IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 May 2023, accepted 21 May 2023, date of publication 29 May 2023, date of current version 10 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3281302

THEORY

A New Unsupervised Validation Index Model Suitable for Energy-Efficient Clustering Techniques in VANET

HAZEM NOORI ABDULRAZZAK[1], GOH CHIN HOCK[1], NURUL ASYIKIN MOHAMED RADZI[1], (Senior Member, IEEE), AND NADIA MEI LIN TAN[1,2], (Senior Member, IEEE)

[1]Institute of Power Engineering, Universiti Tenaga Nasional, Kajang 43000, Malaysia
[2]Key Laboratory of More Electric Aircraft Technology of Zhejiang Province, University of Nottingham Ningbo China, Ningbo 315100, China

Corresponding authors: Goh Chin Hock (Chinhock@uniten.edu.my) and Nurul Asyikin Mohamed Radzi (asyikin.radzi@gmail.com)

This work was supported by Tenaga Nasional Berhad (TNB) and UNITEN through the BOLD Refresh Publication Fund, Centre of Excellence, under Project J510050002-IC-6 BOLDREFRESH2025.

ABSTRACT Clustering evaluation techniques are important to check the clustering algorithm quality. High cluster similarity help to reduce the distance between a node to node within the cluster, also good separation was more important to avoid overlapping clusters. The network performance will increase and the signal will be high. Many researchers proposed different validation indexes such as Davies–Bouldin, Dunn, and Silhouette indexes. These cluster validation indexes focus on the internal or external cluster similarity, and some of them deal with both cases. The employing of graph-based distance to non-spherical clusters and selection of reference points will not be effective all the time because the average distance between reference points and all nodes will be changed dynamically such as in the VANET application. To solve this problem a dynamic sample node should be selected or the similarity of all nodes should be checked. This paper proposes a new Minimum intra-distance and Maximum inter-distance Index (M2I) to improve these indexes. The proposed index checks the internal similarity and the external distance among all nodes from cluster to cluster to ensure that high separation will occur. M2I checks the similarity from node to node within the cluster and cluster to cluster. The proposed index will be an improvement of all high-rank indexes. The proposed index was applied in different scenarios (VANET and real datasets scenarios) and compared with other indexes. The index result shows that the proposed M2I outperforms the others. The M2I accuracy is 100% in the VANET scenario and 89% in the real datasets scenario.

INDEX TERMS Clustering analysis, cluster index validation, energy clustering algorithms, K-means, unsupervised learning, VANET clustering.

ACRONYM

The list of acronyms used throughout this paper is as follows.

C.H.	The first letter of the author's name (Calinski–Harabasz).
CH	Cluster Head.
CM	Cluster Member.
CSP	Compact-Separate Proportion.
CVNN	Clustering Validation index based on Nearest Neighbors.
D.B.	Davies–Bouldin.
DBSCAN	Density Based on Spatial Clustering of Applications with Noise.
DI	Dunn index.
M2I	Minimum intra-distance and Maximum inter-distance Index.
MDL	Minimum Description Length.
ML	Machine Learning.
ND	Normalized Delay.
P.B.M.	The first letter of the author's name (Pakhira–Bandyopadhyay-Maulik).
RSSI	Received Signal Strength Indicator.

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

SC	Silhouette index.
SD	Standard Deviation.
SSE	Sum Square of the Error.
SUMO	Simulation of Urban MObility.
VANET	Vehicular Ad Hoc Network.
V2V	Vehicle to Vehicle.
WD	Within-cluster Distance.

I. INTRODUCTION

Engineering applications need to classify data into different groups, these groups called clusters. In the MANET application, the divergence of each element in an existing cluster is more important to reduce the intra-cluster distance and improve the node connection. The network performance will be high for a minimum distance. In unsupervised machine learning, there are different clustering types, and many clustering algorithms applied in different fields were proposed in the past few decades [1], [2], [3]

The clustering process creates clusters based on similarity among objects. Moreover, the clustering evaluation techniques analyze the clustering results. Objects in the same cluster have the highest similarity, while objects in different clusters have the least similarity. Cluster analysis in unsupervised learning is for learning different systems, different types of image processing, pattern recognition, and statistics [4].

There are many types of clustering algorithms such as partitioning, density-based, and hierarchical clustering algorithms. The main aim of the partitioning clustering algorithm is to divide the data into different clusters like in K-means [5], also based on center location like K-centers [6]. Some new center-based clustering techniques are proposed to avoid the centroid initialization process [7], [8], [9], [10], [11]. In density-based clustering algorithms such as Density Based on Spatial Clustering of Applications with Noise (DBSCAN) [12], the clusters are structured as multi-regions that are separated by low density or free space among the objects. Based on these assumptions the cluster discovering process can be carried out by group shapes. The hierarchical clustering algorithms build its clusters dataset hierarchy and get various clusters by blocks or divisional approaches [13]. Cluster hierarchy can offer high information compared with a single partition. However, the challenges are in analyzing the hierarchies and determining a method to obtain the affected partition from several clustering outputs. The novel model in [14] proposed a solution for multiple outputs of hierarchical clustering.

The validation of clustering algorithms evaluates the cluster result quality based on the cluster analysis roles [15]. For learning systems, the most important research field will be cluster validation. The cluster validation techniques can be divided into two categories: external and internal validations. The difference is in the information that is used in validation as external information or just internal. The external cluster validation measurement should have the cluster labels in advance. It is used mainly for electing the optimal algorithm for an existing data set. The internal cluster validation measurements deal with internal information to select the best algorithm based on the clustering algorithm results without any external information. Many researchers have proposed cluster validation models, such as the Davies–Bouldin (DB) index [16], Calinski–Harabasz (C.H) index [17], Silhouette index [18], and the standard deviation (SD) index [19]. However, those models are effective for spherical data sets clusters only. The authors in [20] proposed a minimum description length (MDL) for the clustering with high synchronization and electing the optimal clustering result. However, the MDL is not successful in multi-patterns clustering algorithms. To evaluate the arbitrary shapes clusters, some new validation indexes were proposed. In [21] authors proposed a grid-distance validation index to enhance the existing indexes.

This paper proposed a new validation index named Minimum intra-distance and Maximum inter-distance Index (M2I) based on minimizing the internal distance among the nodes and maximizing the distance between each node in different clusters. M2I can be used for internal and external validation. The minimum intra-distance and maximum inter-distance improve the quality of connection within the cluster and avoid overlapping among clusters. The intra-cluster and inter-cluster distances are illustrated in Fig.1. Moreover, the proposed index tries to select the optimal number of clusters with high separation which can be used to improve the quality of the K-means clustering algorithm and it can evaluate the K-means clustering results of VANET data sets. The experimental verification results show that M2I can evaluate the clusters with high accuracy. The main contributions of this paper include:

- A proposed new cluster validation index to evaluate unsupervised machine learning clustering algorithm.
- Simulation and testing of more popular cluster validation indexes.
- Comparative testing was designed and executed to show the new index's effectiveness.
- Verify the proposed index result using the highway K-means clustering algorithm for the VANET dataset scenario and real datasets scenario.

This paper is organized as follows: Section II presents the backgrounds of the clustering validation index. Section III summarizes the K-means Clustering Algorithm. Section IV introduces the materials and methods. Section V proposes a new cluster validation index. Section VI illustrates the V2V communication scenario. Section VII is the experimental result with index validation. Finally, the conclusion and suggested future work are provided in Section VIII.

II. LITERATURE STUDY

A validation index aims to evaluate cluster efficiency based on the cluster label and data set elements. The clustering algorithm's goal is to create clusters with high internal similarity and high separation between clusters. Most indexes

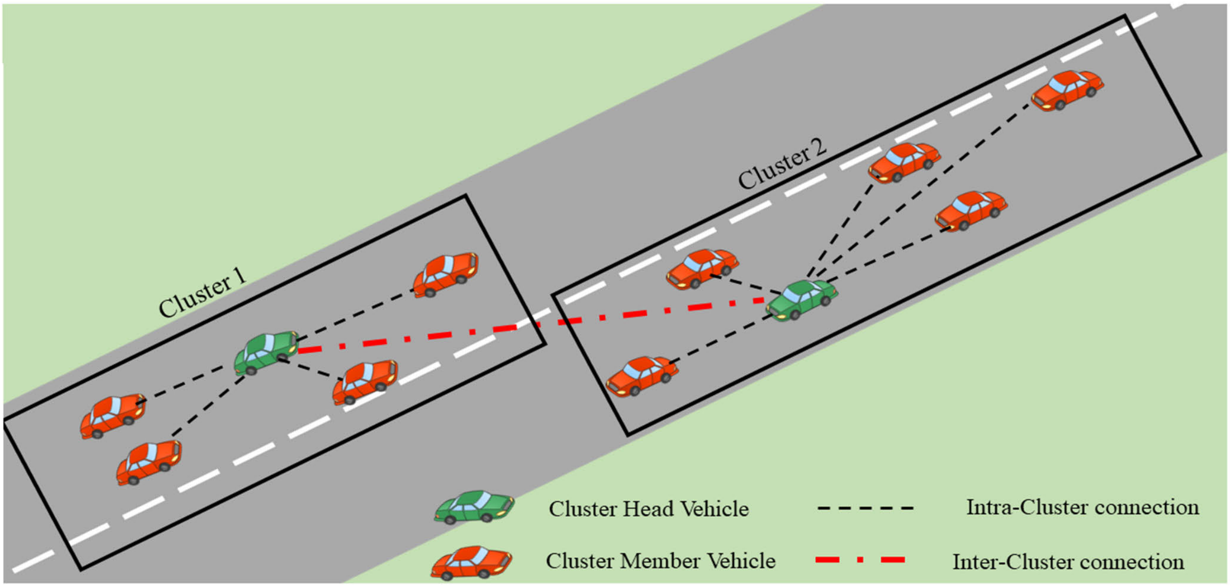

FIGURE 1. VANET intra-cluster and inter-cluster distances.

assume that clusters are separated and diverged, such as D.B. [16], C.H. [17], SD [19], and Dunn Index [22]. D.B. index calculates firstly the cluster similarity as maximizing the inter-cluster distance among clusters. Then, it computes the average similarities to get the cluster index value. The better result will be the smallest index value. C.H. index can evaluate the clustering algorithm depending on the square value of the intra-cluster sum. SD index aims depend on the cluster's separation and account for the average value of scattering. Dunn index criteria calculate the minimum distance between clusters centroid and maximum cluster diameter, this index cannot evaluate the non-spherical cluster. In [18] authors proposed the Silhouette index to evaluate the cluster algorithms by taking a sample object and calculating the intra-cluster distance and inter-cluster distance. This index can ignore some of the characteristics of other objects because it selects a sample of the data set. Some researchers are focusing on computing an optimal number of clusters and defining a new validation index [13], [23], and [24]. In [13] the authors proposed context-independent optimality and partiality index to obtain a good partition in clustering. Recently, there are some new indexes were proposed. In [25] the authors proposed a novel validity index for categorical sequences, while in [26] they generalized a self-organizing map to compute the number of clusters automatically. To address a non-spherical cluster, some validation indexes were proposed. In [20] the author uses the Minimum Description Length (MDL) principle to create the clusters by splitting the data using the values of an attribute (similar to decision tree learning); the clustering technique with MDL will select the best result. In [27] authors proposed a validation index based on arbitrary shapes and object densities. This index tries to minimize the standard deviation of clusters and maximize the separateness density between clusters. In [28] authors proposed an internal validation index, named clustering validation index based on nearest neighbors (CVNN), this index evaluates the clustering result based on the nearest neighbors and can select multiple objects dynamically as representatives for multi-clusters in multi situations. Authors in [29] proposed the compact-separate proportion (CSP) index. The CSP index evaluates the clustering algorithm by using the average distance within the cluster as a minimum value and calculating the stander deviation as the inter-cluster. The index degree is the difference between intra-cluster distance and standard deviation inter-cluster. This index is active only for well-separated clusters. In [30] the authors proposed the PBM index; the term came from the authors' names. The PBM cluster validation index will calculate the maximum inter-cluster distance, and then divide it by the minimum within-cluster distance. The PBM index value should be high for a good clustering algorithm. This paper proposed M2I to evaluate the clustering result as a modification of PBM and take all points to get high index reliability.

Table 1 summarizes the clustering validation indexes, and compares them with the proposed index.

III. K-MEANS CLUSTERING ALGORITHM

The k-means clustering algorithm is one of the most important clustering algorithms; it can be used in different fields, such as database system applications, ad hoc networks, image processing, and wireless sensor networks. In the VANET application, the main goal of K-means is to create clusters with minimum distance between all nodes as well as with cluster head (CH). The number of clusters in the K-means algorithm was calculated based on the Elbow criteria. The first big change of the SSE represents the optimal value of K. In the VANET application, the fast location change will be more effective to calculate the optimal cluster number.

TABLE 1. Summary of cluster validation indexes.

Index	Evaluation Type	Optimal Index Value	Advantages	Limitations
CSP [29]	External	Maximum	Suitable for the following datasets: • Linear datasets • Manifold datasets • Convex datasets	• Not efficient in the dynamic and variable dataset • High complexity • Internal cluster evaluation not supported.
CVNN [28]	Internal	Minimum	• Can suggest and correct the number of clusters • Best data partition on different dataset types. • Support real data sets • Support the data set that has arbitrary cluster shapes	• External cluster evaluation not supported. • evaluation of the K-means clustering algorithm not supported
DB [16]	Internal and External	Minimum	• Support different clustering algorithms • Good index for computing cluster-to-cluster similarity	• Low accuracy in the static dataset (for fixed cluster-to-cluster distance) • The index will be more effective for conditional clustering algorithms that have a threshold cluster-to-cluster distance
DI [22]	Internal and External	Maximum	• Most popular and compatible with different clustering algorithms	• Centroid-to-centroid distance is the main parameter in DI and does not support overlapping clusters.
MDL [20]	Internal	Minimum	• Minimum internal distance is the main goal, good for cluster convergence calculation	• External clustering validation is not supported.
PBM [30]	Internal and External	Maximum	• Good results in the internal and external evaluation. • High reliable clustering index	• Select the maximum distance for the sample node in external evaluation. • Not support the overlapping and dynamic datasets
SC [18]	Internal and External	Maximum	• Good results in internal and external evaluation. • High reliable clustering index	• Select the minimum distance for the sample node in external evaluation. • All nodes from cluster-to-cluster distance evaluation are not supported.
SD [19]	Internal and External	Minimum	• evaluates clustering algorithms based on scattering internally and externally. • High-reliability index.	• Not supporting the uncertainty in data mining techniques. • Need to calculate the similarity among all cluster nodes.
M2I (Proposed herein)	Internal and External	Maximum	• All nodes from cluster-to-cluster distance evaluation are supported. • Good results with dynamic datasets like VANE datasets. • High reliability in internal evaluation because compute node to node distance as well as node-to-centroid distance. • High divergence computation because M2I index compute all cluster-to-cluster similarity.	• Low efficiency in overlapping clustering algorithms, because the nodes in between clusters need to be double-checked in the processing.

However, the calculation of the K value needs to modify and create a strong model. The M2I can be used to select the optimal value of K and improve the clustering quality. The proposed index can be an evaluation index and can be used as a model to select the optimal number of clusters. The main steps of the K-means algorithm can be defined as:

Step 1. Import the K value, which represents the cluster number. The initial value will be selected randomly, in this paper the initial value will be equal to 2.

Step 2. The initial location of each cluster centroid will be selected. The selection process will be done based on the objective function, using the sum of all squared distances between all cluster members and cluster centroids for all clusters, the optimal location objective function is:

$$\begin{aligned}\mathrm{avgmin}_c \sum_{i=1}^{k} \sum_{v_j \in C_i} d(\mathrm{vehicle}_j, \mathrm{centroid}_i) \\ = \mathrm{avgmin}_c \sum_{i=1}^{k} \sum_{v_j \in C_i} \left|v_j - u_i\right|^2 \end{aligned} \tag{1}$$

where $d(\mathrm{vehicle}_j, \mathrm{centroid}_i)$ is the vehicle node to the centroid of its cluster as Euclidean distance; $\mathrm{vehicle}_j$ is the vehicle location and $centroid_i$ is the location of cluster centroid with $i = 2, \ldots, K$; and K is the number of the total cluster.

TABLE 2. Indexes comparison.

Index	Evaluation Steps	Description
SC [32], [33]	• Step 1. Calculate the Intra-clustering distance (F1) by selecting a sample node and compute the average distance with all nodes in the existing cluster. • Step 2. Calculate the minimum Inter-clustering distance (F2) by selecting a sample node and compute the average distance with all nodes in the closest cluster. • Step 3. Compute the SC index. The index value will be in the range {-1, 1}. If the value is close to 1, the clustering algorithm is good and the nodes belong to the optimal cluster.	$F1(v)=\frac{\sum_{v_j\in C_j} d(\text{Sample vehicle}_j, v_j)}{n_j-1}$ $F2(v)=\min\left\{\frac{\sum_{v_k\in C_k} d(\text{Sample vehicle}_j, v_k)}{n_k}\right\}$ $SC(j)=\frac{F2(v)-F1(v)}{\max\{F2(v),F1(v)\}}$ Where:- v_j: Vehicle in the j^{th} cluster, C_j: j^{th} cluster, n_j: total nodes in the j^{th} cluster, Sample vehicle$_j$: Sample node from the j^{th} cluster,: Vehicle in the k^{th} cluster, C_k: k^{th} cluster, and n_k: total nodes in the k^{th} cluster.
DB [32], [33]	• Step 1. Calculate the minimum Intra-clustering distance (Sj) by selecting a sample node and compute the average distance with all nodes in the existing cluster. • Step 2. Calculate the maximum Inter-clustering distance (D_{jk}) by computing the distance between clusters' centroid. • Step 3. Compute (R_{jk}) as dividing the internal distance into different clusters by D_{jk}, and the maximum value is taken. • Step 4. Compute DB as the average value of R_{jk}.	$Sj=\frac{\sum_{v\in cj} d(v_x,v_y)^2}{n_j}$ $D_{jk}=d(\text{Centroid}_j,\text{Centroid}_k)$ $R_{jk}=\frac{Sj+Sk}{D_{jk}}$, $R_j=\max(R_{jk}), j\neq k, 0\le k<n_c$ $DB=\frac{\sum_{j=1}^{n_c} R_j}{n_c}$ Where:- n_c: Number of total clusters, n_j: Number of vehicles in the j^{th} cluster, and v_x&v_y: Two vehicles in the j^h cluster.
DI [32]	• Step 1. Calculate the distance between the centroids as $D_{\text{inter-cluster}}$ • Step 2. Compute the cluster dimension by calculating the maximum distance between two vehicles in the same cluster as $D_{diam.}$ • Step 3. Compute the DI.	$D_{\text{inter cluster}}=\{d(\text{centroid}_i,\text{centroid}_j)\}$ $D_{diam}(C_i)=\max\{d(v_x,v_y)\}$ $DI=\frac{\min(D_{\text{inter cluster}})}{D_{diam}(C_i)}$ Where:- v_x&v_y are two vehicles in the i^{th} cluster.
PBM [30]	• Step 1. Compute the maximum distance between clusters as the maximum distance between different nodes in different clusters (D_{max}) • Step 2. Compute within cluster distance as an intra-cluster distance (WSS) • Step 3. Calculate the PBM Index	$PBM=\frac{1}{k}*\frac{1}{WSS}*D_{max}$ Where:- D_{max}: max. distance between clusters, WSS: within cluster distance, and K: number of clusters.
SSE [32]	• Step 1. Compute the similarity matrix as an internal distance ($S_{Distance}$). • Step 2. Calculate the SSE as a total sum of the similarity matrix. • Step 3. Compute the optimal cluster number as an elbow value.	$S_{Distance}=\begin{bmatrix}0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & 0 & \\ d(n,1) & d(n,2) & d(n,3) & \dots & 0\end{bmatrix}$ $SSE=\sum_{x=1}^{n}\sum_{y=1}^{n} d(v_x,v_y)^2$ Where:- v_x&v_y are two vehicles in the i^{th} cluster.

Step 3. Assigning all vehicles to clusters based on the optimal location of cluster centroid.

Step 4. Take each node and calculate its distance to the centroid. If a node does not belong to the cluster with the nearest centroid, collect this node and switch to the next cluster. Then update the location of the cluster centroid as:

$$\text{centroid}_i=\frac{1}{|C_i|}\sum_{j\in C_i}\text{vehicle}_j,\ \forall i \tag{2}$$

Step 5. Repeat Step 4 until convergence is achieved or until the node distance is unchanged.

Check and verify that all vehicles have been clustered.

Fig. 2 represents the important steps of a k-means clustering algorithm [31].

IV. MATERIALS AND METHODS

There are many indexes to evaluate the clustering algorithm. This paper employs several more popular techniques for evaluating the K-means clustering algorithm and then comparing it with the M2I proposed index. The comparison indexes are SC, DB, DI, PBM, and SSE. Table 2 illustrates the evaluation steps of these indexes with a full description of all formulas.

V. M2I CLUSTER VALIDATION INDEX

In this section, the definition of M2I is presented based on the distance problem matrix. For each clustering technique, the clustering gets a high degree of performance according to the high separation method as well as high internal similarity. M2I is a degree index of clustering that has maximum

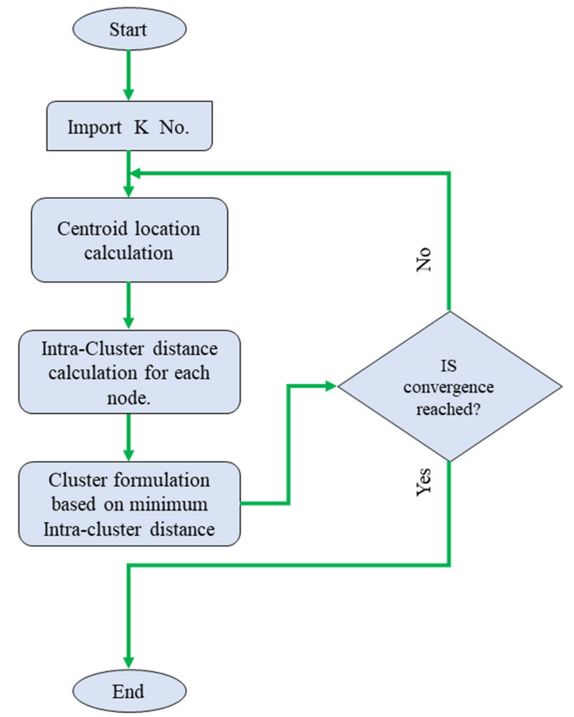

FIGURE 2. K-means algorithm processing steps.

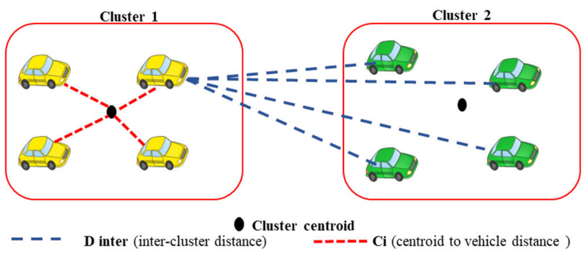

FIGURE 3. Cluster environment.

inter-cluster distance and minimum intra-cluster distance. The proposed model has the following steps:

Step 1: Calculate the centroid matrix for all clusters, the centroid matrix represents the distance between the vehicle and the cluster centroid. For each node (vehicle) in the ith cluster, the centroid matrix can be defined as:

$$C_i = \left[d_{v1,c}, d_{v2,c}, d_{v3,c}, \ldots, d_{vn,c}\right] \tag{3}$$

where: d is the Euclidean distance between the vehicle and the cluster centroid it can be calculated as:

$d_{v_i,c} = \sqrt{(v.x(i) - c.x)^2 + (v.y(i) - c.y)^2}$, for ∀ Vehicles in a cluster with the location (v.x, v.y) and cluster centroid with location (c.x, c.y).

Step 2: Calculate the minimum Euclidean distance between each node (vehicle) with other nodes from the closest cluster as:

$$D_{inter} = \sqrt{(v.x(i) - v.x(j))^2 + (v.y(i) - v.y(j))^2} \tag{4}$$

$$D1 = \min\{D_{inter} \tag{5}$$

where: v_i is the vehicle in the ith cluster and v_j is the vehicle in the jth cluster, i and j are the closest clusters.

The clustering environment of Ci and D1 calculation is illustrated in Fig. 3.

The minimum inter-cluster distance calculation algorithm is shown in Algorithm 1 which it's important to compute the minimum distance between each node with the closest cluster nodes as explained in steps 1 and 2. There are three cases of cluster location in whole clustering sets. Fig. 4A shows the inter-cluster distance calculation range of the first cluster. Firstly, the node's distances in the cluster that is located in the first slot are calculated. All other possible inter-cluster nodes that can communicate with cluster 1 are located in the range between T+1 to Total. Note that, T is the total number of nodes in cluster 1, T+1 represents the first node in cluster 2, and Total represents the total number of nodes. Fig. 4B shows the inter-cluster range consideration for the last cluster. The total nodes of other clusters are from the range 1 to M-1. The total nodes in the last cluster are Total – (M-1). Note that, M-1 is the last node in the last cluster before cluster n also known as cluster n-1. The possible inter-cluster nodes that can communicate with the last cluster (cluster 1 to cluster n-1) can be calculated for the range of nodes (1 to M-1). Finally, Fig. 4C presents the cluster located in between the upper and lower range that was divided into two intervals: left and right. The clusters in the left interval are from (1 to i-1), and the inter-cluster distance range will be from node number 1 to node number M-1. Since the nodes from 1 to M-1 represent the nodes from clusters on the left side. The right interval is the range of clusters from i+1 to n. The inter-cluster distance on this side will be computed from node number N+1 to the last node (Total), where cluster i node range is from (M to N). All loops in Fig. 4 were mentioned in the D1 algorithm.

Step 3: Calculate the maximum Euclidean distance between each node and other nodes in the same cluster. It's the intra-cluster distance and the maximum value we use to evaluate the clustering technique. The M2I model looks for a low rate between the inter and intra as well as a high value of the index. $D_{intra} = \sqrt{(v.x_i - v.x_k)^2 + (v.y_i - v.y_k)^2}$ for i,k: two vehicles in ith cluster

$$D2 = \max\{D_{intra}\} \tag{6}$$

Step 4: Calculate the node index for each node in cluster I based on the high value (logic 1) of the M2I, M2I computes the average two ratios as an important parameter. The first ratio between D1 and D2 shows the minimum value of the inter-cluster distance is higher than the maximum value of the intra-cluster distance to give the first point of the cluster. The second ratio between the minimum inter-cluster distance and the centroid matrix is important to indicate that the clustering has full separations and this technique will be applied to all nodes. The main M2I description is shown:

$$M2I = \left\{\frac{D1}{D2} + \frac{D1}{Ci}\right\} * \frac{1}{2} \tag{7}$$

where:-

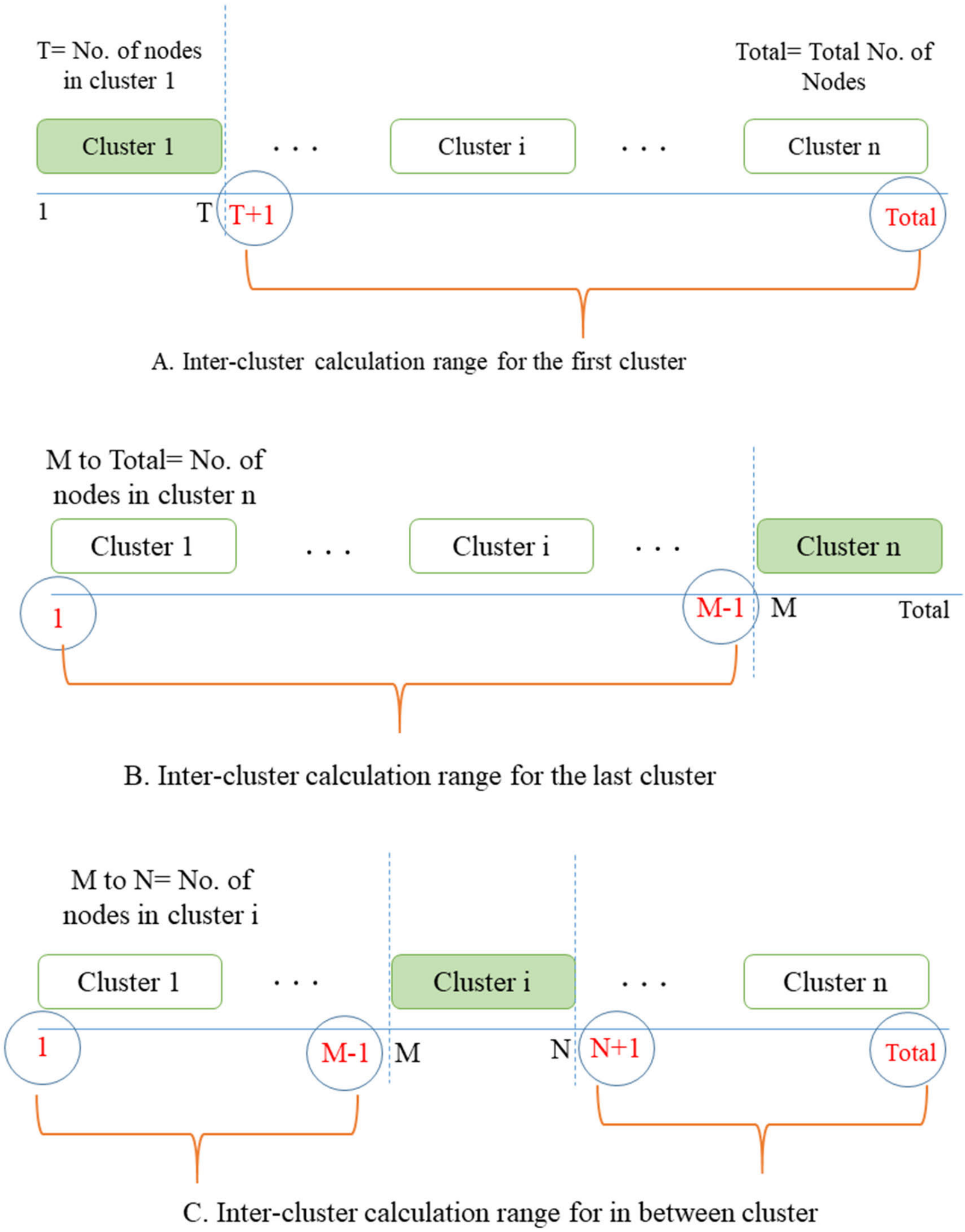

FIGURE 4. The upper and Lower range of the D1 calculation.

Ci: vehicle to centroid distance based on (3)
D1 & D2 were calculated in (5), (6)
By rewriting the formula (7) as:

$$M2I(i) = \frac{C_i D_1 + D_1 D_2}{2 C_i D_2} \tag{8}$$

And the node index will be:

$$node_{index}(i) = \begin{cases} 1, & M2I(i) \geq 1 \\ 0, & M2I(i) < 1 \end{cases} \tag{9}$$

Step 5: In this step, the total cluster index can be calculated as:

$$\text{Cluster M2I} = \frac{\text{No. of high value innode}_{index}}{n} \tag{10}$$

where: n: number of vehicles in ith cluster

Step 6: Finally, the average index of the clustering algorithm will be:

$$\text{Average M2I} = \frac{\sum_{i=1}^{n} \text{Cluster M2I(i)}}{\text{No. of clusters}} \tag{11}$$

Algorithm 1 Minimum Inter-Cluster Distance (D1)

Input: Vehicle Locations, No. Of Clusters (k), No. Of Vehicles per Cluster (n), and Total No. Of Vehicles
Output: D1 Result.
T=0;
M=T+1;
T=T+n;
for i = 1to Total **do**
for j=1to Total **do**
D1(i,j)= Euclidean Distance Among All Vehicles;
End for;
End for;
for s=1 to k **do**
for i =m to T **do**
if s=1 **then** // for the First Cluster
for j=T+1to Total **do**
D1(s)←minimum Distance;
End for;
End if;
if s=k **then** // for the Last Cluster
for j=1 to m-1 **do**
D1(s)←minimum Distance;
End for;
End if;
if s ≠1 && s≠ k **then** // for the in-Between Cluster
for j=1 to m-1 **do**
D1(s)←minimum Distance;
End for;
for j=N+1 to Total **do**
D1(s)←minimum Distance;
End for;
End if;
End for;
End for;
End

M2I value will be in the range {0, 1}, the high value indicates a good clustering algorithm. Algorithm 2 represents the summary of the proposed index steps.

VI. V2V COMMUNICATION SCHEME

The V2V connection helps to understand the quality of clustering and the optimal clustering scheme to get high performance. Fig. 5 shows the connections in a cluster and the way to calculate the average delay and RSSI (Received Signal Strength Indicator). The vehicles will classify into cluster member nodes and one cluster head node. In this scheme, the cluster head node was the nearest node to the cluster centroid.

Definition 1: Let C be a cluster node and {d(v1, ci), d(v2, ci), …, d(vn, ci)} is a set of the distance between the centroid and vehicle node. Assuming that delay∝ distance, the

Algorithm 2 M2I Algorithm

Input: vehicle locations, No. of clusters (k), No. of vehicles per cluster (n), total No. of vehicles, and cluster centroid.
Output: M2I result.
T=0;
M=T+1;
T=T+n;
for i= 1 to k **do**
for j=1 to n **do**
$\mathrm{Cd(i,j)} = \sqrt{(\mathrm{v.x(j)} - \mathrm{c.x})^2 + (\mathrm{v.y(j)} - \mathrm{c.y})^2}$
end for;
end for;
for s=1 to k **do**
for i = m to T **do**
D-intra (s, i)← calculate the intra-cluster distance;
D2 (s,i)← maximum D-intra(s,i);
end for;
end for;
Apply Algorithm 1 to calculate the value of D1
for s=1 to k **do**
for i = m to T **do**
M2I(s, i) ← applying eq. 8
Node index (s, i) ← applying eq. 9
end for;
end for;
for s=1 to k **do**
for i = m to T **do**
cluster M2I(s, i) ← applying eq. 10
average M2I (s) ← applying eq. 11
end for;
end for;
End

normalized delay will be defined as [34], [35]:

$$\mathrm{ND_X(Y)} = \begin{cases} \dfrac{\mathrm{d(x, y)}}{\mathrm{R}}, & \mathrm{d(x, y)} < \mathrm{R} \\ 1, & \mathrm{Others} \end{cases} \tag{12}$$

where: R represents the maximum range for two vehicles to still in connects (sum of the coverage source node area and the coverage distention node area).

According to Fig.5, the number of connections is equal to the number of member nodes. The possible connection will be M, and the average cluster normalized delay will be computed as:

$$\mathrm{Avarage_{ND}} = \frac{\sum_{i=1}^{M} \mathrm{ND_{x,y}}}{\mathrm{M}} \tag{13}$$

Definition 2: Let C: be a cluster node {v1, v2, …, vn} and {d(v1, ci), d(v2, ci), …, d(vn, ci)} is a set of the distance between centroid and vehicle node. Assume the wavelength is λ and the light speed is c. The Intra-cluster V2V RSSI will

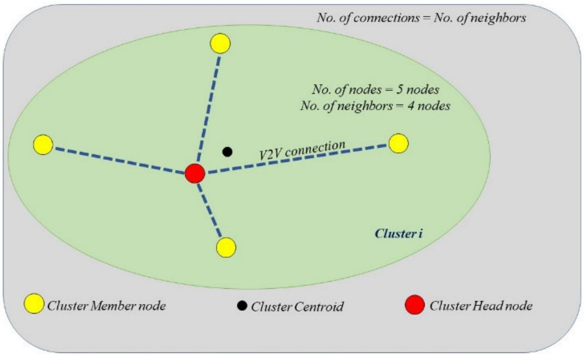

FIGURE 5. Cluster node connection scheme.

be defined as [36], [37]:

$$\mathrm{RSSI} = \mathrm{P_0} - 20\log_{10}\left(\frac{4\pi \mathrm{d}}{\lambda}\right) \tag{14}$$

where:

$\lambda = \frac{c}{f}$, c $= 3 * 10^8$ m/s and f $= 2.4$ GHz

f = carrier frequency and P0 is an empirical constant and its value is set to 31.0 dBm

VII. EXPERIMENTAL RESULTS AND INDEX VALIDATION

The cluster indexes evaluation was presented in this section. There are two different scenarios. The first scenario is the VANET dataset scenario and the second scenario is the real dataset scenario. As an evaluation process, the clusters should be created based on the k-means algorithm and importing the datasets which have been illustrated in subsection A as a datasets initialization stage. The simulation environments of the VANET scenario have been presented in B and the evaluation process steps were illustrated in C. Subsection D discusses the evaluation and testing results. Finally, the evaluation and testing of the second scenario were represented in E.

A. DATASETS INITIALIZATION

One of the most used vehicular simulators that are open-sourced and easy to implement is Simulation of Urban Mobility (SUMO). SUMO will help users to simulate a traffic system with vehicle mobility and different public transportation [38] and [39]. This simulator has multi-support tools which are suitable to implement different tasks such as:

- Searching and detecting the vehicle routes
- Highway network implementation and generation
- Network importing from different databases like Open Street Map
- Tracing the VANET model and exporting the mobility results as an XML file.

The SUMO output-tracing file can be used in different simulators [40]. In this paper, the XML file has been imported into Matlab with different vehicle movement timestamps. A highway scenario was used and the snapshot of SUMO is illustrated in Fig. 6.

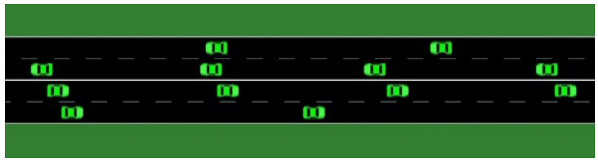

FIGURE 6. SUMO highway scenario.

B. SIMULATION ENVIRONMENTS

This subsection represents the environment of simulation that was implemented in this paper. The laptop processor is core i7-8 550u with 2 × 1.99 GHz and 16 GB RAM. The operating system is Windows 10 Pro. The proposed model and all evaluations and testing algorithms as well as V2V communication were implemented in MATLAB 2018b. The mobility simulation for the VANET highway scenario was simulated using SUMO. Table 3 shows the important parameters setting for the VANET dataset scenario.

C. EVALUATION PROCESS OF VANET DATASET SCENARIO

The evaluation process of the VANET scenario is shown in Fig.7 and can be defined as:

Step 1. Two cases from SUMO datasets will be imported into the main program, applying the K-means clustering algorithm with different K values {2, 3, 4, 5, 6, 7, 8, 9}, and collect the cluster results for both cases.

Step 2. Test and evaluate the clustering results using (SC, DB, DI, PBM, and SSE) indexes, and compute the optimal K value according to the indexes evaluation results.

Step 3. Appling the V2V connection environment and then calculating the RSSI and normalized delay within a cluster in both scenarios (case 1 and case 2).

Step 4. Evaluate the V2V experiment results and select the optimal results with an optimal number of clusters to get an efficient network performance.

Step 5. Compare the results of Steps 2 and 4, the final results of this step represent the correct and efficient K value.

Step 6. Applying M2I proposed model to test and evaluate the clustering results of step 1, the results of this step show the optimal K value passed on M2I.

Step 7. Compare the results of Steps 5 and 6 to evaluate the proposed index model based on previous high-performance indexes and V2V experimental environment.

D. EVALUATION OF TEST RESULTS

1) DIFFERENT VALIDATION INDEXES RESULTS

In this subsection, the SC, DB, DI, and PBM validation indexes were implemented for both cases (Case 1 and Case 2) using the VANET database. These indexes were the main validation indexes and any new model that was compared with it to show the new model's effectiveness. Fig.8 shows the indexes result for Case 1. As shown in Fig.8a, the SC index has the optimal value of 0.8922 at K=3 and the other values are less than 0.8922 for K=2, 4, 5, 6, 7, and 9. Because of the goal of SC, when the K has increased, the

TABLE 3. Parameters and simulation settings.

Parameters	Description	Value
VANET Network Topology	Highway scenario	2 km road length and four lanes
Directions and simulation time	There are two direction road topology	150 Sec. SUMO simulation time
Cases and datasets	Two cases Based on two different timestamps.	Case 1 after 20 Sec. from the simulation starting time Case 2 after 75 Sec. from the simulation starting time
Vehicle Density	-----------	50
Vehicle node speed	-----------	0-33 m/sec.

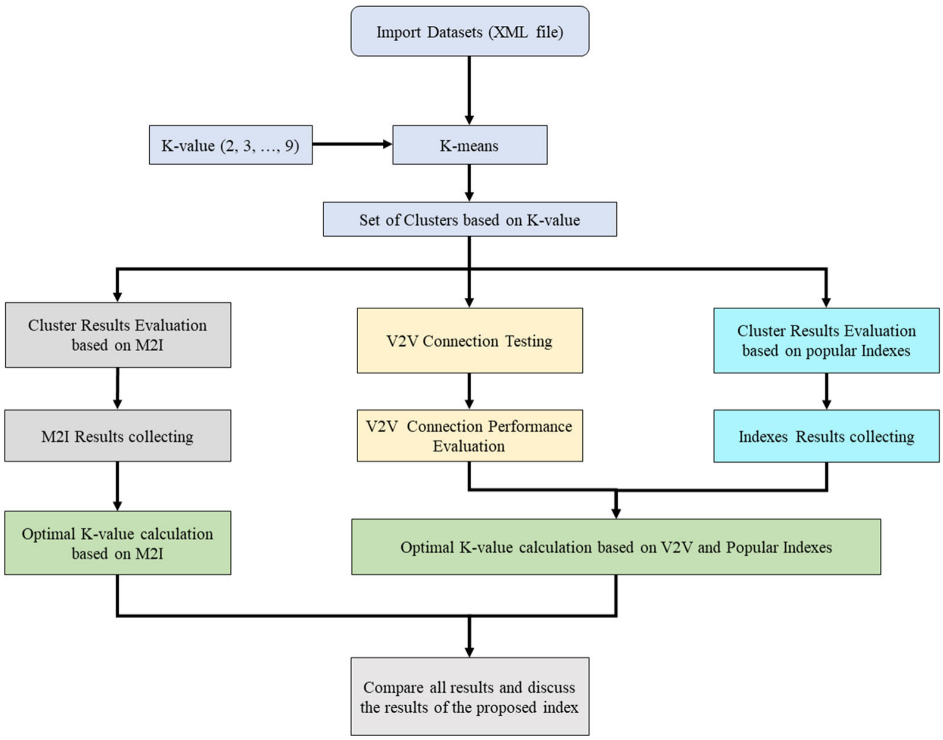

FIGURE 7. Evaluation process of VANET dataset scenario.

distance between clusters was decreased and the SC value was decreased. SC index evaluates the internal and external clustering algorithm. The minimum value of the index result will indicate that the clustering algorithm is good as shown in Fig.8 b for the DB index with the index value 0.218201 at K=3. When the distance between the centroids of the closest clusters is decreased the DB index value will increase and the quality of clustering is decreased. Fig.8c shows the Dunn index results to evaluate the K-means clustering algorithm, the optimal value in this index should be high because DI tries to make the clusters more convergent and increase the distance between clusters (high cluster-to-cluster divergence). The optimal value is 7.7609 with K=3, and the remaining results for K=4, 5, 6, 7, 8, and 9 are decreased because the cluster-to-cluster divergence is low. PBM selects the maximum distance between different nodes in different clusters a concerning index with respect to internal distance, the external distance should be high. Fig. 8 d shows the PBM index value, and the optimal value is 5.545691 with K=3. According to all indexes in Fig. 8, the optimal K value is 3 for case 1.

Fig. 9 shows the indexes result for Case 2. As shown in Fig. 9a, the SC index has the optimal value of 0.95208 at K=2 and the other values are less than 0.95208 for K=3, 4, 5, 6, 7, and 9. Because of the goal of SC, when the K has increased the distance between clusters was decreased and the SC value was decreased. By comparing the results with Case 1, the optimal value is changed from 3 to 2 because the second scenario has high traffic and the clusters are nearby. For the DB index, the minimum value of the index result will indicate that the clustering algorithm is good as shown in Fig. 9b. the minimum value is 0.091531 at K=2. When the distance between the centroids of the closest clusters is decreased the DB index value will increase and the quality of clustering is decreased. The maximum value of the Dunn index indicates that the cluster-to-cluster divergence is high (high separation), as shown in Fig. 9 c, the optimal value is the maximum value of DI is 111.607 at K=2. The remaining values decrease because the cluster-to-cluster distance is decreased when K is increased (low separation). Fig.9d, shows the PBM index evaluation results, the maximum value is 69.5397 at K=2, as discussed, the PBM index value should

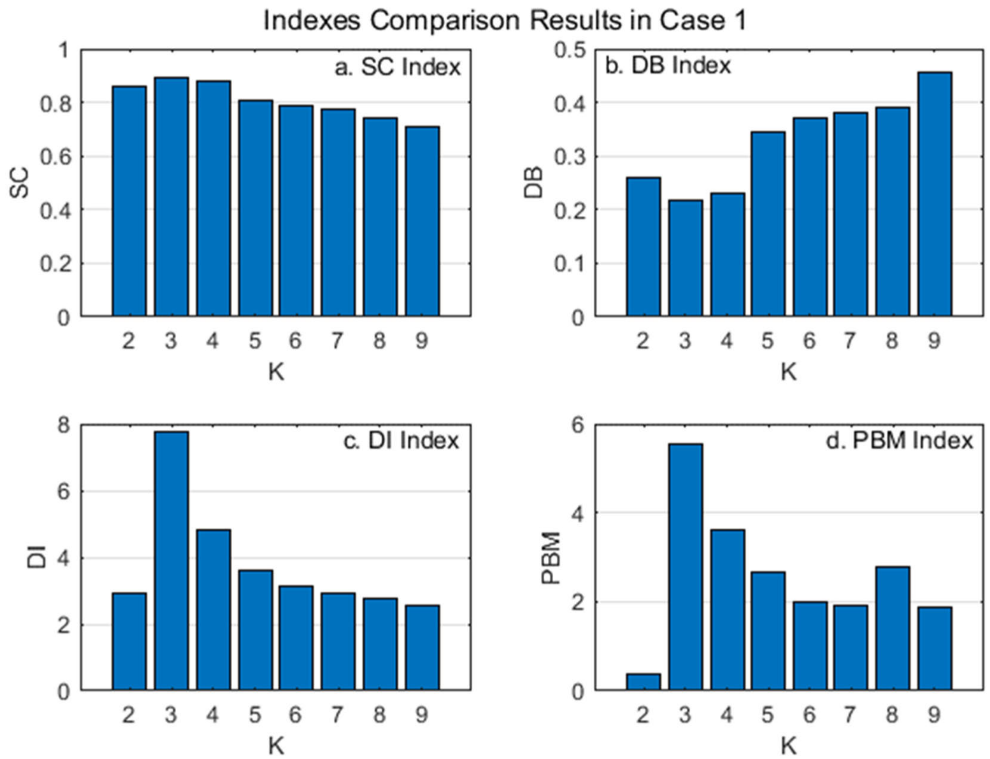

FIGURE 8. Indexes Comparison Results for Case 1 in the VANET dataset scenario: a) SC Index (OIV: maximum). b) DB index (OIV: minimum), c) DI index (OIV: maximum), and d) PBM index (OIV: maximum).

be high to ensure that the inter-cluster distance is high. The other results are decreased when the K value is increased, because of the high traffic in Case 2 and the inter-cluster distance is low. According to all indexes in Fig.9, the optimal K value is 2 for Case 2.

The Sum of Square Error (SSE) represents the similarity in the cluster itself. The values of SSE can help to identify the convergence data in clusters but not the information about the divergence among the clusters. The optimal values of K are 3 for Case 1 and 4 for Case 2 with error values of 153,959 and 20,254 respectively. Fig.10 shows the Elbow curve of the SSE result.

2) V2V TESTING

V2V connection was established to check and evaluate the clustering algorithm. The range of K is {2, 3, 4, 5, 6, 7, 8, and 9}. In each K value, the V2V scenario computes the RSSI value within the cluster and the normalized delay. The optimal value of RSSI and normalized delay represents the real cluster number which can give high network performance. The V2V results can be used to compare with the cluster validation indexes and judge if the index is suitable for the VANET application or not.

As shown in Fig. 11, the optimal values of RSSI are 3 for case 1 and 2 for case 2 Fig. 11a shows the RSSI value of case 1. The first high RSSI value is -67.99445 dBm at K=3, the other value is good but with a high cluster number. In clustering techniques, the minimum number of clusters is better and also can be compared with Fig.12 a, the optimal delay value is at K=3 in case 1. By comparing the two results (RSSI, normalized delay) in Case 1, the optimal and correct value of K is 3 for Case 1. For case 2, Fig. 11 b and Fig. 12 b shows the RSSI and normalized delay in case 2 respectively. The optimal RSSI value is -48.85731 dBm at K=2, and the optimal normalized delay is at K=2. According to the last values of RSSI and normalized delay for case 2, the correct value of K is 2 in Case 2.

3) M2I EVALUATION

Finally, the evaluation result of the proposed index (M2I) is explained in Fig. 13. The M2I computes the convergence and divergence of clusters, and the optimal values represent the exact value that the cluster number is optimal. Fig.13a illustrates the M2I value in Case 1, the values changed based on the distance between clusters, the distance between different nodes in different clusters, and the cluster intra-distance. The maximum value is 0.666667 at K=3. In Fig.13b the maximum value of the M2I index in case 2 is 0.5 at K=2. The M2I index value range is from 0 to 1.

4) RESULTS DISCUSSION

The experimental results for different cluster validation indexes, V2V communication results, and M2I results can be summarized in Table 4. The correct value of K is 3 in Case 1 and 2 in Case 2 based on the V2V communication scenario. All indexes and M2I also get the same result. The

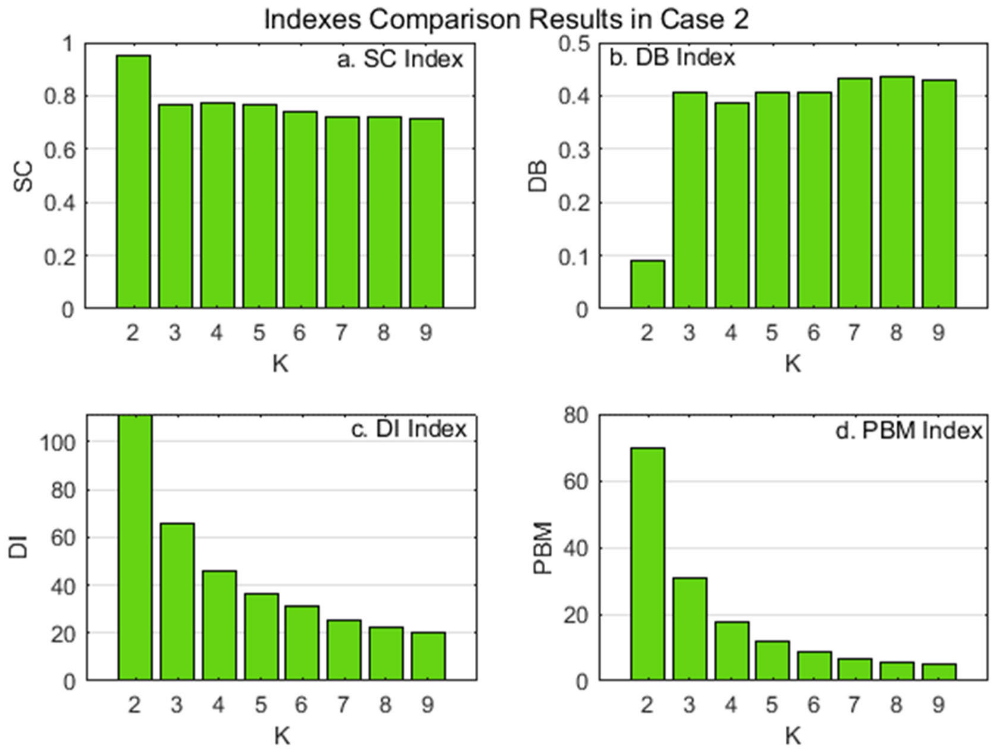

FIGURE 9. Indexes comparison Results for Case 2 in VANET dataset scenario: a) SC Index (OIV: maximum). b) DB index (OIV: minimum), c) DI index (OIV: maximum), and d) PBM index (OIV: maximum).

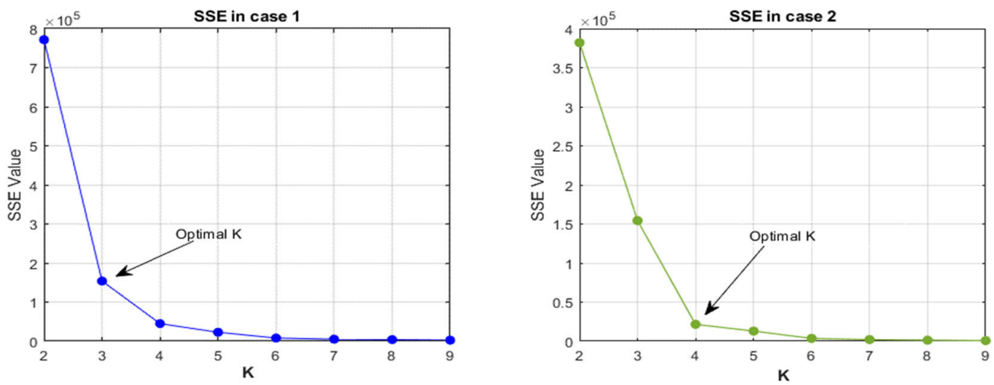

FIGURE 10. The testing value of SSE in case 1 & case 2 in the VANET dataset scenario.

TABLE 4. Summary of result with the optimal value of K.

Index	Case 1			Case 2		
	Index Value	The optimal value of K	Efficient K value-based V2V connection test in case 1	Index Value	The optimal value of K	Efficient K value-based V2V connection test in case 2
SC	0.8922	3		0.95208	2	
DB	0.218201	3		0.091531	2	
DI	7.7609	3	3	111.607	2	2
PBM	5.545691	3		69.5397	2	
SSE	153959	3		20253.74	**4**	
M2I	0.666667	3		0.5	2	

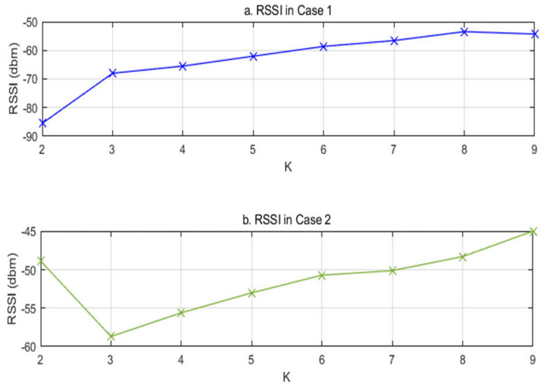

FIGURE 11. RSSI in case 1 & case 2 of the VANET dataset scenario.

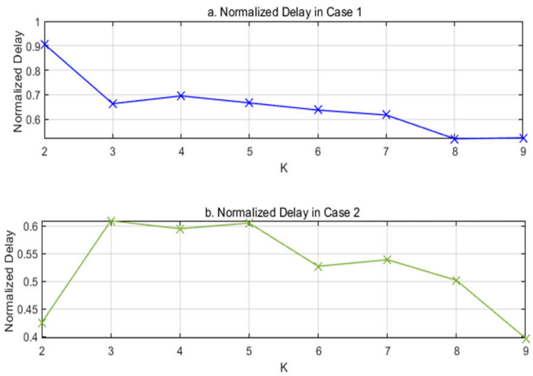

FIGURE 12. Normalized Delay in Case 1 and Case 2 of the VANET dataset scenario.

value of SSE in Case 2 is 4, as the SSE computes the internal error and deals with the inter-cluster distance. The high vehicle density in Case 2 distributes the nodes nearby, so the optimal clustering is to create minimum clusters and increase the cluster member in each cluster. All indexes select K=2 in this case instead of K=4, the optimal value of K is 2 according to internal and external testing. The optimal K value based on SSE is 4 because the clustering evaluation using SSE is done based on the optimal intra-cluster distance only, the high clustering separation was not computed.

E. SECOND SCENARIO (REAL DATASETS)

To better validate the effectiveness of the proposed M2I, several real data sets that are imported from UCI Machine Learning Repository [41] are applied. Many researchers select these data sets to evaluate the clustering validation indexes. In [28], [42], and [43], the UCI real data sets were tested and the validation indexes were evaluated.

Table 5 represents the characteristics of real data sets and the value of attributes, instances, and the corrected cluster number.

As illustrated in Fig. 13 (a), the optimal value of k is 3 which is marked in a red dot. The maximum value of M2I is 0.666 in k=3 and the other value is less than 0.666. The DS 1 dataset has 4 attributes, and the M2I succeeded in obtaining the correct value of k in this dataset. The red dot values in Fig.14 (b) to (f) show the optimal value of K which is

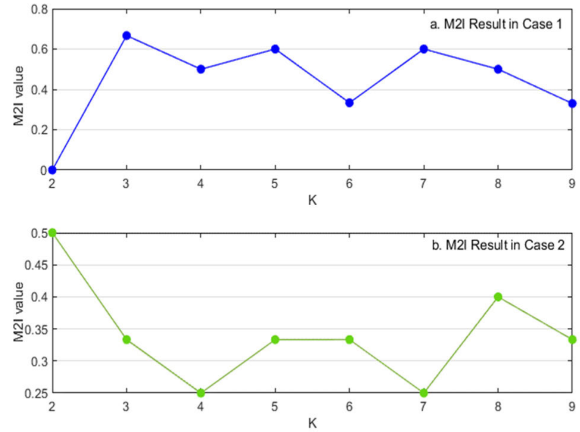

FIGURE 13. The testing value of the M2I index in case 1 & case 2 of the VANET dataset scenario.

TABLE 5. The characteristics of real datasets.

Datasets	Attributes	Instances	Correct Cluster No.
Iris (DS 1)	4	150	3
Magic (DS 2)	10	19020	2
Breast cancer (DS 3)	10	699	2
Glass (DS 4)	9	219	6
Wine (DS 5)	13	178	3
Car (DS 6)	6	1728	4
Control (DS 7)	60	600	6
Page blocks (DS 8)	10	5473	5
Parkinson (DS 9)	22	195	2

calculated using M2I. The M2I values in these datasets are (1, 0.5, 0.833, 1, and 0.75) for DS 2 to DS 6. The M2I results match exactly with the k values in the UCI datasets.

The M2I value of DS 7 is shown in Fig.14 (g) is 1 and the optimal value of k is 5. As mentioned in Table 5, the correct value of k is 6. The M2I cannot select the correct value of K in DS 7. The number of attributes is 60 as shown in Table 5, so the proposed validation index has a little error in overlap datasets and the datasets with a high number of attributes. Finally, the M2I values of DS 8 and DS 9 are equal to 1 in both datasets and the optimal k values are 5 and 2, respectively, which is marked in red dot as shown in Fig.14 (h) and (i).

In this experiment, the following indexes (DB, SC, DI, PBM, CSP, CVNN, and M2I) are applied. The optimal cluster number obtained from these indexes is shown in Table 6, where the underlined bold values represent the correct k value compared with the dataset k value which is imported from the dataset source.

The index accuracy can be calculated as [44]:

$$\text{Index Accuracy} = \frac{\sum_{i=1}^{N} \delta(\bar{c}_i, \hat{c}_i)}{N} \tag{15}$$

where $\hat{c}_i$ is the real cluster label, $\bar{c}_i$ is obtained by matching clustering result c to real clustering labels $\hat{c}$ and then renaming ci according to the best-matched true label as

$$\delta(\bar{c}_i, \hat{c}_i) = \begin{cases} 1, & \text{for } \bar{c}_i = \hat{c}_i \\ 0, & \text{others} \end{cases},$$

The Accuracy comparison results are shown in Fig.15.

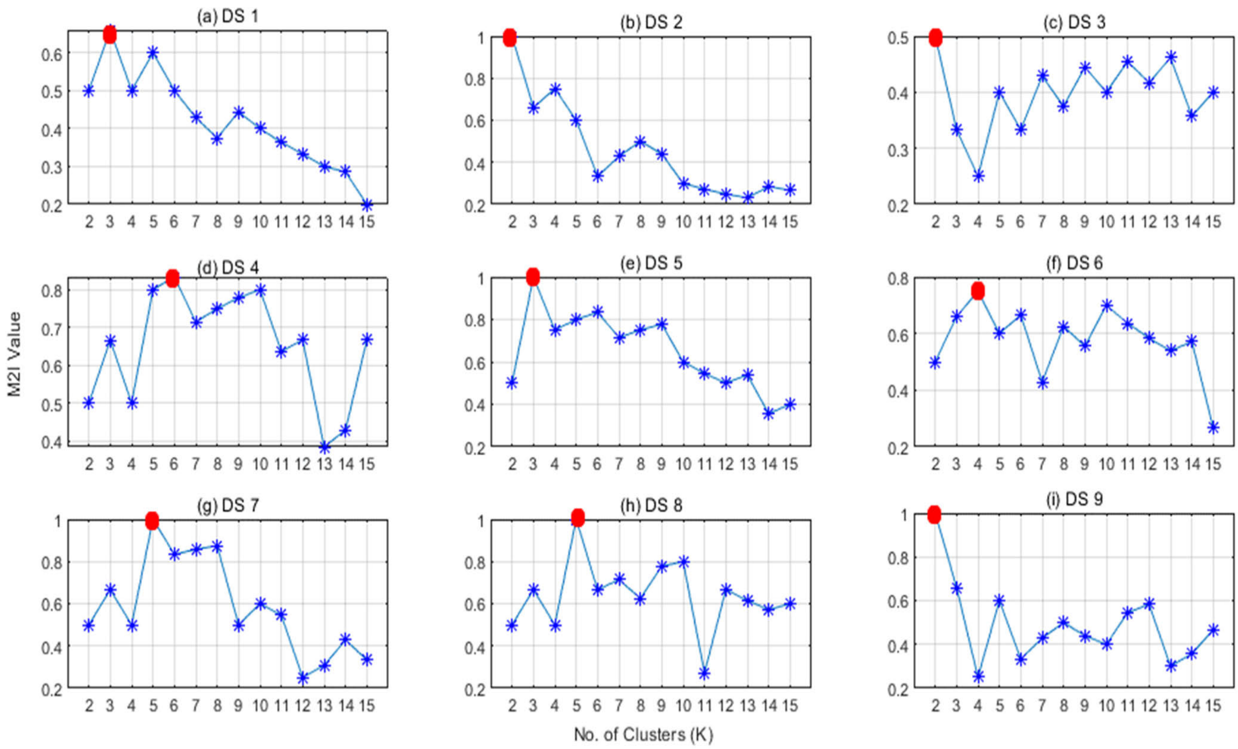

FIGURE 14. M2I results for different real data sets.

TABLE 6. Comparison results of different Indexes over UCI datasets.

Datasets	DB	SC	DI	PBM	CSP	CVNN	M2I
Iris (DS 1)	2	2	**3**	**3**	2	2	**3**
Magic (DS 2)	8	**2**	4	**2**	4	**2**	**2**
Breast cancer (DS 3)	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Glass (DS 4)	2	2	7	5	5	4	**6**
Wine (DS 5)	12	**3**	5	4	2	2	**3**
Car (DS 6)	2	2	7	**4**	**4**	**4**	**4**
Control (DS 7)	3	3	**6**	**6**	3	3	5
Page blocks (DS 8)	2	2	**5**	**5**	2	3	**5**
Parkinson (DS 9)	**2**	**2**	**2**	**2**	14	**2**	**2**

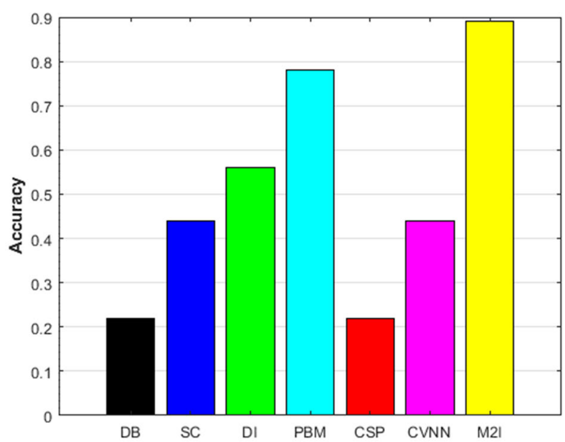

FIGURE 15. Indexes accuracy results.

The M2I accuracy is 89% in the real dataset as presented in Fig. 15. The accuracy of DB, SC, DI, PBM, CSP, and CVNN is 22%, 44%, 56%, 78%, 22%, and 44% respectively.

The proposed index in this paper has improvements compared with the most popular indexes (PBM, DI, SC, and DB):

- 12% improvements compared with PBM
- 37% improvements compared with DI
- 50% improvements compared with SC
- 75% improvements compared with DB

VIII. CONCLUSION

The clustering evaluation index is important to evaluate the clustering algorithm and select the best algorithm, which improves the clustering process and enhances the clustering results. The clustering error rate should be minimum and the nodes will be distributed in a correct cluster. This paper proposed M2I to test and evaluate the clustering algorithm. M2I tries to get minimum intra-cluster distance and maximum inter-cluster distance. Our proposed index has effective results and evaluates the clustering algorithm internally and externally. In this paper, a comparison of the proposed index with the existing indexes was made. The simulation results show that M2I has high correct value clustering results among the others. The proposed index focuses on two ways: firstly, computing the intra-cluster distance and the inter-cluster distance. Secondly, the ratio of the internal and external distance was calculated to ensure that the index value was more efficient and give a high explanation of the evaluation. The proposed index is the new enhancement of the PBM index and increases the accuracy of the index. The accuracy has been increased because the M2I index computes the ratio of all nodes' inter-distance values. A VANET scenario was implemented to test the proposed index and real datasets were imported. The proposed index gave the same results compared with the other indexes and the accuracy of M2I is 100% in the VANET scenario. In real datasets, there are 6 indexes compared with M2I over 9 actual datasets in different applications. The final M2I accuracy is 89% and has good improvements compared with others. The proposed index has 12% improvements compared with PBM which represents a more similar index. The proposed index has 37%, 50%, and 75% improvements compared with DI, SC, and DB respectively. M2I can also be used to evaluate any clustering model in data mining, image processing, energy clustering in wireless sensor networks, or ad hoc network. Our suggestion is to improve the k-means algorithm by using the M2I model as a K-value calculation. A new clustering model with optimal K value selection as future work. The k-means clustering algorithm will be more efficient with a minimum error rate and high cluster separation.

REFERENCES

[1] W. Liang, Y. Zhang, J. Xu, and D. Lin, "Optimization of basic clustering for ensemble clustering: An information-theoretic perspective," *IEEE Access*, vol. 7, pp. 179048–179062, 2019, doi: 10.1109/ACCESS.2019.2950159.

[2] J. Wang, C. Zhu, Y. Zhou, X. Zhu, Y. Wang, and W. Zhang, "From partition-based clustering to density-based clustering: Fast find clusters with diverse shapes and densities in spatial databases," *IEEE Access*, vol. 6, pp. 1718–1729, 2018, doi: 10.1109/ACCESS.2017.2780109.

[3] F. Li, M. Zhou, S. Li, and T. Yang, "A new density peak clustering algorithm based on cluster fusion strategy," *IEEE Access*, vol. 10, pp. 98034–98047, 2022, doi: 10.1109/ACCESS.2022.3205742.

[4] R. Xu and D. WunschII, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005, doi: 10.1109/TNN.2005.845141.

[5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, 1967, pp. 281–297. [Online]. Available: http://books.google.de/books?hl=de&lr=&id=IC4Ku_7dBFUC&oi=fnd&pg=PA281&dq=MacQueen+some+methods+for+classification&ots=nNTcK1IdoQ&sig=fHzdVcbvmYJ-lTNHu1HncmOFOkM#v=onepage&q=MacQueensomemethodsforclassification&f=false

[6] M. Bateni, H. Esfandiari, R. Jayaram, and V. Mirrokni, "Optimal fully dynamic k-centers clustering," 2021, *arXiv:2112.07050*.

[7] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014, doi: 10.1126/science.1242072.

[8] Y. Wang and L. Chen, "K-MEAP: Multiple exemplars affinity propagation with specified K clusters," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2670–2682, Dec. 2016, doi: 10.1109/TNNLS.2015.2495268.

[9] H. Jinlong, Z. Qingsheng, Y. Lijun, and C. Dongdong, "QCC: A novel cluster algorithm based on quasi-cluster centers," in *Proc. Int. Conf. Softw. Eng. Knowl. Eng.*, Jul. 2016, pp. 498–503, doi: 10.18293/SEKE2016-028.

[10] D. Cheng, Q. Zhu, J. Huang, L. Yang, and Q. Wu, "Natural neighbor-based clustering algorithm with local representatives," *Knowl.-Based Syst.*, vol. 123, pp. 238–253, May 2017, doi: 10.1016/j.knosys.2017.02.027.

[11] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007, doi: 10.1126/science.1136800.

[12] M. Daszykowski and B. Walczak, "Density-based clustering methods," *Comprehensive Chemometrics*, vol. 2, pp. 635–654, Jan. 2009, doi: 10.1016/B978-044452701-1.00067-3.

[13] I. Gurrutxaga, I. Albisua, O. Arbelaitz, J. I. Martín, J. Muguerza, J. M. Pérez, and I. Perona, "SEP/COP: An efficient method to find the best partition in hierarchical clustering based on a new cluster validity index," *Pattern Recognit.*, vol. 43, no. 10, pp. 3364–3373, Oct. 2010, doi: 10.1016/j.patcog.2010.04.021.

[14] I. Cattinelli, G. Valentini, E. Paulesu, and N. A. Borghese, "A novel approach to the problem of non-uniqueness of the solution in hierarchical clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1166–1173, Jul. 2013, doi: 10.1109/TNNLS.2013.2247058.

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

[16] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.

[17] T. Caliński and J. Harabasz, "Communications in statistics–theory and methods," *Commun. Statist.*, vol. 3, no. 1, pp. 1–27, 1974.

[18] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[19] M. Halkidi, M. Vazirgiannis, and Y. Batistakis, "Quality scheme assessment in the clustering process," in *Principles of Data Mining and Knowledge Discovery* (Lecture Notes in Computer Science), vol. 1910. Berlin, Germany: Springer, 2000, pp. 265–276, doi: 10.1007/3-540-45372-5_26.

[20] Z. Markov, "MDL-based hierarchical clustering," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 471–474.

[21] S. Yue, J.-S. Wang, T. Wu, and H. Wang, "A new separation measure for improving the effectiveness of validity indices," *Inf. Sci.*, vol. 180, no. 5, pp. 748–764, Mar. 2010, doi: 10.1016/j.ins.2009.11.005.

[22] F. Informa, W. R. Number, M. House, M. Street, and J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," Tech. Rep., Jun. 2013, pp. 37–41.

[23] Y. Jung, H. Park, D.-Z. Du, and B. L. Drake, "A decision criterion for the optimal number of clusters in hierarchical clustering," *J. Global Optim.*, vol. 25, no. 1, pp. 91–111, 2003, doi: 10.1023/A:1021394316112.

[24] L.-F. Chen, "A hierarchical method for determining the number of clusters," *J. Softw.*, vol. 19, no. 1, pp. 62–72, Jun. 2008, doi: 10.3724/SP.J.1001.2008.00062.

[25] G. Guo, L. Chen, Y. Ye, and Q. Jiang, "Cluster validation method for determining the number of clusters in categorical sequences," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2936–2948, Dec. 2017, doi: 10.1109/TNNLS.2016.2608354.

[26] M. B. Gorzalczany and F. Rudzinski, "Generalized self-organizing maps for automatic determination of the number of clusters and their multiprototypes in cluster analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2833–2845, Jul. 2018, doi: 10.1109/TNNLS.2017.2704779.

[27] N. A. Yousri, M. S. Kamel, and M. A. Ismail, "A novel validity measure for clusters of arbitrary shapes and densities," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4, doi: 10.1109/icpr.2008.4761242.

[28] H. Xiong and Z. Li, "Clustering validation measures," *Data Clustering*, vol. 43, no. 3, pp. 571–606, 2019, doi: 10.1201/9781315373515-23.

[29] S. Zhou, Z. Xu, and F. Liu, "Method for determining the optimal number of clusters based on agglomerative hierarchical clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 3007–3017, Dec. 2017, doi: 10.1109/TNNLS.2016.2608001.

[30] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 3, pp. 487–501, Mar. 2004, doi: 10.1016/j.patcog.2003.06.005.

[31] M. Virvou, E. Alepis, and C. Troussas, "Centroid-based clustering for student models in computer-based multiple language tutoring," in *Proc. SIGMAP WINSYS Int. Conf. Signal Process. Multimedia Appl. Wireless Inf. Netw. Syst.*, Jan. 2012, pp. 198–203, doi: 10.5220/0004128201980203.

[32] H. N. Abdulrazzak, G. C. Hock, N. A. Mohamed Radzi, N. M. L. Tan, and C. F. Kwong, "Modeling and analysis of new hybrid clustering technique for vehicular ad hoc network," *Mathematics*, vol. 10, no. 24, p. 4720, Dec. 2022, doi: 10.3390/math10244720.
[33] S. Nawrin, M. Rahatur, and S. Akhter, "Exploreing K-means with internal validity indexes for data clustering in traffic management system," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, pp. 1–9, 2017, doi: 10.14569/ijacsa.2017.080337.
[34] C. He, S. Liu, and S. Han, "A fuzzy logic reinforcement learning-based routing algorithm for flying ad hoc networks," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2020, pp. 987–991, doi: 10.1109/ICNC47757.2020.9049705.
[35] J. Li, M. Wang, P. Zhu, D. Wang, and X. You, "Highly reliable fuzzy-logic-assisted AODV routing algorithm for mobile ad hoc networks," *Sensors*, vol. 21, no. 17, p. 5965, Sep. 2021, doi: 10.3390/s21175965.
[36] K. Heurtefeux and F. Valois, "Is RSSI a good choice for localization in wireless sensor network?" in *Proc. IEEE 26th Int. Conf. Adv. Inf. Netw. Appl.*, Mar. 2012, pp. 732–739, doi: 10.1109/AINA.2012.19.
[37] T. Kitasuka, K. Hisazumi, T. Nakanishi, and A. Fukuda, "Positioning technique of wireless LAN terminals using RSSI between terminals," in *Proc. Int. Conf. Pervasive Syst. Comput.*, May 2014, pp. 47–53, 2005.
[38] K. G. Lim, C. H. Lee, R. K. Y. Chin, K. B. Yeo, and K. T. K. Teo, "SUMO enhancement for vehicular ad hoc network (VANET) simulation," in *Proc. IEEE 2nd Int. Conf. Autom. Control Intell. Syst. (I2CACIS)*, Oct. 2017, pp. 86–91, doi: 10.1109/I2CACIS.2017.8239038.
[39] G. Erdağı, M. A. Silgu, and H. B. Çelikoğlu, "Emission effects of cooperative adaptive cruise control: A simulation case using sumo," *EPiC Ser. Comput.*, vol. 62, pp. 92–100, Jan. 2019, doi: 10.29007/fbb7.
[40] H. N. Abdulrazzak, N. M. L. Tan, and N. A. Mohd. Radzi, "Minimizing energy consumption in roadside unit of zigzag distribution based on RS-LS technique," in *Proc. IEEE Int. Conf. Autom. Control Intell. Syst. (I2CACIS)*, Jun. 2021, pp. 169–173, doi: 10.1109/I2CACIS52118.2021.9495853.
[41] *UCI Machine Learning Repository*. Accessed: May 2, 2023. [Online]. Available: http://archive.ics.uci.edu/ml
[42] Q. Li, S. Yue, Y. Wang, M. Ding, and J. Li, "A new cluster validity index based on the adjustment of within-cluster distance," *IEEE Access*, vol. 8, pp. 202872–202885, 2020, doi: 10.1109/ACCESS.2020.3036074.
[43] J. Xie, Z.-Y. Xiong, Q.-Z. Dai, X.-X. Wang, and Y.-F. Zhang, "A new internal index based on density core for clustering validation," *Inf. Sci.*, vol. 506, pp. 346–365, Jan. 2020, doi: 10.1016/j.ins.2019.08.029.
[44] L. Sun, C. Guo, C. Liu, and H. Xiong, "Fast affinity propagation clustering based on incomplete similarity matrix," *Knowl. Inf. Syst.*, vol. 51, no. 3, pp. 941–963, Jun. 2017, doi: 10.1007/s10115-016-0996-y.

HAZEM NOORI ABDULRAZZAK received the B.Sc. degree in computer engineering from the University of Basrah, Basrah, Iraq, in 2003, and the M.Tech. degree in IT from GGSIP University, Delhi, India, in 2008. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronics Engineering, College of Engineering, Universiti Tenaga Nasional, Kuala Lumpur, Malaysia. He has more than 14 years of experience in the academic field in Iraq. His research interests include ad-hoc networks, WSNs, intelligent systems, and wireless communication systems.

GOH CHIN HOCK received the bachelor's degree in electrical and electronics engineering and the master's and Ph.D. degrees in electrical engineering from Universiti Tenaga Nasional, Malaysia, in 2004, 2008, and 2012, respectively. He is currently a Senior Lecturer with the Department of Electrical and Electronics Engineering, Universiti Tenaga Nasional. He is a Professional Engineer of BEM, an Energy Professional of MEPA, and a Corporate Member of IEM. He is also a Certified Energy Manager (CEM) and a Certified Professional in Measurement and Verification (CPMV). His research interests include advanced material and energy, electromagnetic technology, and smart cities solution.

NURUL ASYIKIN MOHAMED RADZI (Senior Member, IEEE) received the B.E.E.E. (Hons.) and M.E.E. degrees and the Ph.D. degree in engineering from Universiti Tenaga Nasional in the 2008, 2010, and 2013, respectively. She is currently a Senior Lecturer with the Department of Electrical Electronics Engineering, Universiti Tenaga Nasional. Her research interests include optical communication and quality of service. She has contributed 60 technical articles in various journals and conferences. She is a Professional Engineer and a Chartered Engineer for IET.

NADIA MEI LIN TAN (Senior Member, IEEE) was born in Kuala Lumpur, Malaysia. She received the B.Eng. degree (Hons.) from the University of Sheffield, Sheffield, U.K., in 2002, the M.Eng. degree from Universiti Tenaga Nasional, Kajang, Malaysia, in 2007, and the Ph.D. degree from the Tokyo Institute of Technology, Tokyo, Japan, in 2010, all in electrical engineering.

She is currently an Associate Professor with the Department of Electrical and Electronics Engineering, University of Nottingham Ningbo China, China. She was an Associate Professor with the Department of Electrical and Electronics Engineering, Universiti Tenaga Nasional, Kajang, Malaysia, from 2017 to 2021. From 2016 to 2017, she was a Visiting Researcher with the Power Electronics Laboratory, Tokyo Institute of Technology, where she was an Adjunct Associate Professor in 2016. Her research interests include power conversion systems for renewable energy and battery energy storage applications and bidirectional isolated dc–dc converters.

Dr. Tan was a recipient of the 2018 IET Young Engineer Award and the 2018 IPCC IEEE IAS Prize Paper Award. She served as the Vice-Chair of the IEEE Power Electronics Society Malaysia Chapter from 2018 to 2019. She is a Chartered Engineer registered with the Engineering Council, U.K., a Professional Engineer registered with the Board of Engineers Malaysia, and a member of the Institution of Engineering and Technology and the Institution of Engineers Malaysia.