BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 73(2), 2025, Article number: e153426
DOI: 10.24425/bpasts.2025.153426

ELECTRONICS, TELECOMMUNICATION AND OPTOELECTRONICS

Machine learning-based throughput enhancement in fifth-generation networks

Parameswaran Ramesh*, **Bhuvaneswari P T V**, **Ashok R S** and **Veena S**

Department of Electronics Engineering, Madras Institute of Technology, Anna University, Chennai, 600044 India

Abstract. The 5G enhanced mobile broadband (eMBB) category offers faster data rates, network capacity, and user experiences than prior generations. This research aims to boost the 5G uplink user equipment (UE) user data transfer rate. We use Python to build frameworks and analyze data. A 250-m-radius centre-excited picocell base station (PBS) is investigated to support 15 clients. Cell-range Poisson distribution determines user position. All UEs send channel state information (CSI) to the PBS, which evaluates signal transmission channel conditions. The study uses Rayleigh, Rician, free space path, and long-distance route loss models. This inquiry produces a channel state dataset and then it is formulated dataset is dynamic. For service-specific requirements, UEs use k-means clustering. Clustering concatenates bandwidth, enhancing system efficiency and UE sum rate. The research includes observations from simulation findings, in which UEs are grouped by channel gain, achievable data rate, and minimum service-required data rate. Users in cluster 3 achieve the highest cumulative rate of 9.09 Mbps after clustering with an average of 7.16 Mbps. Bandwidth concatenation increased system capacity, meeting each UE service needs. After evaluating performance criteria for different clustering models, k-means remains the best algorithm for the framework. The methodology was carefully designed to satisfy study goals. This paper investigates beamforming and dynamic clustering to improve user fairness and performance.

Keywords: 5G; channel state information; bandwidth; machine learning; clustering.

1. INTRODUCTION

Fifth-generation (5G) wireless technology is the latest mobile communication standard. 5G builds on its predecessors to improve capacity, decrease latency, speed up data transfer, and support a large number of networked devices [1]. This innovative technology can alter transportation, healthcare, telecommunications, and entertainment, as well as improve user experiences through faster downloads and uninterrupted connectivity [2].

5G technologies limited global coverage benefitting metropolitan areas, which is one of the main drawbacks [3] since remote places may not receive 5G for years. Although 5G technologies limit transfer speeds to 100 Mbps, they offer rapid download rates, marking a significant improvement over 4G. Mobile phone battery technology must also improve for 5G to work. 5G supports many use cases, including massive machine type communication (mMTC), enhanced mobile broadband (eMBB), and ultra-reliable low latency communication (URLLC) as the most popular services.

Sum rates are the maximum data rates available to all users in an area at the same time. This essential parameter affects network capacity and user experience. 5G optimizes spectrum efficiency, interference control, and resource allocation to maximize the total rate. This enhancement speeds up data rates, lowers latency, and improves bandwidth-intensive applications like UHD video streaming, virtual reality (VR), augmented reality (AR), and the Internet of Things (IoT). To efficiently serve many customers and supply varied services with reliability, we need high sum rates [4].

The sum rate is critical in 5G as a performance indicator that accurately captures the total data throughput of communication systems, especially in multi-user scenarios. Maximizing the overall rate enhances network capacity and throughput [5] and 5G networks can manage more users, providing higher data rates by optimizing resource allocation to optimize the total rate, and increasing Quality of Service (QoS) metrics like latency and reliability. Sum rate optimization ensures network adaptation in dynamic communication contexts [6]. However, complex optimization methods and advanced interference control solutions make it challenging to achieve the greatest data transfer speeds in 5G networks. The susceptibility of sum rate optimization to channel fluctuations and performance measure concessions may make stability and scalability in large networks difficult. Despite these challenges, sum rate optimization research and development emphasize the need to improve 5G communication systems [7].

The 5G service category of eMBB offers faster data rates, network capacity, and user experiences than prior generations. It aims to achieve speeds of multiple gigabits per second. This accelerates the transfer of huge files, HD videos, and data-intensive applications. It decreases the duration of data transfer between the user device and the network server, especially for online gaming, VR, AR, and interactive multimedia services that have data rate requirements [8].

5G networks can accommodate several users accessing and transmitting data without sacrificing performance. 5G uses sophisticated modulation, massive MIMO, beamforming, and spectrum sharing to maximize spectral efficiency [9]. This opti-

*e-mail: parameswaran0789@mitindia.edu

Manuscript submitted 2024-10-22, revised 2025-01-06, initially accepted for publication 2025-01-07, published in March 2025.

mizes frequency spectrum use, enhancing network capacity and data throughput for improved eMBB services.

High-speed data transmission is in demand, so modern communication systems must optimize sum rates, which indicate data throughput. Heuristics and inflexible resource allocation algorithms may not fully leverage dynamic and complicated communication settings [10]. Machine learning (ML), whose specialty is channel prediction, can optimize communication system components by utilizing data-driven insights to enhance sum rates [11]. ML models can predict channel conditions based on historical data. These models anticipate path loss, fading, and interference to optimize transmission parameters for channel stability and sum rates [12]. Based on real-time feedback and optimization goals, the algorithms can dynamically distribute bandwidth, power, and time slots. Intelligent resource allocation lets ML-driven systems achieve the highest sum rates while also being fair and efficient in extremely dynamic and varied network environments [13]. Advanced machine learning can optimize channel prediction, resource allocation, interference control, and modulation methods to boost communication system total rates [14].

The main objective of this research is to investigate the throughput of users using ML models such as k-means, density-based spatial clustering of applications with noise (DBSCAN), and Gaussian mixture model (GMM) with the aid of clustering. This is achieved by calculating the silhouette score and the Davies Bouldin index (DBI).

The remaining sections of the paper are summarized as follows: Section 2 delves into current research on enhancing throughput in 5G networks, while Section 3 explores the current status of the proposed effort and technique. Section 4 discusses the simulation circumstances and then proceeds to an investigation of clustering using ML models. Finally, Section 5 summarizes the outcomes of the scenarios studied before and after clustering, as well as the future scope.

2. LITERATURE SURVEY

The authors in [15] introduced an approach that utilizes unsupervised machine learning and conditional independence tests (CITs) to detect network performance trends based on data. We assessed the technique by utilizing crowdsourcing data from 5G UEs and a dataset from a long-term evolution (LTE) network, using the k-means clustering algorithm. The findings indicated that the uplink throughput, as assessed, had the greatest impact on the observed performance patterns. The LTE dataset also demonstrated a link between the number of signalling resources assigned in the physical uplink control channel (PUCCH) and the uplink data transfer rate of the user equipment. Deep learning algorithms specifically designed for analyzing time-series data will expand the technique in the future.

In [16], the authors presented a scheduling technique for full-duplex wireless networks that uses reinforcement learning for clustering users to optimize the allocation of network radio resources. The algorithm does not require user-to-user channel estimation. The study introduced a reinforcement learning method for scheduling in OFDMA wireless networks, intending to simplify the scheduling process and improve spectral efficiency. The algorithm exhibited exceptional performance in scenarios involving the clustering of user equipment. The approach fails to address inter-cell interferences and instead concentrates solely on single-cell situations. Future research will focus on situations involving multiple cells and scheduling techniques.

The authors of [17] produced a better k-means clustering method that uses nonorthogonal multiple access (NOMA) for 5G cellular wireless networks. The algorithm aims to achieve the balance between the overall network throughput and fairness among devices, compared to the random-access channel (RACH) in LTE. Devices with higher channel gain are designated as cluster heads to enhance the overall network throughput. The research outperforms typical k-means by achieving a higher sum throughput in the network. We apply the algorithm iteratively to the remaining network to obtain the optimal solution for the cluster formation problem.

In [18], the authors study user clustering and power allocation schemes based on reinforcement learning for the NOMA system. The authors employ the Q-learning technique to optimize power allocation and maximize the aggregate data rate. The authors deploy the k-means algorithm to cluster users based on channel gain, thereby aiding in data rate maximization Extensive simulations confirm that the developed Q-learning technique with user clustering performs better than other scenarios, achieving the highest sum data rate. Additionally, it is capable of overcoming several NOMA constraints, such as transmission power budget limitations and minimum user data rate requirements.

In [19], the authors worked on enhancing user capacity and optimizing power allocation, considering the subchannel assignment constraints in wireless networks. The authors calculated channel data using the Shannon capacity formula and employed multiple linear regression models. We generated test data to predict the sub-channel capacity and use it to solve optimization models. The model used linear regression equations as constraints while treating power and capacity as variables. The study also investigated the enhancement of wireless networks by allocating distinct network segments to users in order to predict network performance.

In [20], the authors examined the application of ML algorithms in green cellular networks for optimizing quality of service (QoS), signal traffic load, and energy efficiency. The work additionally addressed the concept of coordinated transceiver multipoint (CoMP) in TE-advanced networks, which enhances network coverage and improves data rate. The study explored the power efficiency of green cellular communication while considering quality of service (QoS) constraints. It also highlights the power levels required for transmitting bits and analyses the relationship between power consumption and latency caused by bandwidth limitations. The paper also addressed the trade-off between energy efficiency and spectral efficiency for cellular networks.

The authors of [21] suggested a new way to group users in the downlink of the 5G NOMA system that uses artificial neural networks (ANN) to yield optimal results from the system while keeping complexity low. The authors train the ANN model using a historical dataset, which includes transmitting powers, channel gains, and user cluster information of NOMA users. After that, the model is validated to find the best hyper-parameters. This keeps the model from overfitting and lets us accurately predict how clusters will grow. The simulation results of an ANN-based user clustering framework outperform traditional orthogonal multiple access (OMA) techniques, achieving optimal throughput performance in comparison to the Brute force method while maintaining an acceptable level of clustering complexity.

In [22], the authors presented a rapid learning system called extreme learning machine-based user clustering (ELM-UC). The authors designed this scheme to operate in NOMA environments, quickly estimating the optimal formation of user clusters based on their channel gains and powers. The ELM design is well-suited for UC optimization, as it operates as a predictor using significant input data. ELM-UC methodology delivers performance that is almost ideal when compared to the brute-force search (B-FS) method. Furthermore, it surpasses existing clustering strategies such as ANN-UC and dynamic user clustering (DUC). The proposed ELM-UC scheme aims to address the issue of extended learning times in neural network-based UC schemes. It is achieved by solving the output weights in a single step, eliminating the necessity for a time-consuming backpropagation learning process.

The above literature discusses different system models with constraints to improve the 5G sum rate through user clustering, power optimization, beam forming, and other ML algorithms. However, the research has limitations, such as considering a single fading environment and not specifying the exact system scenario for deployment.

To overcome these restrictions, the proposed system scenario incorporates a real-time visible picocell base station in a hexagonal cell. Examined multiple channel conditions and path loss models during dataset construction to match real signal propagation channel circumstances. We studied performance indicators for several models to find the optimal clustering algorithm. Not only does this research attempt to improve the 5G sum rate through user clustering but it also meets the service needs of all eMBB users in the cell, which is novel.

3. PROPOSED METHODOLOGY

This section discusses the deployment of a picocell system and the creation of a tailored dataset. It details the steps for computing channel parameters for users and analyzing channel capacity to meet service-specific requirements, aiming to enhance the sum rate via user clustering algorithms. The chapter covers model construction and assesses the effectiveness of clustering in addressing diverse service needs. It aims to demonstrate the efficacy of optimizing system performance and ensuring seamless service delivery through careful examination and comparative analysis of the proposed methodology. Figure 1 presents the block diagram of the proposed ML-based throughput enhancement system.

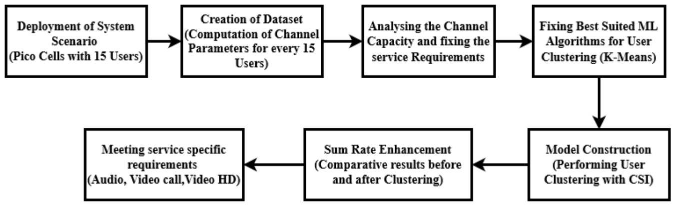

Fig. 1. Block diagram of the proposed work

3.1. System model

Deploying the picocell system with 15 users creates the operational environment for future data collection, analysis, and optimization. It places the work in a real-world context, allowing for practical insights and outcomes.

The dataset construction is crucial as it provides the raw data needed for analysis and optimization. By computing channel parameters for each user, it captures diverse real-world channel conditions. This extensive dataset forms the basis for analysis, enabling informed decision-making and optimization strategies.

The CSI equations analyze free space, log-distance path loss, Rayleigh, and Rician fading channels [23]. The model calculates SINR, communication quality metric, after channel gain. Using SINR values, the model calculates each user channel capacity, which is the maximum data rate under certain conditions.

The path loss can be shown as

$$FSPL = 20 * \log_{10}\left(\frac{4\pi d}{\lambda}\right), \tag{1}$$

$$LDPL = 20 * \log_{10}\left(\frac{4\pi}{\lambda}\left(\frac{d}{d_0}\right)^n\right), \tag{2}$$

where d_0 is the reference distance, FSPL is the free space path loss [24], LDPL is the log distance path loss [24], λ is the wavelength, and then n is the path loss exponent.

The Rayleigh channel can be calculated as shown in equation (3)

$$f(x) = \frac{x}{\sigma^2} e\left(\frac{-x^2}{2\sigma^2}\right). \tag{3}$$

The Rician channel can be calculated as shown in equation below

$$f(x\colon K) = \frac{2(k+1)}{k} e^{-k-1} I_0 \sqrt{k(k+1)x}, \tag{4}$$

where x is a signal power, σ is the Standard deviation, k is the cluster size and K is the normal distribution. The channel gain can be calculated and SINR is

$$H = 10^{-\alpha/20} * \beta, \tag{5}$$

$$\Gamma = 10 \log_{10}\left\{\frac{P_{\text{signal}}}{P_{\text{interfernce}} + P_{\text{noise}}}\right\}. \tag{6}$$

The capacity can be calculated as

$$C = B \log_2 (1 + SINR), \tag{7}$$

where α is the path loss component, β is the signal power, C is the capacity of the user, SINR is the signal-to-interference plus noise ratio and B is the bandwidth.

The system model closely interconnects the derived parameters, namely channel gain, SINR, capacity, and sum rate. Channel gain, determined by path loss and fading factors, directly influences SINR. Concerning interference and noise, the strength of the received signal increases, resulting in a higher SINR value. SINR, in turn, plays a pivotal role in calculating channel capacity, following Shannon's capacity formula, where higher SINR values lead to increased capacity. The individual capacities of user channels, influenced by their respective SINR values, collectively contribute to the overall sum rate of the system. Hence, changes in channel gain can affect SINR and capacity calculations, which in turn affect the sum rate by changing the total data rate of all users in the system. The sum rate can be calculated as

$$R\text{sum} = \sum_{i=1}^{N} c^i, \tag{8}$$

where Rsum is the sum rate and c^i is the capacity of the users varying from 1 to N. Furthermore, the model extends its analysis to encompass the collective performance of all users within the picocell through sum rate calculation. This computation aggregates individual capacities to provide insights into the system overall capacity. This systematic approach ensures a thorough understanding of the system behaviour under diverse channel scenarios, laying the groundwork for subsequent analysis and sum rate enhancement.

3.2. Channel capacity analysis

Analyzing channel capacity and fixing service requirements is critical for ensuring that the system meets users' demands and expectations. This research optimizes the system ability to deliver high-quality communication services by assessing channel capacity and aligning it with service-specific requirements. This step sets clearly defined performance benchmarks and objectives for the optimization process. According to 3GPP standards [25], Table 1 tabulates the minimum data rate required for the services considered under the eMBB application.

Each UE is assigned a fixed service that is demanded by them and their minimum required data rate ranges from 0.064 Mbps to 5 Mbps.

Resource allocation and system optimization depend on choosing the best ML technique for user clustering. We chose k-means clustering over density-based scanning (DB-scan) and the Gaussian mixture model (GMM) because it groups users quickly based on channel characteristics, which maximizes the use of resources and improves system performance. This decision sets the stage for model building, and implements user clustering using machine learning algorithms, particularly k-means clustering, with the aim of improving the sum rate. The suggested study used k-means, DB-scan, and GMM ML techniques to cluster users. The study uses machine learning performance evaluation metrics such as silhouette scores and Davies Bouldin

Table 1
eMBB service requirements

Services	Specifications	
	Data rate	64 Kbps
Audio	Channel	Stereo
	Sample rate	44 KHz
	Format	MP3
Video call	Data rate	1.5 Mbps
	Resolution	720p
	Video compression standard	H265
Video HD	Data rate	5 Mbps
	Resolution	1080p
	Video compression standard	H265

indices to select between these models. This proves that k-means is the best clustering algorithm for the database.

3.3. Model construction and sum rate analysis

Model building is essential for turning raw data into actionable insights. User clustering using channel state information (CSI) data reveals channel condition trends, which facilitates targeted tuning. This phase lays the groundwork for optimizing system performance and sum rate.

Figure 2 depicts the flowchart for the workings of the k-means clustering algorithm in the proposed system scenario, where k denotes the number of clusters ($k = 3$), Ci is the computed capacity of each user, and Cs is the required capacity for specific

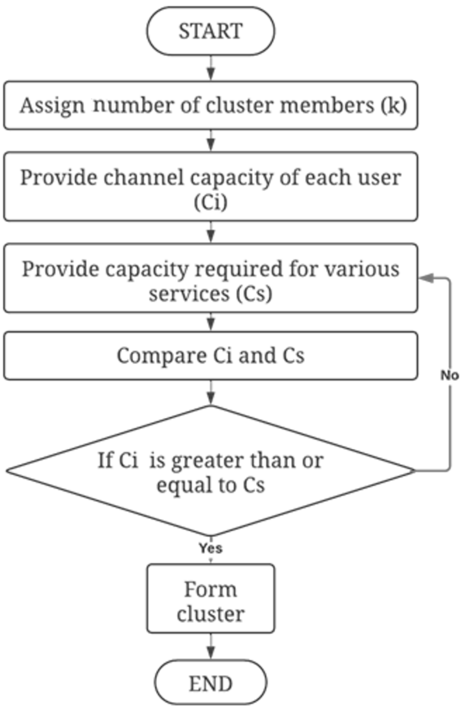

Fig. 2. Working of k-means based user clustering

services. We position the users into a cluster by maintaining the computed cluster capacity above the required data rate.

Evaluating the achieved sum rate before and after clustering provides a quantitative measure of the effectiveness of the optimization strategy and by comparing system performance metrics, the project assesses the impact of user clustering on overall system throughput. This comparative analysis adheres to excellent decision-making and optimization efforts, guiding future enhancements and improvements.

Meeting service-specific requirements is paramount for ensuring user satisfaction and system usability. This research aims to meet the service demands of all eMBB users in the hexagonal cell. Prioritizing the allocation of system capacity toward meeting these requirements, it ensures an optimal user experience across diverse communication modalities. This focus on service quality and user satisfaction underscores the commitment to delivering practical and impactful outcomes.

4. RESULTS AND DISCUSSION

These simulation settings are customized to match real-world conditions and can be performed using Python. Our analysis also includes key performance metrics derived from comprehensive simulations. This detailed analysis compares system performance both before and after clustering. Clustering methods show advantages by enhancing system performance. We also provide a comprehensive review of user clustering machine learning algorithms.

4.1. Simulation parameters

Deploying the picocell system with 15 users creates the operational environment for future data collection, analysis, and optimization. It places the work in a real-world context, allowing for practical insights and outcomes. Table 2 tabulates the considerations for this research.

Table 2
Simulation parameters

S. No	Parameters	Values
1	Number of UEs	15
2	Cell type	Picocell (100–250 m)
3	Path loss	Free space and log distance
4	Fading channels	Rayleigh and Rician
5	Number of pico BS	1
6	Carrier frequency	6 GHz
7	Bandwidth	10 MHz
8	Transmit power	30 dBm

4.2. Performance metrics

This study calculates the capacity and sum rate for each UE in the system under consideration. For the system scenario, we obtain the following simulation parameters for the free-space Rayleigh fading channel. Tables 3 and 4 compute the channel gain, SINR, and Sum rate.

Table 3
Computation of channel gain and SINR

UE ID	Distance (m)	Channel gain	SINR (Watt)
1	131	2.29826E–05	0.007524608
2	137	7.10471E–05	0.003667752
3	147	1.97082E–05	0.017823362
4	155	3.18995E–05	0.007800755
5	137	4.55325E–05	0.000531125
6	162	2.26961E–05	0.011070093
7	138	4.79879E–05	0.01448357
8	147	4.93726E–05	0.02258521
9	158	3.08838E–05	0.00573407
10	145	5.22657E–05	0.004169871
11	156	2.70481E–05	0.013023036
12	134	5.2124E–05	0.011908173
13	141	8.57054E–05	0.016861267
14	159	3.06909E–05	0.007925278
15	145	3.13109E–05	0.004908504

Table 4
Computation of user sum rate

UE ID	Distance (m)	SINR (Watt)	Sum rate (Gbps)
1	131	0.007524608	1.08
2	137	0.003667752	0.52
3	147	0.017823362	2.54
4	155	0.007800755	1.12
5	137	0.000531125	0.07
6	162	0.011070093	1.58
7	138	0.01448357	2.07
8	147	0.02258521	3.22
9	158	0.00573407	0.82
10	145	0.004169871	0.06
11	156	0.013023036	1.36
12	134	0.011908173	1.70
13	141	0.016861267	2.41
14	159	0.007925278	1.13
15	145	0.004908504	0.07

The channel gain influences the signal strength, coverage, and throughput of the user equipment (UEs), ensuring the QoS in the simulated environment. SINR is then used by determining the performance and capacity of the 5G networks to calculate the sum rate of the users, and Table 3 shows the results of the channel gain and SINR value. It achieves a moderated SINR value of 5 dB to 12 dB, making it suitable for mobile broadband users.

The sum rate determines the network overall capacity and it is performed by maintaining QoS to improve user experience and make better use of available resources, and as shown in Table 4, it achieves roughly 2 Gbps for mobile users.

4.3. Performance analysis

This section includes an analysis of performance metrics, specifically the sum rate and ML model evaluation metrics. This includes comparing the sum rate before and after clustering, as well as comparing three different ML models to select the most suitable model for user clustering.

Table 5 lists each user's system capacity, along with the minimum capacity needed to meet the service-specific throughput requirements for that user. Among fifteen users, nine users met their minimum service requirements, whereas the remaining UEs were not able to satisfy the minimum service requirement. Only 60% of the users were able to meet the requirements.

Table 5
Capacity of the system before clustering

UE ID	Distance (m)	Throughput before clustering	Service demanded	Minimum service required	Service met
1	131	1.08	Audio	0.064	Yes
2	137	0.52	Audio	0.064	Yes
3	147	2.54	Video call	1.5	Yes
4	155	1.12	Video_Hd	5	No
5	137	0.07	Audio	0.064	Yes
6	162	1.58	Video call	1.5	Yes
7	138	2.07	Video_Hd	5	No
8	147	3.22	Video call	1.5	Yes
9	158	0.82	Audio	0.064	Yes
10	145	0.06	Video call	1.5	No
11	156	1.36	Video call	1.5	No
12	134	1.70	Video call	1.5	Yes
13	141	2.41	Audio	0.064	Yes
14	159	1.13	Video_Hd	5	No
15	145	0.07	Video call	1.5	No

Figure 3 displays the UEs that belong to each cluster after clustering, along with their coordinates in the picocell. Different shapes denote UEs belonging to three different clusters.

Figure 4 represents a unique cluster of users, classified based bandwidth and SINR, and calculates the total rate. Therefore, clusters with higher sum-rate bars indicate superior network performance in meeting the service requirements of UEs inside such clusters. Furthermore, we compute the average throughput for all clusters to be 7.16 Mbps.

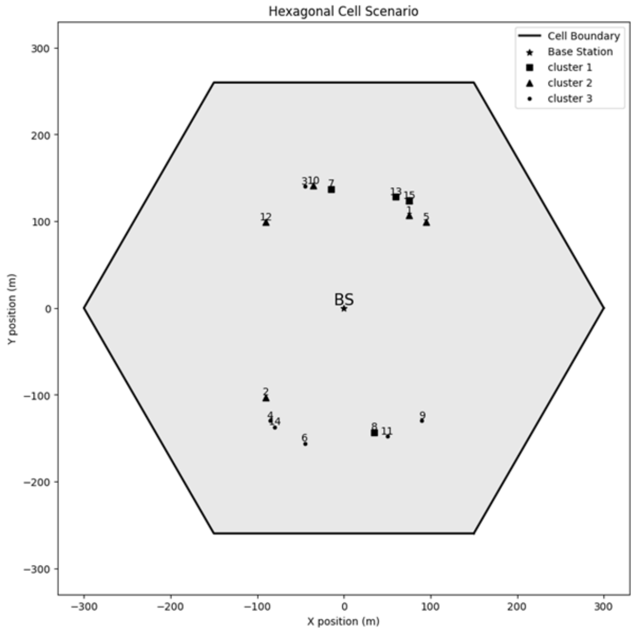

Fig. 3. Picocell scenario after clustering

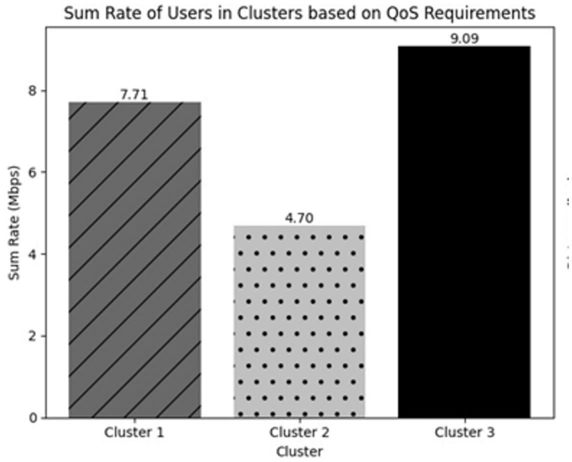

Fig. 4. Capacity after cluster formation

Picocells are used after clustering to optimize network performance. They achieve this by improving the user experience, enhancing user capacity to prevent interference, and ultimately enhancing the quality of service for consumers through throughput enhancement.

Figure 5 shows which cluster the UE belongs to and its distance from the BS.

The distances within the cluster, which comprises three users, are much closer than the other clusters. Cluster 1 has smaller distances; this means the cluster members are likely to be close or far apart depending on the distances given. Cluster 2 is located on the opposite side, and the distance between it and the other two clusters is significantly smaller than that of Cluster 1 and significantly larger than that of Cluster 3. Such an approach could

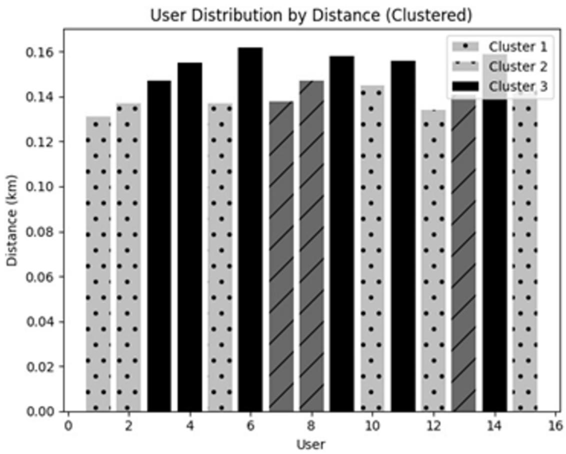

Fig. 5. Distance of user inside picocell

potentially detect trends in user distances, as well as investigate aspects such as user behaviour, regional distribution, and future network construction.

Figure 6 depicts the maximum achievable capacity for each cluster (k) and the minimum required capacity (P) for that cluster to satisfy every UE service-specific requirement.

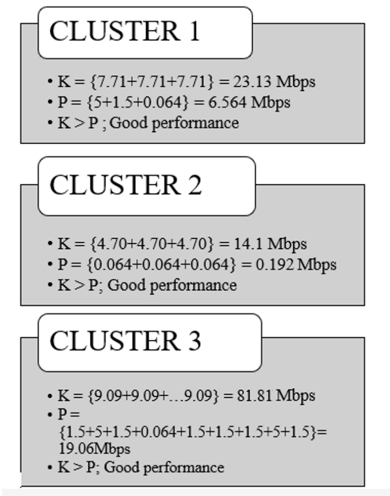

Fig. 6. Maximum achievable capacity for each cluster

The combined SINR values for each cluster are given, providing information on the signal quality that users in these clusters experience. Aggregate SINR for each cluster: [0.053, 0.032, 0.063].

When approaching the first base station, the throughput after clustering proves just 4.70 Mbps – about a factor of two lower than what it is without base stations routing information into place. As the distance from the base station increases, after clustering, throughput seems relatively stable an effect that suggests improving longer distances. Clustering maintains high throughput within a specific range or across all ranges. Even at a further distance of 156 meters, throughput is still 9.09 Mbps, which is quite good for video calls and HD video streaming.

After clustering, the throughput reaches its peak in the process, maintaining or surpassing the minimum service for all users, regardless of their distance from the base station.

Hence, picocell clustering probably enhanced and relayed the capabilities of the network and suppressed interference, thus allowing users, even if they are situated a maximum of 156 meters away, to receive an extremely high throughput that satisfies their service requirements.

Table 6 illustrates how the capacity after clustering enables each user to meet their service requirements, as the minimum required throughput is less than the obtained capacity after clustering. This ensures that clustering increases the overall system capacity, thereby enhancing the throughput of the entire system and meeting service-specific needs for every single user.

Table 6
Capacity of the system after clustering

UE ID	Distance (m)	Throughput after clustering	Service demanded	Minimum service required	Service met
1	131	4.70	Audio	0.064	Yes
2	137	4.70	Audio	0.064	Yes
3	147	9.09	Video call	1.5	Yes
4	155	9.09	Video_Hd	5	Yes
5	137	4.70	Audio	0.064	Yes
6	162	9.09	Video call	1.5	Yes
7	138	7.71	Video_Hd	5	Yes
8	147	7.71	Video call	1.5	Yes
9	158	9.09	Audio	0.064	Yes
10	145	4.70	Video call	1.5	Yes
11	156	9.09	Video call	1.5	Yes
12	134	4.70	Video call	1.5	Yes
13	141	7.71	Audio	0.064	Yes
14	159	9.09	Video_Hd	5	Yes
15	145	4.70	Video call	1.5	Yes

4.4. ML model comparison

We test different ML models for user clustering. We use the silhouette score and Davies-Bouldin index. We test all clustering models, including k-means, DBSCAN, and Gaussian mixture models (GMM), to divide UEs into clusters.

Higher silhouette ratings indicate more distinct clusters and levels of cohesiveness and separation. However, the Davies-Bouldin index evaluates cluster separation by comparing cluster centroids and diameters, and lower values indicate better clustering. A careful analysis and comparison of these measures across multiple clustering models reveals subtle pros and cons of each technique.

Table 7 shows that the k-means algorithm outperforms the other two models by obtaining better evaluation scores for both metrics.

Table 7
Clustering performance of ML models

S. No	ML model	Silhouette score	Davies Bouldin index
1	k-means	0.343	0.8144
2	DBSCAN	0.196	2.006
3	GMM	0.285	1.0598

5. CONCLUSIONS

To conclude the system deployment scenario, we examined four different channel scenarios, each involving combinations of fading channels and path loss models. We developed a thorough dataset with 15 users, each distinguished by specific channel parameters and service requirements tailored to their unique needs. The demonstration of the k-means machine learning method for user clustering showed encouraging results in optimizing system efficiency, improving the overall sum rate, and fulfilling user service requirements.

In the future, researchers may focus on using advanced machine learning techniques, such as extreme learning machines (ELM), k-means, convolutional neural networks (CNNs), and deep neural networks (DNNs), to compare how well different methods work. Additionally, they can explore beamforming systems that focus on improving signal delivery to individual users. Real-time feedback on quality of service (QoS) can guide the implementation of dynamic user clustering processes. This will lead to ongoing enhancements in network performance and user satisfaction.

REFERENCES

[1] R. Dangi, P. Lalwani, G. Choudhary, I. You, and G. Pau, “Study and Investigation on 5G Technology: A Systematic Review,” *Sensors*, vol. 22, p. 26, 2021, doi: 10.3390/s22010026.

[2] P. Ramesh *et al.*, “Software Defined Network Architecture Based Network Slicing in Fifth Generation Networks,” *Inf. Midem-J. Microelectron. Electron. Compon. Mater.*, vol. 54, no. 2, pp. 123–130, 2024.

[3] I. Maulani and C. Johansyah, “The development of 5G technology and its implications for the industry,” *Devotion J. Commun. Serv.*, vol. 4, pp. 631–635, 2023, doi: 10.36418/devotion.v4i2.416.

[4] X. Tian, K. Xiong, R. Zhang, P. Fan, D. Niyato, and K.B. Letaief, “Sum Rate Maximization in Multi-Cell Multi-User Networks: An Inverse Reinforcement Learning-Based Approach,” *IEEE Wirel. Commun. Lett.*, vol. 13, no. 1, pp. 4–8, Jan. 2024, doi: 10.1109/LWC.2023.3292280.

[5] Y. Liu and W. Chen, “Capacity analysis and sum rate maximization for the SCMA cellular network coexisting with D2D communications,” *Chin. Commun.*, vol. 19, no. 10, pp. 55–68, Oct. 2022, doi: 10.23919/JCC.2022.10.004.

[6] C. She, R. Dong, W. Hardjawana, Y. Li, and B. Vucetic, “Optimizing Resource Allocation for 5G Services with Diverse Quality-of-Service Requirements,” *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, USA, 2019, pp. 1–6, doi: 10.1109/GLOBECOM38437.2019.9014271.

[7] Parameswaran R and Bhuvaneswari P.T. V, “Performance Analysis of NOMA under Power Control Mechanism,” *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Kuala Lumpur, Malaysia, 2021, pp. 1–5, doi: 10.1109/i-PACT52855.2021.9696636.

[8] N.H.M. Adnan, I.M. Rafiqul, and A.H. M.Z. Alam, “Massive MIMO for Fifth Generation (5G): Opportunities and Challenges,” *2016 International Conference on Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, 2016, pp. 47–52, doi: 10.1109/ICCCE.2016.23.

[9] J.A. Adebusola, A.A. Ariyo, O.A. Elisha, A.M. Olubunmi, and O.O. Julius, “An Overview of 5G Technology,” *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, Ayobo, Nigeria, 2020, pp. 1–4, doi: 10.1109/ICMCECS47690.2020.240853.

[10] P. Ramesh, S. Sandhiya, S. Sattainathan, Lara Levincy A, Bhuvaneswari P.T.V, and Ezhilarasi S, “Silhouette Analysis Based K-Means Clustering in 5G Heterogenous Network,” *2023 International Conference on Intelligent Technologies for Sustainable Electric and Communications Systems (iTech SECOM)*, Coimbatore, India, 2023, pp. 541–545, doi: 10.1109/iTech-SECOM59882.2023.10435234.

[11] J. Yuan, H.Q. Ngo and M. Matthaiou, “Machine Learning-Based Channel Prediction in Massive MIMO With Channel Aging,” *IEEE Trans Wirel Commun.*, vol. 19, no. 5, pp. 2960–2973, May 2020, doi: 10.1109/TWC.2020.2969627.

[12] J. Hassannataj Joloudari, R. *et al.,* “Resource allocation optimization using artificial intelligence methods in various computing paradigms: A Review,” *EasyChair Preprint 7645, version 2*, 2022, doi: 10.13140/RG.2.2.32857.39522.

[13] M. Kokilavani, R. Siddharth, S. Mohamed Riyaz, P. Ramesh, S. Ezhilarasi and P.T.V. Bhuvaneswari, “Downlink Performance Analysis of 5G PD-NOMA System,” *2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT)*, New Delhi, India, 2022, pp. 1–6, doi: 10.1109/GlobConPT57482.2022.9938331.

[14] T. Xu, M. Zhao, X. Yao, and Y. Zhu, “An improved communication resource allocation strategy for wireless networks based on deep reinforcement learning,” *Comput. Commun.*, vol. 188, pp. 90–98, 2022, doi: 10.1016/j.comcom.2022.02.018.

[15] M. Sousa, P. Vieira, M.P. Queluz, and A. Rodrigues, “Towards the use of Unsupervised Causal Learning in Wireless Networks Operation,” *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 9, p. 101764, 2023.

[16] H. Fawaz, M. El Helou, S. Lahoud, and K. Khawam, “A reinforcement learning approach to queue-aware scheduling in full-duplex wireless networks,” *Comput. Netw.*, vol. 189, p. 107893, 2021.

[17] E. Cabrera and R. Vesilo, “An Enhanced K-means Clustering Algorithm with Non-Orthogonal Multiple Access (NOMA) for MMC Networks,” *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*, Sydney, Australia, 2018, pp. 1–8, doi: 10.1109/ATNAC.2018.8615298.

[18] S. Rezwan, S. Shin, and W. Choi, “Efficient User Clustering and Reinforcement Learning Based Power Allocation for NOMA Systems,” *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju,

Korea (South), 2020, pp. 143–147, doi: 10.1109/ICTC49870.2020.9289376.

[19] P. Adasme, A. Viveros, M.S. Ayub, I. Soto, A.D. Firoozabadi, and D.Z. Rodríguez, "A Multiple Linear Regression Approach to Optimize the Worst User Capacity and Power Allocation in a Wireless Network," *2023 South American Conference on Visible Light Communications (SACVLC)*, Santiago, Chile, 2023, pp. 6–11, doi: 10.1109/SACVLC59022.2023.10347689.

[20] M.M. Saeed, R.A. Saeed, M.A. Azim, E.S. Ali, R.A. Mokhtar, and O. Khalifa, "Green Machine Learning Approach for QoS Improvement in Cellular Communications," *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, Sabratha, Libya, 2022, pp. 523–528, doi: 10.1109/MI-STA54861.2022.9837585.

[21] S.P. Kumaresan, C.K. Tan, and Y.H. Ng, "Efficient User Clustering Using a Low-Complexity Artificial Neural Network (ANN) for 5G NOMA Systems," *IEEE Access*, vol. 8, pp. 179 307–179 316, 2020, doi: 10.1109/ACCESS.2020.3027777.

[22] S.P. Kumaresan, C.K. Tan, and Y.H. Ng, "Extreme Learning Machine (ELM) for Fast User Clustering in Downlink Non-Orthogonal Multiple Access (NOMA) 5G Networks," *IEEE Access*, vol. 9, pp. 130884–130894, 2021, doi: 10.1109/ACCESS.2021.3114619.

[23] Deepak Athipan A.M.B, Subha Dharshini M, P. Ramesh, Shanmugapriya R, "Investigation on the Impact of CQI Index in Performance Enhancement of 5G Downlink Systems," *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2024, pp. 162–166, doi: 10.1109/ICACCS60874.2024.10716972.

[24] A.I. Sulyman, A. Alwarafy, G.R. MacCartney, T.S. Rappaport and A. Alsanie, "Directional Radio Propagation Path Loss Models for Millimeter-Wave Wireless Networks in the 28-, 60-, and 73-GHz Bands," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 10, pp. 6939–6947, Oct. 2016, doi: 10.1109/TWC.2016.2594067.

[25] 3rd Generation Partnership Project (3GPP), "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP TR 38.901, Release 16, March 2019.