Pairing mobile users using K-means algorithm on PD-NOMA-based mmWaves communications system

Litim Abdelkhaliq[1], Bendimerad Mohammed Yassine[2]
[1]LTIT Laboratory, Department of Electrical and Electronic Engineering, Faculty of Technology, Tahri Mohamed University, Bechar, Algeria
[2]LTT Laboratory, Department of Telecommunication, Faculty of Technology, University of Tlemcen, Tlemcen, Algeria

Article Info

Article history:

Received Jan 21, 2025
Revised Apr 10, 2025
Accepted Jul 3, 2025

Keywords:

5G
K-means
ML
NOMA
Pairing users
Rayleigh fading channel
SISO

ABSTRACT

In this research, we study the effectiveness of the K-means machine learning (ML) clustering approach for pairing mobile users on a power domain non-orthogonal multiple access (PD-NOMA) single input single output (SISO) downlink-based millimeter-wave (mmWave) communication system. The basic concept is to pair the mobile users by using a data set that contains essential information about the mobile users in the micro cell base station (BS) (e.g., the SNR, the distance between the mobile users and the BS, the channel gain, and the data rate of each mobile user). The study conducted in this paper demonstrates that the proposed K-means clustering-based scheme achieves a balance between computational complexity and performance metrics. It outperforms single carrier NOMA (SC-NOMA), the conventional NOMA pairing scheme, and time division multiple access (TDMA), offering an effective trade-off between system efficiency and implementation feasibility.

Corresponding Author:

Litim Abdelkhaliq
LTIT Laboratory, Department of Electrical and Electronic Engineering, Faculty of Technology
Tahri Mohamed University
Bechar, Algeria
Email: litim.abdelkhaliq@univ-bechar.dz

1. INTRODUCTION

By 2023, over 8 billion connected devices were using cellular network services, driving a sharp increase in data traffic compared to 2014. Global mobile subscriptions reached approximately 8.9 billion in 2023, up from 8.6 billion in 2022, surpassing 8 billion for the first time in 2019 [1], [2]. Internet of things (IoT), high quality multimedia transmission, video conferencing, cloud computing, and so on, these Recent technological breakthroughs necessitate an increase in data rate [3].

The new generation of mobile cellular networks (fifth-generation (5G) and 6G) enable much higher connection IoT devices, low-latency network response, and faster and better mobile Internet access. Reliability, spectral efficiency, and latency are the main performance measures. These systems offer operators the best possible quality, capacity, and resources for a range of use cases [4]. They are expected to handle faster data rates and provide connectivity to a large number of devices in order to meet this growing demand for data services [5]. These types of mobile wireless networks are required to meet the current bandwidth scarcity problem and can provide peak data rates by utilizing the upper band of the frequency spectrum. However, the communication at these bands is highly sensitive to environmental conditions [6].

In order to evaluate and ensure the performance, quality, and efficiency of a 5G network, there are three fundamental concepts [2], [4], [7]: 1) the enhanced mobile broadband (eMBB) is one of service categories in 5G communication system which allows for downlink providing services of up to 1 Gbit/s. It is anticipated that the spectral efficiency will rise by a factor of 5 to 15 in comparison to 4G. 2) The second service category in the 5G communication system is the ultra-reliable low latency communication (URLLC) which aims to achieve zero latency or very low latency on the order of one millisecond. URLLC enable services for critical applications such as robotic surgeries, e-health, self-driving, vehicle to anything (V2X) which was not possible so far in the legacy technology. 3) The last fundamental concept is massive machine to machine communication (mMTC) which satisfy the massive connectivity requirements for the IoT, the connectivity density aim is at least 1 million devices per square kilometer, which is ten times more than that of 4G.

Millimeter waves (mmWaves) refer to radio frequencies in the range of 30 GHz to 300 GHz. In the context of 5G, the mmWave frequency band typically spans from 24 GHz to 100 GHz. The specific frequency bands used for 5G can vary by region and country due to regulatory differences. The most promising ones are millimeter waves (mm-waves), specifically the 26 GHz and 38 GHz core frequencies [8], [9]. mmWaves suffer from huge propagation loss compared with other communication system in using lower carrier frequencies and are more susceptible to atmospheric absorption and obstacles like buildings, trees, and even rain, which can limit their range. Large bandwidth can be obtained with millimeter wave networking and communications, however because of the differences in propagation characteristics between mmWave and sub-6GHz, it is better suited for the small-cell coverage and short-distance high transmission [10]. The insufficient capacity of 5G wireless networks to accommodate the large number of mobile users is one of their primary issues. To overcome these challenges, 5G networks often use a combination of frequency bands, including both mmWaves for high-capacity, short-range communication in densely populated areas i.e., sub band 6 GHz, and lower-frequency bands for better coverage in suburban and rural areas. By utilizing the vast amount of bandwidth in the mmWave bands, 5G would significantly boost communication capacity [8].

One of the key technologies for distinguishing between 1G and 6G wireless systems is a multiple access scheme. By multiplexing users and sharing resources in terms of time, frequency, or code, multiple access is achieved. This suggests that users utilize the same resource in an organized way to reduce user interference [11]. It could be a waste of resources to distribute resources to users with different rate requirements using the unit of a single resource block. In short, in the upcoming 5G network, the inefficiency of orthogonal multiple access (OMA) might outweigh the advantages of mmWave communication [12].

OMA techniques include frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), and orthogonal frequency division multiple access (OFDMA). These techniques have been utilized in 1G, 2G, 3G, and 4G networks respectively. Due to the limited availability of resources (time, frequency, and code), existing OMA approaches, may face major challenges when deployed beyond 5G (B5G) cellular communication systems [13]. Another issue is that the orthogonality is usually destroyed by the channel-induced impairments, even when orthogonal time-, frequency-, or code-domain resources are used [7]. OMA schemes have simpler implementation processes, but they are unable to provide the high levels of connectivity required for future networks due to their low spectral efficiency and restricted radio resources [14].

In order to accommodate a large number of users with different needs, non-orthogonal multiple access (NOMA) is an attractive option for future wireless networks [15]. In systems designed for the post-5G future, NOMA is a crucial answer to the difficult issue of supporting a large number of IoT devices with a limited amount of radio resource blocks (RRBs) [16]. The primary characteristic that sets NOMA apart from the family of conventional OMA schemes is its ability to use non-orthogonal on a subcarrier by allocating codes or sending data simultaneously on the same frequency at different power levels. resource allocation to handle more mobile users than there are orthogonal resource slots. NOMA concepts have attracted a lot of interest for 5G cellular networks [7], [11], [17]. The fundamental idea behind NOMA is to facilitate the distribution of non-orthogonal resources among users, even if it means increasing receiver complexity in order to separate non-orthogonal signals [7]. NOMA approaches have a major effect on the reduction of latency in simultaneous transmission. This indicates that utilizing the entire bandwidth of the resource improves its spectral efficiency. NOMA has attracted the most research focus among radio access approaches proposed for 5G and beyond. In comparison to traditional OMA, NOMA is a promising solution to increase spectrum efficiency, lower latency, provide high dependability, and permit massive connections [11], [14].

NOMA system has two essential categories namely code-domain and power-domain. Code domain NOMA (CD-NOMA) was inspired by the known CDMA systems that share time-frequency resources among several users. In CD-NOMA, Mobile users are multiplexed by using unique non-orthogonal sequences that have low-density, low cross-correlation, and sparse characteristics for every mobile user [7], [11]. Power domain NOMA (PD-NOMA) gains the attention of the researchers in the recent years and explored with multiple antenna techniques, co-operative communication, device-to-device (D2D) communication, vehicle-to-vehicle (V2V) communication. While PD-NOMA does not have the drawbacks of code domain NOMA, it does require precise channel estimation and an intelligent power allocation plan in order to reach higher capacity [3]. PD-NOMA operates on the principles of successive interference cancellation (SIC) at the receiver and superposition coding (SC) of signals at the transmitter [18]. Power allocation and decoding order optimization are critical in PD-NOMA enabled devices because of the added co-channel interference [16].

In PD-NOMA network SC allows many mobile user's signals to be multiplexed over the same resource blocks (RB) at the same time, but at different transmitter side power levels [5]. Major issues in implementation of PD-NOMA that prevent it from being deployed so far are heavy calculations for power allocation coefficients at the transmitter and the receiver needs a huge amount of computation to run SIC to recognize the intended signal. When low latency and large data flow are needed, the computing performance of SIC becomes even more important [14], [19], [20]. In the case of multi-cell PD-NOMA network inter-cell interference (ICI) becomes a significant problem since it can negatively affect the performance of mobile users [5]. Another problem with PD-NOMA implementation is that, in the absence of preventive measures, the SIC-induced error propagation may significantly impact the error probability when the number of mobile users is sufficiently high. Nevertheless, the error probability can be decreased by utilizing strong channel coding algorithms, sophisticated user pairing techniques and power allocation strategies [7]. As a result, the objective of PD-NOMA should be to either increase the number of users or allocate power equally while maintaining a minimum consumption of power. To maintain user fairness, PD-NOMA increases the power of mobile users with lower channel gains [11]. Multiple mobile users can be identified from one another by taking use of SIC technique at receivers [12], the idea of SIC is illustrated in Figure 1, where the micro-BS sends overlapping signals to two mobile users, with mobile user 02 gaining a higher channel gain than mobile user 01. Mobile user 02 extract first the signal of mobile user 01 then it uses direct decoding to recover his signal.

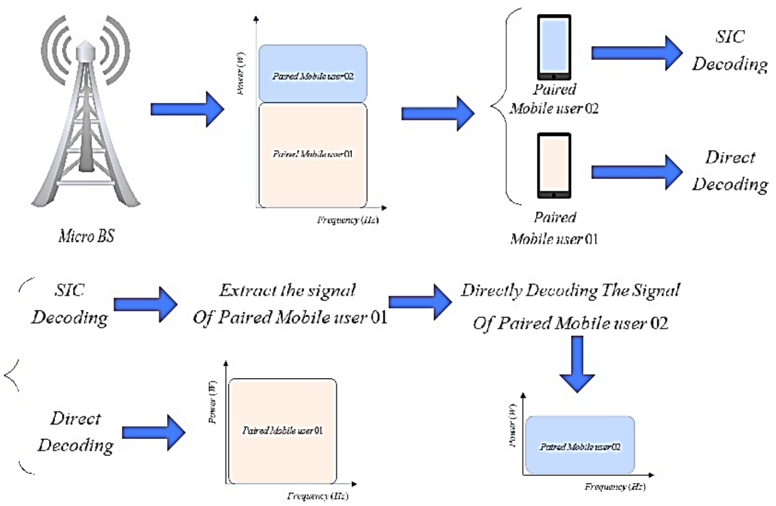


Figure 1. SIC concepts

The potential of PD-NOMA in mmWave systems can be efficiently exploited through user clustering and power allocation [21]. The combination of mmWave and PD-NOMA offers several benefits: i) highly directed transmissions due to strongly correlated user channels and ii) enhanced connectivity and larger capacity. User coupling (or pairing) is key to achieving radio capacity gain in PD-NOMA. Artificial intelligence (AI) tools are increasingly used for user pairing, with the focus on identifying a low-complexity method for pairing mobile users.

One unsupervised learning tool that can assist in extracting data points and automatically determining their relationship or similarity is the K-means algorithm [22]. K-means improves power allocation and reduces user interference by grouping users based on channel conditions, power levels, or other relevant factors. This clustering boosts the performance and capacity of the PD-NOMA system by efficiently grouping users with similar characteristics.

Recent studies primarily focus on optimizing the sum rate using channel state information (CSI) for user clustering and power allocation. However, factors like user consumption rate, inter-user distance, and quality of service (QoS) can also significantly impact user pairing and grouping. This work contributes the following:

i) We developed a method using mobile user information (SNR, distance to BS, data rate, and channel gain) and the K-means algorithm for user pairing.
ii) To meet 5G's low-latency goals, we designed a simplified pairing algorithm to minimize processing time.
iii) Our system is modeled under real-world conditions, considering rain attenuation and oxygen absorption.
iv) We aim to maximize the sum rate and energy efficiency (EE) for all mobile users.
v) Finally, we attempt to reduce the bit error rate (BER) for all mobile users.

This is how the rest of the paper is structured, section 2 present the previous works, section 3 describes our proposed method, results and discussion are presented in section 4. And, lastly, this paper is concluded in section 5.

2. STATE OF ART

This section summarizes previous efforts. Relevant works focus on user pairing methods, such as the fuzzy C-means (FCM) clustering used in [23] where users are paired based on channel correlation, gain differences, and fuzzy membership values. The key distinction between our algorithm and that in [23] is: (i) our algorithm assigns each user to a single cluster, unlike FCM, which allows users to belong to multiple clusters. (ii) Our algorithm is less complex than FCM. Instead of relying on channel conditions, it uses QoS metrics, enabling: (a) SIC decoding order and power allocation design and (b) user scheduling to meet QoS demands [18].

By considering the distribution of mobile users in a BS, Al-Abiad *et al.* [16] took advantage of the user's mmWave channel response's correlation feature to cluster the mobile users by using K-means algorithm. Parihar *et al.* [24] review recent advances, research findings, and machine learning (ML) applications in NOMA systems while addressing future research challenges for NOMA in Beyond 5G (B5G) and beyond. Jiang *et al.* [21], study user pairing and power allocation for multiple cellular users (CUs) in a downlink NOMA system. They first derive closed-form solutions for optimal power allocation on each subchannel, then use the deep Q-network (DQN) algorithm to determine the optimal user pairing scheme. In order to maximize user sum rates and improve spectrum efficiency, Ali *et al.* [25] suggested a double deep Q-network (DDQN) to optimize power allocation and user pairing simultaneously.

Mounchili and Hamouda [17], examined the optimal coupling distance between paired users (Near/Far) to enable a large number of users to participate in PD-NOMA with two users, addressing the demands of massive connectivity. Wang *et al.* [26], used deep deterministic policy gradient (DDPG) to optimize user pairing under soft and hard channel capacity. Combining actor-critic (AC) architecture with the DQN method, DDPG supports policy learning in high-dimensional or continuous action spaces. Their goal was to maximize average spectral efficiency using convolutional neural network (CNN)-based and multi-agent deep reinforcement learning (DRL) approaches with DDPG. Pu *et al.* [27] suggested a DQN-based intelligent downlink secure transmission for power line communication PD-NOMA system user pairing. By addressing policy modifications for user variations, the objective is to optimize the sum rate of nearby users while satisfying the targeted data rate requirements of every user.

Chinnadurai *et al.* [28], analyzed a MIMO-NOMA downlink system to maximize EE using a joint user pairing and dynamic power allocation (JUPDPA) architecture. The proposed pairing scheme aims to minimize inter-user interference. Lin *et al.* [29], used the angle of departure (AoD) of each mobile user as input for three methods: K-means, agglomerative hierarchical clustering (AHC), and density-based spatial clustering of applications with noise (DBSCAN), instead of using path loss. Without prior knowledge of the number of clusters, their goal was to group secondary users without impacting primary users' performance.

3. METHOD

In this section, first, we define the system model, covering user distribution, channel characteristics, and key metrics. Then, we formulate the user pairing problem with optimization criteria. Finally, we propose a K-means-based approach for efficient clustering and compare our method with Conventional methods.

3.1. System model

We examine a downlink system with a Micro BS serving J single antenna mobile users, where mobile users are dispersed at random and the micro-BS is situated in the middle of the cell. The Micro BS sends the data to all mobile users simultaneously and has limited power. The Rayleigh fading model may yield more accurate results since we study mm-waves in 5G communication systems. Since there are J users overall, the wireless communication system that contains J Rayleigh fading channels, the n Rayleigh fading channel coefficients are donated by h_n with 1≤n≤j. we consider that the CSI is well known for all Micro BS and mobile users. In real-world situations, the pilots are periodically broadcast by the Mobile users and the Micro BS use this information to estimate the CSI values [16]. The transmitted signal of Micro cell BS can be expressed by:

$$X_{Noma} = \sqrt{P_T}\,(\sqrt{\alpha}\,x_1 + \sqrt{\beta}\,x_2) \tag{1}$$

where: x_1 and x_2 are the transmitted signals to users ,P_T is the transmitted power from Micro cell BS**,** α and β are the coefficients of transmitted power with α > β and α + β = 1. The signals received at mobile user 1 and mobile user 2 can be written as follows:

$$\begin{cases} Y_1 = h_1\sqrt{P_T}\left(\sqrt{\alpha}\,x_1 + \sqrt{\beta}\,x_2\right) + \omega_1 \\ Y_2 = h_2\sqrt{P_T}\left(\sqrt{\alpha}\,x_1 + \sqrt{\beta}\,x_2\right) + \omega_2 \end{cases} \tag{2}$$

where mobile users 1 and 2's respective additive white Gaussian noise (AWGN) values are represented by the ω_1 and ω_2 variables vectors of zero mean and variance σ^2 .

To maximize the total rate in a traditional NOMA with a single-antenna base station and users, the information of the user with the lower channel gain is typically decoded first [12]. To minimize the SIC complexity at the receiver, we examine a NOMA system that only groups two mobile users together. We are able to retrieve Mobile User 1's desired signal through direct decoding, and it can be stated as (3).

$$Y_1' = \sqrt{P_T}\,\sqrt{\alpha}\,x_1 \tag{3}$$

The signal-to-interference-plus-noise ratio (SINR) of the mobile user 1 can be written as (4).

$$\lambda_1 = \frac{\alpha\, P_T\, |h_1|^2}{B_W\, \boldsymbol{\omega_1}} \tag{4}$$

Using SIC concept, the recovered and desired signal of mobile user 2 can be expressed as follow (5).

$$Y_2' = \sqrt{P_T}\,\sqrt{\alpha}\,x_2 \tag{5}$$

The mobile user 2's SINR can be expressed as (6).

$$\lambda_2 = \frac{\beta\, P_T\, |h_2|^2}{\alpha\, P\, |h_1|^2 + B_W\, \boldsymbol{\omega_2}} \tag{6}$$

The achievable rate of mobile user 1 and mobile user 2 can be stated as (7).

$$\begin{cases} R_1 = B_w\, log_2(1 + \lambda_1) \\ R_2 = B_w\, log_2(1 + \lambda_2) \end{cases} \tag{7}$$

Where: B_w is the total bandwidth of the system.

If we use Ɱ subcarriers in PD-NOMA, then the sum rate of all mobile users can be written as (8).

$$R_{sum} = \sum_{f=1}^{Ɱ} \sum_{i=1}^{2} R_{i,f} \tag{8}$$

The total power consumed by a micro cell BS is given as (9).

$$P_c = P_T + P_P \tag{9}$$

Where: P_P is the circuit power consumption of Micro BS including the power consumed by signal processing and the idle power.

Consequently, as stated in the (10), EE is defined as the ratio of the sum rate of all mobile users to total power consumption [28]:

$$\eta = \frac{R_{sum}}{P_c} \tag{10}$$

An indicator of how well the spectrum is used is spectral efficiency. A crucial performance indicator in wireless communication system design, it shows the maximum amount of data that can be sent over a specific bandwidth.

The spectral efficiency of the system can be stated as (11).

$$\eta_s = \frac{R_{sum}}{B_w} \tag{11}$$

Thus, from (10), the spectral efficiency can be expressed as (12).

$$\eta_s = \frac{\eta \, P_c}{B_w} \tag{12}$$

In the presence of interference and fading, the BER can be affected significantly. The BER of mobile user *i* under SIC can be expressed as (13).

$$\begin{cases} BER_1 = P_{er}(\lambda_1)(1 - P_{er}(\lambda_2)) \\ BER_2 = P_{er}(\lambda_2)(1 - P_{er}(\lambda_1)) \end{cases} \tag{13}$$

Where: $P_{er}(\lambda_i)$ is the probability of error given at the effective SINR for mobile user *i*.

3.2. Problem formulation

Key challenges in PD-NOMA systems include efficiently grouping users into time slots while maintaining signal quality, maximizing the sum rate, and reducing system complexity for signal transmission and decoding [30]. In this paper, the main challenge is to maximize the total rate of the mobile users with known channels while optimizing EE to reduce costs. Additionally, we aim to minimize BER for all mobile users.

The problem formulation can be stated as follow:

$$\begin{cases} \max\limits_{\alpha,\beta} R_{1,2} \\ \max\limits_{P_c, R_{sum}} \eta \\ \min\limits_{P_T} \sum_{i=1}^{J} BER_i \end{cases} \tag{14}$$

Assuming that:

$$\begin{cases} R_1 \geq r_1. \\ R_2 \geq r_2. \\ P_T \geq \sum_{i=1}^{2} p_i \end{cases} \tag{15}$$

where: the minimal rate to mobile users 1 and 2 is indicated by $R_1 \geq r_1$ and $R_2 \geq r_2$, which ensure QoS for all mobile users. Constraint $P_T \geq \sum_{i=1}^{2} p_i$ guaranteeing that the total power assigned for the paired mobile users does not exceed the Total transmitted power from Micro BS.

3.3. Proposed k-means algorithm

In this section, we will describe in full our suggested K-means method for pairing mobile users.

3.3.1. Data set generation

In this study, a dataset was generated by simulating a PD-NOMA system with 10 mobile users, where each user is characterized by its distance from the micro-BS, data rate consumption, SNR, and channel gain. Table 1 presents the ranges of generated data set.

Table 1. Dataset parameters and their ranges

Parameters	Range
Distance (m)	[100, 1000]
Data rate consumption (Mbps)	[0.1, 20]
SNR (dB)	[0, 30]
Channel gain	[1.1*10^{-12}, 8*10^{-10}]

3.3.2. Data preprocessing

After generating the dataset, a standardization method was applied to normalize the features, ensuring that parameters. This preprocessing step improves the performance and stability of the K-means clustering algorithm by eliminating scale differences between features.

3.3.3. K-means clustering algorithm

Algorithm 1 explains our suggested technique for pair mobile users. As mentioned on Algorithm 1, Ɱ represents the collection of ɱ multidimensional data points (e.g., Distance, SNR …) that need to be clustered into a set of K clusters where each cluster have the initial centroids { μ_1, μ_2, …,μ_K}. K-means seeks to minimize the sum of squared distances between each point and its designated centroid, which is known as the total intra-cluster variance or the objective function ζ ($ɱ_i, ɱ_K$). In certain references, the objective function (i.e., the cost function) is referred as the total within-cluster sum of squares (WCSS).

Algorithm 1. Mobile users pairing based on K-means algorithm

The Micro BS select K = 2 (as each mobile user in cluster C_1 should be paired with a mobile user in cluster C_2).

1) We will calculate the minimum Euclidean distance between each centroid and data points then add data point to the cluster that is closest to it. Mathematically, this can be expressed as:

$$ç_{ɱ} = \arg\min ||ɱ - \mu_K||^2 \tag{16}$$

2) Repeat the previous Step until every data point gets a label to cluster.
3) Once all the data points have been assigned to clusters, update the centroids to reflect the mean of all the points in the associated cluster, this can be expressed as:

$$\mu_K = \frac{1}{ɱ_K} \sum_{i=1}^{I} ɱ_i \tag{17}$$

Where: $ɱ_K$ is the number of data points in cluster and I is the total number of data points assigned to each cluster.

4) We repeat steps 2 through 4 to update the cluster members, until minimizing the objective function

Where: the objective function can be stated as:

$$\zeta\,(ɱ_i, \mu_k) = \sum_{k=1}^{2} \sum_{i=1}^{I} |ɱ_i - \mu_k|^2 \tag{18}$$

5) Steps 2 through 5 are repeated until the mobile users are divided into two groups.
6) We will ensure that every cluster will have an equal number of mobile users.
7) each mobile user from cluster C_1 will be paired with mobile user from cluster C_2
8) If ɱ changed (newly mobile users arrived) repeat step 2 through step 8.

Following data preprocessing, we must plot the cost function and extract the optimum value of K using the elbow approach. By taking this step, we can make sure that our selection (K = 2) is almost ideal. After this step, we can use our algorithm to separate our mobile users into two groups.

As mentioned on Algorithm 1, we must confirm that there are an equal number of mobile users in each cluster after grouping the mobile users. Then, each mobile user *(U_i)* from cluster C_1 will be paired with mobile user *(U_j)* from cluster C_2. We can represent mobile user pairs in a pairing matrix P, where each row represents a mobile user pair:

$$P = \begin{bmatrix} U_{i,1} & U_{j,1} \\ U_{i,2} & U_{j,2} \\ \vdots & \vdots \\ U_{i,\frac{J}{2}} & U_{j,\frac{J}{2}} \end{bmatrix} \tag{19}$$

Figure 2 illustrates our approach, processing key mobile user data (e.g., data rate, SNR) to optimize user pairing.

3.3.4. Comparison with conventional methods

To the best of our knowledge, existing studies pair mobile users based on a single parameter: channel correlation [16], [23], [26], distance to BS [17], [27], [28], or AoD [29]. Our proposed method utilizes all available mobile user data with a simple pairing algorithm.

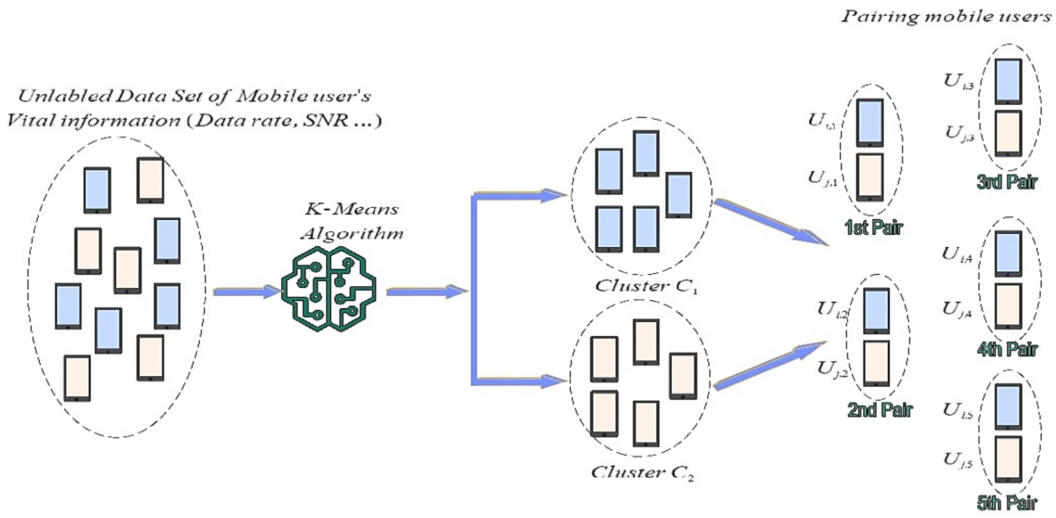


Figure 2. Proposed K-means for pairing mobile users

4. RESULTS AND DISCUSSION

In this section, we evaluate the performance of our system and provide comparative analysis and numerical evidence for the theoretical findings.

4.1. Simulation parameters

The parameters of our simulation are presented in this subsection. We employ MATLAB R2021a. We generate a random distance of [100; 1000] meters between the mobile users and the Micro BS. We employed a low mm Wave carrier frequency f_c (where f_c = 28 GHz). Table 2 is a summary of the simulation parameters. It is worth to note that we use the ITU-R model [31] with frequency-dependent coefficients (ķ and φ) to compute rain attenuation between the Micro BS and mobile user. We consider only the power amplifier's consumption for the Micro BS, assuming other components are optimal. External factors like rain and oxygen absorption attenuation are also accounted for.

Table 2. Crucial parameters configuration

Parameters	Values
Number of mobile users J	10
Path loss exponent ζ	4
Noise power density $\boldsymbol{N_0}$	10^{-17} w/Hz
Rain rate Ŗ	5 mm/h
Oxygen absorption attenuation б	0.15 dB
frequency-dependent coefficients ķ	0.13
frequency-dependent coefficients φ	0.8
The entire bandwidth $\boldsymbol{B}$	50 MHZ
Maximum transmitted power $\boldsymbol{P_T}$	100 mW
Circuit power consumption of Micro cell $\boldsymbol{P_P}$	500 mW
The number of channel realizations	10^5
Total transmitted power of the BS	20 dB
Boltzmann constant	$1.380064852\times10^{-23}$
Temperature	300 K
Speed of light $\boldsymbol{M_{Light}}$	3×10^8

The CSI is assumed to be known at both the base station and the receivers. A fixed power allocation strategy is used (α = 0.75, β = 0.25). Future work may explore ML or deep neural network (DNN) for dynamic power allocation. Mobile users are assumed to move within a predefined area, such as an office building or university, with a mobility factor of M_{Mu} = 2 km/h. The Doppler frequency can be expressed as:

$$f_d = f_c \frac{M_{Mu}}{M_{Light}} \tag{20}$$

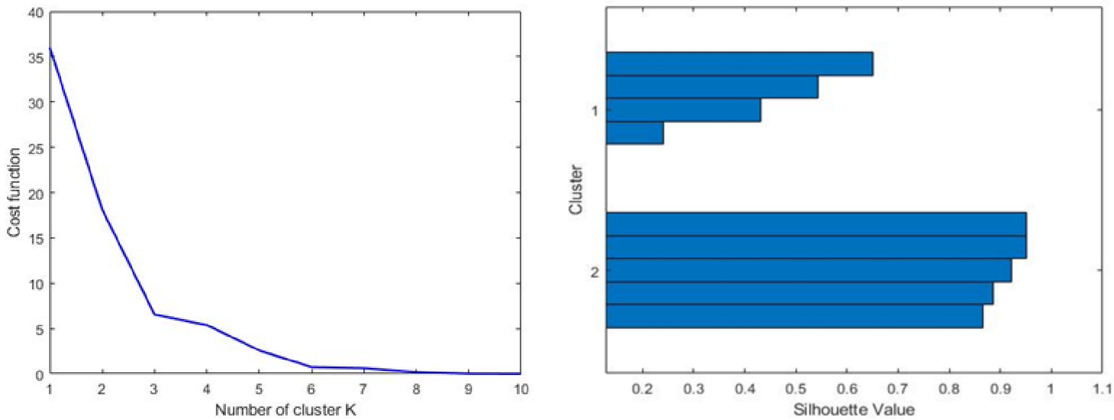
where: M_{Light} is the speed of light. The movement of Mobile users have an impact on their Rayleigh fading channel coefficients. Therefore, we took a sample of the Rayleigh fading channel coefficients of each Mobile user values based on 1/(2 f_d).

4.2. Results

This sub-section presents our simulation results. To evaluate our ML algorithm performance, we use metrics like the silhouette score presented in [32] (clustering quality) and intra-cluster distance (WCSS) (grouping similarity). In order to determine the performance of our system, we calculate the sum rate, energy efficiency, spectral efficiency, average BER, and SNR.

Figure 3 depicts the performance metrics of our algorithm. The cost function, shown in Figure 3(a), measures how well the data points are collected around the centroids of each cluster to determine how compact the clusters are. using elbow method, it is recommended to select (K=3). However, we select (K=2) in order to separate the mobile users into two clusters and pair each mobile user of Cluster C_1 with mobile user of Cluster C_2.

A measure of how well mobile users are clustered is the silhouette plot displayed in Figure 3(b). As shown, the mobile users of Cluster C_1 have lower silhouette values, ranging between 0.2 and 0.6, the mobile users of Cluster C_2 have higher silhouette values, mostly between 0.7 and 0.95. The average of silhouette value is about 0.6573. We can conclude that the clustering is acceptable, but Cluster C_1 needs improvement.



(a) (b)

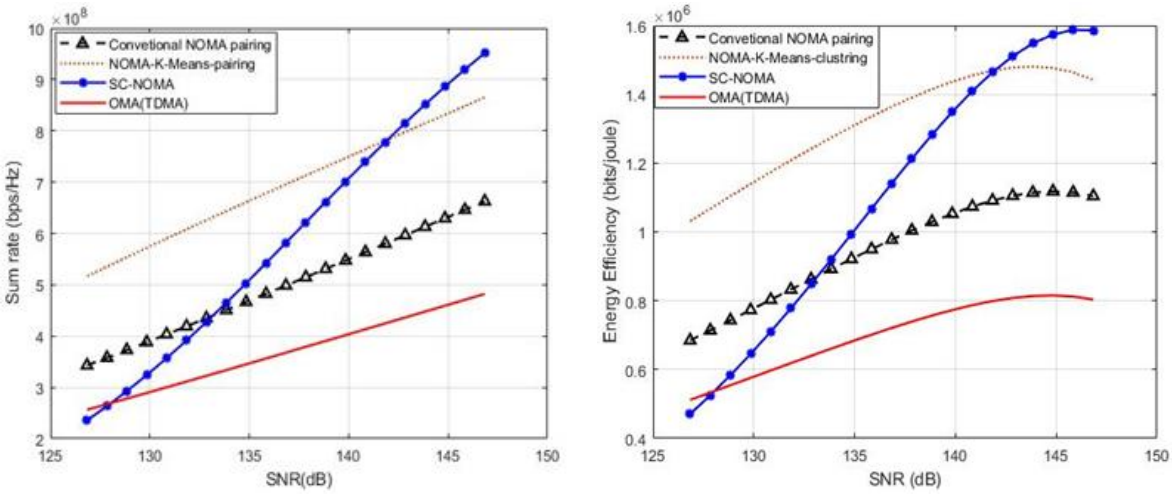
Figure 3. Performance metrics of our algorithm (a) cost function and (b) silhouette plot

Figure 4 displays our simulation's results. As shown in Figure 4(a) NOMA-K-means-pairing scheme is capable of achieving the maximum attainable sum rate, at some point SC-NOMA outperform K-means-pairing scheme at the price of high complexity of receiver. We can deduce that NOMA-K-means pairing performs better than conventional NOMA pairing, proving that system performance is improved by intelligent user clustering.

As seen in Figure 4(b), the EE of three systems (NOMA-K-means-pairing, NOMA Conventional pairing, and TDMA) increases with SNR it indicates that improving SNR enhances energy efficiency due to better error rates. SC-NOMA requires additional power to maintain its energy efficiency, its EE keeps growing but at the cost of the receiver's complexity. We may conclude that NOMA-K-means-pairing scheme provides more accurate results in terms of energy efficiency.

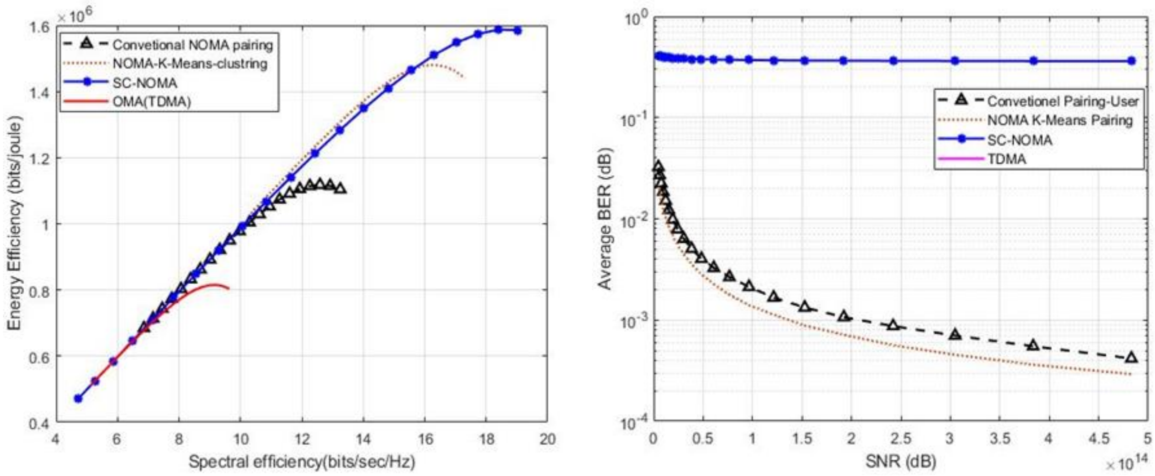
Figure 5 depicts our simulation's outcomes. As illustrated in Figure 5(a), SC-NOMA remains the best in both spectral and energy efficiency, though it is computationally more complex. NOMA-K-means-pairing suggests the optimal trade-off (i.e., gain) between the spectral efficiency and the EE. Furthermore, NOMA-K-means-pairing scheme has the potential of striking a more attractive tradeoff between the spectral efficiency and complexity at the receiver.

As shown in Figure 5(b), we can see that the NOMA-K-means-pairing scheme outperforms both SC-NOMA and NOMA conventional pairing in terms of signal performance at the receiver. Given that NOMA-conventional-pairing scheme suffers from mobile user’s discordant power allocation coefficients and gain channels, which cause noise in the signal, TDMA offers the lowest BER (since all of its values are zero, it is not displayed on the graph) at the cost of total rate and spectral efficiency. Whereas SC-NOMA suffers from signal complexity, which causes the received signal to be distorted.



(a) (b)

Figure 4. Simulation results (a) sum rate VS SNR and (b) energy efficiency VS SNR



(a) (b)

Figure 5. Simulation outcomes (a) energy efficiency vs spectral efficiency and (b) average BER vs SNR

4.3. Discussion

Considering mobile user pairing implications, we chose K=2 as the optimal cluster number. This decision has important implications for the clustering performance and PD-NOMA system efficiency. While K=2 is practical for PD-NOMA pairing, a higher K (e.g., 3 or 4) might provide better separation but at the cost of increased system complexity. we achieve a balance between complexity and performance in our PD-NOMA system. The results indicate that mobile users are reasonably well-clustered based on distance, SNR, channel gain and data rate, but additional refinements in feature selection and clustering metrics could further enhance mobile user pairing efficiency.

The proposed NOMA-K-means pairing demonstrates a better trade-off among sum rate, energy efficiency, and BER compared to conventional NOMA pairing. While TDMA provides the lowest BER, it sacrifices spectral efficiency and sum rate. SC-NOMA achieves the highest throughput and energy efficiency but suffers from increased BER and more complex SIC circuit receiver, making it less reliable. Therefore, NOMA-K-means clustering emerges as an effective approach for optimizing user pairing in NOMA systems, enhancing overall network performance while maintaining a balance between throughput, energy efficiency, and reliability.

5. CONCLUSION

In this study, we presented a K-means clustering approach to pair mobile users on SISO downlink-based mmWave communication system. We start by creating a data collection of important mobile user information, such as the distance between mobile users and Micro BS, SNR, data rate, and channel gain, and then we choose K=2 to obtain two clusters of mobile users. Finally, using a fixed power allocation technique, we connect each mobile user from cluster C_1 with a mobile user from cluster C_2. Simulation result show that NOMA-K-means-pairing outperform the other schemes (i.e., NOMA-Conventional-pairing, SC-NOMA and TDMA). Compared to other methods that depend on a single parameter, our approach is more generalized and characterized by simplicity.

Since the multi-input-multi-output (MIMO) system is the foundation of 5G networks, future study could examine it. In this case, we could evaluate how well our model performs. Additionally, knowledge of CSI is crucial for Micro BS and mobile users in PD-NOMA systems; we can apply ML or DL algorithms to determine the channel response. In order to construct a more accurate system in the real world, we may also employ neural networks to anticipate the power allocation and combine them with clustering approaches for mobile users.

FUNDING INFORMATION

This section should describe sources of funding agency that have supported the work. Authors should state how the research described in their article was funded, including grant numbers if applicable. Include the following (or similar) statement if there is no funding involved: authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Litim Abdelkhaliq	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓
Bendimerad Mohammed Yassine		✓		✓			✓	✓		✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors, upon reasonable request.

REFERENCES

[1] P. Taylor, "Number of mobile (cellular) subscriptions worldwide from 1993 to 2023," Statista. Accessed: Jan. 12, 2025. [Online]. Available: https://www.statista.com/statistics/496125/global-mobile-subscriptions-by-region/

[2] B. Bhushan, S. K. Sharma, R. Kumar, and I. Priyadarshini, *5G and Beyond*. in Springer Tracts in Electrical and Electronics Engineering. Singapore: Springer Nature Singapore, 2023. doi: 10.1007/978-981-99-3668-7.

[3] K. Kavitha, K. Thilagavathi, R. Dhivyapraba, K. Jasmine, and D. Sugumar, "Non-orthogonal multiple access (NOMA) signal detection using SVM over rayleigh fading channel," in *ICDCS 2022 - 2022 6th International Conference on Devices, Circuits and Systems*, IEEE, Apr. 2022, pp. 494–497. doi: 10.1109/ICDCS54290.2022.9780685.

[4] D. A. Milovanovic, Z. S. Bojkovic, and T. P. Fowdur, *Driving 5G mobile communications with artificial intelligence towards 6G*. Boca Raton: CRC Press, 2023. doi: 10.1201/9781003205494.

[5] A. Ebrahim, A. Celik, E. Alsusa, and A. M. Eltawil, "NOMA/OMA mode selection and resource allocation for beyond 5G networks," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, IEEE, Aug. 2020, pp. 1–6. doi: 10.1109/PIMRC48278.2020.9217161.

[6] D. Sharma, K. M. Biradar, S. K. Vipparthi, and R. B. Battula, "HYPE: CNN based hybrid precoding framework for 5G and Beyond," in *International Conference on Advanced Information Networking and Applications*, 2022, pp. 43–54. doi: 10.1007/978-3-030-99587-4_5.

[7] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018, doi: 10.1109/COMST.2018.2835558.

[8] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, Nov. 2015, doi: 10.1007/s11276-015-0942-z.

[9] P. M. Ridoy, A. Saha, K. Y. Fariya, P. Saha, K. M. Elme, and F. Arifin, "Millimeter-wave dual-band slotted antenna for 5G applications," in *Lecture Notes in Networks and Systems*, 2022, pp. 19–30. doi: 10.1007/978-3-030-99587-4_3.

[10] M. Yang, B. Li, Z. Bai, and Z. Yan, "SGMA: semi-granted multiple access for non-orthogonal multiple access (NOMA) in 5G networking," *Journal of Network and Computer Applications*, vol. 112, pp. 115–125, Jun. 2018, doi: 10.1016/j.jnca.2018.03.010.

[11] A. F. M. S. Shah, A. N. Qasim, M. A. Karabulut, H. Ilhan, and M. B. Islam, "Survey and performance evaluation of multiple access schemes for next-generation wireless communication systems," *IEEE Access*, vol. 9, pp. 113428–113442, 2021, doi: 10.1109/ACCESS.2021.3104509.

[12] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X. G. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2961–2974, May 2018, doi: 10.1109/TWC.2018.2804953.

[13] M. A. Masud, A. Al Amin, M. S. Islam, V. Mostafa, and M. Wahiduzzaman, "Optimizing channel capacity for B5G with deep learning approaches in MISO-NOMA-HBF and BFNN," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 36, no. 1, pp. 205–213, Oct. 2024, doi: 10.11591/ijeecs.v36.i1.pp205-213.

[14] N. Ghafouri, N. Movahhedinia, and M. R. Khayyambashi, "A novel user grouping algorithm for downlink NOMA," *Wireless Personal Communications*, vol. 126, no. 2, pp. 1895–1914, Sep. 2022, doi: 10.1007/s11277-022-09827-3.

[15] C. Thokala, K. S. Krishnan, H. Erroju, S. K. Minipuri, and Y. K. Gouti, "Power allocation in NOMA using sum rate-based dwarf mongoose optimization," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 35, no. 2, pp. 683–692, Aug. 2024, doi: 10.11591/ijeecs.v35.i2.pp683-692.

[16] M. S. Al-Abiad, M. Z. Hassan, and M. J. Hossain, "Energy-efficient resource allocation for federated learning in NOMA-enabled and relay-assisted internet of things networks," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 24736–24753, Dec. 2022, doi: 10.1109/JIOT.2022.3194546.

[17] S. Mounchili and S. Hamouda, "Efficient pairing distance for better radio capacity in NOMA systems," in *Proceedings of the International Conference on Advanced Systems and Emergent Technologies, IC_ASET 2020*, IEEE, Dec. 2020, pp. 383–388. doi: 10.1109/IC_ASET49463.2020.9318294.

[18] A. Kumar and K. Kumar, "Multiple access schemes for cognitive radio networks: a survey," *Physical Communication*, vol. 38, p. 100953, Feb. 2020, doi: 10.1016/j.phycom.2019.100953.

[19] T. Gian, V. D. Ngo, T. H. Nguyen, T. T. Nguyen, and T. V. Luong, "Deep neural network-based detector for single-carrier index modulation NOMA," in *Proceedings of 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2022*, IEEE, 2022, pp. 1805–1809. doi: 10.23919/APSIPAASC55919.2022.9980150.

[20] T. Manglayev, R. C. Kizilirmak, Y. H. Kho, N. A. W. A. Hamid, and Y. Tian, "AI based power allocation for NOMA," *Wireless Personal Communications*, vol. 124, no. 4, pp. 3253–3261, Jun. 2022, doi: 10.1007/s11277-022-09511-6.

[21] F. Jiang, Z. Gu, C. Sun, and R. Ma, "Dynamic user pairing and power allocation for NOMA with deep reinforcement learning," in *IEEE Wireless Communications and Networking Conference, WCNC*, IEEE, Mar. 2021, pp. 1–6. doi: 10.1109/WCNC49053.2021.9417564.

[22] A. Ng, "Unsupervised learning, recommenders, reinforcement learning," Coursera. Accessed: Jan. 12, 2025. [Online]. Available: https://www.coursera.org/learn/unsupervised-learning-recommenders-reinforcement-learning

[23] N. W. M. Thet and M. K. Ozdemir, "Partial-beam non-orthogonal multiple access (PB-NOMA) with fuzzy clustering," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1202–1206, Jun. 2021, doi: 10.1109/LWC.2021.3061779.

[24] A. S. Parihar, S. Bisen, V. Bhatia, and O. Krejcar, "Machine learning for non-orthogonal multiple access," in *Intelligent Communication Networks: Research and Applications*, Boca Raton: CRC Press, 2024, pp. 135–153. doi: 10.1201/9781003303114-6.

[25] O. M. Ali, M. A. Damein, M. Elshimy, M. Y. Selim, and A. Nasser, "Reinforcement learning based technique for NOMA user pairing enhancement in RIS assisted hetNets," in *2022 IEEE Global Conference on Artificial Intelligence and Internet of Things, GCAIoT 2022*, IEEE, Dec. 2022, pp. 48–53. doi: 10.1109/GCAIoT57150.2022.10019128.

[26] S. Wang, T. Lv, and X. Zhang, "Multi-agent reinforcement learning-based user pairing in multi-carrier NOMA systems," in *2019 IEEE International Conference on Communications Workshops, ICC Workshops 2019 - Proceedings*, IEEE, May 2019, pp. 1–6. doi: 10.1109/ICCW.2019.8757016.

[27] H. Pu, X. Liu, R. Zhang, Z. Sun, and D. Xu, "Secure resource allocation for NOMA PLC networks with multi-agent DRL," in *International Conference on Communication Technology Proceedings, ICCT*, IEEE, Oct. 2021, pp. 959–964. doi: 10.1109/ICCT52962.2021.9657961.

[28] S. Chinnadurai, P. Selvaprabhu, and M. H. Lee, "A novel joint user pairing and dynamic power allocation scheme in MIMO-NOMA system," in *International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, ICTC 2017*, IEEE, Oct. 2017, pp. 951–953. doi: 10.1109/ICTC.2017.8190822.

[29] Y. Lin, K. Wang, and Z. Ding, "Unsupervised machine learning-based user clustering in THz-NOMA Systems," *IEEE Wireless Communications Letters*, vol. 12, no. 7, pp. 1130–1134, Jul. 2023, doi: 10.1109/LWC.2023.3262788.

[30] R. O. Omslandseter, L. Jiao, Y. Liu, and B. J. Oommen, "User grouping and power allocation in NOMA systems: a novel semi-supervised reinforcement learning-based solution," *Pattern Analysis and Applications*, vol. 26, no. 1, pp. 1–17, Feb. 2023, doi: 10.1007/s10044-022-01091-2.

[31] E. Alozie *et al.*, "A review on rain signal attenuation modeling, analysis and validation techniques: advances, challenges and future direction," *Sustainability*, vol. 14, no. 18, p. 11744, Sep. 2022, doi: 10.3390/su141811744.

[32] C. Hennig, M. Meila, F. Murtagh, and R. Rocci, *Brief history of cluster analysis*. Chapman and Hall/CRC, 2015. doi: 10.1201/b19706.

BIOGRAPHIES OF AUTHORS

Litim Abdelkhaliq completed his Master communication system from Department of Science and Technology at Tahri Mohammed University, Bechar, Algeria, in 2020. And registered for the Ph.D. in Tahri Mohammed University in the year 2022. Currently, he is working as instrumentation senior engineer at C.C (a state construction company). His field of research is on artificial intelligence and the multiple access used in 5G communication systems (NOMA). He can be contacted at email: litim.abdelkhaliq@univ-bechar.dz.

Bendimerad Mohammed Yassine received his B.Sc. degree in electrical and electronic engineering, as well as his M.Sc. and Ph.D. degrees in telecommunication and wireless communication technologies from the University of Tlemcen, Algeria, in 2010, 2012, and 2016, respectively. In 2016, he joined the University of Bechar as an Assistant Professor in the Department of Electrical and Electronic Engineering. Currently, he serves as an Associate Professor in the Telecommunications Department and is a research member of the Digital Communication Team at the LTT Laboratory of Telecommunications at the University of Tlemcen. His research interests include wireless communication and mobile networks, energy efficiency in wireless systems, and artificial intelligence for optimizing communication protocols. He has authored several papers published in international conferences and journals, including those indexed by IEEE. Additionally, he has participated in various national and international research projects, such as the PRFU, PHC-TASSILI and ACADEMY Projects, and collaborates as a reviewer for international journals included in the JCR. He can be contacted at email: yassine.bendimerad@univ-tlemcen.dz.