

Received 18 October 2023, accepted 6 November 2023, date of publication 16 November 2023,  
date of current version 4 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3333753

## RESEARCH ARTICLE

# Clustering of Cities Based on Their Smart Performances: A Comparative Approach of Fuzzy C-Means, K-Means, and K-Medoids

ÖMER N. KENGER<sup>1</sup>, ZÜLAL DİRİ KENGER<sup>1</sup>, EREN ÖZCEYLAN<sup>2</sup>, AND BEATA MRUGALSKA<sup>3</sup> 

<sup>1</sup>Department of Industrial Engineering, Hasan Kalyoncu University, 27010 Gaziantep, Turkey

<sup>2</sup>Department of Industrial Engineering, Gaziantep University, 27310 Gaziantep, Turkey

<sup>3</sup>Faculty of Engineering Management, Poznań University of Technology, 60-965 Poznań, Poland

Corresponding author: Beata Mrugalska (beata.mrugalska@put.poznan.pl)

This work was supported by the Poznan University of Technology under Grant 0811/SBAD/1071.

**ABSTRACT** Smart City is recognized as a potential approach to address serious urban issues such as traffic, pollution, energy use, and waste management. Therefore, it is vital to evaluate how smart cities are in order to put these methods into practice. To offer advice on these matters, numerous reports are created, and one of which is Smart City Index (SCI). The Institute for Management Development (IMD) and the Singapore University of Technology and Design collaborate on SCI every year (SUTD). The report evaluates how locals view the buildings and technological applications that are available in their towns. Although the study offers a thorough examination of the cities for evaluation, the city clusters should be more sensitive and not be created using strict clustering techniques. In order to address this problem, the clustering algorithms namely K-Medoids, Fuzzy C-Means, and K-Means, which outperform hard clustering approaches in terms of robustness to vagueness and knowledge retention, are used. The main goal of this study is to categorize cities using a scientific manner (clustering algorithms) based on SCI data and to present how the chosen approaches work for dealing with the associated problems. The primary innovation of the present study is the use of clustering techniques in reports where the indexes are used. The results indicate that grouping the cities on the basis of their smart indicators would not be as effective as using the three clustering techniques that are suggested in this paper. These results add to the analysis of the dynamic capacities of smart cities and highlight the sustainability of these tactics.


**INDEX TERMS** Comparative analysis, clustering, fuzzy C-Means, K-Means, K-Medoids, smart cities.

## I. INTRODUCTION

In 2021, 56% of people lived in cities, and by 2050, 68% are expected to be urban population according to the World Bank's World Development Indicators (WDI) database [1], [2]. There are actually relatively few nations where it is anticipated that rural shares would surpass urban shares by 2050. Although urbanization has positive impacts (welfare, economic growth, better education, increased healthcare services, etc.) on society [3], [4], rapid urbanization rates have been perceived in many nations as a treat to urban life such as a lack of energy, contamination of the environment, and

traffic congestion, social inequality, lack of public services, and land loss [5], [6], [7]. In order to overcome these negative effects, cities need to be more organized and managed in an efficient and effective way [8].

The "smart city" idea has been developed as a new technology-driven mechanism in the search for answers to these issues. Although smart cities are a very popular term that has been heard and used for years, there is no clear specific definition. According to Almirall et al. [9], transportation, civic entrepreneurship, democratic transparency, sustainable energy, and service delivery are all components of the concept of "Smart Cities", which covers the majority of the functions performed by local governments. Information and communication technology (ICT) are used as

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

a transformative tool to make different regions “smart,” which becomes one area of commonality. In other words, a smart city is a new trend in urban planning that integrates resources to deliver better urban services based on ICT use. These solutions have a high potential for success while are also viable and environmentally benign [10]. Following the definition of a smart city proposed by the European Commission a smart city is a place where traditional networks and services are made more efficient with the use of digital solutions for the benefit of its inhabitants and business [11]. According to smart cities report issued by the World Bank, it is a city that leverages the latest in technology and connectivity to make better decisions and achieve the urban aspirations of its residents [12]. As it can be noticed both of them refer to digitalization towards better life of its inhabitants.

The development of sustainable smart cities faces a range of multifaceted challenges. These encompass financial limitations that impede technology deployment and infrastructure creation, coupled with scarcities in essential resources such as technology, equipment, skilled personnel, and financial backing. Furthermore, resident unawareness of smart city services, compounded by social acceptability issues towards unfamiliar technologies, can hinder successful adoption. Institutions grapple with capacity shortages, hindering the effective management and maintenance of smart city technologies. Moreover, the absence of operational frameworks, standardization disputes, and gaps in strategic planning and governance coordination can thwart cohesive development efforts. These challenges can be addressed through strategies such as fostering partnerships to increase budgets, launching comprehensive public awareness campaigns, facilitating stakeholder consensus-building for unified standards, integrated strategic planning to ensure phased development, boosting capacity through targeted training, establishing streamlined governance structures, implementing standardized system protocols, promoting collaborative content creation, and integrating sustainability principles into all facets of smart city initiatives. By embracing these solutions, cities can navigate these complexities and cultivate more efficient, inclusive, and enduring smart city environments [13].

The issue of delineating and quantifying smart city performance has been always a primary concern for policy developers/city planners, and numerous efforts including indexes have been presented (e.g., [14], [15], [16], [17], [18], [19]). These indices provide the data required to characterize a city's or nation's smart situation in support of public policies, regional, national, and international initiatives and strategies. One of the reports called as SCI is published by combined efforts of the IMD and the SUTD annually [11]. The SCI is the pioneer global index that rates metropolitan regions on the basis of the opinions of their citizens who determine how “smart” a city is. It assesses how locals see the scope and results of urban intelligence initiatives. On the basis of technological, economic, and public assessments of

how “smart” their cities are, cities are ranked according to the SCI. There are three reports published between 2019 and 2021. They were carried out in 118 metropolises in five major dimensions: health and safety, activities, governance, opportunities, and mobility [20]. When the SCI is established, countries are grouped together to find collaborative solutions to the problems. Indexes of this nature serve as helpful resources for overcoming challenges, providing insights into areas that require improvement and enable to concert efforts to address deficiencies.

Finding the differences and similarities among cities are fundamental in determining which cities are suitable/unsuitable to create and put into effect a policy for smartness. Although the aforementioned SCI report provides an extensive analysis to assess the cities, cities' clusters ought to be more delicate and should not be determined using strict clustering techniques. If there are cities that are not clustered correctly, it will be difficult to create road maps for smart urbanization. One of the multivariate statistical techniques, clustering analysis or simply clustering, helps to break up big data sets into subgroups based on similarities and examines various correlations and patterns [21]. The aim of this approach is to partition the data set into as many inputs given to the identical cluster as doable, while avoiding assigning as many items to separate clusters as possible.

In a case of our study we employed soft computing technique (Fuzzy C-Means) in addition to the hard computing techniques (K-Means and K-Medoids). Therefore, this paper aims to frame how cities can be clustered according to their smartness indicators. Its primary goal is to categorize cities using a more scientific approach based on the SCI, and to demonstrate the effectiveness of the selected methods in addressing such issues. Furthermore, our study also utilized soft computing (fuzzy C-means), whereas the SCI relied on hard computing techniques. As the clustering of groups (cities) influences the smart city policy makers' efforts, the research questions considered in this paper are: (i) is it possible to apply soft and hard computing approaches to classify such indexes, (ii) which cities are included by which clusters according to which similarities, (iii) which clustering method provides a better result, and finally (iv) what differences will be identified in comparison to the existing clusters presented by the SCI report?

The structure of the study is as follows. A synopsis of the pertinent literature is provided in Section II. The used clustering approaches are discussed in Section III. The SCI report is given in Section IV along with a clustering of the SCI report's included nations using the suggested methods. The findings are displayed and discussed. Conclusions and limitations of this strategy are discussed in Section V.

## II. LITERATURE REVIEW

This chapter conducts two levels of literary analysis: i) reviews that analyze smart cities in terms of data analytics,

*ii)* current studies that use K-Medoids, Fuzzy C-Means, and K-Means clustering algorithms.

Moustaka et al. [22] applied a systematic review for Smart Cities (SC) by focusing on data harvesting (DH) and data mining (DM) processes. According to their survey findings, “IoT” and “smart mobility” are the themes that researchers are most interested in, although “crowd-sensing” and “smart living” are also hot topics. In addition to the conventional DM approaches, the methods that are most frequently utilized for the harvesting and mining of urban data, are statistics, fuzzy logic, machine learning, visualization, neural networks, web mining, and text mining. Transportation, security/emergency, health, and environmental services are the most popular applications of urban data in smart cities, per the research’s findings. Finally, their research findings revealed that while smart applications collect and depict open and massive data created by sensors or users, applications that evaluate human behavior for multiple objectives are favored. Soomro et al. [23] reviewed the SC domain in reference to big data analytics. Moreover, Souza et al. [24] analyzed the DM and machine learning methods regarding SC domain as well as conducted a network analysis using co-citations from the journal, co-citations from the authors, and co-occurrence of keywords to identify potential future research trends. Predictive analytics, which is mostly employed in the fields of smart mobility and the smart environment, is the technique that is used the most frequently, according to their evaluation results. Ageed et al. [25] investigated the big data usage in order to create more innovative societies. They also examined the advantages and disadvantages of applying big data systems in smart cities. In the study [26] the authors conducted an extensive analysis to determine the primary DM methods applied in SC. Their results revealed classification algorithms such as decision trees, random forest, support vector machine, Bayesian network, K-nearest neighbors, neural network, deep learning and reinforcement learning, are frequently used in smart cities. In addition, the most frequent words and also trend research topics are IoT and big data. Sarker [27] investigated the contributions of SC data science to decision-making in SC systems and services. In order to achieve it, ten application domains were identified such as public safety, transportation, education, healthcare and systems, and services were analyzed in details in these domains.

The clustering methods K-Means, Fuzzy C-Means, and K-Medoids were often employed in a variety of applications. For example, K-Means, K-Medoids, and Fuzzy C-Means algorithms were used for nuclear medicine image segmentation [28]. A density Fuzzy C-Means algorithm was suggested for segmenting brain MR images [29]. For the purpose of detecting brain tumors, the watershed algorithm with the Fuzzy K-Means and Fuzzy C-means algorithms was integrated [30]. Fuzzy C-Means, K-Medoids, and K-Means clustering algorithms were also utilized for cancer prediction and detection [31], [32], [33], [34], [35]. The aforementioned algorithms were also used to cluster the spread of the COVID-19 pandemic, one of the most serious public health problems

that humanity has ever experienced, and to determine the current state of nations [36], [37], [38].

Several studies have also conducted comparative analyses of these three clustering algorithms. Park and H. Jun developed a novel K-Medoids clustering approach that was simpler and faster than existing methods. The outcomes demonstrate that the suggested method succeeds similar or better clustering quality than existing methods while being significantly faster [39]. Furthermore, Jipkate and Gohokar conducted a comparison for the Fuzzy C-Means and K-Means clustering algorithms and found that K-Means were outperformed by Fuzzy C-Means in terms of accuracy and speed [40]. Surya Narayana and Vasumathi proposed new attributes similarity-based K-Medoids clustering technique that achieved better clustering results than traditional K-Medoids algorithm [41]. In their study, Kurniawan et al. also compared K-Means and Fuzzy C-Means with another clustering algorithm, linkage, for the NASA active fire dataset and found that Fuzzy C-Means produced better clustering results [42]. On the other hand, Zhou and Yang investigated the effect of cluster size distribution on clustering and compared K-Means and Fuzzy C-Means algorithms [43].

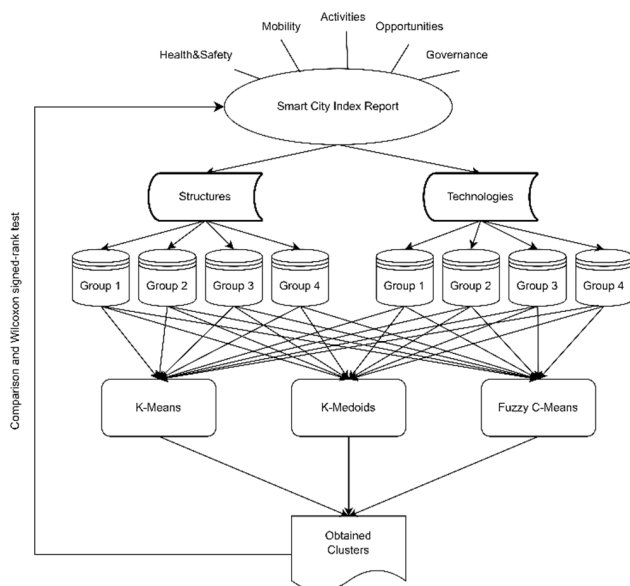
Among the clustering algorithms that are frequently utilized in the context of smart cities are K-Medoids, Fuzzy C-Means, and K-Means [44]. K-Means is a highly well-liked clustering technique in data-driven smart cities because it can scale enormous data sets [26], [27]. Its computational speed and low complexity make it another advantageous algorithm. In order to identify anomalies or cyber dangers in the SC data, the K-Medoids algorithm can be utilized to analyze citizen preferences, behavioral activities, or usage [27], [45]. In terms of being able to provide the probability of belonging to a cluster, the Fuzzy C-Means approach is more realistic [46]. Clustering methods were also been used in intelligent transportation systems for better traffic management [47], [48], [49], [50], [51], wireless networks [52], [53], detection of congestion [54], climate change [21], energy [55], [56], etc.

Considering the literature review presented above, the three most basic contributions of the proposed paper to the relevant literature are: *(i)* This paper presents a methodological framework for clustering cities on the basis of their smartness indicators. *(ii)* The study introduces a scientifically robust approach that employs the Smart City Index to categorize cities. *(iii)* The effectiveness of the selected methods in addressing smart city challenges is empirically demonstrated.

### III. METHODOLOGY

In smart cities, ICT is utilized to boost operational efficiency, enlighten the public, and enhance the standard of public services and citizen welfare. A smart city uses cutting-edge technology and data analysis to improve citizen quality of life, streamline municipal processes, and foster economic development. While countries and cities try to make their communities smart, they need to determine the policies they

will implement accordingly. Some international organizations have developed indexes/reports to determine which policy should be applied or the level of smartness of each community. One of the mentioned reports, and perhaps the most important one, is the SCI, and according to this index, cities are grouped according to a number of criteria. However, hard computations approaches such as simple arithmetic operations are mostly used in such reports. Therefore, in some of the results obtained, grey areas/cities (the places that are left out) may not be grouped correctly. Fuzzy C-Means, K-Medoids, and K-Means are three popular clustering algorithms used in data ensembles with too many features. Each algorithm has strengths and weaknesses, and the choice of the algorithm depends on the specific characteristics of the data and the requirements of the application. Clustering is a fundamental task in machine learning and data analysis. Among the various clustering algorithms available, K-Means, K-Medoids, and Fuzzy C-Means are widely used due to their effectiveness and ease of implementation. This paper aims to apply and compare the Fuzzy C-Means, K-Means, and K-Medoids clustering methods to determine the best method for grouping cities based on the smart city indicators mentioned in the SCI report.



**FIGURE 1.** The flowchart of the applied methodology.

The SCI report assesses individuals' perceptions of two crucial aspects: the urban environment and its technological integration. It employs five core categories—health and safety, mobility, recreational options, opportunities, and governance—to assess each major facet. The cities are categorized into four groups (1-2-3-4) based on their UN Human Development Index score for the economy. In order to perform clustering we used the entire dataset and then compared the outcomes with the present index. In addition, the Wilcoxon Signed Rank statistical test is employed to determine the effectiveness of different clustering approaches.

The flowchart of the applied methodology is presented in Figure 1.

Fuzzy C-Means is a clustering technique that assigns weights to each member of the set, while K-Means is a simple unsupervised learning algorithm that addresses the well-known clustering problem. The K-Means procedure involves classifying a given dataset into a predetermined number of clusters (assumed to be  $k$  clusters). On the other hand, K-Medoids is a clustering technique that can handle objects with high values that may deviate from their distribution [57]. The Fuzzy C-Means clustering algorithm is a useful tool for clustering datasets due to several advantages it offers. Firstly, unlike traditional clustering algorithms that assign data points to only one cluster, Fuzzy C-Means allows data points to belong to multiple clusters. This feature is particularly beneficial when the data suggests overlapping clusters or when multiple interpretations of the same data are possible. Secondly, Fuzzy C-Means is easy to implement and does not require extensive computational resources. This makes it suitable for large datasets where other algorithms may struggle due to limited computing power. Moreover, Fuzzy C-Means is relatively fast and efficient, making it an excellent choice for time-sensitive projects. Furthermore, the algorithm generates a membership value for each data point, indicating the degree to which it belongs to each cluster. It proves useful for various types of analysis. Additionally, Fuzzy C-Means demonstrates robustness against noise and outliers in the data. It can effectively handle non-linearly shaped clusters, thereby establishing itself as a versatile tool for clustering datasets [58], [59]. K-Means clustering is a popular unsupervised machine learning algorithm known for its speed, efficiency, and simplicity. One of its main strengths is its ability to handle large datasets quickly and efficiently. This algorithm is highly flexible and can be applied to various types of data, including numerical, categorical, and binary data. K-Means clustering is also remarkably easy to implement, making it accessible to data analysts and researchers. Its simplicity makes it an excellent starting point for clustering tasks and provides insights into the structure and patterns within a dataset. Overall, the K-Means clustering algorithm is a valuable tool for data analysis, particularly for tasks that require fast and efficient processing of large datasets [60]. The K-Medoids clustering algorithm is a variation of K-Means clustering that offers several advantages over its counterpart. K-Medoids is less sensitive to noise and outliers in the dataset, making it a better choice for datasets with high variability. Additionally, it is capable of handling categorical data, unlike K-Means, which requires numerical data. Another advantage of K-Medoids is its low time complexity, making it more efficient for large datasets. The use of medoids as cluster representatives, instead of the mean value used in K-Means, makes K-Medoids clusters easier to interpret since medoids correspond to actual data points within the cluster. Overall, K-Medoids is a valuable clustering algorithm that offers advantages in terms of interpretability, efficiency, and its ability to handle noisy and categorical data [39], [41].



This section presents the standard Fuzzy C-Means, K-Means, and K-Medoids clustering algorithms, respectively.

#### A. FUZZY C-MEANS CLUSTERING

Hard clustering divides data into discrete clusters, assigning every element of the data to only one cluster. Data elements in fuzzy clustering, also known as soft clustering, can belong to many clusters, and each element has a set of membership levels associated with it. They show the strength of the connection between data element and a specific cluster. These membership levels are established in fuzzy clustering, after which data components are assigned to one or more clusters [40]. Fuzzy C-Means is satisfied as Eq. (1) [42].

$$J_m = \sum_{i=1}^D \sum_{j=1}^N \mu_{ij}^m \|x_i - c_j\|^2 \quad (1)$$

$x_i$  is the  $i$ th data point,  $c_j$  is the center of the  $j$ th cluster, and  $\mu_{ij}$  is the degree of membership of  $x_i$  in the  $j$ th cluster, where  $D$  is the number of data points,  $N$  is the number of clusters,  $m$  variable is a fuzzy partition matrix exponent for determining the degree of fuzzy overlap, and  $m > 1$ . The following lists the steps of the fuzzy C-Means clustering algorithm:

Step 1. Set  $N$ ,  $m$  and  $\varepsilon$

Step 2. Randomly initialize the fuzzy partition matrix  $\mu^{(0)}$ :

Step 3. Calculate the cluster centers by using Eq. (2).

$$c_j = \frac{\sum_{i=1}^D \mu_{ij}^m x_i}{\sum_{i=1}^D \mu_{ij}^m} \quad (2)$$

Step 4. Update the membership matrix through Eq. (3).

$$\mu_{ij} = \frac{1}{\sum_{k=1}^N \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (3)$$

Step 5. Determine  $J_m$ , the objective function

Step 6. Continue to perform steps 3–5 until  $J_m$  improves by less than a predetermined minimum threshold or until a predetermined maximum number of iterations is completed.

Fuzzy C-Means is a clustering algorithm that is used to partition data into clusters and has several strengths that make it an attractive option for certain types of data. Some of the main strengths of FCM are [58], [59]:

- i. Instead of only allowing data points to be a part of one cluster, Fuzzy C-Means allows for them to belong to multiple clusters, which can better reflect the true nature of the data.
- ii. It is relatively simple and easy to implement, and does not require a lot of computational resources.
- iii. It is relatively fast and efficient, especially for large datasets.
- iv. It produces a membership value for each data point, indicating the degree to which it belongs to each cluster, which can be useful for certain types of analysis.
- v. It is robust to noise and outliers in the data.
- vi. It can handle non-linearly shaped clusters.

Overall, Fuzzy C-Means is a useful clustering algorithm that can be applied to a variety of data types and is especially useful for data that does not clearly fit into a single cluster.

For example, Andrejovska analyzed data from 2012 to determine the insolvency of EU countries using several different methodological approaches, including variants of agglomerative hierarchical cluster analysis, the outputs of K-Means and K-Medoids, and Fuzzy C-Means [61]. Moreover, Anggoro et al. clustered the cities/districts in eastern Indonesia through the Fuzzy C-Means algorithm according to ICT vulnerability [62]. The use of the Fuzzy C-Means clustering method to analyze and understand the epidemiological situation of COVID-19 in European countries, and to track changes in this situation over time is discussed in [63].

#### B. K-MEANS CLUSTERING

K-Means clustering, often known as tough Unsupervised and non-hierarchical partitioning techniques for data analysis includes C-Means clustering. K-Means clustering divides data points into a predetermined number of clusters. Based on the locations and separations between different input data points, the approach groups the data [64]. K-Means is mostly used to maximize inter-cluster distance and minimize intra-cluster distance. It computes the objective function using Eq. (4) [43].

$$SSE = \min \sum_{i=1}^c \sum_{x_j \in v_i} \|x_j - v_i\|^2 \quad (4)$$

where  $V_i$  is the  $i$ th cluster and  $x_j$  is the  $j$ th data object. The steps of the K-Means clustering algorithm are listed below [40]:

Step 1. Choose the number  $K$  of desired clusters.

Step 2. Create  $K$  clusters and determine the cluster's centroid.

Step 3. Assign each point to the cluster whose centroid is nearest to it.

Step 4. Recalculate the new  $K$  centroids.

Step 5. Repeat steps 3 and 4 until no point changes its cluster assignment or until the centroids no longer change.

Some of the strengths of K-Means clustering algorithm are given as follows [60]:

- i. It is fast and efficient in terms of computational time.
- ii. It is highly flexible and can be used for a variety of data types
- iii. It is simple and easy to implement
- iv. It can handle large data sets

Latifah et al. [65] used the K-Means and Fuzzy C-Means algorithms to cluster regencies/cities in Central Java Province on the basis of the Human Development Index (HDI). It is an indicator used by the United Nations Development Programme (UNDP) to measure and compare the development levels of different countries and it assesses human development based on three key dimensions: life expectancy (health), education (access to knowledge), and per capita income (standard of living). In [66] there are presented the application and comparisons of the K-Medoids, K-Means, and Fuzzy C-Means algorithms to cluster provinces according to stunting which is widely known as the case of malnutrition suffered by toddlers.

**TABLE 1.** Factor ratings employed in the IMD-SUTD SCI.

STRUCTURES	TECHNOLOGIES
<i>Health &amp; Safety</i>	<i>Health &amp; Safety</i>
Basic sanitation meets the needs of the poorest areas	Online reporting of city maintenance problems provides a speedy solution
Recycling services are satisfactory	A website or App allows to give away unwanted items to other city residents
Public safety is not a problem	Free public wifi improves access to services
Air pollution is not a problem	CCTV cameras make residents feel safer
Medical services provision is satisfactory	A website or App allows effective monitoring of air pollution
	Arranging medical appointments online improves access
<i>Mobility</i>	<i>Mobility</i>
Traffic congestion is not a problem	Car-sharing Apps reduces congestion
Public transport is satisfactory	Apps that direct you to an available parking space reduce journey time
	Bicycle hiring reduces congestion
	Online scheduling and ticket sales make public transport easier to use
<i>Activities</i>	<i>Activities</i>
Green spaces are satisfactory	Online purchasing of tickets to shows and museums makes it easier to attend
Cultural activities (shows, bars, and museums) are satisfactory	
<i>Opportunities (Work &amp; School)</i>	<i>Opportunities (Work &amp; School)</i>
Employment finding services are available	Online access to job listings makes it easier to find work
Most children have access to a good school	IT skills are taught well in schools
Lifelong learning opportunities are provided by local institutions	Online services provided by the city make it easier to start a new business
Businesses create new jobs	
Minorities feel welcome	
<i>Governance</i>	<i>Governance</i>
Information on local government decisions are easily accessible	Online public access to city finances reduces corruption
Corruption of city officials is not an issue of concern	Online voting increases participation
Residents contribute to decision making of local government	An online platform where residents can propose ideas improves city life
	Processing Identification
Residents provide feedback on local government projects	Documents online reduces waiting times

### C. K-MEDOIDS CLUSTERING

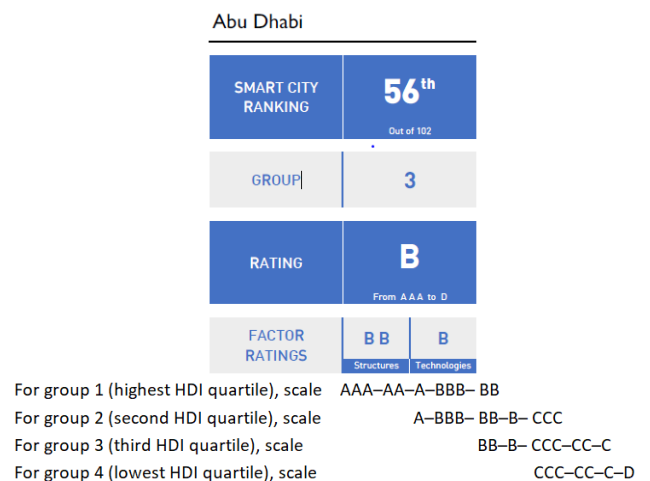
K-Medoids clustering is a method of partitioning a set of data points into k clusters, where each cluster is represented by a medoid [67]. A medoid is a data point within a cluster that is

representative of the other points in the cluster. The medoids are chosen such that the sum of the distances between each data point and its nearest medoid is minimized.

K-Medoids clustering is similar to K-Means clustering, which uses means as the cluster centers. However, K-Medoids is more robust to noise and outliers than K-Means, and is generally more suitable for dealing with non-linearly distributed data [39]. Some strengths of K-Medoids clustering include [39], [41]:

- It is less sensitive to noise and outliers.
- It can handle categorical data.
- K-Medoids has low time complexity, making it more efficient for large datasets.
- The use of medoids as cluster representatives makes K-Medoids clusters easier to interpret.

Özdemir and Kaya used both K-Medoids and Fuzzy C-Means algorithms to classify member countries of the organization for economic co-operation and development on the basis of their CO<sub>2</sub> emissions from fossil fuel consumption [21]. In [68] these methods are used to cluster districts in Riau Province which is one of the provinces that has social inequality in its society with a low welfare index.



**FIGURE 2.** An example of the index result for Abu Dhabi.

### IV. SMART CITY INDEX

The data is obtained from Smart City Index report which results from an annual collaboration between the Institute for Management Development (IMD) and the Singapore University of Technology and Design (SUTD). The IMD-SUTD SCI assesses residents' perceptions concerning the available structures and technology applications in their respective cities. The SCI ranks cities globally by collecting the perceptions of 120 residents from each city. The IMD-SUTD SCI rates people' impressions of two key factors: the city's physical layout and its use of technology. Five primary categories—health and safety, mobility, activities, opportunities, and governance—are used to evaluate each main title. The factor ratings are determined as it is shown in Table 1.

**TABLE 2.** Comparison for group 1 (structures).

HDI QUARTILE	CITIES	INDEX RESULT	K-MEANS RESULT	K-MEDIODS RESULT	FCM RESULT	BB CLUSTER MEMBERSHIP	BBB CLUSTER MEMBERSHIP	A CLUSTER MEMBERSHIP	AA CLUSTER MEMBERSHIP	AAA CLUSTER MEMBERSHIP
HDI 1.GROUP	Amsterdam	A	↑AA	↑AA	●A	0.0923	0.1476	0.4273	0.2697	0.0632
	Auckland	A	●A	↓BBB	↑AA	0.0841	0.1541	0.2099	0.3264	0.2256
	Berlin	BBB	↑AA	●BBB	●BBB	0.2569	0.3079	0.2374	0.1631	0.0347
	Birmingham	BB	↑BBB	●BB	●BB	0.5785	0.2113	0.1038	0.0766	0.0299
	Brisbane	A	●A	↓BB	↑AA	0.1275	0.2390	0.2609	0.2718	0.1008
	Brussels	BB	↑BBB	●BB	●BB	0.4743	0.2223	0.1494	0.1118	0.0422
	Copenhagen	AA	↑AAA	↑AAA	↑AAA	0.0526	0.0901	0.1712	0.2830	0.4032
	Dublin	BB	↑A	●BB	↑BBB	0.2822	0.3482	0.1670	0.1436	0.0589
	Dusseldorf	A	↑AA	↓BBB	↑AA	0.0463	0.0846	0.3031	0.5181	0.0479
	Geneva	AA	↑AAA	↑AAA	↑AAA	0.0819	0.1224	0.1907	0.2577	0.3472
	Glasgow	BB	↑BBB	●BB	●BB	0.5919	0.2147	0.1005	0.0691	0.0238
	Gothenburg	BBB	↑A	↓BB	●BBB	0.2001	0.3879	0.1899	0.1695	0.0526
	Hamburg	A	↑AA	↓BBB	↑AA	0.0643	0.1259	0.3482	0.4180	0.0435
	Hanover	A	↑AA	●A	●A	0.0991	0.1492	0.3548	0.3264	0.0706
	Helsinki	AA	↑AAA	↑AAA	↑AAA	0.0634	0.0923	0.1575	0.2050	0.4819
	Hong Kong	BB	●BB	●BB	●BB	0.3462	0.2433	0.1829	0.1499	0.0776
	Kiel	BBB	●BBB	↓BB	↓BB	0.5124	0.2573	0.1216	0.0843	0.0244
	Lausanne	AAA	●AAA	●AAA	●AAA	0.0476	0.0729	0.1196	0.1690	0.5909
	Leeds	BBB	●BBB	↓BB	↓BB	0.4686	0.2860	0.1307	0.0870	0.0278
	London	BBB	●BBB	↓BB	●BBB	0.3327	0.3369	0.1744	0.1195	0.0365
	Manchester	BBB	↑AA	↓BB	●BBB	0.2004	0.3526	0.2429	0.1616	0.0426
	Melbourne	BBB	↑A	↓BB	●BBB	0.1989	0.4304	0.1814	0.1463	0.0429
	Munich	AA	↑AAA	↓BBB	●AA	0.0670	0.1101	0.2248	0.3854	0.2127
	Newcastle	A	↑AA	●A	●A	0.1025	0.1639	0.3737	0.2838	0.0761
	Oslo	AAA	●AAA	●AAA	●AAA	0.0240	0.0357	0.0597	0.0857	0.7949
	Rotterdam	BBB	↑AA	↑A	↑A	0.2342	0.2383	0.3148	0.1675	0.0453
	Singapore	AAA	●AAA	●AAA	●AAA	0.0679	0.0910	0.1331	0.1615	0.5465
	Stockholm	A	●A	↓BB	↓BBB	0.1375	0.3128	0.2330	0.2441	0.0726
	Sydney	BBB	↑A	↓BB	●BBB	0.1650	0.4818	0.1903	0.1316	0.0313
	The Hague	A	↑AA	●A	●A	0.0549	0.1010	0.5715	0.2388	0.0338
	Zurich	AAA	●AAA	●AAA	●AAA	0.0482	0.0670	0.1035	0.1336	0.6476

**TABLE 10.** The comparative results of clustering methods in terms of their efficiency.

		Fuzzy C-Means				K-Means				K-Medoids	
		MSE	# equal cities	# unequal cities	MSE	# equal cities	# unequal cities	MSE	# equal cities	# unequal cities	
STRUCTURES	HDI 1.GROUP	0.3870	19	12	1.0323	11	20	0.8710	13	18	
	HDI 2.GROUP	0.6896	18	11	0.6207	17	12	0.6897	15	14	
	HDI 3.GROUP	0.2142	22	6	0.2143	22	6	0.7500	13	15	
	HDI 4.GROUP	0.1666	25	5	0.1667	25	5	0.4333	17	13	
TECHNOLOGIES	HDI 1.GROUP	0.8709	10	21	0.4839	16	15	1.6774	9	22	
	HDI 2.GROUP	0.6551	13	16	0.3448	19	10	2.0690	5	24	
	HDI 3.GROUP	0.6785	12	16	0.7500	10	18	1.1786	12	16	
	HDI 4.GROUP	0.1333	26	4	0.0333	29	1	0.1000	27	3	

An illustration of the index result for Abu Dhabi is shown in Figure 2. According to the Rating and its components, the city's place among all other cities is referred to as its "smart city rating." On the basis on the economy's UN Human Development Index score, the cities are ranked in four groups (1-2-3-4) (1-2-3-4). Abu Dhabi falls into "Group 3" here. Cities are awarded a "rating scale" within each group

(AAA to D). Rating of Abu Dhabi is "B". Additively, the factor ratings according to "structures" and "technologies" are "BB" and "B", respectively. Rating scale groups are presented as follows (Figure 2).

The SCI's first edition, which was published in 2019, examines 102 towns across the globe by incorporating the opinions of 120 locals in each location. The SCI's second

**TABLE 11.** Wilcoxon signed-rank test (K-Medoids – Fuzzy C-Means).

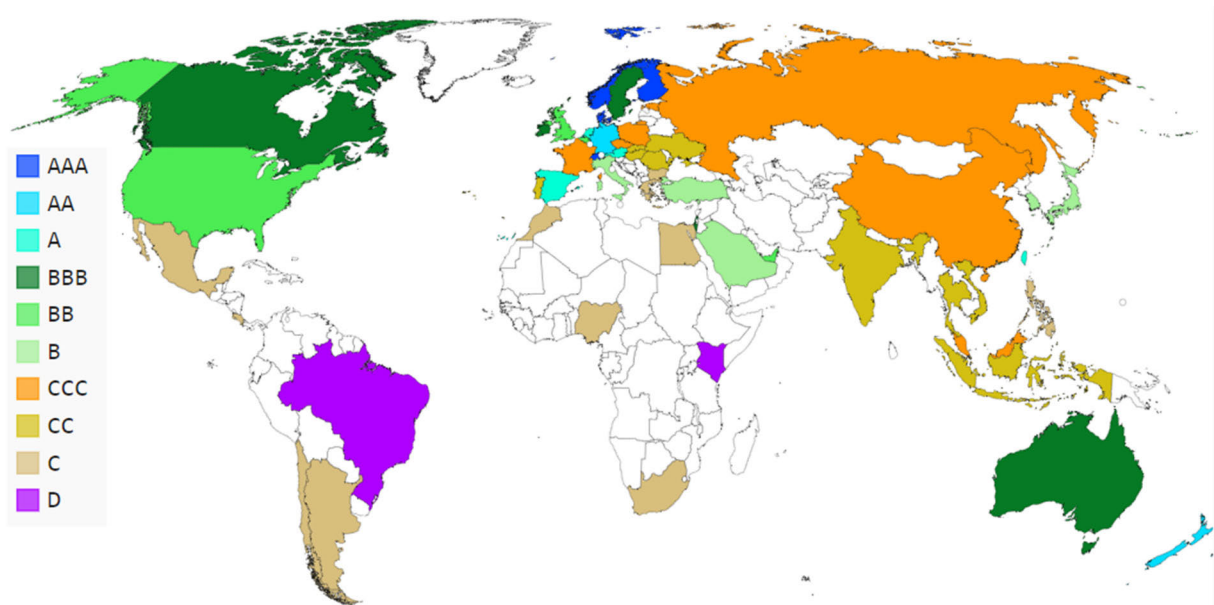
	K-Medoids MSE	Fuzzy C- Means MSE	Difference	Absolute Difference	Rank of Absolute Difference	Positive Ranks	Negative Ranks
STRUCTURES HDI 1.GROUP	0.8710	0.3870	0.4840	0.4840	4	4	-
STRUCTURES HDI 2.GROUP	0.6897	0.6896	0.0001	0.0001	1	1	-
STRUCTURES HDI 3.GROUP	0.7500	0.2142	0.5358	0.5358	6	6	-
STRUCTURES HDI 4.GROUP	0.4333	0.1666	0.2667	0.2667	3	3	-
TECHNOLOGIES HDI 1.GROUP	1.6774	0.8709	0.8065	0.8065	7	7	-
TECHNOLOGIES HDI 2.GROUP	2.0690	0.6551	1.4139	1.4139	8	8	-
TECHNOLOGIES HDI 3.GROUP	1.1786	0.6785	0.5001	0.5001	5	5	-
TECHNOLOGIES HDI 4.GROUP	0.1000	0.1333	-0.0333	0.0333	2	-	-2
TOTAL						34	-2

**TABLE 12.** Wilcoxon signed-rank test (K-Medoids – K Means).

	K-Medoids MSE	K-Means MSE	Difference	Absolute Difference	Rank of Absolute Difference	Positive Ranks	Negative Ranks
STRUCTURES HDI 1.GROUP	0.8710	1.0323	-0.1613	0.1613	3		3
STRUCTURES HDI 2.GROUP	0.6897	0.6207	0.0690	0.0690	2	2	-
STRUCTURES HDI 3.GROUP	0.7500	0.2143	0.5357	0.5357	6	6	-
STRUCTURES HDI 4.GROUP	0.4333	0.1667	0.2666	0.2666	4	4	-
TECHNOLOGIES HDI 1.GROUP	1.6774	0.4839	1.1935	1.1935	7	7	-
TECHNOLOGIES HDI 2.GROUP	2.0690	0.3448	1.7242	1.7242	8	8	-
TECHNOLOGIES HDI 3.GROUP	1.1786	0.7500	0.4286	0.4286	5	5	-
TECHNOLOGIES HDI 4.GROUP	0.1000	0.0333	0.0667	0.0667	1	1	-
TOTAL						33	3

**TABLE 13.** Summary of Wilcoxon Signed-Rank test ( $\alpha = 0.05$ , one-tail).

	K-Medoids – Fuzzy C-Means	K-Medoids – K-Means
Test Statistic	2	3
Critical Value	6	6
Reject $H_0$	Yes	Yes



**FIGURE 3.** Smart cities index in terms of countries (according to structure).

edition evaluated 109 cities by gathering the opinions of 120 locals in each city. The 118 cities included in the previous edition's ranking had the same number of citizens. The

final score was calculated using SCI index utilizing survey responses from the previous three years, with a weight of 3:2:1 for 2021, 2020, and 2019. The final score is calculated



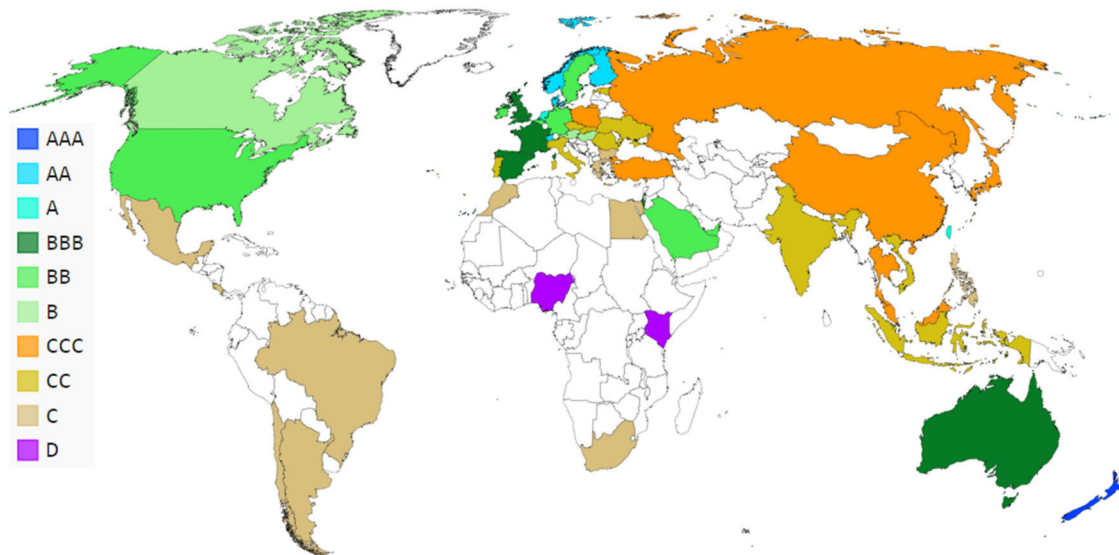


FIGURE 4. Smart cities index in terms of countries (according to technology).

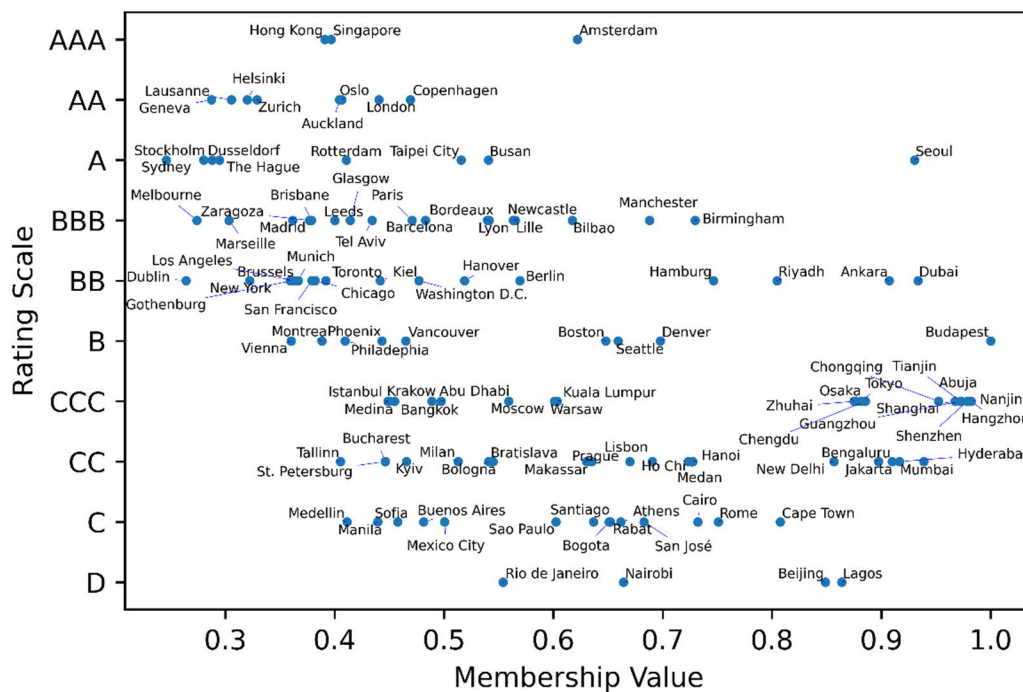


FIGURE 5. Clusters of smart cities (according to technology).

using the same process as the index in this work, and clustering algorithms such K-Means, Fuzzy C-Means, and K-Medoids are suggested.

#### A. OBTAINED CLUSTERS

In this section, clustering algorithms' and the IMD-SUTD SCI results are presented. The results of K-Means, K-Medoids, and Fuzzy C-Means clustering algorithms and also SCI rating scales are represented in eight tables. While Table 2 is provided here, the rest of them (Tables 3 to 9) are

available as supplementary material. In addition, the tables display the class membership values for the Fuzzy C-Means algorithm. Accordingly, it can be said that cities belong to the class that has the largest membership value. For example, the largest value is 0.4273 for Amsterdam in Table 2. So, Amsterdam rating scale is "A". Table 2 and Supp Tables 3 to 5 in the supplementary material show the results for the "structure" main title and Supp Tables 6 to 9 in the supplementary material show the results for the "technologies" main title, respectively. Table 2 and Supp Table 6 show the

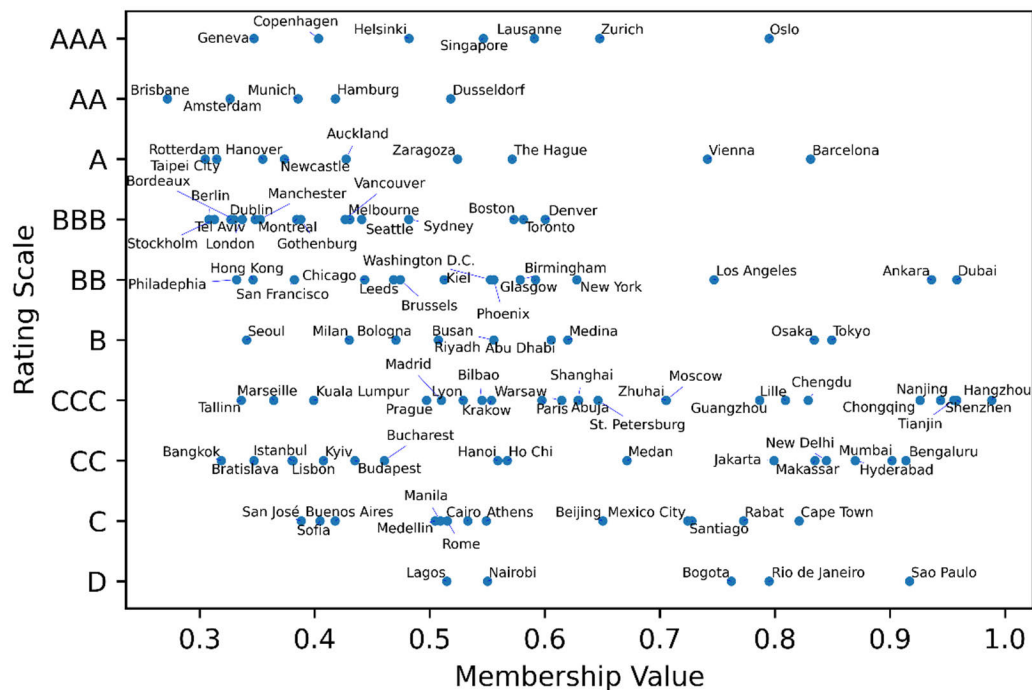


FIGURE 6. Clusters of smart cities (according to structure).

results for cities in the highest HDI quartile. Table 3 and Supp Table 7 display the results for second HDI quartile. Supp Table 4 and 8 show the results for third HDI quartile and finally, Supp Table 5 and 9 present the results for lowest HDI quartile. The symbols ( $\uparrow$ ,  $\downarrow$ , and  $\bullet$ ) represent the change (increased, decreased and stable) in ranking compared to the current report (index result).

According to the results from both SCI and the other three algorithms, Lausanne, Oslo, and Zurich have the highest scores under the “structure” title as indicated in Table 2. Supp Table 3 in the Supplementary Material reveals that Singapore has the highest score among the cities under the “technology” title, based on the results of the index and algorithms.

In this section, we present the analysis of the clustering methods on the basis of their effectiveness in clustering data. Specifically, we evaluate the minimum square error (MSE), the number of cities with similar rating scales, and the number of cities with dissimilar rating scales, as compared to the SCI report. The MSE is a measure of the overall quality of the clustering, where a smaller value indicates a better fit of the clustering model to the data. The number of cities with similar rating scales indicates the ability of the clustering method to group cities with similar characteristics together. On the other hand, the number of cities with dissimilar rating scales indicates the clustering method’s ability to differentiate between cities with different characteristics. By evaluating the clustering efficiency using these three metrics, we can gain a more comprehensive understanding of the strengths and limitations of each method. This information will be invaluable in select-

ing the most appropriate clustering algorithm for a particular dataset and research question.

The comparative results of the clustering methods are given in Table 10. According to the results, the Fuzzy C-Means algorithm outperformed K-Means and K-Medoids on SCI data for groups 1 and 2 in “structure” title, as it produced the results that were closest to the SCI results. In contrast, both K-Means and K-Medoids had less favorable performance. On the other hand, the Fuzzy C-Means and K-Means clustering algorithms showed similar performance for groups 3 and 4 in the “structure” title. In the “technology” title, K-Means generally had the highest performance among the tested algorithms. However, for group 3, both Fuzzy C-Means and K-Medoids outperformed K-Means.

The Wilcoxon Signed Rank statistical test is used to see which clustering approach is effective. We suggest two hypotheses to test this situation. Hypothesis theses are as follows:

$H_0$ : The Fuzzy C-Means and K-Means algorithms produce results that aren’t more similar to the index than K-Medoids algorithm.

$H_a$ : The Fuzzy C-Means and K-Means algorithms produce results that are more similar to the index than K-Medoids algorithm.

While Tables 11 and 12 show the Wilcoxon signed-rank test results of K-Medoids compared to Fuzzy C-Means and K-Means, respectively, summarized results are given in Table 13.

According to the Wilcoxon signed-rank test, if the test value is less than critical value, null hypothesis ( $H_0$ ) is

rejected (Table 13). Reject  $H_0$  refers there is enough evidence at the 5% level of significance to support the claim that the Fuzzy C-Means and K-Means algorithms produce results that are more similar to the index than K-Medoids algorithm.

The status of countries is determined by averaging the fuzzy membership values of the cities within each country. A country's rating scale is determined according to the scale with the highest membership value among the calculated averages. Figure 3 illustrates the rankings of countries under the "structure" title. Accordingly, the countries that receive the highest rating (AAA) in descending order of membership value are as follows: Norway, Singapore, Switzerland, Finland, and Denmark. Figure 4 shows the rankings of countries under the "technology" title. According to this, the countries with the highest rating of membership value are Hong Kong, New Zealand, and Singapore.

Two images were created to enhance the understanding of the Fuzzy C-Means algorithm results for the "technology" and "structure" titles. The primary objective of them was to facilitate the identification of which city belongs to which rating. In other words, these images provide answers to questions regarding which city is included in which scale, as well as their corresponding membership degrees. For example, according to Figure 5 both Hong Kong and Singapore is part of the AAA scale with a membership rating of approximately 0.4, whereas Amsterdam is included in the AAA scale with a higher membership rating of approximately 0.6. As a result, Amsterdam is ranked higher than Hong Kong and Singapore in the context of smart cities. Figure 5 depict the clusters of smart cities as represented by their fuzzy membership values according to "technology" title. As a result, while Hong Kong, Singapore and Amsterdam are rated as AAA, Rio de Janeiro, Nairobi, Beijing and Lagos are rated as D in terms of the "technology" title. According to the "structure" title the cluster of smart cities is illustrated in Figure 6 through their fuzzy membership values. While Singapore is rated as AAA, Rio de Janeiro, Nairobi, and Lagos are rated as D for both "technology" and "structure" titles.

## V. CONCLUSION

The desire for cities to offer the improved public services that might affect people's daily lives has been currently rising. Although cities only make up 3% of the earth's land, they are the backbone of our economy, utilize 75% of the world's energy and generate 80% of the extra greenhouse emissions produced worldwide [67]. The so-called smart cities have made efficient use of resources and improved living conditions via the application of technology. The 2021 SCI, published by the IMD in partnership with SUTD, includes significant insights on how urban inhabitants will handle new duties. The report groups the cities into different clusters to interpret smart indicators better and more effectively. Although the study offers a thorough examination of the cities for evaluation, the city clusters should be more sensitive and not produced using strict clustering techniques.

This paper proposes K-Means, Fuzzy C-Means, and K-Medoids clustering algorithms that are much more information and flexible than any hard clustering methods and computes the final score using the same method as the index. The effectiveness of different clustering methods is evaluated in grouping data. Our analysis focuses on three metrics: minimum square error (MSE), the number of cities with similar rating scales, and the number of cities with dissimilar rating scales, with comparison to the SCI report. A lower MSE is generally considered as advantageous in various contexts. The MSE measures the average squared difference between predicted values and actual values in a dataset. A lower MSE indicates that the predicted values are closer to the actual values, implying higher accuracy in the model's predictions. Within this framework K-means demonstrates superior performance with a MSE of 0.033 for the "technology" category, group 4. Our results indicate that, for groups 1 and 2 in the "structure" title, the Fuzzy C-Means algorithm demonstrated superior performance compared to K-Means and K-Medoids. This is evidenced by the Fuzzy C-Means algorithm producing results that closely match those of the SCI report, while K-Means and K-Medoids had less favorable performance. In contrast, for groups 3 and 4 in the "structure" title, both Fuzzy C-Means and K-Means algorithms showed similar performance. In the "technology" title, K-Means generally demonstrated the highest performance among the tested algorithms. However, for group 3, both Fuzzy C-Means and K-Medoids outperformed K-Means. These findings suggest that the selection of the most appropriate clustering algorithm may depend on the specific research question and dataset characteristics. In summary, it can be inferred that the Fuzzy C-Means and K-Means clustering algorithms demonstrated comparable performance. The Fuzzy C-Means and K-Means algorithms produce results that are more similar to the index than K-Medoids algorithm. While conducting the study, some limitations have been encountered. The first of these is that the study is limited to only 118 cities, while the other is that some of the data in the report (SCI) used as a source are obtained through a survey.

The following open research questions should be looked into for additional studies: (i) embedding other temporal and spatial indicators should be strengthen the assessment process, (ii) calculating the weights of indicators should be done multi-criteria decision making approaches, and finally (iii) different clustering algorithms should be tested in the similar studies. For future studies, other clustering techniques such as the K-rms algorithm, the Fast Markov clustering algorithm, or meta-heuristics data clustering algorithms could be employed and compared with the existing work.

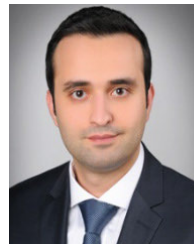
## REFERENCES

- [1] World Bank. (2022). *Urban Population (% of Total Population)*. Accessed: Jan. 9, 2023. [Online]. Available: <https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS>

- [2] United Nations. (2022). *United Nations World Urbanization Prospects*. Accessed: Jan. 9, 2023. [Online]. Available: <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>
- [3] I. Turok and G. McGranahan, "Urbanisation and economic growth: The arguments and evidence for Africa and Asia," *Urbanisation*, vol. 4, no. 2, pp. 109–125, Nov. 2019.
- [4] C. Zhong, R. Hu, M. Wang, W. Xue, and L. He, "The impact of urbanization on urban agriculture: Evidence from China," *J. Cleaner Prod.*, vol. 276, Dec. 2020, Art. no. 122686.
- [5] L. Shen, Z. Huang, S. W. Wong, S. Liao, and Y. Lou, "A holistic evaluation of smart city performance in the context of China," *J. Cleaner Prod.*, vol. 200, pp. 667–679, Nov. 2018.
- [6] R. Ali, K. Bakhsh, and M. A. Yasin, "Impact of urbanization on CO<sub>2</sub> emissions in emerging economy: Evidence from Pakistan," *Sustain. Cities Soc.*, vol. 48, Jul. 2019, Art. no. 101553.
- [7] X. Qi, Y. Han, and P. Kou, "Population urbanization, trade openness and carbon emissions: An empirical analysis based on China," *Air Qual., Atmos. Health*, vol. 13, no. 5, pp. 519–528, May 2020.
- [8] H. Kumar, M. K. Singh, M. P. Gupta, and J. Madaan, "Moving towards smart cities: Solutions that lead to the smart city transformation framework," *Technol. Forecasting Social Change*, vol. 153, Apr. 2020, Art. no. 119281.
- [9] E. Almirall, J. Wareham, C. Ratti, P. Conesa, F. Bria, A. Gaviria, and A. Edmondson, "Smart cities at the crossroads: New tensions in city transformation," *California Manage. Rev.*, vol. 59, no. 1, pp. 141–152, Nov. 2016.
- [10] M. Ibrahim, A. El-Zaart, and C. Adams, "Smart sustainable cities roadmap: Readiness for transformation towards urban sustainability," *Sustain. Cities Soc.*, vol. 37, pp. 530–540, Feb. 2018.
- [11] IMD. (2022). *Smart City Index*. Accessed: Jan. 9, 2023. [Online]. Available: <https://www.imd.org/smart-city-observatory/home/>
- [12] World Bank. *World Development Report 2016: Digital Dividends*, The World Bank, Washington DC, USA, 2016.
- [13] H. H. Khan, M. N. Malik, R. Zafar, F. A. Goni, A. G. Chofreh, J. J. Klemes, and Y. Alotaibi, "Challenges for sustainable smart city development: A conceptual framework," *Sustain. Develop.*, vol. 28, no. 5, pp. 1507–1518, Sep. 2020.
- [14] J. H. Lee, R. Phaal, and S.-H. Lee, "An integrated service-device-technology roadmap for smart city development," *Technol. Forecasting Social Change*, vol. 80, no. 2, pp. 286–306, Feb. 2013.
- [15] F. Pumomo and H. Prabowo, "Smart city indicators: A systematic literature review," *J. Telecommun., Electron. Comput. Eng. (JTEC)*, vol. 8, pp. 161–164, 2016.
- [16] A. Sharifi, "A typology of smart city assessment tools and indicator sets," *Sustain. Cities Soc.*, vol. 53, Feb. 2020, Art. no. 101936.
- [17] A. Sharifi and Z. Allam, "On the taxonomy of smart city indicators and their alignment with sustainability and resilience," *Environ. Planning B, Urban Analytics City Sci.*, vol. 49, no. 5, pp. 1536–1555, Jun. 2022.
- [18] D. J. Clement and P. N. Crutzen, "How local policy priorities set the smart city agenda," *Technological Forecasting Social Change*, vol. 171, Oct. 2021, Art. no. 120985.
- [19] A. Abu-Rayash and I. Dincer, "Development of integrated sustainability performance indicators for better management of smart cities," *Sustain. Cities Soc.*, vol. 67, Apr. 2021, Art. no. 102704.
- [20] F. Skultéty, D. Benová, and J. Gnap, "City logistics as an imperative smart city mechanism: Scrutiny of clustered EU27 capitals," *Sustainability*, vol. 13, no. 7, p. 3641, Mar. 2021.
- [21] O. Ozdemir and A. Kaya, "K-Medoids and fuzzy C-Means algorithms for clustering CO<sub>2</sub> emissions of Turkey and other OECD countries," *Appl. Ecol. Environ. Res.*, vol. 16, no. 3, pp. 2513–2526, 2018.
- [22] V. Moustaka, A. Vakali, and L. G. Anthopoulos, "A systematic review for smart city data analytics," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–41, Sep. 2019.
- [23] K. Soomro, M. N. M. Bhutta, Z. Khan, and M. A. Tahir, "Smart city big data analytics: An advanced review," *WIREs Data Mining Knowl. Discovery*, vol. 9, no. 5, Sep. 2019, Art. no. e1319.
- [24] J. Souza, A. Francisco, C. Piekarski, and G. Prado, "Data mining and machine learning to promote smart cities: A systematic review from 2000 to 2018," *Sustainability*, vol. 11, no. 4, p. 1077, Feb. 2019.
- [25] Z. S. Ageed, S. R. M. Zeebaree, M. M. Sadeeq, S. F. Kak, Z. N. Rashid, A. A. Salih, and W. M. Abdullah, "A survey of data mining implementation in smart city applications," *Qubahan Academic J.*, vol. 1, no. 2, pp. 91–99, Apr. 2021.
- [26] A. Kousis and C. Tjortjis, "Data mining algorithms for smart cities: A bibliometric analysis," *Algorithms*, vol. 14, no. 8, p. 242, Aug. 2021.
- [27] I. H. Sarker, "Smart city data science: Towards data-driven smart cities with open research issues," *Internet Things*, vol. 19, Aug. 2022, Art. no. 100528.
- [28] D. Borys, P. Bzowski, M. Danch-Wierzchowska, and K. Psiuk-Maksymowicz, "Comparison of k-means related clustering methods for nuclear medicine images segmentation," *Proc. SPIE*, vol. 10341, pp. 232–236, Mar. 2017.
- [29] H.-X. Pei, Z.-R. Zheng, C. Wang, C.-N. Li, and Y.-H. Shao, "D-FCM: Density based fuzzy c-means clustering algorithm with application in medical image segmentation," *Proc. Comput. Sci.*, vol. 122, pp. 407–414, Jan. 2017.
- [30] A. Zaman, K. Ullah, R. Ullah, H. H. Imtiaz, and L. Yu, "Image segmentation of MRI image for brain tumor detection," *Int. J. Eng. Appl. Sci. Technol.*, vol. 4, pp. 50–55, Jan. 2019.
- [31] T. Velmurugan and A. Dharmarajan, "Clustering lung cancer data by k-means and k-medoids algorithms," *Int. J. Data Mining Techn. Appl.*, vol. 3, no. 2, pp. 95–98, 2014.
- [32] A. Dharmarajan and T. Velmurugan, "Lung cancer data analysis by k-means and farthest first clustering algorithms," *Indian J. Sci. Technol.*, vol. 8, no. 15, pp. 1–8, Jul. 2015.
- [33] U. Ojha and S. A. Goel, "A study on prediction of breast cancer recurrence using data mining techniques," in *Proc. 7th Int. Conf. Cloud Comput., Data Sci. Eng.*, Jan. 2017, pp. 527–530.
- [34] A. Bose, A. Sarkar, and U. Maulik, "Lesion segmentation using entropy based membership," in *Proc. 2nd Int. Conf. Electron., Mater. Eng. Nano-Technology (IEMENTech)*, May 2018, pp. 1–4.
- [35] K. Karthikeyan, A. J. Jabeen, K. S. Kumar, and K. Venkatalakshmi, "Analysis of diverse optimisation algorithms in breast cancer detection," *Int. J. Image Mining*, vol. 3, no. 1, pp. 4–21, 2018.
- [36] O. N. Kenger and E. Ozceylan, "A comparative analysis of Fuzzy C-Means, K-Means, and K-Medoids clustering algorithms for analysis countries' COVID-19 risk," in *Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation*, C. Kahraman, S. Cebi, S. C. Onar, B. Oztaysi, A. C. Tolga, and I. U. Sari, Eds. Berlin, Germany: Springer, 2022, pp. 27–35.
- [37] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the COVID-19 outbreak in Indonesia," *ILKOM J. Ilmiah*, vol. 13, no. 1, pp. 31–35, Apr. 2021.
- [38] N. Nurdin, S. Fitriani, Z. Yunizar, and B. Bustami, "Clustering the distribution of COVID-19 in Aceh province using the Fuzzy C-Means algorithm," *JTAM*, vol. 6, no. 3, pp. 665–677, Jul. 2022.
- [39] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Exp. Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, Mar. 2009.
- [40] B. R. Jipkate and V. V. Gohokar, "A comparative analysis of Fuzzy C-means clustering and K means clustering algorithms," *Int. J. Comput. Eng.*, vol. 2, pp. 737–739, 2012.
- [41] G. Surya Narayana and D. Vasumathi, "An attributes similarity-based K-medoids clustering technique in data mining," *Arabian J. Sci. Eng.*, vol. 43, no. 8, pp. 3979–3992, Aug. 2018.
- [42] M. Kurniawan, R. R. Muhima, and S. Agustini, "Comparison of clustering K-means, Fuzzy C-means, and linkage for Nasa active fire dataset," *Int. J. Artif. Intell. Robot. (IJAIR)*, vol. 2, no. 2, pp. 34–41, Dec. 2020.
- [43] K. Zhou and S. Yang, "Effect of cluster size distribution on clustering: A comparative study of k-means and fuzzy c-means clustering," *Pattern Anal. Appl.*, vol. 23, no. 1, pp. 455–466, 2019.
- [44] M. Hosseinzadeh, A. Hemmati, and A. M. Rahmani, "Clustering for smart cities in the Internet of Things: A review," *Cluster Comput.*, vol. 25, no. 6, pp. 4097–4127, Dec. 2022.
- [45] L. Rokach, "A survey of clustering algorithms," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA, USA: Springer, 2010, pp. 269–298.
- [46] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, Jan. 2020.



- [47] O. A. D. Ghoneim, "Traffic jams detection and congestion avoidance in smart city using parallel K-means clustering algorithm," in *Proc. Int. Conf. Cognition Recognit.*, D. S. Guru, T. Vasudev, H. K. Chethan, and Y. H. S. Kumar, Eds. Singapore: Springer, 2018, pp. 21–30.
- [48] K. Pholsena and L. Pan, "Traffic status evaluation based on possibilistic fuzzy c-means clustering algorithm," in *Proc. IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC)*, Jun. 2018, pp. 175–180.
- [49] Z. Liu, H. Jia, and Y. Wang, "Urban expressway parallel pattern recognition based on intelligent IoT data processing for smart city," *Comput. Commun.*, vol. 155, pp. 40–47, Apr. 2020.
- [50] H. E. Pérez, A. L. Mederos, D. G. Lio, L. E. Hurtado, D. G. Duarte, and G. A. Atemez, "A system for traffic events detection using fuzzy c-means," in *Proc. Knowl. Graphs Semantic Web*, B. Villazon-Terrazas, F. Ortiz-Rodríguez, S. Tiwari, A. Goyal, and M. Jabbar, Eds. Cham, Switzerland: Springer, 2021, pp. 71–83.
- [51] Z. Lv, L. Qiao, K. Cai, and Q. Wang, "Big data analysis technology for electric vehicle networks in smart cities," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1807–1816, Mar. 2021.
- [52] F. Xiao, X. Xie, Z. Li, Q. Deng, A. Liu, and L. Sun, "Wireless network optimization via physical layer information for smart cities," *IEEE Netw.*, vol. 32, no. 4, pp. 88–93, Jul./Aug. 2018.
- [53] R. Sharma, V. Vashisht, and U. Singh, "Performance analysis of evolutionary technique based partitioned clustering algorithms for wireless sensor networks," in *Soft Computing: Theories and Applications*, M. Pant, T. K. Sharma, O. P. Verma, R. Singla, and A. Sikander, Eds. Singapore: Springer, 2020, pp. 171–180.
- [54] A. Mohanty, S. K. Mohanty, B. Jena, A. G. Mohapatra, A. N. Rashid, A. Khanna, and D. Gupta, "Identification and evaluation of the effective criteria for detection of congestion in a smart city," *IET Commun.*, vol. 16, no. 5, pp. 560–570, Mar. 2022.
- [55] E. T. Al-Shammari, S. Shamshirband, D. Petkovic, E. Zalnezhad, P. L. Yee, R. S. Taher, and Z. Cojbasic, "Comparative study of clustering methods for wake effect analysis in wind farm," *Energy*, vol. 95, pp. 573–579, Jan. 2016.
- [56] S. Ansari, A.-K. Hamid, and N. A. A. Hindawi, "Optimal placement of grid-connected solar photovoltaic systems using artificial intelligence methods," in *Proc. 14th Int. Conf. Develop. eSystems Eng. (DeSE)*, Dec. 2021, pp. 252–258.
- [57] T. Velmurugan, "Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data," *Appl. Soft Comput.*, vol. 19, pp. 134–146, Jun. 2014.
- [58] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," in *Pattern Recognition With Fuzzy Objective Function Algorithms*. Boston, MA, USA: Springer, 1981.
- [59] J. Nayak, B. Naik, and H. S. Behera, "Fuzzy c-means (FCM) clustering algorithm: A decade review from 2000 to 2014," in *Computational Intelligence in Data Mining (Smart Innovation, Systems and Technologies)*, vol. 32, L. C. Jain, H. S. Behera, J. K. Mandal, D. P. Mohapatra, Eds. New Delhi, India: Springer, 2015, pp. 133–149.
- [60] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.
- [61] A. Andrejovska, "Meta-analysis combining cluster analysis and multidimensional scaling—Categorisation of signs of the European union countries' insolvency," *J. Appl. Econ. Sci.*, vol. 8, no. 26, pp. 416–425, 2013.
- [62] F. Anggoro, R. E. Caraka, F. A. Prasetyo, M. Ramadhani, P. U. Gio, R.-C. Chen, and B. Pardamean, "Revisiting cluster vulnerabilities towards information and communication technologies in the eastern island of Indonesia using fuzzy C means," *Sustainability*, vol. 14, no. 6, p. 3428, Mar. 2022.
- [63] A. Luczak and S. Kalinowski, "Fuzzy clustering methods to identify the epidemiological situation and its changes in European countries during COVID-19," *Entropy*, vol. 24, no. 1, p. 14, Dec. 2021.
- [64] S. Ghosh and S. Kumar, "Comparative analysis of K-Means and Fuzzy C-Means algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 4, pp. 35–39, 2013.
- [65] U. W. Latifah, S. Surono, and S. Suparman, "K-Means and Fuzzy C-Means algorithm comparison on regency/city grouping in central Java province," *Desimal, J. Matematika*, vol. 5, no. 2, pp. 155–168, 2022.
- [66] G. R. Suraya and A. W. Wijayanto, "Comparison of hierarchical clustering, K-Means, K-Medoids, and Fuzzy C-Means methods in grouping provinces in Indonesia according to the special index for handling stunting," *Indonesian J. Statist. Appl.*, vol. 6, no. 2, pp. 180–201, 2022.
- [67] T. Velmurugan and T. Santhanam, "Computational complexity between K-Means and K-Medoids clustering algorithms for normal and uniform distributions of data points," *J. Comput. Sci.*, vol. 6, no. 3, pp. 363–368, Mar. 2010.
- [68] S. Farafisha, A. Fauzan, and A. Novianti, "Implementation of Fuzzy C-Means and K-Medoids in grouping people's welfare indicators (case study in Riau province in 2020)," in *Proc. 1st Int. Seminar Teacher Training Educ. Purwokerto*, Indonesia: EAI, Jul. 2021, pp. 17–18.
- [69] K. Kuru and D. Ansell, "TCitySmartF: A comprehensive systematic framework for transforming cities into smart cities," *IEEE Access*, vol. 8, pp. 18615–18644, 2020.



**ÖMER N. KENGER** was born in May 1988. He received the bachelor's degree in industrial engineering from Sakarya University and the master's and Ph.D. degrees in industrial engineering from Gaziantep University.

Before his current position as a Faculty Member of Hasan Kalyoncu University, in 2023, he was the Production Planning and Procurement Manager. He has extensive experience in the field and brings practical industry knowledge to his academic work. He has contributed to academia by publishing many articles and conference papers. His main research interests include machine learning, data mining, and artificial intelligence.

Dr. Kenger benefited from the support of the YÖK 100/2000 Scholarship Program throughout the Ph.D. studies.



**ZÜLAL DİRİ KENGER** was born in Gaziantep, Turkey, in April 1988. She received the bachelor's, master's, and Ph.D. degrees in industrial engineering from Gaziantep University.

After completing the Ph.D. degree, she was a Faculty Member with the Department of Industrial Engineering, Hasan Kalyoncu University, in 2021. She is actively involved in teaching industrial engineering subjects to students and conducting research in her field. She has published several articles in national and international journals and presented her research at conferences. Her publications contribute to the current topics and innovations in the field of industrial engineering. She actively participates in academic conferences, sharing her research with the academic community. Her research interests include the supply chain management, additive manufacturing technology, integration of assembly and disassembly line balancing, and vehicle routing problems.

Dr. Kenger received support from TÜBİTAK and YÖK 100/2000 Scholarship Programs during the Ph.D. studies.



**EREN ÖZCEYLAN** received the bachelor's and master's degrees in industrial engineering and the Ph.D. degree in computer engineering from Selçuk University, in 2007, 2010, and 2013, respectively.

He was a Postdoctoral Fellow with Northeastern University, in 2019. He is currently a Professor with the Industrial Engineering Department, Gaziantep University, Turkey. He has published over 100 peer-reviewed publications in prestigious academic journals, including *Transportation*

*Research Part D: Transport and Environment*, *Transportation Research Part E: Logistics and Transportation Review*, the *Journal of the Operational Research Society*, *International Journal of Production Research*, *International Journal of Production Economics*, and *Computers & Industrial Engineering*. His research interests include the application of mathematical optimization techniques to sustainable transportation, supply chain, and logistics management problems, more specifically to location-allocation and disassembly line balancing problems.

Dr. Özceylan was a recipient of the Science Academy Young Scientists Award (BAGEP), in 2016, and Highly Cited Researcher Award from Gaziantep University, in 2017.



**BEATA MRUGALSKA** received the M.Sc. degree in management and marketing from the Faculty of Mechanical Engineering and Management, Poznań University of Technology, Poland, in 2001, the Ph.D. degree in machine construction and operation from the Faculty of Computer Science and Management, Poznań University of Technology, in 2009, and the D.Sc. (Dr.Habil.) degree in mechanical engineering from the Faculty of Mechanical Engineering, Poznań University of

Technology, in 2019.

She is currently an Associate Professor and the Head of the Division of Applied Ergonomics, Institute of Safety and Quality Engineering, Faculty of Management Engineering, Poznań University of Technology. She has over 100 publications focused on modern organizational management concepts and diagnosis and prognosis of state of technical objects. She serves as a member for eight editorial boards for international journals and a guest editor for four special issues of journals. She was honored with the top peer-reviewer by Publons for her contribution to the preparation of reviews of papers.

...