Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 238 (2024) 306–313

Procedia Computer Science

www.elsevier.com/locate/procedia

The 15th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 23-25, 2024, Hasselt, Belgium

Unsupervised Machine Learning-Based Clustering of High-Frequency Radio Channel Properties: Analysis of Sector Communication Efficiency

Djamila Talbi[a, *], Zahra Daoui[b], Zoltan Gal[a]

[a]*Faculty of Informatics, University of Debrecen, Debrecen, Hungary*
[b]*Beijing Institute of Technology, Beijing, China*

Abstract

Machine learning has emerged as a powerful tool for both engineering and geo-localization applications. In this study, we investigate the Terabit/sec bandwidth wireless technology application using specialized ns-3 simulation tools. Through extensive simulations, we explore various scenarios with diverse parameters, including population density, topology types, and overlapping ratios among consecutive radio sectors centered around a single access point. To extract meaningful insights from the data, we employ the DBScan unsupervised learning method, enabling us to identify the optimal number of classes for sector efficiency features. Our optimization approach considers both the number of outliers and the minimum number of elements within each radio sector. By analysing a synthetic dataset generated from the simulation cases, we uncover valuable insights and establish the optimal working point for the system.

Keywords: DBScan clustering, Terahertz communication, Unsupervised machine learning, Beyond 5G/6G

1. Introduction

The advent of wireless communication technologies beyond 5G (B5G) or 6G promises terabit-per-second (Tbps) data rates, enabling a multitude of smart services, accommodating a larger number of mobile users, and emphasizing

* Corresponding author. Tel.: +36 20 274 6179.
E-mail address: talbi.djamila@inf.unideb.hu

10.1016/j.procs.2024.06.029

the communication properties such as high reliability, energy efficiency and the latency [1]. Machine learning (ML) offers tremendous potential for the next generation, including self-driving vehicles, voice assistants, and the effective utilization of vast amounts of stored data [2].

The rapid advancement of high-speed wireless communication systems, such as 6G, has generated significant interest in the application of unsupervised learning methodologies [3]. These sophisticated techniques facilitate the extraction of valuable insights from unlabelled data, without the need for explicit supervision, thereby presenting potential solutions to complex challenges in system optimization, resource allocation, fault diagnosis, and adaptive design. Unsupervised learning, as a prominent branch of machine learning, provides a powerful framework for discovering intricate patterns and extracting meaningful knowledge from unstructured data, eliminating the reliance on explicit labels or guidance. In the context of 6G, this cutting-edge technology exhibits great potential, empowering networks to autonomously analyse and comprehend intricate data patterns, ultimately fostering improved network optimization, efficient resource allocation, and the development of intelligent security solutions. Highlights of the paper are the following:

- Simulation-based analysis and MAC mechanism for THz band communication.
- Population density and overlapping ratios impact the collision rate.
- The use of unsupervised ML-based DBScan clustering and high-speed MAC mechanism evaluation with it.

In the following section, we highlight a selection of papers that are relevant to the subject matter at hand. Section three describes the features of THz radio communications with a focus on the collisions. Moreover, in section four we analyse the behavior of THz radio communication sectors in addition to the unsupervised learning-based clusterization application. Then we conclude with a summary of the results in section five.

2. Related Work

In a recent study [4], an overview of promising research directions in AI for 5G was presented, focusing on network optimization, resource allocation, physical layer acceleration, and end-to-end optimization. Another study [5] addressed the challenge of limited labeled data in machinery fault diagnostics through a deep learning-based method that leveraged unsupervised data and employed a three-stage training scheme, achieving promising diagnostic performance and overcoming data sparsity issues. The Multi-Scale Convolutional Recurrent Encoder-Decoder method [6] effectively captured temporal dependencies, inter-correlations, and noise robustness in multivariate time series analysis, outperforming state-of-the-art methods. Furthermore, in the field of routing for Opportunistic IoT (OppIoT) networks, a proposed protocol called GMMR [7] utilized unsupervised learning, specifically Gaussian Mixture Models, to automate routing decisions.

Neural networks were also explored in various applications, such as energy consumption reduction in Wireless Sensor Networks (WSNs) [8], where their parallel distributed computation, distributed storage, and data robustness offered intelligent solutions for dynamic power management and dimensionality reduction. Additionally, a DUL-based approach [9] was proposed in an IRS-assisted wireless-powered communication network, achieving high throughput without the computational complexity of GA. Moreover, the study [10] introduced CatAAE, a clustering-based approach for unsupervised fault diagnosis in rolling bearings, eliminating the need for labeled data and prior feature knowledge, and demonstrating satisfactory performance, robustness, and high clustering indicators in experimental evaluations.

In the realm of high-speed communication, prior research has investigated the potential of the terahertz (THz) band for achieving terabit-per-second communications in 6G networks [11]. While previous studies have focused on physical layer solutions, there is an emphasis on the need for a comprehensive communication stack design to address network challenges. This work provides an overview of obstacles and considerations in integrating the terahertz spectrum into mobile networks. Another research effort in the field of 6G/B5G has proposed a novel 3D space-time-frequency model for wireless channel modeling, enabling the characterization of channels in diverse scenarios [12]. The model can be customized for specific frequencies and scenarios, including 5G, and its accuracy and practicality have been confirmed through simulations and measurements. Moreover, researchers have emphasized the significance of the THz band in 6G networks, highlighting its potential for high-speed transmission and wide bandwidth [13].

The exploration of channel characteristics, atmospheric attenuation, propagation mechanisms, measurements, and models for specific scenarios, as well as the identification of future research challenges, are addressed. Furthermore, the importance of next-generation 6G systems in meeting the demands of high-speed, low-latency applications is discussed [14]. The potential of millimeter wave (mmWave) and THz bands, including propagation characteristics, channel models, design considerations, applications, and ongoing standardization efforts, is explored. Another related study focuses on THz-based self-backhauling in ultra-dense networks for achieving near-wired transmissions and reducing costs [15]. Addressing challenges such as time-slot scheduling, sub-band allocation, and power allocation, a joint scheme, JTSP, is proposed to optimize resource allocation and achieve significant throughput gains. Lastly, the potential of the THz band for 6G wireless networks, encompassing propagation characteristics, measurement capabilities, modeling techniques, and future research outlook, is investigated [16], highlighting its ability to support high data rates and capacity.

Motivated by related works, we adopted DBScan unsupervised machine learning to delve deeper into the MAC mechanism, aiming to extract valuable insights and pinpoint optimal efficiency features of the Adaptive Directional Antenna Protocol for THz networks (ADAPT).

3. Features of THz Radio Communication

To analyse 6G efficiently, we employ a specialized simulation tool called TeraSim [17], which simulates THz frequencies and proposes a new Medium Access Control (MAC) mechanism compatible with the IEEE 802.11.3d standardization of THz band communication schemes called ADAPT. We used the provided example of a macroscale scenario where the Access Point (AP) is placed at the center of a cell with a radius of 18 meters ($r = 18\,m$). The cell is divided into equally sized sectors, enabling the transmission and reception of data to meet the requirements of high-frequency THz communication. In our case, the number of sectors is set to 30 ($n = 30$).

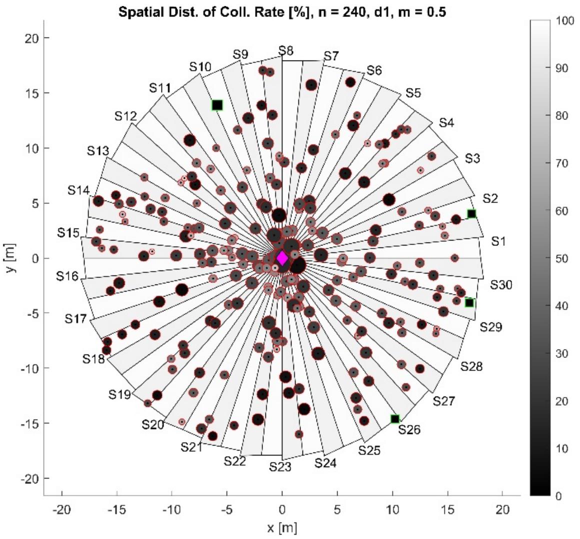

Fig. 1. Spatial distribution of collision rates in a medium congested centered topology (d1).

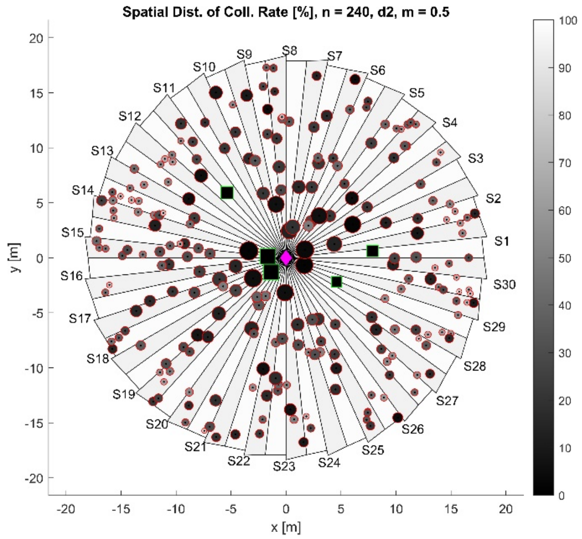

Fig. 2. Spatial distribution of collision rates in a medium congested uniform topology (d2).

Different numbers of Mobile Terminals (MTs) are utilized in each scenario N ∈ {60, 120, 240, 480, 960} for a total of five different population densities. We used the introduced new parameter in [18], the overlapping ratio (m), to visualize the radio channel property. The overlapping ratio is defined as:

$$m = A_{i+1} / A_i \tag{1}$$

where A_i represents the area of the sector $i = 1, \dots, N$. In our case, $m \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

This approach results in $M = 5 * 2 * 5 = 50$ simulation cases. There were used two spatial distributions of collision rates for both centered (d1) and uniform random (d2) distributions in a high population density scenario with an overlap ratio of $m = 0.5$ (see Fig. 1 & Fig. 2). The overlapping of sector areas was utilized to evaluate the impact of potential multiple transmissions by an MT in consecutive time frames. The probability of such consecutive transmissions is directly proportional to the parameter m. This phenomenon is relevant in real-world systems that involve burst traffic generated by MTs, which is commonly observed in high-speed communication systems.

The size of the bubble radius corresponds to the overall count of control frames transmitted by the Mobile Terminals, while the darkness level indicates the collision rate. A square marker denotes the absence of collisions at the Mobile Terminal level.

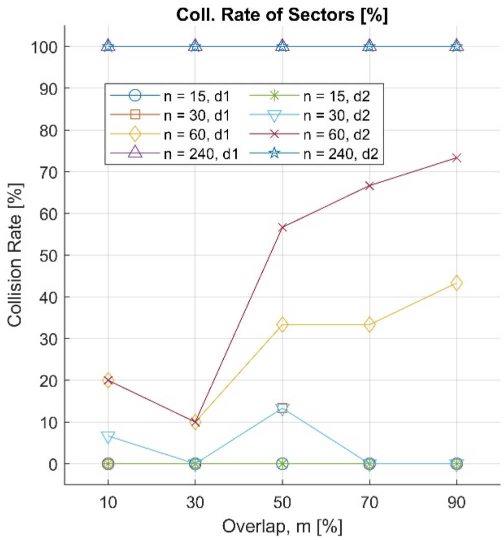

Fig. 3. Collision rate of the sectors.

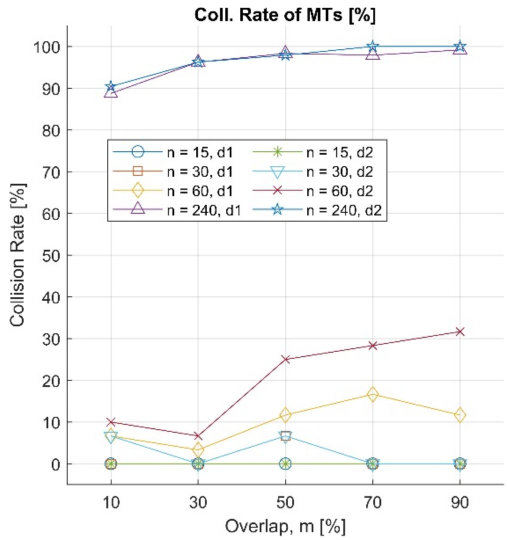

Fig.4. Collision rate of the Mobile Terminals.

The relationship between population density and collision rate is observed, (see Fig. 3 and Fig. 4). As the population density increases, the collision rate also increases accordingly. To better compare the impact of population density, the curves are normalized. It is worth noting that at very low population densities, no collisions occur. This highlights the direct correlation between population density and the occurrence of collisions in the studied system.

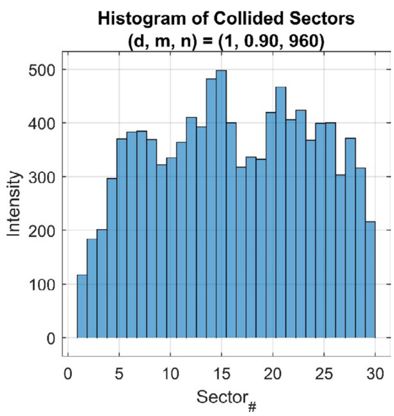

Fig.5. Histogram of the collided sectors.

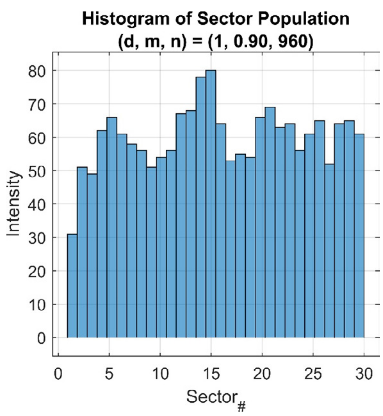

Fig.6. Histogram of the sector population.

To analyse the collision rates in the system, we differentiated between collision rates for sectors and collision rates for mobile terminals. Within a given sector, multiple MTs can collide simultaneously. Due to the overlapping sector areas, determined by the parameter m, collisions can occur in the overlapped regions of two consecutive sectors attempting to communicate concurrently. The collision rate for sectors is consistently present, especially at maximum population density (as observed in Fig. 3). It is evident that increasing the overlapping parameter leads to a higher collision rate, emphasizing the direct relationship between the size of the overlapping areas and collision occurrences.

The intensity range of collided sectors is $[100, 500]$ (Fig. 5 & Fig. 6), In contrast, the sector populations encompass a range of 30 to 80. The observed difference in these ranges can be attributed to the dynamic variations in the activity levels of Mobile Terminals within each individual sector. This indicates that the intensity of collisions is influenced by the number of active Mobile Terminals in a given sector, leading to variations in sector populations.

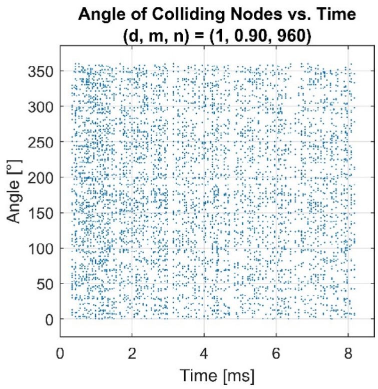

Fig.7. Dependence of the angle of the colliding nodes on time.

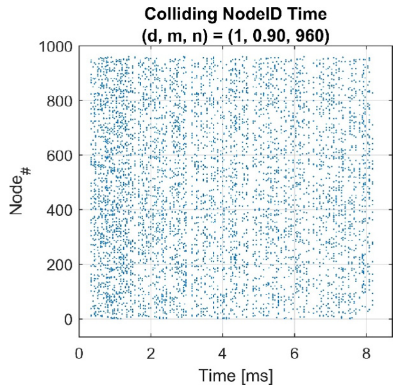

Fig.8. Dependence of the nodes ID on time.

By analysing the temporal changes in the angle of colliding nodes and the node ID (as depicted in Fig. 7 and Fig. 8), it becomes evident that both distributions possess attributes of independence and identical distribution (i.i.d.) within the two-dimensional (2D) space. This observation indicates that nodes within each sector exhibit comparable random behavior, thereby reinforcing the notion of homogeneous action among nodes in the same sector. The consistent i.i.d. nature of these distributions supports the assumption of a uniform and unbiased distribution of node activity within sectors, bolstering the understanding of their collective behavior.

4. The Behavior of THz Radio Sectors

To gain deeper insights into the unique characteristics of THz radio channels, we conducted a comprehensive analysis of sector vectors in 50 simulation cases (refer to Fig. 9 & Fig. 10).

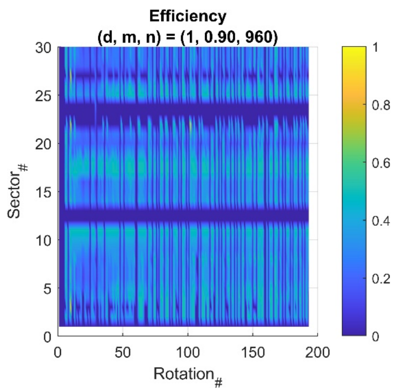

Fig.9. Sector's efficiency vs. rotations.

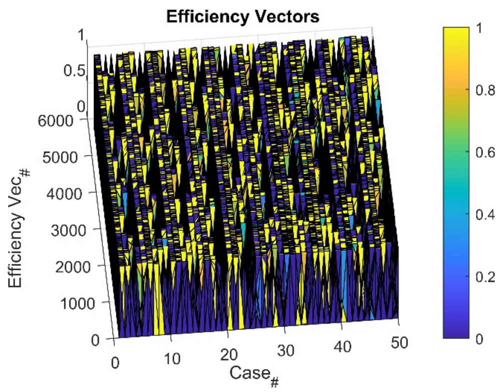

Fig.10. Dependence of efficiency vectors on the number of cases.

By exploring the efficiency of sector vectors plotted against sectors and rotations, as well as the efficacy vectors plotted against the number of cases, we gained valuable insights into their performance and behavior. This examination allows us to better understand the efficiency and efficacy of sector vectors, contributing to the optimization and design of THz radio communication systems. The results provide crucial information for improving channel utilization, optimizing resource allocation, and enhancing the overall performance of THz-based wireless networks.

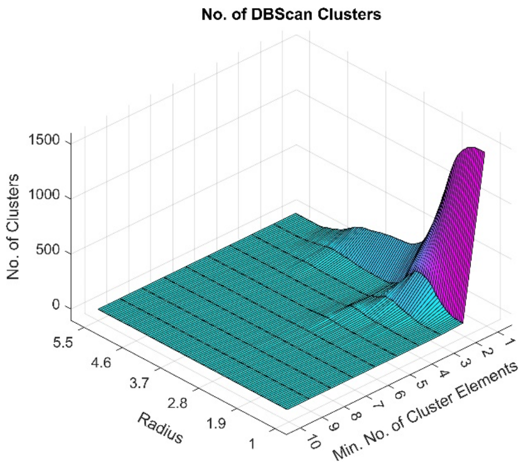

Fig.11. Dependence of the number of the clusters on the radius and the number of cluster elements.

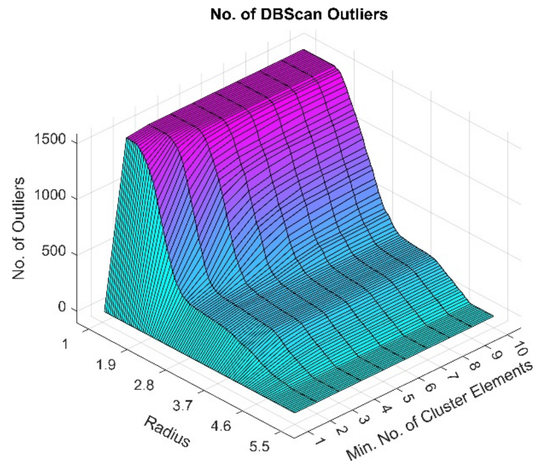

Fig.12. Dependence of the number of outliers on the radius and the number of cluster elements.

In the context of unsupervised machine learning, the DBScan algorithm utilizes two crucial parameters: the cluster radius and the minimum number of elements (e_min) required in each cluster (see Fig. 11). Results from the analysis (depicted in Fig. 12) consistently unveil a well-defined clustering solution, establishing a clear association between the number of outliers and the chosen values for the radius and minimum number of elements.

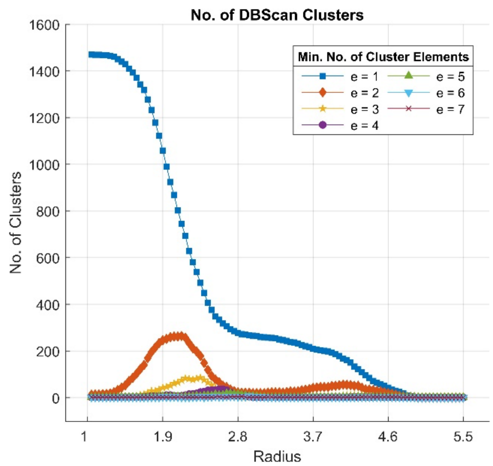

Fig.13. Dependence of the number of DBScan clusters on the radius.

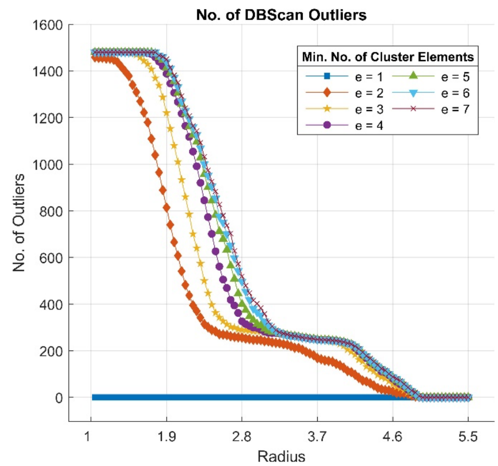

Fig.14. Dependence of the number of DBScan outliers on the radius.

This underscores the robustness of the DBScan algorithm and emphasizes the critical dependence on these parameters, offering valuable insights for achieving effective data clustering across diverse applications. The analysis demonstrates a distinct correlation between the number of clusters and the cluster radius (see Fig. 13). Each curve represents a specific minimum number of elements required in the cluster. Additionally, it is worth noting that all efficiency vectors identified as outliers are concentrated within cluster 0. This finding emphasizes the significance of the cluster radius in determining the clustering outcome and highlights the distinct separation of outlier data points from the main clusters. These insights provide valuable information for understanding the clustering structure and identifying anomalous data points within the analysed dataset. The relationship between increasing the cluster radius and the resulting impact on the number of clusters and outliers is clearly evident (Fig. 14). As the radius expands, there is a noticeable decrease in both the number of clusters formed and the presence of outliers.

This observation further emphasizes the critical importance of finding the optimal balance between the cluster radius and the minimum number of elements required for each cluster. Careful consideration of these parameters is essential for achieving accurate and effective data clustering while minimizing the presence of outliers in the analysed dataset. Striking the right balance ensures that the clustering results are meaningful, representative, and reliable, enabling valuable insights and knowledge extraction from the data. To determine the optimal working point, we investigated the relationship between the number of outliers and the minimum number of cluster elements (see Fig. 15). It is evident that when $e\_min = 1$, each element forms an individual cluster, resulting in the number of clusters equating to the number of elements. Consequently, the number of outliers is zero in this scenario (refer to Fig. 15).

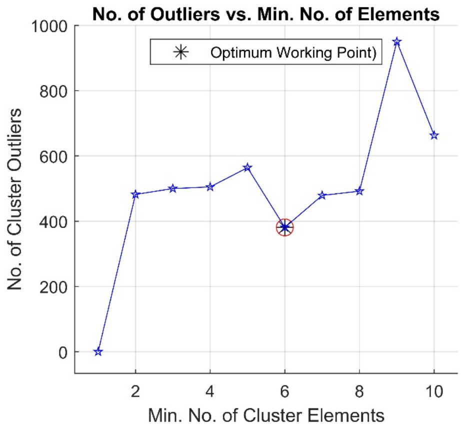

Fig.15. Number of outliers vs. number of elements.

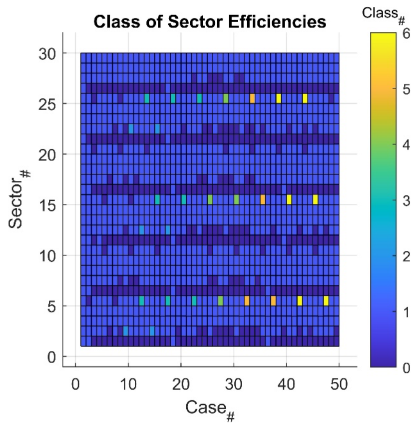

Fig.16. Classes of sector efficiencies.

Through our analysis, we determined that the optimal working point is achieved by requiring a minimum of 6 elements in each cluster of sector efficiency vectors, as it corresponds to the absolute minimum on the curve for $e\_min > 1$. Accordingly, we established 7 distinct classes $(C_0, C_1, C_2, C_3, C_4, C_5, C_6)$ to categorize the sector efficiencies. Outliers are enrolled into class C_0. This careful selection of the working point and class categorization enhances the precision and effectiveness of the data clustering process, providing meaningful insights into the sector efficiency patterns within the analysed dataset.

Given the relatively high number of outliers (383 out of 1500, which is approximately 25.5% of the data), it is notable that Fig. 16 comprises a significant number of low values (dark blue rectangles belonging to C_0). However, it remains an open question as to why sectors with identifiers $Sector\_ID \in \{5, 15, 25\}$ and case identifiers are consistently localized within clusters identified as $Cluster\_ID \in \{C_4, C_5, C_6\}$. This intriguing observation presents an opportunity for further investigation and exploration to determine the underlying factors contributing to this specific clustering pattern, potentially unveiling unique characteristics or dependencies specific to these sectors and cases.

5. Conclusions

In this paper, we presented a method for unsupervised machine learning-based clustering of local properties in Terahertz radio communication technology using 50 simulation cases. The 2D spatial localization was performed for 30 sectors surrounding the AP. We determined that the optimal number of sector efficiency classes is 7, identified through the DBScan clustering algorithm. Further research is needed to address questions such as the unequal distribution of sectors among classes and the applicability of labeling proposed in this work for supervised machine learning.

Acknowledgements

This work has been supported by the QoS-HPC-IoT Laboratory and project TKP2021-NKTA of the University of Debrecen, Hungary. Project no. TKP2021-NKTA-34 has been implemented with the support provided by the National Research, Development, and Innovation Fund of Hungary, financed under the TKP2021-NKTA funding scheme.

References

[1] Taleb, T., Aguiar, R. L., Yahia, I. G. B., Chatras, B., Christensen, G., Chunduri, U., Clemm, A., Costa, X., Dong, L., Elmirghani, J., Yosuf, B., Foukas, X., Galis, A., Giordani, M., Gurtov, A., Hecker, A., Huang, C.-W., Jacquenet, C., Kellerer, W., Zorzi, M. (2020) "White Paper on 6G Networking" [White paper], *6G Research Visions* **6**: 1-36.

[2] Ali, S., Saad, W., & Steinbach, D. (2020) "White Paper on Machine Learning in 6G Wireless Communication Networks" [White paper], 6G Research Visions **7**: 1-36.

[3] V. P. Rekkas, S. Sotiroudis, P. Sarigiannidis, G. K. Karagiannidis and S. K. Goudos, "Unsupervised Machine Learning in 6G Networks -State-of-the-art and Future Trends," 2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST), Thessaloniki, Greece, 2021, pp. 1-4, doi: 10.1109/MOCAST52088.2021.9493388.

[4] You, X., Zhang, C., Tan, X., et al. (2018) "AI for 5G: research directions and paradigms" in Science China Information Sciences **62**: 1-13.

[5] Li. Xiang, Li. Xu, Ma. Hui (2020), "Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery" *Mechanical Systems and Signal Processing* **143**.

[6] Chuxu Zhang, Dongjin Song, Yuncong Chen, Xinyang Feng, Cristian Lumezanu, Wei Cheng, Jingchao Ni, Bo Zong, Haifeng Chen, and Nitesh V. Chawla, (2019) "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data" *AAAI Press* **179**: 1-8.

[7] V. Vidushi, Ch. Anshuman, Sh. Kumar (2019) "GMMR: A Gaussian mixture model based unsupervised machine learning approach for optimal routing in opportunistic IoT networks" *Computer Communications* **134**: 138-148.

[8] Enami, Neda & Askari, Reza & Dadashtabar, Kourosh & Mojtaba, H. (2010) "Neural Network Based Energy Efficiency in Wireless Sensor Networks: A Survey" *International Journal of Computer Science & Engineering Survey* **1 (1):** 39-55

[9] M. Ahsan, W. Omer, UR. Rahman, M. Mahboob, (2022) "Throughput maximization of an IRS-assisted wireless powered network with interference: A deep unsupervised learning approach" *Physical Communication* **51**.

[10] L. Han, Xu. Yanhe, Z. Yang, P. Xuanlin, J. Wei, (2018) "Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks" *Neurocomputing* **315**: 412-424.

[11] M. Polese, J. M. Jornet, T. Melodia and M. Zorzi (2020) "Toward End-to-End, Full-Stack 6G Terahertz Networks," *IEEE Communications Magazine*, **58 (11)**: 48-54, doi: 10.1109/MCOM.001.2000224.

[12] C. -X. Wang, J. Wang, S. Hu, Z. H. Jiang, J. Tao and F. Yan (2021) "Key Technologies in 6G Terahertz Wireless Communication Systems: A Survey," *IEEE Vehicular Technology Magazine*, **16 (4)**: 27-37, doi: 10.1109/MVT.2021.3116420.

[13] J. Wang, C. -X. Wang, J. Huang and Y. Chen, "6G THz Propagation Channel Characteristics and Modeling: Recent Developments and Future Challenges" *IEEE Communications Magazine*, doi: 10.1109/MCOM.001.2200403.

[14] Tripathi, S., Sabu, N.V., Gupta, A.K., Dhillon, H.S. (2021). "Millimeter-Wave and Terahertz Spectrum for 6G Wireless" *6G Mobile Wireless Networks. Computer Communications and Networks. Springer*, Cham. https://doi.org/10.1007/978-3-030-72777-2_6.

[15] M. Yu, A. Tang, X. Wang and C. Han (2020) "Joint Scheduling and Power Allocation for 6G Terahertz Mesh Networks" *International Conference on Computing, Networking and Communications (ICNC)*, Big Island, HI, USA, 631-635, doi: 10.1109/ICNC47757.2020.9049790.

[16] D. Serghiou, M. Khalily, T. W. C. Brown and R. Tafazolli (2022) "Terahertz Channel Propagation Phenomena, Measurement Techniques and Modeling for 6G Wireless Communication Applications: A Survey, Open Challenges and Future Research Directions," *IEEE Communications Surveys & Tutorials*, **24 (4)**; 1957-1996, doi: 10.1109/COMST.2022.3205505.

[17] M. Daniel and J. Josep, "ADAPT: An Adaptive Directional Antenna Protocol for medium access control in Terahertz communication networks", Ad Hoc Networks, vol. 119, May 2021.

[18] D. Talbi and Z. Gal (2020) "Impact of Multi-Layer Recurrent Neural Networks in the Congestion Analysis of TeraHertz B5G/6G MAC Mechanism," *International Conference on Software, Telecommunications and Computer Networks*, 1-6, doi: 10.23919/SoftCOM55329.2022.9911500.