Cluster Computing (2022) 25:4097–4127
https://doi.org/10.1007/s10586-022-03646-8

Clustering for smart cities in the internet of things: a review

Mehdi Hosseinzadeh[1] · Atefeh Hemmati[2] · Amir Masoud Rahmani[3]

Received: 19 January 2022 / Revised: 28 March 2022 / Accepted: 4 June 2022 / Published online: 27 June 2022

Abstract

Nowadays, internet of things (IoT) applications, especially in smart cities, are fast developing. Clustering is a promising solution for handling IoT issues such as energy efficiency, scalability, robustness, mobility, load balancing, and so on. The clustering method, which can be applied in IoT, groups sensor nodes into clusters with one node operating as the cluster head. This paper intends to determine the usage of clustering in IoT as a case study for smart cities. Furthermore, this study discusses clustering algorithms on IoT, open issues, and future challenges of clustering in the context of the smart city, and also existing research papers selected by the systematic literature review technique published between 2017 and 2021. Also, we provide a technical taxonomy for clustering categorization in IoT, which includes algorithm, architecture, and application. According to the statistical analysis of 51 chosen research articles in the domain of clustering in IoT, the results show that the number of clusters has a high percentage of 24%, the energy factor has 23%, the execution time factor has 18%, the accuracy has 14%, the delay has 9%, the lifetime has 6%, and throughput has 6%.

Keywords Internet of Things · Clustering algorithm · Smart city · Systematic literature review

1 Introduction

The IoT refers to devices that collect and share data with an internet connection. IoT connects the world around us to make it smarter and more responsive and connect the digital and physical worlds. Cluster analysis, also referred to as clustering, is the procedure of organizing a group of things so that things in the same group, known as a cluster, are similar to things in other groups or clusters.

The Internet is currently thoroughly interconnected with the future. The discovery of data through data mining cases like clustering will play a significant role in the field of intelligent systems, allowing for the provision of appropriate services and environments. Data mining approaches are also utilized in wireless sensor networks to cluster equipment and identify clusters [1–3].

Clustering is the concept of connecting your IoT devices to a gateway in a specific scope. Clustering is a method for dealing with the difficulties of managing the resources of IoT objects. Clustering helps minimize energy consumption while also improving the scalability and reliability of object networks. IoT items are combined into clusters to address scalability, energy efficiency, and robustness challenges. Wireless sensing networks may have a higher level of hierarchy thanks to clustering. Clustering, in other terms, is a technique for identifying structure in unlabeled data. A clustering method can find groupings of items in a dataset you do not know anything about by comparing the average distances between members of one cluster to members of other clusters.

Smart city apps use data on both real and virtual system elements. The primary purpose of data collecting is to obtain accurate and complete information. These data, which come from various sources, including sensors, are raw and noisy, and they need to be processed by apps before they can be turned into useful information [4].

Because a smart city's primary demand is a dependable, energy-efficient, and continuous power supply, the complexity of ensuring such a supply grows exponentially. The

✉ Amir Masoud Rahmani
rahmania@yuntech.edu.tw

1 Mental Health Research Center, Psychosocial Health Research Institute, Iran University of Medical Sciences, Tehran, Iran

2 Department of Computer Science, University of Human Development, Sulaymaniyah, Iraq

3 Future Technology Research Center, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou 64002, Yunlin, Taiwan

clustering approach is used to overcome the scalability of power grids and other cyber-physical systems (CPSs) for future smart cities prone to problems and adverse events such as cyber-attacks and physical hazards [5].

Techniques based on big data have turned into a noteworthy data analytics tool in IoT, with data clustering methods recognized as a vital element for data analysis, the quick progress of big data, and IoT. Because the IoT connects sensors and other devices to the Internet, it is critical for the growth of smart services. Alternatively, dynamic entities gather various data types from their real-world surroundings [6].

Smart city devices create data regularly. IoT data is generally diverse and unorganized when it is collected in the cloud. Clustering proves to be a promising solution for resolving IoT-related issues. Depending on each IoT application's problems, different clustering strategies are required [7–9].

Big Data technologies have emerged as key analytics tools for evaluating IoT data due to the tremendous data growth in many IoT areas, such as healthcare and Smart City IoT. Data clustering is an important method of processing IoT data among the Big Data technologies. However, it is currently unclear how to choose an appropriate clustering technique for IoT data [10–12].

There have been few literature surveys on clustering in the smart city context. According to the gathered information, no complete and systematic literature review on clustering in the smart city context has been published. Some studies did not give a taxonomy on clustering in the smart city context. In addition, several studies failed to mention or focus on open issues and future research challenges.

We examined and categorized the articles in clustering in the smart city context, divided into three major categories: algorithm, architecture, and application for answering research questions stated for the proposed study.

This survey paper focuses on five significant contributions, which are as follows:

- Presenting a systematic study due to clustering in the smart city.
- Presenting some research questions (RQs) about clustering in the smart city.
- Reviewing evaluation environments are used to evaluate clustering in the smart city.
- Designing a taxonomy to organize various aspects of clustering in IoT.
- Discussing and focusing on open issues and future research challenges in the clustering in the smart city.

The following sections are included in the rest of the article: the summaries of related work are discussed in Sect. 2. Section 3 addresses the purpose of this study and gives more information on how to choose related research articles. Section 4 describes the classification of research articles, and Sect. 5 highlights existing research activities and open issues. Finally, Sect. 6 brings the paper to a conclusion.

2 Related work and definitions

2.1 Related work

Relevant studies on clustering in IoT summarize in this section. The summary features of each survey in clustering in IoT are defined in Table 1.

Kousis et al. [4] conducted a bibliometric study to offer a thorough overview of works relating to Data Mining technologies utilized in smart cities. The survey paper aimed to find out the primary DM approaches utilized in the smart city domain and the evolution of the research area of data mining for smart cities through time. They investigated the problem using both qualitative and quantitative methodologies. They utilized the Biliometrix library, written in R, for the bibliometric analysis. The findings reveal that various data mining techniques are employed in each stage of a smart city design. According to the bibliometric data, data mining in smart cities is a rapidly emerging academic topic. Scientists from all around the globe are keen to conduct studies and collaborate on this interdisciplinary scientific topic.

Tharwat et al. [5] offered a complete study of artificial intelligence-based clustering methods for smart city cyber-physical systems. Their work is the first thorough survey of AI-based clustering. They mention and compared clustering techniques and algorithms. They also present a taxonomy of an Artificial Intelligence Perspective. And they discussed clustering objectives. They did not present open issues, and the Paper selection process is unclear.

Bangui et al. [6] surveyed big data techniques and clustering methods, outlining the benefits and drawbacks of each method. They have also linked their review to IoT research and highlighted the connections between big data, clustering methods, and IoT. They suggested a set of examination challenges addressing rising research problems and concerns in IoT data clustering, big data application dynamics, and machine learning applications. This study, in particular, has stressed the necessity of clustering techniques in big data in IoT.

Mahyastuty et al. [7] studied wireless sensor network clustering algorithms (WSN). IoT applications can benefit from this clustering strategy. They also recommended IoT-friendly clustering approaches. The clustering method clearly explains the IoT difficulties, particularly energy savings. According to the findings of this study, different

Table 1 Related studies of clustering and IoT

Reference	Main topic	Publication year	Publisher	Review type	Paper selection process	Open issue	Taxonomy	Covered year
Kousis et al. [4]	Data mining algorithms for smart cities	2021	MDPI	Comprehensive overview	Not clear	Not presented	Not presented	2013–2021
Tharwat et al. [5]	Clustering techniques for smart cities	2020	Springer	Comprehensive survey	Not clear	Not presented	Presented	2005–2021
Bangui et al. [6]	Big data clustering algorithms for IoT applications	2019	–	Not clear	Not clear	presented	Not presented	2001–2019
Mahyastuty et al. [7]	Clustering techniques in IoT architecture	2020	IEEE	Survey	Not clear	Not presented	Presented	2004–2020
Nguyen et al. [8]	Clustering on smart cities	2020	Elsevier	Not clear	Not clear	Presented	Presented	1990–2020

clustering approaches can be employed in IoT. The authors classify clustering approaches and give a taxonomy of clustering techniques in this paper, although they do not address open issues.

Nguyen et al. [8] presented a brief introduction of space–time series clustering, divided into three types:

- Hierarchical
- Partitioning-based
- Overlapping clustering

They also go into detail about each category's remedies. Furthermore, they demonstrate these techniques in urban traffic data collected in two Chinese smart cities. Views on great topics and research challenges are also explored, allowing for greater knowledge of the various space–time series clustering algorithms' logic, limits, and advantages. They presented a taxonomy of their work and also discussed open issues.

2.2 Clustering methods

(a) Density-based methods

As shown in Fig. 1 in density-based methods, the clusters are in denser areas separated from the less densely populated areas. In these methods, points within a certain range of each other are placed in a cluster. A minimum density is usually considered in density-based methods, and clustering is performed in areas where this minimum is observed. These methods are inherently defined for continuous space. In the figure below, we have some data—the data cluster forms in each region where the data density is higher.

The following methods can be mentioned as density-based methods:

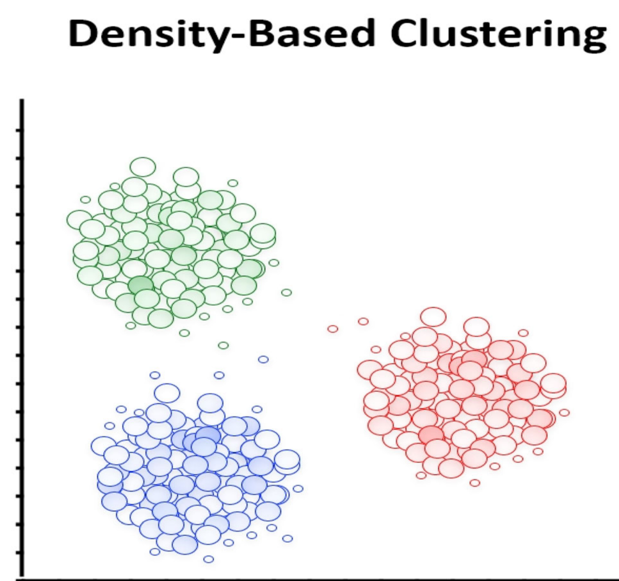

Fig. 1 Density-based method

- Density-based clustering (DENCLUE)
- Density-based spatial clustering of application with noise (DBSCAN)
- Ordering points to identify the clustering structure (OPTICS)

(b) Partitioning methods

As shown in Fig. 2 in Partitioning methods, we divide a subset of the desired data set by the number K into a predetermined set of data subsets. They are suitable for obtaining groups with a spherical shape or a maximum convex shape and can be used in small or medium-sized datasets. This method performs clustering operations based on n observations and K groups. Thus, the number of clusters or groups in this algorithm is already known.

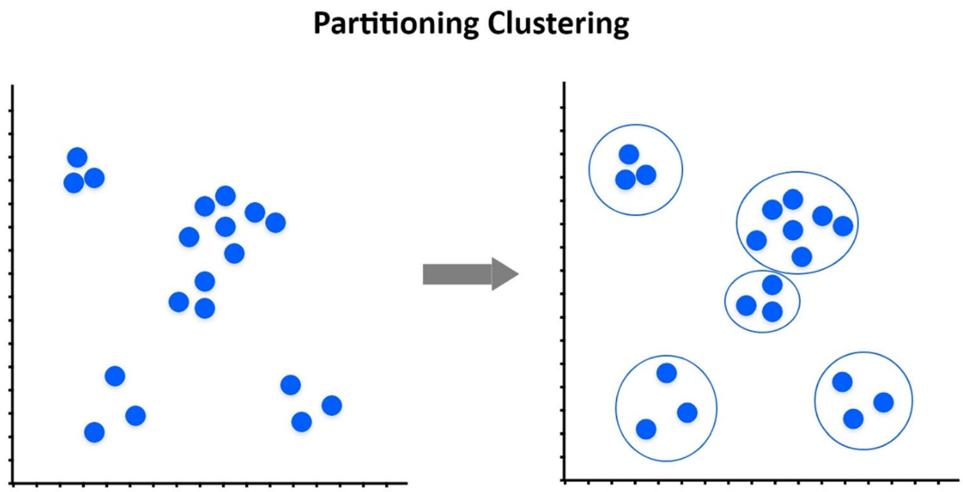

Fig. 2 Partitioning method

During the Partitioning clustering process, each object will belong to only one cluster, and no cluster will remain without a member.

The following methods can be mentioned as partitioning methods:

- K-means
- k-median
- Fuzzy C means
- Partitioning around medoids (PAM)
- Clustering large application (CLARA)
- Clustering large application based on randomised search (CLARANS)

(c) Hierarchical methods

As shown in Fig. 3, hierarchical methods, the tree method divides the data into subgroups. There is no need to specify the number of subgroups in these algorithms. Types of hierarchical methods include divisive hierarchical methods and agglomerative hierarchical methods.

The divisive hierarchical method is a top-down clustering method that starts from the whole data and ends with the smallest component.

The agglomerative hierarchical method is exactly the opposite of the divisive hierarchical method. A bottom-up method starts with the smallest component and ends with all the data in one category.

The following methods can be mentioned as hierarchical methods:

- Agglomerative nesting method (AGNES)
- Divisive analysis method (DIANA)
- Balanced iterative reducing and clustering using hierarchies (BIRCH)
- Clustering using representative (CURE)
- Robust clustering algorithms (ROCK)

(d) Grid methods

As illustrated in Fig. 4, Grid methods are a subset of density-based methods in which each area of the data space that is searched is organized in a network-like structure. For example, the points given on the coordinate page are plotted, and then the page is divided into networks, and the points in a network are in a cluster. This method has a lower accuracy percentage than other methods, but a very good time in It has clustering.

The following methods can be mentioned as Grid methods:

- Statistical information grid-based method (STING)
- Wave cluster

(e) Model-based clustering

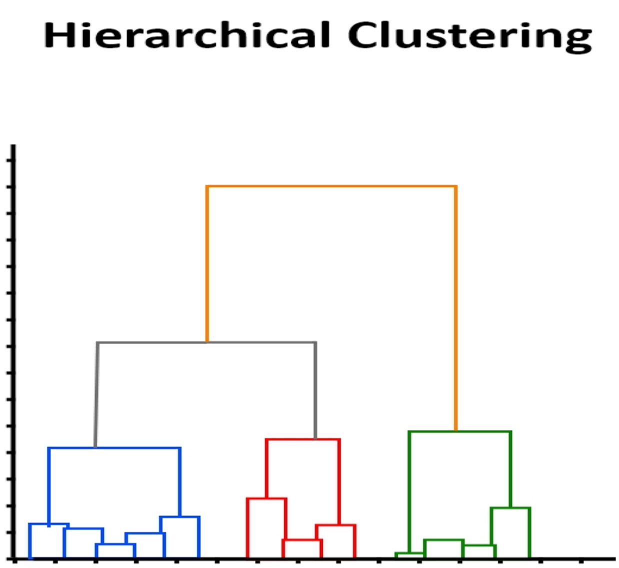

Fig. 3 Hierarchical method

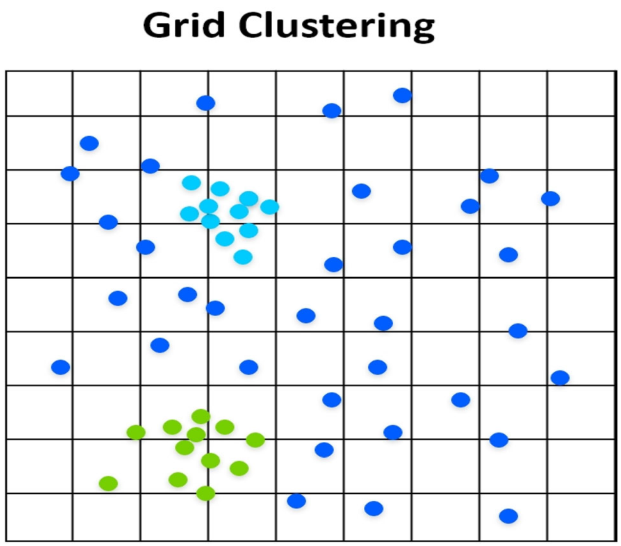

Fig. 4 Grid method

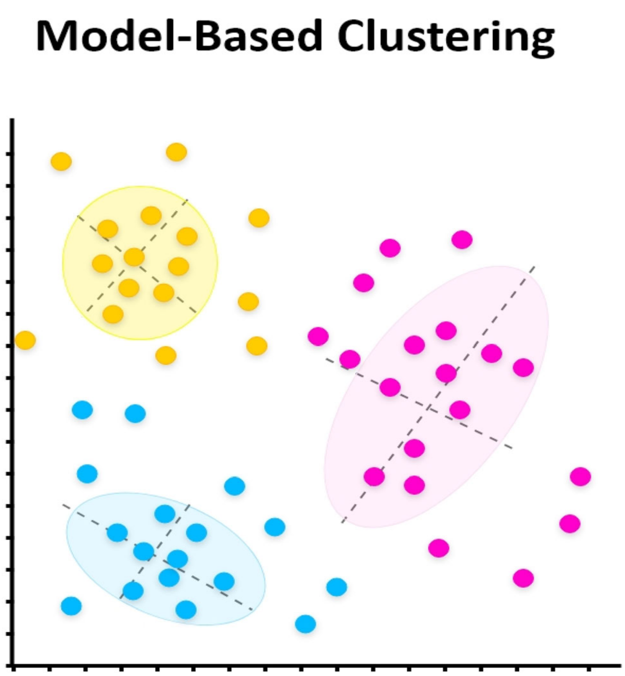

Fig. 5 Model-based method

As shown in Fig. 5 in the model-based clustering method, a statistical distribution for the data is assumed. The purpose of performing model-based clustering is to estimate the statistical distribution parameters and the hidden variable introduced as the label of clusters in the model.

The following methods can be mentioned as model-based methods:

- Expectation maximization (EM) method
- Self-organizing maps (SOM)

2.3 Evaluation factors definition

Execution time: The length of time that the system requires to complete a task.

Number of clusters: Determining the number of clusters in a data set.

Lifetime: The time it will take for the first sensor's energy to run out.

Throughput: The amount of data transmitted from source to destination in a given amount of time.

2.4 Machine learning methods in smart city

In today's technology, machine learning means minimizing human influence wherever possible. This means that the data can discover patterns independently and make decisions without the need for new coding. On the other hand, the IoT aims to connect and interact with various things - over the Internet. The similarities and ambiguity between IoT and machine learning make it challenging to identify the two fields in many cases. However, the IoT is an important infrastructure for developing machine learning algorithms and artificial intelligence. The IoTs' effectiveness comes from their ability to generate and transmit data, which machine learning critically needs. Machine learning algorithms benefit from the information provided by the IoT ecosystem, which improves their accuracy and efficiency. It is necessary to categorize and analyze the data received by IoT devices at regular intervals. Also, make sure it sends information back to the machine so it can make a choice. Each machine learning algorithm is responsible for completing a specific task, so we must first determine what must be completed. Data sorting helps us discover tasks and apply the appropriate algorithm quickly [13–15].

Cities may employ real-time data and intelligent surveillance systems to respond more carefully to the IoT. Smart city devices constantly generate data. Data management is essential for a long-term IoT strategy. When acquired in the cloud, IoT data is typically heterogeneous and unstructured. The key to establishing IoT smart applications is intelligent data processing and analysis. On the IoT, we've chosen Smart City as our major application. Cities are looking for strategies to handle the difficulties of large-scale development as the population rises and the city's infrastructure becomes more complex. Advanced techniques like machine learning and statistical analysis are used in data analysis. Machine learning algorithms look for trends in past sensor data stored in a big data warehouse and build prediction models. Control applications that give commands to the drivers of IoT devices use these models [16–18].

A training set is a collection of samples that a learning algorithm uses as input. There are four types of machine learning: supervised, unsupervised, semi-supervised, and reinforcement.

- In supervised learning, the instructional set includes examples of input vectors and corresponding relevant vector vectors, often known as labels, that are associated with them.
    - Regression: The training data generates a single output value in regression.
    - Classification: It entails classifying the data.
    - Naive Bayesian model: It's a direct acyclic graph-based method for assigning class labels.
    - Decision trees: A decision tree is a flowchart-like architecture that includes conditional control statements and decisions and their likely outcomes.
    - Support vector machines: SVM is a classifier that can discriminate between two classes.
- In unsupervised learning, the training set does not need to be labeled.
    - Clustering: the purpose is to locate subgroups within the data; the grouping is based on similarities.
    - Reduced dimensionality: the goal is to summarize the data in a smaller number of dimensions.
- Reinforcement learning is concerned with the challenge of learning the most profitable activity or a list of activities in a given scenario.
    - Q learning: is a value-based strategy for providing information to assist an agent in making decisions.
- Semi-supervised learning is a hybrid of unsupervised and supervised learning.

Choosing which tasks should be performed is essential to make the best conclusions possible when analyzing smart data. The smart city can be considered a set of healthcare, transportation, and smart grid areas where machine learning plays an important role. Figure 6 shows machine learning methods in the IoT and smart city domain.

(a) Healthcare

Identification and diagnosis of diseases: Diagnosis is normally difficult; machine learning plays an essential role in identifying a person's illness, monitoring their health, and suggesting measures to prevent it. The disease can include minor illnesses or acute illnesses such as cancer that are difficult to diagnose early.

(b) Smart grid

Several machine learning approaches have been employed in smart grid applications due to the continual advancement of computational technologies, particularly in data management and analysis. It's the final component of the smart grid system, which is powered by data collection, analysis, and decision-making. Machine learning approaches enable the smart grid to perform as intended by providing an efficient way to analyze and then make suitable grid-running decisions.

(c) Smart transportation

Machine learning has a lot of potential in the transportation industry. Machine learning approaches have become a feature of smart transportation in recent years. Using deep learning, machine learning investigated the complicated interplay between roads, highways, traffic, environmental components, collisions, and other factors. In addition to daily traffic management and data collection, machine learning has a lot of promise.

(d) Smart home

Smart home devices are becoming more interconnected. Instead of depending solely on direct orders or manually programmed routines, the goal of incorporating machine learning into smart homes is to make the home more valuable and responsive by anticipating users' requirements and patterns. Environmental sensors detect things like temperature, light, and the presence of a user. This data and the power usage of some of them during the day can be recorded and transferred to a remote server. This data can be utilized to train machine learning models to give smart home users enhanced convenience, efficiency, and security.

3 Method of research selection

This section provides a summary of SLR-based clustering in the IoT. Also, explain the research aims, database selection, and search terms.

This section aims to explicitly indicate the RQs that are expected to be discovered in "clustering in the context of smart cities." Because of the significance of the selected topic and other factors, such as the lack of a comprehensive paper on the chosen topic, this SLR paper attempts to know the following RQs.

RQ1 Which clustering algorithms were used in the smart city domain?

RQ2 How might data transmission be reduced by using clustering in the smart city?

RQ3 What evaluation environments are used to evaluate clustering in the smart city?

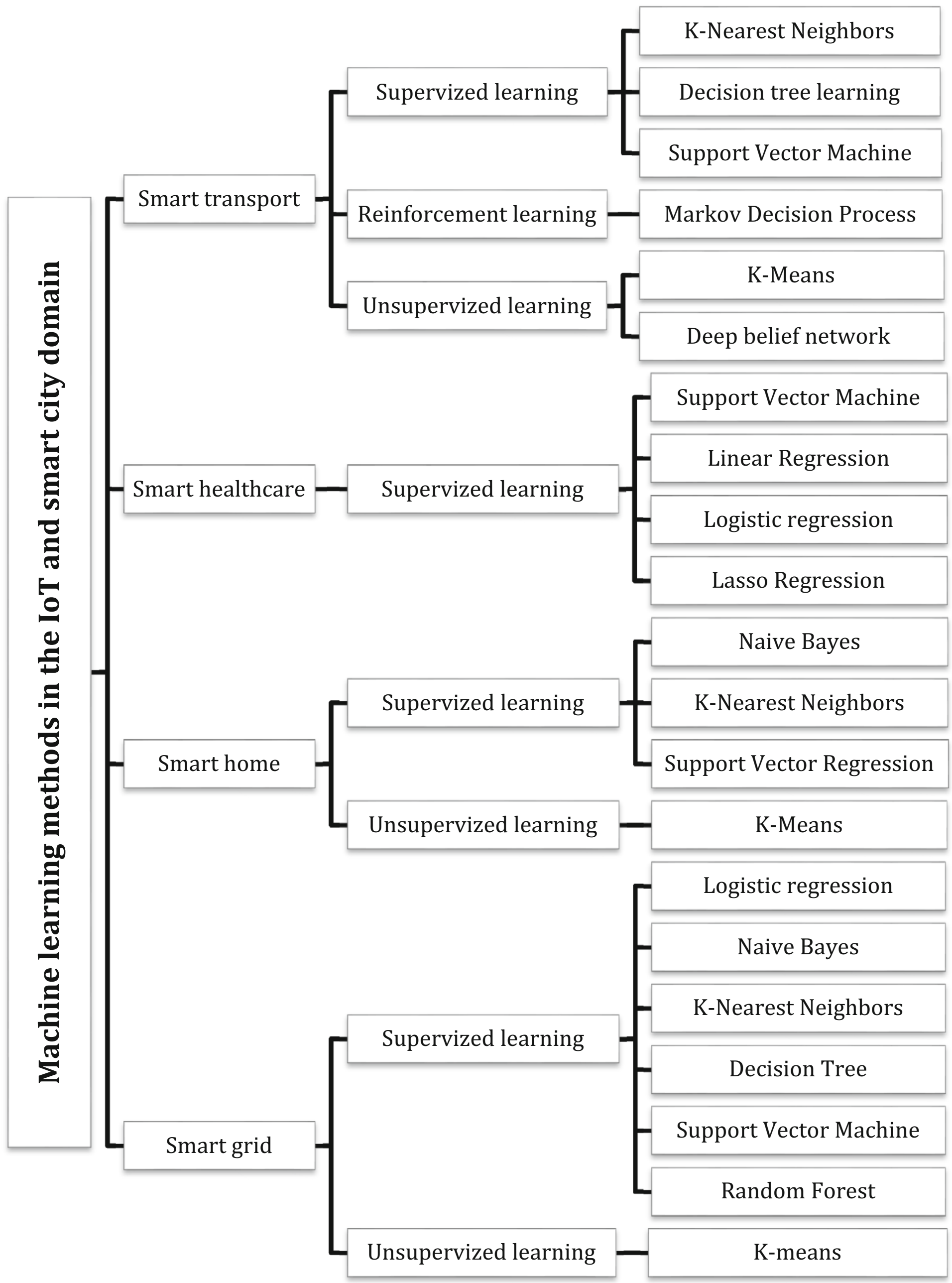

Fig. 6 The taxonomy of machine learning methods in the IoT and smart city domain

RQ4 What are the evaluation factors in clustering domains in the smart city context?

RQ5 What are the open issues and future challenges of clustering in the smart city?

The following string demonstrates how the exploration string was expressed in terms of synonyms and other alternatives for the necessary key components:

("clustering" OR "clustering algorithm") AND ("internet of things" OR "IoT") AND ("smart city") AND ("Survey" OR "Review").

Figure 7 shows the results of research analyses on journal citations and analysis methods published by prestigious publishers such as Elsevier, IEEE, Springer, and others. It shows the number of papers published between 2017 and 2021 in the clustering and IoT domain.

The following inclusion standards are considered when choosing the final articles:

- Articles that were published after 2017
- Articles with titles that contain keywords
- Research articles
- Articles with relevant content

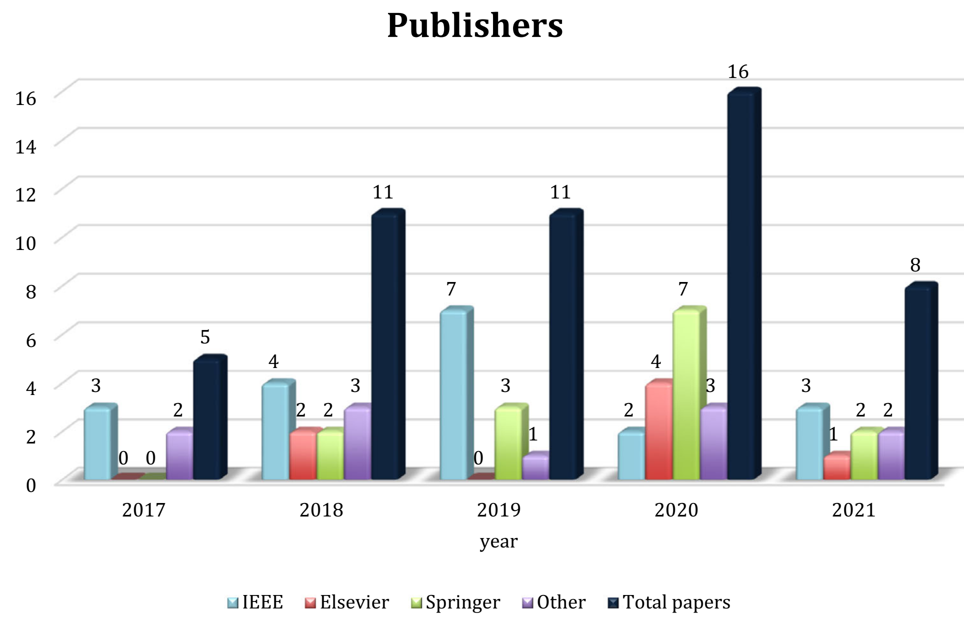

Fig. 7 Distribution of research papers by publisher

The following exclusion standards are considered when choosing the final articles:

- Articles that were published before 2017
- Articles with titles that do not contain keywords
- Not written in English
- Unknown publishers

Subsequently, as presented in Sect. 5, 51 articles were selected to answer the previously defined RQs.

The primary objective of our study, which is predicated on a systematic review, is to examine key capabilities and advancements in the field of clustering in the context of the smart city while also presenting open issues for future research.

The research strategy is demonstrated in Fig. 8, which is categorized into five stages.

A comprehensive taxonomy of clustering in the IoT context is shown in Fig. 9, including algorithm, architecture, and application.

Figure 10 shows a taxonomy of clustering algorithms in the context of smart cities due to a literature review.

4 Organization of clustering in the smart city

4.1 Algorithm

Table 2 shows the classification of the listed articles of the algorithm in the clustering and IoT domain.

In the Hadoop distributed framework, Srinivas et al. [19] developed an evolutionary computing-assisted K-Means clustering technique for MapReduce processing. The suggested solution uses a genetic algorithm to improve centroid estimation and clustering, resulting in superior clustering for MapReduce support. In order to achieve the above centroid estimate and clustering improvement, the suggested GA-based K-Means clustering was used over Hadoop-MapReduce. As the goal function, the silhouette coefficient was employed. GA-k means was used in this case to estimate optimized centroid and clusters across Mapper and Reducer simultaneously, resulting in a quicker and more accurate total calculation. Better centroid estimation, clustering enhancement, achieving higher accuracy, and low computational time are the advantages of the idea of the article.

Alazab et al. [20] developed an enhanced model for selecting the cluster head using wireless sensor networks (WSN) factors like distance, latency, and energy. The suggested FA-ROA model was used to accomplish the cluster head selection procedure. The answers are divided into two groups depending on the most beneficial fitness value in the suggested method. The initial set, named Fitness Averaged-ROA, was changed with the help of averaged values of bypass and follower riders. In contrast, the new group, called Fitness Averaged-ROA, was modified using the averaged values of the attacker and overtake riders. The suggested FA-performance ROA was validated by comparing multiple optimization concepts for living nodes and normalized energy. Expanding the suggested

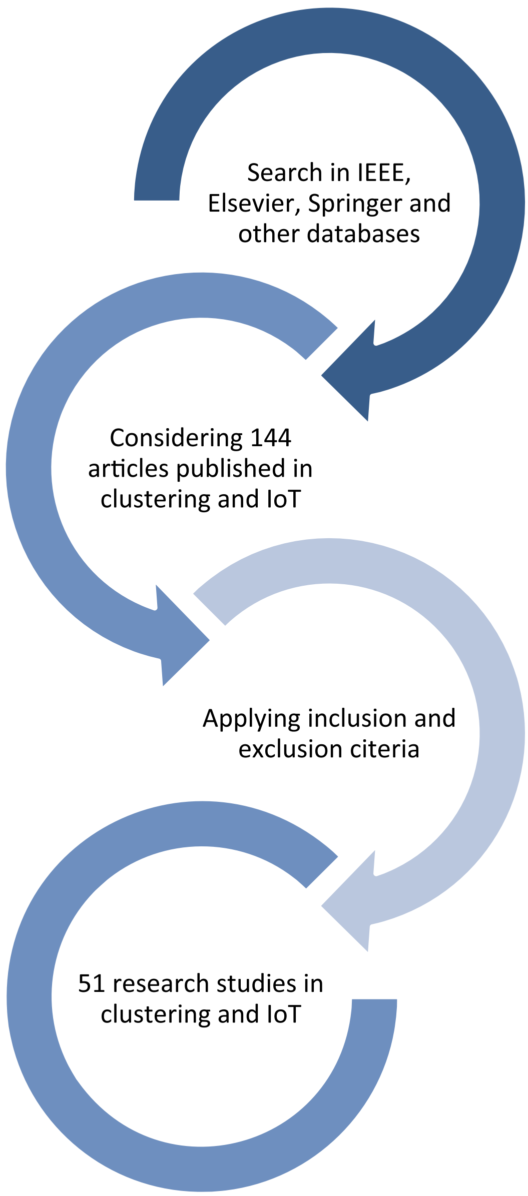

Fig. 8 Research study selection criteria and assessment chart

technique by looking at additional performance parameters is a good idea for the future.

Qin et al. [21] used an enhanced Top-K method to investigate the topic of edge server deployment in edge computing settings for smart cities. This algorithm considers the distance between base stations and edge servers to decrease task access delays and edge server installation costs, load-balancing between edge servers. The results suggested that the method for deployment described in this work outperforms other approaches in terms of server utilization, load balancing, cost and outperforms different algorithms in terms of latency by a small margin. They want to consider the diversity of servers and develop connectivity among edge servers as future work.

Liang et al. [22] considered the focusing fuzzy clustering algorithm as a solution to the issue of the cascaded fault diagnosis based on the weighted computer network's poor dynamic adaptive ability in the smart city. The defect analyses on the cascaded similarity measure around the surroundings of smart cities objects may be performed appropriately and sensitively by incorporating the historical contact proof frame and the timing parameter. The simulation findings suggest that the proposed technique has constant cascaded measurements capability in a dynamic and uncertain smart city, laying a slightly excellent framework for cascaded connectivity, cascaded management, and cascaded decisions based on cascaded measurements.

Ghoneim et al. [23] presented a traffic control model. They used the K-means clustering method to propose a parallel in-memory computing model. Consequently, the data set utilized in this model, the suggested model, offered a complete view of the traffic situation in Aaruthu city. This model made city transit more convenient because it assists citizens in avoiding traffic congestion and saving time. The proposed method provides real-time traffic information for the smart city. The proposed technology-assisted drivers in preventing traffic jams are one of the safest and greenest uses since it reduces energy usage, gas emissions, and fuel consumption. This program was created with H2O open source, an in-memory computing technology that provides quick results for real-time applications.

Feng et al. [24] presented the contribution as follows:

1. They presented an enhanced K-means algorithm for clustering the network and a weighted evaluation function for optimizing the cluster structure.
2. Data fusion was employed to increase the energy usage rate of cluster heads during the data transmission phase. They presented an approach to address the transmission delay problem that data fusion causes.
3. Compared to other algorithms, the proposed method successfully reduces network energy consumption, and the data fusion tree construction reduces data fusion process transmission time.

According to simulation experiment findings, the proposed asymmetrical clustering routing protocol increases network performance and balances network energy consumption, extending network lifespan. The suggested technique is particularly well suited to IoT applications with time constraints. The proposed technique is beneficial for the IoT delay-constraint application.

Mydhili et al. [25] introduced a unique multi-scale parallel K-means ++ clustering method, which was further refined by using machine learning approaches appropriate to challenge optimize energy and scalability

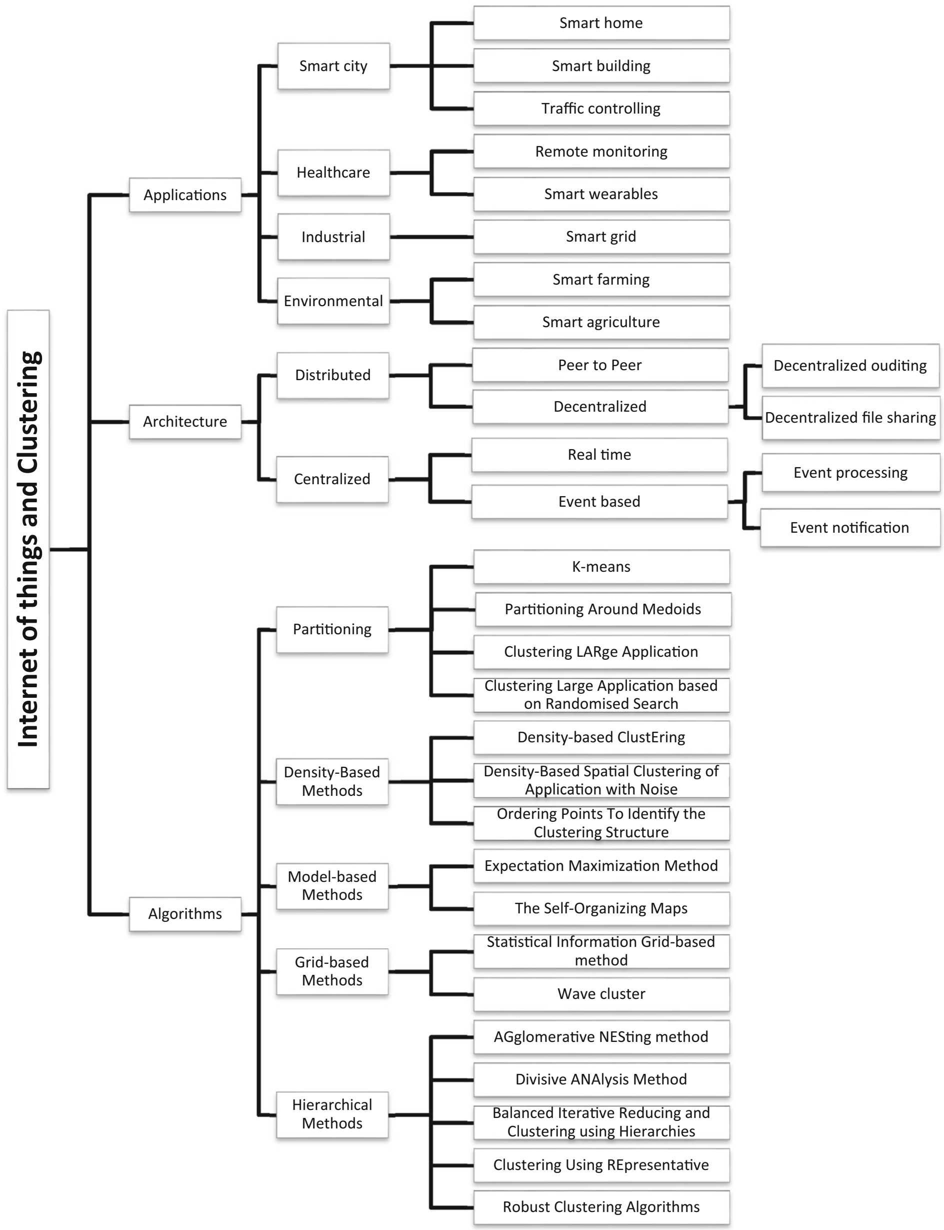

Fig. 9 The taxonomy of clustering and IoT

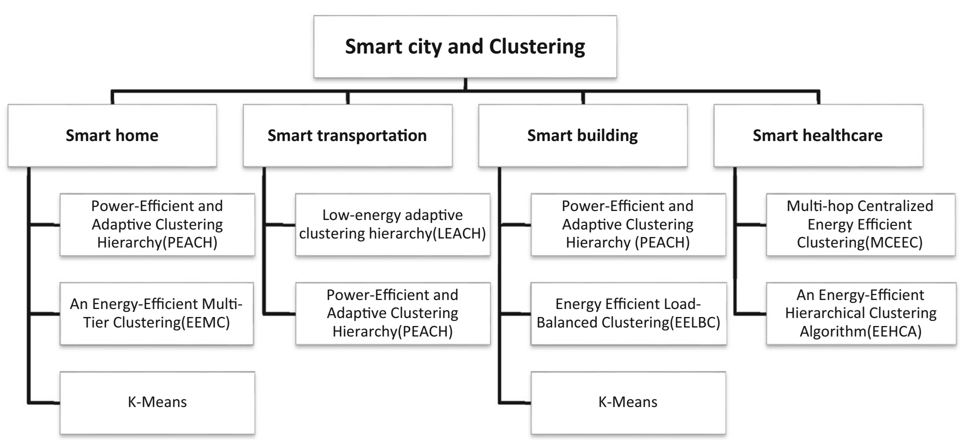

Fig. 10 The taxonomy of clustering in the context of smart cities

Table2 Analyzed articles in the algorithm classification

Research	challenge	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Srinivas et al. [19]	The population of cities is growing exponentially, creating new needs	Developing a K-Means clustering algorithm	Better centroid estimation Clustering enhancement Achieving higher accuracy Low computational time	–	Simulation and implementation	Real-world dataset	K-Means	–
Alazab et al. [20]	Improving performance and energy efficiency in smart cities	Implementing a clustering algorithm	Effectiveness	Not mention enough detail about the advantage of the proposed algorithm	Simulation and implementation using MATLAB 2018a	–	–	Expanding the suggested technique by looking at additional performance parameter
Qin et al. [21]	Edge server deployment in smart city IoT applications includes imbalanced server load and low server utilization	Used an enhanced Top-K algorithm to investigate the topic of edge server deployment for smart cities	Better adaptability lower deployment cost	Not consider the heterogeneity of servers	Simulation	–	Top-K K-Means	Take into account the diversity of servers and develop connectivity among edge servers
Liang et al. [22]	The cascaded fault diagnosis based on the weighted computer network has low dynamic adaptive capacity in the smart city	Cascaded fault diagnosis based on weighted computer network in smart city	Effectiveness Good dynamic adaptability	–	Simulation	–	Fuzzy clustering algorithm	Refine the approach proposed in this study and research the cascaded connection using cascaded metrics
Ghoneim et al. [23]	Traffic management	Presented a traffic control model	Making city transportation more convenient Avoiding the traffic jams Time-saving	–	Implementation	CityPulse	K-means	–
Feng et al. [24]	Unbalanced energy consumption	Proposed an improved K-means algorithm	Balancing network energy consumption Extending the network lifetime	–	Simulation using MATLAB R 2016a and implementation	–	K-means	-
Mydhili et al. [25]	Optimizing energy and scalability problems in wireless sensor networks	Introduced a unique multi-scale parallel K-means ++ clustering method	Better performance	–	Simulation using network simulator (NS2) tool	–	K-means ++	–

Table2 (continued)

Research	challenge	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Han et al. [26]	For a massive network, a single mobile charger is insufficient	Presented a collaborative charging technique for wireless rechargeable sensor networks	Properly refill the network's energy Improving the network's stability	The proposed algorithm cannot handle large-scale locations	Simulation using Matlab	–	Mean-shift clustering algorithm	To handle large-scale locations, contemplate collaborating with numerous MWCVs in a network simultaneously
Wang et al. [27]	Traffic overload	Presented a clustered routing technique that is energy efficient	Increasing the lifespan of a network Increasing throughput Increasing energy efficiency	–	Simulation	–	LEACH	–
Azri et al. [28]	Data retrieval and analytics	proposed 3D dendrogram clustering	Effectiveness	Not mention enough details about the experiment	Implementation	–	K-means	–
Baniata et al. [29]	Energy efficiency	Presented a technique for energy-efficient clustering	Enhancing the energy balance	–	Simulation using Matlab	–	Multihop routing algorithm	–
Azri et al. [30]	Energy efficiency	Developed a novel approach for organizing information from wireless sensor networks in a geographical database	Balancing energy consumption High flexibility and stability	–	Simulation	–	K-means ++	–
Bindhu et al. [31]	Connectivity and sparsity factors	Used subspace clustering In edge servers	Improving accuracy	–	Simulation	Traditional picture datasets	Finding close neighbors	–
Fouladlou et al. [32]	Sensor devices are inexpensive and plentiful	Presented a new routing algorithm	Better performance	–	Implementation using the OPNET modeler simulator	–	Genetic algorithm	–
Xu et al. [33]	Balancing the energy consumption	Clustering routing algorithm	Better performance	–	Simulation	–	–	–
Kumar et al. [34]	Quality of service (QoS) and organizing the resources	K-Means clustering based resource allocation model for IoT	Effectiveness	–	Simulation	–	K-Means	–
Khan et al. [35]	Bandwidth and other network resources	Verification and assessment of the query control mechanism's performance	Effectiveness	–	Simulation	–	Query control mechanism	–

problems in wireless sensor networks. The suggested algorithm's isotropy was demonstrated globally and locally, and the results were empirically confirmed compared to algorithms of reasonable complexity. They used network simulator (NS2) tool to simulate their idea. The suggested method performed better, according to the results.

Han et al. [26] presented an approach based on network density clustering (CCA-NDC) in wireless sensor networks (WRSNs). This technique clusters nodes using a mean-shift algorithm based on density. The experimental findings show that the suggested method may successfully replenish the network's energy and improve its stability. To handle large-scale locations, they want to contemplate working with numerous MWCVs in a network simultaneously in the future. The disadvantage of the presented solution is that it cannot handle large-scale locations.

Wang et al. [27] suggested a clustered routing algorithm that is energy efficient. They offered an unequal cluster formation technique for task scheduling and energy efficiency based on traffic distribution. Furthermore, they suggested a decentralized cluster head (CH) rotating method optimize energy usage within each cluster. They developed a dynamic clustered routing algorithm between CH nodes to circumvent the energy gap issue in lengthy transfers to the base station. The effectiveness of their suggested algorithm was confirmed according to simulation results.

Azri et al. [28] introduced a three-dimensional analytics strategy based on dendrogram clustering. This approach would be used to arrange data, and many outputs and evaluations would be performed to demonstrate the structure's effectiveness for data management in three aspects of smart cities. 3D dendrogram clustering was presented to create a hierarchical tree structure for data recovery and analysis.

Baniata et al. [29] presented an energy-efficient unequal chain length clustering (EEUCLC) protocol with a routing algorithm to decrease cluster head load and a based on probabilities cluster head selection mechanism to extend network lifespan. Compared to other analogous protocols, simulation findings proved that the EEUCLC mechanism improved energy balance and extended network longevity.

Azri et al. [30] developed a new method for organizing information from wireless sensor networks in a geographical database. A specific technique, 3D geo-clustering, was employed to address many challenges with sensor positioning in a three-dimensional metropolitan environment in a smart city. The method was created to reduce group cluster overlap. The method was evaluated through a series of tests in this study. Based on simulation findings, this method can balance node energy consumption and extend the network's life lifetime. Several tests were carried out to ensure that the approach for measuring database performance was effective.

Bindhu et al. [31] used artificial intelligence models to install subspace clustering in edge servers for the quick response and cost savings. The unnecessary and erroneous connections are pruned at the representation coefficient matrix using a pruning approach based on near neighbor (PSCN), a post-process method to improve clustering efficiency and decrease noise effects. Close neighbors with greater coefficients and common neighbors have stronger ties. A sparse matrix may be obtained by reversing the coefficients and ensuring the intra-subspace connections. The suggested technique is helpful in IoT-based earth observation systems and outperforms existing models.

Fouladlou et al. [32] introduced a novel routing method for IoT WSN devices to increase energy efficiency. In order to achieve optimal routing and network longevity, a well-known genetic algorithm is used to cluster sensor devices. Experimental findings showed that the suggested routing algorithm performs better than the IEEE 802.15.4 protocol.

Xu et al. [33] enhanced the LEACH protocol's clustered routing algorithm to balance the energy consumption of wireless sensor network nodes while extending the network's life cycle. The results showed that the modified LEACH algorithm consumes more energy than the contrast enhancement, has the minimum cluster head node power consumption, and has enhanced network connection and dependability.

Kumar et al. [34] presented a K-Means clustering-based resource allocation model for IoT to increase the quality of service (QoS) and coordinate resources. The response time achieved for transmission of a message from the beginning until the end was the main criterion for evaluation. The model's usefulness is revealed through simulation.

Khan et al. [35] offered a metric for measuring the performance of several query control strategies. It looked at how well the query control mechanism (QCM) performed for QoS-enabled layered-based clustering for reactive flooding in the IoT. This study used statistical methods to determine that the QCM algorithm beat the existing techniques for identifying and eliminating duplicate flooding queries.

4.2 Architecture

Table 3 shows the classification of the listed articles of the architecture in the clustering and IoT domain.

Gao et al. [36] created a comprehensive approach based on neural collaborative filtering (NCF) and fuzzy clustering to address QoS prediction in IoT environments. They created a fuzzy clustering algorithm to cluster contextual data and then developed a novel way for composite computing

Table 3 Analyzed articles in the architecture classification

Research	Challenges	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Gao et al. [36]	Quality-of-service (QoS)	Creating a comprehensive framework for tackling QoS prediction in an IoT environment	Effectiveness	–	Implementation	WSDream dataset	Fuzzy clustering algorithm	Exploring the influence of various neural network-based models in the QoS prediction task
Sekar et al. [37]	Traffic management	Proposing a technique for forecasting heavily crowded roadways	Effectiveness	–	Simulation and Implementation	CityPulse Dataset	Density-based clustering algorithm	–
Mohapatra et al. [38]	The essential criteria of every smart city project, which Wireless Sensor Networks meet, are unattended surveillance and data collecting	Proposing a rational dynamic cluster head scheme	Giving better classification accuracy Reducing false alarm rates	–	Simulation using OMNET ++	–	LEACH	Examination of existing fault diagnostic techniques for integrating dynamic defects in the network
Bharti et al. [39]	This is an unrealistic job in terms of interactions between humans and technologies	Proposed four-layered framework	More exact similarity searches Consuming less search time Takes minimum CPU throughput Increasing CPU's efficiency	Less security	Simulation and implementation using MATLAB	Collect a dataset	Agglomerative Fuzzy K-means (AFKM)	This architecture may be enhanced in the future to include security, privacy, and trust for real-world dispersed environments where resource connection is virtual. Additionally, software agents may be added to the security coordination and communication framework across several stations, allowing for resource allocation before query processing
Shuja et al. [40]	The processing of many geotextual data necessitates resource-efficient methods and frameworks	Proposed a clustering framework	Reducing leading towards memory and time efficiency	–	Implementation	Twitter API containing tweets dataset	–	Aim to examine 3D clustering while including social IoT temporal data
Azevedo et al. [41]	High network traffic is required to convey data from devices to a central node	Presented a centralized IoT data clustering approach	Reducing network traffic	–	Simulation and implementation	Ground truths dataset	–	–

Table 3 (continued)

Research	Challenges	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Kumar et al. [42]	Coordination between IoT devices to deliver diverse applications to the real world	Proposed two clustering algorithms based on heuristic and graph	Effectiveness	Not mention enough detail	Simulation using Cooja simulator	–	–	–
Jabeur et al. [43]	Making the most use of limited resources	Presented a novel firefly-based clustering method for IoT applications	Effectiveness	–	Simulation	–	Bio-inspired	Build the ASFiCA algorithm and evaluate the results of recommended techniques to that of current clustering algorithms like LEACH Working on fine-tuning the firefly approach so that RWTs may attract other devices based on different characteristics
Shukla et al. [44]	Creating a reliable and energy-efficient routing protocol	Proposed a multi-tiered clustering architecture	Better network lifetime Scalability Better energy efficiency	Not mention enough detail	Simulation	–	–	–
Effah et al. [45]	Self-configured sensor nodes	Proposed a multihop routing framework	Higher network energy-savings Scalability	Not mention enough detail	Simulation using MATLAB R2019b	–	–	–
Jiang et al. [46]	Many IoT systems need clustering toward a dynamic data stream	Exemplar-based data stream clustering	Effectiveness	Not mention enough detail No Simulation or implementation	No simulation or implementation	–	e-expansion move	-
Sharif et al. [47]	Network traffic	Presented an IoT-enabled routing strategy based on clusters	Choosing cluster heads with as few overheads as possible	–	Implementation using a 64bit system with a Core i3 processor and Simulation using NS 3		–	–
Yuan et al. [48]	Renewable energy source	Suggested a neural network-based DFIG	Increasing the system's stability is	–	Simulation using MATLAB/ Simulink	–	–	–

Table 3 (continued)

Research	Challenges	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Qureshi et al. [49]	Technological obstacles, system control and administration, scalability, and security problems	proposed three solutions for IoV challenges	Increasing data delivery ratio Increasing data throughput	–	Simulation using MATLAB	–	–	Create new solutions for other smart city applications to improve data connectivity, optimization, security, and privacy for users
Yin et al. [50]	Intelligent stadium system development	Created a structure that is centered on the athlete	Improving the management efficiency	–	Implementation	–	K-means	–

similarity. A new NCF model was created that used both local and global properties. Powerful tests were carried out on two real-world datasets, and the results showed that the suggested framework was successful. They plan to investigate the impact of various neural network-based models in the QoS prediction job as future work. They also aimed to examine the role of time in QoS prediction.

Sekar et al. [37] presented a method to anticipate traffic congestion on heavily populated highways based on current and previous traffic congestion. Alternative routes for the specified source and destination are also suggested by this method. After fuzzifying the data, density-based clustering provides weightage to the heavily packed segment on the route map. The weighting of a path at a specific point helps choose the ideal route. Floyd's approach was used to identify the shortest collection of alternative pathways from one point to another. The system then calculates the most convenient path for travel at the provided time and proposes it to the user. The authors provided the idea by displaying the numerous streets in the route and markings on the map for easy comprehension.

Mohapatra et al. [38] suggested a rational dynamic cluster-head approach, in which the cluster-head, like other nodes in the network, is prone to mistakes. At the end of each round, the LEACH protocol was changed to include smart, dynamic cluster-head selection. The suggested procedure improves accuracy and lowers false alarm rates for the various probability of periodic and irreversible failures compared to LEACH. Examining existing fault diagnostic techniques for integrating dynamic defects in the network will focus on future work. Under a comparable dynamically defective environment, the suggested fault diagnosis and distribution technique will be severely assessed.

Bharti et al. [39] proposed a four-layered framework:

1. An ontology was used to find resources and their related services automatically.
2. Manages resources by producing and representing knowledge.
3. Gives effective methods for indexing resources based on their highest similarity match
4. The duty of finding the near-optimal resource from a list of indexed resources delegated

The collected findings demonstrate that the framework can do in less time, better reliable match analyses. The framework was discovered to be consistent in providing accurate erroneous parameters resources and in assisting in the discovery of the correct resource with the computation of maximum resources. The framework used the smallest CPU throughput possible to execute requests, increasing CPU efficiency and reducing server load. This architecture may be enhanced in the future to include security, privacy, and trust for real-world dispersed environments where resource connection is virtual. Additionally, software agents may be added to the framework for security coordination and communication across several stations, allowing for resource allocation before query processing.

Shuja et al. [40] provided a two-step clustering of geo-textual data. Initially, all of the data points in the dataset were grouped based on their geographical placements simply using Euclidean distance. The text-similarity metric was founded on the Word-to-Vec model's unsupervised neural network training was utilized in the second stage to construct geo-textual clusters inside the geographical clusters. The two-step hierarchical clustering reduced clustering dimensionality. On the other hand, the hierarchical clustering framework only concedes a small amount of clustering quality. They want to take the hierarchical clustering framework variously. Furthermore, they intend to examine 3D clustering while using Social IoT temporal data.

Azevedo et al. [41] Proposed a centralized IoT data clustering approach requiring minimum network traffic and device processing power. The suggested technique reduces network traffic by using a data grid to aggregate information at the devices. Following data transmission, the presented technique employs a clustering algorithm designed to analyze data in a summary representation. Experiments revealed proof that the proposed technique minimizes network traffic while producing results equivalent to DBSCAN and HDBSCAN, two robust centralized clustering methods.

Kumar et al. [42] developed an application-based two-layer IoT architecture comprised of a sensing layer and an IoT layer Sensing devices are essential for every real-time application. Both of these layers are essential for the development of IoT-based applications. Any IoT-based application's effectiveness was reliant on the devices' and data gathered by the devices' efficient communication and utilization at both levels. The grouping of these devices aids in this goal, resulting in clusters of devices at different levels. They also offered two clustering techniques, one based on heuristics and the other on graphs. The suggested clustering algorithms are tested on an IoT platform with standard settings.

Jabeur et al. [43] provided a novel clustering technique for IoT applications based on fireflies. Clusters were refined during a micro clustering phase, in which they competed to incorporate small nearby clusters. They expand techniques to enable IoT clusters to self-adapt by employing current events and their anticipated effect on the network and deployment area. Results showed that clusters tend to stabilize irrespective of network density. In the future, they want to construct the ASFiCA algorithm and compare the effectiveness of their methods to that of specific current clustering algorithms, such as LEACH.

Shukla et al. [44] Proposed a multi-tiered clustering architecture and a unique energy-efficient routing protocol based on the subdivision method. They also created a zone and cluster creation technique that varies regarding the network's size to address a scalable, energy-efficient network. Communication across inter clusters at a short distance in a multi-hopping method for a broad area network was a key advantage of the suggested protocol. The scalable and energy-efficient routing protocol (SEEP) has been examined for network lifespan and energy depletion statistics for various distribution scenarios.

Effah et al. [45] presented a total communication cost metric-based energy-efficient multihop routing architecture and applied it in a multihop clustering-based agricultural IoT network. As the power-constrained clustering-based agricultural IoT network expands, a more real, power multihop route-actuating design will be necessary to meet the stated objective metrics in next-generation IoT applications. The suggested framework specified where IoT devices should be placed and how much electricity they should use. Compared to single-hop routing algorithms under identical conditions, results demonstrate that the proposed framework guarantees better performance.

Jiang et al. [46] Focused on dynamic exemplar-based clustering methods. They first presented a coherent interpretation for two prominent exemplar-based clustering models using the maximal a priori principle (AP). A novel exemplar-based data stream clustering technique known as DSC was presented. The suggested algorithm DSC has the particular advantage of allowing us to reuse the framework of the enhanced-expansion move algorithm by just changing the definitions of numerous variables, eliminating the need to develop a new optimization mechanism. Furthermore, the DSC method can handle two situations of similarity. Compared to AP and enhanced expansion move, results show that algorithm DSC can handle real-world IoT data streams.

Sharif et al. [47] suggested a cluster-based IoT-enabled routing approach according to a hybrid scenario incorporating static and dynamic transportation. Cluster-based routing is a mechanism for effectively sending packets through a network, even with low vehicle density. They implemented their approach using a 64bit system with a Core i3 processor and Simulated their approach using NS 3.

Yuan et al. [48] presented a DFIG based on neural networks and a super capacitor energy storage system (SCESS). The analysis was done in MATLAB/Simulink, and the results show a considerable increase in the system's stability. With the decline of non-renewable fuel sources, there has been a rise in non-renewable energy sources. Conventional wind turbine generation systems are not

Table 4 Analyzed articles in the application classification

Research	Challenge	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Yang et al. [51]	Due to the variable placement of sharing bike stations in different places, there is an unbalance in utilization	Presented a hierarchical model for sharing bike prediction	Improving the root-mean-squared logarithmic error Improvement in return prediction	Not mention enough detail for the proposed idea No simulation or implementation	No simulation or implementation	open-source bike-sharing datasets	XGBoost	–
Famila et al. [52]	Analyze enormous amounts of multimedia data, resulting in maximum energy consumption	Proposed an improved artificial bee colony optimization based clustering method	Reducing energy consumption rate Maximum lifetime	–	Simulation and implementation using Network Simulator (ns-2.34)	–	Artificial bee colony (ABC) algorithm	A sink node with mobility, or a large number of sink nodes, can help to decrease data collection and aggregation delays dramatically
Yang et al. [53]	Road traffic	Presented a high-performance computing model	Optimal planning of traffic network Low cost	–	Implementation	Use a deep belief network model	K-means	–
Muntean et al. [54]	Park occupancy problem	Found alternate solutions to the parking IoT occupancy problem	Effectiveness	Not mention enough detail about the proposed method No simulation or implementation	No simulation or Implementation	BHMBCCMKT01 dataset	k-Nearest K-means	–
Qureshi et al. [55]	Commodity single board computers	Design a framework for smart parking	Low-cost need Low energy usage	Security issue	Implementation	IoT datasets	–	Fix the prototype's dependability and security shortcomings
Zou et al. [56]	Flow data processing concept and its use in smart cities	Handling flow data from various sources in the smart city	Dynamic load balancing Recalculate the lost information specific time interval	Not mention enough detail	Implementation	Resilient distributed datasets	–	Evaluate the suggested method's performance in a greater number of cases

Table 4 (continued)

Research	Challenge	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Logesh et al. [57]	The integration of bio-inspired clustering to create optimum suggestions has yet to be investigated	Proposed a new user clustering method based on QPSO	Accuracy and efficiency		Implementation	Yelp TripAdvisor	PSO QPSO K-Means	Investigate the potential and consequences of user clustering models Nature-inspired intelligent optimization techniques for recommender system combinational optimization issues should be investigated
Lücken et al. [58]	Traffic routing	Presented a comprehensive strategy that integrates smart street lighting with traffic sensor technologies	Effectiveness	–	Implementation using Monte Carlo simulations	–	DBSCAN	Ultrasonic sensor arrays might be used to get directional and velocity data in the future
Tabatabai et al. [59]	Network overhead	Proposed an algorithm that will be used by the SDN controller	Better load distribution	Not mention enough detail	Simulation	–	–	They want to put the proposed heuristic to the test using real IoT data in large-scale tests on a cluster of computers in the future. They also plan to look at the feasibility of creating an online algorithm that guarantees worst-case achievement
Zhang et al. [60]	Current clustering approaches, which can just manage static data, can no longer cluster many data in dynamic industrial applications	Suggested an incremental clustering algorithm based on ICFSKM	Better accuracy and computational time	Not supported large-scale cloud platform	Implementation	UCI datasets	CFS clustering ICFSKM	The proposed techniques will be evaluated further on the large-scale cloud platform
Guo et al. [61]	Leakage of private details during the data-handling process	Suggested a K-means technique with privacy protection	High performance	–	Simulation using Python	Collect a dataset	K-means	Create a new K-means algorithm to minimize communication complexity using a novel privacy concept

Table 4 (continued)

Research	Challenge	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Chithaluru et al. [62]	Sensor-based connectivity in smart cities should enhance	Proposed an improved-adaptive ranking based energy-efficient opportunistic routing protocol	Maximizing the network lifetime	–	Simulation and implementation using MATLAB	–	NR-LEACH	Machine learning methods can be used regularly by cluster member nodes to categorize sensor data based on similarity
Xu et al. [63]	Wireless Sensor networks have limited energy resources	Demonstrated a smart clustering technique	Better performance	-	Simulation using MATLAB	–	LEACH	Enhance their algorithm so that it can be used in more realistic systems: Energy storage situation with a wide range of possibilities Communication interfaces on nodes that are heterogeneous The financial cost and the balance between energy efficiency and quality of life
Rani et al. [64]	Energy conservation	Suggested a GA-based dynamic clustering routing technique	Better throughput	Not mention enough details	Simulation using MATLAB	–	Genetic algorithm	–
Bellaouar et al. [65]	Developed effective vehicle communication	Proposed a VANET clustering solution	Reducing the communication Increasing the level of stability	–	Simulation using OMNET simulator	–	–	–
Darabkh et al. [66]	Resource consumption	Proposed an energy-aware clustering and routing protocol	Increasing network lifespan Better resource consumption	–	Simulation using MATLAB	–	Genetic algorithm	In the future, adapting and assessing alternative software-defined network controllers will be a fantastic addition to this effort. Furthermore, determining the impact of IoT sensor energy heterogeneity on the lifetime of IoT networks would be a helpful study issue
Zheng et al. [67]	In cognitive radio sensor networks, manage communications	A novel clustering protocol	Better network stability and energy consumption	–	Simulation	–	–	–

Table 4 (continued)

Research	Challenge	Main idea	Advantage	Disadvantage	Assessment environments	Dataset	Used algorithm	Future work
Tripathi et al. [68]	Data analyzing tools and algorithms with high performance are required	Provided a novel meta-heuristic-based clustering algorithm	More accuracy	–	Simulation	Susy Pokerhand DLePM sIoT IoT_Bonet	–	Future research will include the application of the suggested technique to other real-world IoT and big data applications
Zhang et al. [69]	The lifetime of the industrial wireless sensor network	A grid-based clustering algorithm	Balancing energy depletion effectively Extending the whole network lifetime	–	Simulation	–	–	–

ecologically friendly; thus, this study presented a smart city-based system paradigm to address this issue.

Qureshi et al. [49] attempted to address at least three aspects of smart cities by designing three nature-inspired solutions. The dragonfly, moth flame, and ant colony optimization methodologies were used to develop these solutions. A simulator is used to verify the efficiency of the offered solutions. These solutions will aid future investigators in exploring nature-inspired solutions to address smart cities' new and complicated systems.

Yin et al. [50] used a K-means clustering technique of customer classification to develop an athlete-centered system to increase the athletes' service quality. The findings suggested that using IoT technology to construct and design an intelligent stadium system may assist in achieving intelligent stadium management and increase the intelligent stadium's management efficiency and provide a favorable application benefit. The study's findings demonstrated that the construction and design of an intelligent stadium system could benefit from IoT technology.

4.3 Application

Table 4 shows the classification of the listed articles of the application in the clustering and IoT domain.

Yang et al. [51] suggested a hierarchical method for predicting the returns numbers from each sharing bike station to accomplish resource resharing. The proposed model is made up of two stages:

1. Sharing bicycle station clustering by learning network display based on movement patterns and geographical location details of sharing bicycles among stations
2. Bicycle sharing station clustering by learning to display the network based on migration trends and geographical location information of bicycle sharing between stations.

The total number of bike-sharing stations was projected in the hierarchical prediction stage. Eventually, each station's returns number may be calculated. Their strategy was tested against numerous baseline methods using two publicly accessible sharing data sets. Findings showed that the suggested hierarchy could produce the best forecast outcomes.

Famila et al. [52] presented an Improved artificial bee colony optimization based clustering (IABCOCT) method by combining the advantages of the Grenade explosion method (GEM) with the Cauchy operator. They used network simulator (ns-2.34) to evaluate their presented idea. Reducing energy consumption rate and maximum lifetime are the advantages of the proposed algorithm. In future work, a sink node with mobility, or many sink nodes, can be applied to dramatically decrease data collection and aggregation delays.

Yang et al. [53] provided a high-performance computing model for dynamic traffic planning. They suggested transportation planning based on real-time IoT, which DBN and K-means analyze to provide a correct answer close to reality and fit the needs of computing with superior efficiency and low cost. This research used the problem of hotels in Tianjin as an example to determine the best dynamic traffic network design solution. The result revealed that the model, which was based on super high computing performance, was beneficial for optimal traffic network planning in real-time mass data circumstances at a cheap cost and stimulating the building and growth of smart cities.

Muntean et al. [54] proposed alternate solutions to the Birmingham car park overcrowding problem. Their method entails clustering the dataset first to extract key time intervals during a day and then forecasting data within these clusters. The greatest forecast rates are achieved by dividing data into six groups and using the k-Nearest Neighbor approach to estimate car park occupancy. Their results demonstrate that dividing data into six groups using the k-Nearest Neighbor approach to estimate car park occupancy yields the best forecast rates. This technology may also be used to address other smart city challenges like public transportation, traffic management, and smart lighting.

Qureshi et al. [55] presented a container-based architecture for a smaller footprint because single-board computers (SBCs) are resource-constrained devices. They used a proof-of-concept SBC-based Edge cluster for a smart parking application as a prospective IoT use-case to verify methodology. Their work demonstrates that the proposed architecture provides a low-cost, low-energy green computing option. The suggested framework may be used for other cloud-based applications. The prototype's successful deployment highlights the low-cost advantage of installing a cluster like this in smart city applications. In future work, they plan to fix the prototype's dependability and security shortcomings.

Zou et al. [56] examined the concept of stream data management and its applicability in smart cities by applying a cluster analysis technique. Cluster computing was used to manage data flow from diverse sources, and this suggested research delivers data flow in the smart city utilizing cluster computing. The findings compare the programs in the same environment on both Apache Spark and Hadoop technologies. The protocols of the real clusters were reviewed, and the difficulties they face when deployed in limited situations. In the future, they will test the performance of the proposed method in a larger number of scenarios.

Logesh et al. [57] developed a novel user clustering strategy based on quantum-behaved particle swarm optimization (QPSO) for a recommender system based on collaborative filtering. The collected findings show that the suggested strategy outperforms current peer research. They presented a novel mobile recommendation framework for urban trip recommendations in smart cities. The assessment findings show that the generated recommendations are useful and that users are satisfied with the suggested recommendation technique. They want to include users' online activity in user clustering in the future, and we'll investigate the potential and consequences of such user clustering models. They also intend to conduct a comprehensive investigation to develop a powerful clustering ensemble model that incorporates user behavior to generate reliable recommendations.

Lücken et al. [58] integrated smart street lighting with traffic sensing technologies in an integrated strategy. Multilane traffic participant identification and categorization are performed using infrastructure-based ultrasonic sensors installed with a street light system. In the past, using infrastructure-based ultrasonic sensors in reflection moment settings constituted an unsolved difficulty for many acoustic detection technologies, limiting their use. They propose a method based on an algorithm that blends statistical standardization with unsupervised learning clustering approaches. The assessment was based on data taken on a vehicle test track and real-world deployments across Europe.

Tabatabai et al. [59] presented a software-defined networking (SDN) controller technique for minimizing the load discrepancies across brokers while maintaining a restriction on reconfiguration in defense of data and decision integration scenarios. The limitation of inseparable topics minimized the variation in brokers' load within a reconfiguration cost. The suggested algorithm was assessed using realistic simulated traffic traces compared to a baseline heuristic based on instantaneous topic data. The suggested heuristic performs better load distribution. They want to test the proposed heuristic in the future using real IoT data in large-scale trials on a cluster of computers. They also intend to investigate the possibility of developing an algorithm that ensures worst-case performance.

Zhang et al. [60] proposed an incremental clustering technique based on quickly discovering and searching density peaks (ICFSKM) based on k medoids. Cluster forming and cluster merging are specified in the proposed method to merge the recent pattern with the old one for the ultimate clustering result. Finally, tests were carried out to verify the proposed approach regarding clustering accuracy and processing time. The suggested techniques' performance was first examined on a cloud platform with just 10 servers in this research. The results proved the suggested methods' potential scalability on the cloud platform. Future work will further validate the proposed approaches in the large-scale cloud platform.

Guo et al. [61] offered a mutual privacy-preserving K-means technique (M-PPKS) based on homomorphic encryption to ensure that participants' privacy and cluster center's private data will not be exposed. The suggested M-PPKS method separates each iteration of a K-means algorithm into two steps:

1. locating the cluster center closest to each participant
2. calculating a new cluster center

The cluster center was kept secret from participants in both stages, and no analyst had access to any of the participants' personal information. The suggested M-PPKS technique may reach high performance, according to extensive privacy analysis and performance assessment findings. Furthermore, it may effectively generate approximation clustering results while maintaining private data.

Chithaluru et al. [62] suggested an improved-adaptive ranking based energy-efficient opportunistic routing protocol (I-AREOR) according to geographical density, distance, and surviving energy. As a result, the proposed technique considered the geographical density, distance, and remaining energy of the sensor nodes to give a solution to lengthen the period of first node death. The I-AREOR protocol examines energy for each round depending on the dynamic threshold. The exhibited findings suggest that, compared to current techniques, the I-AREOR clustering methodology is more effective in optimizing network lifespan.

Xu et al. [63] presented a clustering technique (Smart-BEEM) based on BEE(M) to achieve energy-efficient and QoE-enabled connectivity in cluster-based IoT networks. It's a user-centered and context-aware technique that makes it easier for IoT devices to select the best platforms for interaction and cluster headers for data transfer. The findings showed that Smart-BEEM could boost the effectiveness of BEE and BEEM in terms of penetration sensitive durability.

Rani et al. [64] described a variety of IoT applications to investigate that dynamic clustering-based IoT may effectively serve real-world applications. The suggested technique uses a dynamic clustering-based approach and framework base stations to select the cluster node with the more preferred sensor node. Experiments show that the presented approach outperforms the dynamic clustering relay node clustering algorithm.

Bellaouar et al. [65] presented the "QoSCluster" VANET clustering solution, which strived to keep the clusters while meeting the network's quality of service standards. They used the Veins platform, the OMNET simulator, and the SUMO realistic mobility model to test their protocol. Reducing communication and increasing the level of stability are the main advantages of the presented idea.

Darabkh et al. [66] introduced LiMAHP-G-C, an energy-aware clustering and routing protocol that tries to mitigate the obstacles that may arise during the communication process of heterogeneous smart devices. The simulation findings reveal that their suggested protocol beats other current studies in terms of network lifespan, resource consumption, and scalability measures. Adapting and analyzing different software-specified network controllers will greatly complement this work in the future. Furthermore, evaluating the influence of IoT sensor energy heterogeneity on IoT network longevity would be a worthwhile research topic. Finally, implementing the mobility option will pose a significant challenge and serve as a case study for many IoT researchers.

Zheng et al. [67] presented a technique for CRSNs that is network stability-aware clustering (NSAC). The protocol design of NSAC incorporates both spectrum dynamics and energy usage simultaneously. According to extensive simulations, the proposed NSAC protocol clearly surpasses existing techniques in terms of network stability and energy usage.

Tripathi et al. [68] provided a novel meta-heuristic-based clustering algorithm that takes advantage of MapReduce's strengths to handle massive data challenges. The suggested approaches use the military dog squad's searching capabilities to find the best centroids and the MapReduce architecture to handle large data sets. In addition, a MapReduce-based parallel variant of the suggested approach (MR-MDBO) was provided for clustering large datasets generated by industrial IoT. Furthermore, the performance of MR-MDBO was investigated. Experiments showed that the provided MR-MDBO-based clustering beats the other clustering accuracy and computation time methods.

Zhang et al. [69] presented a grid-based clustering approach for industrial IoT using load analysis. The network load is statistically studied first, followed by creating a load model. A series of phrases is also inferred to represent the network load distribution. The total number of packets delivered at each level is proportional. Finally, the network was divided into unequal grids based on the ideal cluster size, with each grid's nodes forming a cluster. Results showed that the suggested approach performs better.

Table 5 shows the evaluation factors in the clustering and IoT domain.

Table 5 Evaluation factors in clustering and IoT

Research	Execution time	Number of Clusters	Energy	Delay	Lifetime	Throughput	Accuracy
Srinivas et al. [19]	✓	✓					
Alazab et al. [20]			✓	✓			
Qin et al. [21]	✓			✓			
Liang et al. [22]	✓						
Ghoneim et al. [23]		✓					
Feng et al. [24]		✓			✓	✓	
Mydhili et al. [25]	✓		✓				✓
Han et al. [26]		✓		✓			
Wang et al. [27]			✓				
Azri et al. [28]							✓
Baniata et al. [29]			✓		✓		
Azri et al. [30]		✓	✓				
Bindhu et al. [31]							✓
Fouladlou et al. [32]			✓	✓		✓	
Xu et al. [33]			✓				
Kumar et al. [34]		✓					
Khan et al. [35]				✓			
Gao et al. [36]							✓
Sekar et al. [37]			✓				
Mohapatra et al. [38]	✓		✓			✓	✓
Bharti et al. [39]	✓					✓	
Shuja et al. [40]	✓						
Azevedo et al. [41]							✓
Kumar et al. [42]		✓					
Jabeur et al. [43]		✓					
Shukla et al. [44]			✓		✓		
Effah et al. [45]			✓				
Jiang et al. [46]		✓	✓				
Sharif et al. [47]		✓					
Yuan et al. [48]				✓			✓
Qureshi et al. [49]		✓					
Yin et al. [50]							✓
Yang et al. [51]							✓
Famila et al. [52]		✓	✓				
Yang et al. [53]	✓						
Muntean et al. [54]	✓	✓					
Qureshi et al. [55]	✓						
Zou et al. [56]	✓	✓					
Logesh et al. [57]		✓					✓
Lücken et al. [58]	✓						
Tabatabai et al. [59]	✓						
Zhang et al. [60]	✓						✓
Guo et al. [61]		✓					
Chithaluru et al. [62]		✓	✓				
Xu et al. [63]		✓	✓				
Rani et al. [64]						✓	
Bellaouar et al. [65]					✓		
Darabkh et al. [66]			✓		✓		
Zheng et al. [67]			✓				
Tripathi et al. [68]		✓		✓			
Zhang et al. [69]			✓				

5 Discussion and comparison

The approach for analyzing the selected research in clustering and IoT was outlined in the previous sections. Furthermore, we categorized and evaluated the research depending on several factors, including the algorithm used, simulation or implementation, future work, datasets, advantages, disadvantages, evaluation factors, etc. Clustering and IoT categories are broken down statistically and compared in this section. We also discuss IoT intrusion detection systems as future work direction of IoT in this section.

5.1 IoT intrusion detection systems (IDS)

The IoT attempts to connect queries through the Internet, whereas SDN provides managed services coordination via detaching the handle plane and the data plane.

The IoT has arisen as a distinct type of wireless sensor network (WSNs) in the last decade. Software-defined networking (SDN) is a new paradigm for organizing and controlling the enormous volumes of data created by IoT devices. It decouples the data plane from the control plane of network devices, allowing for simple deployment and maintenance. In addition, to improve and protect the SDN-IoT network, network function virtualization (NFV) has evolved. It allows network devices to be virtualized and implemented as software components.

Furthermore, SDN-WSN is vulnerable to security threats, negatively affecting the system's vulnerability and QoS performance. As a result, the massive implementation of WSNs is complex and problematic; managers must employ unique adaptation methods when using specific applications that necessitate adaptability and administration. The integration of SDN into WSNs is proposed, and a new model has been created and appears to address these issues.

One future research work direction is considering intrusion detection systems in IoT [70, 71].

An IDS is a tool that monitors data flow to detect and guard against an invasion that compromises an information system's confidentiality, integrity, and availability. An IDS's operations may be broken down into three parts. The stage of observation, which uses network-based or host-based sensors, is the initial step. The second stage is the evaluation step, which depends on feature extraction or pattern recognition algorithms. The final level is the detection stage, which depends on the oddity or abuse discovery of intrusions. Security and privacy are critical issues in every IoT architecture based on every smart infrastructure. Intelligent devices are in danger due to security issues in IoT-based platforms. In conclusion, IDSs are crucial for combating security attacks related to IoT that target a handful of these security weaknesses. Traditional IDSs may not be a solution for IoT devices' limited CPU and storage capabilities, and the particular protocols employed [72, 73].

The security of IoT applications has become a key problem as the variety of providers and users in IoT networks grows. As IoT technologies and smart environments are combined, smart things become more effective. However, in essential smart environments, the consequences of IoT security flaws are quite serious. There will be a concern for systems in IoT-based smart environments without sufficient security mechanisms.

There are many ways to classify IDS types, especially IDS for IoT, because most of them are still under research.

- Anomaly-based IDS (AIDS)
- Host-based IDS (HIDS)
- Network-based IDS (NIDS)
- Distributed IDS (DIDS)

- IoT IDS deployment strategies

IDS can also be classified according to how it detects IoT threats. IDS can be categorized as distributed, centralized, or hierarchical in IDS deployment strategies [74].

Distributed IDS: A distributed IDS consists of multiple IDS scattered over a large IoT ecosystem that connects with a central server. Distributed architectures are used by a number of IDSs. A portion of the network is tasked with inspecting the other nodes. Compared to centralized IDS, distributed IDS provides various benefits to incident analysts. The primary benefit is recognizing attack types across a whole IoT ecosystem. This might help boost the speed with which IoT attacks are detected and prevented [74].

Centralized IDS: The IDS is deployed on central devices in the centralized IDS location. As a result, the packets exchanged between IoT devices and the network may be checked by an IDS placed on a boundary switch. However, inspecting network packets passing through the border switch is insufficient to detect abnormalities affecting IoT devices. In a centralized IDS, network traffic is monitored. Different network data sources are used to extract this traffic from the network. A network-based IDS can keep track of all the machines on a network [74].

Hierarchical IDS: The network is divided into clusters in a hierarchical IDS. Sensor nodes that are close together are usually part of the same cluster. Each cluster has a cluster head, which observes the network nodes and participates in network-wide investigations [74].

- Synchronization between multiple IoT IDS

The IoT is predicted to transform consumers' daily lives by allowing data exchange between widespread objects

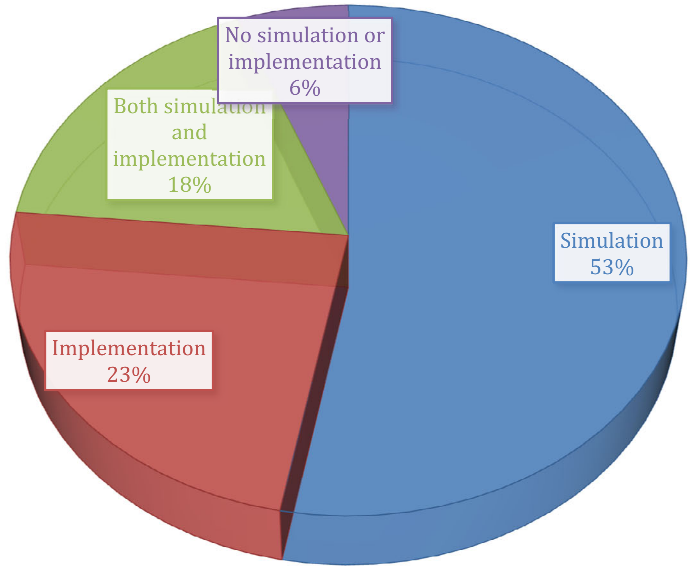

Fig. 11 Percentage of evaluation environments presented

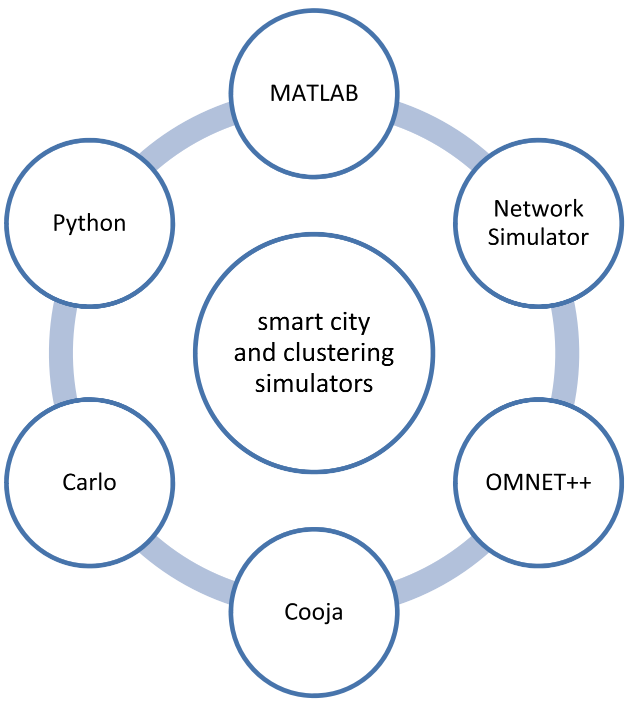

Fig. 12 Simulator platforms

over the Internet. However, because networks are entities with widely different resources, such a broad goal imposes prohibitive limits on applications requiring time-synchronized activity for historical data sorting or simultaneous execution of specific activities. On the one hand, current time synchronization solutions are difficult to adapt to resource-limited devices. On the other hand, solutions for restricted systems do not scale well to diverse placements. Time synchronization is a critical middleware service for distributed systems and the distributed IDS [75].

5.2 Research questions

In addition, the following analytical reports on the Sect. 2 research questions were posed:

5.2.1 RQ1: Which clustering algorithms were used in the smart city domain?

According to studies and literature reviews, clustering algorithms used in the smart city are listed below.

- K-means
- K-means ++
- K-medoids
- CFS clustering
- ICFSKM
- Mean-shift clustering algorithm
- Low-energy adaptive clustering hierarchy(LEACH)
- NR-LEACH algorithm
- Fuzzy clustering algorithm
- Agglomerative Fuzzy K-means (AFKM)
- Bio-inspired
- Density-based spatial clustering of applications with noise(DBSCAN)
- Artificial bee colony (ABC) algorithm
- Density-based clustering algorithm

5.2.2 RQ2: How might data transmission be reduced by using clustering in the smart city?

Data mining is one of the approaches for analyzing large amounts of data and extracting information and knowledge from it. Various techniques extract the pattern from this data, each with its own set of applications. Clustering

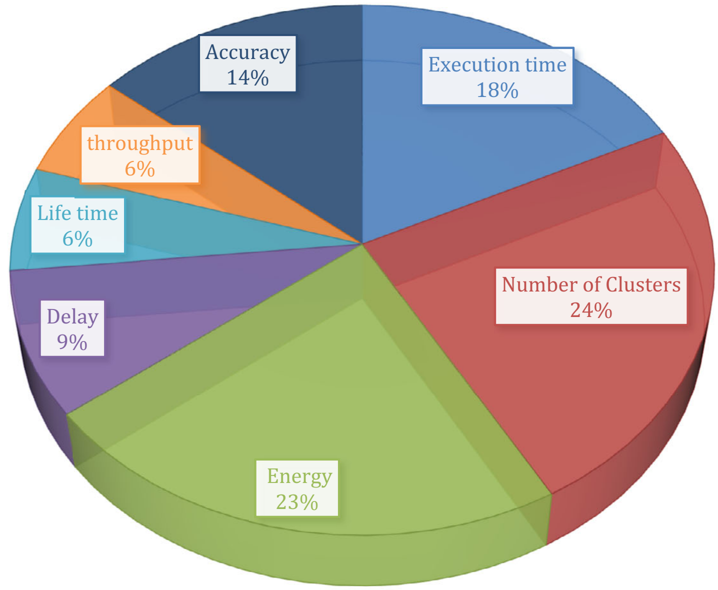

Fig. 13 Evaluation factors in clustering in the context of the smart city

algorithms are one of the most essential and extensively used data mining tools. Similar inherent similarities are used to group similar data in clustering or categorization. And it identifies the pattern and extraction hidden in the essence of data based on this categorization and similarity. Finding these patterns also makes data administration for various applications relatively simple. Clustering is a type of unsupervised learning. A database containing data without labeling the target or group to which the data belongs is referred to as an unsupervised learning approach. It is used to find a meaningful structure or pattern for categorizing data in general [34, 35, 40].

Clustering divides a group of data points into several categories. The data points in each group are the most similar and are not comparable to data points in other groups. A set of items or data is sorted into groups based on their similarity and dissimilarity; this is a significant task for data mining and evaluating large amounts of data. Clustering algorithms divide data into distinct groups called clusters based on their related qualities. When presented with a tiny collection with minimal features, classifying it is simple. Consider the following scenario: you are faced with a massive collection of thousands of data sets, and you are tasked with categorizing them; this is a challenging and time-consuming task for humans. Clustering algorithms are the greatest instrument for handling such issues since classification work with many characteristics is beyond human tolerance. When there are a lot of data attributes, this method is used. Clustering algorithms perform analysis that is highly different from human data categorization. These algorithms have a thorough grasp of the data, cluster formation and discover a cluster quickly. Clusters with small distances between cluster members, clusters with high data density, distances, and certain statistical distributions are among the qualities on which these algorithms form these clusters [22, 39, 57].

5.2.3 RQ3: What evaluation environments are used to evaluate clustering in the smart city?

We discovered that 23% of the research studies had implemented the proposed idea, and 53% of the study papers used simulation methods to evaluate the presented idea. We can point to MATLAB simulation platforms as popular simulators. In addition, 6% of the papers did not include any implementation or simulation, but 18% of articles use both simulation and implementation methods, as shown in Fig. 11.

Figure 12 shows the simulator platforms that are used in this concept.

5.2.4 RQ4: What are the evaluation factors in clustering domains in the smart city context?

Figure 13 compares the evaluation factors in the domains of clustering in the smart city context. The statistical percent of evaluation factors demonstrates that the number of clusters has a high percentage of 24%, the energy factor has 23%, the execution time factor has 18%, the accuracy has 14%, the delay has 9%, the lifetime has 6%, and throughput has 6%.

5.2.5 RQ5: What are the open issues and future challenges of clustering in the smart city?

(a) Maintenance of clusters

Network maintenance is the desired role in the IoT network architecture. Therefore, finding the nearest nodes in a given network using an effective synchronization mechanism is complicated. Because nodes in the network are not static and do not have a set position, cluster stability is one of the challenges. Because the environment in IoT networks is always changing, it is vital to do network maintenance. The most challenging difficulty in a network with movable nodes is determining the best cluster head to maintain a long-lasting network.

(b) Redundancy of data

Because numerous sensor nodes in IoT approaches may send redundant data to the cluster head, the rapid expansion of IoT devices and sensors in smart cities creates massive data. Using a data fusion/aggregation approach may mitigate this type of problem.

(c) Multiple paths

Another issue that arises when using cluster-based methods to transmit data is that when several pathways exist, the capability of the paths is generally reduced due to the interference created by the close paths. As a result, the packet loss ratio rises, causing network performance to suffer.

(d) Management of data

The amount of data created by the IoT and its applications like the smart city is enormous and constantly changing. As a result, obtaining and managing valuable data in such a large setting is too tough. Handling this dynamic, as well as the variability of data types, is a significant challenge. Because data is the foundation of every IoT industry, it must be kept safe, secure, and well-managed.

(e) Multiple connectivity

Every device in the IoT applications must be connected. As a result, numerous devices are connected in actual IoT scenarios, each with its communication standard. It is challenging to establish seamless connectivity due to the heterogeneity of communication protocols and interfaces.

6 Conclusion

This survey focuses on clustering and IoT, particularly the smart city approach. The SLR-based method was discovered by an exploratory search of 144 publications published between 2017 and 2021. Finally, we investigated 51 papers about IoT and clustering. It is possible that not every research on the SLR-based method has been examined. We conducted a significant study on clustering and IoT using the findings of over 100 authors and papers. Due to the increasing number of studies completed in this field, it is impossible to verify that all studies have been covered by the research ended in 2021.

Acknowledgements Not applicable.

Funding No funding was received.

Data availability Not applicable.

Declarations

Conflict of interest There is no conflict of interest.

References

1. Balakrishna, S., Thirumaran, M.: Semantics and clustering techniques for IoT sensor data analysis: a comprehensive survey. Princ. Internet Things (IoT) Ecosyst. **2020**, 103–125 (2020)
2. Bangui, H., Ge, M., Buhnova, B.: A research roadmap of big data clustering algorithms for future internet of things. Int. J. Organ. Collect. Intell. (IJOCI) **9**(2), 16–30 (2019)
3. Mahdavinejad, M.S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., Sheth, A.P.: Machine learning for internet of things data analysis: a survey. Digit. Commun. Netw. **4**(3), 161–175 (2018)
4. Kousis, A., Tjortjis, C.: Data mining algorithms for smart cities: a bibliometric analysis. Algorithms **14**(8), 242 (2021)
5. Tharwat, M., Khattab, A.: Clustering techniques for smart cities: an artificial intelligence perspective. In: Smart Cities: A Data Analytics Perspective, pp. 113–134. Springer, Cham (2021)
6. Bangui, H., Ge, M. and Buhnova, B.: Exploring big data clustering algorithms for internet of things applications. In: IoTBDS. pp. 269–276, 2018
7. Mahyastuty, V.W., Hendrawan, H., Iskandar, I. and Arifianto, M.S.: Survey of clustering techniques in internet of things architecture. In: 2020 14th International Conference on Telecommunication Systems, Services, and Applications TSSA, pp. 1–4. IEEE, 2020
8. Belhadi, A., Djenouri, Y., Nørvåg, K., Ramampiaro, H., Masseglia, F., Lin, J.-W.: Space–time series clustering: algorithms, taxonomy, and case study on urban smart cities. Eng. Appl. Artif. Intell. **95**, 103857 (2020)
9. Sowmya, R., Suneetha, K. R.: Data mining with big data. In: 2017 11th International Conference on Intelligent Systems and Control (ISCO), pp. 246–250. IEEE, 2017
10. Luckey, D., Fritz, H., Legatiuk, D., Dragos, K., Smarsly, K.: Artificial intelligence techniques for smart city applications. In: International Conference on Computing in Civil and Building Engineering, pp. 3–15. Springer, Cham (2020)
11. Lytras, M.D., Visvizi, A., Sarirete, A.: Clustering smart city services: Perceptions, expectations, responses. Sustainability **11**(6), 1669 (2019)
12. Sholla, S., Kaur, S., Begh, G.R., Mir, R.N., Chishti, M.A.: Clustering internet of things: a review. J. Sci. Technol. **3**(2), 21–27 (2017)
13. Torabi, M., Hashemi, S., Saybani, M.R., Shamshirband, S., Mosavi, A.: A Hybrid clustering and classification technique for forecasting short-term energy consumption. Environ. Prog. Sustain. Energy **38**(1), 66–76 (2019)

14. Taherei Ghazvinei, P., Hassanpour Darvishi, H., Mosavi, A., Yusof, K.B.W., Alizamir, M., Shamshirband, S., Chau, K.W.: Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. Eng. Appl. Comput. Fluid Mech. **12**(1), 738–749 (2018)
15. Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., Mosavi, A.: Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. IEEE Access **8**, 150199–150212 (2020)
16. Bonaccorso, G.: Machine Learning Algorithms. Packt Publishing Ltd, Birmingham (2017)
17. Puliafito, A., Tricomi, G., Zafeiropoulos, A., Papavassiliou, S.: Smart cities of the future as cyber physical systems: challenges and enabling technologies. Sensors **21**(10), 3349 (2021)
18. Bernard, A.: Solving interoperability and performance challenges over heterogeneous IoT networks: DNS-based solutions. PhD diss., Institut polytechnique de Paris, 2021
19. Srinivas, K. G., Hosahalli, D.: Evolutionary computing assisted K-Means clustering based mapreduce distributed computing environment for IoT-driven smart city. In: 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 192–200. IEEE, 2021
20. Alazab, M., Lakshmanna, K., Reddy, T., Pham, Q.-V., Maddikunta, P.K.R.: Multi-objective cluster head selection using fitness averaged rider optimization algorithm for IoT networks in smart cities. Sustain. Energy Technol. Assess. **43**, 100973 (2021)
21. Qin, Z., Xu, F., Xie, Y., Zhang, Z., Li, G.: An improved top-K algorithm for edge servers deployment in smart city. Trans. Emerg. Telecommun. Technol. **32**, e4249 (2021)
22. Liang, T.: Cascaded fault diagnosis based on weighted computer network in smart city environment considering focusing fuzzy clustering algorithm. In: 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), pp. 178–182. IEEE, 2019
23. Ghoneim, O.A.: Traffic jams detection and congestion avoidance in smart city using parallel K-Means clustering algorithm. In: Proceedings of International Conference on Cognition and Recognition, pp. 21–30. Springer, Singapore (2018)
24. Feng, X., Zhang, J., Ren, C., Guan, T.: An unequal clustering algorithm concerned with time-delay for internet of things. IEEE Access **6**, 33895–33909 (2018)
25. Mydhili, S.K., Periyanayagi, S., Baskar, S., Mohamed Shakeel, P., Hariharan, P.R.: Machine learning based multi scale parallel K-means++ clustering for cloud assisted internet of things. Peer-to-Peer Netw. Appl. **13**(6), 2023–2035 (2020)
26. Han, G., Wu, J., Wang, H., Guizani, M., Ansere, J.A., Zhang, W.: A multicharger cooperative energy provision algorithm based on density clustering in the industrial internet of things. IEEE Internet Things J. **6**(5), 9165–9174 (2019)
27. Wang, Z., Qin, X., Liu, B.: An energy-efficient clustering routing algorithm for WSN-assisted IoT. In: 2018 IEEE wireless communications and networking conference (WCNC), pp. 1–6. IEEE, 2018
28. Azri, S., Ujang, U., Abdul Rahman, A.: Dendrogram clustering for 3D data analytics in smart city. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **42**, 247–253 (2018)
29. Baniata, M., Hong, J.: Energy-efficient unequal chain length clustering for wireless sensor networks in smart cities. Wirel. Commun. Mob. Comput. (2017). https://doi.org/10.1155/2017/5790161
30. Azri, S., Ujang, U., Abdul Rahman, A.: 3D geo-clustering for wireless sensor network in smart city. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **42**, 11–16 (2019)
31. Bindhu, V., Ranganathan, G.: Hyperspectral image processing in internet of things model using clustering algorithm. J. ISMAC **3**(02), 163–175 (2021)
32. Fouladlou, M., Khademzadeh, A.: An energy efficient clustering algorithm for wireless sensor devices in internet of things. In: 2017 Artificial Intelligence and Robotics (IRANOPEN), pp. 39–44. IEEE, 2017
33. Xu, Y., Yue, Z., Lv, L.: Clustering routing algorithm and simulation of internet of things perception layer based on energy balance. IEEE Access **7**, 145667–145676 (2019)
34. Kumar, S.. Raza, Z.: A K-means clustering based message forwarding model for internet of things (IoT). In: 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 604–609. IEEE, 2018
35. Khan, F.A., Noor, R.M., Kiah, M.L., Ahmedy, I., Mohd Yamani Idna, I., Soon, T.K., Ahmad, M.: Performance evaluation and validation of QCM (query control mechanism) for QoS-enabled layered-based clustering for reactive flooding in the internet of things. Sensors **20**(1), 283 (2020)
36. Gao, H., Yueshen, Xu., Yin, Y., Zhang, W., Li, R., Wang, X.: Context-aware QoS prediction with neural collaborative filtering for internet-of-things services. IEEE Internet Things J. **7**(5), 4532–4542 (2019)
37. Sekar, E.V., Anuradha, J., Arya, A., Balusamy, B., Chang, V.: A framework for smart traffic management using hybrid clustering techniques. Clust. Comput. **21**(1), 347–362 (2018)
38. Mohapatra, A.D., Sahoo, M.N., Sangaiah, A.K.: Distributed fault diagnosis with dynamic cluster-head and energy efficient dissemination model for smart city. Sustain. Cities Soc. **43**, 624–634 (2018)
39. Bharti, M., Jindal, H.: Optimized clustering-based discovery framework on internet of things. J. Supercomput. **77**, 1739–1778 (2021)
40. Shuja, J., Humayun, M.A., Alasmary, W., Sinky, H., Alanazi, E., Khan, M.K.: Resource efficient geo-textual hierarchical clustering framework for social IoT applications. IEEE Sens. J. **21**, 25144 (2021)
41. Azevedo, R.D., Machado, G.R., Goldschmidt, R.R., Choren, R.: A reduced network traffic method for IoT data clustering. ACM Trans. Knowl. Discov. Data **15**(1), 1–23 (2020)
42. Kumar, J.S., Zaveri, M.A.: CCstering approaches for pragmatic two-layer IoT architecture. Wirel. Commun. Mob. Comput. (2018). https://doi.org/10.1155/2018/8739203
43. Jabeur, N., Yasar, A.-H., Shakshuki, E., Haddad, H.: Toward a bio-inspired adaptive spatial clustering approach for IoT applications. Future Gener. Comput. Syst. **107**, 736–744 (2020)
44. Shukla, A., Tripathi, S.: A multi-tier based clustering framework for scalable and energy efficient WSN-assisted IoT network. Wirel. Netw. **26**(1), 23 (2020)
45. Effah, E., Thiare, O., Wyglinski, A.: Energy-efficient multihop routing framework for cluster-based agricultural internet of things (CA-IoT). In: 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), pp. 1–5. IEEE, 2020
46. Jiang, Y., Bi, A., Xia, K., Xue, J., Qian, P.: Exemplar-based data stream clustering toward internet of things. J. Supercomput. **76**(4), 2929–2957 (2020)
47. Sharif, A., Li, J.P., Saleem, M.A., Saba, T., Kumar, R.: Efficient hybrid clustering scheme for data delivery using internet of things enabled vehicular ad hoc networks in smart city traffic congestion. J. Internet Technol. **21**(1), 149–157 (2020)
48. Yuan, Z., Wang, W., Fan, X.: Back propagation neural network clustering architecture for stability enhancement and harmonic suppression in wind turbines for smart cities. Comput. Electr. Eng. **74**, 105–116 (2019)
49. Qureshi, K.N., Ahmad, A., Piccialli, F., Casolla, G., Jeon, G.: Nature-inspired algorithm-based secure data dissemination framework for smart city networks. Neural Comput. Appl. **33**(17), 10637–10656 (2021)

50. Yin, H., Xu, J., Luo, Z., Xu, Y., He, S., Xiong, T.: Development and design of intelligent gymnasium system based on K-Means clustering algorithm under the internet of things. In: International conference on Big Data Analytics for Cyber-Physical-Systems, pp. 1568–1573. Springer, Singapore (2020)
51. Yang, J., Guo, B., Wang, Z., Ma, Y.: Hierarchical prediction based on network-representation-learning-enhanced clustering for bike-sharing system in smart city. IEEE Internet Things J. **8**(8), 6416–6424 (2020)
52. Famila, S., Jawahar, A., Sariga, A., Shankar, K.: Improved artificial bee colony optimization based clustering algorithm for SMART sensor environments. Peer-to-Peer Netw. Appl. **13**(4), 1071–1079 (2020)
53. Yang, J., Han, Y., Wang, Y., Jiang, B., Lv, Z., Song, H.: Optimization of real-time traffic network assignment based on IoT data using DBN and clustering model in smart city. Future Gener. Comput. Syst. **108**, 976–986 (2020)
54. Muntean, M.V.: Car park occupancy rates forecasting based on cluster analysis and kNN in smart cities. In: 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), pp. 1–4. IEEE, 2019
55. Qureshi, B., Kawlaq, K., Koubaa, A., Saeed, B. and Younis, M.: A commodity SBC-edge cluster for smart cities. In: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), pp. 1–6. IEEE, 2019
56. Zou, X., Cao, J., Sun, W., Guo, Q., Wen, T.: Flow data processing paradigm and its application in smart city using a cluster analysis approach. Clust. Comput. **22**(2), 435–444 (2019)
57. Logesh, R., Subramaniyaswamy, V., Vijayakumar, V., Gao, X.Z., Indragandhi, V.: A hybrid quantum-induced swarm intelligence clustering for the urban trip recommendation in smart city. Future Gener. Comput. Syst. **83**, 653–673 (2018)
58. Lücken, V., Voss, N., Schreier, J., Baag, T., Gehring, M., Raschen, M., Lanius, C., Leupers, R., Ascheid, G.: Density-based statistical clustering: enabling sidefire ultrasonic traffic sensing in smart cities. J. Adv. Transp. (2018). https://doi.org/10.1155/2018/9317291
59. Tabatabai, S., Mohammed, I., Al-Fuqaha, A. and Salahuddin, M.A.: Managing a cluster of IoT brokers in support of smart city applications. In: 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1–6. IEEE, 2017
60. Zhang, Q., Zhu, C., Yang, L.T., Chen, Z., Zhao, L., Li, P.: An incremental CFS algorithm for clustering large data in industrial internet of things. IEEE Trans. Ind. Inform. **13**(3), 1193–1201 (2017)
61. Guo, X., Lin, H., Yulei, Wu., Peng, M.: A new data clustering strategy for enhancing mutual privacy in healthcare IoT systems. Future. Gener. Comput. Syst. **113**, 407–417 (2020)
62. Chithaluru, P., Al-Turjman, F., Kumar, M., Stephan, T.: I-AREOR: an energy-balanced clustering protocol for implementing green IoT in smart cities. Sustain. Cities Soc. **61**, 102254 (2020)
63. Xu, L., O'Hare, G.M., Collier, R.: A smart and balanced energy-efficient multihop clustering algorithm (smart-beem) for mimo iot systems in future networks. Sensors **17**(7), 1574 (2017)
64. Rani, S., Ahmed, S.H., Rastogi, R.: Dynamic clustering approach based on wireless sensor networks genetic algorithm for IoT applications. Wirel. Netw. **26**(4), 2307–2316 (2020)
65. Bellaouar, S., Guerroumi, M., Moussaoui, S.: QoS based clustering for vehicular networks in smart cities. In: International Conference on Dependability in Sensor, Cloud, and Big Data Systems and Applications, pp. 67–79. Springer, Singapore (2019)
66. Darabkh, K.A., Wafaa, K.K., Ala'F, K.: LiM-AHP-GC: life time maximizing based on analytical hierarchal process and genetic clustering protocol for the internet of things environment. Comput. Netw. **176**, 107257 (2020)
67. Zheng, M., Chen, Si., Liang, W., Song, M.: NSAC: a novel clustering protocol in cognitive radio sensor networks for Internet of Things. IEEE Internet Things J. **6**(3), 5864–5865 (2019)
68. Tripathi, A.K., Sharma, K., Bala, M., Kumar, A., Menon, V.G., Bashir, A.K.: A parallel military-dog-based algorithm for clustering big data in cognitive industrial internet of things. IEEE Trans. Ind. Inform. **17**(3), 2134–2142 (2020)
69. Zhang, J., Feng, X., Liu, Z.: A grid-based clustering algorithm via load analysis for industrial Internet of things. IEEE Access **6**, 13117–13128 (2018)
70. Ja'afreh, M.A., Adhami, H., Alchalabi, A.E., Hoda, M., El Saddik, A.: Toward integrating software defined networks with the Internet of Things: a review. Clust. Comput. **2021**, 1–18 (2021)
71. Alomari, Z., Zhani, M.F., Aloqaily, M., Bouachir, O.: On minimizing synchronization cost in nfv-based environments. In: 2020 16th International Conference on Network and Service Management (CNSM), pp. 1–9. IEEE, 2020
72. Elrawy, M.F., Awad, A.I., Hamed, H.F.: Intrusion detection systems for IoT-based smart environments: a survey. J. Cloud Comput. **7**(1), 1–20 (2018)
73. Spadaccino, P., Cuomo, F.: Intrusion detection systems for iot: opportunities and challenges offered by edge computing. Accessed from https://arxiv.org/abs/2012.01174 (2020)
74. Khraisat, A., Alazab, A.: A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. Cybersecurity **4**(1), 1–27 (2021)
75. Yiğitler, H., Badihi, B., Jäntti, R.: Overview of time synchronization for IoT deployments: clock discipline algorithms and protocols. Sensors **20**(20), 5928 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mehdi Hosseinzadeh received the B.Sc. degree in computer hardware engineering, from Islamic Azad University, Dezfol, Iran in 2003. He also received the M.Sc. and Ph.D. degrees in computer system architecture from the Science and Research Branch, Islamic Azad University, Tehran, Iran in 2005 and 2008, respectively.,He is the author/co-author of more than 120 publications in technical journals and conferences. His research interests include SDN, Information Technology, Data Mining, Big data analytics, E-Commerce, E-Marketing, and Social Networks.

Atefeh Hemmati received her B.S. degree in Computer engineering, Information technology from Central Tehran Branch, IAU University, Iran, in 2020. Currently, she is an MSc candidate in Computer Engineering, Software at Science and Research Branch, IAU, Tehran, Iran. Her research interests include Fog computing, Cloud computing, the Internet of Things and, Security.

Amir Masoud Rahmani received his B.S. in Computer Engineering from Amir Kabir University, Tehran, in 1996, the MS from Sharif University of Technology, Tehran, in 1998, and the PhD degree in computer engineering from IAU University Tehran, in 2005. Currently, he is a professor in the department of computer engineering. He is the author/co-author of more than 350 publications in technical journals and conferences. His research interests are in distributed systems, the Internet of things, and evolutionary computing.