrated the results computed where, the optimal value of $\mathcal{C}$ determined by various validity indices over varying $\mathcal{C}$=$[c_{\min}, ..., c_{\max}]$; $c_{\min} = 2$; $c_{\max} = \sqrt{N} \approx 29$. The proposed validity index $VI_{DSO}$ achieves its minimum value at $\mathcal{C}$=3; thus, it indicates $\mathcal{C}$=3 as optimal number of clusters which can also be verified with distribution of data shown in Fig. 1(c). In contrast, $V_{PC}$ [7] and $V_{PE}$ [6] index indicate $\mathcal{C}$=2, $V_{CWB}$




(a) Iris (b) Wine (c) Vehicle

(d) Seeds (e) Glass (f) Bupa

Fig. 2. Comparison between various validity indices $V_{PC}$, $V_{PE}$, $V_{CWB}$, $v_{os}$, $VI_{DSO}$ indicating the optimal number of clusters $\mathcal{C}$ at $m = 2$ for Iris, Wine, Vehicle, Seeds, Glass and Bupa data sets

[3] index indicates $\mathcal{C}$=5 and $v_{os}$ [2] index indicates $\mathcal{C}$=29 as optimal number of clusters. Thus, $VI_{DSO}$ index is only able to correctly recognize the presence of three clusters expect this all other indices fails to determine the correct value of $\mathcal{C}$.

Fig. 1(d), represents the plot of Seeds data set contain 210 samples distributed in two dimensions i.e. length of kernel groove and length of kernel. Thus, spread of data in two dimensions, indicates the presence of 3 as true number of clusters ($\mathcal{C}$). Fig. 2(d), demonstrated the results computed where, the optimal value of $\mathcal{C}$ determined by various validity indices over varying $\mathcal{C}$=$[c_{\min}, ..., c_{\max}]$; $c_{\min} = 2$; $c_{\max} = \sqrt{N} \approx 14$. The proposed validity index $VI_{DSO}$ and the existing index $V_{CWB}$ [3] achieves its minimum value at $\mathcal{C}$=3; thus, it indicates $\mathcal{C} = 3$ as optimal number of clusters which can also be verified with distribution of data shown in Fig. 1(d). Thus, $VI_{DSO}$ and $V_{CWB}$ correctly recognize the presence of three clusters. In contrast, $V_{PC}$ index [7] and $V_{PE}$ index [6] indicates $\mathcal{C}$=2 and $v_{os}$ index [2] indicates $\mathcal{C}$=14 as optimal number of clusters. Thus, it fails to correctly determine the true number of clusters.

The scatter plot of glass data set contain 214 samples distributed in two dimensional space shows the presence of 4 natural cluster according to the data distribution as presented in Fig 1(e). In Fig 2(e), we have reported the result corresponding to Glass data set where the optimal value of $\mathcal{C}$ is determined by various validity indices over varying $\mathcal{C} = [c_{\min}, ..., c_{\max}]$; $c_{\min} = 2$; $c_{\max} = \sqrt{N} \approx 14$. The proposed Index $VI_{DSO}$ and the existing index $V_{CWB}$ [3] attain its minimum value at $\mathcal{C}$=4 and thus indicates $\mathcal{C}$=4 as the optimal number of clusters which also be inferred with distribution of data shown in Fig. 1(e). These two indices only able to correctly discern the presence of 4 natural clusters. In contrast, the other indices such as $V_{PC}$ [7], $V_{PE}$ [6] indicate $\mathcal{C}$=2 and $v_{os}$ [2] indicate $\mathcal{C}$=14 as the optimal value of $\mathcal{C}$. Thus, it can be inferred from the distribution of data that the presence of 4 natural cluster is being correctly identify only by the $VI_{DSO}$ and $V_{CWB}$ index whereas the other indices $V_{PC}$, $V_{PE}$ and $v_{os}$ fails to identify correct number of clusters.

The distribution of Bupa data set contain 345 samples dispersed in two dimensions shows the presence of 2 natural clusters as presented in Fig 1(f). We have reported the results in Fig 2(f), representing the optimal value of $\mathcal{C}$ determined by various validity indices over varying $\mathcal{C} = [c_{\min}, ..., c_{\max}]$;

TABLE III. OPTIMAL NUMBER OF CLUSTERS $\mathcal{C}$ DETERMINED BY VALIDITY INDICES $V_{PC}$, $V_{PE}$, $V_{CWB}$, $v_{os}$ AND $VI_{DSO}$ ON (A) IRIS (B) WINE (C) VEHICLE (D) SEEDS (E) GLASS (F) BUPA DATA SETS WITH DIFFERENT VALUES OF $m \in [1.5, .., 2.5]$ AT STEP SIZE OF 0.2.

(A) Iris

| m | $V_{PC}$ | $V_{PE}$ | $V_{CWB}$ | $v_{os}$ | $VI_{DSO}$ |
|---|---|---|---|---|---|
| 1.5 | **2** | **2** | 2 | 12 | **2** |
| 1.7 | **2** | **2** | 2 | 12 | **2** |
| 1.9 | **2** | **2** | 2 | 12 | **2** |
| 2.1 | **2** | **2** | 2 | 12 | **2** |
| 2.3 | **2** | **2** | 3 | 12 | **2** |
| 2.5 | **2** | **2** | 3 | 12 | **2** |

(B) Wine

| m | $V_{PC}$ | $V_{PE}$ | $V_{CWB}$ | $v_{os}$ | $VI_{DSO}$ |
|---|---|---|---|---|---|
| 1.5 | **2** | **2** | 2 | 11 | **2** |
| 1.7 | **2** | **2** | 2 | 13 | **2** |
| 1.9 | **2** | **2** | 2 | 12 | **2** |
| 2.1 | **2** | **2** | 2 | 13 | **2** |
| 2.3 | **2** | **2** | 4 | 13 | **2** |
| 2.5 | **2** | **2** | 4 | 13 | **2** |

(C) Vehicle

| m | $V_{PC}$ | $V_{PE}$ | $V_{CWB}$ | $v_{os}$ | $VI_{DSO}$ |
|---|---|---|---|---|---|
| 1.5 | 2 | 2 | 3 | 29 | **3** |
| 1.7 | 2 | 2 | 3 | 29 | **3** |
| 1.9 | 2 | 2 | 3 | 29 | **3** |
| 2.1 | 2 | 2 | 6 | 29 | **3** |
| 2.3 | 2 | 2 | 20 | 29 | **3** |
| 2.5 | 2 | 2 | 22 | 29 | **3** |

(D) Seeds

| m | $V_{PC}$ | $V_{PE}$ | $V_{CWB}$ | $v_{os}$ | $VI_{DSO}$ |
|---|---|---|---|---|---|
| 1.5 | 2 | 2 | 2 | 14 | **3** |
| 1.7 | 2 | 2 | 2 | 14 | **3** |
| 1.9 | 2 | 2 | 2 | 14 | **3** |
| 2.1 | 2 | 2 | 3 | 14 | **3** |
| 2.3 | 2 | 2 | 3 | 14 | **3** |
| 2.5 | 2 | 2 | 3 | 14 | **3** |

(E) Glass

| m | $V_{PC}$ | $V_{PE}$ | $V_{CWB}$ | $v_{os}$ | $VI_{DSO}$ |
|---|---|---|---|---|---|
| 1.5 | 2 | 2 | 4 | 13 | **4** |
| 1.7 | 2 | 2 | 3 | 14 | **4** |
| 1.9 | 2 | 2 | 4 | 14 | **4** |
| 2.1 | 2 | 2 | 4 | 14 | **4** |
| 2.3 | 2 | 2 | 4 | 14 | **4** |
| 2.5 | 2 | 2 | 4 | 14 | **4** |

(F) Bupa

| m | $V_{PC}$ | $V_{PE}$ | $V_{CWB}$ | $v_{os}$ | $VI_{DSO}$ |
|---|---|---|---|---|---|
| 1.5 | **2** | **2** | 3 | 18 | **2** |
| 1.7 | **2** | **2** | 3 | 18 | **2** |
| 1.9 | **2** | **2** | 5 | 18 | **2** |
| 2.1 | **2** | **2** | 5 | 18 | **2** |
| 2.3 | **2** | **2** | 7 | 18 | **2** |
| 2.5 | **2** | **2** | 11 | 18 | **2** |

$c_{\min} = 2$; $c_{\max} = \sqrt{N} \approx 18$. These indices attain its minimum value at $\mathcal{C}$=2 and thus, indicate 2 as the optimal number of cluster for this data set. Thus, the presence of 2 natural clusters according to the data distribution is correctly recognized by $V_{PC}$ [7], $V_{PE}$ [6] and $VI_{DSO}$ index. In contrast, other indices $V_{CWB}$ [3] and $v_{os}$ [2] achieves its minimum value at $\mathcal{C}$=5 and 18, thus fails to identify the true number of clusters according to the data distribution.

However, Pal and Bezdek [13] suggested that the FCM algorithm provided the best results for $m \in [1.5, .., 2.5]$. Also, they suggested [13] that validity index is considered reliable when it is insensitive with change in values of $m$. Therefore, we reported the results on six data sets in order to judge the reliability of proposed validity index in comparison with other validity indices by varying $m$ in the range of $[1.5, .., 2.5]$ with a step size 0.2. In TABLE III(A), III(B) and III(F), we reported the results for Iris, Wine, Bupa data sets which shows that the $VI_{DSO}$, $V_{PC}$ and $V_{PE}$ index correctly identify optimal number of clusters ($\mathcal{C}$=2) for all values of $m$ which can also be inferred from the distribution of data shown in Fig. 1(a), 1(b), 1(f) indicate the presence of 2 natural clusters. Thus, proposed index $VI_{DSO}$ along with other two indices $V_{PC}$ and $V_{PE}$ is considered reliable for Iris, Wine, Bupa data sets because optimal value of $\mathcal{C}$ does not change with change in values of $m$. Similarly results reported for Vehicle, Seeds, Glass data set in TABLE III(C), III(D), III(E) the proposed index $VI_{DSO}$ is only able to correctly identify $\mathcal{C} = 3$ for Vehicle, Seeds data set and $\mathcal{C}$=4 for Glass data sets as the optimal number of cluster for all the values of $m$. In contrast, the optimal value of $\mathcal{C}$ determined by other indices for $m$=$[1.5, .., 2.5]$ vary with change in values of $m$. Thus, it can be inferred from the above discussion that, the proposed index $VI_{DSO}$ is the only index which correctly identify the presence of true number of clusters on all the above stated data sets at each value of $m$. Also optimal value of $\mathcal{C}$ determined by $VI_{DSO}$ index for all the data sets is insensitive with change in values of $m$. Hence, the $VI_{DSO}$ considered as the most reliable index above all the other indices [2], [3], [6], [7] stated above.

## V. Conclusion

In this paper, we proposed a new cluster validity index to find the optimal clusters from the number of clusters generated by FCM algorithm where, the generated clusters is of hyper-ellipsoid or hyper-spherical shape [14]. The proposed ($VI_{DSO}$) index jointly handle three measures i.e. the intra-cluster compactness, inter-cluster separation and overlap measure to determine the optimal number of clusters ($\mathcal{C}$). The proposed intra-cluster compactness measure based on the concept of relative variability therefore, it evaluates the dispersion of data points in all the dimensions. Then, it considers the dimension in which data points have maximum dispersion. This is required to minimize the over all dispersion of data points in all the dimensions. Thus, value of $\mathcal{C}$ which minimizes the intra-cluster dispersion and increase the overall compactness of data points in all the dimensions is considered as optimal value of $\mathcal{C}$. The proposed inter-cluster separation measure evaluate the separation between clusters using distance measure in fuzzy set therefore, it considers the fuzzy clusters with minimum separation. Hence, the value of $\mathcal{C}$ which maximize the inter-cluster separation among the less separated clusters corresponds to the optimal number of clusters. The proposed inter-cluster overlap measure evaluates the degree of overlapping among every pair of two fuzzy clusters. Hence, the value of $\mathcal{C}$ which minimize the inter-cluster overlap among the highly overlapped clusters corresponds to the optimal number of clusters. Therefore, the value of $\mathcal{C}$ at which $VI_{DSO}$ index achieves its minimum value corresponds to an optimal fuzzy partition and an optimal value of $\mathcal{C}$. In other way, optimal value of $VI_{DSO}$ index shows that, the intra-cluster compactness is minimized, the inter-cluster separation is maximized and the inter-cluster overlap is minimized. Further, the experimental results demonstrate the effectiveness and reliability of proposed $VI_{DSO}$ index in comparison with various other indices on six benchmark data sets. As presented in TABLE III, the proposed index for all the six data sets always correctly determine optimal number of clusters ($\mathcal{C}$) independent of change in the values of $m$ whereas, the optimal number of clusters determined by other indices changes with change in values of $m$. Therefore, the proposed index is considered more effective and reliable over other indices.

## Acknowledgment

The author would like to thank the anonymous reviewers and our colleagues for their constructive comments and suggestions, which were very helpful in improving the quality and presentation of this paper.

## References

[1] W. K. Lung, and M. S. Yang, "A cluster validity index for fuzzy clustering," *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1275-1291, Jul. 1995.

[2] K. D. Won, K. H. Lee, and D. Lee, "On cluster validity index for estimation of the optimal number of fuzzy clusters," *Pattern Recognition*, vol. 37, no. 10, pp. 2009-2025, Oct. 2004.

[3] M. R. Rezaee, B. P.F. Lelieveldt, J. H.C. Reiber, " A new cluster validity index for the fuzzy c-mean," *Pattern Recognition Letters*, vol. 19, no. 3-4, pp. 237-246, Mar. 1998.

[4] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," *Advanced Applications in Pattern Recognition*, USA, 1981.

[5] R. Krishnapuram, O. Nasraoui, and H. Frigui, "The fuzzy c spherical shells algorithm: A new approach," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 663-671, Sep. 1992.

[6] J. C. Bezdek, "Cluster validity with fuzzy sets," *Journal of Cybernetics*, vol. 3, no. 3, pp. 58-73, 1973.

[7] J. C. Bezdek, "Numerical taxonomy with fuzzy sets," *Journal of Mathematical Biology*, vol. 1, no. 1, pp. 57-71, 1974.

[8] X. L. Xie, and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841-847, Aug. 1991.

[9] Y. Fukuyama, M. Sugeno, "A new method of choosing the number of clusters for the fuzzy c-means method," *Proceedings of the Fifth Fuzzy Systems Symposium*, pp. 247-250, 1989.

[10] H. L. Kwang, Y. S. Song, K. M. Lee, "Similarity measure between fuzzy sets and between elements," *Fuzzy Sets and Systems*, vol. 62, no. 3. pp. 291-293, Mar. 1994.

[11] J. C. Bezdek, *Partition structures: A tutorial In: Bezdek, J.C. (Ed.)*, The Analysis of Fuzzy Information, CRC Press, Boca Raton, FL., 1987.

[12] C. Blake, E. Keogh, and C. J. Merz, "UCI Repository of Machine learning Databases," *Dept. Inf. Comput. Sci., Univ. California Irvine*, Irvine, CA, 1998[Online]. Available: http://www.ics.uci.edu/ mlearn/MLRepository.html

[13] N. R. Pal, J.C. Bezdek, "On cluster validity for the Fuzzy C-Means model," *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370-379, Aug. 1995.

[14] B. Rezaee, "A cluster validity index for fuzzy clustering," *Fuzzy Sets and Systems*, vol. 161, no. 23. pp. 3014-3025, July 2010.