Received 12 June 2023, accepted 25 June 2023, date of publication 28 June 2023, date of current version 3 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3290482

RESEARCH ARTICLE

A DL-Enabled Relay Node Placement and Selection Framework in Multicellular Networks

IOANNIS A. BARTSIOKAS[1], PANAGIOTIS K. GKONIS[2], DIMITRA I. KAKLAMANI[1], AND IAKOVOS S. VENIERIS[3]

[1]Microwave and Fiber Optics Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Zografou, 15780 Athens, Greece

[2]Department of Digital Industry Technologies, National and Kapodistrian University of Athens, Dirfies Messapies, 34400 Athens, Greece

[3]Intelligent Communications and Broadband Networks Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Zografou, 15780 Athens, Greece

Corresponding author: Panagiotis K. Gkonis (pgkonis@dind.uoa.gr)

This work was supported in part by the Hellenic Academic Libraries (HEAL-Link).

ABSTRACT The ever-increasing and diverse user demands as well as the need for uninterrupted access to the medium with minimum latency in dense machine type communication networks, are the key driving forces to a holistic network redesign. In this context, fifth-generation and beyond (5G/B5G) networks, incorporate various advanced physical layer techniques, such as relaying-assisted transmission, aiming to improve network performance and extend the coverage area of multicellular orientations. However, the deployment of such techniques in a cellular environment characterized by high interference levels and multi-variate channel representations, leads to increased computational complexity for radio resource management (RRM) tasks. Machine learning (ML), and especially Deep Learning (DL), is proposed as an efficient way to support end-to-end user applications in highly complex environments, since ML/DL models can relax the RRM-associated computational burden. In this paper, we consider the joint problem of relay node (RN) placement and selection subject to subcarrier allocation and power management constraints in 5G/B5G networks. Various DL-based methods are examined and combined to solve both sub-problems. The performance of these schemes is evaluated for various relaying-assisted transmission approaches, either considering known channel state information (CSI) or not. According to the derived results, total system energy efficiency (EE) and spectral efficiency (SE) can be improved by up to 30%, when considering only the DL-based RN placement scheme compared to state-of-the-art non-ML schemes. The deployment of the reinforcement learning (RL) model for RN selection, can improve EE up to 80%, while SE can be improved up to 75%, compared to a system with only DL-enabled RN placement.

INDEX TERMS Relay assisted transmission, machine learning, deep learning, Q-learning, 5G networks, system level simulations.

ACRONYMS

3D	Three-dimensional.
5G	Fifth Generation.
A&F	Amplify-And-Forward.
AI	Artificial Intelligence.
ANN	Artificial Neural Networks.
AR	Augmented Reality.
B5G	Beyond Fifth Generation.
BER	Bit Error Rate.
BS	Base Station.
CSI	Channel State Information.
D&F	Decode-And-Forward.
DL	Deep Learning.
DNN	Deep Neural Networks.
DQL	Deep Q-Learning.
EE	Energy Efficiency.
FEC	Forward Error Correction.
FNN	Feed-Forward Neural Network.

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

HetNet	Heterogenous Network.
IoT	Internet Of Things.
KPI	Key Performance Indicator.
L3	Layer 3.
LOS	Line of Sight.
LSTM	Long-Short Memory Network.
MANET	Mobile Ad Hoc Network.
MAPE	Mean Absolute Percentage Error.
MaC	Macro Cell.
MC	Monte Carlo.
ML	Machine Learning.
m-MIMO	Massive Multi-Input-Multiple-Output.
mmWave	Millimeter Wave.
mMTC	Massive Machine Type Communications.
MRC	Maximal Ration Combining.
MSE	Mean Squared Error.
NOMA	Non-Orthogonal Multiple Access.
OFDMA	Orthogonal Frequency Division Multiple Access.
OP	Outage Probability.
PA	Power Allocation.
QoE	Quality of Experience.
QoS	Quality of Service.
RA	Resource Allocation.
RB	Resource Block.
ReLU	Rectified Linear Unit.
RL	Reinforcement Learning.
RN	Relay Node.
RNN	Recurrent Neural Network.
RRM	Radio Resource Management.
SC	Small Cell.
SE	Spectral Efficiency.
SER	Symbol Error Rate.
SNIR	Signal-To-Noise-Plus-Interference-Ratio.
TD	Temporal Difference.
UAV	Unmanned Aerial Vehicle.
UE	User Equipment.
URLLC	Ultra-Reliable-Low-Latency Communication.
UxV	Unmanned ground, air, surface or undersea Vehicle.
VR	Virtual Reality.
WANET	Wireless Ad Hoc Network.
WSN	Wireless Sensor Network.

I. INTRODUCTION

In recent years, fifth-generation (5G) and beyond (B5G) wireless communications systems have been established to support the exponential growth rate of mobile data traffic [1]. Moreover, the rapid evolution of wireless services, such as high-definition video streaming, Internet of Things (IoT) applications, augmented/virtual reality (AR/VR), wireless or mobile ad-hoc networks (WANETs/MANETs) and unmanned aerial vehicles (UAVs), drove 5G/B5G standardization process to deal with different telecommunication service categories, such as ultra-reliable low-latency-communications (URLLC), enhanced mobile broadband (eMBB) and massive machine type communications (mMTC) in mass access environments [2], [3]. In this context, various novel physical layer technologies have been introduced over the last years to cope with the increasing challenges in the wireless communications domain, such as massive multi-input- multiple-output (m-MIMO) configurations, millimeter Wave (mmWave) transmission, as well as non-orthogonal multiple access (NOMA) [4]. However, the aforementioned advanced physical layer technologies, when applied in a cellular environment characterized by high interference levels and complex channel approximations, along with increased connection density and near-random user mobility patterns, maximize the computational cost to support strict users' requirements and demands. Machine learning (ML) algorithms are proposed as an efficient way to tackle these considerations, due to their ability to utilize data generated by the network itself in improving network performance and efficiency [5], [6]. ML algorithms are trained using either data generated by the wireless network under test or by similar ones. In this way, complex channel calculations are encapsulated in ML models' layers, which leads to the decrease of computational cost and complexity after successful training rounds. Moreover, some ML algorithms (e.g., Reinforcement Learning (RL) ones), can directly interact in real-time and support low-latency requirements of modern era networks. It is important to highlight that when applying ML models in physical layer optimization tasks, both ML metrics and network metrics (such as Quality of Experience (QoE), energy efficiency (EE), spectral efficiency (SE), user fairness, achieved throughput, active users and blocking probability, etc.) should be jointly examined and evaluated [5].

The 5G/B5G network architecture is based on the heterogenous networks (HetNets) model in order to meet the increased network capacity and ultra-density requirements. HetNets involve the composition of a number of smaller, simpler, and lower-power base stations (BSs), with different characteristics (transmission capacities, coverage areas, carrier frequencies, etc.) to improve cell-edge coverage and enhance the network key performance indicators (KPIs) [7]. Several recent research efforts have been conducted on HetNets' performance evaluation under various radio resource management (RRM)-related tasks. Indicatively, Beshley *et. al.* in [8], proposed a QoE-enabled RRM technique for 5G/B5G multi-layer networks to reduce energy consumption in such topologies, while distributing the service process between Macro and small cells (MaCs, SCs). In fact, by switching SCs to sleep mode when no active users are served, radio usage efficiency is improved by 25% and energy consumption by 8.7% compared to a system where SCs are constantly operating. On the same context, authors in [9] propose a delay-aware RRM scheme to guarantee user fairness, minimize losses and provide low-latency services in eMBB 5G HetNet scenarios, taking also into consideration Quality of

Service (QoS) levels maintenance. The proposed algorithm has three main phases, which are summarized as follows: a) delay-based resource block (RB) matrix formulation, b) flow prioritization based on a greedy algorithm, c) data rate calculation and adjustment based on channel state information (CSI) and available data to be scheduled. System's performance is evaluated with extensive system level simulations, which indicate that the proposed solution achieves a 4-5% throughput gain, a ∼65% delay decrease and a guaranteed user fairness compared to other state-of-the-art approaches. Similar conclusions have been presented by authors in [10], where a secure RRM optimization problem for HetNets is formulated, with the objective to maximize the achieved throughput for both cell center and edge users. The proposed Utility-based Resource Scheduling Algorithm shares resources between least-delay users of each application, aiming to maintain fairness and reduce cross layer interference for real and non-real time applications. Performance evaluation indicates that the adopted approach overperforms other state-of-the-art ones in terms of achieved throughput, user fairness and SE.

The indicative research efforts presented in [8], [9], and [10], are focused on the direct communication link between a BS (either this is a Macro, micro or even smaller one) and the user equipments (UEs). However, there are cases even in HetNet topologies, where the coverage area of each cell should be extended for more users to be served by the network. Such scenarios are of significant interest in Unmanned Vehicles (UxVs) scenarios or military/defense networks. In this context, relay nodes (RNs) have been proposed as a "retransmission technology", that can relax transmission burden from centralized BSs. Thus, by using RN-assisted RRM mechanisms, total system's performance can be upgraded by further improving data rates, mitigating interference levels and extending network coverage [11].

An open research field concerning the use of RNs in multicellular 5G/B5G networks is the optimal relay placement within each cell's area, to improve various network KPIs, such as total served users, achieved throughput and signal-to-noise-plus-interference-ratio (SNIR), total transmitting power and blocking probability. Moreover, the selection of the optimal RN, from the candidate ones, to serve each active UE in the topology is, also, of the same interest.

Compared to an established one-hop downlink communication link between a BS (transmitter) and a UE (receiver), an n-hop relaying-assisted BS-UE link's complexity is increased due to the following reasons:

(a) The use of RNs was introduced in 3GPP release 16 [12], while the beginning of standardization process can be found in release 17 (latest stable edition of 3GPP documents) [13]. This means that there are no detailed channel, pathloss and mobility models for RNs.
(b) The effectiveness of the RN-UE connection is based, also, on the quality and stability of the BS-RN link.
(c) Shared resource management should be performed, as RNs are a layer 3 (L3) entity, which needs BSs' assistance in performing advanced RRM tasks.

Recently, ML is proposed as an efficient approach to deal with the abovementioned problems of optimal RN placement and optimal RN selection between candidate RNs [14]. The key characteristic of ML-based approaches is that -using data generated by existing systems-, they can accurately estimate the examined system's behavior, with the minimum computational cost. In this way, complexity is reduced, and accurate predictions can be performed leading to real-time responses.

In this paper, we focus on solving the joint problem of RN placement and selection by utilizing different ML-based techniques, focusing on Deep Learning (DL) and RL. Our contributions are the following:

- We first formulate the problem of RN placement to maximize the number of active users in each cell of the cellular topology. Thus, given only the number of the RNs per cell to be deployed and a set of potential geographical positions (x-y coordinates of potential RNs), the k best-performing RNs are selected to serve the active users in each cell. The aforementioned selection is performed subject to three main constraints. The first one is the minimization of pathloss for each accepted user, the second refers to the minimization of the total transmitted power by each deployed RN, while the latter is the maximization of the total accepted users in the topology. Moreover, we test our algorithm in two different simulation scenarios. The first one considers the presence of ideal CSI, while the latter considers no CSI at all.
- To tackle the aforementioned problem, we propose two efficient offline RN placement ML/DL algorithms which focus on fast response times taking into consideration the constraints previously presented. The fist DL algorithm considers a Deep Artificial Neural Network (ANN) orientation, while the latter considers a Long-Short Memory Network (LSTM) one.
- After the optimal placement of the RNs in each cell's coverage area, we formulate the problem of optimal RN selection for each accepted user in the topology. In other words, for users not served directly by the BSs, either for pathloss or power consumption reasons, the optimal RN (from the k eligible) should be selected to serve them.
- To solve the aforementioned problem, we propose an energy efficient RL-based algorithm to select the optimal beam (RN) to serve each accepted user in the topology. A Deep Q-Learning (DQL) RL algorithm is utilized for this scope. In this context, EE and SE are the KPIs that determine algorithm's transitions.
- Finally, all presented approaches are evaluated by extensive system level simulations in different usage scenarios. Performance evaluation indicates that the joint DL-based RN placement and selection scheme can overperform state-of-the-art approaches in improving various network KPIs, such as EE and SE.

- To sum up, the utilization of DL/DRL schemes, both for efficient RN placement in each cell of the cellular topology and for RN selection, which forms a full ML/DL-assisted RRM framework focusing on both EE and SE, is the key novelty of this paper.

The rest of the paper is organized as follows. In Section II, we provide a brief overview of the existing literature on the field of RN placement and/or selection in 5G/B5G HetNets. In Section III, the 5G/B5G system model is introduced. Moreover, the joint NP-Hard relay placement and selection problem is mathematically formulated as well. Thus, after the optimal placement of the required number of RNs in each cell's coverage area, each accepted user -which cannot be served by the primary BSs system- is associated with the RN that maximizes EE and SE. In Section IV, the proposed DL-based algorithm for RN placement with the corresponding deep neural networks (DNNs) structures are presented. The performance of these algorithms is evaluated assuming both guaranteed knowledge of CSI and not. In Section V, the DQL algorithm for RN selection (with the corresponding state and transition tables), is presented. Hense, the joint RN placement and selection problem is completely formulated. In Section VI, the performance of the proposed ML/DL-based algorithms is evaluated using a MATLAB 5G/B5G link and system level network simulator that has been developed in our lab. Finally, in Section VII, concluding remarks are presented.

II. ML-BASED RELATED WORKS

Algorithms for RN placement and selection is an active area of research in wireless communications, especially in 5G/B5G cellular communication networks. In fact, the implementation of relaying-assisted communications is proposed in different usage scenarios in modern era wireless systems, such as MANET/WANET networks, supply chain management and manufacturing. Moreover, the joint utilization of RN-assisted communication and modern multiple access schemes, e.g., NOMA, are also of high research interest nowadays. In these cases, the NP-Hard optimization problem [14] of RN placement and/or selection is solved through either distance-based techniques with the use of graph theory or game theory [15], [16], or via extensive search algorithms (e.g., using ergodic capacity analysis) [17]. Furthermore, moving RNs are, also, under research, due to the growing interest in UAV communications [18].

In this framework, we investigate ML-based solutions as an efficient way to deal with the NP-Hardness of the aforementioned problem. In this Section, current research activities on the field of ML-assisted RN placement and selection are presented. The concept of ML-assisted RN placement is introduced in [19]. Authors presented an optimal RN positioning method, aiming to improve system's performance in uncertain and dynamic-changing multicellular topologies. Consequently, channel quality prediction in both BS-RN and RN-UE link is of primary concern in defining RN positioning. Authors proposed a learning-based and a distance-based method for channel prediction based on mobility patterns of RN under test. The achieved connectivity levels are used as the basic KPI during performance evaluation. A heuristic optimization algorithm is used for optimal RN positioning, outperforming a recently developed relay positioning algorithm.

In [14], an approach for deploying the minimum accepted number of RNs -as a subset of given potential locations- is considered with respect to QoS requirements in multi-hop wireless systems. A hop count boundary is inserted to ensure a certain blocking probability in the BS-RN link. To deal with the NP-Hardness of the RN placement problem, a polynomial time approximation algorithm using shortest path trees and heuristically pruning the relay nodes used until the hop count bound is violated, is proposed. Performance evaluation indicated that this approach efficiently solves the above-mentioned problem in various randomly generated network scenarios. More specifically, optimal solutions are given in over 90% of the tested scenarios. Afterwards, authors used random graph techniques to derive an upper bound on the average case approximation ratio for the used algorithms based on the number of source nodes, and the hop count bound. This average case analysis was the first one in RN placement literature.

On the other hand, authors in [20] face RN placement problem as a clustering one. The scenarios of interest consider Wireless Sensor Networks (WSNs), where RNs are used as mediators between users and applications' servers by assisting messages transmission. A k-means clustering ML approach is activated for link restoration whenever it is necessary based on transmitting power, number of packets lost in a RN-BS link and BS-UE distance. Thus, for each BS, the corresponding RN is deployed at the most frequently used route in the network. Numerical evaluation indicated that the proposed method outperforms existing distance-based methods on the basis of various KPIs such as residual energy, end-to-end delay and the number of hops required in the network from source to destination. Moreover, k-means clustering algorithm implementation can reduce the total number of used RNs.

Considering industrial WSNs, authors in [21] studied the placement of RNs in a realistic three-dimensional (3-D) factory space based on the satisfaction of various physical, performance and energy-related KPIs. The study was performed using IEEE 802.15.4e low latency deterministic network mode in order to achieve low latency and highly reliable communications in harsh factory environments, which are suffering from noise, interference and multipath fading. Hense, frequent packet losses are reported. The authors proposed the joint incorporation of RN nodes and forward error correction (FEC) techniques leading to enhanced communication reliability. More specifically, on the one hand an efficient and pragmatic relay-placement strategy based on rainbow product ranking algorithm for a 3-D factory space, and on the other hand an adaptive RL transmission scheme (using Q-learning techniques), which incorporates cooperative diversity and Reed Solomon block codes, are proposed.

A real-world case study is performed in order to evaluate the correctness and effectiveness of the presented solution. The proposed RN placement strategy has improved performance in terms of cost reduction and total number of deployed RNs compared to other state-of-the-art approaches. Moreover, the used Q-learning method efficiently utilized the resources in terms of relays and BSs, making the transmission scheme more generic in terms of, not only adopting to versatile factory environments but also accommodating the dynamic behavior per link in the factory space.

A more complex scheme considering not only RN localization but also power management in 5G networks is presented in [22]. A mathematical analysis for defining expressions and minimum threshold for end-to-end average symbol error rate (SER) and outage probability (OP) is performed for in amplify-and-forward (A&F) RNs, introducing those KPIs as the main ones for the problem definition. As a result, the high correlation between BS-RN and RN-UE links, is described. Afterwards, the joint power allocation (PA) and RN placement problem is considered. Results indicated that RN placement optimization is more efficient than PA. Thus, an ML implementation of the proposed convex optimization problem is investigated. The joint problem is translated to a regression ML problem and authors propose a feed-forward neural network (FNN) approach (2 and 3 hidden layers are considered). ML-models' performance is controlled using the mean absolute percentage error (MAPE) metric which reached over 90% score. The simulation results demonstrated a compromise between MAPE and computation times for the FNN-based joint PA-RL optimization.

Authors in [23] proposed a combined RN selection and resource allocation (RA) algorithm. A key drawback in existing approaches is the need for a large number of relays to forward signals transmitted on multiple subcarriers. However, signal generation in multi-hop scenarios increases the complexity of combined RN selection compared to that of per-subcarrier relay selection, when the number of relays increases. In dense 5G networks, the impact of that problem is even bigger. To deal with this drawback, authors proposed a supervised ML method. The training phase is implemented off-line, leading to a considerable reduction to the RN selection complexity and the processing latency. An ANN scheme is used for the best couple of relays to be selected. In each epoch the least accuracy criterion is checked to continue simulations. Accuracy and Mean Squared Error (MSE) are the two considered KPIs for ANN's performance. Numerical evaluation indicated that the proposed supervised ML approach can provide near-optimal performance with lower computing latency, which nearly reaches the optimal relay selection in a per-subcarrier manner.

Over the last years, RL-assisted RN selection and RA have attracted scientific research interest as well. In this context, Geng et al. in [24] studied the joint outage probability minimization, RN selection optimization and transmission power reduction problem in RN-assisted 5G networks, where the existence of accurate CSI is extremely difficult. Thus, the authors proposed an RL prioritized experience replay aided framework, acting in optimal solution finding to the above-mentioned problem without any prior knowledge of CSI. The proposed approach is compared to other RL-based solutions, and performance evaluation indicated that communication success rate can be improved by about 5%. On the same context, the authors in [25] proposed an RN selection algorithm to succeed in providing guaranteed reliability, low latency, and power consumption levels in large scale multi-hop 5G topologies. The proposed scheme uses Q-learning RN selection based on SNIR levels. Q-learning is an RL approach, which consists of an agent, the environment, agent's states, actions, as well as rewards or penalties. In learning stage, the agent learns the optimal allocation policy to maximize the reward [26]. From network's perspective, the BS knows the optimal RN to select and transmit the signal. The used RNs are decode-and-forward (D&F) ones, and the system uses orthogonal frequency division multiple access (OFDMA) techniques. Finally, the proposed scheme tries to utilize optimal allocation policy based on the learning outcomes of the previous stage, based on the SNIR. Performance evaluation indicated that the proposed approach achieves the same bit error rate (BER) levels as conventional RN selection schemes in the literature. The basic advantage of the proposed approach is the selection of fewer RNs when the target BER is satisfied. Consequently, system's latency is improved.

The aforementioned research efforts describe some aspects of the utilization of Artificial Intelligence (AI)/ML methods for the optimization of either the problem of RN placement or RN selection -over a set of available RNs- in wireless systems. However, most of these works evaluate the proposed algorithms in single-cell orientations, or by limiting the number of active UEs in the topology. Our motivation is to extend these works and present a global ML-based framework to both train an offline ML model to place the RNs based on the performance of simulated UEs in the topology, but also, propose an RL method to interact with the cellular environment and select the best-performing RN for each accepted UE in the topology.

III. SYSTEM MODEL

In this section, the studied system model under investigation is presented. Subsection A refers to the overall system description, subsection B refers to the formulation of the RN placement problem, while subsection C analyses the RN selection problem.

A. SYSTEM OVERVIEW

The downlink of a cooperative wireless OFDMA 5G/B5G multicellular HetNet is considered, as illustrated in Fig. 1. The studied system has two different levels of base entities. The first one, Macro-BSs, forms the primary system where UEs can directly access and request service. The latter one, RNs, form the secondary system, aiming to assist the primary system in improving capacity and coverage area, by serving UEs that have been initially rejected by the primary system.

We should state that in our analysis and simulations A&F RNs are considered.

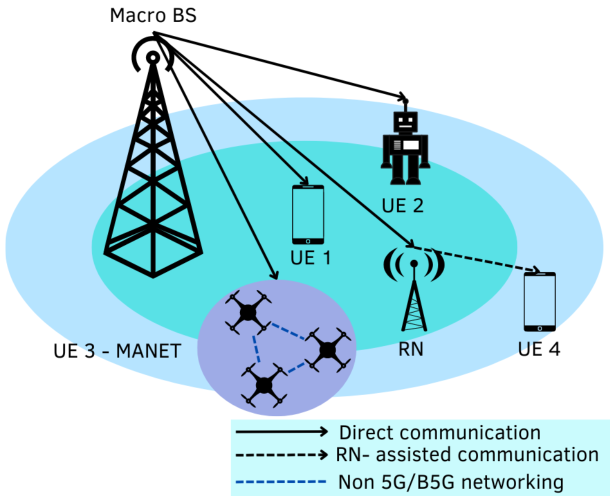


FIGURE 1. Two-hop 5G/B5G HetNet with A&F RNs.

Thus, the cooperative system consists of M BSs, R RNs and N uniformly distributed UEs. The set of BSs is denoted as $S_b = \{b_1, b_2, \ldots, b_M\}$, the set of RNs is denoted as $S_r = \{r_1, r_2, \ldots, r_R\}$, while the set of UEs is denoted as $S_u = \{u_1, u_2, \ldots, u_N\}$, respectively. The potencial link between a BS and a UE is denoted as $L_{b,u}$ where $b \in S_b$ and $u \in S_u$, the potential link between a BS and a RN is denoted as $L_{b,r}$ where $b \in S_b$ and $r \in S_r$, while the potential link between a RN and a UE is denoted as $L_{r,u}$ where $r \in S_r$ and $u \in S_u$. Note that the Cartesian coordinate system is used to locate all enrolled entities both considering 2D or 3D space.

B. RN PLACEMENT PROBLEM FORMULATION

As previously stated, RNs are deployed to assist the primary communication's system to support UEs that are initially rejected due to high pathloss or other power allocation reasons. The goal of this sub-problem is to select the N_{CRN} positions (set of x-y-z coordinates) for the best-performing RNs to be deployed in each cell's coverage area. Best performing RNs are selected to optimally meet user requirements and maximize each cell's performance. Thus, a predefined number of potential RNs are placed in different positions inside each cell's coverage area, declared as RN_{can}. Fig. 2 provides an illustration of such a topology for a single 5G/B5G cell where $RN_{can} = 10$ candidate RNs deployed.

Thus, the N_{CRN} best-performing RNs are selected out of S_{rc}, where $S_{rc} = \{rc_1, rc_2, \ldots, rc_{\mathrm{RN}_{can}}\}$ is the set of candidate RNs is each cell, subject to the following constraints:

- (C1): $min(PL_n), \forall n \in N$, where PL_n is the pathlos between each accepted UE by the secondary sytem, and the RN that is assigned to.
- (C2): $min(P_{t,r}), \forall r \in R$, where $P_{t,r}$ is the total transmitted power by each deployed RN.

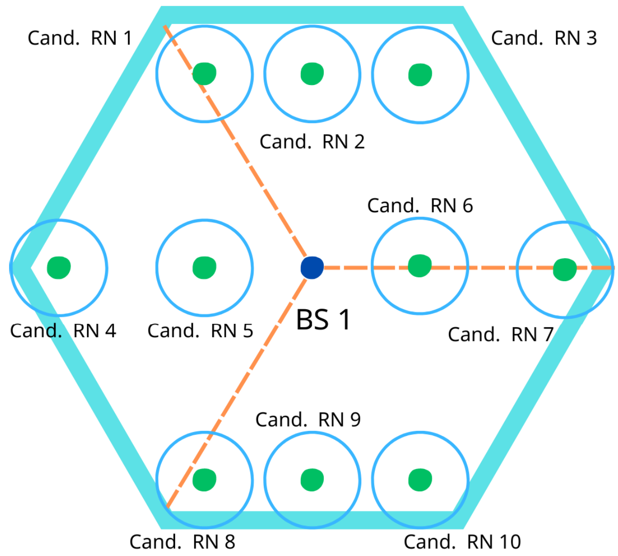


FIGURE 2. 5G/B5G system's cell with candidate RNs.

- (C3): $max(AN_r), \forall r \in R$, where AN_r denotes the total accepted UEs served by RNs.

Two offline ML-enabled methods are proposed to solve the aforementioned problem utilizing DL principles and techniques, as presented in Section IV. Moreover, both methods are examined either under known CSI/channel conditions or under fully unknown CSI. These aspects are also discussed in Section IV.

C. RN SELECTION PROBLEM FORMULATION

As shown in Fig. 3, for each cell of the cellular topology, there is an M_t antenna source -which is located at the BS of each cell, $N_{u_{RN}}$ UEs -where $N_{u_{RN}} \leq N$ are the initially rejected UEs from the primary system, that request RN assistance, equipped with M_r antennas, and N_{CRN} RNs in the two-hop wireless relay network. As it was previously described, the direct link between source and destinations does not exist due to high pathloss effect. Therefore, A&F relays are used to process the received signal and support communication. Each UE is connected only to one RN and orthogonal channels are used to achieve full set gain and mitigate co-channel interference. This sub-problem's goal is to optimally select the most suitable RN out of N_{CRN} candidate ones for each UE $n \in N_{u_{RN}}$, with respect to the active user maximization for each cell of the topology.

In a two-hop 5G/B5G wireless communications system, like the one depicted in Fig. 1, the total bandwidth, BW, is divided into N_{sc} subcarriers to be allocated to the accepted UEs. There are two classes of accepted UEs. On the one hand, the first class contains UEs that are directly accepted by the primary system (BS-UE direct communication). The SNIR for the n^{th} UE ($1 \leq n \leq N$) of this category, associated with the l^{th} subcarrier ($1 \leq l \leq N_{sc}$) for a specific channel realization and assuming independent BS-UE links,

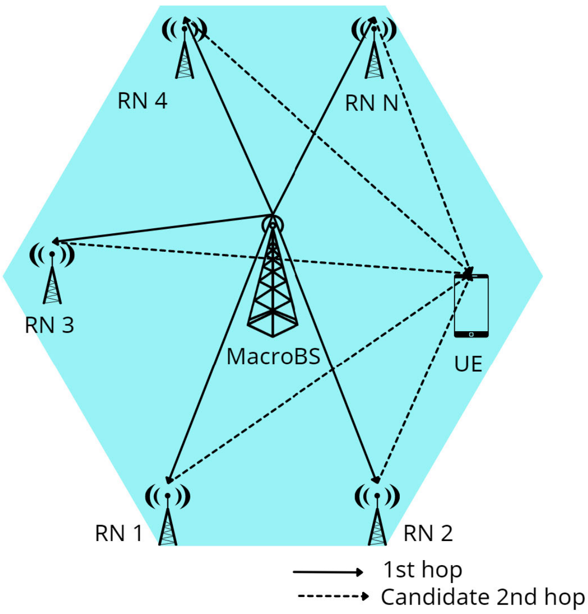


FIGURE 3. Two-hop 5G/B5G Relay Selection.

is denoted as follows [27]:

$$SNIR_{n,l} = \frac{G_{n,n,l}}{\mathbf{r}_{n,l}{}^{H}\mathbf{r}_{n,l}I_0 + \sum_{m\neq n,l\in S_m} G_{n,m,l}} \tag{1}$$

where $G_{n,m,l} = p_{n,l}\mathbf{t}_{m,l}{}^{H}\mathbf{H}_{n,sec(m),l}{}^{H}\mathbf{r}_{n,l}{}^{H}\mathbf{r}_{n,l}\mathbf{H}_{n,sec(m),l}\mathbf{t}_{m,l}$, $\mathbf{H}_{n,sec(n),l}$ represents the $M_r \times M_t$ channel matrix (flat Rayleigh fading) for the l^{th} subcarrier of the n^{th} UE relevant to its serving sector $sen(n)$, $\mathbf{t}_{n,l}$ is the $M_t \times 1$ transmission vector in diversity combining transmission mode, $\mathbf{r}_{n,l}$ is the Maximal Ratio Combing (MRC) multiplying vector [28] and $p_{n,l}$ denotes the transmission power allocated to the l^{th} subcarrier of the n^{th} UE. Moreover, the set S_n indicates the subcarriers allocated to the n^{th} UE and I_0 is the thermal noise level. Finally, $\boldsymbol{A}^H$ denotes the conjugate transpose of matrix $\mathbf{A}$.

On the other hand, the second class contains UEs that are directly connected to the secondary system (RN-UE connection). Thus, through relaying, BS-UE communication link is established through multi-hop communication. In this two-hop connection between BSs and UEs, RNs can be defined as UEs in the BS-RN link, and as BSs in the RN-UE link. In this case, for the n^{th} UE ($1 \leq n \leq N$) of this category equation (1) is modified as follows:

$$SNIR_{n,l(RN)} = \frac{G_{n,n,l(RN\text{-}UE)}}{\mathbf{r}_{n,l}{}^{H}\mathbf{r}_{n,l}I_0 + I_{BSn,l} + I_{RNn,l}} \tag{2}$$

where $I_{BSn,l} = \sum_{b=1}^{N_{BS}} \sum_{m\in UE_b, l\in S_m} G_{n,m,l}$ and $I_{RNn,l} = \sum_{r=1}^{N_{RN}} \sum_{j\in UE_r, l\in S_j} G_{n,j,l}$ are the cumulative interference levels for the l^{th} subcarrier of n^{th} UE served by the b^{th} BS or r^{th} RN. Moreover, N_{BS}, N_{RN} are the total number of BSs and RNs in the topology, respectively, UE_r denotes the set of UEs served by the r^{th} RN, while the notation x-y indicates all possible link connections.

Total system throughput is given by [29] for the whole two-hop wireless communication 5G/B5G system:

$$\begin{aligned} R &= \sum_{n=1}^{N}\sum_{s\in S_n} r_{n,s} \\ &= B_{\text{SC}}\{\sum_{b=1}^{N_{BS}}\sum_{n\in UE_b}\sum_{s\in S_n} log_2(1+SNIR_{n,s\,(BS)}) \\ &\quad + \sum_{r=1}^{N_{RN}}\sum_{m\in UE_r}\sum_{s\in S_m} log_2(1+SNIR_{m,s\,(RN)})\} \end{aligned} \tag{3}$$

where $|S_n|$ indicates the length of the set S_n, $r_{n,s}$ is the corresponding data rate for the s^{th} subcarrier and B_{SC} is the bandwidth per subcarrier. Using (3), system's EE is defined as:

$$EE = \frac{R}{\sum_{n=1}^{N}\sum_{s\in S_n} p_{n,s}} \tag{4}$$

On the same context, system's SE is defined as:

$$SE = \frac{R}{BW} \tag{5}$$

Therefore, the overall RRM-policy that will be applied to the system targets to maximize EE and SE subject to the following system and power constraints:

- (C1): $\sum_{s\in S_n} p_{n,s} \leq p_m$, where p_m denotes the maximum power limit per UE.
- (C2):$\sum_{n\in UE_b}\sum_{l\in S_n} p_{n,l} \leq P_m$, where P_m denotes the maximum power limit per BS and the set UE_b consists of the UEs that are served by the b^{th} BS.
- (C3):$SNIR_{n,l} \geq SNR_{thr}$, where SNR_{thr} is the SNIR threshold for acceptable QoS.
- (C4): $\sum_{n\in UE_b} |S_n| \leq N_{SC}$, as all BSs have equal access to all available subcarriers.

A DQL scheme is proposed in Section V to solve the aforementioned problem of selecting the suitable RN for each accepted UE of the secondary system. DQL extends the classic frameworks by utilizing ANNs to help software agents to learn how to define actions and rewards. In other words, a DQL framework optimizes underlying function approximation by the use of one (or sometimes two) ANNs to map states and actions to the rewards they lead to. Consequently, it is visible that such an approach can be quickly and dynamically adjustable based on the environment. In our case, Section V proposes a dynamic DQL framework to select the best performing RN -out of the available ones- for each UEs of the secondary (RN-assisted) 5G/B5G system.

IV. DL-BASED RN PLACEMENT

In this Section the proposed DL-based schemes for optimal RN placement are presented. Subsection A refers to the used dataset construction, subsection B refers to the two proposed models (DNNs) structure, layers, and optimization, while subsection C analyzes the performance of both schemes in the following two scenarios:

- *Scenario 1*: The channel coefficient matrix sub-table is known for the link of each UE and the corresponding BS. Moreover, the channel coefficient sub-tables are known, also, for the RN-UE link.
- *Scenario 2*: There is no CSI information available both for the BS-RN and the RN-UE link. Thus, dataset construction and algorithms are based only on geographical, pathloss and topology parameters.

A. DATASET CONSTRUCTION

As expressed in [5], a key procedure for building ML models is the validation and training stage. For this purpose, datasets used for learning objectives should be accurate, up-to-date and should always be evaluated. In this paper, we use a MATLAB 5G/B5G link and system level network simulator to construct our dataset after adequate Monte-Carlo (MC) simulation rounds. This simulator is based on the work in [11], where both different Inband and Outband A&F RN scenarios are considered. The implemented simulator takes into consideration all physical layer aspects such as small and large scale fading, interference management and cluster definition for each user of interest, etc. However, the following improvements have been made:

- The deployment of more RNs per cell has been included. In [11] all scenarios (both Inband and Outband ones) consider the deployment of up to three RNs per cell, mainly deployed in cell edges. In the updated version of the 5G simulator, an increased number of RNs per cell is considered, so that the best performing RN can be selected for each accepted user in an unbiased manner.
- Moreover, channel modelling has been updated according to the newest 3GPP specifications (basically the latest version of 3GPP TS 138 211 regulation) by integrating mobility parameters, existence of Line of Sight (LOS) propagation and outage probability estimation [30].
- The Algorithm 1 of [11] is extended to select the best RN (out of the deployed ones in each cell) based on both minimum pathloss and energy consumption. Hence, this algorithm, which combines MIMO and OFDMA principles in two-hop 5G/B5G cellular orientations, tries to maximize both the EE and SE for the overall orientation under evaluation.

Such a topology is presented in Fig. 2 concerning 10 potential RNs. Thus, by performing numerous MC simulations we store the performance of various UEs, both indoor and outdoor ones, and both moving and static ones. Then, we are focusing only on the UEs that are served by the secondary system. MC simulations are finalized only after adequate UEs' performance is simulated. For the NP-Hard RN placement problem simulating 10.000 to 100.000 UEs defines an adequate number of UEs' performance evaluation. The values that are stored for each UE, which is accepted by the secondary system in the simulation round, concern both location/localization parameters (x, y and z-axis position), serving BS, pathloss, total losses and MIMO parameters (channel coefficient matrix). All related variables and metrics of interest for each simulated UE are presented in Table 1, forming the dataset used for DL-model training.

TABLE 1. Dataset features.

Feature	Description
UE_x	x-axis position of the UE
UE_y	y-axis position of the UE
UE_z	z-axis position of the UE
BS_{serve}	ID of the BS that serves that UE (related to the RN to be selected)
UE_{sec}	Serving Sector of the UE
$UE\text{-}BS_{angle}$	Angle between BS and UE
PL_{mat}	$RN_{can} \times 1$ matrix with the Pathloss between the UE and all the potential RNs
TL_{mat}	$RN_{can} \times 1$ matrix with the total losses between the UE and all the potential RNs
$\mathbf{H}_{matrix}$	$M_r \times M_t$ channel matrix coefficient (used only in *Scenario 1*)
RN_{serve}	ID of the RN that serves the UE **(response variable)**

In order to form the final dataset, the last feature ($\mathbf{H}_{matrix}$) is decomposed into $M_r \times M_t$ different features, each for a specific cell of the $\mathbf{H}_{matrix}$. Thus, the whole dataset feature number is the following:

$$Dataset_{size} = \begin{cases} 6 + 2 \times RN_{can} + M_r \times M_t & Scenario1 \\ 6 + 2 \times RN_{can} & Scenario2 \end{cases} \tag{6}$$

B. DL MODELS STRUCTURE AND FINAL RN PLACEMENT

We propose two DL models to predict the best performing RN for the UEs in the constructed dataset. We should note that for hyperparameter tuning and selection the Exhaustive Grid Search method has been utilized in both models' design [31]. According to that method multiple search rounds are performed over all possible hyperparameter configurations in order for the best performing ones to be selected based on the KPIs of interest (in this case accuracy and F1-score).

The first one is a Deep ANN, also known as DNN network, with the following structure (see also Fig. 4):

- Feature input later with z-score normalization of the input, where $Dataset_{size}$ features are inserted to the DNN.
- A fully connected layer with 50×1 output size, multiplies the input (feature input layer) by the corresponding weight matrix and, also, adds the bias vector.
- A batch normalization layer, to normalize data across all observations for each channel independently, making training of the NN faster through re-centering and re-scaling.
- A (rectified linear unit) ReLU layer, using rectified activation function to force the input directly to the output if it is positive, otherwise, to zero output.

- Another, fully connected layer similar to the previous one with output size $num_{Classes} \times 1$, where $num_{\text{Classes}} = RN_{can}$.
- A soft maximization (sofmax) layer to predict the multinomial probability distribution of the data. These layers are commonly used in multi-class classification tasks, as the one that is examined in this work.
- The classification's output layer, which produces the predicted best-performing RN for each UE.

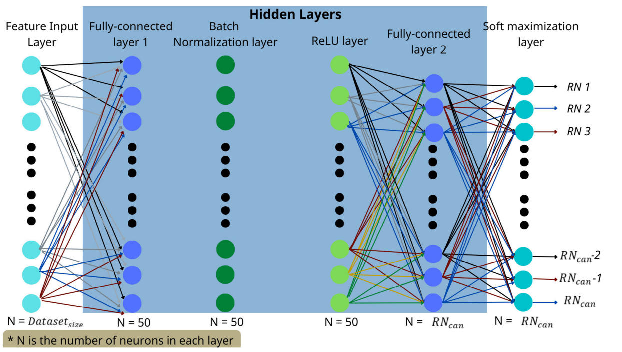


FIGURE 4. Proposed DNN's structure.

The second one is a Recurrent NN (RNN), LSTM, with the following structure (see also Fig. 5):

- Feature input layer with z-score normalization of the input, where $Dataset_{size}$ features are inserted to the DNN.
- An LSTM layer with 52 hidden units. This layer is categorized as an RNN layer, which learns long-term dependencies between data features. Additive interactions between features are used to improve gradient flow over long sequences of data throughout training phase.
- A dropout layer with 0.2 probability to randomly set input elements to zero.
- Another LSTM layer with 40 hidden units.
- Another dropout layer, as the previous one.
- Another LSTM layer with 15 hidden units.
- Another dropout layer, as the previous ones.
- A fully connected layer like the previous one with $num_{Classes} \times 1$, where $num_{Classes} = RN_{can}$.
- A soft maximization layer.
- The classification's output layer, which produces as output the predicted best-performing RN for each UE.

C. PERFORMANCE EVALUATION OF THE PROPOSED MODELS

We consider the downlink of a wireless multicellular 5G orientation, where extensive use of RNs takes place. A 2-tier and 19 cell topology is considered, where UEs are uniformly distributed. Concerning the used MIMO antenna configurations, both BSs, RNs and UEs are equipped with 2 antennas. Each BS's antenna lays at 25 m, while each RN's one lays at 12,5 m and each UE's one at 1,5 m. The relevant antenna gains are 18/9/4 dB for BSs, RNs and UEs respectively. Each

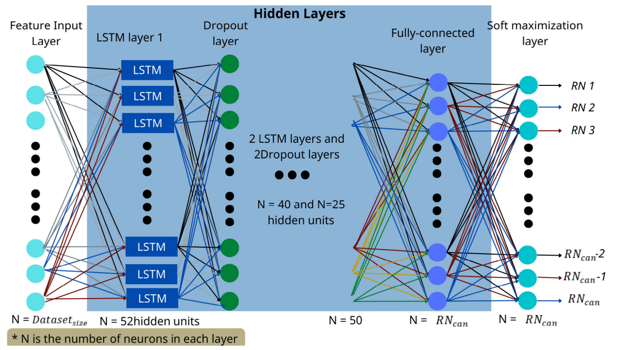


FIGURE 5. Proposed LSTM network's structure.

accepted UE requests 6 subcarriers in each timeslot, while 132 subcarriers are available to be allocated to UEs from each BS. Subcarrier spacing is set to 60 kHz. A significant configuration parameter is the possibility of direct LOS connection between BSs - UEs, BSs- RNs and RNs -UEs. The first two parameters are defined by [30], while the RN-UE LOS existence, which is not regulated, is set to 10%.

We simulate the performance of a large number of UEs (50.000 indoor/outdoor -with 80/20% probability- moving/static UEs) to construct the dataset used for training of our DL models. Moreover, 10 candidate RNs are deployed in each cell's coverage area, as shown in Fig. 2. An adaptive modulation scheme is used based on each UE's demands for QoS levels. Three different modulation levels (QPSK, 16-QAM and 64-QAM) are considered along with their respective threshold values. All simulation parameters are summarized in Table 2.

TABLE 2. Dataset simulation parameters.

Parameter	Value/Assumption
Tiers/Number of cells	2/19
Carrier frequency	28 GHz
Simulated UEs	50.000
Number of antennas per BS/RN/UE	2/2/2
Cell radius	$500\sqrt{3}$=288.68
BS antenna height	25 m
UE antenna height	1.5 m
Indoor to Outdoor UE ratio	80%/20%
LOS BS-UE(RN-UE)/BS-RN Probability	[30] Table 7.4.2-1/10%
Maximum allowed pathloss (dB)	320 dB
Antenna gains BS/RN/UE in dB	18/9/4
Number of requested subcarriers per UE	6
Number of subcarriers per BS or Cluster	132
Subcarrier spacing	60 kHz
SNIR threshold levels QPSK/16-QAM/64-QAM [30]	9.6/16.4/22.7 dB

Using the parameters presented in Table 2, MATLAB 5G/B5G system and link level simulator produces the dataset that is used as input to the two proposed DL models. During the training phase of both DL-based approaches, an 80%-20% training-test set split has been used, as well as a 10-fold cross validation procedure to split the dataset into training, validation and test set. The problem of optimal RN placement is examined as a classification one, by selecting the best performing RN out of the 10 potential RNs for each UE. The performance of the abovementioned classifiers is evaluated, using the accuracy and F1-score metrics. Accuracy is the percentage of the total number of the correct predictions divided by the total number of observations. In other words, accuracy is the sum of True Positive (TP) and True Negative (TN) predictions, divided by the number of the total predictions (TP + TN + False Positive (FP) + False Negative (FN)). Then, Precision, Recall and F1-score are given by the following formulas:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

Tables 3 and 4 summarize the performance of the two proposed DL models in the RN placement based on test set classification accuracy, precision, recall and F1-score for both *Scenario 1* and *Scenario 2*.

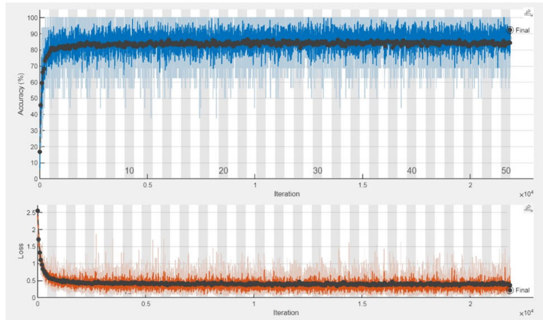
TABLE 3. DNN performance.

DNN	***Scenario 1***	***Scenario 2***
Accuracy	0.9260	0.9387
Precision	0.9251	0.9343
Recall	0.8951	0.9194
F1-score	0.9099	0.9268
Training time	**4 min. 13 s.**	**3 min. 30 s.**

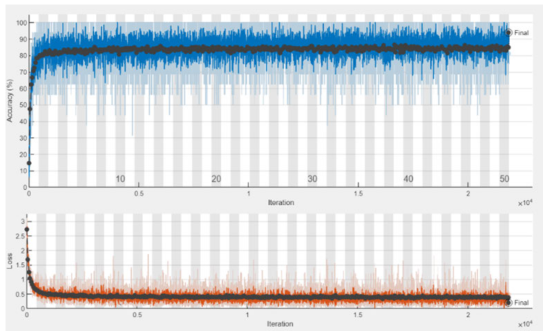
TABLE 4. RNN(LSTM) performance.

LSTM	***Scenario 1***	***Scenario 2***
Accuracy	**0.9513**	**0.9660**
Precision	0.9321	0.9618
Recall	0.9259	0.9502
F1-score	0.9290	0.9560
Training time	4 min. 53 s.	5 min. 12 s.

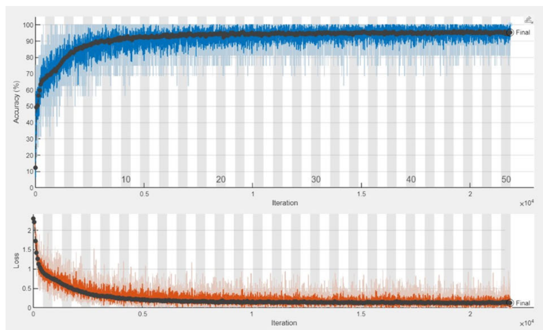
As it can be observed from Tables 3 and 4 LSTM algorithm's performance is better that DNN's performance (both accuracy, precision, recall and F1-score) when CSI is known and is included in training set's features. Similarly, LSTM algorithm's performance is better (both concerning accuracy, precision, recall and F1-score) when there is no CSI knowledge. However, when fast, low-latency responses are considered in 5G/B5G networks, it is vital to examine the trade-off between ML metrics and training time required, as also depicted in [5]. In that perspective, it is visible from Tables 3 and 4 that LSTM networks need some more time to train relative to DNNs. However, comparing training times of these two approaches (LSTM, DNN), we can state that training times are similar in both Scenario 1 and Scenario 2. Thus, it is fair to say, that the overall performance of the LSTM networks is better than DNN, concerning ML/DL performance and metrics-training time trade off. The same conclusions can be drawn from Fig. 6 as well, where the accuracy and loss versus training epochs are displayed both for DNN and LSTM algorithms in the two examined scenarios.



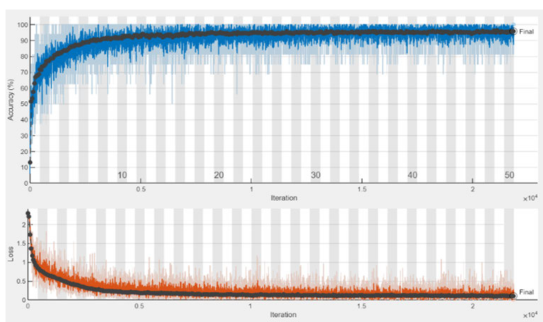
(a) DNN – *Scenario 1*



(b) DNN – *Scenario 2*



(c) RNN – *Scenario 1*



(d) RNN – *Scenario 2*

FIGURE 6. Accuracy and Loss per training iteration and epoch.

After evaluating the two proposed DL models (ANN, RNN) based on the ML classification KPIs, we use these two approaches to identify the *k* best performing RNs out of the 10 candidate active ones placed in each cell of the topology (see Fig. 2). This is achieved using our lab's MATLAB 5G/B5G link and system level network simulator, as follows:

- Simulate the performance of 100.000 UEs in the cellular topology of Fig. 2, configured with the parameters in Table 2.
- Select the best-performing RN for each UE using the two proposed ML/DL models, both in Scenario 1 and Scenario 2.
- Find the k potential RNs serving the most UEs, and thus, assign them as deployed ones. The proposed models are evaluated with k=3 deployed RNs in each cell of the cellular topology.

Table 5 presents the k=3 best-performing RNs out of the 10 potential ones, that are selected to be deployed in each cell of the cellular topology illustrated in Fig. 2, which is also used in overall system's performance evaluation.

TABLE 5. Deployed RNs.

Rank	Scenario 1		Scenario 2	
	DNN	*RNN*	*DNN*	*RNN*
1	RN-10	RN-10	RN-10	RN-10
2	RN-6	RN-6	RN-6	RN-6
3	RN-5	RN-5	RN-5	RN-4

V. DEEP QL FRAMEWORK FOR RN SELECTION

A. Q-LEARNING AND DEEP Q-LEARNING PRINCIPLES

RL is an ML category which is based on the interaction and communication with the learning environment to train and validate effective models. This is achieved by the utilization of a learning entity called software agent [5]. As also described in [5], *"the information feedback that the agent returns to the model is called rewards (positive case) or penalty (negative case)"*. In that way, the agent creates a policy to set up its own learning scheme and decide which actions to choose in a certain situation. The scope of an effective RL model is to maximize the cumulative rewards over time [26]. There are several known RL schemes such as state-action-reward-state-action [32], Q-learning [33], DQL [34], deep deterministic policy gradient [35] and asynchronous advantage actor-critic algorithm [36]. However, the most widely used RL algorithms are Q-Learning and deep Q-Learning, which combines Q-learning and neural networks.

Q-Learning algorithm has been proposed as an efficient way to deal with rapidly changing and non-linear environments. For this purpose, Q-Learning fits perfectly in the 5G/B5G wireless network domain. The cellular environment is characterized by complex propagation models, increased interference levels, dense connections and high user mobility, making Q-Learning a promising approach to solve complex optimization problems which have to do among others with resource allocation, power management and RN or BS selection.

A typical Q-Learning environment is depicted in Fig. 7. The agent (Q-function) collects feedback from the environment and takes some action that will later affect the environment. In other words, there is a set of potential states and a set of potential actions that can be performed. The agent specifies the transitions between states, based on the actions, aiming to maximize reward. Q-function is mathematically formulated as follows [5], [37], [38]:

$$Q'(s_t, a_t) \leftarrow Q(s_t, a_t) + a \times (r_t + \gamma \times max_b(Q(s_{t+1}, b) - Q(s_t, a_t)), b \in A \quad (10)$$

where Q' is the updated Q value, s_t is the state at the current time interval and s_{t+1} is the state at the next time interval. Moreover, α is the learning rate and r_t is the reward received from the network when moving from the state s_t to state s_{t+1} and A is the Q-table that stores all the actions. Moreover, γ is the discount factor which determines the importance of future rewards. In fact, $0 \leq \gamma \leq 1$, where a zero value means that only current rewards are considered, while a discount factor close to one means that long-term high rewards are of interest.

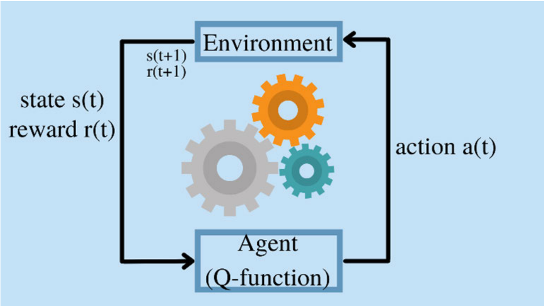


FIGURE 7. Q-learning environment.

However, in 5G/B5G RRM problems (such as RN placement and selection), when utilizing m-MIMO antenna configurations and advanced physical layer techniques such as OFDMA or NOMA, advance precoding and beamforming, the Q-table with the full set of potential actions, states and rewards can be large enough. This can exponentially increase the optimization problem's complexity, which comes against the major 5G/B5G requirement for low latency responses. In such cases, a NN can be trained to map the set of states with the best-performing action or in other words to perform the Q-function approximation. This RL technique is called DQL and is widely proposed due to its' ability to decrease the amount of the state-action duplets of the tabular-based Q-Learning algorithm, and, thus, produce more generalized models in optimization tasks [39], [40].

Based on Equation (10), a DQL agent aims to gather all the related information from the environment by minimizing the so-called temporal difference (TD) function [41], between the next Q-value $r_t + \gamma \times max_b(Q(s_{t+1}, b)), b \in A$ and the current Q-value $Q(s_t, a_t)$. For this purpose, the DQL's basic characteristic is to utilize two approximators (NNs). The first one is used to estimate the current best action, while the latter is used to predict the next action. A typical DQN structure is depicted in Fig. 8.

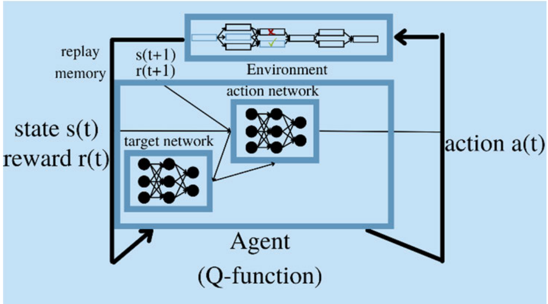


FIGURE 8. DQL Methodology [37].

B. PROPOSED DQL ALGORITHM FOR RN SELECTION

There are several DQL schemes, which are classified according to the algorithm's calculative iterations. The first category is the centralized DQL schemes, where a single software agent is used to perform the information gathering and processing from different sources placed in the environment. In the wireless communications domain, such an agent can be placed to the core network or on a server in a BS and collect information from different BSs and/or RNs. The other category is the decentralized DQL schemes, where multiple software agents are utilized and each one of them is responsible for communication and information gathering from a specific subset of the overall environment. Such agents can be placed in different BSs and/or be responsible for a subset of the total accepted users in the topology.

In this paper, a semi-centralized DQL framework is proposed to solve the RN selection problem subject to EE and SE maximization. The term semi-supervised, refers to the presence of multiple similar agents, one in each BS/cell of the topology.

The general state, action and reward of the proposed scheme are defined as follows:

1) STATE SPACE

Assuming that there are E number of episodes for DQL agent training, the system state is described as $S = \{S_1, S_2, \ldots, S_E\}$. At any time step, assuming t, the state is described by the following information about each UE, denoted as u, served by the secondary system (RN-assisted communication): a) the ID of the BS which serves UE u, b) the cell sector where UE u is positioned, c) the set of CSI information (channel coefficient matrices) between each one of the active RNs in the cell where UE u is located, declared as $H_u = \{H_{u,r_1}, H_{u,r_2}, \ldots, S_{u,r_k}\}$, where k is the total number of active RNs in the considered cell.

2) ACTION SPACE

The taken actions in each one of the E algorithm's episodes are noted as $A = \{A_1, A_2, \ldots, A_E\}$. At any time step, assuming t, and assuming that the $k_s, k_s \in (1, 2, \ldots, k)$ RN is currently selected for a UE u, the software agent can select the next, the previous or the same RN as the next action. In other words, the action that is taken at time t is denoted as $A_t = [a_{1,u_1}, a_{2,u_2}, \ldots, a_{u,u_u}, \ldots, a_{N,N}]$, where N denotes the total number of UEs that are served by the secondary system, $a_{m,u_m} \in \{RN_{step}, -RN_{step}, 0\}$ is the selection of the serving RN for UE m and RN_{step} is the change of RN for each UE under test. Thus, the serving RN update rule for each episode for UE m is calculated as follows:

$$RN_{m,u_m}(t) = RN_{m,u_m}(t-1) + a_{m,u_m}(t) \tag{11}$$

3) REWARD

After taking an action, as described previously, the DQL system transits into a new state thus leading to alternate RN selection for the UEs of the secondary system. The feedback received at time t focuses on EE and SE levels maximization and is expressed by:

$$r_t = \{r_{t_{EE}}(S_{t\text{-}1}, A_{t\text{-}1}), r_{t_{SE}}(S_{t\text{-}1}, A_{t\text{-}1})\} \tag{12}$$

where,

$$r_{t_{EE}}(S_{t\text{-}1}, A_{t\text{-}1}) = \begin{cases} \dfrac{EE_t - EE_{t\text{-}1}}{EE_{t\text{-}1}} \times 100, & if EE_t > EE_t \\ 0, & otherwise \end{cases} \tag{13}$$

$$r_{t_{SE}}(S_{t\text{-}1}, A_{t\text{-}1}) = \begin{cases} \dfrac{SE_t - SE_{t\text{-}1}}{SE_{t\text{-}1}} \times 100, & if SE_t > SE_t \\ 0, & otherwise \end{cases} \tag{14}$$

Regarding the action selection strategy, the ϵ-greedy method is used to balance the DQL algorithm's exploration and exploitation phases (with probabilities ϵ and $1-\epsilon$ respectively). Exploration refers to the DQL phase of improving knowledge about each action, whereas exploitation refers to the phase of maximizing the reward function by exploiting the set's action-value estimation.

Fig. 9 depicts the proposed semi-centralized DQN algorithm, where one agent is deployed per cell/BS. Each agent's training is performed only for the coverage area of the cell that is located into. This means that each agent is responsible only for a subject of the total UEs of the network. Thus, this approach considers C (performance evaluation considers $C = 19$) DQL agents, equal to the total cells of the topology. Each DQL agent optimizes performance in the coverage area assigned to the BS that is located. In order to ensure the global (for all cells) optimization of EE and SE performance, a global reward is defined for the whole cellular topology by the addition of all the single rewards of the C deployed agents.

To conclude with the DQL model, the global reward-based state transition is performed by a set of similar NNs, where the input layer includes the space state's triplet (serving BS ID, sector and channel coefficient between each UE of each cell and active RNs in a cell) for each of the C different agents (thus the corresponding BSs). The NN includes $C \times N_c \times 3$ neurons, where N_c is the number of active secondary system UEs in the cell $c \in C$. The output layer is one of the three possible Q-value results for each UE (select the next RN,

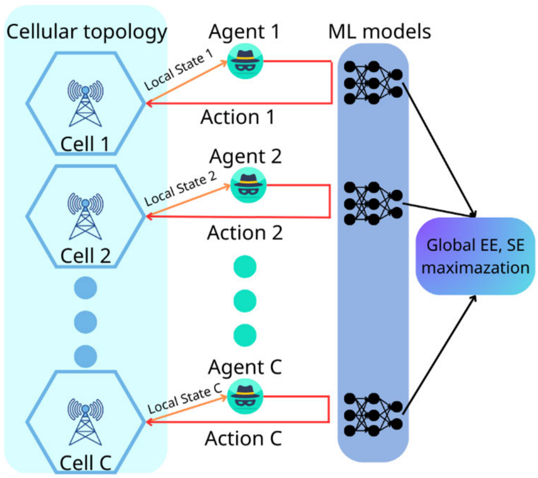


FIGURE 9. Proposed DQL scheme.

select the previous RN or select the same RN) concerning system's EE and SE maximization. Afterwards, a global reward optimization step is performed in order to define if the total system will change state or not. The NN structure is depicted in Fig. 10.

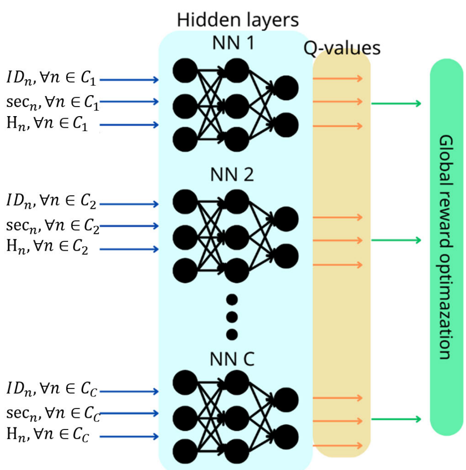


FIGURE 10. Proposed NNs architecture.

VI. PERFORMANCE EVALUATION

In this section, the performance of the proposed ML algorithms for RN placement and selection is presented and evaluated concerning the downlink of a 2-tier wireless multicellular 5G/B5G orientation. In all cases, algorithms' performance is compared to a state-of-the-art non-ML approach, presented in [11], as well as to a reference system where no RNs are deployed. The deployed RNs are layer 1 RNs (A&F) regarding the OSI level of deployment. Both Inband and Outband RNs are considered. When Inband RNs are used both BS-RN and RN-UE links share the same spectrum resources. On the other hand, when Outband RNs are used, additional spectrum resources are -a priori- exclusively for RN usage [42].

All the simulation setups in this section were implemented in MATLAB (R2022b release [43]) environment using among others the Communications Toolbox, the Statistics and Machine Learning Toolbox and Deep Learning Toolbox.

This section is spitted in two subsections. The first refers to the performance evaluation of the ML/DL-based RN placement algorithm in different RN implementation scenarios. The second subsection, refers to the performance evaluation of the overall system, where both the ML/DL-based RN placement scheme and the DQL RN selection algorithm are deployed.

A. DL-BASED RN PLACEMENT PERFORMANCE EVALUATION

In this subsection, the two proposed DL-schemes (DNN, LSTM), that have been presented in section IV for the problem of RN placement are evaluated. A 2-tier (19 BSs, 54 sectors) cellular orientation is considered, with network and simulation parameters as depicted in Table 2. UEs are uniformly distributed in the coverage area, while the number of requested subcarriers varies to either 6, 8 or 11.

Regarding RN implementation, five scenarios are examined in our simulations (including reference basis of no RN deployment), as follows: (1) *No-RN*: No RNs are deployed, (2) *SME-I*: Inband RNs are deployed in the middle edge of each sector, (3) *SME-O*: Outband RNs are deployed in the middle edge of each sector, (4) *MLP-I*: ML/DL-based Inband RN placement, (5) *MLP-O*: ML/DL-based Outband RN placement. It should be noted that in our simulation environment Outband RN scenarios use an additional bandwidth of ∼55MHz to serve initially rejected UEs, leading to interference mitigation and increased capacity gains over Inband ones [11]. It should be also noted that as LSTM's performance is slightly improved compared to DNN's performance, as depicted in section IV, we pick LSTM as the implemented ML/DL technique for RN placement for the simulations of this section.

Extensive MC simulations were performed, where the extracted mean values are presented for all considered KPIs. To this end, total system's EE is presented in Fig. 11, while the corresponding SE is presented in Fig. 12 for the aforementioned RN implementation scenarios. It should be noted that the best-performing candidate RNs are the same for the two scenarios that are discussed in section IV (CSI presence or not).

As it can be observed from Fig. 11, the use of RNs can significantly improve network's metrics, such as EE. Moreover, EE is increasing for increasing number of subcarriers per UE. In fact, for 6 subcarriers per UE, EE can reach up to 35.45/61.45 Mbps/W for the *SME-I*/*SME-O* scenarios, respectively. The EE values for the DL-enabled scenarios are 42.54/79.89 Mbps/W for the *MLP-I*/*MLP-O* scenarios, respectively. In the reference No-RN scenario, EE is limited to 23.45 Mbps/W. These numbers indicate that RN usage can improve total system's EE from ∼50-240%.

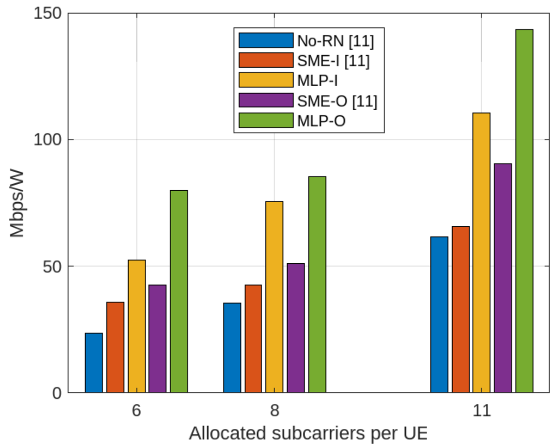


FIGURE 11. Mean total EE for various RN implementations.

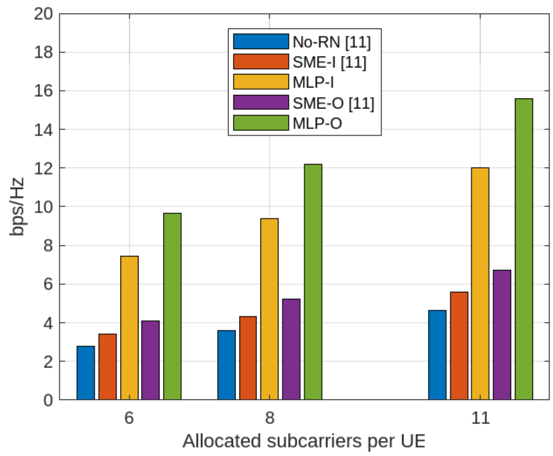


FIGURE 12. Mean total SE for various RN implementations.

When considering 11 subcarriers per UE the corresponding values are 52.38/75.34/110.34/90.45/143.45 Mbps/W for the *No-RN*/*SME-I*/*SME-O*/*MLP-I*/*MLP-O* scenarios, respectively, which lead to a ∼43-170% EE improvement. Similar conclusions can be drawn for SE as well, as depicted in Fig. 12, leading to a ∼20-200% SE improvement.

It can be witnessed, also, from Figs. 11, 12 that the use of the DL scheme for RN placement further improves networks' KPIs, such as EE and SE. In fact, comparing the non-ML Inband scenario (*SME-I*) with the ML/DL-enabled Inband scenario (*MLP-I*), a ∼20% improvement in both EE and SE for both 6 and 11 subcarriers per UE is depicted. Similarly, comparing the non-ML Outband scenario (*SME-O*) with the ML/DL-enabled Inband scenario (*MLP-O*), a ∼30% improvement in both EE and SE for both 6 and 11 subcarriers per UE is achieved.

B. OVERALL PERFORMANCE EVALUATION (RN PLACEMENT AND RN SELECTION)

For the overall system's performance evaluation, we consider the same 2-tier 5G/B5G network orientation as described in the previous subsection, with the parameters depicted in Table 2. In this subsection the DQL scheme for RN selection, which is analyzed in section V, is, also, enabled, acting additively to the RN placement DL scheme presented in both section IV and subsection A. The DQN parameters, that are used for the simulations of this subsection, are depicted in Table 6.

TABLE 6. DQN/DQL parameters.

Parameter	Value	Component
Number of hidden layers	4	DQN
Activation function (input and hidden layers)	ReLU	DQN
Activation function (output layer)	Linear	DQN
Memory size	10000	DQN
Mini-batch size	128	DQN
Optimizer	Adam	DQN
Loss function	Huber	DQN
Number of episodes	40000	DQL
Learning rate (α)	0.001	DQL
Discount factor (γ)	0.8	DQL

A key procedure that has to be performed when evaluating every ML-based scheme is hyperparameter tuning [5], which refers to extensive simulations with different ML parameter values. The scope of this procedure is the selection of the optimal set of parameters for the proposed ML schemes. These parameters are selected based on the overall system's performance optimization, based on KPIs of interest. The hyperparameters that have been selected after various simulations rounds are the following: a) number of episodes for the DQL algorithm, which affects the total training time, b) learning rate (α), which refers to the contribution percentage between the current and the previous Q-values, c) the discount factor (γ), which is linked to the significance of the future rewards.

As far as RN implementation is considered, five scenarios are examined in our simulations (including reference basis of no RN deployment), as follows: (1) *No-RN*: No RNs deployed, (2) *SME-I*: Inband RNs are deployed in the middle edge of each sector, (3) *SME-O*: Outband RNs are deployed in the middle edge of each sector, (4) *MLP-I*: ML/DL-based Inband RN placement and DQL RN selection, (5) *MLP-O*: ML/DL-based Outband RN placement and DQL RN selection.

To this end, total system's EE is presented in Fig. 13, while the corresponding SE is presented in Fig. 14 for the aforementioned RN implementation scenarios. As it can be observed from Fig. 13, the use of DRL-based RN placement can significantly improve network metrics, such as EE compared to the reference scenario where no RNs are deployed. In fact, for 6 subcarriers per UE, EE can reach up to 35.45/61.45 Mbps/W for the *SME-I*/*SME-O* scenarios, respectively. The EE values for the DQL scenarios are 76.95/139.79 Mbps/W for the *MLP-I*/*MLP-O* scenarios, respectively. In the reference *No-RN* scenario, EE is limited to 23.45 Mbps/W. These numbers indicate that DQL RN selection utilization can improve total EE from ∼140-500%. When considering 11 subcarriers per UE the corresponding values are 52.38/75.34/110.34/166.35/252.45 Mbps/W for the *No-RN/SME-I/SME-O/MLP-I/MLP-O* scenarios, respectively, which lead to a ∼200-500% EE improvement. Similar conclusions can be drawn for SE as well, as depicted in Fig. 14, leading to a ∼145-505% SE improvement.

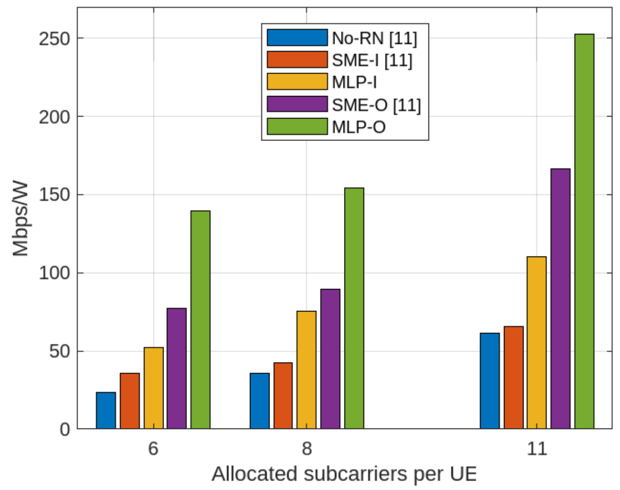


FIGURE 13. Mean total EE for various RN implementations (with DQL RN selection).

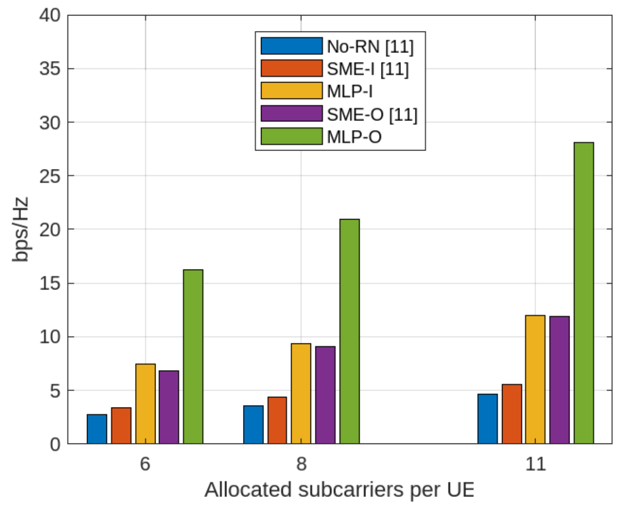


FIGURE 14. Mean total SE for various RN implementations (with DQL RN selection).

It can be witnessed, also, from Figs. 11, 12 that the use of the DQL RN selection scheme further improves networks' KPIs, such as EE and SE, compared to the case where only the DL RN placement algorithm is enabled. In fact, comparing ML/DL Inband scenario (*SME-I*) in these two occasions (only DL-based RN placement or DL-based RN placement and DQL RN selection), a ∼75% improvement in both EE and SE is depicted for 6 subcarriers per UE. For 11 subcarriers per UE the improvement is ∼80%. Similarly, comparing ML/DL Inband scenario (SME-I) in these two occasions, a ∼79% improvement in both EE and SE is depicted for 6 subcarriers per UE. For 11 subcarriers per UE the improvement is about ∼83%. Thus, it is visible that DQL-based RN selection can further improve overall system's performance.

From the above presented analysis and, also, from Figs. 11 - 14 the following outcomes can be witnessed:

- The proposed joint RN placement and selection DL/DRL-based framework can improve the performance of 5G/B5G networks, by the improvement of key network metrics, such as EE and SE.
- Concerning comparison with other state-of-the-art approaches we evaluate the proposed models in a two-level basis. More specifically:
  - The first level concerns the comparison of the proposed DL-enabled RN placement models with a 5G/B5G system where RNs are statically deployed and non-ML optimization techniques are utilized. From the subsection A it is visible that both EE and SE levels are improved by ∼30% compared to such a system (as described in [11]).
  - The second level concerns the comparison of the joint RN placement and selection framework with a 5G/B5G system where RNs are statically deployed and non-ML optimization techniques are utilized. It is derived by the analysis in subsection B that the DRL-based RN selection algorithm contributes even more on the EE and SE improvement. In fact, these KPIs can be improved by up to ∼80% compared to [11].
  - Moreover, our DL/DRL approach overperforms other state-of-the-art approaches that are not utilizing ML/DL models for RN placement and/or selection. For example, the proposed scheme in [44] reaches up to ∼50% improvement in EE levels compared to a state-of-the-art-approach. Moreover, our approach has similar or better performance compared to recently proposed ML-based schemes. For example, the proposed scheme in [45] reaches about ∼80% EE improvement compared to a non-ML state-of-the-art-approach.
- Finally, it should be mentioned at this point that, in general, Outband RN orientations overperform Inband ones in all scenarios under test. However, in Outband cases, extra bandwidth has been pre-allocated to RNs. Thus, despite the aforementioned gain over Inband ones, in real-world scenarios Outband RNs have extremely high deployment costs, due to the external resources and necessary hardware needed.

C. COMPUTATIONAL COMPLEXITY CONSIDERATIONS

A key aspect when designing AI/ML algorithms is the computational complexity gain that is achieved compared to traditional optimization (non-ML) approaches. In our case this is achieved in the following ways:

- As it is presented in Tables 3, 4 both DL models need ∼4 to 5 minutes for the training phase. After this phase, the response to select the best performing RNs deployment is instant, based on the constraints described in the subsection B of Section IV. In this analysis we should add the time for dataset generation which is ∼2 hours. In our approach presented in [11], ∼1 hour is needed for a round of ∼100 MC simulations.

- Finally, in the same context, as far as the DRL RN placement scheme is considered, each one of the C cells needs some time (∼1 to 2 minutes as the NNs there are lightweight) for the models' training, while dataset generation is performed online and, thus, there is no need for extra time there. As is visible, the aforementioned computation time is extremely lower than the one of a full MC simulation.

VII. CONCLUSION

In the present work, the joint optimization problem of RN placement and selection in 5G/B5G networks is investigated. The two sub-problems are considered as different ones and, thus, different approaches are proposed for each one of them. However, in all cases, DL-enabled techniques are utilized, forming an overall system's DL-assisted framework for metrics optimization.

In fact, for the RN placement problem two offline ML/DL methods (one DNN and one RNN/LSTM network) are proposed to select the k best locations for placing RNs inside each cell's coverage area. ML KPIs evaluation indicated that the LSTM network is the best-performing solution in achieving optimal trade-off between ML KPIs (accuracy, precision, recall, F1-score) and required training time. Moreover, both approaches propose the same candidate RNs as the best-performing ones in maximizing accepted UEs in each cell of the 2-tier cellular topology under test.

The second sub-problem refers to the selection of the best-performing RN for the UEs of the secondary system (the primary system refers to the UEs served directly by the BS), subject to the joint maximization of the overall EE and SE levels of the system. For this purpose, a DQN-based framework is proposed for improving the system-level EE and SE of two-tier 5G heterogeneous cells with multi-channel transmissions. Following extensive simulations using a 5G/B5G system and link level simulator implemented in our lab, we showed that by only enabling ML/DL-based RN placement, a 30% improvement in both EE and SE levels can be achieved, compared to the non-ML approach of [11]. Additionally, if the DQN-based RN selection is also enabled, the EE improvement can reach 80%, while SE can be improved by 75% compared to the non-ML approach.

REFERENCES

[1] CISCO, "Cisco annual internet report (2018–2023)," White paper, Cisco Syst., Inc. San Jose, CA, USA, White Paper, 2020.

[2] *Service Requirements for the 5G System*, document 3GPP TS 22.261, Version 17.11.0, Release 17, 2022.

[3] *NR and NG-RAN Overall Description*, document 3GPP TS 38.300, Version 17.0.0, Release 17, 2022.

[4] P. K. Gkonis, "A survey on machine learning techniques for massive MIMO configurations: Application areas, performance limitations and future challenges," *IEEE Access*, vol. 11, pp. 67–88, 2023, doi: 10.1109/ACCESS.2022.3232855.

[5] I. A. Bartsiokas, P. K. Gkonis, D. I. Kaklamani, and I. S. Venieris, "ML-based radio resource management in 5G and beyond networks: A survey," *IEEE Access*, vol. 10, pp. 83507–83528, 2022, doi: 10.1109/ACCESS.2022.3196657.

[6] Y. Azimi, S. Yousefi, H. Kalbkhani, and T. Kunz, "Applications of machine learning in resource management for RAN-slicing in 5G and beyond networks: A survey," *IEEE Access*, vol. 10, pp. 106581–106612, 2022, doi: 10.1109/ACCESS.2022.3210254.

[7] B. Agarwal, M. A. Togou, M. Marco, and G. Muntean, "A comprehensive survey on radio resource management in 5G HetNets: Current solutions, future trends and open issues," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2495–2534, 4th Quart., 2022, doi: 10.1109/COMST.2022.3207967.

[8] M. Beshley, N. Kryvinska, and H. Beshley, "Energy-efficient QoE-driven radio resource management method for 5G and beyond networks," *IEEE Access*, vol. 10, pp. 131691–131710, 2022, doi: 10.1109/ACCESS.2022.3228758.

[9] N. K. M. Madi, M. M. Nasralla, and Z. M. Hanapi, "Delay-based resource allocation with fairness guarantee and minimal loss for eMBB in 5G heterogeneous networks," *IEEE Access*, vol. 10, pp. 75619–75636, 2022, doi: 10.1109/ACCESS.2022.3192450.

[10] R. Gatti, G. B. A. Kumar, K. N. S. Kumar, S. Palle, and T. R. Gadekallu, "Optimal resource scheduling algorithm for cell boundaries users in heterogenous 5G networks," *Phys. Commun.*, vol. 55, Dec. 2022, Art. no. 101915, doi: 10.1016/j.phycom.2022.101915.

[11] K. A. Psilopanagiotis, I. A. Bartsiokas, P. K. Gkonis, and D. I. Kaklamani, "On relay-based subcarrier allocation and power management in 5G multi-cellular networks," in *Proc. IEEE 95th Veh. Technol. Conf., (VTC-Spring)*, Jun. 2022, pp. 1–6, doi: 10.1109/VTC2022-Spring54318.2022.9861029.

[12] *5G NR: NR and NG-RAN Overall Sescription Stage-2*, document 3GPP TS 38 300, Version 16.5.0, Release 16, 2021.

[13] *5G NR: Sidelink Relay Adaptation Protocol (SRAP) Specification*, document 3GPP TS 38 351, Version 17.0.0, Release 17, 2023.

[14] A. Bhattacharya and A. Kumar, "A shortest path tree based algorithm for relay placement in a wireless sensor network and its performance analysis," *Comput. Netw.*, vol. 71, pp. 48–62, Oct. 2014, doi: 10.1016/j.comnet.2014.06.011.

[15] R. A. Safiee, N. I. A. Apandi, N. A. Muhammad, W. W. Sheng, and M. A. Sarijari, "Relay node placement in wireless sensor network for manufacturing industry," *Bull. Electr. Eng. Informat.*, vol. 12, no. 1, pp. 158–166, Feb. 2023.

[16] A. A. As'ari, N. I. A. Apandi, N. A. Muhammad, R. A. Rashid, M. A. Sarijari, and J. Salleh, "Energy efficiency scheme for relay node placement in heterogeneous networks," *Bull. Electr. Eng. Informat.*, vol. 12, no. 1, pp. 187–195, Feb. 2023.

[17] N. H. Khattab, M. S. Darweesh, and S. S. Soliman, "Relay selection in NOMA-based cooperative wireless backhaul networks," in *Proc. 17th Wireless Demand Netw. Syst. Services Conf. (WONS)*, Mar. 2022, pp. 1–8, doi: 10.23919/WONS54113.2022.9764466.

[18] E. Yanmaz, "Positioning aerial relays to maintain connectivity during drone team missions," *Ad Hoc Netw.*, vol. 128, Apr. 2022, Art. no. 102800.

[19] J. Kim, P. Ladosz, and H. Oh, "Optimal communication relay positioning in mobile multi-node networks," *Robot. Auto. Syst.*, vol. 129, Jul. 2020, Art. no. 103517, doi: 10.1016/j.robot.2020.103517.

[20] M. Singh and G. Anudeep, "Link restoration and relay node placement in partitioned wireless sensor network," in *Design and Development of Efficient Energy Systems*. Beverly, MA, USA: Scrivener Publishing LLC, Jul. 2022, pp. 101–117.

[21] H. Kapil and C. S. R. Murthy, "A pragmatic relay placement approach in 3-D space and Q-learning-based transmission scheme for reliable factory automation applications," *IEEE Syst. J.*, vol. 12, no. 1, pp. 823–833, Mar. 2018, doi: 10.1109/JSYST.2016.2524695.

[22] H. Amiriara, M. R. Zahabi, and V. Meghdadi, "Power-location optimization for cooperative nomadic relay systems using machine learning approach," *IEEE Access*, vol. 9, pp. 74246–74257, 2021, doi: 10.1109/ACCESS.2021.3079171.

[23] S. Dang, J. Tang, J. Li, M. Wen, S. Abdullah, and C. Li, "Combined relay selection enabled by supervised machine learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3938–3943, Apr. 2021, doi: 10.1109/TVT.2021.3065074.

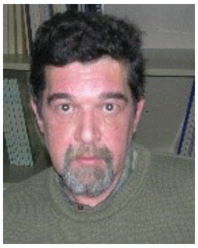
[24] Y. Geng, E. Liu, R. Wang, Y. Liu, J. Wang, G. Shen, and Z. Dong, "Deep deterministic policy gradient for relay selection and power allocation in cooperative communication network," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1969–1973, Sep. 2021, doi: 10.1109/LWC.2021.3088894.

[25] M.-J. Paek, Y.-J. Na, W.-S. Lee, J.-H. Ro, and H.-K. Song, "A novel relay selection scheme based on Q-learning in multi-hop wireless networks," *Appl. Sci.*, vol. 10, no. 15, p. 5252, Jul. 2020, doi: 10.3390/app10155252.
[26] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, "Q-learning algorithms: A comprehensive classification and applications," *IEEE Access*, vol. 7, pp. 133653–133667, 2019.
[27] P. K. Gkonis, M. A. Seimeni, N. P. Asimakis, D. I. Kaklamani, and I. S. Venieris, "A new subcarrier allocation strategy for MIMO-OFDMA multicellular networks based on cooperative interference mitigation," *Sci. World J.*, vol. 2014, pp. 1–9, Jan. 2014, doi: 10.1155/2014/652968.
[28] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[29] Y. Li, E. Pateromichelakis, N. Vucic, J. Luo, W. Xu, and G. Caire, "Radio resource management considerations for 5G millimeter wave backhaul and access networks," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 86–92, Jun. 2017.
[30] *5G NR Physical Channels and Modulation*, document 3GPP TS 138 211, Version 15.3.0, Release 17, 2023.
[31] F. S. Samidi, N. A. M. Radzi, K. H. M. Azmi, N. M. Aripin, and N. A. Azhar, "5G technology: ML hyperparameter tuning analysis for subcarrier spacing prediction model," *Appl. Sci.*, vol. 12, no. 16, p. 8271, Aug. 2022.
[32] Q. Liu, C. F. Kwong, S. Zhou, T. Ye, L. Li, and S. P. Ardakani, "Autonomous mobility management for 5G ultra-dense HetNets via reinforcement learning with tile coding function approximation," *IEEE Access*, vol. 9, pp. 97942–97952, 2021, doi: 10.1109/ACCESS.2021.3095555.
[33] J. Clifton and E. Laber, "Q-learning: Theory and applications," *Annu. Rev. Statist. Appl.*, vol. 7, pp. 279–301, Mar. 2020.
[34] F. Tan, P. Yan, and X. Guan, "Deep reinforcement learning: From Q-learning to deep Q-learning," in *Proc. Neural Inf. Process., 24th Int. Conf. (ICONIP)*, Guangzhou, China: Springer, Nov. 2017, pp. 475–483.
[35] C. Qiu, Y. Hu, Y. Chen, and B. Zeng, "Deep deterministic policy gradient (DDPG)-based energy harvesting wireless communications," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8577–8588, Oct. 2019, doi: 10.1109/JIOT.2019.2921159.
[36] L. Liu, J. Feng, Q. Pei, C. Chen, Y. Ming, B. Shang, and M. Dong, "Blockchain-enabled secure data sharing scheme in mobile-edge computing: An asynchronous advantage Actor–Critic learning approach," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2342–2353, Feb. 2021, doi: 10.1109/JIOT.2020.3048345.
[37] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.
[38] A. Salh, L. Audah, M. A. Alhartomi, K. S. Kim, S. H. Alsamhi, F. A. Almalki, Q. Abdullah, A. Saif, and H. Algethami, "Smart packet transmission scheduling in cognitive IoT systems: DDQN based approach," *IEEE Access*, vol. 10, pp. 50023–50036, 2022.
[39] A. Giannopoulos, S. Spantideas, N. Kapsalis, P. Karkazis, and P. Trakadas, "Deep reinforcement learning for energy-efficient multi-channel transmissions in 5G cognitive HetNets: Centralized, decentralized and transfer learning based solutions," *IEEE Access*, vol. 9, pp. 129358–129374, 2021, doi: 10.1109/ACCESS.2021.3113501.
[40] M.-A. Dittrich and S. Fohlmeister, "A deep q-learning-based optimization of the inventory control in a linear process chain," *Prod. Eng.*, vol. 15, no. 1, pp. 35–43, Feb. 2021.
[41] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[42] A. BenMimoune and M. Kadoch, "Relay technology for 5G networks and IoT applications," in *Internet of Things: Novel Advances and Envisioned Applications*. Cham, Switzerland: Springer, 2017, pp. 3–26.
[43] *Release Notes for MATLAB, MATLAB & Simulink, MathWorks*. Accessed: May 9, 2023. [Online]. Available: https://www.mathworks.com/help/MATLAB/release-notes.html
[44] M. Rahman, Y. Lee, and I. Koo, "Energy-efficient power allocation and relay selection schemes for relay-assisted D2D communications in 5G wireless networks," *Sensors*, vol. 18, no. 9, p. 2865, Aug. 2018.
[45] Q. W. Ahmed, S. Garg, A. Rai, M. Ramachandran, N. Z. Jhanjhi, M. Masud, and M. Baz, "AI-based resource allocation techniques in wireless sensor Internet of Things networks in energy efficiency with data optimization," *Electronics*, vol. 11, no. 13, p. 2071, Jul. 2022.

IOANNIS A. BARTSIOKAS was born in Athens, Greece, in 1997. He received the M.Eng. degree in electrical and computer engineering (ECE) from the National Technical University of Athens (NTUA), in 2020, where he is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering. His current research interests include wireless networks, radio resource management in 5G/B5G networks, machine learning and deep learning in wireless systems, adaptive modulation, and MIMO antennas. He is a member of the Technical Chamber of Greece.

PANAGIOTIS K. GKONIS received the Diploma and Ph.D. degrees in electrical and computer engineering and the M.Sc. degree in engineering-economic systems program from the National Technical University of Athens (NTUA), Greece, in 2005, 2009, and 2009, respectively. From 2010 to 2015, he was a Postdoctoral Researcher with the Intelligent Communications and Broadband Networks Laboratory (ICBNET), NTUA. He was a Scientific Associate with the Department of Electrical Engineering, and the Department of Aircraft Engineering, Technological Institute of Sterea Ellada (Former). He is currently an Assistant Professor with the Department of Digital Industry Technologies, National and Kapodistrian University of Athens. He has authored/coauthored more than 60 scientific publications in international journals, conferences, and book chapters in the areas of wireless communications and computational modeling of cellular networks. As a researcher, he has participated in various national and European-funded projects.

DIMITRA I. KAKLAMANI was born in Athens, Greece, in 1965. She received the Ph.D. degree in electrical and computer engineering (ECE) from the National Technical University of Athens (NTUA), in 1992. She was elected as a Lecturer, an Assistant Professor, an Associate Professor, and a Professor with the School of ECE, NTUA, in April 1995, April 2000, October 2004, and February 2009, respectively. She has more than 330 publications in the fields of software development for information transmission systems modeling, microwave networks, mobile, and satellite communications. She has coordinated the NTUA activities in the framework of several EU and national projects in the same areas. She is an Editor of one international book by Springer-Verlag in applied CEM, in 2000, and a reviewer for several IEEE journals.



IAKOVOS S. VENIERIS has been a Professor with the School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), since 1994, and the Director of the Intelligent Communications and Broadband Networks (ICBNet) Laboratory. He has participated in and has successfully led several national and international projects. His research interests include distributed systems, security and privacy, software and service engineering, agent technology, multimedia, mobile communications and machine learning, internetworking, signaling, resource scheduling and allocation for network management, modeling, performance evaluation, and queuing theory. He has more than 370 publications in the above areas and has received several national and international awards for academic achievement.