

## ORIGINAL RESEARCH

# Machine learning approach of multi-RAT selection for travelling users in 5G NSA networks

Nurudeen O. Salau  | Sanaullah Manzoor | Muhammad Z. Shakir

School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley, Scotland, UK

**Correspondence**

Nurudeen O. Salau  
Email: [Nurudeen.Salau@uws.ac.uk](mailto:Nurudeen.Salau@uws.ac.uk)

**Funding information**

Qatar National Research Fund., Grant/Award Number: NPRP13S-0130- 200200

**Abstract**

The rapid increment of mobile device usage and the corresponding huge data volume generated afterwards, necessitated the utilisation of the 5G network spectrum. This is deployed today in terrestrial communication in a non-stand-alone (NSA) architectural mode; where 5G networks are supported by 4G LTE networks. Hence, the current 5G implementation with the gargantuan number of mobile subscribers, poses challenges to the choice of network Radio Access Technology (RAT) selection between 4G and 5G networks, among available multiple base-stations to mobile (travelling) users, with respect to their location, bandwidth requirement, and mobility style. Hence, to address the scenario presented above, the authors record live signal measurements of 4G and 5G networks by a travelling user, that transversed multiple 5G NSA base stations. RAT selection implementations were carried out with support vector machine (SVM), deep neural network (DNN), and eXtreme Gradient Boosting (XGBoost) algorithms to select an appropriate RAT between 4G and 5G RATs, for effective resource allocation for travelling users' requirements. Evaluation of results with standard classification metrics shows XGBoost with overall outstanding accuracy performance at 99.64%.

**KEYWORDS**

4G mobile communication, 5G mobile communication, data analysis, resource allocation

## 1 | INTRODUCTION

The current generation of mobile communication; 5G networks, which is currently deployed as non-stand-alone (NSA) architecture in multiple radio access technology (Multi-RAT) fashion, introduced a paradigm shift towards a user-centric technology framework. This requires an efficient and effective RAT selection scheme, which supports three essential use cases: enhanced mobile broadband (eMBB), massive machine type communication (mMTC), and ultra-reliable low latency communications (URLLC), with stipulated ultra-high data rate, ultra-low latency, and abundant bandwidth. This scheme needs to be responsive to users' demands, mobility, network requirements, and backhaul capabilities, which are imperative due to heavy traffic across the networks. Radio access technology (RAT) is the underlying physical connection method for a radio communication network. It simply indicates the

kind of cellular network access technology deployed by a mobile network operator (MNO) for example, 3G or 4G LTE or 5G New Radio (5G NR), each representing a single RAT and a combination referred to as Multi-RAT. However, RAT selection fundamentally requires the following factors: traffic pattern information of a mobile device, network coverage level information, and cell loading consideration of each RAT [1], which informed our choice of parameters; the geographical location of a travelling user recorded as latitude and longitude, data rate experienced by the user, received signal strength indicator (RSSI) and bandwidth of each RAT, in multiple 5G NSA base stations scenario as depicted in Figure 1. In general, RAT technologies can be characterised based on deployment, coverage, performance, and security [2]. Hence, from the literature, RAT selection that supports multiple-connectivity in 5G networks can be grouped into four main types: (i) Cellular: which consists of 4G LTE and 5G NR, (ii)

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *IET Networks* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

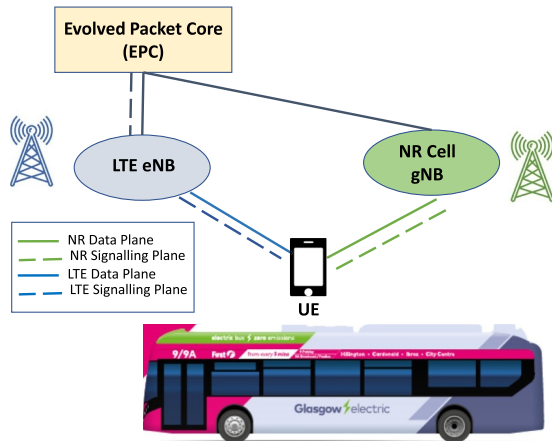


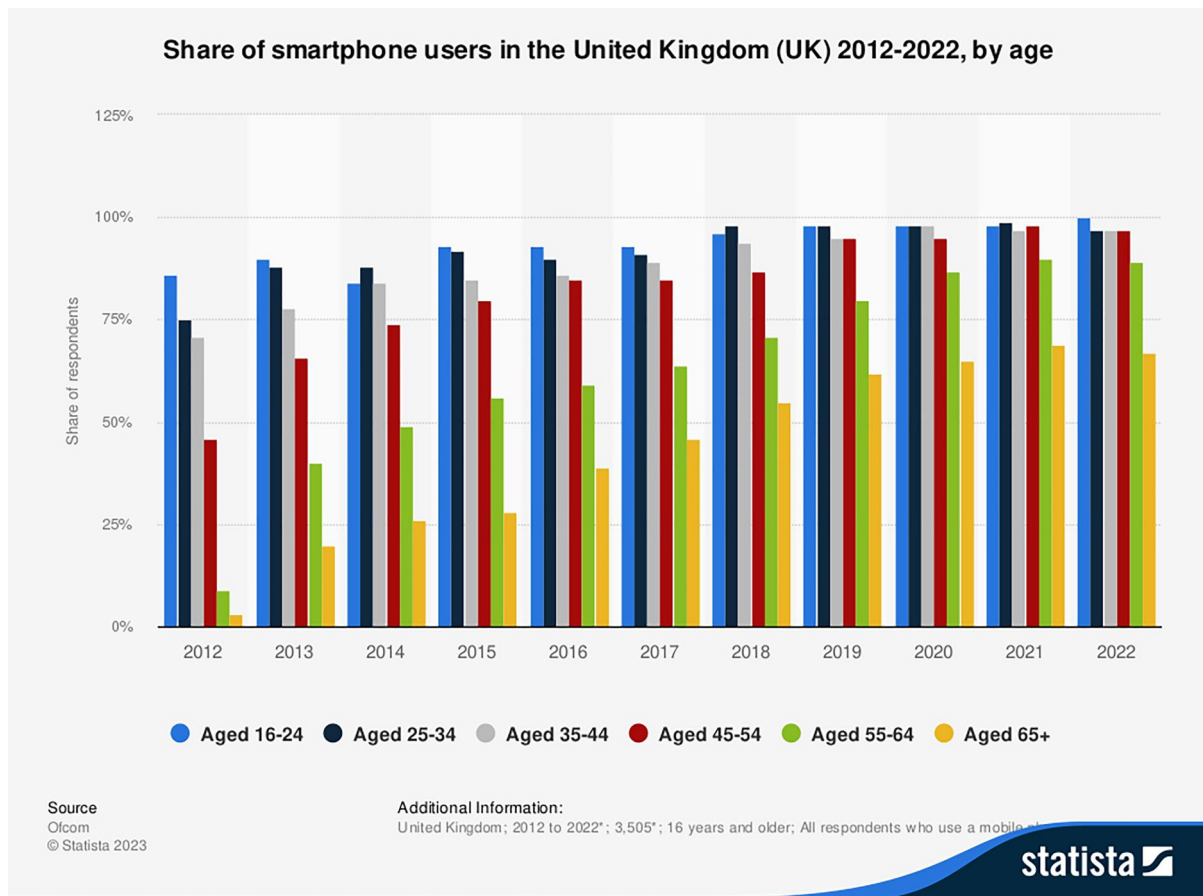
FIGURE 1 Research (Multi-RAT) topology.

Wireless Fidelity (WiFi): which includes IEEE 802.11ac and IEEE 802.11p, (iii) Low-Power Wide-Area (LPWA): which comprises of Narrowband Internet of things (NB-IoT) and Long Range (LoRA), and (iv) Satellite communication at Low Earth Orbit (LEO). RAT selection can also be viewed as: (a) User-centric or (b) Network-centric RAT selection. Hence, the peculiarity of our study, that combines both network and user parameters in cellular 5G and 4G LTE networks. Meanwhile, this study adopts the machine learning (ML) technique due to its ability to manage large data efficiently, and its capacity to minimise computational complexities (e.g. in voice or action recognition [3]) and prediction time. Hence, as part of our motivation, we specifically address challenges faced by mobile users travelling in public buses (since the use of a mobile phone while driving a car is prohibited), which has not been researched before in previous literature. These groups of users account for a high percentage of mobile subscriptions in the UK [4]; the country of our chosen investigated areas. Figure 2 shows smartphone users' age groups in UK between 2012 and 2022. Therefore, the purpose of this study is to develop an intelligent predictive model using a combination of parameters: Geo-location (latitude and longitude), data rate experience of the mobile user travelling in a public bus, network coverage capacity in terms of RSSI and available bandwidth, that will select an appropriate RAT (4G/5G) among multiple 5G NSA base stations for efficient management of large number of subscribers in 5G and future networks to achieve effective resource allocation.

## 1.1 | Related work

In ref. [5], a 5G multi-layered radio access network (RAN) selection scheme, that considered the direction of movement of user equipment (UE) and the location of the candidate cells had been proposed, the scheme utilised the velocity of UE and type of cell for connection, to reduce the number of handovers. However, the authors considered 5G network RANs alone and also neglected bus-travelling users, rather considered pedestrian and car users only in their simulation. Likewise, the authors in

ref. [6] developed a deep machine learning model based on gated recurrent neural networks (RNN) that uses a sequence of previous beams of serving base station, to predict the most likely base station a mobile user will connect with next. To address reliability, and latency challenges during coordination among multiple base stations in millimetre wave (mmWave) wireless systems. The authors only considered a car mobile user moving at random speeds, within a 160 m street, served by two mmWave base stations in their simulation. Also in ref. [7], a convolutional neural network (CNN)-based approach, to predict the mobile traffic of base stations had been proposed. The authors leveraged on collected 4G mobile data in highway scenarios only, using the data rate and longitude of seven base stations. These are in contrast with our live collected data from the intra-city public transport system, and utilisation of both 4G and 5G RATs in 5G NSA scenario, with consideration to latitude, longitude, data rate experience of the user, as well as RSSI and bandwidth of 47 base stations as parameters. Meanwhile, CNN requires a lot of training data and usually fails to encode the position and orientation of objects. Similarly, in ref. [8], the authors proposed a Multi-RAT mobility management (MM) algorithm, where network topology, radio frequency (RF) conditions, and discount factors were used as parameters, for appropriate RAT selection. The proposed system was modelled as a Markov decision process (MDP), to guide the handover policy between 4G and 5G RATs. However, the authors did not address radio resource management (RRM) issues, and they utilised the MDP approach type of reinforcement learning. Likewise in ref. [9], a pricing-based network selection scheme to address traffic allocation issues in dense Multi-RAT had been proposed. Where dynamic decisions of network controller-defined policy, pricing, and congestion control the decisions of UEs in directing their traffic through available RATs. Although the authors considered WiFi, 3G, 4G, and 5G-NR networks in their simulation, they utilised a proportionally fair bandwidth allocation process, instead of artificial intelligence (AI) or ML algorithms for RAT selection decisions. Furthermore, in ref. [10] a software defined network (SDN)/ML-based scheme referred to as; K-nearest neighbour (A2T-KNN) algorithm had been proposed, to address handover between conventional high-powered macro and low-powered small base stations, in millimetre wave (mmWave)-based heterogeneous networks (HetNets) in London. HetNet features and UE movement information were used as parameters for a KNN model, that was trained based on generated vehicle and base station information. Although the authors' simulation was based on a two-tier setup, they restricted their study to mmWave only and utilised KNN; one of the weakest predictive ML algorithm. Lately in ref. [11], an online RAT selection algorithm in a 5G Multi-RAT network, using Constrained Markov Decision Process (CMDP)—a reinforcement-based algorithm was proposed to address sub-optimal utilisation of network resources. Several networks, channel conditions, and priority of users were used as criteria for RAT selection between 5G-NR and Wireless Fidelity (WiFi) networks; an offloading process, in an NS-3 simulated approach, unlike inter-generational handover between 4G and 5G in a live scenario RAT selection. Also,



**FIGURE 2** UK Smartphone users by Age in 2012–2022 [4].

recently in ref. [12], a 5G Multi-RAT URLLC and eMBB dynamic task offloading with multiaccess edge-computing (MEC) resource allocation using distributed deep reinforcement learning (DRL) had been proposed, to address Quality-of-Service (QoS) requirement. The authors proposed UE to make optimal offloading decisions while the MEC server dynamically adjusts the server resources based on offloading requests from multiple UEs using DRL technology, to minimise the energy consumption of the UEs, while maximising the system utility (SU) performance. However, DRL has high computational cost because its models often involve a large number of parameters and require massive amounts of data to train effectively. Similarly in ref. [13], a 5G non-standalone (NSA) mode, a traffic steering mechanism based on deep Q-learning, was adopted to maintain a seamless user experience by choosing appropriate RAT (5G or LTE) dynamically. The proposed method was compared with a heuristic-based algorithm and Q-learning-based traffic steering, for load-balancing purposes. The authors reported success in an optimal way as whenever the high load is induced to a particular RAT, traffic is steered to another RAT dynamically. However, in deep Q-Learning, the Q-function is typically non-linear and can have many local minima. This can make it difficult for the neural network to converge to the correct Q-function. Furthermore, leveraging on the Multi-RAT facilities, in ref. [14], a regional

Multi-RAT dual connectivity management for reliable 5G Vehicle-to-Anything (V2X) communications was carried out, based on release 16 of the Third Generation Partnership Project (3GPP) standards for V2X specifications. The authors presented Multi-RAT dual connectivity (MR-DC) aware, regional Multi-RAT management (MRM) entity, for reliability improvement of 5G V2X through the use of redundant transmissions. Moreover, in ref. [15], Unmanned Aerial Vehicles (UAVs) were researched in different areas and diverse use cases for UAV parcel transport in an emergency service. The facilities of 4G and 5G networks technology in Multi-RAT architecture were used, as an alternative solution for scenarios, where the 5G network is unstable or not yet fully deployed. Authors conclusion showed that the 5G network can provide the necessary QoS for UAV operations, even at low received signal power in the order of  $-90$ , unlike 4G network which although has a reliable link but can not provide the required latency as 5G. Consequentially, in ref. [16], a comparative analysis of four superior mobility predictors: Deep Neural Network (DNN), Extreme Gradient Boosting Trees (XGBoost), Semi-Markov and Support Vector Machine (SVM) had been done. Synthetic dataset of 84 mobile users, generated through Self-Similar Least Action Walk (SLAW) mobility model, was used to predict the future location of mobile users. The authors utilised an LTE simulator to generate the network topology of seven macro cells

and rated XGBoost highest in performance. Hence, in furtherance to our previous study on RAT selection by a pedestrian, that utilised a single base station [17], where XGBoost was also rated with utmost performance. We therefore chose XGBoost as a benchmark model, then carried out RAT implementation of this study, with classification algorithms: deep neural network (DNN)-(Multi-Layer Perceptron (MLP) type), and Support Vector Machine (SVM), using data rate along with Geo-location of UE (longitude and latitude), RAT bandwidth, and RSSI, which were uniquely measured from the live network of 5G NSA base-stations, unlike simulated ones in literature as basic parameters. To the best of our knowledge, these machine learning algorithms have not been used for RAT selection in multiple RAN or base station scenarios before. Our observation revealed that XGBoost showed the highest training accuracy, and model performance accuracy and also displayed its ability to predict future (unknown) data-set among large data-set without over-fitting problems.

## 1.2 | Our contributions

In this study, live signal measurements of multiple 5G NSA network base stations from a travelling user were utilised; our research topology is as depicted in Figure 1, where the 5G NSA architecture allows the 5G NR to utilise 5G for radio to UE communication (down-link) while relying on 4G LTE for the UE to radio-head (up-link) communication, before connecting to the Evolved Packet Core (EPC); 5G NSA option 3 [19] Hence, our main contributions are as follows:

- We implemented RAT selection using live signal measurements of 5G NR, and underlying 4G LTE networks, from multiple base stations in contrast to simulated ones in previous studies.
- We used Geo-location (longitude, latitude), the data rate of UE, in addition to RSSI and bandwidth of base-stations as basic parameters, different from previously used ones in the literature.
- We implemented RAT selection with XGBoost as a benchmark to other classification machine learning algorithms that got their classification capability acknowledged by the computer vision community [16]. The algorithm of implementation is presented in Section 5.
- We also performed a comparative analysis of these three highly rated predictors in machine learning; support vector machine (SVM), deep neural networks (DNN), and extreme gradient boosting (XGBoost), in RAT selection application.
- We computed training accuracy and further calculated the execution time of each algorithm considered, to obtain their respective speed of computation.
- We evaluated each model performance using the following standard classification evaluation metrics: accuracy, precision, recall, and F-score.
- In addition, the cross-validation score function was used for validation purposes, to determine model proficiency on future datasets as presented in Section 6.4.

The rest of the paper is organised as follows: Section 2 presents our methodology, Section 3 presents system model and problem formulation, Section 4 elucidates data measurement procedure, while Section 5 presents the Multi-RAT implementation. Performance evaluation and validation are presented in Section 6 and paper conclusion in Section 7.

## 2 | OUR METHODOLOGY

Live 5 G/4G radio measurement data were carefully recorded, across a total of 47 (multiple) 5G NSA base stations, between Paisley (Sub-urban) and Glasgow (Urban) terrains, over a distance of 12.6 miles (20.28 km). Table 1 shows the key data collation information. Areas of investigation were chosen due to the availability of 5G and 4G footprints of the selected UK mobile service providers in these regions. Samsung Galaxy A52s 5G enabled phone equipped with Cellular-z software application was used by a travelling user in a public bus, moving at an average speed of 35 miles/hr (56.33 km/h) [20]. Through the cellular-application software, network information, such as serving cell, signal strength, neighbouring cells information, geographical location (latitude and longitude), cell type, frequency band of operation and other parameters can be recorded. Hence, field UE measurements were taken at 50 m apart, over a total distance of 12.6 miles (20.28 km), between High Street in Paisley as the starting point, and Buchanan bus station in Glasgow as the destination. Data were methodologically collected using the walk tests technique [21], between September 2022 and February 2023, in areas of investigation as shown in Figure 3. The collected signal measurements in the form of log files from the application were then exported into Microsoft Excel for collation, saved in CSV format, and analysed with a machine learning tool; anaconda-Python programming language, whose models were used to train the data from Excel. Geo-location (latitude, longitude), data rate, RSSI, and bandwidth were set as objective functions (input parameters). The ML models were trained with the input parameters after good performance training stops, otherwise model is updated with new sets of parameters (iteration). The trained model with experience is now tested with fresh data to select either 4G LTE or 5G RAT.

**TABLE 1** Data collation information.

Item	Details
Duration	Sept. 2022 - Feb 2023
Distance covered	12.6 miles (20.28 km)
Average speed of bus	35 miles/hr (56.33 km/h)
No. of User(s)	1
No. of 4G RATs	47
No. of 5G RATs	47
Measuring device	Samsung galaxy A52s
ML platform	Anaconda-python
ML algorithms	SVM, DNN & XGBoost

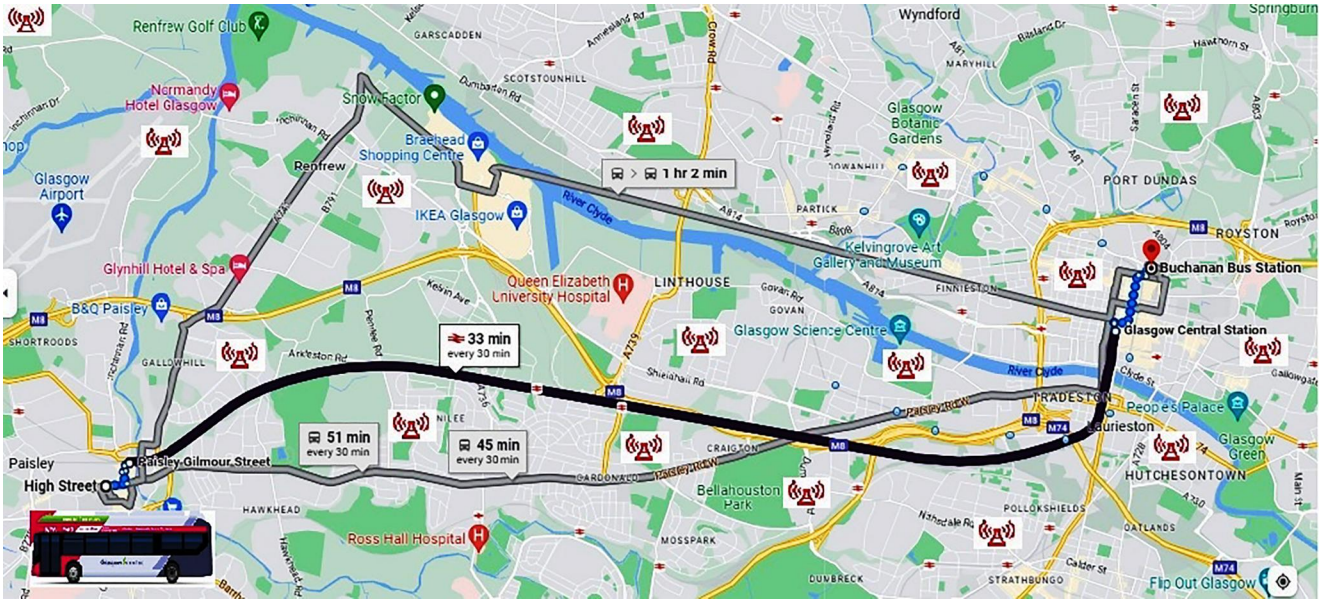


FIGURE 3 Investigated Area [starting point: High street Paisley, endpoint: Buchanan bus station Glasgow, distance: 12.6 miles] [18].

### 3 | SYSTEM MODEL AND PROBLEM FORMULATION

Mobility depicts the movement of mobile users to their location, velocity, and direction over some time. These are captured as the geographical location of the travelling user (latitude and longitude), a public bus travelling at the average speed of 35 miles/hr (56.33 km/h) from Paisley High Street to Buchanan bus station in Glasgow, UK, covering a distance of 12.6 miles (20.28 km) at an estimated time of 33 min per trip as shown in Figure 3. Hence, considering our Multi-RAT system model as depicted in Figure 1, where travelling (mobile) user(s) can be sequentially handed over from one RAT to another, for example, from 4G (RAT-A) to 5G NR (RAT-B) or vice versa, based on network coverage, their location and mobility, which guided our input parameters; geographical location a travelling user recorded as latitude and longitude, data rate at UE, received signal strength indicator (RSSI) from multiple 5G NSA base stations and bandwidth of each network to select an output variable RAT. Let the number of travelling users transverse through the set of base station (BS) be denoted as  $U = U_1, U_2, \dots, U_i$ , while, the set of NSA base stations be denoted as  $BS = B_1, B_2, \dots, B_j$ , where each contains 4G (RAT-A) and 5G NR (RAT-B) respectively. Following our study field observation, a user can only associate with a single BS at a time. Hence, the association matrix between the base stations and travelling user  $U_i$ , in the public bus is denoted as  $A = A_{11}, A_{12}, \dots, A_{ij}$ , where the association variable between travelling user  $U_i$  and base station  $B_j$  is expressed as  $A_{ij}$ , and it can be either 0 or 1. Hence, to perform connection action, where the set of all possible association action ( $A$ ), in choosing either RAT-A or RAT-B in all base stations is represented as follows:  $\{a_0^1, a_1^1\}, \{a_0^2, a_1^2\} \dots \{a_0^j, a_1^j\}$ , where the number of base station =  $1toj$ . Therefore,  $\{a_0^1, a_1^1\}$ , denotes the action of dropping a RAT or choosing the alternative RAT for connection

in base station  $B_1$ . Similarly,  $\{a_0^2, a_1^2\}$ , denotes the action of dropping a RAT or choosing the other RAT for connection in base station  $B_2$ , and so on. Let the network coverage level ( $C$ ) of each RAT be considered as  $\{c_1^j, c_2^j\}$ , where it is assumed that network transmission is reliable. Meanwhile, parameter  $B_w$  also stands for available bandwidth, while  $V_b$ , represents the average speed of a travelling bus in miles per hour. Hence, the connection status ( $S$ ) of the travelling user can be defined by the following vector of features:  $\{c_1^j, c_2^j, V_b, B_w\}$ . Therefore, the vector ( $S$ ) will be utilised to determine the most appropriate RAT connection (association) action:  $A \in a_0^j, a_1^j$ ; which represents the event of dropping a RAT or using the alternative appropriate RAT in a base station. We can observe that the RAT selection problem formulation above is a classification type, we therefore adopted the following machine learning classification algorithms for implementation.

#### 3.1 | Machine learning for Multi-RAT selection

In broad terms, machine learning (ML) is the application of artificial intelligence (AI), that utilises statistical methods, to build a mathematical model, and train the model with sample data known as training data; thereby providing systems the ability to automatically learn from historical data [22], and improve from experience, to make predictions or decisions without being explicitly programmed for the task. Mathematically, this can be viewed as mapping input variables  $x$  to an output variable  $y$ , and can be expressed as follows:

$$y = f(x) \quad (1)$$

Machine learning is categorised as supervised, unsupervised and reinforcement machine learning. Supervised machine learning, can be further divided into classification and regression algorithms, while unsupervised machine learning can also be divided into clustering and association algorithms. The reinforcement machine learning also has Markov Decision Process (MDP) and Q learning as two essential learning models.

Hence, in this study, our field data via the Python ML tool are fed into the following supervised classification algorithms, whose models can be found in ref. [23]:

- Support Vector Machine (SVM): SVM is a widely used ML classification algorithm, that finds the decision boundary between any two classes by creating a separation line (hyperplane), which divides the classes in the best possible manner, for example, 4G or 5G.
- Deep neural network (DNN): Deep learning is a subset of ML, which makes the computation of a multilayer neural network structure that can learn and make intelligent decisions on its own feasible. DNNs basically are of three types: (1) Multi-Layer Perceptron (MLP), (2) Recurrent neural network (RNN) and (3) Convolutional neural network (CNN). MLP, just like every other type is based on an artificial neural network structure, and is a non-parametric estimator that can be used for classification and regression purposes. Since RNN and CNN had already been used in literature, hence MLP is adopted in this study, where MLPClassifier setting used is stated as follows: solver = 'Adam', activation = 'relu', and hidden\_layer\_sizes = (64,64).
- Extreme Gradient Boosting (XGBoost): XGBoost is an ensemble, gradient-boosted decision tree (GBDT) ML algorithm. It is a scalable, end-to-end tree-boosting system, which comprises of a set of classification and regression trees (CART). It reduces the error rate as its trees grow one after another, hence its adoption as our proposed machine learning tool based on outstanding performance. XGBClassifier used is stated as follows: tree\_method = gpu\_hist, enable\_categorical = True, use\_label\_encoder = False. It can be mathematically expressed as follows:

$$\hat{y}_i = \sum_{k=1}^K f_k x_i, f_k \in F \quad (2)$$

where  $y_i$  represents predicted output,  $K$  represents the number of trees, each  $f_k$  corresponds to an independent tree structure, and  $F$  stands for the set of all possible CARTs [24].

## 4 | FIELD DATA MEASUREMENT

### 4.1 | Base station parameters

- Frequency of operation: It was observed during our research, that the UK-MNO utilised bands 1, 3, 7, 20, and

bands n1, n3, n7, n20 for 4G LTE and 5G NR RATs respectively. Hence, these frequency bands according to 3GPP data [25, 26], have the following equivalent frequencies of operations: 2100, 1800, 2600, and 800 MHz respectively, all operating in frequency duplex division (FDD) mode.

- Mode of transmission: Frequency division duplexing (FDD) was utilised, and it is a method of establishing a full-duplex communication link that uses two different radio frequencies for transmitter and receiver operation. That is, the FDD operation usually assigns the transmitter and receiver to different communication channels.
- Bandwidth: It refers to the data transmission capacity of a communication channel in Hertz (Hz), that is, as bandwidth increases, more information per unit of time can pass through the channel. Hence, it connotes the maximum amount of spectrum available for data transmission on a communication channel in a specified time. From our study, since bands 1, 3, 7, and 20 are used in both 4G LTE and 5G NR, their equivalent bandwidths are 60, 75, 70, and 30 MHz respectively as presented in Table 2.
- Data rate: Bandwidth specification normally gives an idea of the capacity of a communication channel, but the data rate defines the actual (present) data transmission speed in the channel. Data rates are usually expressed in bits per second or bytes per second, hence data rate can be viewed as the actual data transmission speed in bits/sec. As one of the benchmarks used in assessing the performance of a wireless communication channel, hence its adoption is one of our parameters. We therefore computed data rates based on the Shannon theorem formula. According to Shannon's theorem, the maximum data transmission rate possible in bits per second is given by the following equation:

$$Data\_rate = B \log_2 \left( 1 + \frac{S}{N} \right) \quad (3)$$

where  $B$  denotes channel bandwidth measured in MHz.

$S/N$  represents signal to noise ratio.

The Shannon Equation (3) is then used to compute data rates experience of the travelling user as he transverses across base stations.

TABLE 2 Base stations parameters.

4G band	5G band	Frequency (MHz)	Bandwidth (MHz)	Duplex mode
1	n1	2100	60	FDD
3	n3	1800	75	FDD
7	n7	2600	70	FDD
20	n20	800	30	FDD

## 4.2 | Radio measurement parameters

The following measured parameters: RSRP, RSRQ, and RSSI [27], associated with 4G LTE; cell reference signal (CRS) were measured during our study:

- Signal-to-noise and Interference ratio: Typical reporting ranges are as tabulated in Table 3. Values recorded were used in Equation (3) to compute the data rate of 4G RATs.
- Reference Signal Received Power (RSRP): The typical reporting range is from  $-140$  to  $-44$  dBm.
- Reference Signal Received Quality (RSRQ): Its reporting range is from  $-19.5$  to  $-3$  dB.
- Received Signal Strength Indicator (RSSI): RSSI is a measure of power and can be mathematically expressed as it relates with RSRP and RSRQ:

$$RSSI = N * \left( \frac{RSRP}{RSRQ} \right) \quad (4)$$

where  $N$  is the number of resource blocks (RBs) per channel bandwidth.

However, the following parameters: SS-SINR, SS-RSRP, and SS-RSRQ associated with 5G networks; synchronisation signal (SS) and channel state information (CSI), as defined by 3GPP [28] were measured as follows:

- Synchronisation Signal Signal-to-noise and Interference Ratio (SS-SINR): Typical reporting ranges are as shown in ref. [29]. Measurements recorded were substituted in Equation (3) to calculate the data rate of 5G RATs
- Synchronisation Signal Reference Signal Received Power (SS-RSRP). Its reporting range is from  $-140$  to  $-44$  dBm
- Secondary Synchronisation Signal Reference Signal Received Quality (SS-RSRQ). Its reporting range is from  $-19.5$  to  $-3$  dB.
- RSSI can also be mathematically expressed as follows:

$$RSSI = N * \left( \frac{SS - RSRP}{SS - RSRQ} \right) \quad (5)$$

where  $N$  is the number of resource blocks in the  $RSSI$  measurement bandwidth.

TABLE 3 Signal strength indicator.

SINR	Signal strength
$\geq 20$ dB	Excellent
13–20 dB	Good
0–13 dB	Fair to poor
$\leq 0$ dB	No signal

## 4.3 | Geo-location parameters

- Latitude (lat.): This is the first number listed in geographical coordinates and its between  $-90$  and  $90$ . Usually expressed in decimals. For instance latitude is  $54.961628$  in  $(54.961628, -3.171315)$ .
- Longitude (long.): This is the second part of the geographical coordinates and it is between  $-180$  and  $180$ . It is also expressed in decimals. For example, longitude is  $-3.171315$  in  $(54.961628, -3.171315)$ .

### Algorithm 1 Machine learning RAT selection for travelling user

---

**Require:** data  
**Ensure:** RAT

```

begin
  start timer
  Load data (Busdata.csv) in CSV file format
  Define input (X) and output (y) training and
  test data-sets
  Call instance of a model: (DNN or XGBoost
  or SVM)
  Split data: X - train, X - test, y - train,
  y - test
  Xk1mno = <lat, long, data_rate, RSSI, Bw>
  Yi = <4G or 5G>
  Train the model: X-train, y-train
  Define evaluation metrics
  Total number of BS ← j
  BS ← 1
  while BS ≤ j do
    if lat = Xk, long = Xl, data_rate = Xm,
    RSSI = Xn and BW = Xo then
      Yi ← 4G
    else if lat = X'k, long = X'l, data_rate
    = X'm, RSSI = X'n and BW = X'o then
      Yi ← 5G
    end if
    Test the model: X - test
    Predict (select) RAT: y - test
    BS ← BS + 1
  end while
  Score model
  stop timer
  Print results
End

```

---

## 5 | MULTI-RAT IMPLEMENTATION

The following parameters: longitude, latitude, data rate, RSSI (which is common to 5G NR and 4G LTE in 5G NSA architecture depending on the RAT a user is connected), and bandwidth ( $B_w$ ), were compiled and saved as CSV (comma-

separated values) file. The CSV format permits data to be saved in a table structured format (422 rows by 11 columns), suitable for machine learning tool analysis.

As shown in Algorithm 1, the ML-based pseudo-code can be expressed in nested if-else if the sequence in a while loop, to compute parameters for all base stations, in order to obtain either 4G or 5G as an output (Ensure) is defined as RAT, after checking input (Require) data parameters conditions. The algorithm begins with a start timer, where the execution time of the algorithm is initialised. The CSV file named (Busdata.csv) is then imported (Load) into Jupyter Notebook; a ML environment programming language. Furthermore, the input data (lat, long, data\_rate, RSSI,  $B_w$ ), and output data (RAT) which can either be 4G or 5G are defined as X and y parameters respectively. The input X and output y were further divided into training and testing datasets; X-train, y-train, and X-test, y-test respectively. An instance of a model of Deep Neural Network (DNN), or eXtreme Gradient Boosting (XGBoost), or Support Vector Machine (SVM) is also initiated (called) for prediction (selection) of a suitable RAT. The Geo-location (lat, long), data\_rate, RSSI, and  $B_w$  are used as input data (features);  $X_{klmno}$  to obtain the suitable RAT;  $y_i = 4G$  or 5G (labelled output). Thereafter, the model is then trained with the input and output training data-sets; X-train and y-train so as to learn data patterns and gain experience to be able to predict the y-test, meant to select either 4G or 5G RAT network. In addition, the performance evaluation metrics (accuracy, precision, recall, and F-Score) of classification algorithms and cross-validation scores used to evaluate and validate the model performance respectively are defined accordingly. The total number of base station (BS) is represented as  $j$ , this is followed by an initial value 1 assigned to BS for a while-loop operation, which is accomplished with ( $BS = BS + 1$ ) increment, until the total number of BS is computed. The input parameters are tested in the following if-else if conditions to select suitable RAT;  $y_i$  as 4G or 5G, after the ML model is tested with defined X\_test ab initio. The evaluation and validation metrics are executed at the score model. After this, the timer stops, and results follow (Print results). Output results of training accuracy, execution time, RAT selected, performance evaluation, and validation metrics are displayed. Full dataset available in ref. [29].

## 6 | PERFORMANCE EVALUATION

In this study, we based our performance evaluation on: the training accuracy of algorithms, execution time of each machine learning algorithm, classification metrics evaluation results, and cross-validation scores. During implementation, iteration was carried out 10 times for each algorithm and an average of collated values were calculated to obtain performance evaluation results as highlighted in the following sections.

### 6.1 | Training accuracy of algorithms

During machine learning implementation, 80% of collated data were used as training data-set while the remaining 20% were used as testing data-set for evaluation. However, we observed that SVM, DNN, and XGBoost have the following training accuracy: 0.6568, 0.9907, and 0.9982, respectively. Hence, SVM is rated as having the lowest training accuracy, followed by DNN and XGBoost in ascending order respectively. This is as presented in Table 4 and captured in Figure 4.

### 6.2 | Execution time of algorithms

Similarly, we calculated the execution time which shows the computational speed of each algorithm implored from the equation:

$$execution\_time = stop\_time - start\_time \quad (6)$$

We observed from our study that SVM, DNN, and XGBoost have the following execution time: 0.2803, 1.55, and 0.2767 respectively. Therefore, they are rated accordingly from highest computational speed to the lowest as follows: XGBoost - fastest; showing its dexterity in computing, SVM - fast; displaying its high performance and DNN algorithm rated lowest, due to the number of layers naturally required to achieve good performance. Values collated are as presented in Table 4 and captured in Figure 5.

### 6.3 | Analysis of evaluation results

In general, the performance of the classification algorithms is analysed in terms of accuracy (ACC), precision (P), recall (R), and F-score (F), which are calculated using the following formulas [30]:

TABLE 4 Training accuracy and execution rating.

Algorithm	Training accuracy	Execution time	Exec rating
SVM	0.6468	0.2803	Fast
DNN	0.9907	1.55	Lowest
XGBoost	0.9982	0.2767	Fastest

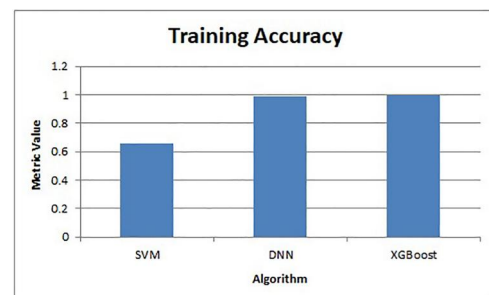


FIGURE 4 Training accuracy.



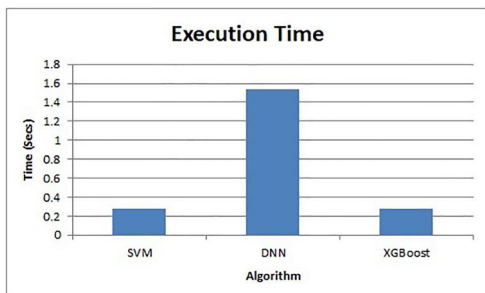


FIGURE 5 Execution time of algorithms.

- Accuracy: It is the ratio of correct predictions that is, true positive (TP) results and true negative (TN) results, to the total number of instances evaluated. It can be calculated as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

where  $FP$  stands for false positive (FP) and  $FN$  represents false negative (FN). Our results revealed the accuracy of algorithms as follows: SVM 0.6893, DNN 0.9929, and XGBoost 0.9964 respectively.

- Precision: It is referred to as the number of true positive (TP) results divided by the number of all positive results, including results not correctly identified. It can be calculated as follows:

$$P = \frac{TP}{TP + FP} \quad (8)$$

During our study, the following precision values of algorithms were obtained: SVM 0.6880, DNN 0.9925, and XGBoost 0.9970.

- Recall: It refers to the number of true positive (TP) results divided by the number of all sample data-sets that should have been recognised as positive. It can be calculated as follows:

$$R = \frac{TP}{TP + FN} \quad (9)$$

Hence, our observed results: SVM 0.8820, DNN 0.9920 and XGBoost 0.9940.

- F-Score: It combines precision and recall as an overall measure of the model's performance. It can be calculated as follows:

$$F = \frac{2}{1/P + 1/R} \quad (10)$$

The following values were recorded during our study: SVM 0.8100, DNN 0.9915, and XGBoost 0.9955. Results as presented in Table 5.

Evaluation metrics Accuracy, Precision, Recall and F-score are presented in Figure 6, 7, 8 and 9, respectively, and further compared pictorially in Figure 10. Observation revealed that

TABLE 5 Evaluation and validation metrics values.

Algorithm	Accuracy	Precision	Recall	F-score	Cross-val
SVM	0.6893	0.6880	0.882	0.8100	0.7293
DNN	0.9929	0.9925	0.9920	0.9915	0.9906
XGBoost	0.9964	0.9970	0.9940	0.9955	0.9857

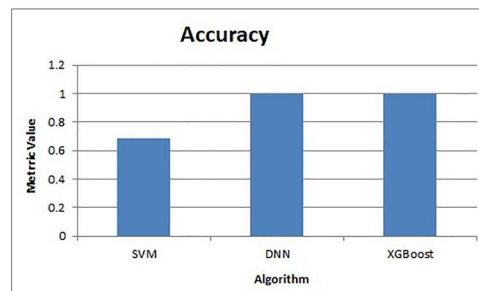


FIGURE 6 Accuracy evaluation metric of algorithms.

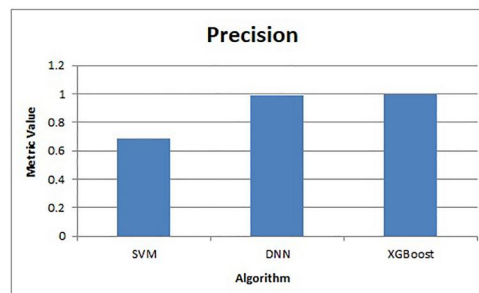


FIGURE 7 Precision evaluation metric of algorithms.

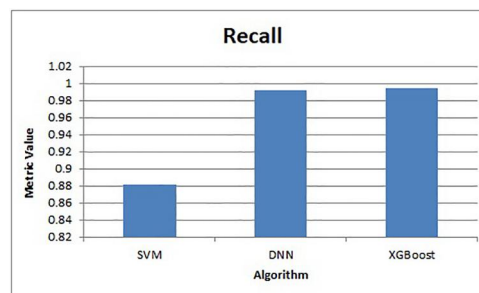


FIGURE 8 Recall evaluation metric of algorithms.

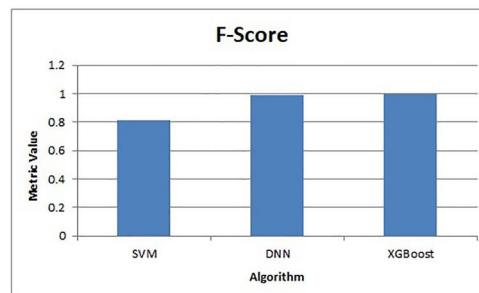


FIGURE 9 F-Score evaluation metric of algorithms.

the XGBoost algorithm had the overall best evaluation performance, with DNN coming at close range in performance to XGBOOST, while SVM took the least among the three algorithms used for implementation, hence we propose and adopted XGBOOST model for RAT selection in multiple base station scenario.

## 6.4 | Validation metric

We further validated models, with cross-validation score metrics, whose model is already described in ref. [23]. The cross-validation score metric is used in machine learning to evaluate the proficiency of any machine learning model on unseen future data. This is achieved by using a limited sample of training data-set to evaluate the model performance in totality when used to make predictions on data-set not used during initial training. The numerical results are shown in Table 5, where we observed that SVM, DNN, and XGBoost have the following cross-validation score values: 0.7293, 0.9906, and 0.9857 respectively. Both DNN and XGBoost displayed remarkable results as shown in Figure 11. DNN has the ability to learn from data and generalise while XGBoost is fast in interpreting when handling large-sized datasets. Meanwhile, the SVM algorithm is not suitable for large and noisy data sets.

## 7 | CONCLUSION

In this paper, we presented the RAT selection algorithm for efficient resource allocation for public bus travelling user(s), utilising multiple 5G NSA base stations. Supervised learning algorithms; support vector machine (SVM), deep neural

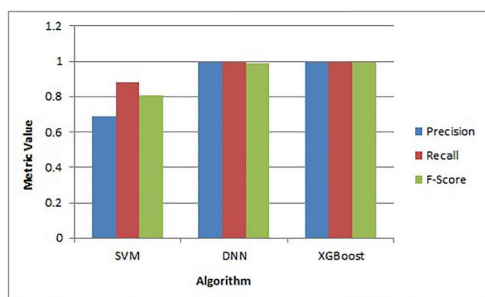


FIGURE 10 Comparison of metrics evaluation of algorithms.

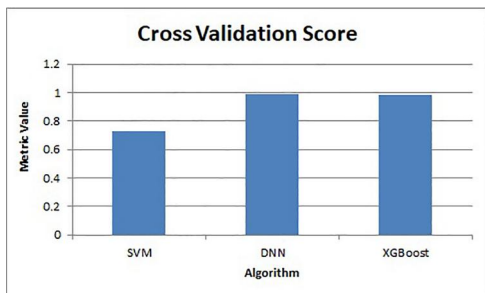


FIGURE 11 Cross validation score of algorithms.

network (DNN), and extreme gradient boosting (XGBoost), were deployed to select an appropriate RAT. The geographical location (longitude and latitude) of a travelling user, data rate experienced by the user from the network, network coverage in terms of received signal strength indicator (RSSI), and available bandwidth on each RAT were used as parameters for RAT selection. Collated live 5G NSA data were divided into training and testing data-sets, the training (input) data-sets were used to train models of the supervised machine learning algorithms deployed; SVM, DNN, and XGBoost. Furthermore, the trained models were tested with input test data-sets to predict or select the appropriate RAT (4 G/5G) as labelled output. Evaluation of results showed a measure of the accuracy of our proposed model; XGBoost at the optimal level of 99.64%, which was further cross-validated at 98.57%, when compared with other algorithms for its effectiveness on future data and mitigation ability. Therefore, recommended for optimisation purposes in multiple base stations environments, for proper resource allocation on the part of phone manufacturers, vendors, and service providers that will impact users' experience positively, in order to assist the continually increasing request of network devices, which will consequently enhance network efficiency. This study is limited to a travelling user in a public bus traversing a defined route. However, future work will be geared towards developing an ML algorithm with consideration of different routes, paths, and RAT selections for mobile users.

## AUTHOR CONTRIBUTIONS

**Nurudeen Oladehinbo Salau:** Conceptualisation; Data curation; Investigation; Methodology; Validation; Writing – original draft. **Sanaullah Manzoor:** Visualisation; Writing – review & editing. **Muhammad Zeeshan Shakir:** Funding acquisition; Project administration; Supervision.

## ACKNOWLEDGEMENTS

This publication was made possible by NPRP13S-0130–200200 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the author[s].

## CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest or competing interests to the work reported in this paper.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ["zenodo"] at <https://doi.org/10.5281/zenodo.8401367>.

## ORCID

**Nurudeen O. Salau**  <https://orcid.org/0009-0002-9053-7903>

## REFERENCES

1. Method and System for Rat Selection in a Multi-Rat Network: Verizon White paper, (2018)

2. Sylla, T., et al.: Multi-connectivity for 5G networks and beyond: a survey. *Sensors* 22(19), 7591 (2022). <https://doi.org/10.3390/s22197591>
3. Shahid, A.R., et al.: View-adaptive graph neural network for action recognition. *IEEE Transact. Cognit. Developm. Syst.* 15(2), 969–978 (2023). <https://doi.org/10.1109/TCDS.2022.3204905>
4. Share of smartphone users in the United Kingdom (UK) 2012–2022, by age [Graph]. Ofcom, January 19, (2023). [Online]. <https://www.statista.com/statistics/300402/smartphone-usage-in-the-uk-by-age/>
5. Kishida, A., et al.: Cell selection scheme for handover reduction based on moving direction and velocity of UEs for 5G multi-layered radio access networks. In 2018 International Conference on Information Networking (ICOIN) (pp. 362–367). (2018) IEEE
6. Alkhateeb, A., Beltagy, I. and Alex, S.: Machine learning for reliable mmwave systems: blockage prediction and proactive handoff. In 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp. 1055–1059). (2018) IEEE
7. Liang, D., et al.: Mobile traffic prediction based on densely connected CNN for cellular networks in highway scenarios. In 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP) (pp. 1–5). (2019) IEEE
8. Mumtaz, T., et al.: Dual connectivity-based mobility management and data split mechanism in 4G/5G cellular networks. *IEEE Access* 8, 86495–86509 (2020). <https://doi.org/10.1109/access.2020.2992805>
9. Passas, V., et al.: Pricing based distributed traffic allocation for 5G heterogeneous networks. *IEEE Trans. Veh. Technol.* 69(10), 12111–12123 (2020). <https://doi.org/10.1109/tvt.2020.3012189>
10. Alablani, I.A., Arafah, M.A.: An SDN/ML-based adaptive cell selection approach for HetNets: a real-world case study in London, UK. *IEEE Access* 9, 166932–166950 (2021). <https://doi.org/10.1109/access.2021.3136129>
11. Roy, A. et al.: Online radio access technology selection algorithms in a 5G multi-RAT network. *IEEE Trans. Mobile Comput.*, 22, (2), 1110–1128, (2023), <https://doi.org/10.1109/TMC.2021.3096968>
12. Yun, J. et al.: 5G multi-RAT URLLC and eMBB dynamic task offloading with MEC resource allocation using distributed deep reinforcement learning. *IEEE Internet Things J.*, 9, (20), 20733–20749, (2022), <https://doi.org/10.1109/JIOT.2022.3177425>
13. Habib, Md.A., et al.: Traffic steering for 5G multi-RAT deployments using deep reinforcement learning. 2023 IEEE 20th Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA pp. 164–169 (2023). <https://doi.org/10.1109/CCNC51644.2023.10060026>
14. Garcia-Roger, D., et al.: Regional multi-RAT dual connectivity management for reliable 5G V2X communications. 2022 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), pp. 320–325. Grenoble, France (2022). <https://doi.org/10.1109/EuCNC/6GSummit54941.2022.9815600>
15. Resende, de, Carvalho, H., Nunes Pinheiro, F.: 4G/5G performance of a multi-RAT UAV for medical parcel delivery 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC), pp. 268–271. Las Vegas, NV, USA (2022). <https://doi.org/10.1109/CCNC49033.2022.9700528>
16. Gebrie, H., Farooq, H. Imran, A.: What machine learning predictor performs best for mobility prediction in cellular networks? In 2019 IEEE International Conference on Communications Workshops (ICC Workshops) (pp. 1–6). IEEE.(2019)
17. Salau, N.O. Shakir, M.Z.: Machine learning analysis of multi-radio access technology selection in 5G NSA network. In 2023 IEEE Wireless Communications and Networking Conference (WCNC) (pp. 1–6). (2023) IEEE
18. <https://www.google.co.uk/maps/@55.966047,-3.1749638,17z?entry=ttu>
19. GSMA: Road to 5G: Introduction and Migration (2018)
20. <https://www.google.co.uk/maps/dir/High+Street,+Paisley/Buchanan+Bus+Station>
21. Joseph, I., Rotimi, K.: On the distributional-based modeling and assessment of radio system networks reliability: two or three parameter weibull distribution model. *BIU J. Basic Appl. Sci.* 6(1), 42–53 (2021)
22. Amit, K.T., et al.: Machine learning, blockchain technologies and big data analytics for IoT's: methods, technologies and applications. IET (2022). (PDF) Book. <https://doi.org/10.1049/PBSE016E.978-1-83953-340-2>
23. <https://scikit-learn.org/>
24. Developers, X.: aXgboost documentatiON (2016)
25. <https://www.techplayon.com/lte-eutra-bands/>
26. <https://www.techplayon.com/5g-nr-frequency-bands/>
27. 3GPP, “TS 36.214. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer; Measurements”, version 12.1.0 Release 12
28. 3GPP: 5G NR; Physical Layer Measurements. 3GPP TS 38, 215
29. Salau, N.O., Manzoor, S., Shakir, M.Z.: Machine Learning Approach of Multi-RAT Selection for Travelling Users in 5G NSA Networks (2023). <https://doi.org/10.5281/zenodo.8401367>
30. Jiaqi, L., Zhifeng, Z. Rongpeng, L.: Machine learning-based IDS for software-defined 5G network. *IET Netw.*, 7, (2), 53–60, (2018) <https://doi.org/10.1049/iet-net.2017.0212>

**How to cite this article:** Salau, N.O., Manzoor, S., Shakir, M.Z.: Machine learning approach of multi-RAT selection for travelling users in 5G NSA networks. *IET Netw.* 1–11 (2024). <https://doi.org/10.1049/ntw2.12124>