2024 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)

K-Means Artificial Intelligence Clustering Technique for 6G Propagation Channels

Saud Alhajaj Aldossari*
*Department of Electrical Engineering, Prince Sattam bin Abdulaziz University, Wadi Addwasir, Saudi Arabia
Email:s.alhajaj@psau.edu.sa

***Abstract*—The global contemporary revolution is accelerating in our modern environment. Mobile communications have undergone a generational transition at regular periods. While fifth-generation (5G) structures are presently beginning to be financially suitable and with the benefaction of high MmWave and terahertz (THz) frequencies, there is a great deal of interest in systems past 5G. It is known as the 6Th generation (6G) remote system. In recent years, the development of Artificial Intelligence (AI) has rapidly grown, leading to its integration into various fields, including wireless communications. This integration of AI in wireless communication technologies can potentially revolutionize the future of 6G technology. The widespread use of new generation Information and Communications Technology (ICT) such as AI, blockchain innovation, virtual reality (VR), augmented reality (AR), etc. With the usage of high MmWave, the structure and propagation of the wireless signal is unknown. This paper studies the behavior of Channel State Information (CSI) via unsupervised learning method such as K-Means clustering methods in 145 GHz bands.**

***Index Terms*—6G, artificial intelligence clustering, K-Means, MmWave, CSI.**

I. Introduction

Wireless communication frameworks have seen substantial revolutionary progress in recent years. Fifth-generation (5G) is the most recent attempt that refreshes diverse communications innovation and 5G is expected to reach one-third of the world's population by 2025 [1]. While, 6G will be a linked framework of several diverse organizations: neighborhood, mobile cell, sea, satellites, and organizations [7].

Various flourishing applications have demonstrated 5G's enormous capability, which continues to develop and adapt to various arising use cases. Nonetheless, as cultural requirements grow continuously, there has been an observable expansion in the quantity of arising use cases that 5G cannot deal with [3]. 6G introduces plausible new ideas, models, and solutions to recent drawbacks. Thus, 6G will be capable of overcoming 5G's in terms of delay, data rate, auto driving, and other applications that are related to artificial intelligent devices [4] [5]. The 6G framework comprises an intelligent network, deep availability, holographic availability, and omnipresent availability. Intelligent availability reflects all aspects of the communication framework's mental prowess: the organization's components and design, the intellectual prowess of the associated end device, and the data conveyed to assist the intelligent assistance. Holographic communication anywhere and anytime, high loyalty, and constant inclusion AR/VR/XR are the characteristics of a holographic network. A pervasive network is a multi-layered off-road and all-space inclusion association [6]. 5G is in the beginning phases of large-scale marketing. Its linked specialist capabilities should be improved and advanced regularly. Research in IoT business approaches and industrial application scenarios should also be conducted [7]. The most recent portable communications framework includes the ten-year cycle.

It is not a surprise that there is much mistrust regarding Artificial Intelligence. Many people still rely on traditional mathematical dependent optimization and find it hard to move to data-dependent deep learning. Autonomy is becoming the norm as people transition into the new era of mass digital connectivity. Different things, from vehicles to home appliances, are interconnected to ease daily life. However, there is much distrust resulting from a lack of explainability. This mistrust will further decrease as 6G starts to manage most of our daily activities. Fundamentally, the expectation of 6G is extended to enhance the coverage including regular cellular, satellite, and underwater communications. Furthermore, 6G will significantly improve the transmission rate, latency, security and QoS's [8]. Due to the new behavior of the signal propagation in the MmWave, researchers are working to investigate these behaviors such as [**?**] which explored multiple clustering algorithmic techniques for big data applications. [10] used clustering method for V2V communications. While [11] enhanced a new methodology for Clustream based on data streaming for Multi path Components (MPC) of time-variant radio channels. Moreover, [12] focused on the Gaussian mixture model (GMM) to study the structure and similarity of the wireless channel multipath.

The primary contributions of this paper are as follows:

- Introducing the state of the art of 6G.
- Presenting K-Means clustering method
- Show how to involve Unsupervised learning in the Mmwave
- Shows a case of how AI is involved in assisting wireless channel modeling
- Learn the behavior of CSI

Furthermore, the structure of this manuscript starts with a general introduction. Then, six generations of cellular systems

This project was funded by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University award number 2023/SDG/01.

(6G) in section two, and section three explain the standards of 6G. Section four goes over the intuitive architecture of 6G and followed by section five on the advent of 6G networks. Then section six regarding the methods for developing to 6G and followed by section seven explores the 6G concerns, challenges and opportunities. Section eight introduces artificial intelligence in 6G and section 9 focuses on how to involve AI in channel state information using unsupervised learning methods. Finally, future works and conclusions are in sections ten and eleven, respectively.

Finally, we conclude by discussing the challenges and future research directions in the integration of AI in 6G wireless communication technology. The results of this study provide insights into the potential benefits of AI in the future wireless communications 6G and highlight the need for further research in this area.

II. Artificial Intelligence and 6G

As a significant knowledge enabling innovation, machine learning (ML) is currently employed in the wireless framework to address issues and prepare for future emerging technologies such as vehicular communications. Near to the future systems management in 6G, AI calculations are commonly utilized for network radio resource management, transmit power control, correspondence booking, dynamic directing, traffic dumping, and associate enrolment. This section examines the related investigation, particularly the recently stated proposition of utilizing ML for automobile framework organization.

A. Machine Learning in the Asset Section

Unlike the standard wireless organization, where asset allotments are typically investigated, the vehicular organization has unique imperatives and requirements for intelligent asset identification methods. The intelligent ML-based asset portion has recently received much attention for conquering the three basic demands of asset allotment in the automotive organization. ML strategies have, as of late, been broadly used in strong asset designation in automotive companies. In highly powerful vehicular organizations, directed learning and support learning are readied for rapid choice and reaction, state evaluation and displaying asset recuperation, and self-adaption.

B. Machine Learning for Network Traffic Management

Vehicle flexibility affects both communication availability and directing planning in the vehicular organization. Recently, ML has expanded adaptability expectation, capable association, dynamic directing, speedy handoff, and planned block avoidance for vehicular association shrewd traffic light.

- Intelligent Prediction-Based Network Traffic Management: Due to the high movement of vehicles, the built associations and steering paths are frequently disrupted. An essential arrangement is to lay out and maintain start-to-finish connections in vehicular organizations, including vehicle versatility and vehicular traffic change patterns in the substantial organization steering issue in a vehicular organization [13]. The estimating feature makes the administered ML useful for data recuperation in range and space and state expectancy in time. Several ML-based anticipation methodologies are provided to anticipate vehicle movement and plan network traffic: the verified directions, speed-based vehicles' area prediction, and the authentic vehicles' stream-based future traffic load estimate. For the essential method, time-series-based ML calculations like a fake cerebrum association and well are used for vehicle development and region expectation.
- Correlation Recuperating Based Network Traffic Signal: One more usage of AI in the network traffic light is the recuperation of data from fragmented information. Traded flagging, for example, signals create significant overburden channels in the wireless organization. Flagging, including security and region data, is more sequential and easier to block for the vehicular organization. Furthermore, cars in remote places find it difficult to obtain comprehensive global data because of the vast geography of the vehicular organization. By recovering the steering data from verified traffic designs acquired only from edge hubs, profound learning might be used in an association for a wise traffic signal. The data gained covers the connections between traffic plans and guiding choices. To pick the best action for a wise traffic signal, we should initially figure out the connection between its enormous boundaries (e.g., hub thickness, spatial dissemination, and wireless channel conditions), typically displayed with nonlinear conditions, and deal with heuristic calculations. In order to address such an issue, an ML-based calculation is used to deal with the comprehension connections and assessing choices in a solitary opening utilizing controlled ML to push the envelope in multihop broadcast conventions for car associations.
- Unresolved problem: The varied and massive scope structure makes future vehicular communications challenging to implement with productive ML computations. For a long time, the ML-based network capabilities have been arranged, which have been spread to change under the cut association essential and close by features of various subnetworks. The associations between scattered network capacities are a tedious and massive cost.

III. Results of Investigating 5G+/6G Channel State Information via Unsupervised AI Learning

In this section, channel state information (CSI date has been generated as mentioned in [14] and AI techniques such as clustering have been applied to investigate the 6G CSI. The environment that was used to conduct a wireless channel experiment was urban micro with a frequency of 145 GHz. The following CSI data were used T-R Separation Distance (m), Time Delay (ns), Received Power (dBm), Phase (rad), Azimuth AoD (degree) Elevation AoD (degree), Azimuth AoA (degree), Elevation AoA (degree), Path Loss (dB) and RMS Delay Spread (ns). The methodology used to investigate the wireless channel is unsupervised learning. Moreover, this section presents analysis models for binary variable analysis.

The objective is to assess the performance of clustering models and gain insights into the underlying data patterns. Through the use of metrics and visualizations, this manuscript aims to provide a comprehensive understanding of the data and facilitate informed decision-making. Thus, we explore three models: k-means, GMM and Agglomerative. Evaluation metrics such as silhouette score, Calinski-Harabasz score, and Davies-Bouldin score are utilized to measure the quality of the clusters. Visualizations, including cluster proportions and scatter plots comparing features, enhance the interpretation of the results. The use of metrics and visualizations enables a thorough evaluation of model performance and enhances our understanding of the binary variable analysis.

Exploratory Data Analysis (EDA) and data cleaning have been performed. The EDA function provides a comprehensive report with a boxplot analysis chart as shown in figure 1 to gain a better understanding of the data and obtain a comprehensive view of its key features and patterns. The box plot provides insights into data concentration. It is noteworthy that Time Delay, Azimuth AoD, and Azimuth AoA exhibit a wider range of values compared to other variables, confirming their high variability. Additionally, the researcher observed that the Received Power variable consists solely of negative values which represents the dB decibel. Furthermore, we can notice the path loss values from 30-180 dB representing the power loss between Tx and Rx.

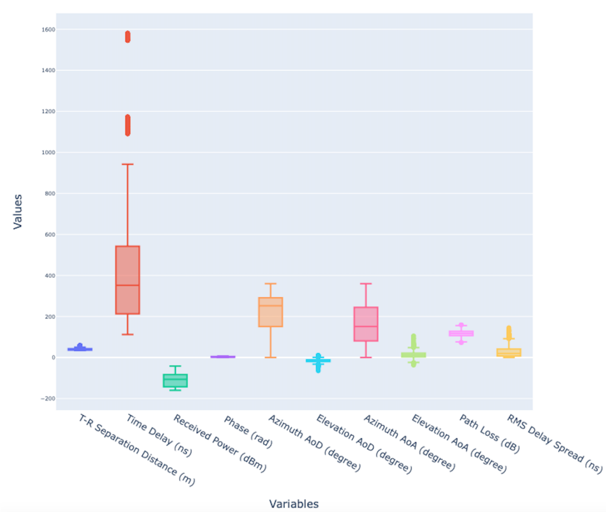

Fig. 1. Boxplot Analysis

A. K-Means

The K-Means algorithm is commonly used in the field of clustering to group data into sets based on their similarity [15]. Kmeans is based on minimizing the sum of squared distances between data points and cluster centroids. Formally, the K-Means algorithm can be described using the following equations:

Algorithm 1 explores the procedure of K-means as it starts with the initialization of the centroid deterministically and

Algorithm 1 K-Means Algorithm

Input: The CSI data X_i
Output: Clustering the provided data X_i
Initialisation : K initial centroids are randomly
for $\forall\ X_i$ **do**
E_d to each centroid C_j is calculated
if $(E_d \neq 0)$ **then**
$X_i \rightarrow \forall\ C_j$.
end if
$C_J = \text{argmin}\,||X_i - C_j||^2$
end for
return

then assigns each point to the closest centroids based on the Euclidean distance E_d to each centroid C_j. Then, once all points have been assigned to a certain centroid, the centroids are updated by recalculating their position as the average of all points assigned to that centroid which is continued until a convergence criterion is met, such as centroid stability or the maximum number of iterations.

The K-means algorithm aims to find the optimal configuration of centroids that minimizes the sum of squared distances within each cluster. Although widely used, it is important to consider its limitations, such as sensitivity to centroid initialization and inefficiency in the presence of unevenly sized or dense clusters.

The clustering metrics that will be involved in evaluating clustering the channel state information of the 6G propagation scenario are the Silhouette Score, Calinski-Harabasz Score, and Davies-Bouldin Score. The Silhouette Score measures the degree of separation between clusters. A higher score indicates better-defined clusters. While the Calinski-Harabasz Score evaluates the ratio of between-cluster dispersion to within-cluster dispersion. A higher score suggests well-separated and compact clusters. The Davies-Bouldin Score measures both the separation and compactness of clusters. A lower score indicates better clustering results. Overall, these metrics can guide further analysis and potential improvements, such as enhancing cluster separation for better-defined clusters.

Clustering Algorithm	Silhouette Score	Calinski-Harabasz Score	Davies-Bouldin S
KMeans	0.174574	1120.308620	2.043847
GMM	0.122913	649.720022	2.278664
Agglomerative	0.150975	902.502864	2.268344

TABLE I
AN ASSESSMENT OF THE CLUSTERING ALGORITHMS

To illustrate the metrics more, the Silhouette Score measures the degree of separation between clusters. A higher score indicates better-defined clusters. While Calinski-Harabasz Score evaluates the ratio of between-cluster dispersion to within-cluster dispersion. A higher score suggests well-separated and compact clusters. The high value obtained here indicates distinct and low within-cluster variance clusters. Then, the Davies-Bouldin Score is responsible for measuring clusters' separation and compactness. A lower score indicates better clustering results. Based on these results obtained from comparing cluster model metrics, we can make the following comments:

- KMeans: The KMeans model demonstrates the best performance in categorizing the binary variable 'Classes' of the CSI of 6G data. With a Silhouette Score of 0.175, it indicates a reasonable level of cluster cohesion and separation as can be seen in figure 2. The Calinski-Harabasz Score of 1120.309 suggests good inter-cluster separation and intra-cluster similarity, supporting its effectiveness in distinguishing the two binary categories. Furthermore, the Davies-Bouldin Score of 2.044 indicates that the clusters formed by KMeans have reasonably well-defined boundaries for the 'Classes' variable.

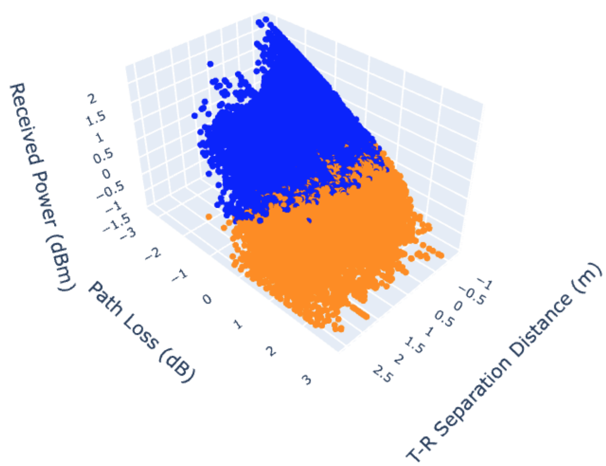

Fig. 2. K-Means Clustering

In general, the cluster models exhibit moderate performance in categorizing the data. The average Silhouette Score across all models is approximately 0.152, indicating a moderate level of cluster cohesion and separation. The average Calinski-Harabasz Score is around 910.396, suggesting reasonable inter-cluster separation and intra-cluster similarity, which implies the models' potential for distinguishing different clusters. The average Davies-Bouldin Score of 2.192 reflects a moderate level of cluster quality overall. It is worth noting that the K-Means model achieved the highest Silhouette Score and Calinski-Harabasz Score among the models, indicating stronger cluster separation and cohesion. On the other hand, the GMM model obtained the lowest scores, suggesting less distinct clustering. However, all models have shown reasonable performance in forming clusters and can be considered for categorizing the data of 6G.

IV. Future Works

Future works of the integration of ALI and wireless communication systems are wide. The role of AI in improving the performance of wireless networks, such as enhancing the quality of service, increasing the network capacity, and reducing latency. Moreover, researchers may examine how AI can be used to optimize network resource allocation, including spectrum allocation, power allocation, and bandwidth allocation. Furthermore, how AI improves wireless security, including intrusion detection, threat prediction, and prevention.

V. Conclusions

Nowadays, B5G New Radio is expected to have more progress beyond 2023 to reach the level of 6G requirements that will be settled by ITU. It is expected that the normal development of another communications era will not stop at 5G. The primary 6G association is supposed to be conveyed in 2030, given the tremendous interest in delivering 6G from the scholastic local area and industry. 6G is currently under standardization and development. 6G will oblige the use cases and applications given in 5G, for instance, IoT, Industry 4.0, PC-created reproduction, and modified driving, with further developed support in a more expense-capable, energy-powerful, and resource-useful strategy. In the interim, it will empower exceptional use cases that 5G cannot support, for example, holographic correspondences, inescapable knowledge, and worldwide ubiquitous connectivity, as well as other complex uses that we cannot yet envisage. 6G will be updated and practical for the new use cases presented in 5G, empowering their far-reaching reception. At the same time, it will empower novel use cases that we cannot yet wholly imagine or portray. The development of adaptable cells to verticals, which started with the presentation of minimal expense IoT upgrades in 4G and very strong low-lethargy IIoT in 5G, will keep turning out to be more significant in 6G. The quick progression of AI/ML development and its feasibility in tending to difficulties in a few regions shows the requirement for a 6G system to exploit these new capacities to improve execution by better adjusting to the practical environment. Thus, in this manuscript, unsupervised learning has been conducted to investigate the channel state information of the expected wireless channel modeling of 6G. Clustering algorithms such as K-Mean, and Gaussian Mixture Model have been used and compared with Agglomerative Hierarchy to explore CSI of high MmWave bands.

References

[1] Rahman, M. & Khatun, F. Challenges and Prospective of AI and 5G-Enabled Technologies in Emerging Applications during the Pandemic. *Industry 4.0-Perspectives And Applications*. (2023)

[2] Lu, Y. Security in 6G: The prospects and the relevant technologies. *Journal Of Industrial Integration And Management*. **5**, 271-289 (2020)

[3] Akyildiz, I., Kak, A. & Nie, S. 6G and Beyond: The Future of Wireless Communications Systems. *IEEE Access*. **8** pp. 133995-134030 (2020)

[4] Bhat, J. & Alqahtani, S. 6G ecosystem: Current status and future perspective. *IEEE Access*. **9** pp. 43134-43167 (2021)

[5] Raj, V. & CA, A. Understanding the future communication: 5G to 6G. *International Research Journal On Advanced Science Hub*. **3** pp. 17-23 (2021)

[6] Lu, Y. & Zheng, X. 6G: A survey on technologies, scenarios, challenges, and the related issues. *Journal Of Industrial Information Integration*. **19** pp. 100158 (2020)

[7] Lu, Y. & Zheng, X. 6G: A survey on technologies, scenarios, challenges, and the related issues. *Journal Of Industrial Information Integration*. **19** pp. 100158 (2020), https://www.sciencedirect.com/science/article/pii/S2452414X20300339

[8] Wang, C., You, X., Gao, X., Zhu, X., Li, Z., Zhang, C., Wang, H., Huang, Y., Chen, Y., Haas, H. & Others On the road to 6G: Visions, requirements, key technologies and testbeds. *IEEE Communications Surveys & Tutorials*. (2023)

[9] He, R., Ai, B., Molisch, A., Stuber, G., Li, Q., Zhong, Z. & Yu, J. Clustering enabled wireless channel modeling using big data algorithms. *IEEE Communications Magazine*. **56**, 177-183 (2018)

[10] Yang, M., Ai, B., He, R., Chen, L., Li, X., Li, J., Zhang, B., Huang, C. & Zhong, Z. A cluster-based three-dimensional channel model for vehicle-to-vehicle communications. *IEEE Transactions On Vehicular Technology*. **68**, 5208-5220 (2019)

[11] Wang, Q., Ai, B., He, R., Guan, K., Li, Y., Zhong, Z. & Shi, G. A framework of automatic clustering and tracking for time-variant multipath components. *IEEE Communications Letters*. **21**, 953-956 (2016)

[12] Li, Y., Zhang, J., Ma, Z. & Zhang, Y. Clustering analysis in the wireless propagation channel with a variational Gaussian mixture model. *IEEE Transactions On Big Data*. **6**, 223-232 (2018)

[13] Letaief, K., Chen, W., Shi, Y., Zhang, J. & Zhang, Y. The Roadmap to 6G: AI Empowered Wireless Networks. *IEEE Communications Magazine*. **57**, 84-90 (2019)

[14] Aldossari, S. Predicting Path Loss of an Indoor Environment Using Artificial Intelligence in the 28-GHz Band. *Electronics*. **12**, 497 (2023)

[15] Wu, J. Advances in K-means clustering: a data mining thinking. (Springer Science & Business Media,2012)