COMPUTERS AND INFORMATICS
2757-8259 2024, Volume 4 Issue 1
DOI: 10.62189/ci.1260288 Research Article

Optimal method to monitor network for IoT devices based on anomaly detection

Umar Ali *
TTG International Ltd., Reşadiye Mah. Reşadiye Cad. 110/1 34799 Çekmeköy - İstanbul, Türkiye, umar.ali@ttgint.com
Cenk Calis
TTG International Ltd., Reşadiye Mah. Reşadiye Cd. 110/1 34799 Çekmeköy - İstanbul, Türkiye, umar.ali@ttgint.com

Submitted: *31.03.2023*
Accepted: *01.01.2024*
Published: *01.06.2024*

* *Corresponding Author*

Abstract:

Many challenges have been identified to monitor, manage, process, and store the big data that accumulates from different sources in the IoT concept. The focus of this paper is very significant and limited to solving the problem of monitoring classified big data. Detection of anomalies in a grouping of classified data made it easy to monitor and help to make decisions for action to operate. There is no need to store, process, or manage the redundant data further that is already within the range of the group. So, the main concern is abnormal values in the groups that need to be processed further and require focus. The method proposed in this paper serves as an optimal solution designed to address the visualization challenges associated with dense and high-volume datasets. Our approach involves a strategic process of categorizing data into groups and pinpointing anomalies within these groups. This systematic classification not only enhances data organization but also plays a pivotal role in simplifying the visualization of intricate data patterns. Additionally, this method brings about significant cost efficiencies by strategically optimizing the expenses incurred in processing operations and the allocation of storage space for the equipment.

Keywords: *Anomaly detection, Big Data, Data monitoring, IoT*

Cite this paper as: Ali, U., & Calis, C., Optimal method to monitor network for IoT devices based on anomaly detection, *Computers and Informatics*, 2024; 4(1); 41-50, https://doi.org/10.62189/ci.1260288

1. INTRODUCTION

To manage heterogeneous devices, the Internet of Things (IoT) strategy establishes a framework, such as smart devices, sensors, and online services. Every type of device can communicate and connect with one another via the Internet, or any other network means [1]. The fundamental benefit of an IoT design is that it allows for connectivity with a wide range of devices and end-users can choose from several sources and services. With the current development in high-speed communication networks, sensor technology, and embedded microcomputers, IoT is taking one step ahead in the advancement of technologies.

With the emergence of the Internet of Things (IoT), a networked environment is anticipated in which a wide range of gadgets would transform all aspects of life. Several writers have stated that new IoT designs require redefined computing capabilities since Fog/Edge devices are among the most common sources of huge data streams [3-6]. The difficulty is in extracting valuable insights from this massive amount of unstructured data and then efficiently handling, evaluating, and keeping an eye on it. IoT big data visualization may take many different shapes, such as pie charts, heat maps, bubble plots, box-and-whisker plots, graphs, histograms, and donut charts. However, there are significant challenges in using Real-Time (RT) and Non-Real Time (NRT) data visualization for performance monitoring and alert raising.

In this research, an ideal method for tracking the massive amount of big data coming from different IoT device sources is presented. The approach focuses on avoiding duplicate data that is already categorized into established groups, so processing, managing, or retaining it is not necessary. As a result, the primary focus shifts to discovering anomalous values inside these defined groupings, necessitating more processing and focused attention. This novel approach not only simplifies the presentation of dense IoT data but also successfully tackles the problem of controlling anomalies and lessens the processing and storage load related to redundant data.

When examining the difficulties posed by IoT big data, a thorough analysis frequently centers on the 7Vs framework—Volume, Velocity, Variety, Veracity, Value, Variability, and Visualization—which is explained in [8] and covered in detail in the part that follows. But within this range, visualization is the precise subject of this paper's main attention. We focus on the targeted use of thresholds and limits to provide anomaly detection that is particular to a certain area. This novel method offers the best possible strategy to improving visualization skills. The approach defines cutoff points and bounds, which makes it easier to visualize dense, large-volume data by carefully classifying it while also pointing out anomalies. This not only helps to maximize visualization but also results in lower processing expenses and less space needed for equipment storage. marking a significant advancement in managing and comprehending intricate IoT big data.

2. RELATED CHALLENGES AND STUDY

In the context of IoT (Internet of Things) and big data, "7vs challenges" may refer to a set of challenges that arise from the characteristics of big data, commonly referred to as the "7Vs": Volume, Velocity, Variety, Veracity, Variability, Visualization, and Value. Each of these Vs presents specific challenges in handling and analyzing IoT big data. These challenges include: To begin with the background, there are 3Vs (Volume, Velocity, and Variety) and discussed in [7] by Gartner. While another challenge identified by some authors is “Value” in [9] afterward 5th V was introduced “Variability” in [10, 11]. Exploring more challenges couldn’t end and 6th V “Veracity” explicated in [12, 13]. As authors dug out more, they found the 7th V “Visualization” that is discussed in [14, 15].

Study shows that a Smart City has a lot of data flowing from many devices and sources. in terms of user requirements, this un-structural and unclassified (RAW) data is meaningless. As a result, the initial step is to convert this raw data into usable information using different Machine Learning strategies.

After processing data, interpretation at the application layer is sometimes easy but in the IoT smart city concept as shown in Figure 1, it becomes difficult. In this case, structural or unstructured data cannot be managed in the traditional way and the quality of visualization becomes low due to the high rate of image change, information loss, and visual noise [16].

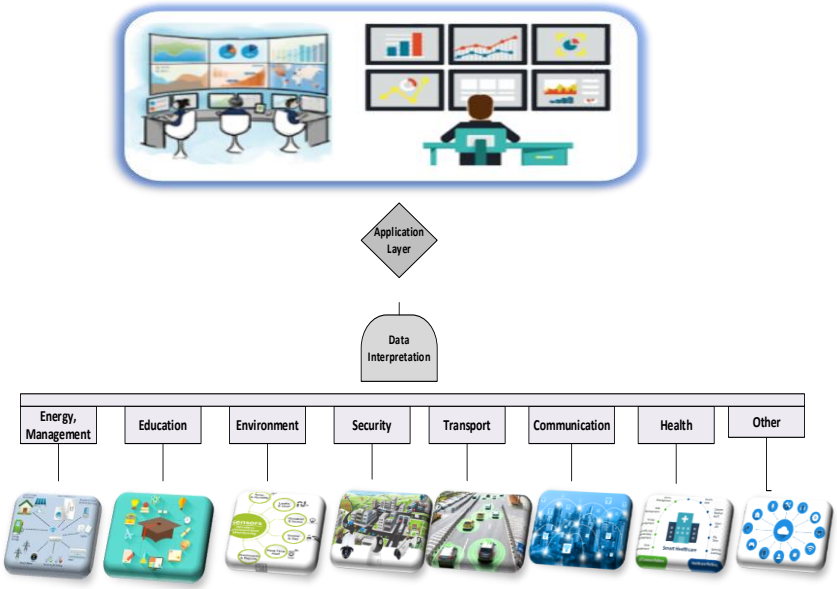

Figure 1. IoT based smart city data monitoring concept.

The increased usage of IoT devices suggests that new frameworks and solutions are required and should be created by organizations, academic institutions, and businesses. There are several IoT frameworks and platforms that are used to gather and analyze data in our daily lives. Several publications examine and survey these issues [17], [18], [19]. Even though various IoT frameworks and solutions have been suggested and created, there are still issues and features missing in this domain. IoT systems nowadays require privacy and security precautions that should be scalable, distributed, and lightweight [20].

As mentioned in [34], Mark Altaweel has created a GIS map displaying air quality values from Internet of Things devices. However, the reading of individual sensor data values is an inherent problem. The way these numbers are presented is challenging, making it more difficult to understand and evaluate the individual data from each sensor on the map.

As recommended in [35], Riyadh Arridha established a system of categorization for pollution levels, grouping them into standard, lightly polluted, polluted, and heavily polluted categories. Nevertheless, there is no way for redundant data to be handled by the process considering big data to visualize. When significant amounts of data fall into the categories of "lightly polluted," "polluted," or "heavily polluted," the current system is unable to handle, analyze, report, and store important data within these ranges in an efficient manner.

Our contribution proposes to combine a strategic grouping strategy with threshold-based anomaly detection algorithms. Our method intelligently identifies redundant or repetitive sensing data, categorizing it as non-critical for immediate reporting, storage, and subsequent processing, especially within the realm of big data visualization. This systematic approach not only streamlines data organization but also plays a pivotal role in simplifying the visualization of complex data patterns.

3. BASIC IDEA OF PROCESSING UNIFIED DATA IN IOT SYSTEM

IoT is assisting developed nations in their transition to smart city models, for instance, Smart transportation and communication [21], smart security [22], smart environment [23], Smart energy and management [24,25], smart education [26-28], smart finance [29], smart healthcare system [30, 31], and others are still in the works.

In the context of a Smart City, a vast number of diverse devices with various sensing capacities provide raw data. These devices are often dispersed around the city, allowing for real-time monitoring and the creation of a comprehensive picture of the city's current situation [32-33]. According to the proposed solution, to monitor and analyze the data, there are a few steps involved as shown in Figure 2. Fog/Edge devices such as sensors, web services, and actuators produce massive amounts of unstructured data [2]. For better support, decision, and planning, all data needs to be collected in one place to be treated equally.

Considering diverse standards and structures, there's a need to separate Real-Time (RT) and Non-Real Time (NRT) data due to their distinct processing requirements. Following filtration, mediation assumes a pivotal role by harmonizing differently formatted data into a unified structure. Consequently, during the classification process, non-identical data must be segregated. For instance, data on air quality and smart parking can't be analyzed together due to the absence of comparability between them.

The next process is grouping and processing through Artificial Intelligence (AI) techniques, this process creates groups of different ranges. The purpose of grouping is to identify the anomalies that are discussed in the next section in detail. At last, the decision for operation will be made once the abnormal data.

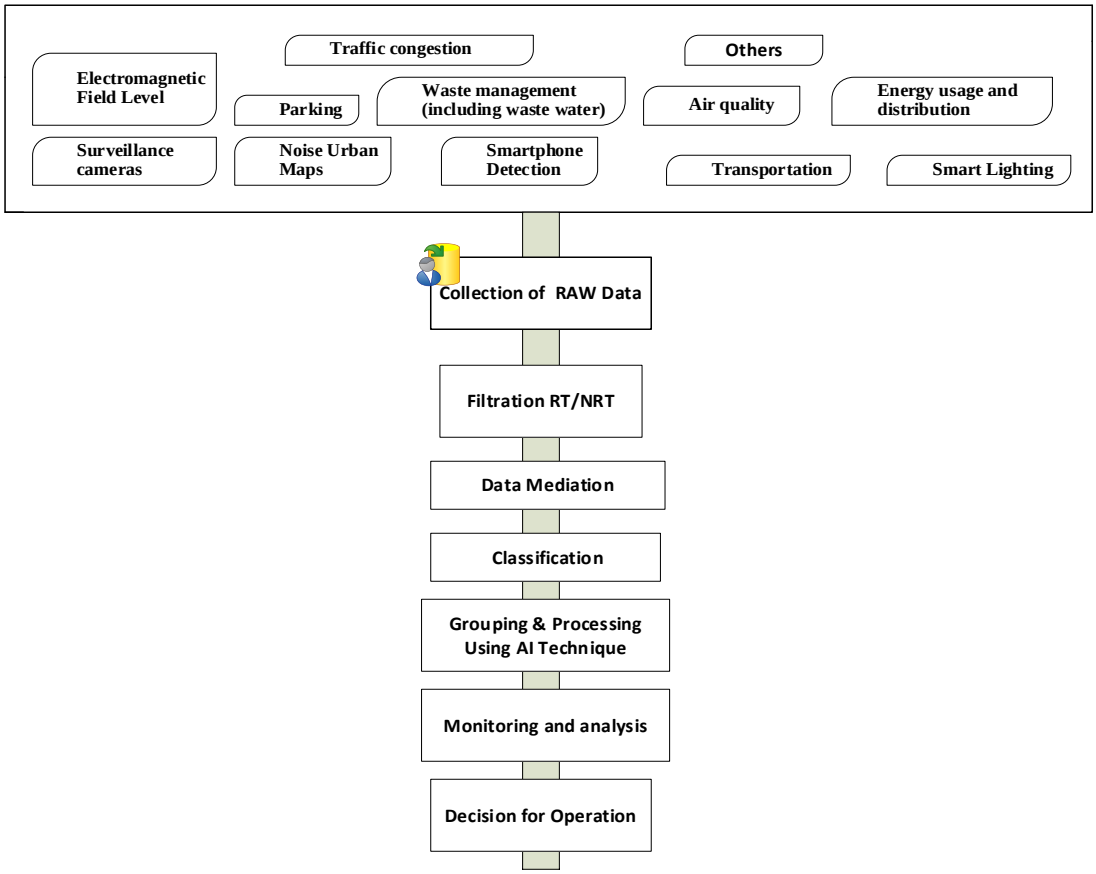

Figure 2. Basic idea of unified IoT data processing steps

4. PROPOSED GROUPING AND ANOMALY DETECTION TECHNIQUE

In various domains, the ability to classify group data and detect anomalies is significant for gaining insights, identifying patterns, and making informed decisions. By applying classification techniques to group similar data points together and subsequently detecting anomalies within those groups, valuable information can be extracted from complex datasets. This process enables a better understanding of normal behavior, facilitates anomaly identification, and contributes to improved data analysis. Here is an overview of the steps involved in classified group data and anomaly detection:

- *Data Preparation: Start by gathering and readying the dataset for examination. This entails guaranteeing the integrity of the data, addressing any missing values, and modifying variables as needed. Furthermore, take into account the pertinent features and attributes that depict the traits of the data points.*
- *Classification: In order to categorize data points into different classes based on their features, a classification algorithm is applied. The objective of this step is to recognize comparable patterns and cluster data points with common characteristics. In this study, counter based data was utilized.*
- *Anomaly Detection within Groups: After classifying the data points into distinct groups, the subsequent task involves identifying anomalies within each group. Anomaly detection methods are utilized to pinpoint instances that exhibit a substantial deviation from the anticipated or standard behavior within their respective groups.*
- *Anomaly Evaluation: Assess and verify the detected anomalies to ascertain their importance and pertinence. This entails contrasting the anomalies with domain expertise, expert evaluation, or past data. It is crucial to maintain an equilibrium between recognizing authentic anomalies and preventing incorrect categorizations.*
- *Visualization and Interpretation: Gain valuable insights and effectively interpret the results by visualizing the classified data and identifying anomalies. Employing various visualization techniques, such as scatter plots, bar charts, or GIS maps (in a spatial context), enables a comprehensive understanding of patterns, relationships, and spatial distributions within different data groups and anomalies. These visual representations play a crucial role in uncovering hidden trends and effectively communicating findings to stakeholders.*
- *Iterative Refinement: Conduct an analysis of the outcomes, and if required, adjust the classification and anomaly detection process. Enhance the model or parameters to enhance accuracy and performance. This iterative refinement process enables ongoing improvement and adaptation to the unique features of the dataset.*

The process of following these steps involves a thorough analysis of categorized group data, with the aim of uncovering any anomalies within these groups. This method involves delving into the intricate patterns embedded within the data, which facilitates the identification of regular or normal behaviors exhibited by the categorized entities. Furthermore, this approach is crucial in identifying instances that deviate significantly from the expected norms within these groups, indicating potential anomalies or irregularities. By identifying these abnormal occurrences, it prompts a closer inspection or further investigation to discern the underlying causes or implications. The key objective is to provide domain experts and decision-makers with a comprehensive understanding of the dataset's dynamics. This empowers them to draw accurate conclusions, make informed decisions, and initiate appropriate actions based on the insights derived from this detailed analysis.

The data categorization step organizes and classifies all the data obtained from the various sensor categories. The classification depends on the nature, value, properties, and standards of the collected data from different sources and areas in the smart city concept. Grouping of identical data after the classification process reduces the effort of computation and the complexity of analysis. In Figure 3, the data flow architecture has been divided mainly into two parts. The upper part of the architecture shows the different groups with different ranges. Here C_i denotes the classified different values between a range of x and y, while G represents the groups, and all ranged values are the members of the relevant

group as shown below. The number of groups varies with respect to the volume of the data and requirements.

$$x_1 \leq c_i \leq y_1 \rightarrow \epsilon\, G_1 \quad (1)$$

$$x_2 \leq c_i \leq y_2 \rightarrow \epsilon\, G_2 \quad (2)$$

$$x_n \leq c_i \leq y_n \rightarrow \epsilon\, G_n \quad (3)$$

The lower segment of the architecture delineates a threshold-based anomaly detection system. When an abnormal value is identified within any group, it triggers a process for reporting and necessary operational actions. The continual iterations span from time T_o (indicating the initial decision time) to T_K (marking the final decision time). This allocated time frame is specifically implemented to counteract abnormal values; if no action is taken within this timeframe, the anomaly is officially flagged as detected.

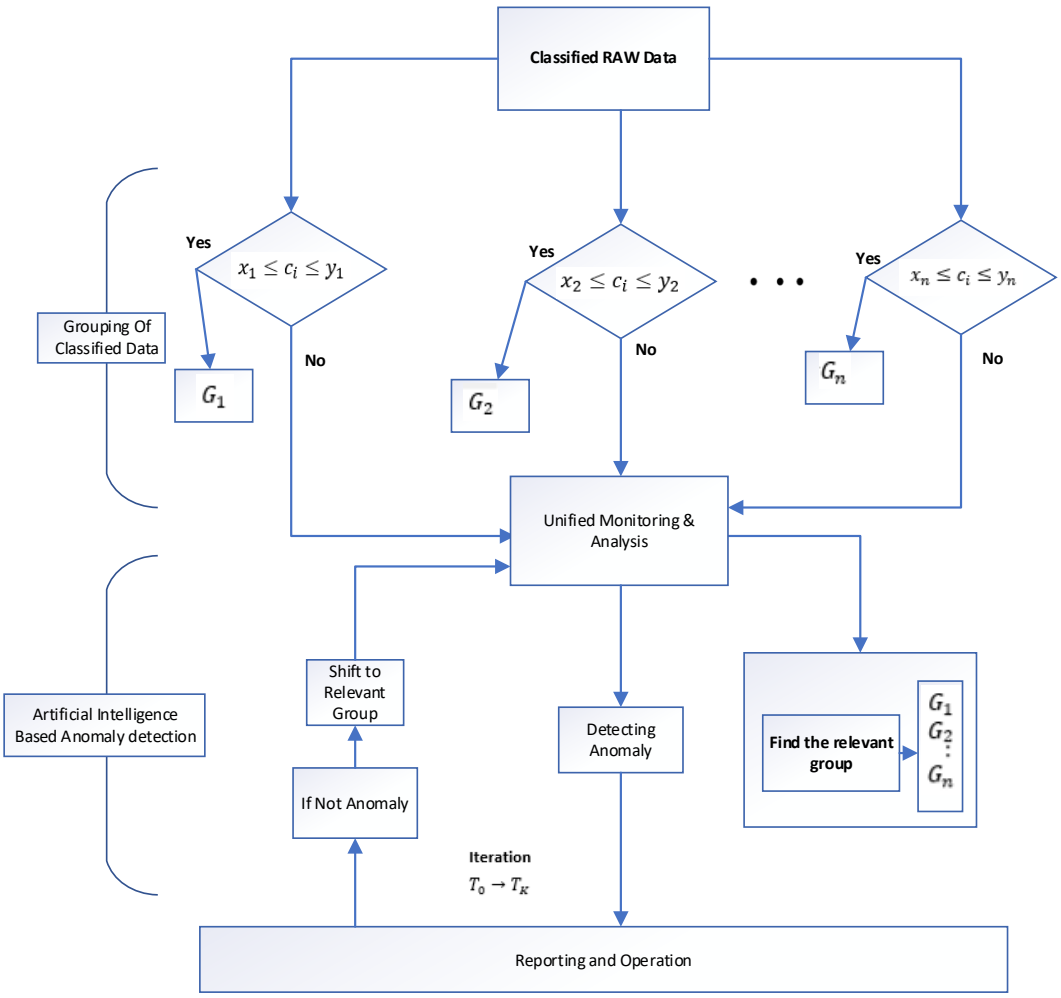

Figure 3. Data flow architecture

4.1. Results on GIS Based IoT data.

We conducted the experiment utilizing Sigma Plot and Maptive software tools. Figure 4 illustrates the processing of raw data from Geographic Information Systems (GIS)-based Internet of Things (IoT) sources. Part (a) of the figure depicts the collection of classified and quantified data at regular intervals from the source or edge devices. The process of monitoring and analyzing this data is particularly

challenging due to its complexity. Hence, to effectively manage this data complexity, various groups with well-defined quantified ranges have been established, as depicted in (b), (c), (d) parts. Collecting similar characteristic data and utilizing it for grouping purposes enhances data analysis and decision-making. Leveraging similar characteristic data for grouping provides valuable insights and opportunities for optimized operations, improved outcomes, and enhanced understanding of complex datasets. After performing data filtration and grouping, it is essential to identify any outliers or unusual values that deviate significantly from the expected patterns. So, it has been noticed after filtration of grouped values; a few countered values were detected as unusual which is highlighted in the graph (k) and (f).

The process involves setting a specific timeframe, from T_0 to T_k, to make decisions about anomalies. This timeframe is established because it's not practical or feasible to act for every abnormal value that's detected. Sometimes, the anomaly detection mechanism might identify a value as abnormal that doesn't align with the criteria set for anomalies within the decision process.

When such instances occur, and a value identified as an anomaly doesn't fit the criteria, there's a need to reassign the source device or edge to the appropriate group. This step ensures that the detected value, which was initially considered an anomaly but doesn't align with the anomaly criteria, is moved, or shifted to the relevant group for accurate classification or processing. This helps maintain the integrity and accuracy of the anomaly detection system.

Hence, the repetitive data or extensive redundant or has been filtered out. Only the pertinent data, picked out through iterations and abnormalities, has been extracted for visualization. This refined dataset requires less processing, facilitates easier analysis, simplifies reporting, and demands less storage capacity.

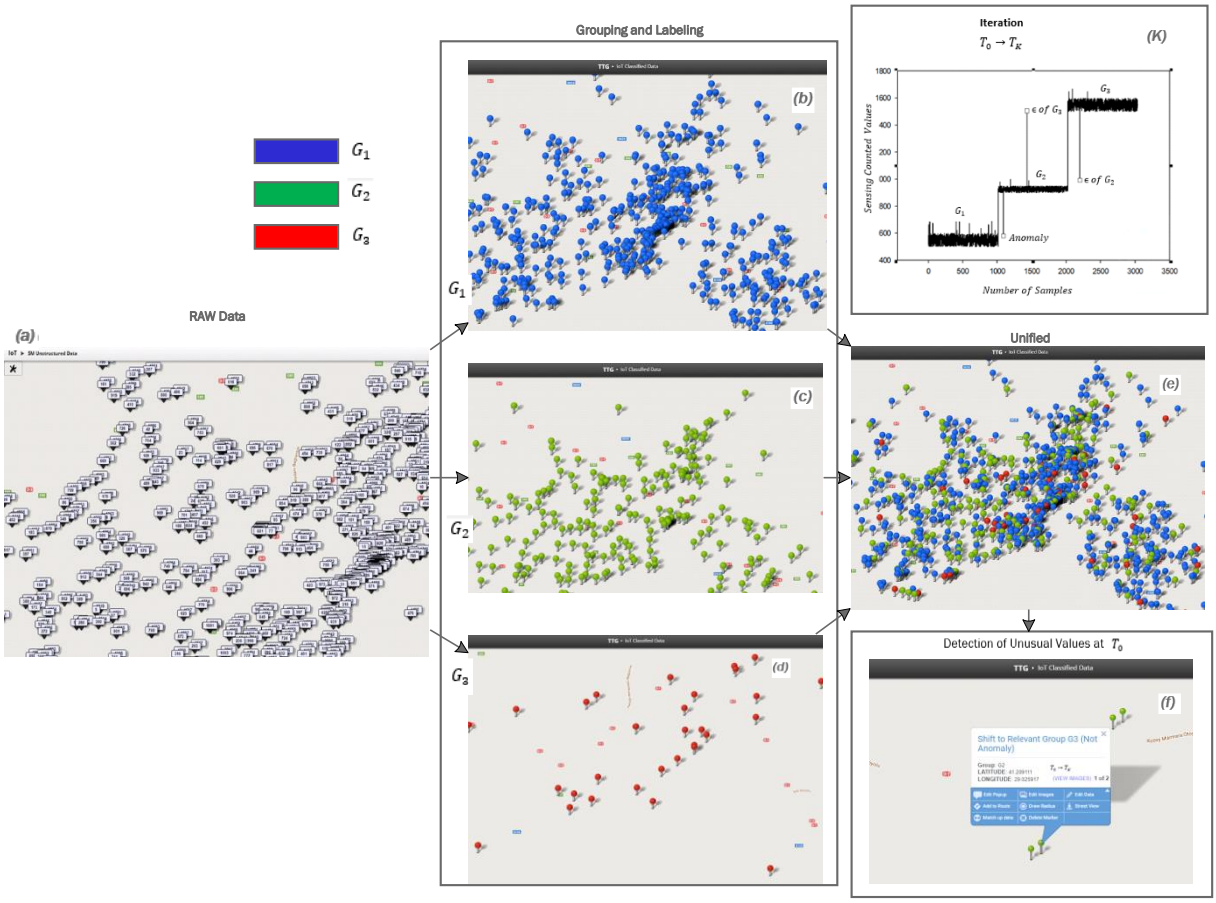

Figure 4. GIS IoT data processing by implementing proposed method.

5. CONCLUSION

Managing, processing, storing, and monitoring big data is never easy. A traditional way of processing big data consumes a lot of storage space and computational time that directly affects the cost of the hardware. Nevertheless, the efficiency of analyzing the data couldn't improve. Research on different IoT big data frameworks, challenges, and solutions has been proposed and is still in process.

This paper presents a pioneering approach geared toward IoT big data, introducing a novel grouping and anomaly detection technique. This method significantly streamlines the monitoring and analysis of classified data within the IoT sphere. Our proposed solution deliberately avoids advocating the storage and processing of redundant data originating from IoT fog/edge devices. Instead, it emphasizes the identification and management of exceptional data instances that warrant storage and operational processing. By adopting this optimized technique, not only are storage demands diminished, but processing time is also mitigated, offering an efficient framework for overseeing extensive IoT datasets.

Future work in GIS-based IoT network monitoring using anomaly detection can focus on several areas to enhance the efficiency, accuracy, and scalability of the system. Advanced Anomaly Detection Techniques: Investigate and develop more sophisticated anomaly detection algorithms that can handle complex and evolving patterns in IoT network data. This may include machine learning approaches such as deep learning, ensemble methods, or anomaly detection algorithms specifically tailored for spatial and temporal data. Develop techniques that can detect anomalies based on their spatial relationships, such as detecting sudden spikes or outliers in a specific geographic region or identifying spatial patterns of anomalies.

Acknowledgement

This research project has been funded by TTG International Türkiye for R&D purposes. TTG International is very active in the field of big data monitoring and analysis, especially for centralized network solutions. It provides OSS products and solutions to IT and Telecom clients all around the world. We are thankful to TTG International for supporting and encouraging us in research and development work.

REFERENCES

[1] Sinaeepourfard, A., Hierarchical distributed fog-to-cloud data management in smart cities (Doctoral dissertation, *Universitat Politècnica de Catalunya (UPC)).* 2017; DOI: 10.5821/dissertation-2117-114435.

[2] Sarkar, S., Chatterjee, S., & Misra, S., Assessment of the suitability of fog computing in the context of internet of things. *IEEE Transactions on Cloud Computing*. 2015; 6(1); 46-59. DOI: 10.1109/TCC.2015.2485206.

[3] Bonomi, F., Milito, R., Zhu, J., Addepalli, S., Fog computing and its role in the Internet of things. In: *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*. 2012; 13-16. DOI: 10.1145/2342509.2342513.

[4] Khan, R., Khan, SU., Zaheer, R., Khan, S., Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges. *10th International Conference on Frontiers of Information Technology*. 2012; 257-260. DOI: 10.1109/FIT.2012.53.

[5] Yi, S., Hao, Z., Qin, Z., & Li, Q., Fog computing: Platform and applications, *Third IEEE Workshop on Hot Topics In Web Systems And Technologies (HotWeb)*. 2015; 73-78. DOI: 10.1109/HotWeb.2015.22.

[6] Munir, A., Kansakar, P., & Khan, S, U., IFCIoT: Integrated Fog Cloud IoT: A novel architectural paradigm for the future Internet of Things. *IEEE Consumer Electronics Magazine*. 2017; 6(3); 74-82. DOI: 10.1109/MCE.2017.2684981.

[7] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U., The rise of "big data" on cloud computing: Review and open research issues. *Information systems*. 2015; 98-115. DOI: 10.1016/j.is.2014.07.006

[8] Pandey, K. K., Challenges of big data to big data mining with their processing framework, *8th International Conference on Communication Systems and Network Technologies,* 2018; 89-94. DOI: 10.1109/CSNT.2018.8820282.

[9] Uddin, M. F., & Gupta, N., Seven V's of Big Data understanding Big Data to extract value. *In Proceedings of the 2014 zone 1st conference of the American Society for Engineering Education*. 2014; 1-5. DOI: 10.1109/ASEEZone1.2014.6820689.
[10] Samuel, S. J., Rvp, K., Sashidhar, K., & Bharathi, C. R., A survey on big data and its research challenges. *ARPN Journal of Engineering and Applied Sciences*. 2015; 10(8); 3343-3347.
[11] Chen, C. P., Zhang, C. Y., Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*. 2014; 314-347. DOI: 10.1016/j.ins.2014.01.015.
[12] Rossi, R., Hirama, K., Characterizing big data management. *Issues in Informing Science and Information Technology*. 2015. DOI: 10.48550/arXiv.2201.05929.
[13] Demchenko, Y., Ngo, C., de Laat, C., Membrey, P. and Gordijenko, D., Big security for big data: Addressing security challenges for the big data infrastructure. *In Secure Data Management: 10th VLDB Workshop, SDM 2013, Trento, Italy, Proceedings* 10. 2014; 76-94. DOI: 10.1007/978-3-319-06811-4_13.
[14] Narasimhan, R., & Bhuvaneshwari, T, Big data-a brief study. Int. J. Sci. Eng. 2014; 5(9); 350-353.
[15] Eileen, M., DATACONOMY. Understanding big data: the seven V's [Internet]. Accessed June 2022. Available from: http://dataconomy.com/seven-vs-big-data/.
[16] Wang, L., Wang, G., & Alexander, C. A., Big data and visualization: methods, challenges and technology progress. *Digital Technologies*. 2015; 1(1); 33-38.
[17] Mineraud, J., Mazhelis, O., Su, X., & Tarkoma, S., A gap analysis of Internet-of-Things platforms. *Computer Communications*. 2016; 8; 5-16. DOI: 10.1016/j.comcom.2016.03.015.
[18] Perera, C., Liu, C. H., Jayawardena, S., & Chen, M., A survey on internet of things from industrial market perspective. *IEEE Access,* 2014; 1660-1679. DOI: 10.1109/ACCESS.2015.2389854.
[19] Perera, C., Liu, C. H., & Jayawardena, S., The emerging internet of things marketplace from an industrial perspective: A survey. *IEEE Transactions on Emerging Topics in Computing*, 2015; 3(4); 585-598. DOI: 10.1109/TETC.2015.2390034.
[20] Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P., Blockchain for IoT security and privacy: The case study of a smart home. *In 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops)*. 2017; 618-623. DOI: 10.1109/PERCOMW.2017.7917634.
[21] Kakani, P. N., & Rajendran, L., Flexible Communication Technologies Utilized in Developing Smart Cities. *In Smart Cities*. 2022; 245-267.
[22] Minoli, D., Sohraby, K., & Occhiogrosso, B., IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems. *IEEE Internet of Things Journal*. 2017; 4(1); 269-283. DOI: 10.1109/JIOT.2017.2647881.
[23] Vermesan, O., & Friess, P. (Eds.)., Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems. River publishers. 2013.
[24] Han, D. M., & Lim, J. H., Smart home energy management system using IEEE 802.15. 4 and zigbee. *IEEE Transactions on Consumer Electronics*. 2010; 56(3); 1403-1410. DOI: 10.1109/TCE.2010.5606276.
[25] Chen, C., Duan, S., Cai, T., Liu, B., & Hu, G., Smart energy management system for optimal microgrid economic operation. *IET Renewable Power Generation*. 2011; 5(3); 258-267. DOI: 10.1049/iet-rpg.2010.0052.
[26] Zhu, Z. T., Yu, M. H., & Riezebos, P., A research framework of smart education. *Smart Learning Environments*. 2016; 3; 1-17. DOI: 10.1186/s40561-016-0026-2
[27] Jeong, J. S., Kim, M., & Yoo, K. H., A content oriented smart education system based on cloud computing. International *Journal of Multimedia and Ubiquitous Engineering*. 2013; 8(6); 313-328. DOI: 10.14257/ijmue.2013.8.6.31
[28] Tikhomirov, V., Dneprovskaya, N., & Yankovskaya, E., Three dimensions of smart education. *In Smart Education and Smart e-Learning*. Springer International Publishing. 2015; 47-56. DOI: 10.1007/978-3-319-19875-0_5.
[29] Lehmann, E. E., Seitz, N., & Wirsching, K., Smart finance for smart places to foster new venture creation. *Economia e Politica Industriale*. 2015; 44; 51-75. DOI: 10.1007/s40812-016-0052-7.
[30] Catarinucci, L., De Donno, D., Mainetti, L., Palano, L., Patrono, L., Stefanizzi, M. L., & Tarricone, L., An IoT-aware architecture for smart healthcare systems. *IEEE Internet Of Things Journal*. 2015; 2(6); 515-526. DOI: DOI: 10.1109/JIOT.2015.2417684.
[31] Al Mamun, S. A., & Valimaki, J., Anomaly detection and classification in cellular networks using automatic labeling technique for applying supervised learning. *Procedia Computer Science*. 2018; 140; 186-195. DOI: 10.1016/j.procs.2018.10.328.
[32] Wang, J., Tang, Y., Nguyen, M., & Altintas, I., A scalable data science workflow approach for big data bayesian network learning. *In 2014 IEEE/ACM International Symposium on Big Data Computing*. 2014; 16-25. DOI: 10.1109/BDC.2014.10
[33] Kolias, V., Anagnostopoulos, I., & Kayafas, E., A Covering Classification Rule Induction Approach for Big Datasets. *In 2014 IEEE/ACM International Symposium on Big Data Computing*. 2014; 45-53. DOI: 10.1109/BDC.2014.17.

[34] Mark, A., The Spatial Internet of Things. [Internet]. Accessed June 2022. Available from: https://www.gislounge.com/the-spatial-internet-of-things/

[35] Arridha, R., Sukaridhoto, S., Pramadihanto, D., & Funabiki, N., Classification extension based on IoT-big data analytic for smart environment monitoring and analytic in real-time system. *International Journal of Space-Based and Situated Computing*. 2017; 7(2); 82-93. DOI: 10.1504/IJSSC.2017.086821.