IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 December 2022, accepted 15 January 2023, date of publication 19 January 2023, date of current version 24 January 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3237851*

RESEARCH ARTICLE

# A2T-Boost: An Adaptive Cell Selection Approach for 5G/SDN-Based Vehicular Networks

**IBTIHAL AHMED ALABLANI[1,2] AND MOHAMMED AMER ARAFAH[1]**

[1]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Computer Technology, Technical Digital College, Technical and Vocational Training Corporation, Riyadh 11472, Saudi Arabia

Corresponding author: Mohammed Amer Arafah (arafah@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research at King Saud University through the Research Group under Grant RG-1440-122.

**ABSTRACT** Heterogeneous ultra-dense networks (HUDNs) are one of the key enabling technologies for the fifth-generation (5G) networks. They aim to provide high capacity, low installation cost, and distributed traffic loads. The cell selection is a challenging issue in HUDNs, due to the different characteristics of base stations (BSs) and the existence of a large number of them. Thus, the traditional cell selection scheme is not applicable in such a network. In this paper, a novel adaptive cell selection strategy is proposed, called adaptive two-tier based on adaptive boosting (A2T-Boost). It can adapt to the various characteristics of base stations, as well as the different movement features of mobile stations such as vehicles and pedestrian. It is a software-defined networking (SDN)/machine learning (ML)-based scheme. A real-world case is considered in the downtown of Los Angeles city. Simulation results demonstrate that A2T-Boost achieves high prediction performance and it outperforms other related schemes in terms of average number of handovers (HOs) by up to 50%. Moreover, it enhances the average achievable downlink sum-rates and network energy efficiency achieved by vehicles by up to 33.76%. Furthermore, the average packet delay is decreased using the proposed scheme by up to 12.87%.

**INDEX TERMS** 5G, small cells, SDN, Los Angeles, machine learning, HUDNs, adaptive selection.

## I. INTRODUCTION

5G wireless cellular network is the current generation technology that developed by the 3rd Generation Mobile Partnership Project (3GPP) community. 5G networks have developed to meet the increasing needs for higher data rates, lower delays, efficient energy consumption, and reliable connectivity [1]. 5G technology will change our life, work, and the way of communication between with each other. It will support emerging services and new applications such as autonomous vehicle, smart home and factory, and remote surgery [2].

Vehicular networks are emerging technology that provides low-cost and reliable solution for the intelligent transport system (ITS) [3]. Vehicle-to-everything (V2X) technology is an evolution towards the intelligent transportation system. It aims to enhance road safety, the reliability of communications, and traffic efficiency [4], [5]. 5G revolution and beyond will support V2X communication to allow a vehicle to be connected to an entity such as a pedestrian, another vehicle, infrastructure and a network to provide a robust transportation solution [6].

Building a sustainable communication of a vehicular network is a critical requirement for the future Connected and Autonomous Vehicle (CAV). A stable vehicular network infrastructure has an important role in driving safety such as collision warning, slippery road detection, and traffic lights warning signs [7].

Software-defined networking is an emerging solution that is used to handle network management [8]. The major idea of the SDN architecture is the separation of the data and the control planes to perform fast data forwarding and to achieve central control [9]. Network devices have the responsibility of data forwarding, while SDN controllers are responsible for manage network operations [10]. Southbound interface

The associate editor coordinating the review of this manuscript and approving it for publication was Jie Gao.

connects the data and the control planes [11]. In addition, there is an application plane that composes of end user applications which interact with the SDN controllers via the northbound interface to perform specific tasks such as mobility management, routing, access control, security, file transfer, and supervising [12], [13], [14].

Machine learning is a branch of artificial intelligence (AI) that focuses on learning computers how to use data to provide a solution of a problem [15]. It can find a simple solution to a complex problem by analyzing a large amount of data and predicting a solution in a short time and with high accuracy. 5G cellular networks are becoming complex due to new types of services and a large number of connected devices [16]. Machine learning techniques should be used to make 5G network operations more effective [17]. There are three classes of learning techniques, which are supervised, unsupervised, and reinforcement learnings. Supervised learning uses labeled training datasets to identify patterns or behaviors, while unlabeled training datasets are used by unsupervised learning. In reinforcement learning, rewards and punishments are used as signals for correct and wrong actions, respectively [15], [18].

Heterogeneous ultra-dense network is one of the most promising technology for 5G cellular networks. HUDNs refers to networks that combine a very dense deployment of small cells with traditional Long-Term Evolution (LTE) macrocell. In other words, HUDNs are a multi-tier networks that involves very dense low-power small cells and high-power legacy macrocells [19]. Small cells can be installed on utility pole and street light, and light poles [20], [21].

In 5G cellular networks, small cells are used to increase the achievable throughput and minimize the energy consumption [22]. The traditional scheme of cell selection is based on the received signal strength indicators. In 5G HUDNs, this method is inefficient and leads to ping-pong effect [23]. Ping-pong effect means the number of handovers in a specific period exceeds a threshold [24]. The cell selection in HUDNs is a challenging task and it faces many issues as shown in Figure 1. The existence of high-density of 5G small cells in addition to the legacy macrocells increases the complexity of selection process. The macro cells are still required in HUDNs for high-speed mobile stations (MSs), while the small cells are used by low- and medium-speed MSs [25], [26]. 5G HUDNs differ in terms of density, distribution, and sizes of cells. In additions, city roads differ in their types and associated features such as speed limits, lane widths, and traffic volumes. Furthermore, the moving stations have different velocities and directions based on moving behavior and the followed path. Due to these differences, there is a high probability that a non-ideal base station will be selected. A non-ideal base station means a base station that leads to ping-pong effect or a failing in handover [23].

The cell selection decision can be coordinated by using SDN solution [27], and the combination of SDN and ML creates a new network management approach [28], [29].

### A. MAIN CONTRIBUTIONS

The main contributions of this paper are as follows:

1) Proposing an adaptive cell selection scheme called A2T-Boost for mobile stations, such as vehicles, bikes and pedestrian. It has the ability to adapt to different characteristics of HUDNs and mobile station movements. In addition, it aims to maximize the dwell time of a vehicle within a serving cell based on time estimation by relying on four parameters.
2) Modeling the heterogeneous ultra-dense networks based on a real-world dataset that was collected in the city of Los Angeles, where a high-density area of small base stations is selected to be studied.
3) Creating a new network management solution by combining of SDN and ML techniques. AdaBoost model is trained and it achieves prediction accuracy of 99.58%.
4) Modeling vehicle movement behaviour on the LA map based on the principle of reackoning, where the geographical coordinated, direction, and speed of a MS are considered.
5) Achieving superiority over the traditional and recent related works in terms of average number of handovers by up to 50%. In addition, the average achievable downlink sum-rates and network energy efficiency achieved by vehicles are enhanced by up to 33.76%. Furthermore, the average packet delay is decreased using the proposed scheme by up to 12.87%.

### B. PAPER OUTLINE

The rest of this paper is organized as follows: Section II provide a review of related works. The proposed A2T-Boost approach is described in detail in Section III. Section IV explains the methodology of the study in terms of simulation tool, datasets, and system model. The performance evaluation of the proposed approach is discussed in Section V. The paper is concluded in Section VI.

## II. RELATED WORKS

In this section, recent related cell selection (CS) methods are explained. Some of these methods depend on machine learning techniques to predict the optimal wireless cell, while others do not use ML to solve the selection issue. At the end of this section, the limitations of the existing works are given and the recent trends that have been followed by our proposed approach are discussed.

### A. NON ML-BASED CELL SELECTION STRATEGIES

In [30], topology-aware skipping methods are introduced by Arshad et al. to solve the problem of unnecessary handovers. The Poisson point process is used to model a single-tier networks, while the Poisson cluster process is utilized to provide a model for two-tier networks. According to the location of a user and/or the size of a wireless cell, the handover decision is initiated. Simulation results demonstrate that the proposed strategies have superiority over the traditional cell selection scheme in terms of the mean user throughput by up to 47%.

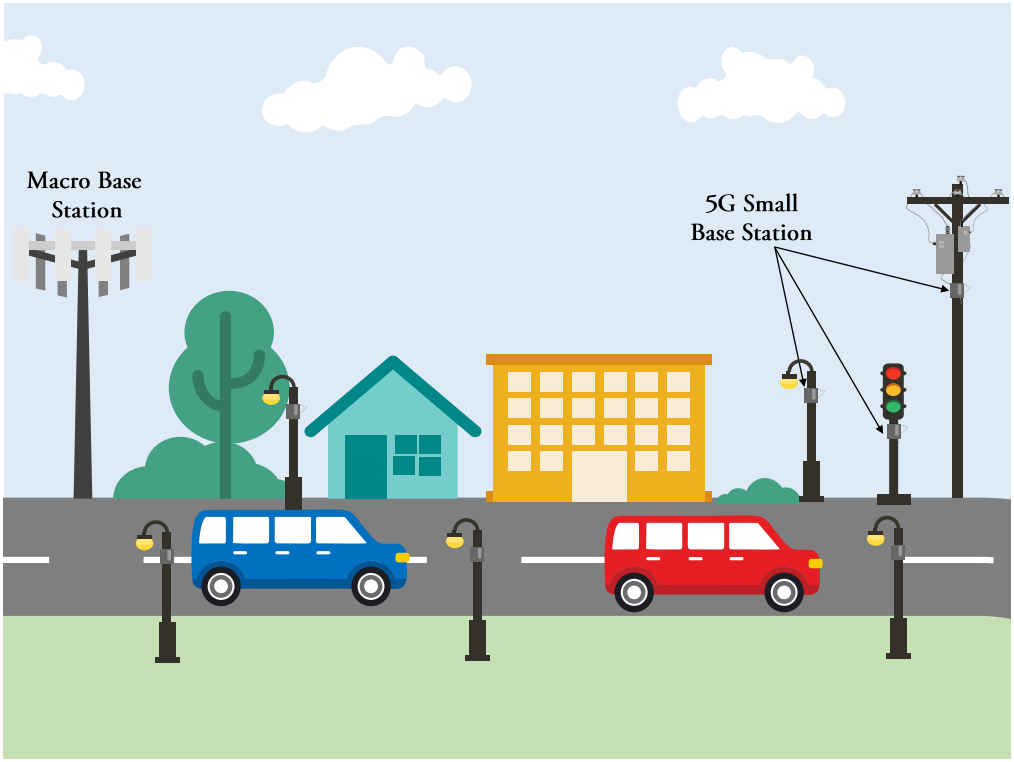



**FIGURE 1.** Cell selection issue in HUDNs.

Tesema et al. proposed a fast cell select scheme in [31], where the serving wirless cell is selected according to the value of the signal-to-interference-plus-noise ratio. The proposed scheme supports multi-connectivity in order to enhance the reliability of the communication in 5G networks. Simulation results show that achievable sum rate is improved and the radio link failures issue is overcome.

In [23], Kapoor et al. proposed four cell selection strategies for urban wireless networks. The first strategy is called minimum distance in the vehicle direction (MD-VD), while the second one is known as minimum load in the vehicle direction (ML-VD). The third and the fourth strategies are next neighbor on the same street (NN-S) and on the opposite street (NN-O). To perform the selection task based on Kapoor's methods, K-nearest base stations are selected, which are known as shortlisted stations. Then, BSs that are existed in the vehicle direction and belong to shortlisted stations are chosen. After that, based on the aim of the proposed strategy, the serving BS is selected. A base station that has the minimum distance and the minimum load in the vehicle direction will be chosen by MD-VD and ML-VD approaches, respectively. Based on NN-S approach, the BS that is located in the range of 0 to $\pi/2$ and has the minimum angle is selected. The base station that is existed in the range of $\pi/2$ to $\pi$ and has the minimum angle will be chosen according to the NN-O approach. Simulation results prove that the MD-VD approach is the worst one in term of the average number of handovers.

Cacciapuoti introduced a cell selection strategy for 5G networks in [32] where a polynomial-time association algorithm is proposed. It considers user mobility and cell loads and thus the proposed strategy mitigates the congestion at base stations. In addition, it avoids the issues of millimeter-wave communications such as directionality, blockage and the effects of non-line-of-sight propagation. The spatial distribution is considered when the 5G small base stations were distributed. The numerical result displays that the proposed strategy has superiority over the conventional cell selection scheme and the handovers rate is reduced.

In [33], a parallel dynamic cell selection method was proposed by Naderializadeh et al. It based on the received signal-to-interference-plus-noise ratio (SINR) and priority of users. A set of transmission points and users are randomly distributed in an area of 200 by 200 meters$^2$. The proposed method passes through four steps to achieve the cell selection task. The fist step is the association between a user and a transmission point based on the SINR value. The second step is users ordering according to their priorities, while the third step is users scheduling which performed when a transmission point works on multiple-users mode. The last step is link scheduling to mitigate the interference effect by deactivating a set of transmission points. The simulation results proved the superiority of the proposed dynamic method in terms of throughput and network coverage compared to some benchmark methods.

A cell selection approach was proposed by Kishida et al. in [34] for 5G multi-layered Radio Access Networks (RANs). It considers the direction and velocity of UE movement to reduce the number of frequent handovers. The final decision is based on the value of SINR, whereby the BS that has the maximum SINR value will be selected. The RAN network is modeled based on actual locations of base stations in Shinjuku, Tokyo. Simulation results proved that the proposed approach achieved an approximate 30% improvement in the number of handovers while maintaining the average flow time.

Elkourdi et al. developed a Bayesian-based user association scheme in [35] for 5G heterogeneous networks. The first player is user equipment (UE) that is classified into delay sensitive and tolerant users. The second player of the Bayesian game is access nodes. A two-tier heterogeneous network is considered with random distribution of low-power node (LPN) in an service area of 400 $\times$ 400 meters$^2$ with density of 20 LPNs/km$^2$. The UEs are distributed based on a homogeneous PPP. Simulation results demonstrates that the proposed scheme has superiority over the conventional and cell-range-expansion (CRE) approaches in terms of end-to-end latency and the appropriate association probability.

A cell selection scheme is developed by Liu et al. in [36] that achieves the selection task based on multiple attributes decision-making (MADM) and fuzzy logic algorithms. The selection of the serving cell is taken based on many factors which are SINR, reference signal receiving power (RSRP), and jitter. The process of generating the proper fuzzy membership function is performed based on subtractive clustering algorithm. The simulation results show that the proposed scheme outperforms the RSSI-based approaches in terms of handover rates by 90% and it minimizes the ratio of ping-pong effects by 10%, while the quality of service level is maintained.

In [37], Sun et al. introduced two cell selection approaches for HUDNs that based on the concept of coordinated multipoint (CoMP). The first approach is known as movement-aware CoMP handover (MACH), while the second one is called improved MACH (iMACH). According to the MACH strategy, a set of base stations are chosen that have the highest signal power and residence time greater than a certain threshold. Then, the handover is initiated when the farthest base station in the set becomes the closest. The iMACH strategy is an improvement of MACH strategy where the nearest base station is added to the selected set of BSs, instead of the station that has the lowest signal power in the set. After that, the handover is triggered when the closet base station becomes the farthest one. Simulation results show that the average achievable sum rate, number of handovers, and coverage probability are improved by both MACH and iMACH approaches.

In [38], a cell selection method, known as called Handover based on Resident Time Prediction (HO RTP), was proposed by Qin et al. It was designed for 5G ultra-dense networks and the main idea of the HO RTP method is the estimation of the residence time within the serving cell. Then, the cell that has the strongest receiving power with a residence time longer than a certain threshold will be selected. Simulation results show that the HO RTP method outperforms the conventional RSSI-based method in terms of average achievable user sum-rates.

Alablani and Arafah developed an adaptive cell selection (ADA-CS) strategy in [26], for 5G heterogeneous ultra-dense networks. The selection of the serving BS is based on user's movement and network characteristics. To perform the cell selection task, six phases should be achieved which are (1) configuration, (2) decision-making, (3) filtering, (4) narrowing, (5) selecting, and (6) handover triggering. Simulation results proved that the ADA-CS method outperforms the traditional and recent related strategies in terms of the mean achievable downlink data rates. In addition, the spectral efficiency and handover rate are enhanced.

Table 1 represents a comparison among recent non ML-based cell selection schemes in terms of cell selection factors and system model.

### B. ML-BASED CELL SELECTION STRATEGIES

Dilranjan et al. introduced a cell selection method for 5G cellular networks in [40]. A Recurrent Neural Network (RNN) is used to predict the optimal BS that a mobile user will be associated with. To train the proposed RNN model, received signal strength (RSS) values are used. The proposed RNN architecture has three layers, which are input, hidden, and output. The RNN model has 640 neurons and the used activation functions are sigmoid and tanh. To evaluate the performance of the proposed approach, Google's Python- based Tensorflow library was utilized. The learning rate is set to 0.0003 and the model training took 35 minutes. An area of 36 km$^2$ is considered that has eight base stations that are distributed randomly. A mobile node, which can be a pedestrian or a vehicle, can connect with three nearest base stations. Simulation results demonstrate that the proposed RNN-based method yielded 98% prediction accuracy of the serving cell.

Perez et al. developed in [41] an ML-based strategy for 5G heterogeneous networks that aims to perform the user association task. The proposed strategy uses the Q-learning algorithm and it based on BS load, the index of the base station, and SINR value to predict the serving BS. Simulation results proved that the proposed strategy has superiority over other existed methods.

In [42], Zhang et al. proposed an ML-assisted cell selection approach in wireless cellular networks for drones. The prediction method of the serving cell is based on a conditional random field model that uses SINR values. Simulation results proved that the proposed ML-based approach achieved prediction accuracy of 90% and it has superiority over two heuristic-based approaches.

**TABLE 1.** A comparison among recent non ML-based cell selection schemes.

| Ref | Year | Authors | Cell selection factors | System model |
|---|---|---|---|---|
| [30] | 2016 | Arshad et al. | User position, cell size | PPP and PCP distributions |
| [31] | 2016 | Tesema et al. | RSRP, noise power | Hexagonal grid |
| [23] | 2017 | Kapoor et al. | Distance, vehicle direction, cell load, azimuth | Deterministic distribution |
| [32] | 2017 | Cacciapuoti | Cell load, distance | Spatial distribution |
| [33] | 2018 | Naderializadeh et al. | SINR, UE priority | Uniform random distribution |
| [34] | 2018 | Kishida et al. | SINR, direction and velocity of UE | Metropolitan case study in Shinjuku, Tokyo |
| [35] | 2018 | Elkourdi et al. | SINR, UE priority | Random distribution |
| [36] | 2019 | Liu et al. | RSRP, SINR, Jitter | Deterministic distribution |
| [39] | 2020 | Zhang et al. | Channel gain, cell load, UEs fairness | Random distribution |
| [37] | 2021 | Sun et al. | Dwell time, distance | PPP distribution |
| [38] | 2021 | Qin et al. | RSSI, residence time | Tyson polygon distribution |
| [26] | 2021 | Alablani and Arafah | RSSI, speed, azimuth, load, and cone angle | Hexagonal grid |

Zappone et al. developed a user association strategy for massive multiple-input and multiple-output (mMIMO) networks in [43] to enhance the user throughput. It based on deep learning where a feed-forward artificial neural network is used to predict the optimal serving cell. The neural network consists of four layers and the input to the neural network is geographical locations of users. Rectified Linear Unit (ReLU) and sigmoid activation functions are used. The adaptive moment (ADAM) is utilized as an optimizer. The numbers of training and testing sets are 140,000 and 15,000 samples, respectively. Numerical results demonstrate that the proposed strategy the computational complexity of the user association process is reduced in comparison to the conventional RSSI-based approach.

In [44], two hidden Markov-model (HMM) based ML techniques suggested by Balapuwaduge et al. were introduced to address a cell selection issue. The key objectives of the suggested HMM-based ML methods were the accessibility and reliability of network resources. The proposed strategies outperformed a random cell selection method in simulations in terms of channel availability and dependability.

An ML-based cell selection method for automobiles in mmWave networks was introduced by Khan et al. in [45]. The vehicle association problem is solved using distributed deep reinforcement learning (DDRL). The Markov decision process is used to formulate the reinforcement learning problem. This framework, known as Asynchronous Advantage Actor Critic (A3C), combines actors and critics. Roadside units (RSUs) transmit their actions to a centralized body that determines the RSUs' compensation. When compared to existing complex methods, the suggested technique has lower control overhead and computational complexity. The suggested DDRL-based approach outperforms existing cell selection schemes in terms of possible sum rate by up to 15%, according to numerical results.

Zhang et al. created an intelligent machine-learning-based user association for 5G heterogeneous networks in [39]. A cross-entropy technique was used to label the ideal base station to be associated with the challenge as a supervised learning assignment. To address the user association problem while adhering to the cell load constraint, a U-Net convolutional neural network (CNN) was trained. The outputs of the ML model were the user association matrices whereas the inputs of CNN were channel gain matrices that were translated onto pictures. Simulation findings showed that the suggested strategies improved computing speed and network robustness.

In [46], Anand et al. proposed an ML-based cell selection approach for a single-tier LTE environment called Machine Learning-Network Selector (ML-NetSel). It aims to improve the quality-of-service for video applications. Two machine learning predictors are trained to perform the network selection task which are Support Vector Machine (SVM) and Random Forest (RF). The input features of the ML models are QoS parameters which include throughput, packet loss ratio (PLR), and delay. Two LTE base stations and up to 50 users are considered in the study. The SVM predictor achieves an accuracy of 92.4, while the RF predictor achieves an accuracy of 97.1%. Simulation results demonstrate that the ML-NetSel

enhances the network performance in terms of throughput, delay and packet loss ratio.

### C. LIMITATIONS OF RECENT CELL SELECTION WORKS

The limitations of recent cell selection works that are presented in this section are:

- The majority of modern works choose the serving base stations in a static, non-adaptive manner. Given that HUDNs have many tiers, adaptive selection, which may be carried out by establishing certain thresholds to transition between the network tiers, is favored. Macro-BSs are preferred for vehicles that have very high speed to preserve performance of the network. The serving BSs for low- and medium-speed vehicles, however, will be small BSs.
- Recent studies prioritize BSs with the highest receiving power to increase the throughput that can be achieved. The nearest BS with the best reception strength will be far away in mobility situations while the user is going forward. Because of the numerous handovers caused by relying on this theory, network performance suffers.
- Some works rely on the estimation of the mobile station's cell dwell duration, which is a crucial element in choosing the serving cell. However, these studies estimate the dwell time by assuming the mobile station is near the cell's edge for simplicity's sake, which is incorrect.
- While there are less ML-based works than non-ML-based works, forecasting serving BSs must be based on ML techniques to minimize computational complexity and thus shorten the cell selection delay. A machine learning model's input characteristics should also be properly chosen in order for the trained model to effectively handle the cell selection issue. To develop an effective ML model, training and testing datasets should also be created or acquired using trusted methods.
- It is better to use a cell selection approach in a real-world setting to examine the efficacy of the suggested technique. Applying these works to real-world scenarios will result in suboptimal network performance because they were designed for specific typologies.

Given the aforementioned restrictions, it is necessary to use cell selection techniques that can be modified to choose the serving cell to be linked with in order to preserve network performance. Additionally, depending on machine learning techniques is a modern trend that we ought to use in order to decrease the computational complexity and prediction time. Additionally, it is preferable to apply a suggested cell selection algorithm in a real-world setting to assess the efficiency and viability of the suggested approach.

## III. THE PROPOSED A2T-BOOST APPROACH

In this section, the proposed adaptive cell selection scheme is explained in terms of problem formulation, the proposed SDN/ML-Based model building process, and the framework of the cell selection approach. At the end of this section, a case study is given to show how the cell selection process is performed based on this proposed method in comparison to other recent related methods.

### A. OPTIMIZATION PROBLEM AND FORMULATION

The macro- and small base stations (referred to as $\mathbb{BS}s_{macro}$ and $\mathbb{BS}s_{small}$), respectively, are included in the set of base stations given by the notation $\mathbb{BS}s = \{B_1, B_2, \ldots, B_I\}$. Total network vehicles are spread throughout a HUDN and are denoted as $V_1, V_2, \ldots, V_J$. A single BS can only have one vehicle linked at once. The association matrix between base stations and vehicles is represented by the notation $\mathbb{A} = \{A_{11}, A_{12}, \ldots, A_{IJ}\}$, where $A_{ij}$ represents the association variable between base station $B_i$ and vehicle $V_j$ and has two possible values: 0 and 1.

$$A_{ij} = \begin{cases} 1 & 1 \; if \; V_j \; associates \; with \; B_i \\ 0 & otherwise \end{cases} \tag{1}$$

The constrained optimization problem is posed for the vehicle association in a HUDN, whose objective function is to maximize the dwell time of a vehicle within a serving cell within the association duration (i.e., $t \in \{t_0, t_2.., t_n\}$). Therefore, the optimization problem is shown in equation 2.

$$max \sum_{i=1}^{I} \sum_{t=t_0}^{t_n} A_{ij}(t) \; Td_{ij}(t), \quad \forall \; V_j \in \mathbb{V} \; B_i \in \mathbb{BS}. \tag{2}$$

### B. VEHICLE ASSOCIATION CONSTRAINTS

The vehicle association process is restricted by a set of constraints which are

1) Number of simultaneous association constraint: A vehicle can connect with only one base station at a time.

$$C1 : \sum_{I} A_{ij} = 1, \quad \forall \; V_j \in \mathbb{V}. \tag{3}$$

2) Maximum BS Load constraint: The maximum load of a BS $B_i$, which is referred to as $L_{max}$, should not be exceeded.

$$C2 : \sum_{J} A_{ij} = L_i, \quad \forall B_i \in \mathbb{BS}; \; 0 \leq L_i \leq L_{max} \tag{4}$$

3) Transmission power constraint: The maximum transmission power ($p_{tx_{max}}$) of vehicle $V_j$ should be considered. Therefore, we have

$$C3 : \sum_{J} A_{ij} p_{tx_{ij}} \leq p_{tx_{max}}, \quad \forall B_i \in \mathbb{BS}. \tag{5}$$

4) Quality of Service constraint: The minimum achievable sum rate ($R_{min}$) for a vehicle $V_j$, must be maintained.

$$C4 : \sum_{I} A_{ij} r_{ij} \geq R_{min}, \quad \forall V_j \in \mathbb{V}. \tag{6}$$

### C. THE PROPOSED SDN/ML-BASED MODEL BUILDING PROCESS

To build the proposed SDN/ML-based model, five main stages have been passed through, as shown in Figure 2 which are:

- **Stage 1: Data Preparation:** The objective of this stage is to gather, create, and prepare data from vehicles as well as from large and small base stations. The data needed to train and test the suggested machine learning model will be ready at the conclusion of this phase.
  1) **Dataset Collecting:** The proper dataset for the macro and small base stations should be gathered in this step. The BSs dataset, which includes data on both macro and minor BSs, may be downloaded from the Internet as a single file. The information on macro and mini BSs may be discovered in other hands as two independent databases. The BS dataset must contain the geographic location data for BSs in the form of latitude and longitude coordinates. In addition, vehicle dataset is collected based on the selected study area. Some applications can be used to accomplish this task such as Google Maps, MediaQ and Ultra GPS Logger. We may find vehicle datasets available for download on the Internet that have been collected by other researcher. The most important things that must be available in the vehicle dataset is the geographical locations, directions and speeds of vehicle samples.
  2) **Data Cleaning:** The data that is not utilized by the suggested cell selection technique to forecast the serving BS is eliminated in the cleaning step.
  3) **Data Labeling:** The adaptive two-tier (A2T) method is used to accomplish the labelling process. Algorithm 1 shows the pseudocode of A2T labeling algorithm which is used to assign a label of a serving base station ID for each sample in the vehicle dataset to train the machine learning model. As shown in Algorithm 1, the proposed cell selection scheme targets the small BS located in the service range that has longest dwell time and with load lower than the load threshold, if a vehicle speed is lower than the predetermined speed threshold. Otherwise, the proposed A2T scheme aims to prevent frequent handovers by selecting a serving BS from the macro BS tier.
     The small and macro- BSs are represented by $\mathbb{BS}s_{small}$ and $\mathbb{BS}s_{macro}$. The vehicle speed threshold, the received signal strength indicator threshold, and the BS's load threshold are expressed by $\hat{S}$, $R\hat{S}SI$, and $\hat{L}$, respectively. The cell radius is denoted by $R$ and the dwell time of a vehicle within a cell is represented by $T_{d_{ij}}$. The distance and azimuth between a base station $B_i$ and vehicle $V_j$ are represented by $d_{ij}$ and $\theta_{ij}$, respectively. Threshold function is applied to speed-up the searches for the serving base station where the BSs that pass the RSSI threshold condition will be checked [47]. Our proposed A2T-Boost considers the traffic load of base stations to avoid blocking probability, which refers to the probability that a connection request is blocked due to lack of resources. The adaptability of the proposed protocol is clear by its ability to switch to the macro BS tier if the speed of vehicle is greater than the predefined speed threshold, where the macro BS that has the largest RSSI value is selected to be the serving BS. This adaptability feature of the proposed protocol helps to maintain a good performance in terms of average of dwell time, number of HOs, sum rate, and packet delay. The dwell time of a vehicle within a serving base station is estimated based on Equation 7. The time estimation depends on four parameters which are (1) the radius of a cell, (2) the distance between the current location of a vehicle and the BS, (3) the angle between the vehicle direction and BS, and (4) the vehicle's speed. Considering all factors influencing the staying time will help to choose the serving cell with the longest residence duration.

$$T_d = (d\ cos(\theta) + \sqrt{R^2 - d^2\ sin^2(\theta)})/s \quad (7)$$

     where $s$ is vehicle speed. The raduis of the serving cell is expressed by $R$.
  4) **Data Splitting:** The samples of data are split into two datasets: a training set for building ML models and a testing set for evaluating models. An 80/20 (training/testing set) ratio was applied in this study. Items from the entire dataset were

**Algorithm 1** A2T Labeling Algorithm Pseudocode

**input :** $\mathbb{BS}s_{small}$, $\mathbb{BS}s_{macro}$, $\mathbb{V}$.
**output:** $\mathbb{BS}$
Set $\hat{L}$, $\hat{S}$, $R\hat{S}SI$, *and* $R$.
**for** *each* $V_j \in \mathbb{V}$ **do**
  **if** $V_j.kspeed < \hat{S}$ **then**
    **=== Determine a serving small BS ===**
    $\mathbb{BS}s'_{small}$= *Thresholding* ($\mathbb{BS}s_{small}$, $R\hat{S}SI$);
    $Z = \{B_i | B_i \in \mathbb{BS}s'_{small}$ *and load* $< \hat{L}\}$;
    $T_{d_{ij}} = \frac{d_{ij}\ cos(\theta_{ij}) + \sqrt{R^2 - d_{ij}^2\ sin^2(\theta_{ij})}}{s_i}$; $\forall\ B_i \in Z$
    $\mathbb{BS} = \{B_i | B_i \in Z$ & *has* $max(T_{d_{ij}})\}$;
  **end**
  **if** *Length* ($\mathbb{BS}$==*0)* **then**
    **=== Determine a serving macro BS ===**
    $\mathbb{BS}s'_{macro}$= *Thresholding* ($\mathbb{BS}s_{macro}$, $R\hat{S}SI$);
    $Z = \{B_i | B_i \in \mathbb{BS}s'_{macro}\}$
    $\mathbb{BS} = \{B_i | B_i \in Z$ & *has* $max(RSSI_{ij})$;
  **end**
**end**

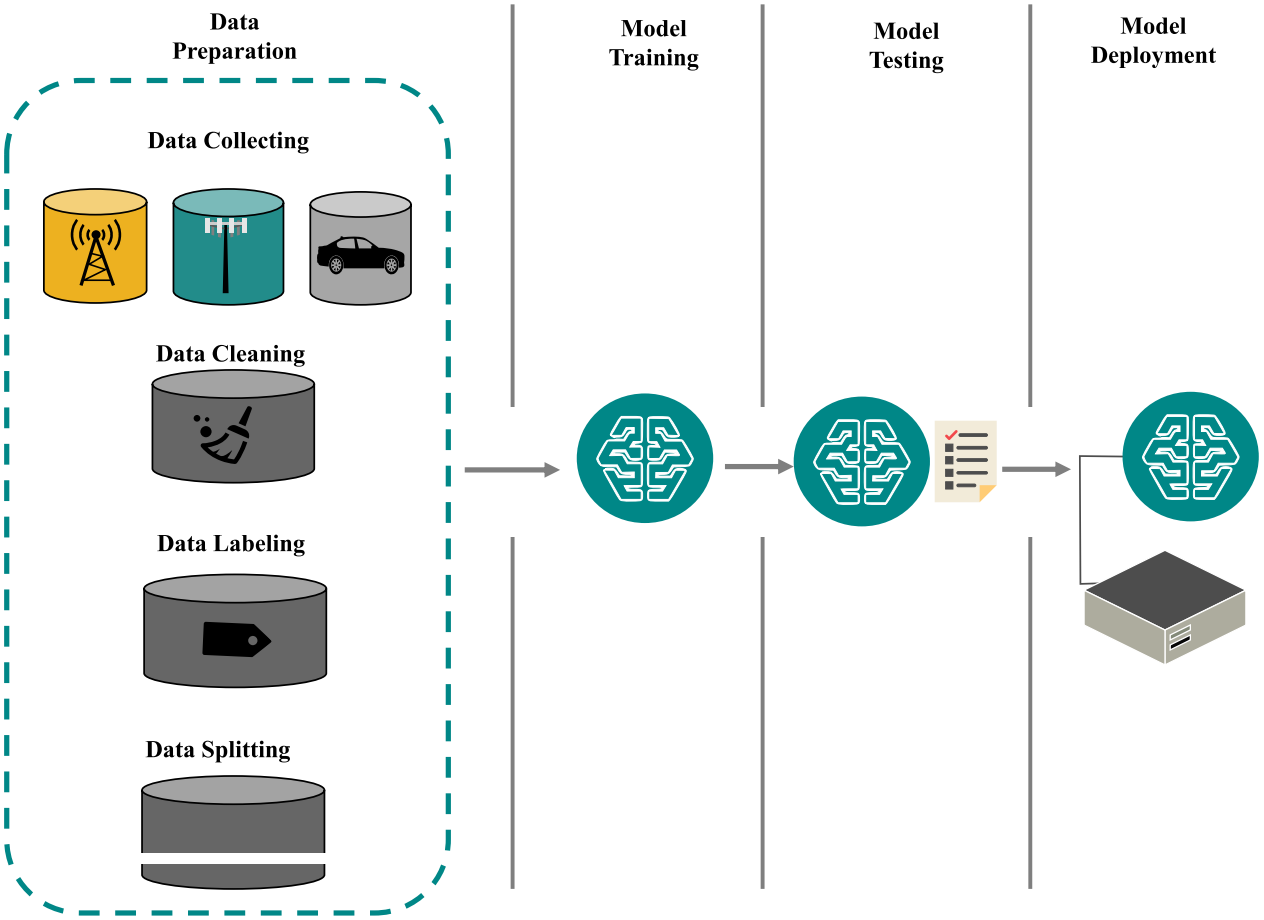



**FIGURE 2.** Stages of construction for the proposed SDN/ML-based model.

randomly chosen for the training sample, while the remaining items were used for the testing sample. It is worth mentioning that all classes should be existed in training and testing sets.

- **Stage 2: ML Model Training:** The machine learning model was trained using the training samples at this stage. In this work, K-Nearest Neighbors (KNN), SVM, Naive Bayes (NB), Artificial Neural Network (ANN), XGBoost, RF, Decision Tree (DT), Adaptive Boost (AdaBoost) classifiers were trained to perform classification task based on supervised learning.

- **Stage 3: ML Model Testing:** The testing samples were used to evaluate the trained ML classifiers. In Section V-A, the effectiveness of the trained ML models is described in detail.

- **Stage 4: ML Model Deployment:** AdaBoost classifier, a trained machine learning model with the highest prediction accuracy, is deployed on the SDN controller. An SDN-based vehicular network's central part, the SDN controller, is physically coupled to other network components, such as wireless base stations [48], [49]. It is in charge of carrying out the adaptive cell selection task using the set up ML model. Algorithm 2 gives the pseudocode for the proposed A2T-Boost scheme. The inputs for the trained adaptive boost model (*AdaBoostMdl*) are the latitude ($V.lat$) and the longitude ($V.lon$) coordinates of the moving vehicle, the azimuth between the direction of the vehicle and the north ($V.Azimuth$) and the vehicle speed ($V.kspeed$). According to the given vehicle information, AdaBoostMdl can predict the serving BS to be connected with.

The idea of turning off some BSs to reduce the power consumption can be applied using our proposed A2T-Boost protocol if a list of sleeping BSs is given to be excluded when performing the prediction task using the trained machine learning model.

### D. THE ARCHITECTURE OF THE PROPOSED SDN/ML-BASED CELL SELECTION STRATEGY

As seen in Figure 3, the architecture is built by involving SDN and ML solutions. SDN is used to carry out effective cell selection and traffic management duties based on the

**Algorithm 2** Pseudocode for A2T-Boost Algorithm

**input :** $V.lat,\ V.lon,\ V.azimuth,\ V.kspeed.$
**output:** Output:$\mathbb{BS}$.
**while** *Vehicle V moves* **do**
  **if** $RSSI < Th \ || \ Length(\mathbb{BS}) == 0$ **then**
    $Input =$
    $[V.lat,\ V.lon,\ V.azimuth,\ V.kspeed,\ ];$
    $\mathbb{BS} =$ AdaBoostMdl $(Input)$;
    *Trigger handover to BS#* $\mathbb{BS}$;
  **end**
**end**

installed trained machine learning model. The geographic coordinates of the vehicle (i.e., latitude (*LAT*) and longitude (*LON*)), the azimuth between the direction of the vehicle and the north (*AZIMUTH*), and the vehicle speed in kilometres per hour are the inputs for the ML-model (*KSPEED*). The ML model can forecast the ideal BS to be associated with, whether it is a macro- or small BS, based on the provided vehicle information.

## IV. THE METHODOLOGY

In this section, the methodology for performing the cell selection task in 5G HUDNs is presented in detail in terms of simulation tool, datasets, and system model.

### A. SIMULATION TOOL

Due to its robust capabilities, the MATLAB 2021b simulator was utilized to model and analyze the effectiveness of the suggested cell selection technique. Additionally, the cell selection process can be carried out in a realistic setting using one of the various tool boxes available in the MATLAB simulator. The simulation was carried out using a powerful gaming machine that has AMD Rayzan 7 processor, 64GB DDR RAM, and NVIDIA EVGA GeForce RTX 2070 Super GPU.

### B. DATASETS

In this work, three datasets were used for to built and evaluate the proposed cell selection scheme in the city of Los Angeles. The first two are for macro and small BSs and the third one is for vehicles.

- **Macro BSs Dataset:** This dataset stores information of 5,248 microwave towers in the country of LA [50]. It was published by LA GeoHub governmental website on September 16, 2016 and it was updated on November 10, 2020. It includes many columns, the most important of which are the BS object identifier (*OBJECTID*) and the latitude and longitude coordinates of the macro BSs. Figure 4 shows a snapshot of the used LA macro BSs dataset.
- **Small BSs Dataset:** The city of Los Angeles's 1,796 5G small base stations mounted on streetlight poles are included in this collection. The primary website for LA, data.LAcity.org, is the one that publishes it [51]. On November 10, 2020, metadata was created, and on October 8, 2021, it was updated [52]. It has two columns: small BS IDs (*SLID*) and the geometry of SBSs (*the_geom*). The small BSs' latitude and longitude coordinates are listed in the *the_geom* column.
  The distributions of the macro and small BSs in Los Angeles based on the mentioned BSs datasets are shown in Figure 5.
- **Vehicles Dataset:** A dataset of vehicles in a high-density area of small cells in Los Ang is used in this work, which was proposed in [53]. It called Vehicle Dataset in the city of LA (VehDS-LA). It was generated using Google Maps and the MATLAB simulator. The VehDS-LA has 74,170 samples that are existed on fifteen LA streets and each vehicle sample has four features. The vehicle dataset has five columns which are street name, latitude, longitude, azimuth and speed.

**TABLE 2. Numbers of macro and small BSs and vehicle samples.**

| | | |
|---|---|---|
| Number of macro BSs | 38 | |
| Number of small BSs | 198 | |
| Number of vehicle samples | 74,170 | Training (80%): 59,336 |
| | | Testing (20%): 14,834 |

### C. SYSTEM MODEL

#### 1) NETWORK MODEL

In this paper, a two-tier mmWave-based heterogeneous ultra-dense network is considered. The first tier, which consists of traditional LTE macro-cells, runs at a carrier frequency of 2 GHz with a bandwidth of 10 MHz. The second tier, which includes 5G small cells, works at a carrier frequency of 28 GHz with a bandwidth of 500 MHz. The system model represents the distribution of base stations in selected areas in Los Angeles that have a high-density of small cells. Figure 6 shows the system model in the selected area in LA, showing the positions of macro and small cells and the distribution of vehicle samples on the surrounding streets. Table 2 gives the numbers of macro and small BSs and vehicle samples in the selected area in LA. The vehicles samples were divided into 80% for training and 20% for testing purposes.

The architecture of the proposed SDN-based heterogeneous network is represented in Figure 7. The network is split into many areas and a distributed set of SDN controllers are working in a flat formation. Each SDN controller manages a specific area and performs the cell selection task based on the installed trained ML model. Organizing the SDN controllers in flat fashion reduces the control delay and improves the resiliency of the system. SDN switch is responsible for directing the flow of data in the determined area. Eastbound/westbound interfaces are used to connect all SDN controllers to exchange network information.

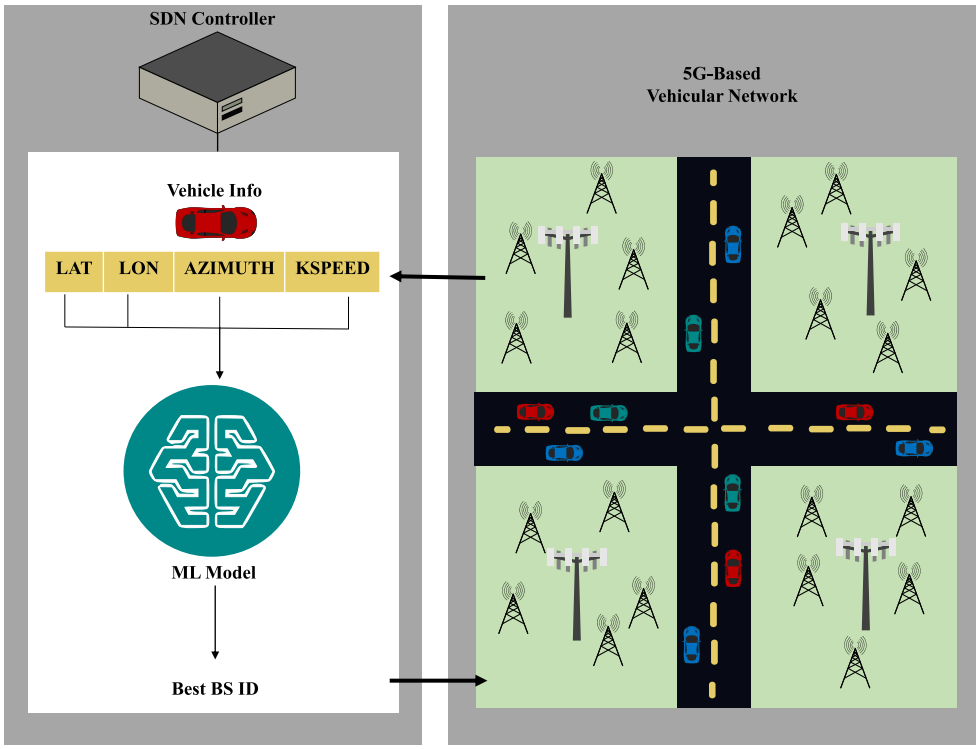



**FIGURE 3.** The architecture of the Proposed SDN/ML-Based solution.

| | OBJECTID | city | state | latitude | longitude |
|---|---|---|---|---|---|
| 1 | 39706 | Los Angeles | CA | 33.40222222 | -118.4138889 |
| 2 | 39707 | Los Angeles | CA | 33.34305556 | -118.3208333 |
| 3 | 39708 | Los Angeles | CA | 33.40222222 | -118.4138889 |
| 4 | 39709 | Los Angeles | CA | 33.34305556 | -118.3208333 |
| 5 | 39736 | Los Angeles | CA | 34.35166667 | -117.6744444 |
| 6 | 39737 | Los Angeles | CA | 34.32416667 | -118.5816667 |
| 7 | 39741 | Los Angeles | CA | 34.35166667 | -117.6744444 |
| 8 | 39742 | Los Angeles | CA | 34.32416667 | -118.5816667 |
| 9 | 39922 | Los Angeles | CA | 34.26777778 | -118.237 |
| 10 | 39936 | Los Angeles | CA | 34.32638889 | -118.5870278 |
| 11 | 39942 | Los Angeles | CA | 34.325 | -118.5775833 |
| 12 | 39943 | Los Angeles | CA | 34.325 | -118.5775833 |
| 13 | 39959 | Los Angeles | CA | 34.325 | -118.5775833 |
| 14 | 39961 | Los Angeles | CA | 34.35222222 | -117.6778333 |
| 15 | 39972 | Los Angeles | CA | 34.31777778 | -118.50925 |

**FIGURE 4.** A snapshot of the LA macro BSs dataset.

### *D. PROPAGATION AND TRAFFIC MODELS*

In this study, path-loss (PL), fading, and shadowing are considered, which are the three essential parts of the propagation channel model. The principal losses that impact the strength of wireless transmissions are those mentioned above. To determine the received signal power at a specific separation from the serving base station, the 3GPP path loss models are employed. According to 3GPP Technical Report (TR) 38.901 version 16.1.0 [54], the urban macro-cell-non-line-of-sight (UMa-NLOS) PL model is used by the macro-BSs

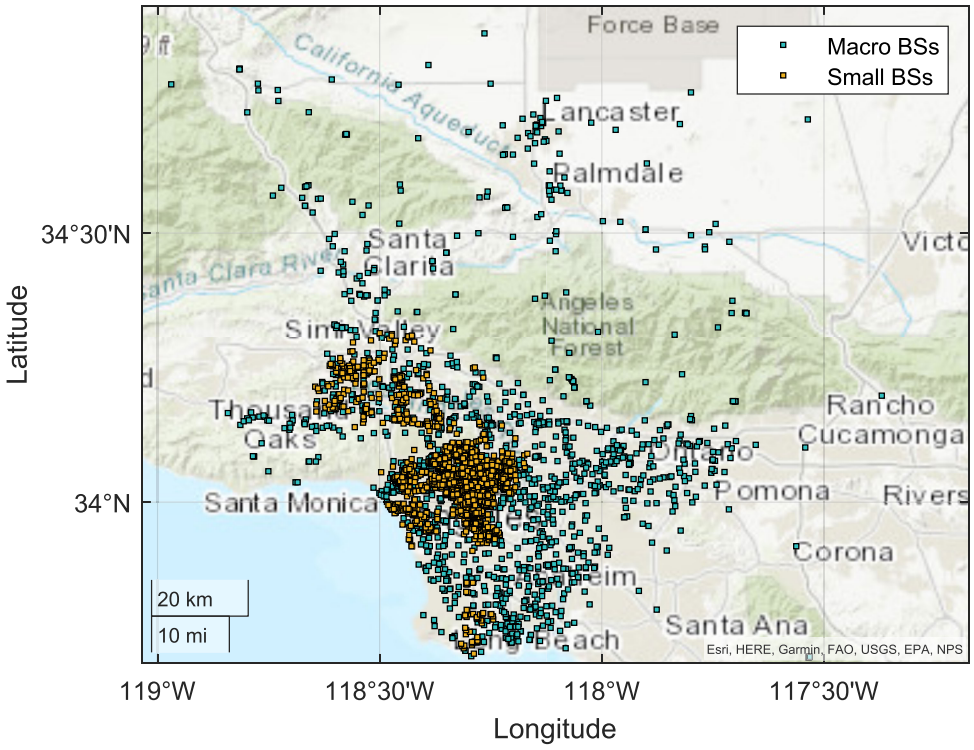



**FIGURE 5.** Distribution of BSs in Los Angeles.

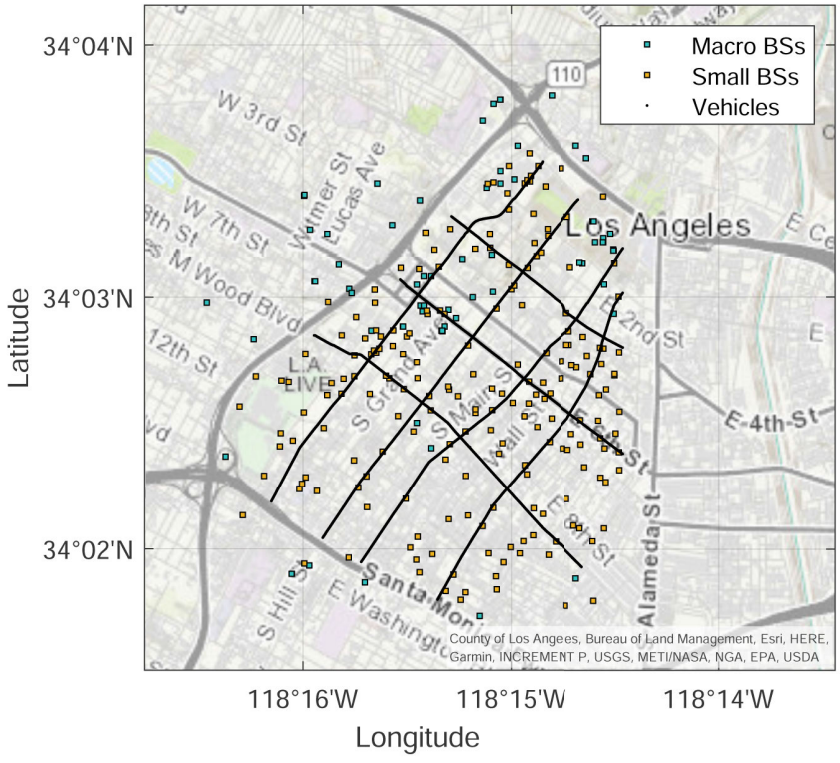



**FIGURE 6.** System model in Los Angeles.

tier, whereas the urban microcell-line-of-sight (UMi-LOS) (street canyon) model is used by the small BSs tier. Rayleigh fading is taken into account in this study because it provides a close approximation of true wireless channel conditions. It has a unit mean and follows an independent exponential distribution [55]. Additionally, because it is frequently

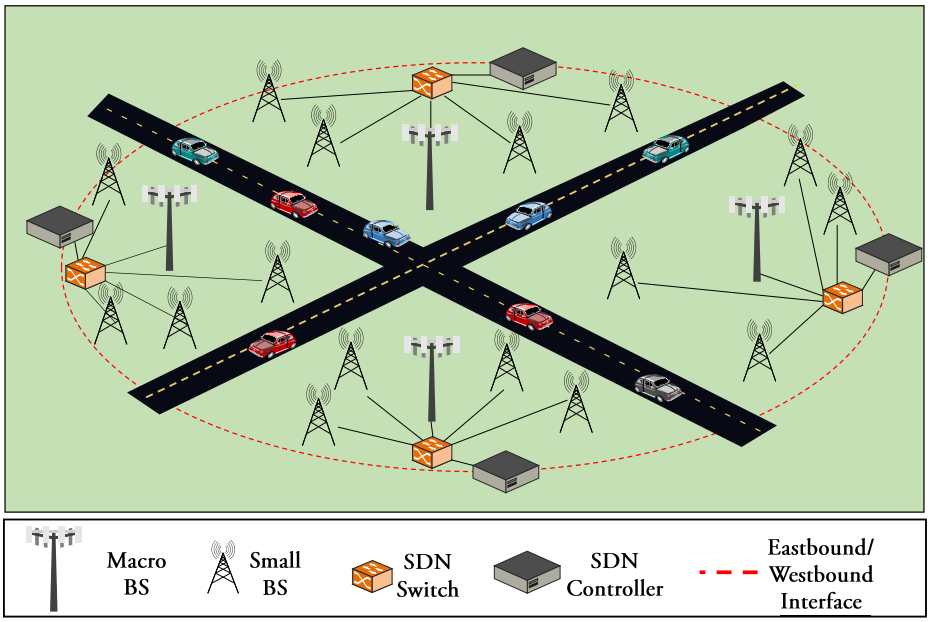



**FIGURE 7.** The architecture of the proposed SDN-based HUDN.

employed to describe the link between RSSI and range, log-normal shadowing is taken into account [56]. It has a standard deviation of 4 dB and 6 dB for small BS tier and macro-BS tier, respectively based on the used 3GPP specifications. The formula of RSSI is given in Equation 8.

$$RSSI_{ij} = p_{tx_i} - \zeta_{ij}(d) \;\; \forall \, V_j \in \mathbb{V} \; and \; \forall \, B_i \in \mathbb{BS}s. \qquad (8)$$

For the traffic model, file transfer requests are considered which follow a Poisson point process that has arrival rate $\lambda$.

#### 1) MOVEMENT BEHAVIOUR MODEL

Vehicle movement behaviour is modeled based on a process called *reckoning*. The new geographical position of a vehicle is reckoned based on the current vehicle location (latitude and longitude coordinates), the movement distance, and the azimuth between vehicle direction and the north (Figure 8). In MATLAB simulator, reckon function belongs to Mapping Toolbox, which is a toolbox that includes many algorithms and functions for transforming and visualizing the geographic information. Figure 9 shows vehicle movement process on S San Perdo and the 6th streets based on the reckoning process.

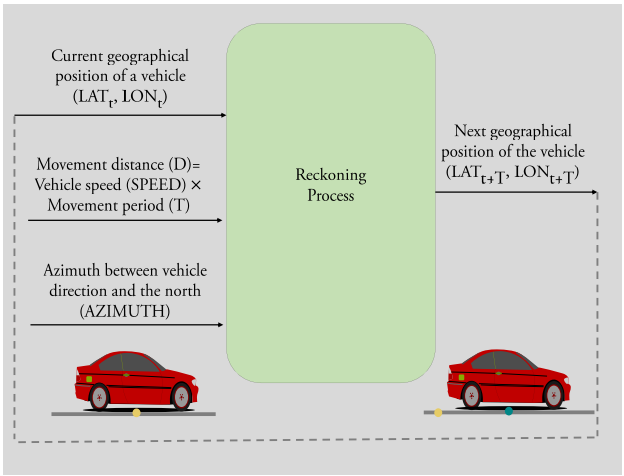



**FIGURE 8.** Modeling vehicle movement behaviour based on reckoning process.

## V. PERFORMANCE EVALUATION

In this section, the performance metrics, which are the key performance indicators (KPIs) that are used to evaluate the trained machine learning models and the proposed cell selection scheme, are discussed.

### A. EVALUATION OF THE TRAINED ML MODELS

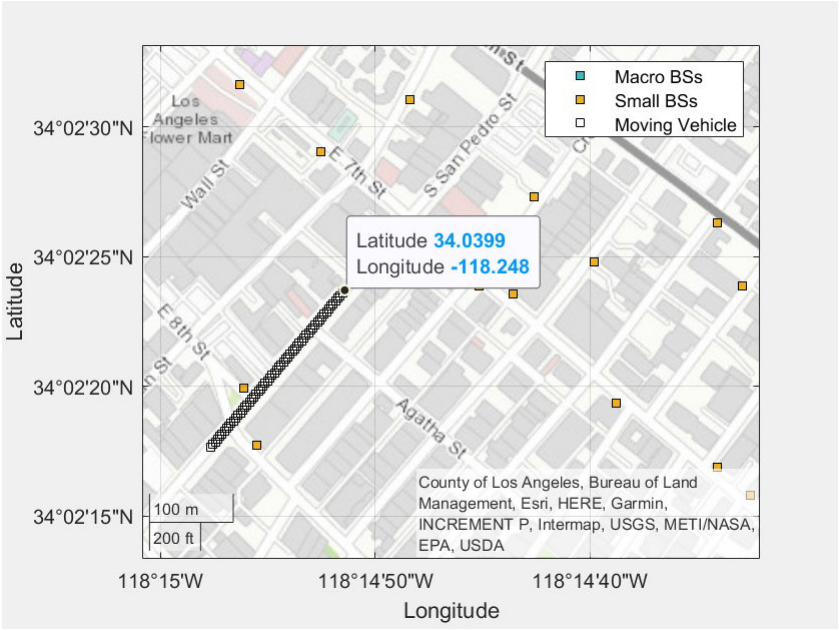
Common metrics for measuring prediction mistakes include root mean square error (RMSE) and mean absolute error (MAE) [57], [58]. The root mean square error calculates the deviation between the predicted and actual values. The average of the absolute difference between the predicted and the target values is calculated using the mean absolute error. Equations (9) and (10) show the definitions of RMSE and MAE, respectively.

$$RMSE = \sqrt{\frac{1}{\mathcal{N}} \Sigma_{i=1}^{\mathcal{N}} (\hat{y}_i - y_i)^2} \qquad (9)$$

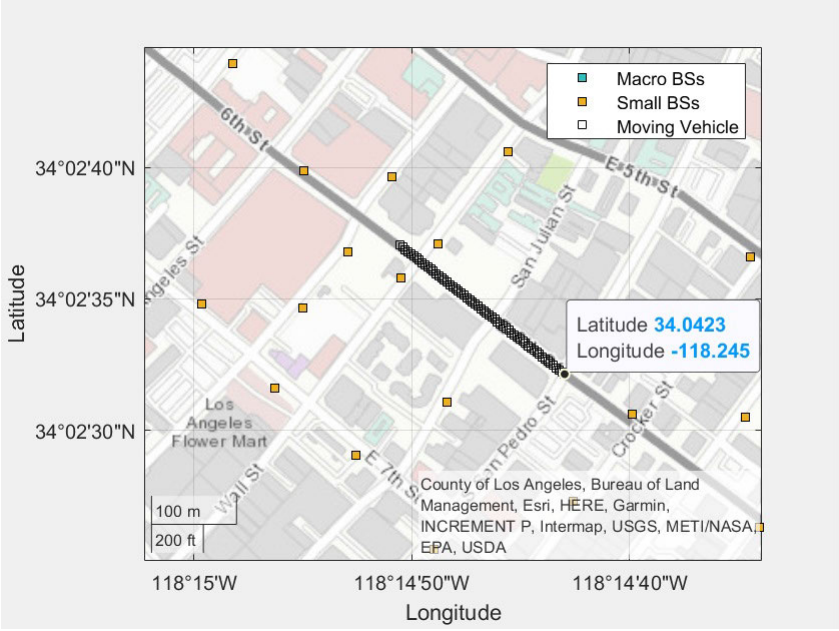
$$MAE = \frac{1}{\mathcal{N}} \Sigma_{i=1}^{\mathcal{N}} |y_i - \hat{y}_i|, \qquad (10)$$

where the expected and the target small BSs are represented by $\hat{y}$ and $y$, respectively, and the number of testing samples is given by $\mathcal{N}$.

A confusion matrix, also known as a contingency table, is built to assess the performance of a trained ML-based




(a) A moving vehicle on S San Perdo Street.




(b) A moving vehicle on the 6th Street.

**FIGURE 9.** **Displaying vehicle movement on some LA streets based on reckoning process.**

model [59]. An efficient tool for reporting the percentages of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) is the confusion matrix [60]. The TP refers to the number of samples that are predicted as positive and they are actually positive, while the TN refers to the number of samples that are predicted as negative and the target class is also negative. When the ML model classifies a sample as positive but the target class is negative, FP classification occurs. If a sample is classified as negative and it is actually positive, the classification is considered as FN [61]. Accuracy refers to ratio of the number of samples that are classified correctly to the total number of testing samples [62], as given in Equation (11). Sensitivity is defined as the ratio of true positives to the summation of true positives and false negatives (Equation (12)). Specificity means the ratio of true negatives to the summation of false positives and true negatives (Equation (13)) [63]. Precision is the ratio of true positives to all positives (Equation (14)) [64]. F-score

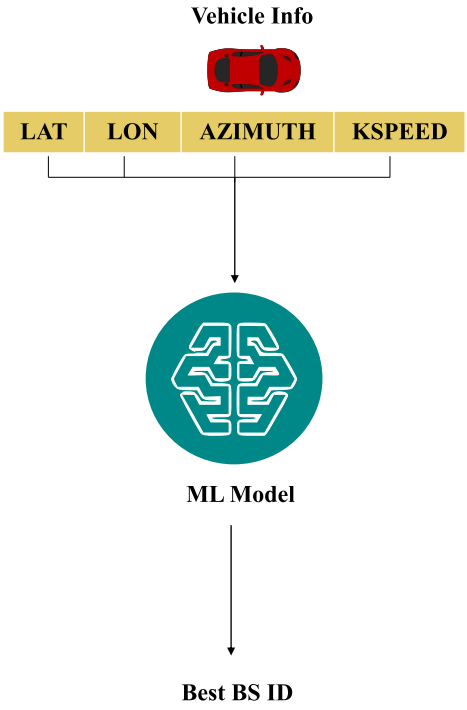



**FIGURE 10.** The prediction process of the serving BS.

is a combination of precision and sensitivity, as shown in Equation (15) [65]. Geometric mean (G-mean) refers to the square root of the product of sensitivity and specificity [66], as given in Equation (16).

$$Accuracy = \frac{Number\ of\ correctly\ classified\ samples}{Total\ number\ of\ testing\ samples} \tag{11}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{12}$$

$$Specificity = \frac{TN}{FP + TN} \tag{13}$$

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$F\text{-}score = \frac{2TP}{2TP + FP + FN} \tag{15}$$

$$G\text{-}mean = \sqrt{(Sensitivity \times Specificity)}. \tag{16}$$

In this paper, KNN, NB, SVM, ANN, DT, RF, AdaBoost, and XGBoost are trained to perform the prediction of the best serving base station based on the proposed scheme, as shown in Figure 10. Based on the selected area in LA, the number of BS classes is 178 of which 142 are for small BSs and the rest are for macro BSs. A grid search (GS), which is the process of finding the optimal combination of the hyperparameters of the targeted ML model [67], is performed. In addition, cross-validation (CV) is used, which is the standard method for testing the validity of the trained ML model by splitting the data into K folds (10 in this work) and then the model is trained K times using all data folds except one [68]. The parameters of the trained ML models are given in Table 3.

**TABLE 3.** The parameters of the trained ML models.

| ML models | Parameters |
|---|---|
| AdaBoost | max_depth=20<br>num_estimators=100<br>learning rate = 0.1 |
| RF | max_depth=20<br>num_estimators=300<br>learning rate = 0.1 |
| DT | max_depth=20 |
| XGBoost | max_depth=20<br>n_estimators=300<br>learning rate = 0.1 |
| KNN | k=1<br>metric='minkowski'<br>p=2 |
| SVM | C=1000<br>gamma =1<br>kernel = 'rbf' |
| ANN | batch size = 16<br>epochs = 200<br>optimization algo= 'adam'<br>learning rate = 0.01<br>num_neurons=10<br>activation func='ReLU' |

The evaluation values of the trained ML models in the selected area in LA is provided in Table 4. Figure 11 represents the prediction accuracy of the trained ML classifiers. As illustrated in the results, the AdaBoost model achieves a high prediction performance with low percentages of error compared with the other trained models. Therefore, it is selected as the classifier model in this work.

Inference time, which is the time taken to predict a serving base station ID of a vehicle test sample, has been calculated using the MATLAB commands. We found that our proposed AdaBoost model takes 0.23612 milliseconds to predict the BS ID based on the specifications of the gaming computer that are mentioned in IV-A. Based on our findings, the prediction duration is very short and the maximum round-trip end-to-end delay is less than 10 ms. Therefore, our proposed A2T-Boost cell selection scheme is suitable for 5G networks because it meets the latency requirement.

### B. EVALUATION OF THE PROPOSED A2T-BOOST

#### 1) EVALUATION METRICS

To evaluate the performance of our proposed A2T-Boost approach, nine performance metrics were utilized. These metrics are the average of (a) dwell time, (b) number of HOs, (c) number of HO failures and unnecessary HOs, (d) downlink sum rate, (e) achievable spectral efficiency, (f) network energy efficiency, (g) packet delay, (h) radio link failure rate, and (i) handover interruption time.

- **Average Dwell Time** The average dwell time of a vehicle within a serving base station ($E(T_d)$) is calculated

**TABLE 4.** The trained ML models evaluation values in Los Angeles.

| Performance Metrics | AdaBoost | Random Forest | Decision Tree | XGBoost | KNN | NB | SVM | ANN (1 Layer) | ANN (2 Layers) |
|---|---|---|---|---|---|---|---|---|---|
| RMSE | 5.18 | 5.24 | 5.95 | 7.18 | 7.43 | 60.81 | 35.25 | 32.35 | 36.12 |
| MAE | 0.28 | 0.29 | 0.36 | 0.50 | 0.57 | 17.43 | 12.14 | 10.88 | 13.42 |
| Accuracy (%) | 99.58 | 99.56 | 99.39 | 99.19 | 98.87 | 85.86 | 79.78 | 79.15 | 75.92 |
| Sensitivity (%) | 99.58 | 99.56 | 99.39 | 99.19 | 98.87 | 85.86 | 79.78 | 79.15 | 75.92 |
| Specificity (%) | 99.998 | 99.997 | 99.996 | 99.995 | 99.99 | 99.92 | 99.88 | 99.88 | 99.86 |
| Precision (%) | 99.58 | 99.56 | 99.39 | 99.19 | 98.87 | 85.86 | 79.78 | 79.15 | 75.92 |
| F-score (%) | 99.58 | 99.56 | 99.39 | 99.19 | 98.87 | 85.86 | 79.78 | 79.15 | 75.92 |
| G-mean (%) | 99.79 | 99.78 | 99.69 | 99.59 | 99.43 | 92.62 | 89.27 | 88.91 | 87.07 |

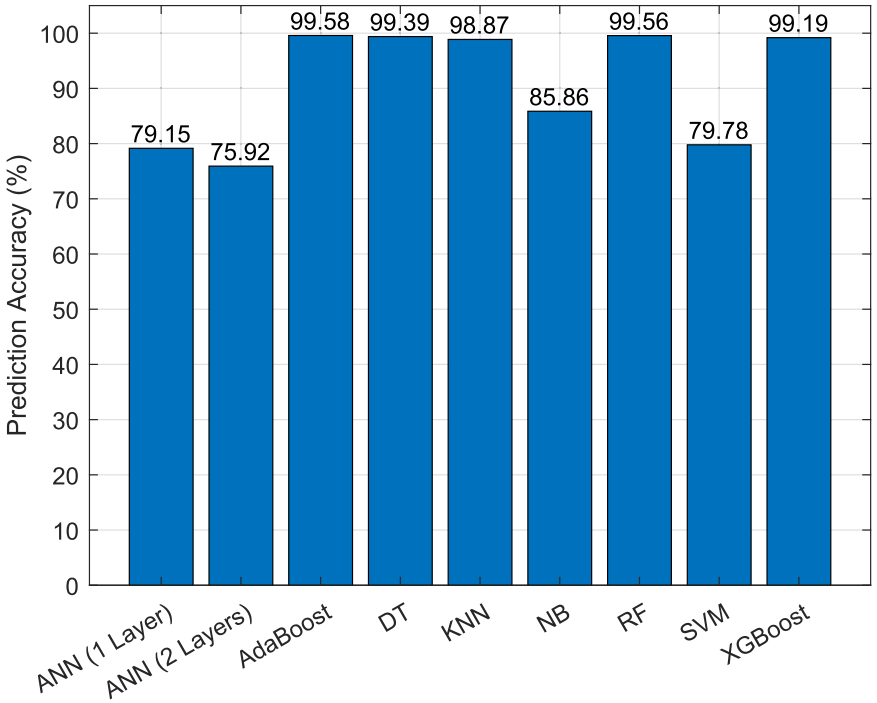



**FIGURE 11.** Accuracy of the trained ML classifiers.

based on Equation (17).

$$E(T_d) = \frac{\sum_J (\sum_I T_d / N_{HO})}{J} \tag{17}$$

where $J$ stands for the number of vehicles and $T_d$ is the amount of time a vehicle $V_j \in J$ spends within a serving base station $B_i \in I$. The symbol $N_{HO}$ represents the number of a vehicle handovers.

- **Average Number of Handovers** Handover occurs when the wireless connection from a current serving BS is switched to another BS. The average number of HOs ($E(N_{HO})$) is calculated based on Equation (18).

$$E(N_{HO}) = \frac{\sum_J N_{HO}}{J} \tag{18}$$

- **Average Number of Handover Failures** HO failure happens when a vehicle's dwell time inside a cell is smaller than the HO delay($\tau_i$) [69]. The probability of HO failure ($P_f$) can be calculated based on the following lemma.

*Let HO latency to move into a cell be denoted by $\tau_i$ and the time threshold of HOF*

*be denoted by $T_1$; then,*

$$P_f = \begin{cases} \frac{2}{\pi}[\sin^{-1}(\frac{s\tau_i}{2R}) - \sin^{-1}(\frac{sT_1}{2R})] & 0 \leqslant T_1 \leqslant \tau_i \\ 0 & \tau_i < T_1 \end{cases} \quad (19)$$

$$T_1 = \frac{2R}{s} sin(sin^{-1}(\frac{s\tau_i}{2R}) - \frac{2}{\pi}P_f); \quad 0 \leqslant P_f \leqslant 1 \quad (20)$$

where $s$ is the speed of a vehicle and $T_1$ is the time threshold of HO failures. The acceptable value of $P_f$ in this study is 0.02.
The calculation of the average number of HO failures, $E(N_f)$, is presented in Equation (21).

$$E(N_f) = P_f \times E(N_{HO}) \quad (21)$$

- **Average Number of Unnecessary Handovers** Unnecessary HO, often referred to as false HO, is a-inessential procedure that occurs when the total amount of time taken to move into ($\tau_i$) and out ($\tau_o$) of a small cell exceeds the time spent there [70]. The following lemma gives the calculation of the probability of unnecessary HO ($P_u$).
*Let HO latencies to move into a cell and to move out of the cell be $\tau_i$ and $\tau_o$ and the time threshold of UHO be denoted by $T_2$; then,*

$$P_u = \begin{cases} \frac{2}{\pi}[\sin^{-1}(\frac{s(\tau_i+\tau_o)}{2R}) - \sin^{-1}(\frac{sT_2}{2R})] \\ \quad 0 \leqslant T_2 \leqslant (\tau_i+\tau_o) \\ 0 \quad (\tau_i+\tau_o) < T_2 \end{cases} \quad (22)$$

$$T_2 = \frac{2R}{s} \sin(\sin^{-1}(\frac{s(\tau_i+\tau_o)}{2R}) - \frac{2}{\pi}P_u) \quad ; 0 \leqslant P_u \leqslant 1 \quad (23)$$

Based on Equation (24), the mean of unnecessary of HO, $E(N_u)$, is estimated. The allowable value of $P_u$ in this study is 0.04

$$E(N_u) = P_u \times E(N_{HO}) \quad (24)$$

- **Average Downlink Sum Rate** The sum-rate of a moving vehicle $V_j$ is obtained by summing the data rate values achieved by the vehicle over the network [37], as given in Equation (25).

$$SumRate_j = \sum_i R_{ij}, \quad \forall\, V_j \in \mathbb{V} \;\&\; \forall\, B_i \in \mathbb{BS}s. \quad (25)$$

where $R_{ij}$ is the achievable DL data rates of the vehicle $V_j$ serving by small base stations $B_i$. It can be calculated based on Shannon's formula, as given in Equation (26).

$$R_{ij} = BW \;\; log_2(1+\gamma_{ij}) \quad (26)$$

The signal-to-interference-plus-noise ratio is calculated by dividing the strength of the signal received from the serving BS by the sum of interference power from other BSs and the power of the noise [71]. The SINR at vehicle $V_j$ associated with the base station $B_i$, which is expressed as $\gamma_{ij}$, can be calculated as given in Equation (27). The thermal noise is considered as an additive white Gaussian noise (AWGN) with noise power spectral density ($N_0$), and channel bandwidth ($BW$).

$$\gamma_{ij} = \frac{p_{tx_i}\zeta_{ij}(d)h_{ij}}{\sum_{k\neq i}(p_{tx_k}\zeta_{kj}(d)h_{kj}) + N_0BW} \quad \forall\, V_j \in \mathbb{V} \; and \; \forall\, B_i \in \mathbb{BS}s. \quad (27)$$

The transmitted power of BSs is expressed as $p_{tx}$. The path loss, $\zeta(d)$, is based on the 3GPP PL model. The channel gain, which is denoted by $h$, composes of the effects of log-normal shadowing and Rayleigh fading. A graph-coloring algorithm is used in this work to manage frequency resources and to control interference issue between cells [72]. It is an efficient way to manage interference problems in a random topology network, such as ultra-dense networks [72], [73]. The available frequency resources are modelled as colors. The graph is represented by the equation $\mathbb{G} = (n, l)$, where $n$ denotes BS node nodes and $l$ denotes links between BSs. Each node must have a color, with the requirements that two neighbouring nodes use distinct frequency resources and that there be a minimum number of colors. As a result, interference can be prevented by allocating various colors (resources) to nearby BSs.

- **Average Downlink Spectral Efficiency** Spectral efficiency (SE) refers to the achievable data rate in bits per second (bps) per channel bandwidth in hertz [74]. The SE of vehicle $V_j$ provided by BS $B_i$, is shown in Equation (28).

$$SE_{ij}(bps/Hz) = \frac{R_{ij}}{BW} \quad (28)$$

- **Average Network Energy Efficiency** An important consideration while creating cell selection methods is energy efficiency (EE). It is the entire data rates that can be achieved divided by the total power used [75]. Equation (29) provides the energy efficiency formula, which is expressed in bits/joules.

$$EE(bits/joule) = \frac{Achievable\ sum\ rate(bps)}{Total\ consumed\ power(watts)} \quad (29)$$

- **Average Packet Delay** One of the QoS requirements for 5G networks is to reduce the amount of delays [76], [77]. A packet's downlink delay is a result of delays in transmission, propagation, processing, and queuing [78], [79]. It can be written as Equation (30) shows, where $D_{ijk}$ stands for the vehicle $V_j$ 's download delay when downloading packet $k$ through BS $B_i$.

$$D_{ijk} = D^{trans}_{ijk} + D^{prop}_{ijk} + D^{proc}_{ijk} + D^{queu}_{ijk} \quad (30)$$

Equation (31) illustrates that the transmission delay, $D^{trans}_{ijk}$, is equal to the packet length ($L_k$) divided by the available transmission rate ($R_{ij}$) [80]. The file transfer

protocol (FTP) paradigm is taken into consideration in this study with packet sizes up to 1500 Bytes.

$$D_{ijk}^{trans} = \frac{L_k}{R_{ij}} \tag{31}$$

The amount of time needed for a packet to travel between a base station and a vehicle is referred to as the propagation delay, or $D_{ijk}^{prop}$. It can be calculated by dividing the speed of light ($c$) by the distance between a small BS $B_i$ and a vehicle $V_j$ ($d_{ij}$), as shown in Equation (32).

$$D_{ijk}^{prog} = \frac{d_{ij}}{c} \tag{32}$$

The processing delay, $D_{ijk}^{proc}$, is the time taken to process a packet and it is measured in microseconds [81]. A packet must wait in line at a small base station for a certain amount of time before being sent, or Dqueu. We take into account an M/M/1 queuing system with an average service rate $\mu$ and an average traffic arrival rate $\lambda$. Equation (33) illustrates how Little's theorem, a well-known law in queuing theory, is used to compute queuing delay [82].

$$D_{ijk}^{queu} = \frac{1}{\mu_n - \lambda_n} \tag{33}$$

Equation (34) can be used to calculate the system's average downlink latency.

$$E(D) = \frac{\sum_J (\sum D_{ijk})}{J} \quad \forall\, V_j \in \mathbb{V} \; and \; \forall\, B_i \in \mathbb{BS}s. \tag{34}$$

- **Average Radio Link Failure Rate** SINR is a term used to describe the radio link's quality. When the SINR of a vehicle $V_j$ from the serving base station $B_j$ drops below the out-of-synchronization threshold ($SYN_{out}$) for an RLF detection time, which is also referred to as $T_{RLF}$, radio link failure (RLF) occurs. The vehicle faces an RLF issue when the $T_{RLF}$ timer, which is also known as T310, has run out and the value of SINR does not enhanced to be above the in-synchronization threshold ($SYN_{in}$). The conditions of a radio link failure and recovery, are shown in equations (35) and (36) [83], [84].

$$C_{RLF} : \gamma_{ij} < SYN_{out}; \;\; for\; t_{out} > T_{RLF} \tag{35}$$

$$C_{Recovery} : \gamma_{ij} > SYN_{in}; \;\; for\; t_{in} > T_{RLF} \tag{36}$$

- **Handover Interruption Time** A crucial parameter for evaluating the effectiveness of cell selection strategies is handover interruption time (HIT), which is also known as HO delay. The time when the vehicle's connectivity is lost while the handover operation is being carried out is referred to as the HIT [85]. The formula for HIT is provided in equation (37) and is the product of the break time ($T_{Break}$), processing time ($T_{Proc}$), interruption time ($T_{Interrupt}$), radio access channel time ($T_{RACH}$), and handover completion time ($T_{HC}$). The 3GPP has accepted 50 millisecond to be the typical HO interruption time [86], [87].

$$T_{HIT} = T_{Break} + T_{Proc} + T_{Interrupt} + T_{RACH} + T_{HC} \tag{37}$$

**TABLE 5. Simulation parameters.**

| Simulation Parameters | Values | |
|---|---|---|
| | Macro BS | Small BS |
| Number of BSs in LA | 38 | 198 |
| Carrier frequency (GHz) | 2 | 28 |
| System bandwidth (MHz) | 10 | 500 |
| Transmit power (dBm) | 46 | 30 |
| Path loss model (dB) | 3GPP UMa Model | 3GPP UMi Model |
| Standard deviation of shadow factor (dB) | 6 | 4 |
| Base station height (meters) | 25 | 10 |
| Cell radius (meters) | 1400 | 600 |
| $SYN_{out}$ (dB) | -8 | -12 |
| Number of training samples | 59,336 | |
| Number of testing samples | 14,834 | |
| Vehicle speeds (km/h) | [10-40] | |
| Vehicle height (meters) | 1.8 | |
| RSSI threshold (dBm) | -90 | |
| Speed threshold (km/h) | 30 | |
| Load threshold (%) | 90 | |
| Thermal noise density (dBm/Hz) | -174 | |
| Shadowing | Log-normal | |
| Fast fading | Rayleigh fading | |
| Handover delay (ms) | 50 [89] | |
| $T_{RLF}$ (sec) | 1 | |
| Simulation time (sec) | 500 | |

### 2) RESULTS AND DISCUSSION

The simulation results are given and discussed in this section. In comparison to the conventional max-RSSI, Qin et al. HO RTP [38], and Zappone et al. ANN-based [43] techniques, the performance of the new A2T-Boost is evaluated. In addition, Kapoor et al. cell selection approaches [23], which are MD-VD, ML-VD, NN-S and NN-O, are compared to our proposed scheme. The simulation parameters that were utilized in the evaluation of the cell selection strategies are shown in Table 5. Figure 12 depicts the correlation between the average number of handovers and vehicle speeds, whereas Figure 13 illustrates the average dwell duration of vehicles at various driving speeds. The findings indicate that when a vehicle's speed increases, the dwell duration lowers, leading to an increase in the frequency of HOs. In terms of average dwell duration and average number of HOs, the suggested A2T-Boost algorithm outperforms the conventional scheme, which is based on the maximum RSSI values, Qin et al. HO RTP, Zappone et al. ANN-based, and Kaproor et al. approaches. The suggested A2T-Boost chooses the small BS with the longest dwell time when it is implemented for this reason. When the speed threshold is exceeded, the nearest macro BS is chosen to prevent needless HOs. The NN-O scheme selects a BS in the vehicle direction that has the smallest azimuth in the range between /2 and, making it the second-best technique in terms of average stay time and number of HOs. In other words, the little BS that is chosen

is the one that is closest to the vertical line from the car's location and is in the left quadrant of the vehicle. As a result, the car is linked to the serving BS for a longer time, which lowers the average HO rateHowever, the average dwell time and the quantity of HOs are superior than the NN-O scheme by 44.47% and 28.57%, respectively, for the suggested A2T-Boost. The BS with the smallest azimuth within the NN-O scheme's allowed range does not always result in the longest dwell time. Due to the fact that the ML-VD scheme depends on selecting the cell with the least amount of load facing the direction of the vehicle, the cell selection will not be fixed regardless of the vehicle density on the streets. In terms of average dwell duration and the number of HOs, our suggested strategy outperforms the ML-VD scheme by 45.78% and 34.78%, respectively. However, depending solely on traffic load does not ensure the longest life of the vehicle within the serving cell. Due to their reliance on a similar method for selecting the serving small cell, the HO RTP and MD-VD schemes perform equally in terms of average dwell time and number of horizontal HOs. Both plans give the closest cell to the vehicle top priority, however relying on this idea will shorten dwell times and increase the number of HOs. In terms of the average number of HOs, our A2T-Boost technique exceeds others by 45.91%. The serving cell is chosen using the NN-S method based on which cell has the smallest angle between 0 and pi. Thus, the cell in the right quadrant for the vehicle that is closest to the lowest horizontal line will be chosen. When the car is travelling forward, picking the cell that is on the lowest horizontal line will result in more HOs. In terms of average dwell duration and the quantity of HOs, our suggested approach performs better than it by 46.27% and 46.43%, respectively. In terms of average dwell time and quantity of HOs, the conventional max-RSSI and Zappone et al. ANN-based systems perform the poorest. Without taking into account the direction and speed of a vehicle, the max-RSSI approach chooses a base station that has the highest RSSI values. Based on the idea of boosting the achievable sum-rate by relying on the shortest distance between a base station and a vehicle, regardless of the direction and speed of the vehicle, the Zappone et al. method uses a trained FF-ANN model to predict the next base station. As a result, when the speed threshold is not surpassed, our A2T-Boost performs better than the conventional and Zappone et al. techniques in terms of average dwell time and number of HOs by 53.49% and 50%, respectively. Due to the adaptive feature of the proposed algorithm, the suggested AT2-Boost also achieves additional improvements in terms of the average staying duration and quantity of handovers with vehicles that exceed the speed threshold.

Figures 14 and 15 show the typical amount of HO failures and unnecessary HOs at various speeds in the chosen areas of Los Angeles. It is obvious that increasing speed causes an increase in the average number of unnecessary and unsuccessful HOs. However, we discovered that when compared to the conventional max-RSSI, Qin et al. HO RTP, Zappone et al. ANN-based, and Kapoor et al. techniques, the suggested

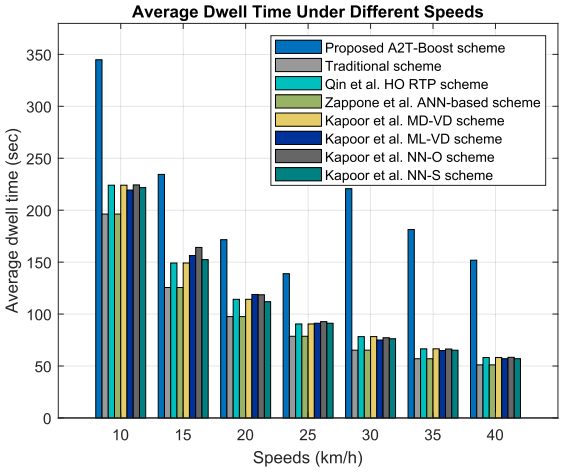



**FIGURE 12.** Average dwell time vs vehicle speed.

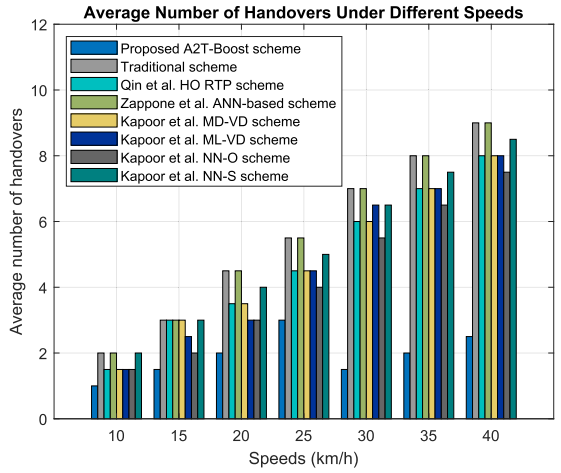



**FIGURE 13.** Average number of handovers vs vehicle speed.

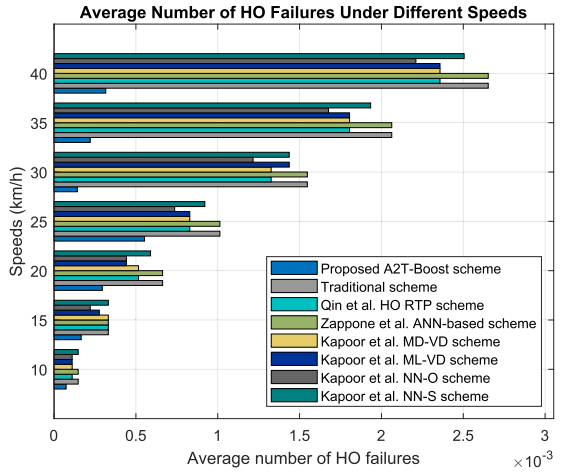



**FIGURE 14.** Average HO failures vs vehicle speed.

A2T-Boost algorithm obtains the lowest average numbers of HO failures and unnecessary HOs. This is because the A2T-Boost technique depends on precisely calculating dwell

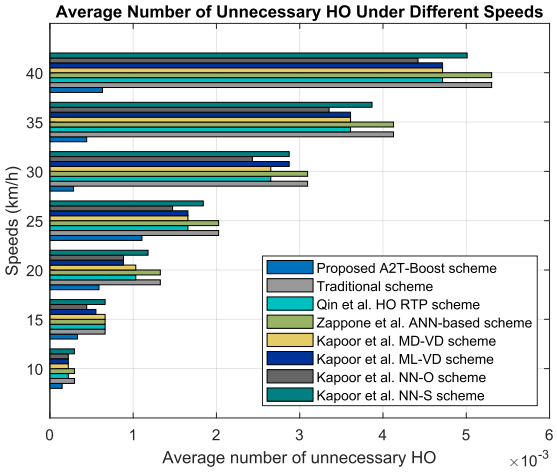



**FIGURE 15.** Average unnecessary HOs vs vehicle speed.

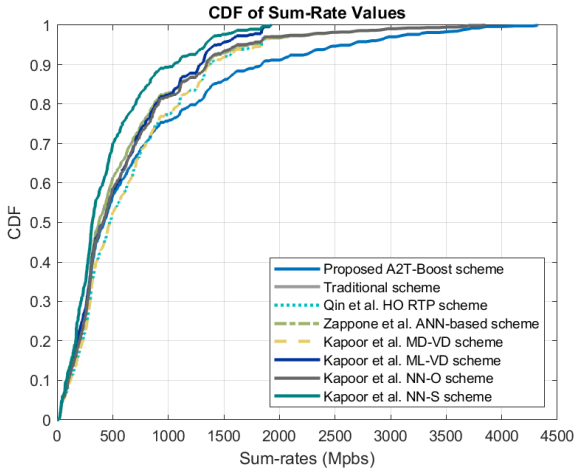



**FIGURE 16.** CDF of achievable downlink data rates by a vehicle during simulation time.

time and has the adaptability to switch between multiple BSs tiers based on a particular speed threshold. As a result, when the vehicle speed is below the speed threshold, the suggested strategy is 28.05% better than the NN-O scheme. The A2T-Boost performs 49.57% better than the classic max-RSSI and Zappone et al. ANN-based approaches, and it performs 39.18% better than the HO RTP and MD-VD methods. Additionally, the proposed strategy outperforms the ML-VD approach by 34.44%. When the vehicle speed exceeded the threshold as a result of the association upgrading to the macro-BS tier, the A2T-Boost method achieved further improvements. Figure 16 shows the cumulative distribution function (CDF) of a vehicle's possible downlink data rate during the simulation period at a particular speed (15 km/h in this case). The proposed A2T-Boost protocol, as shown in the figure, reaches data rate peaks that are not reached by any other techniques. When a vehicle moves forward while using the A2T-Boost protocol, it approaches close to the middle of a serving wireless cell where the BS is situated, allowing the A2T-Boost to achieve the highest possible value of the data rate. However, relying on signal strength values and giving them a high priority does not ensure that the highest data rates will be reached when a vehicle. Additionally, because of the HO latencies required to complete the handover process, handover across cells results in lower attainable datarates. According to the previously discussed average number of HOs, our A2T-Boost method is the best one. In terms of the average achievable sum rate, the A2T-Boost scheme outperforms the max-RSSI and Zappone et al. ANN-based methods by 5.19% and the HO RTP and MD-VD approaches by 6.84%. is moving. In addition, it is superior to NN-O, NN-S and ML-VD schemes by 3.08%, 33.76%, and 20.96%, respectively. Figures 17 and 18 illustrate the CDF of spectral efficiency and network energy efficiency during the simulation time. As the spectral efficiency is the achievable sum-rate divided by channel bandwidth and the energy efficiency is the achievable sum data rate divided by total consumed power, the A2T-Boost algorithm is superior to the other cell selection schemes. The reason behind this is that it outperforms them in terms of the total achieved downlink data rate, as illustrated in Figure 16. The percentage of improvement in terms of average spectral efficiency and network energy efficiency achieved by vehicles is 5.19% compared with the conventional max-RSSI and Zappone et al. ANN-based methods, while A2T-Boost outperforms the HO RTP by 6.84%. Furthermore, it is superior to Kapoor's et al. schemes by up to 33.76%. Figure 19 displays the average packet delay by utilizing three packet sizes: 500, 1000, and 1500 Bytes. We found that the average delay and packet size are directly related. Due to the delay in packet transmission, the downlink latency grows as the packet size does as well. Because it reduces the average propagation time, the suggested A2T-Boost system outperforms the other approaches in terms of mean packet delay. The propagation delay is a function of the separation between the attached vehicle and the serving BS. When the vehicle is close to the cell centre, as is the case with the A2T-Boost system, the propagation delay is at its lowest value, although it may not be the case with other approaches. Additionally, since a packet's transmission delay is based on the data rate, which is reaching its maximum values, that delay is also lowered. The packet delay will also be improved by lowering the number of HOs. As a result, the suggested method outperforms both conventional and Zappone et al. ANN-based systems by 6.05%. Additionally, it exceeds ML-VD, NN-S, and NN-O by 11.55%, 12.87%, and 2.81%, respectively, superior to HO RTP, MD-VD, and both by a margin of 11.40%.

Figure 20 displays the CDF of SINR values that a vehicle received during the simulation period. As a result, our A2T-Boost system enhances downlink SINR since it produces high SINR values that are not possible with the other cell selection techniques. Using our A2T-Boost system and the other cell selection scheme, we discovered that RLFs happen when SINR values fall below $SYN_{out}$ for $T_{RLF}$, which does

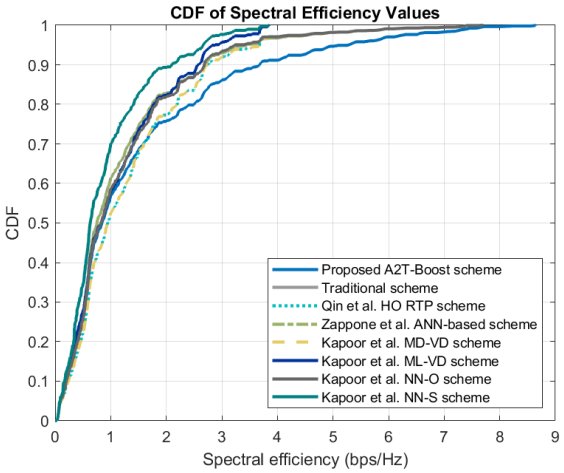



**FIGURE 17.** CDF of achievable downlink spectral efficiency by a vehicle during simulation time.

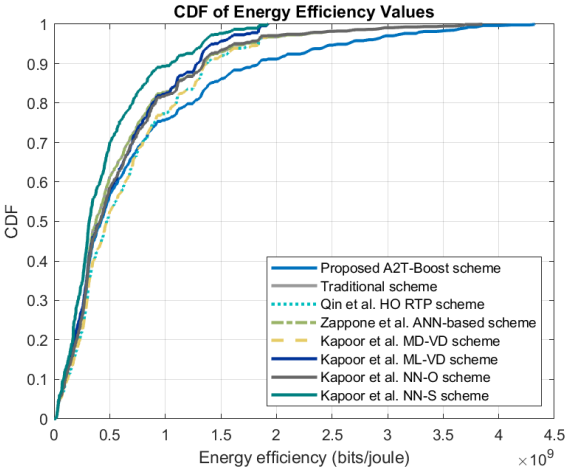



**FIGURE 18.** CDF of network energy efficiency achieved by a vehicle during simulation time.

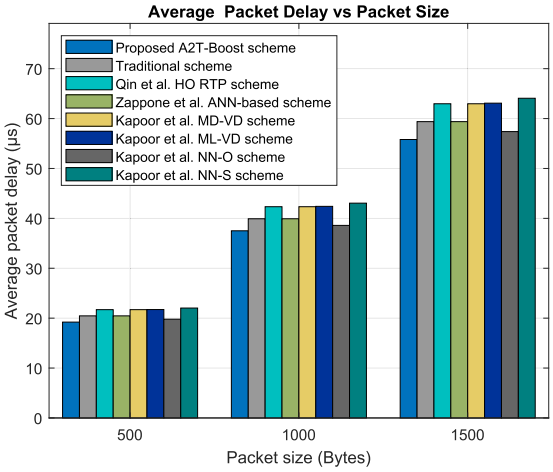



**FIGURE 19.** Average DL packet delay vs packet size.

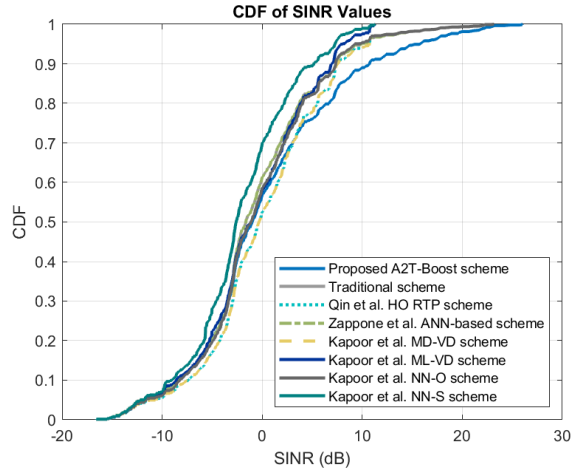



**FIGURE 20.** CDF of received SINR values by a vehicle during simulation time.

not exceed 5.43%. This is because all simulation experiments rely on the establishment of an RSSI threshold, and the cell selection procedure is only applied if the received RSSI value is below this threshold. In fact, by improving the SINR values through the use of interference mitigation measures, radio connection failures can be prevented. The RLF rate can also be decreased by relying on soft handover, which entails connecting to the new BS before breaking the old one.

Figure 21 depicts the relationship between average cumulative handover interruption time and vehicle speed. The graphic illustrates how our suggested A2T-Boost system performs better than previous approaches since it tries to increase the dwell time of cars inside serving cells and thereby reduce the cumulative HIT. To prevent frequent HOs, the suggested approach also applies switching between the small BS tier and macro BS tier if the vehicle speed is higher than the predetermined speed threshold. The Max-RSSI and Zappone et al. ANN-based schemes, according to our research, are the worst cell selection techniques in terms of cumulative HIT. They choose the next BS based on the strongest received RSSI value, which leads to a rise in HOs and a worsening of the cumulative HIT. In terms of cumulative HIT, our procedure performs better than their by 50%. The HO RTP and MD-VD approaches outperform the max-RSSI and Zappone et al. ANN-based systems in terms of cumulative HIT because they provide the cell with the strongest signal strength the highest priority while also taking the vehicle direction into account. As a result, the A2T-Boost is 40% better than the HO RTP and MD-VD techniques. The NN-O scheme is the second-best method in terms of the cumulative HIT because it achieves a long stay time based on selecting a BS that is located in the left quadrant for the vehicle with the nearest vertical line with respect to the vehicle's location. Our proposed scheme outperforms NN-O by 28.57%. The ML-VD does not guarantee decreasing the HO rate of vehicles, but is governed by the distribution of loads between cell. The NN-S method is

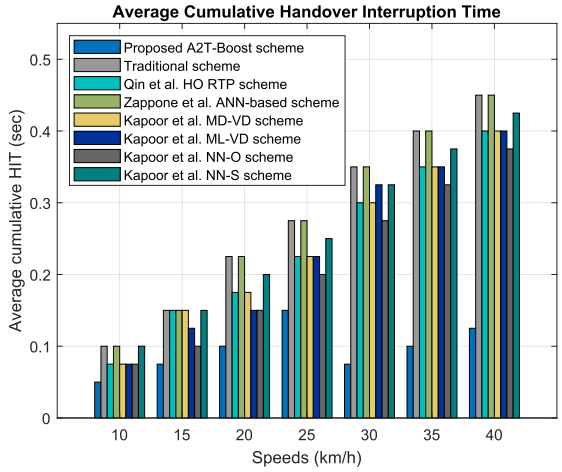



**FIGURE 21.** Average cumulative HIT vs vehicle speed.

based on selecting cells located on the lowest horizontal line with respect to the vehicle and this way of selection does not guarantee a reduction in the number of HOs. Therefore, our proposed A2T-Boost outperforms ML-VD and NN-S in terms of cumulative HIT by 34.78% and 46.43%, respectively.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an adaptive cell selection scheme known as A2T-Boost, which has the ability to adapt to the different features of base stations and of mobile stations. The proposed A2T-Boost protocol can adapt to the changes in (1) cells' size, load, and geographic location, and (2) vehicles' geographic location, speed, and direction. It is an SDN/ML-based strategy that can solve the cell selection issue using a trained AdaBoost model. The downtown of Los Angeles city is taken as a study case in this paper. Simulation results prove that our A2T-Boost can predict the next base station with high prediction performance comparing with other related strategies in terms of average number of handovers by up to 50%. In addition, it improves the average achievable downlink sum-rates and network energy efficiency achieved by vehicles by up to 33.76%. Moreover, the proposed strategy decreases the average packet latency by up to 12.87%. For future work, further performance metrics can be evaluated and additional study cases will be studied to show the applicability of our proposed A2T-Boost scheme.

## APPENDIX

## LIST OF ABBREVIATIONS

The following table gives a list of abbreviations. Table 6 gives a list of all abbreviations.

## ACKNOWLEDGMENT

The authors would like to thank the Deanship of Scientific Research and RSSU at King Saud University for their technical support.

**TABLE 6.** List of main abbreviations.

| Abbreviation | Meaning |
|---|---|
| 3GPP | Third Generation Partnership Project |
| 5G | Fifth-Generation |
| A2T | Adaptive two-Tier |
| A2T-Boost | Adaptive two-Tier based on Adaptive Boosting |
| A3C | Asynchronous Advantage Actor Critic |
| AdaBoost | Adaptive Boost |
| ADA-CS | Adaptive Cell Selection |
| ADAM | Adaptive Moment |
| AI | Artificial iIntelligence |
| ANN | Artificial Neural Network |
| AWGN | Additive White Gaussian Noise |
| BS | Base station |
| CAV | Connected and Autonomous Vehicle |
| CDF | Cumulative Distribution Function |
| CNN | Convolutional Neural Network |
| CoMP | Coordinated Multipoint |
| CRE | Cell-Range-Expansion |
| CS | Cell Selection |
| CV | Cross-Validation |
| DDRL | Distributed Deep Reinforcement learning |
| DT | Decision Tree |
| EE | Energy Efficiency |
| FN | False Negatives |
| FP | False Positives |
| FTP | File Transfer Protocol |
| G-mean | Geometric mean |
| GS | Grid Search |
| HMM | Hidden Markov-Model |
| HO | Handover |
| HO RTP | Handover based on Resident Time Prediction |
| HUDNs | Heterogeneous Ultra-Dense Networks |
| iMACH | improved MACH |
| ITS | Intelligent Transport System |
| KNN | K-Nearest Neighbors |
| KPIs | Key Performance Indicators |
| LA | Los Angeles |
| LPN | Low-Power Node |
| LTE | Long-Term Evolution |
| MACH | Movement-Aware CoMP Handover |
| MADM | Multiple Attributes Decision-Making |
| MAE | Mean Absolute Error |
| MD-VD | Minimum Distance in the Vehicle Direction |
| ML | Machine Learning |
| ML-NetSel | Machine Learning-Network Selector |
| ML-VD | Minimum Load in the Vehicle Direction |
| mMIMO | massive Multiple-Input and Multiple-Output |
| NB | Naive Bayes |
| NN-O | Next Neighbor on the Opposite street |
| NN-S | Next Neighbor on the Same street |
| PLR | Packet Loss Ratio |
| RANs | Radio Access Networks |
| ReLU | Rectified Linear Unit |
| RF | Random Forest |
| RMSE | Root Mean Square Error |
| RNN | Recurrent Neural Network |
| RSRP | Reference Signal Receiving Power |
| RSS | Received Signal Strength |
| RSUs | Roadside Units |
| SDN | Software-Defined Networking |
| SE | Spectral Efficiency |
| SINR | Signal-to-Interference-plus-Noise Ratio |
| SVM | Support Vector Machine |
| TN | True Negatives |
| TP | True Positives |
| UE | User Equipment |
| UMa-NLOS | Urban Macro-cell-Non-Line-of-Sight |
| UMi-LOS | Urban Microcell-Line-of-Sight |
| V2X | Vehicle-to-everything |
| XGBoost | eXtreme Gradient Boosting |

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

[1] H. Hui, Y. Ding, Q. Shi, F. Li, Y. Song, and J. Yan, "5G network-based Internet of Things for demand response in smart grid: A survey on application potential," *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 113972.

[2] J. D. Chimeh, "Compelling services for 5G creation," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2020, pp. 1–6.

[3] J. Dias, J. Rodrigues, V. Soares, J. Caldeira, V. Korotaev, and M. Proença, "Network management and monitoring solutions for vehicular networks: A survey," *Electronics*, vol. 9, no. 5, p. 853, May 2020.

[4] S. Chen, J. Hu, Y. Shi, Y. Peng, J. Fang, R. Zhao, and L. Zhao, "Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G," *IEEE Commun. Standards Mag.*, vol. 1, no. 2, pp. 70–76, Jul. 2017.

[5] N. Raza, S. Jabbar, J. Han, and K. Han, "Social vehicle-to-everything (V2X) communication model for intelligent transportation systems based on 5G scenario," in *Proc. 2nd Int. Conf. Future Netw. Distrib. Syst.*, Jun. 2018, pp. 1–8.

[6] H. Peng, L. Liang, X. Shen, and G. Y. Li, "Vehicular communications: A network layer perspective," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1064–1078, Feb. 2019.

[7] B. Zhang, M. Ma, C. Liu, and Y. Shu, "Delay guaranteed MDP scheduling scheme for HCCA based on 802.11 p protocol in V2R environments," *Int. J. Commun. Syst.*, vol. 30, no. 15, p. e3307, Oct. 2017.

[8] M. Alsaeedi, M. M. Mohamad, and A. A. Al-Roubaiey, "Toward adaptive and scalable OpenFlow-SDN flow control: A survey," *IEEE Access*, vol. 7, pp. 107346–107379, 2019.

[9] C. D. Cajas and D. O. Budanov, "SDN applications and plug-ins in the OpenDaylight controller," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (EIConRus)*, Jan. 2020, pp. 9–13.

[10] A. Hussein, I. H. Elhajj, A. Chehab, and A. Kayssi, "SDN security plane: An architecture for resilient security services," in *Proc. IEEE Int. Conf. Cloud Eng. Workshop (IC2EW)*, Apr. 2016, pp. 54–59.

[11] F. Baskoro, R. Hidayat, and S. B. Wibowo, "LACP experiment using multiple flow table in RYU SDN controller," in *Proc. 2nd Int. Conf. Appl. Inf. Technol. Innov. (ICAITI)*, Sep. 2019, pp. 51–55.

[12] S. B. H. Natanzi and M. R. Majma, "Secure northbound interface for SDN applications with NTRU public key infrastructure," in *Proc. IEEE 4th Int. Conf. Knowl.-Based Eng. Innov. (KBEI)*, Dec. 2017, pp. 452–458.

[13] M. Nkosi, A. Lysko, L. Ravhuanzwo, T. Nandeni, and A. Engelberencht, "Classification of SDN distributed controller approaches: A brief overview," in *Proc. Int. Conf. Adv. Comput. Commun. Eng. (ICACCE)*, Nov. 2016, pp. 342–344.

[14] M. Karakus and A. Durresi, "A survey: Control plane scalability issues and approaches in software-defined networking (SDN)," *Comput. Netw.*, vol. 112, pp. 279–293, Jan. 2016.

[15] S. Mirjalili, H. Faris, and I. Aljarah, *Evolutionary Machine Learning Techniques*. Cham, Switzerland: Springer, 2019.

[16] V. P. Kafle, Y. Fukushima, P. Martinez-Julia, and T. Miyazawa, "Consideration on automation of 5G network slicing with machine learning," in *Proc. ITU Kaleidoscope, Mach. Learn. 5G Future*, 2018, pp. 1–8.

[17] I. A. Alablani and M. A. Arafah, "Enhancing 5G small cell selection: A neural network and IoV-based approach," *Sensors*, vol. 21, no. 19, p. 6361, Sep. 2021.

[18] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities," *J. Internet Services Appl.*, vol. 9, no. 1, pp. 1–99, Dec. 2018.

[19] B. U. Kazi and G. A. Wainer, "Next generation wireless cellular networks: Ultra-dense multi-tier and multi-cell cooperation perspective," *Wireless Netw.*, vol. 25, no. 4, pp. 2041–2064, May 2019.

[20] B. Gholampooryazdi, H. Hämmäinen, S. Vijay, and A. Savisalo, "Scenario planning for 5G light poles in smart cities," in *Proc. Internet Things Business Models, Users, Netw.*, Nov. 2017, pp. 1–7.

[21] A. Pandharipande and P. Thijssen, "Connected street lighting infrastructure for smart city applications," *IEEE Internet Things Mag.*, vol. 2, no. 2, pp. 32–36, Jun. 2019.

[22] X. Ge, S. Tu, G. Mao, and C. X. Wang, "5G ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.

[23] S. Kapoor, D. Grace, and T. Clarke, "A base station selection scheme for handover in a mobility-aware ultra-dense small cell urban vehicular environment," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–5.

[24] Y. Aghazadeh, H. Kalbkhani, M. G. Shayesteh, and V. Solouk, "Cell selection for load balancing in heterogeneous networks," *Wireless Pers. Commun.*, vol. 101, no. 1, pp. 305–323, Jul. 2018.

[25] D. Bega, M. Gramaglia, C. J. Bernardos Cano, A. Banchs, and X. Costa-Perez, "Toward the network of the future: From enabling technologies to 5G concepts," *Trans. Emerg. Telecommun. Technol.*, vol. 28, no. 8, p. e3205, Aug. 2017.

[26] I. A. Alablani and M. A. Arafah, "An adaptive cell selection scheme for 5G heterogeneous ultra-dense networks," *IEEE Access*, vol. 9, pp. 64224–64240, 2021.

[27] A. Papazafeiropoulos, P. Kourtessis, M. Di Renzo, J. M. Senior, and S. Chatzinotas, "SDN-enabled MIMO heterogeneous cooperative networks with flexible cell association," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2037–2050, Apr. 2019.

[28] E. Ganesan, I.-S. Hwang, A. T. Liem, and M. S. Ab-Rahman, "SDN-enabled FiWi-IoT smart environment network traffic classification using supervised ML models," *Photonics*, vol. 8, no. 6, p. 201, Jun. 2021.

[29] I. A. Alablani and M. A. Arafah, "An SDN/ML-based adaptive cell selection approach for HetNets: A real-world case study in London, UK," *IEEE Access*, vol. 9, pp. 166932–166950, 2021.

[30] R. Arshad, H. Elsawy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Handover management in 5G and beyond: A topology aware skipping approach," *IEEE Access*, vol. 4, pp. 9073–9081, 2016.

[31] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Fast cell select for mobility robustness in intra-frequency 5G ultra dense networks," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–7.

[32] A. S. Cacciapuoti, "Mobility-aware user association for 5G mmWave networks," *IEEE Access*, vol. 5, pp. 21497–21507, 2017.

[33] N. Naderializadeh, H. Nikopour, O. Orhan, and S. Talwar, "Feedback-based interference management in ultra-dense networks via parallel dynamic cell selection and link scheduling," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[34] A. Kishida, Y. Morihiro, T. Asai, and Y. Okumura, "Cell selection scheme for handover reduction based on moving direction and velocity of UEs for 5G multi-layered radio access networks," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2018, pp. 362–367.

[35] M. Elkourdi, A. Mazin, and R. D. Gitlin, "Towards low latency in 5G HetNets: A Bayesian cell selection/user association approach," in *Proc. IEEE 5G World Forum (5GWF)*, Jul. 2018, pp. 268–272.

[36] Q. Liu, C. F. Kwong, S. Zhang, L. Li, and J. Wang, "A fuzzy-clustering based approach for MADM handover in 5G ultra-dense networks," *Wireless Netw.*, vol. 28, no. 2, pp. 1–14, 2019.

[37] W. Sun, L. Wang, J. Liu, N. Kato, and Y. Zhang, "Movement aware CoMP handover in heterogeneous ultra-dense networks," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 340–352, Jan. 2021.

[38] Z. Qin, W. Feng, Z. Yue, and H. Tian, "A handover management strategy using residence time prediction in 5G ultra-dense networks," in *Signal and Information Processing, Networking and Computers*. Singapore: Springer, 2021, pp. 808–816.

[39] Y. Zhang, L. Xiong, and J. Yu, "Deep learning based user association in heterogeneous wireless networks," *IEEE Access*, vol. 8, pp. 197439–197447, 2020.

[40] D. S. Wickramasuriya, C. A. Perumalla, K. Davaslioglu, and R. D. Gitlin, "Base station prediction and proactive mobility management in virtual cells using recurrent neural networks," in *Proc. IEEE 18th Wireless Microw. Technol. Conf. (WAMICON)*, Apr. 2017, pp. 1–6.

[41] J. S. Perez, S. K. Jayaweera, and S. Lane, "Machine learning aided cognitive RAT selection for 5G heterogeneous networks," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, Jun. 2017, pp. 1–5.

[42] S. Q. Zhang, F. Xue, N. A. Himayat, S. Talwar, and H. T. Kung, "A machine learning assisted cell selection method for drones in cellular networks," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[43] A. Zappone, L. Sanguinetti, and M. Debbah, "User association and load balancing for massive MIMO through deep learning," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1262–1266.

[44] I. A. M. Balapuwaduge and F. Y. Li, "Hidden Markov model based machine learning for mMTC device cell association in 5G networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[45] H. Khan, A. Elgabli, S. Samarakoon, M. Bennis, and C. S. Hong, "Reinforcement learning-based vehicle-cell association algorithm for highly mobile millimeter wave communication," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 4, pp. 1073–1085, Dec. 2019.

[46] D. Anand, M. A. Togou, and G.-M. Muntean, "A machine learning solution for automatic network selection to enhance quality of service for video delivery," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Aug. 2021, pp. 1–5.

[47] S. S. Moghaddam and K. S. Moghaddam, "Efficient base-centric/user-centric clustering algorithms based on thresholding and sorting," in *Proc. 14th Int. Conf. Innov. Inf. Technol. (IIT)*, Nov. 2020, pp. 131–136.

[48] I. Yaqoob, I. Ahmad, E. Ahmed, A. Gani, M. Imran, and N. Guizani, "Overcoming the key challenges to establishing vehicular communication: Is SDN the answer?" *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 128–134, Jul. 2017.

[49] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, and J. Ren, "SDN/NFV-empowered future IoV with enhanced communication, computing, and caching," *Proc. IEEE*, vol. 108, no. 2, pp. 274–291, Feb. 2020.

[50] *Microwave Towers*. Accessed: Jan. 5, 2022. [Online]. Available: https://geohub.lacity.org/datasets/lacounty::microwave-towers/explore?location=33.805000%2C-118.295000%2C7.99

[51] *Small Cell Locations*. Accessed: Oct. 10, 2021. [Online]. Available: https://data.lacity.org/City-Infrastructure-Service-Requests/Small-Cell-Locations/3nrm-mq6k

[52] *Small Cell Locations*. Accessed: Oct. 10, 2021. [Online]. Available: https://catalog.data.gov/dataset/small-cell-locations

[53] I. A. Alablani and M. A. Arafah, "A new vehicle dataset in the city of Los Angeles for V2X and machine learning applications," *Appl. Sci.*, vol. 12, no. 8, p. 3751, Apr. 2022.

[54] *Technical Specification Group Radio Access Network; Study on Channel Model for Frequencies From 0.5 to 100 GHz (Release 16)*, 3GPP, document no. 38.901, V16.1.0 (Release 16), 2019.

[55] K. Zia, N. Javed, M. N. Sial, S. Ahmed, A. A. Pirzada, and F. Pervez, "A distributed multi-agent RL-based autonomous spectrum allocation scheme in D2D enabled multi-tier HetNets," *IEEE Access*, vol. 7, pp. 6733–6745, 2019.

[56] C. K. Sung, S. Li, M. Hedley, N. Nikolic, and W. Ni, "Skew log-normal channel model for indoor cooperative localization," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–5.

[57] P. Fan, J. Zhao, and I. Chih-Lin, "5G high mobility wireless communications: Challenges and solutions," *China Commun.*, vol. 13, no. 2, pp. 1–13, 2016.

[58] R. Wang, X. Liang, X. Zhu, and Y. Xie, "A feasibility of respiration prediction based on deep bi-LSTM for real-time tumor tracking," *IEEE Access*, vol. 6, pp. 51262–51268, 2018.

[59] D. Chicco, V. Starovoitov, and G. Jurman, "The benefits of the Matthews correlation coefficient (MCC) over the diagnostic odds ratio (DOR) in binary classification assessment," *IEEE Access*, vol. 9, pp. 47112–47124, 2021.

[60] A. Bhatnagar and S. Srivastava, "A robust model for churn prediction using supervised machine learning," in *Proc. IEEE 9th Int. Conf. Adv. Comput. (IACC)*, Dec. 2019, pp. 45–49.

[61] C.-Y. Hsu, S. Wang, and Y. Qiao, "Intrusion detection by machine learning for multimedia platform," *Multimedia Tools Appl.*, vol. 80, no. 19, pp. 29643–29656, Aug. 2021.

[62] J. D. C. Rodrigues, P. P. R. Filho, E. Peixoto, and V. H. C. de Albuquerque, "Classification of EEG signals to detect alcoholism using machine learning techniques," *Pattern Recognit. Lett.*, vol. 125, pp. 140–149, Jul. 2019.

[63] J. Lee, J. Woo, A. R. Kang, Y.-S. Jeong, W. Jung, M. Lee, and S. H. Kim, "Comparative analysis on machine learning and deep learning to predict post-induction hypotension," *Sensors*, vol. 20, no. 16, p. 4575, Aug. 2020.

[64] A. Porto and K. L. Voje, "ML-morph: A fast, accurate and general approach for automated detection and landmarking of biological structures in images," *Methods Ecol. Evol.*, vol. 11, no. 4, pp. 500–512, 2020.

[65] J. Lee, U. Lee, and H. Kim, "PASS: Reducing redundant notifications between a smartphone and a smartwatch for energy saving," *IEEE Trans. Mobile Comput.*, vol. 19, no. 11, pp. 2656–2669, Nov. 2020.

[66] M. Khanna, M. Aggarwal, and N. Singhal, "Empirical analysis of artificial immune system algorithms for aging related bug prediction," in *Proc. 7th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2021, pp. 692–697.

[67] G. Lindner, S. Shi, S. Vučetić, and S. Mišković, "Transfer learning for radioactive particle tracking," *Chem. Eng. Sci.*, vol. 248, Feb. 2022, Art. no. 117190.

[68] W. W. Tso, B. Burnak, and E. N. Pistikopoulos, "HY-POP: Hyperparameter optimization of machine learning models through parametric programming," *Comput. Chem. Eng.*, vol. 139, Aug. 2020, Art. no. 106902.

[69] X. Yan, Y. A. Sekercioglu, and N. Mani, "A method for minimizing unnecessary handovers in heterogeneous wireless networks," in *Proc. Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2008, pp. 1–5.

[70] R. Hussain, S. A. Malik, S. Abrar, R. A. Riaz, H. Ahmed, and S. A. Khan, "Vertical handover necessity estimation based on a new dwell time prediction model for minimizing unnecessary handovers to a WLAN cell," *Wireless Pers. Commun.*, vol. 71, no. 2, pp. 1217–1230, Jul. 2013.

[71] M. Bakshi, B. Jaumard, and L. Narayanan, "Optimal aggregated ConvergeCast scheduling with an SINR interference model," in *Proc. IEEE 13th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2017, pp. 1–8.

[72] D. Qu, Y. Zhou, L. Tian, and J. Shi, "User-centric QoS-aware interference coordination for ultra dense cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[73] A. R. Elsherif, W.-P. Chen, A. Ito, and Z. Ding, "Adaptive resource allocation for interference management in small cell networks," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2107–2125, Jun. 2015.

[74] G. R. Hessler, A. W. Eckford, and P. J. Thomas, "Linear noise approximation of intensity-driven signal transduction channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[75] M. Javad-Kalbasi, Z. Naghsh, M. Mehrjoo, and S. Valaee, "A new heuristic algorithm for energy and spectrum efficient user association in 5G heterogeneous networks," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–7.

[76] M. J. F. Alenazi, A. Almutairi, S. Almowuena, A. Wadood, and E. K. Cetinkaya, "NFV provisioning in large-scale distributed networks with minimum delay," *IEEE Access*, vol. 8, pp. 151753–151763, 2020.

[77] I. A. Alablani and M. A. Arafah, "EE-UWSNs: A joint energy-efficient MAC and routing protocol for underwater sensor networks," *J. Mar. Sci. Eng.*, vol. 10, no. 4, p. 488, Apr. 2022.

[78] O. A. Egaji, A. Griffiths, and M. S. Hasan, "Delay optimisation for multimedia application in a wireless network control system (WNCS)," in *Proc. 11th IEEE Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Nov. 2020, pp. 267–275.

[79] I. Alablani and M. Alenazi, "EDTD-SC: An IoT sensor deployment strategy for smart cities," *Sensors*, vol. 20, no. 24, p. 7191, Dec. 2020.

[80] I. A. Alablani and M. A. Arafah, "Applying a dwell time-based 5G V2X cell selection strategy in the city of Los Angeles, California," *IEEE Access*, vol. 9, pp. 153909–153925, 2021.
[81] Y. Zhong, M. Haenggi, F.-C. Zheng, W. Zhang, T. Q. S. Quek, and W. Nie, "Toward a tractable delay analysis in ultra-dense networks," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 103–109, Dec. 2017.
[82] H. Chen, Q. Chen, R. Chai, and D. Zhao, "Utility function optimization based joint user association and content placement in heterogeneous networks," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2017, pp. 1–6.
[83] Z.-H. Huang, Y.-L. Hsu, P.-K. Chang, and M.-J. Tsai, "Efficient handover algorithm in 5G networks using deep learning," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.
[84] X. Ba and Y. Wang, "Load-aware cell select scheme for multi-connectivity in intra-frequency 5G ultra dense network," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 354–357, Feb. 2019.
[85] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sep. 2017.
[86] S. Singh, D. Kedia, N. Rastogi, T. Velmurugan, and P. Prakasam, "P2P mobility management for seamless handover using D2D communication in B5G wireless technology," *Peer-Peer Netw. Appl.*, vol. 14, no. 4, pp. 1–10, 2021.
[87] H. Park, Y. Lee, T. Kim, B. Kim, and J. Lee, "ZEUS: Handover algorithm for 5G to achieve zero handover failure," *ETRI J.*, vol. 44, no. 3, pp. 361–378, Jun. 2022.
[88] A. Alhammadi, M. Roslee, M. Y. Alias, I. Shayea, and A. Alquhali, "Velocity-aware handover self-optimization management for next generation networks," *Appl. Sci.*, vol. 10, no. 4, p. 1354, Feb. 2020.

**IBTIHAL AHMED ALABLANI** received the B.S. degree in computer science from Princess Nourah Bint Abdulrahman University, in 2008, and the M.S. and Ph.D. degrees in computer engineering from King Saud University, in 2013 and 2022, respectively. She is currently an Assistant Professor with the Computer Technology Department, Technical Digital College, Technical and Vocational Training Corporation (TVTC). Her research interests include wireless sensor networks, 5G heterogeneous networks, smart cities, and machine learning. She received the Certificate of Excellence from the Saudi Ministry of Education as one of the top students in Saudi Arabia at high school, in 2004.

**MOHAMMED AMER ARAFAH** received the B.S. degree from King Saud University and the M.S. and Ph.D. degrees from the University of Southern California, USA. He is an Associate Professor in computer engineering. His areas of research interests include computer network modeling and simulation, wireless sensor networks, cooperative relay networks, fault tolerance, 5G heterogeneous networks, and high-speed networks.