Received: 21 February 2023 | Revised: 8 September 2023 | Accepted: 12 October 2023

DOI: 10.4218/etrij.2023-0065

ORIGINAL ARTICLE

ETRI Journal WILEY

Deep reinforcement learning for base station switching scheme with federated LSTM-based traffic predictions

Hyebin Park | **Seung Hyun Yoon**

Telecommunications & Media Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

Correspondence
Hyebin Park, Telecommunications & Media Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea.
Email: hb0390@sookmyung.ac.kr

Funding information
The Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government (MSIT)
(no. 2021-0-00851, On-demand data based network intelligence framework technology development).

Abstract

To meet increasing traffic requirements in mobile networks, small base stations (SBSs) are densely deployed, overlapping existing network architecture and increasing system capacity. However, densely deployed SBSs increase energy consumption and interference. Although these problems already exist because of densely deployed SBSs, even more SBSs are needed to meet increasing traffic demands. Hence, base station (BS) switching operations have been used to minimize energy consumption while guaranteeing quality-of-service (QoS) for users. In this study, to optimize energy efficiency, we propose the use of deep reinforcement learning (DRL) to create a BS switching operation strategy with a traffic prediction model. First, a federated long short-term memory (LSTM) model is introduced to predict user traffic demands from user trajectory information. Next, the DRL-based BS switching operation scheme determines the switching operations for the SBSs using the predicted traffic demand. Experimental results confirm that the proposed scheme outperforms existing approaches in terms of energy efficiency, signal-to-interference noise ratio, handover metrics, and prediction performance.

KEYWORDS
base station switching, deep reinforcement learning, federated learning, LSTM, traffic forecasting

1 | INTRODUCTION

As mobile traffic demands increase, small base stations (SBSs) have been densely overlapped to meet traffic requirements in high-traffic areas. Densely deployed SBSs can improve mobile services, but can cause serious energy consumption problems. According to an Ericsson Mobility Report, data traffic from mobile networks is expected to grow by a factor of seven between 2020 and 2026 [1]. Accordingly, the energy consumption from the dense deployment of SBSs will also increase to eventually occupy a significant portion of the energy consumption in information and communications technology. In addition, whole base stations (BSs) consume approximately 70%–80% of the total energy consumption of cellular networks [2]. To address this issue, several BS-switching operation strategies have been studied to minimize BS energy consumption [3].

BS switching operation methods that switch BS status ON/OFF, that is, deactivation or sleep strategies, have recently been studied [4–8]. The BS switching operation is considered one of the most efficient methods for

reducing energy consumption because the current network architecture does not need to be changed. Although the BS-switching operation strategy can significantly reduce energy consumption, it can degrade the quality of service (QoS) for mobile users. A deactivated BS causes handovers for the users it serves. Consequently, these users are associated with a suboptimal BS. If BSs frequently change their mode between active and deactivated in a high-traffic area, mobile users may experience serious QoS degradation from frequent handovers. To control the degradation in QoS, it is necessary to study BS switching operation techniques while considering traffic demand and user mobility.

In this study, we introduce a deep reinforcement learning (DRL)-based BS switching operation scheme with a federated traffic prediction model to optimize system energy efficiency. We first formulate an energy-efficiency optimization problem that minimizes energy consumption and QoS degradation. To minimize the QoS degradation, we adopted a traffic demand forecasting model using a federated learning-based long short-term memory (F-LSTM) model. Finally, to obtain efficient switching operations in dynamic environments, we optimized energy efficiency using DRL-based BS switching operations with the forecasted traffic demand. The main contributions of our study are summarized as follows.

- We formulated an objective to maximize the energy efficiency by developing an energy efficiency model. BS switching operations can reduce energy consumption, whereas they cannot maintain the QoS because of handovers from the deactivated SBS. Therefore, the energy efficiency model consists of an energy consumption model and a data rate model. Subsequently, we established an equation that maximizes energy efficiency while guaranteeing QoS stability.
- To implement the traffic-aware BS switching operation method, we proposed the F-LSTM model, which predicts user trajectories. The trajectories from each user have spatiotemporal-dependent patterns. Hence, we use the trajectory to determine and forecast traffic demand. However, users cannot predict traffic accurately using only their own data because the size of the dataset is not large enough. Since there is a privacy problem with sharing user trajectory data, we used a federated learning approach that does not share trajectory data. In addition, our model needs the ID of the previous cell, which provides the user trajectory and user communication conditions. The predicted trajectories can be used to determine the user traffic demand for the next time slot.
- The proposed traffic-aware BS switching operation scheme determines the switching operations using the forecasted traffic data. In our method, the DRL model determines the switching operations by considering the energy and QoS degradation factors. Our scheme makes decisions that can minimize energy consumption by switching OFF inefficient BSs. It can also minimize QoS degradation by reducing frequent and repetitive switching operations and handovers. Through the proposed scheme, we achieved significant performances in the experiments.

The remainder of this paper is organized as follows. In Section 2, we review the related studies on BS switching operations. In Section 3, we describe the system model and define the problem. In Section 4, the design of the traffic-aware BS switching operation method using the F-LSTM model is described. The simulation results and a discussion of the proposed scheme are presented in Section 5. Finally, the conclusion of this study is presented in Section 6.

2 | RELATED STUDIES

BS switching operations are a promising method for addressing energy consumption problems in dense mobile networks. However, it is an NP-hard problem to determine the optimal BS switching operation that minimizes the energy consumption with various constraints in polynomial time [9]. Therefore, heuristic and greedy methods have been adopted to determine optimal BS switching operations [4, 5, 7, 10]. In Oh and Krishnamachari [4], a dynamic BS switching operation was proposed that considered the handover traffic of the neighboring BSs. This operation determines the BS switching operations by comparing the calculated handover traffic with the switching threshold. In Oh and others [5], a BS switching operation method was proposed that minimized the effect of switching operations that increased the load on neighboring BSs. It defines a network impact that considers the additional load caused by the handover users of deactivated BSs. By treating the network impact as a decision metric, it can also reduce the signaling and implementation overhead. To improve energy performance while ensuring full coverage, a BS switching scheduling scheme for both the uplink and downlink was proposed [7]. A set of BS switching patterns at the global system level offers full coverage by applying suitable scheduling schemes.

However, traditional optimization methods have computational complexity problems because they must ensure the data rate and QoS for users. Hence, some studies have defined dynamic BS switching operations as a Markov decision process (MDP). In Li and others [11],

reinforcement learning (RL) was applied to solve the MDP problems. In particular, an actor-critic approach was used that incorporates the strengths of both policy and value-based RL approaches [12]. In El-Amine and others [13], an energy-efficient strategy was introduced for BSs with multiple sleep mode (SM) levels to reduce energy consumption. It focuses on multilevel SM, where the BSs can switch between several SM levels. To optimize the balance between energy savings and delay, a Q-learning algorithm was proposed to adapt the BS states based on user locations and velocities. Additionally, to strike a balance between energy savings and system delay, an advanced algorithm was proposed to determine the optimal SM level for BSs by assessing various SM options [14]. The calculation of the system delay takes into account the wake-up time associated with each SM level, aiding in the selection of the most suitable switch-off SM policy. A BS switching operation scheme using the Q-learning method was proposed to minimize the energy consumption and data loss in El Amine and others [15]. This scheme determined the operations based on the interference of the mobile user, expected throughput, and buffer size of each BS. The method in Masoudi and others [16] utilized an online RL technique, specifically SARSA, to develop an algorithm for determining the appropriate SM based on factors such as time and BS load. To demonstrate the effectiveness of the algorithm, actual mobile traffic data collected from a BS located in Stockholm were employed.

Unfortunately, increasing the dynamics of the network environment significantly increases the state-action space of existing RL models. This requires an exponentially increasing space and degrades the performance of the RL model. To handle a high-dimensional state-action space, DRL-based approaches were researched in earlier studies [17, 18]. In Ye and Zhang [19], a traffic demand-aware DRL approach was studied to incorporate the spatial and temporal correlations of traffic arrivals. To assist in the exploration of the DRL process, a cost-greedy action refinement procedure was defined to address the inefficiency of the random exploration procedure. In Ju and others [20], the challenge was to efficiently determine the optimal active mode/SM for BSs in ultradense networks. The proposed method introduced a DRL-based approach to reduce energy consumption. A crucial aspect of this approach is the utilization of a decision selection network to streamline the selection process and decrease the complexity of the action space. In addition, to dynamically consider changing traffic demands, forecasting the traffic demand of cells has also been studied in BS switching operations [21, 22]. Because operating a BS in SM causes handovers for the users it serves, their QoS significantly degrades. In Wu and others [23], a traffic-aware dynamic BS switching operation was proposed that jointly adopts a convolutional-LSTM method and DRL methods. In the convolutional-LSTM model, both the geographical and semantic spaces were considered as traffic similarity graphs. Using the predicted traffic, the DRL model performs a training process to reduce the energy and QoS degradation costs.

Nevertheless, the approaches mentioned above cannot handle the degradation in QoS during handover. The BS switching operation causes frequent handovers for users; therefore, robust connectivity is an important issue [24, 25]. The handover users experience QoS degradation because deactivating the optimal BS forces users re-associate with another BS and obtain services from a sub-optimal BS. Thus, a BS-switching operation method that considers mobility and traffic forecasting is required to minimize the degradation in QoS. However, the mobility dataset of each user is not sufficient to train the model; therefore, the dataset must be shared on a centralized server. To handle the data privacy problems caused by data sharing and improve the performance of the forecasting model in fast-changing mobile networks, a federated learning-based mobility forecasting model must be examined [26–28].

In this paper, we introduce a DRL-based BS switching operation scheme with the F-LSTM model to predict traffic that optimizes energy efficiency. The proposed F-LSTM model consists of a global model and multiple local models, with each model consisting of various LSTM layers. The trained global model can predict user locations; therefore, the predicted locations can be calculated as traffic. We used the predicted traffic data to reduce the degradation in QoS for the handover users. By expressing the changes in traffic as user trajectories, the change in QoS of the handover users can be considered. Therefore, our proposed DRL-based BS switching scheme determines operations by considering the predicted traffic, current traffic demand, status of the BSs, and active time of the BSs. Moreover, we determine the BS operations by considering the active time to minimize the frequent switching that causes handovers.

3 | SYSTEM MODEL AND PROBLEM DEFINITION

In this study, a heterogeneous network (HetNet) that consists of **M** macro BSs (MBSs) and **S** SBSs is considered, as illustrated in Figure 1. In addition, **U** users move continuously and are preferentially associated with the SBS that provides the highest signal-to-interference noise ratio (SINR). The sets of MBSs, SBSs, and users are denoted by $\mathbf{M} = \{1, ..., M\}$, $\mathbf{I} = \{1, ..., I\}$, and

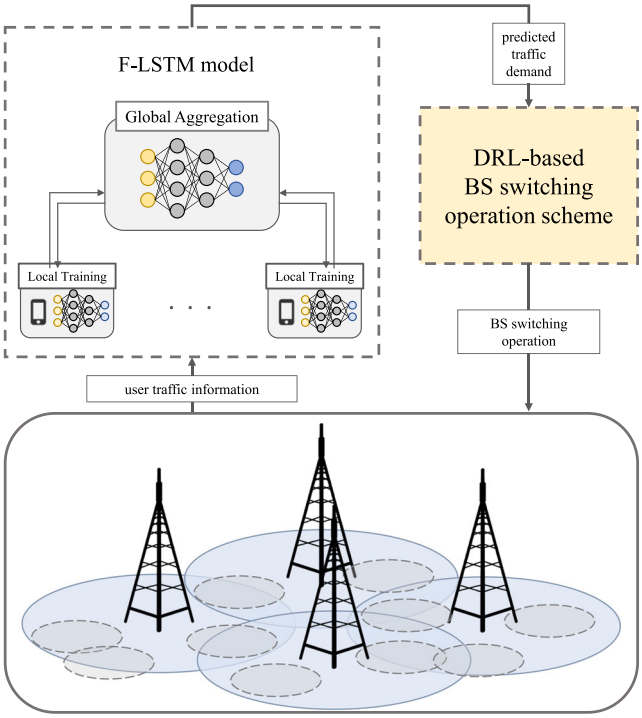

FIGURE 1 Overview of our proposed scheme and the heterogeneous network architecture considered in this study, which consists of multiple macro base stations (MBSs) and small base stations (SBSs). The MBSs are deployed in each coverage region, whereas the SBS coverage overlaps with the MBS coverage.

$\mathbf{U} = \{1, ..., U\}$, respectively. To provide services to users in areas with high-traffic demand, SBSs are deployed and have coverage overlaps. The active status indicator of an SBS is denoted by a_s, where 1 represents the active mode and 0 represents the SM. Time T is divided into t time slots during which the active status of the SBSs remains unchanged.

We define the energy consumption model of an SBS using two parts: a fixed energy part and a load-dependent energy part. SBSs in SM only consume the fixed energy, as reported in Chang and others [29]. Therefore, we express the energy consumption, P_I^t, of the SBSs in timeslot t as follows:

$$P_I^t = \sum_{i \in \mathbf{I}} \left(a_i \left(\phi_1 + \Delta \sum_{u \in \mathbf{U}} b_i^u p_i \right) + (1 - a_i)\phi_0 \right), \quad (1)$$

where ϕ_1 is the fixed energy consumption of an active mode SBS, ϕ_0 is the fixed energy consumption of a sleep mode SBSs, and Δ is the load-dependent power consumption of the SBSs. b_i^u is the association indicator for user equipment (UE) u with SBS i and takes a value of 1 for an associated UE and 0 for a non-associated UE. Finally, p_i is the transmission power of SBS i.

We aimed to minimize the energy consumption of the SBSs while minimizing the QoS degradation caused by switching operations. In other words, minimizing the energy consumption of the SBSs with minimal QoS degradation can be solved as an energy efficiency maximization problem. "Energy efficiency" refers to the ability to obtain an achievable data rate per unit of energy consumed. This means that the user QoS can be guaranteed while the energy consumption is minimized. Therefore, the achievable data rate in timeslot t can be calculated according to the Shannon capacity as follows:

$$R^t = W \cdot \sum_{u \in \mathbf{U}} \left(\log_2 \left(1 + \tau_u^t \right) \right), \quad (2)$$

where W represents the system bandwidth and τ_u^t is the SINR of UE u served by the BS during time slot t. Incorporating the AWGN channel model, the SINR τ_u^t is computed as follows:

$$\tau_u^t = \frac{p_{u,i} \cdot h_{u,i}}{\sum_{j \in \mathbf{I} \setminus i} p_{u,j} \cdot h_{u,j} + \sigma^2}, \quad (3)$$

where $p_{u,i}$ is the transmission power of SBS i to UE u, $h_{u,i}$ is the channel gain between UE u and its serving BS, σ^2 represents the noise power, and the summation runs over all SBSs except i. This way, the energy efficiency of the system is expressed as follows:

$$EE^t = \frac{R^t}{P_I^t}. \quad (4)$$

To determine the optimal BS switching operation strategy for maximizing energy efficiency, we formulated the problem as an MDP that can be defined as a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$. Here, $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the state transition function, $\mathcal{R}$ is the reward function, and γ is the discount factor, $\gamma \in [0,1]$. The MDP design is described in detail in the following.

1. **State**: The state s^t at each time slot t consists of four factors: the forecasted traffic demand $^{D^t}$, the traffic demand D^{t-1} at time slot $t-1$, the active status of SBSs i, a_i^{t-1}, and the active time of SBS $i, t_{a,i}^t$. Thus, the state s^t can be expressed as

$$s^t = \{^{D^t}, D^{t-1}, a_i^{t-1}, t_{a,i}^t\}. \quad (5)$$

2. **Action**: The action a^t is the decision to perform switching operations for an SBS at time slot t. According to the state transition function $P(s^{t+1}|s^t, a^t)$ with a given state s^t and action a^t, the current state s^t

transitions to the next state s^{t+1} when action a^t is executed. Action a^t can be represented as

$$a^t \in \{0,1\}, \tag{6}$$

where 0 is for the deactivating operation and 1 is for the activating operation.

3. **Reward**: Executing action a^t on SBS i affects the neighboring BSs N_i. Hence, we defined the reward by considering neighboring BSs. Reward r^t is composed of three parts: the energy cost of SBS $i, P^t(N_i)$, where neighboring BSs N_i include SBS i; the cost of QoS degradation, $C_i^t(a^{t-1},a^t)$; and the fixed value for switching penalty, $\varphi^t(a^{t-1},a^t)$. The switching penalty is considered to minimize continuous and frequent switching operations. Reward r^t can be expressed as

$$r^t = C_i^t\left(a^{t-1},a^t,N_i\right) - P^t(N_i) - \varphi^t\left(a^{t-1},a^t\right). \tag{7}$$

The cost of QoS degradation $C_i^t(a^{t-1},a^t,N_i)$ consists of handover factors and the achievable data rates of the UEs at neighboring BSs N_i. $P^t(N_i)$ can be calculated only for N_i, not for all SBS, using (1). To reduce the OFF operation for SBSs that service many UEs, we used handover factor h_i^t. This means the UEs from a deactivated SBS i are handed over to neighboring BSs N_i. Hence, the handover factor h_i^t can be calculated as

$$h_i^t = \sum_{u \in \mathbf{U}} \left(b_{i,t-1}^u - b_{i,t}^u\right), \tag{8}$$

where $b_{i,t}^u$ and $b_{i,t-1}^u$ are the association indicators for UE u served by SBS i at time slot t. Moreover, the cost of QoS degradation $C_i^t(a^{t-1},a^t)$ can be defined as follows:

$$C_i^t\left(a^{t-1},a^t,N_i\right) = \left(h_i^t\right)^{-1} \cdot \frac{R(N_i)}{\sum_{n,i' \in N_i} b_{i'}^u}, \tag{9}$$

where $R(N_i)$ is the achievable data rate of UEs who are associated with neighboring BSs N_i.

To maximize the system energy efficiency, we adopted the state-action value function as the expected value at a given state s^t when action a is executed. The state-action value function $Q(s^t, a)$ is defined as

$$Q(s^t, a) = r^t + \gamma^q \cdot \min_{a'} \mathbb{E}\left[Q\left(s^{t+1}, a'\right)\right]. \tag{10}$$

Therefore, our objective, which is to maximize the energy efficiency, can be used to determine the action with the given state s^t as $a = \arg\max_a Q(s^t, a)$ for all SBSs, which can be defined as follows:

$$\max_{a_1,\ldots,a_i} EE, \tag{11}$$

$$\begin{aligned} \text{s.t.} \quad & C1: \tau^{\min} \le \tau_u, \ \forall u \\ & C2: p_i^u \le p^{\max}, \ \forall u,i \\ & C3: p_m^u \le p^{\max}, \ \forall u,m \\ & C4: R_{\min} \le R_u, \ \forall u \end{aligned} \tag{12}$$

where $\tau^{\min}$ is the minimum threshold of SINR, p_i^u and p_m^u are, respectively, the transmission powers of the SBSs and MBS to UE u, $p^{\max}$ is the maximum transmission power, R_u is the achievable data rate of UE u, and $R_{\min}$ is the minimum threshold of the achievable data rate. $C1$ is the minimum SINR constraint of UE, $C2$ and $C3$ are the maximum transmission power constraints, and $C4$ ensures the data rates of the UEs.

4 | PROPOSED BS SWITCHING OPERATION SCHEME

In this section, the DRL-based BS switching operation scheme with traffic prediction is introduced. A handover occurs when a BS switches from active mode to SM and this can significantly impact the QoS. During a handover, the connection between the UE and the associated BS is severed, and a new association is established with the new BS. The more handovers a UE experiences, the more likely it is that the QoS will be impacted. This is because each handover requires time for a new association and can result in disruptions to the communication link. This can degrade communication quality, increasing congestion and reducing performance. In conclusion, minimizing the number of handovers is important for maintaining a high QoS in mobile communication environments. This can be achieved through user trajectory prediction, which can be represented by user mobility and handover predictions.

To achieve this, we propose an F-LSTM model to predict each user trajectory to consider future traffic demand. The F-LSTM model consists of local models for each UE and a global model of the system. To focus not only on predicting traffic demand but also on using the traffic predicted by a DQN agent, we set the local models to obtain the LSTM model to predict the traffic that consists of time-series data. The LSTM model, which is a type of recurrent neural network, can outperform other methods in the prediction of time-series data. While robust time-series forecasting models are available, the limited amount of trainable data and the need for quick learning and transmission to a central server in federated

learning led us to employ a stacked LSTM model in an auto-encoder architecture.

To obtain an optimal global model, each local model learns its local data and uploads its local weights. After uploading the local weights, the global model aggregates and averages the local weights at the end of each communication round. The optimal global model, acquired after the completion of all rounds, is then employed to predict the user trajectories in the DRL-based BS switching operation scheme. Subsequently, the DRL-based BS switching operation scheme determines the switching operations by considering the user traffic predicted for each predicted trajectory. An overview of the proposed scheme is depicted in Figure 1.

4.1 | F-LSTM-based traffic forecasting model

To consider future traffic demand and reduce the disadvantages of BS deactivation, we applied the F-LSTM model in multiple communication rounds to predict user trajectories. To reduce the local model's complexity to minimize the computational load, we predicted the next-located cell ID $\hat{d}_u^t$ of UE u as a trajectory. Then, the predicted next-located cell IDs of the UEs can be calculated for the predicted traffic demand as $\hat{D}^t$.

In the federated learning framework, the models are trained by the system and UEs during the communication rounds $\Lambda, \Lambda = \{1, 2, ..., \Lambda\}$. In the λth communication round, each participant UE downloads the global weights w_g^λ from the system. Then, each UE participant initializes the local weights w_u^λ with the global weights w_g^λ. The participant UEs train their local model to minimize the defined loss function, which is defined as

$$w_u^{\lambda *} = \arg\min\{L(w_u^\lambda)\}, \tag{13}$$

where $L(w_u^\lambda)$ is the categorical cross-entropy, that is, in the case of classification with local weights w_u^λ in communication round λ.

During the training process, we preprocess the input dataset with normalization to obtain the training data. The training data consist of four main features: the longitude and latitude of the GPS coordinates long_u^l, lat_u^l, timestamp ts_u^l, and current cell ID d_u^l. Thus, the input X_u for the local model of UE u can be expressed as

$$X_u = \{X_1, ..., X_L\}, \tag{14}$$

where $X_u = \{(\text{long}_u^l, \text{lat}_u^l), ts_u^l, d_u^l\}$, and L is the length of the input.

At every epoch, the validation loss is checked using the training data to prevent overfitting. After the training process, the testing process begins. The weights can be updated by backpropagation, which is more commonly used than the stochastic gradient descent algorithm, while the participant UEs upload their local weights.

With a set of uploaded local weights W^λ, $W^\lambda = \{w_1^\lambda, ..., w_u^\lambda\}$, the system aggregates the set of local weights W^λ as global weights $w_g^{\lambda+1}$ in communication round $\lambda+1$. The global weights $w_g^{\lambda+1}$ in communication round $\lambda+1$ can be aggregated according to the federated averaging algorithm (FedAvg) in McMahan and others [30], and it can be defined as

$$w_g^{\lambda+1} = \sum_{u \in \mathbf{U}} \omega_u w_u^\lambda, \tag{15}$$

where ω_u is the aggregation coefficient. When the λth communication round is terminated, the global model obtains aggregated weights $w_g^{\lambda+1}$. Finally, the participant UEs obtain the aggregated global weights at the beginning of each communication round.

4.2 | DRL-based BS switching operation scheme

RL is a well-known approach for solving MDP problems [31]. RL does not require complex mathematical models, and it solves the problem by learning the model and interacting with the environment. However, since environments can become more complex and have higher dimensions, the DRL method, which is a deep neural network adapted to Q-learning, has been studied. In addition, a continuously increasing environment can cause overestimation problems in the learning policy. To improve the DRL performance, double-deep Q-learning (DDQN) was developed to be robust and stable [32]. Thus, we apply the DDQN method to the BS switching scheme. The proposed DRL-based BS switching operation scheme is shown in detail in Figure 2.

In DDQN, the deep neural network represents the action and state spaces, and the state-action value function, that is, the Q-value function, is approximated. The DDQN applies the *online network* and *target network*, and the *online network* estimates the Q-value $Q(s, a)$ of the state-action pair $\langle s, a \rangle$, while the *target network* produces the approximated true value y. As a result, the DQN updates its weights to minimize the loss function $L(\theta)$, which is defined as follows:

$$L(\theta) = \mathbb{E}\left[(y - Q(s, a|\theta))^2\right], \tag{16}$$

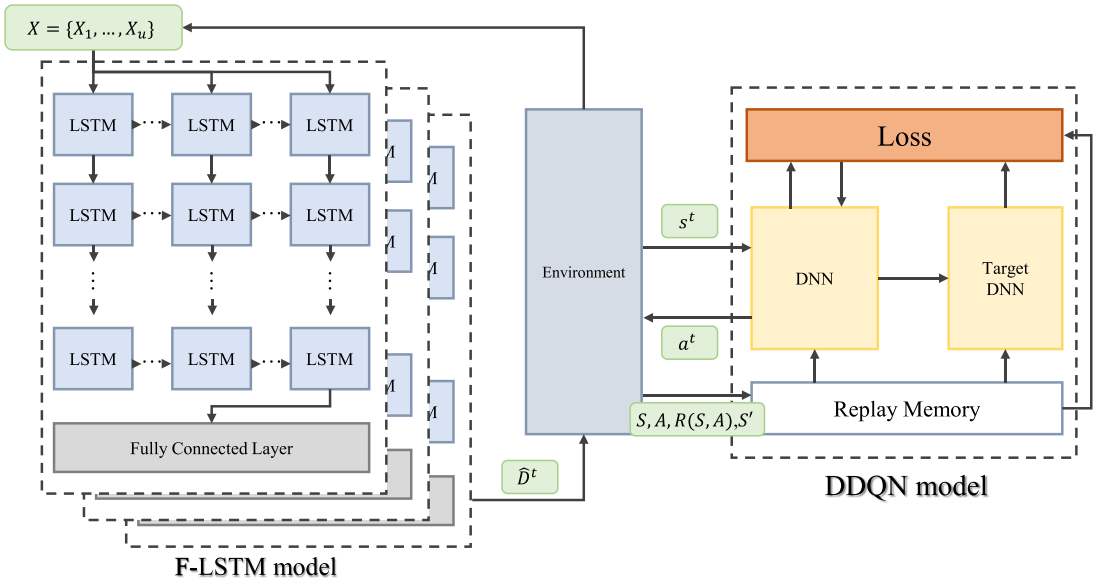

FIGURE 2 Proposed deep reinforcement learning (DRL)-based base station (BS) switching operation scheme.

where θ represents the weights of the *online network*. In addition, the approximated true value y can be calculated as

$$y = r(s, a) + \gamma^q Q\left(s', \arg\max_{a' \in A} Q(s', a'|\theta)|\theta^-\right), \quad (17)$$

where θ^- represents the weights of a *target network*.

The action from *online network* $Q(s, a|\theta)$ can be selected using a simple ϵ-greedy policy. The *target network* is a duplicate of the *online network* $Q(s, a|\theta)$. While updating the weights of the *online network*, the weights of the *target network* are fixed. To stabilize learning, an experience replay memory strategy that stores the transition of the experience is used. Therefore, the DDQN can train the model by sampling minibatches of the experience-replay memory. The next state s' can be obtained from both the *online network* and *target network* by calculating the optimal value $Q(s', a'|\theta)$. Then, the target value y is calculated using the discount factor γ^q and the current reward $r^t(s, a)$. Finally, the error is obtained from (17) and backpropagated to update the weights.

In our proposed DRL-based BS switching operation scheme, traffic demand $\hat{D}^t$ are aggregated with the predicted traffic data from the proposed F-LSTM model at the beginning of each time slot t. With the observed current state $s^t = \{\hat{D}^t, D^{t-1}, a_i^{t-1}, t_{a,i}^t\}$ from our environment, the agent decides on action a_i^t, which represents the active or SM operations for SBS i, and the system executes the actions for all SBSs. At the end of time slot t, the agent stores the experiences that include multiple experiences $(s_i, a_i, r(s_i, a_i), s')$ in the replay memory. Then, the agent randomly samples a mini-batch from the replay memory to update the network. The mechanism of the proposed scheme is summarized in Algorithm 1.

Algorithm 1 DRL-based BS switching operation scheme with traffic forecasting model

Initialize the replay memory B
Initialize *online network* $Q(s, a \mid \theta)$ with weights θ, *target network* $Q\left(s', a' \mid \theta^-\right)$ with weights θ^-.
for each episode $e \in \{1, 2, \ldots\}$ **do**
  for each time slot $t \in \{1, 2 \ldots, T\}$ **do**
    Calculate the traffic volume $\hat{D}^t$ from the predicted
    traffic data of each UE.
    Set current states $\{s_1^t, \ldots, s_i^t\}$ for the SBSs.
    Choose actions $\{a_1^t, \ldots, a_i^t\}$ for each state and execute
    the actions.
    Observe new states $\{s_1', \ldots, s_i'\}$ and
    rewards $\{r_1^t, \ldots, r_i^t\}$.
    Store transition $\left(s_i^t, a_i^t, r_i^t, s_i'\right)$ in B.
    Get a mini-batch sample randomly from
    replay memory B.
    Get y and update weights θ.
    Replace target weights $\theta^- = \theta$
    in every given *target network* replacement step.
  end for
end for

5 | EXPERIMENT RESULTS AND DISCUSSION

In this section, the performance of our proposed scheme is evaluated by comparing it with existing approaches. To set the simulation settings, a multicell HetNet environment is considered that consists of multiple MBSs and multiple SBSs. The coverage radii of the MBSs and SBSs were 1000 m and 100 m, respectively. The maximum transmission powers of the MBSs and SBSs were 20 W (43 dBm) and 1 W (30 dBm), respectively. The network parameters for the simulations were set according to 3GPP specifications [33, 34]. To consider the spatial and temporal characteristics of user mobility in urban areas, we utilized the mobility trace dataset from San Francisco [35]. The traffic applied to the dataset is shown in Figure 3. The other parameters and hyperparameters are summarized in Table 1.

To train the model of the proposed scheme, we used the TensorFlow platform with the Adam optimizer and a backpropagation algorithm [36]. We also use the Flower framework to build the simulation environment for federated learning using the FedAvg algorithm [37]. To predict the unlabeled trajectory data, we utilized an LSTM auto-encoder model, which consists of two encoder layers and two decoder layers. To avoid overfitting and accelerate training, we adopted a batch normalization enhancement method [38]. In the LSTM model, we set the learning rate, time step, and batch size to 0.001, 5, and 128, respectively. The architecture of our LSTM auto-encoder model is given in Figure 4. In the DRL model, we set the learning rate decay ratio, batch size, and target network update frequency to 0.98, 32, and 100 episodes, respectively.

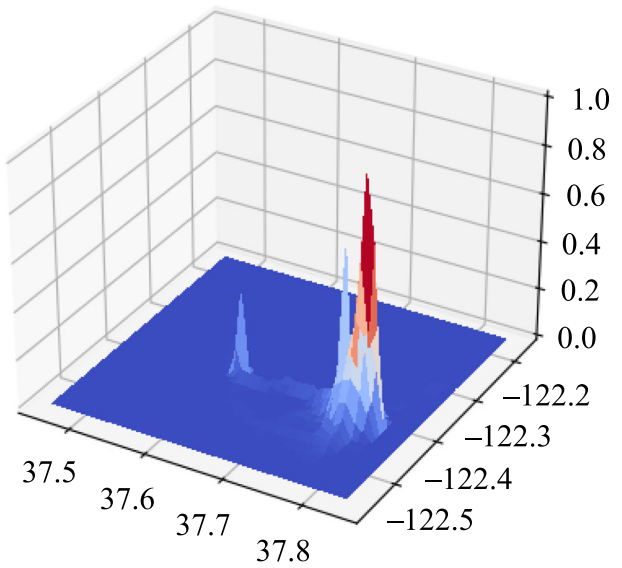

FIGURE 3 Normalized traffic demand according to geographical space in San Francisco.

TABLE 1 Simulation parameters.

Parameter	Value
Time slot (t)	30 min
Number of MBSs	15
Number of SBSs	45
Number of UEs	50, 100
Fixed energy consumption of active SBSs (π_1)	160 W
Fixed energy consumption of sleep SBSs (π_0)	24 W
Load-dependent power consumption of the SBSs (Δ)	216 W
Bandwidth (W)	20 MHz
Learning rate of the LSTM model	0.001
Time step of the LSTM model	5
Batch size of the LSTM model	128
Learning rate decay ratio in the DRL model	0.98
Target network update frequency of the DRL model	100 episodes
Batch size of the DRL model	32

Abbreviations: DRL, deep reinforcement learning; LSTM, long short-term memory; MBS, macro base station; SBS, small base station.

Layer		Output Shape
Encoder	LSTM, ReLu	(batch, timestep,128)
	BatchNormalization, Dropout	(batch, timestep,128)
	LSTM, ReLu	(batch, 64)
	BatchNormalization, Dropout	(batch, 64)
RepeatVector		(batch, timestep,64)
Decoder	LSTM, ReLu	(batch, timestep,64)
	BatchNormalization, Dropout	(batch, timestep,64)
	LSTM, ReLu	(batch, 128)
	BatchNormalization, Dropout	(batch, 128)
Dense		(batch, #SBS)

FIGURE 4 Summary of the long short-term memory (LSTM) auto-encoder model architecture.

To evaluate our proposed scheme, we compared it with existing approaches. We chose two BS switching operation approaches to achieve fair comparisons: the

DRL model with the convolutional-LSTM model and the optimization method LSTM model. The existing approaches are as follows.

1. **Full activation**: "Full Activation" was adopted to show the performance of BS switching operations. In Full Activation, all BSs always maintain the active mode.
2. **DQN with prediction (Compare1)**: "Compare1" was adopted to compare the performance of the considered rewards and prediction model [23]. In Compare1, the convolutional neural network and LSTM model, referred to as the C-LSTM model in this study, is used to capture the relationships of semantics and geography. It minimizes the cost that consists of the energy cost and service delay cost.
3. **Optimization with prediction (Compare2)**: "Compare2" was used to effectively compare DRL-based BS switching operations [22]. In Compare2, the LSTM model is used to predict future traffic demand by considering the positions and traffic of users. It formulates the BS switching operation problem as a Lyapunov optimization problem, and it solves the problem by selecting modes that minimize the objective function.
4. **Q-learning (Compare3)**: "Compare3" is a model that uses Q-learning to make decisions on BS switching operations and is employed to demonstrate the effectiveness of the proposed algorithm [15]. This method uses dropping rate and delay constraints to minimize QoS degradation caused by frequent handovers. In contrast to the methods proposed in our paper and other related papers, it controls power using four modes, not just simple ON/OFF, but with three levels of SM.

The proposed F-LSTM model trains the LSTM models using a federated learning algorithm; thus, the hyperparameters for the LSTM model must be set for fast convergence and accurate prediction. To set the step size for the proposed LSTM model, we simulated the LSTM model with various time-step sizes. The time-step size is an important factor for improving the prediction score of the LSTM model. Therefore, the selection of the optimal time step affects the performance of the prediction model. The user trajectory has spatiotemporal dependencies, and the user trajectory in urban areas can change in a number of different cases. As a result, setting the time-step value can determine the length of the trajectory that should be considered in the prediction. In Figure 5, the accuracy is represented with time step values of 1, 3, 4, 5, and 10 to determine the optimal value of our F-LSTM model. The proposed model with time steps 1, 3, 4, 5, and 10 converged with accuracies of approximately 74%, 77%, 87%, 82%, and 69%, respectively. The proposed model with a time step of 5 outperformed the other models. This means that the time step should be set to a value that is neither too long nor too short. Therefore, we set the time step value to 5 because it is the optimal value for the proposed model.

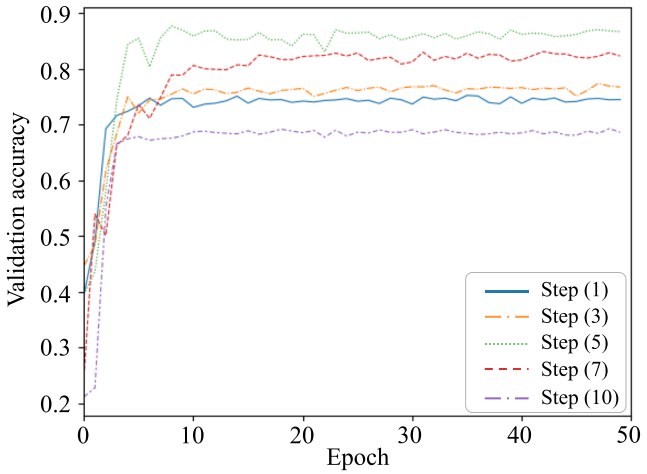

FIGURE 5 Comparison of accuracies with various step sizes.

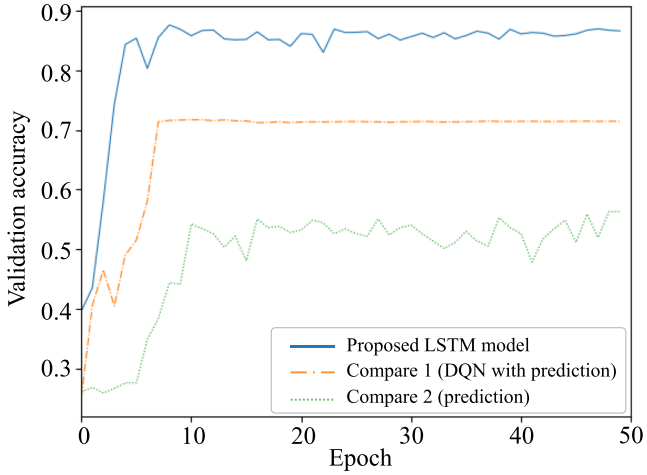

FIGURE 6 Comparison of accuracies obtained using the proposed scheme and existing approaches.

To show the prediction performance of our proposed LSTM model, we compared the prediction accuracy of the proposed LSTM model and existing models in Figure 6. As illustrated in Figure 6, the proposed LSTM model converges faster and obtains higher accuracy than other existing prediction models in the training processes. The proposed LSTM model, LSTM model in Compare1, and LSTM model in Compare2 converged after approximately 6, 8, and 10 training epochs, with accuracies of

approximately 87%, 71%, and 56%, respectively. We can see that the proposed LSTM model after training has an accuracy that is higher than that of Compare1 and Compare2 by approximately about 22.5%–26.7%. The user trajectory data have historical characteristics; therefore, our proposed LSTM model, which was constructed with the auto-encoder architecture, is better able to train using the important features of the input data.

In federated learning, the local models upload their trained weights to the global model after every local update epoch. If the local update is too low or too high, the global model converges very slowly or cannot converge. Therefore, the local update epoch is the most important hyperparameter affecting the convergence time. In Figure 7, we compare the validation losses with various local update epochs to determine the optimal value. In Figure 7, we can see that the validation loss of epoch 1 converges more slowly than other epoch values. This indicates that an epoch that is too low will update using weights that have not been sufficiently learned, thus disturbing the convergence of the global model. Therefore, we set the number of local update epochs to 10 to achieve fast convergence.

In Figures 8 and 9, the comparisons of average energy efficiencies for 1000 episodes on scenarios containing 50 and 100 UEs, respectively, are displayed. We can see that in the figures, there are obvious increments in energy efficiency when using the BS switching operation. With the proposed scheme, the energy efficiency increases further compared with other existing

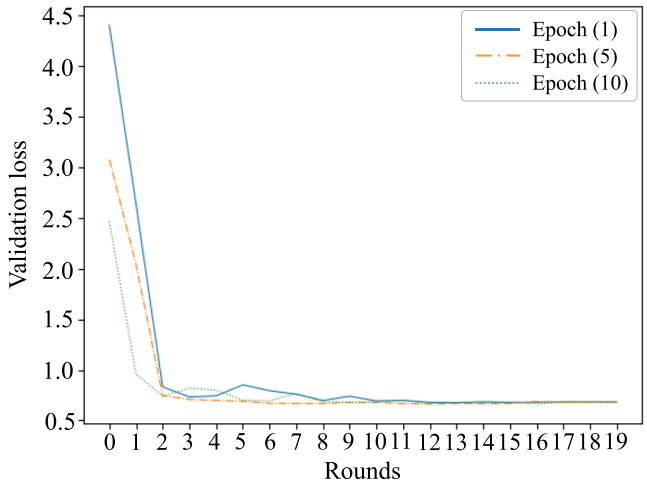

FIGURE 7 Comparison of validation losses with respect to various local update epochs.

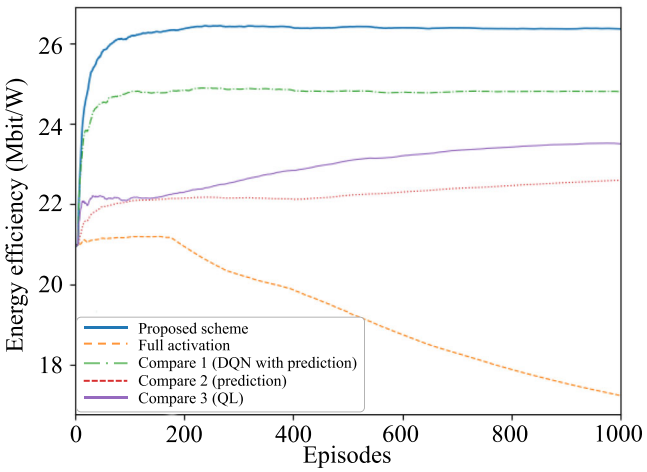

FIGURE 8 Comparison of average energy efficiencies with 50 UEs.

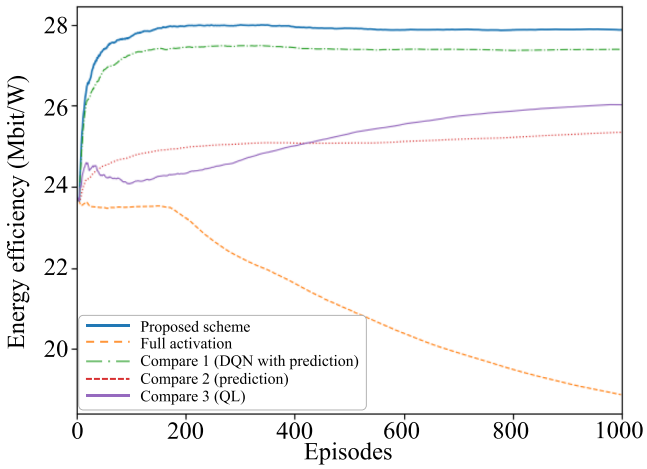

FIGURE 9 Comparison of average energy efficiencies with 100 UEs.

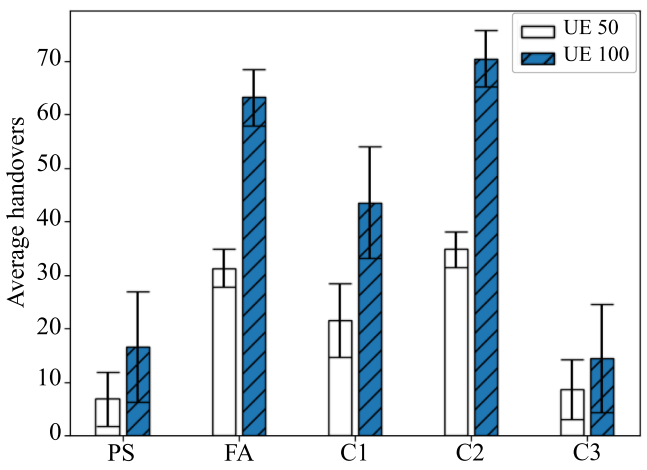

FIGURE 10 Comparison of average handovers with various UE scenarios.

approaches. Specifically, our proposed scheme in the 50-UE scenario outperformed Full Activation by 36%, Compare1 by 6%, Compare2 by 18%, and Compare3 by 15%, respectively. The proposed scheme in the 100-UE scenario performed better than Full Activation by 31.7%, Compare1 by 1.8%, Compare2 by 11%, and Compare3 by 11%, respectively. Moreover, high-traffic areas were created by increasing the number of episodes and decreasing the energy efficiency of Full Activation. However, the BS switching operation schemes can ensure energy efficiency. In addition, the proposed scheme converges to higher values. This indicates that the proposed scheme can learn the optimal BS switching policy better than Compare1. Compare3 outperformed Compare2, which does not utilize the RL method, but it exhibited lower energy efficiency than the DQN-based Compare1. In realistic environments, there are more states to consider, and ON/OFF actions that do not take future traffic demands into account may result in users not receiving service from the optimal BS.

To show the impact of the designed rewards on the reduction of QoS degradation from handovers, we compare average handovers for the 50- and 100-UE scenarios in Figure 10. We refer to the proposed scheme, Full Activation, Compare1, and Compare2 as PS, FA, C1, and C2, respectively. We can see that in the figure, there are significant reductions in handovers when the proposed scheme is utilized. When comparing the proposed scheme with Compare1, both approaches reduced handovers by considering the QoS degradation. However, the proposed scheme reduce more handovers by using a reward that considers the QoS degradation from handovers. The proposed scheme determines the active mode for the SBSs that perform with low energy efficiency and cause fewer UE handovers. As a result, it achieved the lowest average handover with increased energy efficiency. In addition, Compare3 utilized the drop rate and delay constraints as metrics to minimize handovers, obtaining results that are comparable to those of our proposed scheme.

To show the impact of the proposed scheme, we compared the average SINRs using boxplots in Figure 11. This figure includes the smallest and largest observations, lower and upper quartiles, and the median. The variance and skew in the distributions of the SINRs are shown. Regarding the performance results of the SINRs, the proposed scheme performed better by approximately 39%, 10%, 26%, and 24% than the Full Activation, Compare1, Compare2, and Compare3 methods, respectively. This is because the proposed scheme determines the modes of the SBSs by considering the reduction in QoS degradation due to handovers. The users associated with BSs are serviced with the optimal SINR in our environment, and hence, the handover decreases the SINRs of the users because of suboptimal BSs. Nevertheless, Compare3 also aimed to minimize handovers, and the average SINR was lower that that of both our proposed method and Compare1. This is because we considered a switching penalty to prevent frequent mode switching, which can cause critical QoS degradation and additional energy consumption. Therefore, Figures 10 and 11 prove that the proposed scheme tends to determine the modes of the SBSs that reduce QoS degradation.

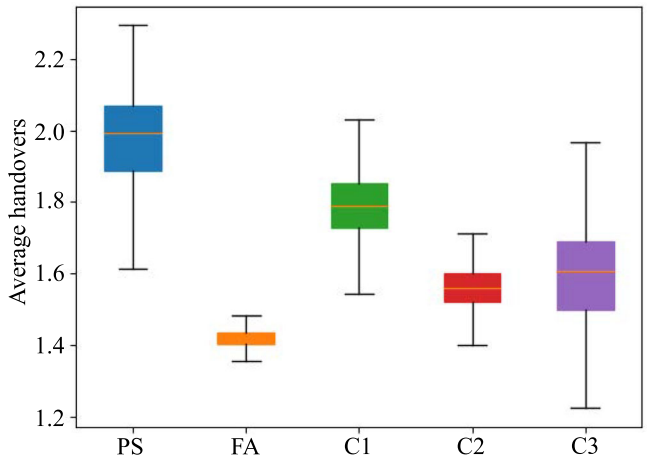

FIGURE 11 Comparison of average signal-to-interference noise ratios (SINRs) with 50 UE scenarios.

6 | CONCLUSION

In this paper, we introduced a DRL-based BS switching scheme with the F-LSTM traffic prediction model to enhance energy efficiency. Our F-LSTM model predicts future traffic by learning user trajectories through federated learning, ensuring rapid convergence and private training. The DRL model uses these predictions to optimize SBS operation modes, minimizing energy consumption and QoS degradation. We considered factors such as energy use, QoS degradation (data rate and handover impact), switching penalty, and SBS active time to reduce repetitive handovers. Simulation results demonstrate the enhanced energy efficiency when maximizing energy savings while minimizing QoS degradation.

In real mobile networks, varying the amount of user data for training can lead to communication delays in federated learning due to data uploading. To improve federated learning performance, we will extend our approach by adjusting each user's training data volume. Additionally, we will enhance BS switching by allocating transmission power based on diverse communication conditions for individual users.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Hyebin Park https://orcid.org/0000-0003-0317-4017

REFERENCES

1. Ericsson, Ericsson mobility report, 2020. https://www.ericsson.com/4ac68e/assets/local/reports-papers/mobility-report/documents/2020/june2020-ericsson-mobility-report.pdf
2. H. Holtkamp, G. Auer, S. Bazzi, and H. Haas, *Minimizing base station power consumption*, IEEE J. Sel. Areas Commun. **32** (2014), no. 2, 297–306.
3. C. Liu, B. Natarajan, and H. Xia, *Small cell base station sleep strategies for energy efficiency*, IEEE Trans. Veh. Technol. **65** (2015), no. 3, 1652–1661.
4. E. Oh and B. Krishnamachari, *Energy savings through dynamic base station switching in cellular wireless access networks*, (IEEE Global Telecommunications Conference GLOBECOM 2010, Miami, FL, USA), 2010, pp. 1–5.
5. E. Oh, K. Son, and B. Krishnamachari, *Dynamic base station switching-on/off strategies for green cellular networks*, IEEE Trans. Wirel. Commun. **12** (2013), no. 5, 2126–2136.
6. S. Song, Y. Chang, X. Wang, and D. Yang, *Coverage and energy modeling of HetNet under base station on-off model*, ETRI J. **37** (2015), no. 3, 450–459.
7. A. Kumar and C. Rosenberg, *Energy and throughput trade-offs in cellular networks using base station switching*, IEEE Trans. Mob. Comput. **15** (2016), no. 2, 364–376.
8. M. Feng, S. Mao, and T. Jiang, *Base station on-off switching in 5G wireless networks: Approaches and challenges*, IEEE Wireless Commun. **24** (2017), no. 4, 46–54.
9. W.-T. Wong, Y.-J. Yu, and A.-C. Pang, *Decentralized energy-efficient base station operation for green cellular networks*, (IEEE Global Communications Conference (GLOBECOM), Anaheim, CA, USA), 2012, pp. 5194–5200.
10. Y. Yang, Z. Liu, H. Zhu, X. Guan, and K. Y. Chan, *Energy minimization by dynamic base station switching in heterogeneous cellular network*, Wireless Netw. **2022** (2022), 1–16.
11. R. Li, Z. Zhao, X. Chen, J. Palicot, and H. Zhang, *Tact: a transfer actor-critic learning framework for energy saving in cellular radio access networks*, IEEE Trans. Wireless Commun. **13** (2014), no. 4, 2000–2011.
12. R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*, MIT press, 2018.
13. A. El-Amine, H. A. Haj Hassan, M. Iturralde, and L. Nuaymi, *Location-aware sleep strategy for energy-delay tradeoffs in 5G with reinforcement learning*, (IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey), 2019, pp. 1–6.
14. A. El-Amine, M. Iturralde, H. A. Haj Hassan, and L. Nuaymi, *A distributed q-learning approach for adaptive sleep modes in 5G networks*, (IEEE Wireless Communications and Networking Conference, Marrakesh, Morocco), 2019, pp. 1–6.
15. A. El Amine, J.-P. Chaiban, H. A. H. Hassan, P. Dini, L. Nuaymi, and R. Achkar, *Energy optimization with multi-sleeping control in 5g heterogeneous networks using reinforcement learning*, IEEE Trans. Netw. Service Manag. **2022** (2022), 1–1.
16. M. Masoudi, M. G. Khafagy, E. Soroush, D. Giacomelli, S. Morosi, and C. Cavdar, *Reinforcement learning for traffic-adaptive sleep mode management in 5G networks*, (IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK), 2020, pp. 1–6.
17. Y. Li, *Deep reinforcement learning: an overview*, arXiv preprint, 2017, DOI: 10.48550/arXiv.1701.07274. arXiv preprint arXiv: 1701.07274.
18. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, and G. Ostrovski, *Human-level control through deep reinforcement learning*, Nature **518** (2015), no. 7540, 529–533.
19. J. Ye and Y.-J. A. Zhang, *Drag: deep reinforcement learning based base station activation in heterogeneous networks*, IEEE Trans. Mob. Comput. **19** (2019), no. 9, 2076–2087.
20. H. Ju, S. Kim, Y. Kim, and B. Shim, *Energy-efficient ultra-dense network with deep reinforcement learning*, IEEE Trans. Wireless Commun. **21** (2022), no. 8, 6539–6552.
21. Y. Zhu and S. Wang, *Joint traffic prediction and base station sleeping for energy saving in cellular networks*, (ICC 2021-IEEE International Conference on Communications, Montreal, Canada), 2021, pp. 1–6.
22. G. Jang, N. Kim, T. Ha, C. Lee, and S. Cho, *Base station switching and sleep mode optimization with lstm-based user prediction*, IEEE Access **8** (2020), 222711–222723.
23. Q. Wu, X. Chen, Z. Zhou, L. Chen, and J. Zhang, *Deep reinforcement learning with spatio-temporal traffic forecasting for data-driven base station sleep control*, IEEE/ACM Trans. Netw. **29** (2021), no. 2, 935–948.
24. S. Kumari and B. Singh, *Data-driven handover optimization in small cell networks*, Wireless Netw. **25** (2019), no. 8, 5001–5009.
25. K. Tan, D. Bremner, J. Le Kernec, Y. Sambo, L. Zhang, and M. A. Imran, *Intelligent handover algorithm for vehicle-to-network communications with double-deep q-learning*, IEEE Trans. Veh. Technol. **71** (2022), no. 7, 7848–7862.
26. K. Qi, T. Liu, and C. Yang, *Federated learning based proactive handover in millimeter-wave vehicular networks*, (15th IEEE International Conference on Signal Processing (ICSP), Beijing, China), 2020, pp. 401–406.
27. J. Feng, C. Rong, F. Sun, D. Guo, and Y. Li, *PMF: A privacy-preserving human mobility prediction framework via federated learning*, Proc. ACM on Interact., Mobile, Wear. Ubiquitous Technol. **4** (2020), no. 1, 1–21.
28. C. Koetsier, J. Fiosina, J. N. Gremmel, J. P. Müller, D. M. Woisetschläger, and M. Sester, *Detection of anomalous vehicle trajectories using federated learning*, ISPRS Open J. Photogr. Remote Sens. **4** (2022), 100013.
29. K.-C. Chang, K.-C. Chu, H.-C. Wang, Y.-C. Lin, and J.-S. Pan, *Energy saving technology of 5G base station based on internet of things collaborative control*, IEEE Access **8** (2020), 32935–32946.
30. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, *Communication-efficient learning of deep networks from decentralized data*, (Artificial Intelligence and Statistics. PMLR), 2017, pp. 1273–1282.
31. P. Geibel, *Reinforcement learning for MDPS with constraints*, (European Conference on Machine Learning, Berlin, Germany), 2006, pp. 646–653.

32. H. Van Hasselt, A. Guez, and D. Silver, Deep reinforcement learning with double Q-learning, (Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA), 2016.
33. 3GPP, Evolved universal terrestrial radio access (E-UTRA): Further advancements for E-UTRA physical layer aspects. 36.814. 3rd Generation Partnership Project (3GPP), 2017. Version 9.2.0.
34. 3GPP, Technical specification group radio access network: Small cell enhancements for E-UTRA and EUTRAN- Physical layer aspects. 36.872. 3rd Generation Partnership Project (3GPP), 2013. Version 12.1.0.
35. MatthiasGrossglauser Michal Piorkowski Natasa Sarafijanovic Djukic, Crawdad dataset epfl/mobility, 2009. (v. 2009 2024).
36. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. Software available from https://www.tensorflow.org/
37. D. J. Beutel, T. Topal, A. Mathur, X. Qiu, T. Parcollet, and N. D. Lane, *Flower: A friendly federated learning research framework*, arXive preprint, 2020, DOI: 10.48550/arXiv.2007.14390
38. S. Ioffe and C. Szegedy, *Batch normalization: accelerating deep network training by reducing internal covariate shift*, (International Conference on Machine Learning. PMLR, Lille, France), 2015, pp. 448–456.

AUTHOR BIOGRAPHIES

Hyebin Park received her B.S. degree and Ph.D. degree in IT engineering from Sookmyung Women's University, Seoul, Republic of Korea, in 2017 and 2022, respectively. She is currently a researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea. She is interested in wireless networks, energy-efficient networks, 6G and optimization with reinforcement learning and federated learning.

Seunghyun Yoon received his B.S., M.S., and Ph.D. degree in Industrial Engineering from Sungkyunkwan University, Seoul, Korea, in 1991, 1993, and 1997. He is currently with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. He is interested in computer networks, cloud computing, and optimization in networks and computers.

How to cite this article: H. Park and S. H. Yoon, *Deep reinforcement learning for base station switching scheme with federated LSTM-based traffic predictions*, ETRI Journal (2024), 1–13. DOI 10.4218/etrij.2023-0065