International Journal of Basic and Applied Sciences, 14 (1) (2025) 291-303

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/5y394g91
Research paper

Advancing communication networks: integrating enhanced knowledge mapping with hybrid deep recurrent neural networks for dynamic spectrum access

Yelithoti Sravana Kumar [1] *, Tapaswini Samant [2], Swati Swayamsiddha [2]

[1] Research Scholar, School of Electronics, KIIT Deemed to be University, Bhubaneswar, India
[2] Associate Professor, School of Electronics, KIIT Deemed to be University, Bhubaneswar, India
**Corresponding author E-mail: 1981083@kiit.ac.in*

Received: April 17, 2025, Accepted: May 9, 2025, Published: May 13, 2025

Abstract

By merging knowledge mapping using Hybrid Deep Recurrent Neural Networks (RNNs), our suggested method maximizes dynamic spectrum access in diverse networks. Optimizing the assignment of spectrum resources while enabling different device characteristics and network circumstances is our approach to addressing the issues of spectrum allocation in different network situations. Improved processing capabilities at the network's edge enable real time monitoring of patterns in spectrum consumption. In order to dynamically allocate spectrum according to user demands and network dynamics, our dynamic spectrum access technology employs reinforcement learning algorithms. Hybrid Deep RNNs take use of both deep learning as well as recurrent neural networks to enhance feature extraction and behavioral dependence modeling in spectrum data. In order to guarantee the system's reliability and resilience in real world applications, assessment indicators are used to analyze its performance and efficiency. Consistent with our hypothesis, the results demonstrate significant gains in spectrum utilization effectiveness and allocation accuracy, validating our approach to maximizing resource consumption and facilitating faultless functioning in diverse network settings.

Keywords: *Hybrid Deep Recurrent Neural Networks; Deep Learning; Knowledge Map; Heterogeneous Networks; Dynamic Spectrum Access; Reinforcement Learning; Real Time Analysis.*

1. Introduction

Research into how to make heterogeneous wireless networks work together peacefully has recently exploded in popularity, thanks to the learning based dynamic spectrum access strategy's impressive capacity to achieve optimal resource allocation in next generation wireless communication systems [1]. Every day brings new developments in the fast-paced world of mobile device manufacturing and the explosion of wireless applications. In order to handle such heavy demand while enhancing QoS, it is necessary to leverage wireless spectrum employing various Radio Access Devices and to create unique network selection strategies [2]. The development of spectrum access methods for efficient distribution of spectrum holes is a fundamental technology in the IoT, and its integration with cognitive radio may help alleviate spectrum scarcity, which is a problem for widespread IoT implementation. Secondary users in a Cognitive Radio IoT network have a hard time avoiding interferences and getting to the spectrum fast because of the partly viewable channels and the rising user count [3].
Flexibly providing effective spectrum access service and maintaining stable operation in extremely dynamic environments are two capabilities of reconfigurable wireless networks. In order to provide priority queuing for both main and secondary users, the evolution behavior is captured by modeling the spectrum Markov state [4]. The issue of underutilized spectrum may be addressed by the use of dynamic spectrum access (DSA). The majority of current DSA research, however, disregards the prospect of spectrum sharing owing to users' mobility in favor of analyzing the effect of secondary users' spectrum sensing findings on opportunistic access to spectrum gaps [5]. The need for wireless spectrum is growing in importance due to factors such as the fast expansion of several technologies, the proliferation of wireless devices, and the advancements in 5G technology. 5G communication systems have recently emerged as a result of the fast advancements in wireless communication technology.
In CRN, spectrum detection is the main paradigm used to dynamically access the spectrum. Spectrum sensing scenario has been the focus of many academics, who have developed a variety of approaches to the problem of sharing spectrum resources. In order to detect the signal, most approaches use decision statistics that compare and contrast signal and noise characteristics [6]. When it comes to aeronautical interactions, legacy systems often make scant use of the frequency bands that were previously allotted to them. It is important for new systems like the L-band Digital Aeronautical Communications Network Air-Air to coexist with older systems and avoid interfering too much [7]. For digital inclusion, dependable and inexpensive internet connection is essential for global connectedness. A number of technologies exist that may contribute to more cheap connection, and shared spectrum solutions are only one of them. Upcoming dynamic heterogeneous secondary user systems will share the accessible spectrum (channels) from Primary Users. There is a higher chance of wasteful spectrum

and power utilization and a bad SU experience due to the increased difficulty in coordinating these fleeting resources in dynamic SU network circumstances.

This is why we're using IEEE 802 [8], propagation models based on topography, and reinforcement learning. Because of the problem of multiple access that arises in heterogeneous wireless systems when several nodes attempt to use the same wireless channel, a Medium Access Control method is required to manage the data transmission of all the nodes using that channel [9]. To handle the need for an exceptionally high data throughput, the 5G system has used mmWave. Nevertheless, network densification is necessary due to the limited coverage caused by the severe propagation characteristics of mmWave signals. In light of this, the 3GPP has developed the Integrated Access and Backhaul design to facilitate the deployment and operation of networks at low cost. While most network designs rely on wired backhaul lines to transmit data, IAB employs wireless backhaul links instead [10]. As a component of 5G technology, the third generation partnership project introduced the Narrowband Internet of Things in Release-13. Low Power Wide Area Networks are what these networks are officially called. While transmitting messages over great distances, it consumes less electricity. Automating processes in homes, industries, and the environment is possible with the NB-IoT.

According to the comparison between NB-IoT and normal Long Term Assessment Machine Type Communication, and the uplink repetition is 128. The Maximum Coupling Loss is severely exceeded by 164 dB [11]. A broad range of latency, reliability, energy efficiency, etc., needs have emerged in 5G mobile technology due to the proliferation of the IoT. An increase in spectrum efficiency with a decrease in power consumption is necessary for a network of this size. To improve system efficiency, numerous users are merged into one frequency using the non-orthogonal multiple access technology. For heterogeneous IoT networks based on NOMA, energy efficient resource allocation issue has been introduced [12].

By allowing SUs to strategically access the licensed bands while main users are idle, dynamic spectrum access has been seen as a potential solution to the severe spectrum scarcity issue in 6G networks. Partially sensing the spectrum with an appropriate sensing window is seen as a viable solution for identifying available idle bands in the presence of hardware constraints. It is worth noting that the number of accessible bands might be determined by the SW selection, and that network efficiency after access could be used to direct the SW selection [13]. Integrating energy harvesting, cognitive radio, as well as nonorthogonal multiple access (NOMA) techniques can improve the energy as well as spectral efficiency of IoT networks. Unmanned aerial vehicles are a quick and adaptable way to improve coverage performance [14]. Due to the acute spectrum limitation and massive traffic demands, it is vital for the 5G and beyond systems to exploit and share unlicensed spectrum resources across cellular and WiFi networks. The current method is not flexible enough to handle wireless network traffic in dynamic environments with varied quality of service requirements, leading to high consensus overhead and poor QoS levels, even though distributed consensus using block chain has been contemplated as a means to achieve efficient and fair spectrum sharing [15]. The proliferation of wireless gadgets has led to a jammed spectrum. In both licensed and unlicensed bands, dynamic spectrum access is gaining traction as a way to provide secondary users white space. Having the ability to sense the spectrum is crucial for dynamic spectrum access [16].

Users without the proper license may access licensed providers' radio spectrum and use the designated channel for data transfer in vehicular communication while using cognitive radio to help with the Internet of Vehicles. Cognitive IoV networks may make better use of their spectrum resources by improving channel access [17]. Determining whether a device should transmit or receive data from a base station is the primary task of the first synchronization phase in wireless network architecture. Typically, a RA method is used to dynamically distribute and allot radio resources, powering this process. The IoT has been expanding at a dizzying rate in recent years, with new services and varying needs popping up all over the globe [18]. This has been particularly true in the telecommunications sector. Cognitive radio provides a great deal of efficiency and versatility in the application of radio spectrum, and it is a pioneering technology in the realization of DSA [19]. Through the use of DSA, secondary users are able to take advantage of principal user channels during their brief times of inactivity, thus increasing spectrum usage. Due to the need for comprehensive network state information, previous research on spectrum access techniques that maximize utility may not be applicable. Q-learning and other model free RL based approaches provide encouraging adaptive solutions that don't need full network knowledge [20].

1.1. Contributions of the work

One promising solution to the problem of inefficient spectrum use is dynamic spectrum access, or DSA. The basic "sense-and-avoid" method and other conventional DSA techniques fall short in many cases when it comes to providing adequate performance. So, instead of using oversimplified methods, a combination of sensing and deep reinforcement learning (DRL) has shown promise. In contrast to more conventional forms of reinforcement learning, DRL does not need the costly and time-consuming process of explicitly estimating transition probability matrices. Also, to find the best channel access policy via online learning, Deep Recurrent Q-Networks (DRQN) have been suggested, as many learning methods fail to solve the online Partially Observable Markov Decision Processes (POMDP). The suggested network model may effectively close the spectrum gaps in the communication of current users through acquiring the access mode from past data.

Long Short-Term Memory (LSTM) as well as Gated Recurrent Unit (GRU) are the two components of a proposed hybrid Recurrent Neural Network (RNN) that would serve this function. Specifically, it allows the cognitive user to simultaneously communicate across many channels, for instance via multi carrier technology. This allows for optimal use of the spectrum resource and maximum aggregate throughput of the HetNet. According to the simulation findings, the suggested algorithm outperforms the traditional methods (the Hybrid LSTM-GRU methods for unsaturated traffic and the Whitley index policy for saturated traffic) in terms of throughput. Furthermore, it demonstrates the improved resilience to time varying communications by attaining a near optimal outcome in dynamic situations with shifting main users.

2. Related works

By merging deep reinforcement learning with a memory module, the authors of [21] suggested a multiple access control system that might achieve high network throughput. To make use of the data from changing environmental observations at each time step, they develop the bidirectional gated recurrent unit for deep Q-learning. Moreover, they implement the strategy in a real-world highway dataset situation, where DQL nodes compete for the same wireless channel. The findings show that the suggested method does not need any complicated mechanisms or priors to learn an optimum policy. In addition, authors take into account practical scenarios where the uplink traffic patterns of nodes on a highway stretch are either saturated or unsaturated, as well as the online training tactics of the DQL node in proximity to roadside amenities. Considering the unique properties of various applications, the authors of [22] presented a distributed architecture for

dynamic network allocation at the edge and resource allocation at the RAN level. More specifically, their system makes use of a deep Multi Agent Reinforcement Learning method to optimize the experience quality of the edge nodes, while also prolonging the nodes' battery life and making use of adaptive compression strategies. Data can be efficiently and cheaply sent to the cloud from the network's edge nodes which have multi-RAT capabilities using their architecture, all while meeting the quality-of-service criteria of the many apps that use it. They found that their method reduces energy usage, delay, and cost compared to state-of-the-art network selection methods.

The study of dispersed heterogeneous wireless networks is explored in [23], along with a spectrum access strategy that investigates the equitable allocation of channels among users according to their request levels. Taking into account that various users' request levels have varying priorities that is, that certain packets are more essential to the user than others help bring the model closer to reality. A reinforcement learning-based approach is suggested to estimate the network state using the history of successful packet transmissions in the network. That is necessary since the issue is dispersed and users cannot coordinate among each other prior to transmission. Because of the adaptable nature of the reinforcement learning approach, their system may function in diverse environments with users using different media access control protocols. A Cognitive Radio network that makes use of multi agent Deep Reinforcement Learning for dynamic spectrum access is described in full in the study [24].

In order to establish the optimal time and frequency for transmission, every node in the network uses a neural network model. To reduce the impact of sluggish online training time, the algorithms are trained offline in simulation. They also suggest using entropy-based exploration to dynamically find out when the wireless network needs further training. They provide over the air measurement data for the throughput and channel utilization obtained from a large-scale software defined radio testbed, in contrast to earlier work that has only focused on comparable methods in theory and simulation. DRL based DSA techniques abound in the literature, although the vast majority of them have focused on relatively straightforward heterogeneous cognitive radio networks [25]. They provide a distributed DSA approach for complicated heterogeneous CRN that combines two common kinds of DRL networks: ResNet and Long short-term memory. Achieving maximum channel usage while keeping SU interference to PU to a minimum is the goal of the method. In order to enhance the capacity to forecast the spectrum state, the LSTM is implemented to extract the temporal properties of the sequence of historical spectrum data. However, ResNet enhances training accuracy while simultaneously fixing the performance deterioration issue with deep neural networks induced by network depth. By regulating the interference between SU and PU, the algorithm is able to greatly enhance spectrum usage, according to the simulation findings.

Explores the DSA approach for multiuser wireless networks with poor feedback in [26]. In order to send packets with a certain probability, each user chooses an orthogonal channel inside a specific time window. Depending on their local observations, users who have sent packets will get an ACK signal in the following time slot. Because wireless networks are always evolving, it seems sense to use a combination of approaches to ensure successful DSA. In order to maximize the utility of the network, that research seeks to propose a distributed system based on Deep Reinforcement Learning. While it is generally believed that the feedback packet received is always accurate in traditional DRL frameworks, noise and other interference may cause ACK packets to be lost or damaged in wireless networks. To solve the hidden node issue that arises when aeronautical communication protocols coexist, a new deep learning-based solution was suggested in [27].

Distance Measuring Equipment is an older technology that has to share spectrum with the more current L-band Digital Aeronautical Communications technology in Air-Air mode. All additional planned aviation systems that will coexist with DME in the spectrum must avoid interfering with it since it is safety critical. To get beyond the restrictions of static methods and get dynamic spectrum access, cognitive radio approaches have been suggested as a solution for LDACS A/A recently. That was accomplished by training a Recurrent Neural Network to anticipate periods of inactivity inside the respective frequency ranges, where the two systems function. One possible approach is to use trends in DME's spectrum access to forecast the number of idle resources. Read about a dueling deep recurrent Q-network (Dueling DRQN)–based deep reinforcement learning technique for flexible multichannel access in heterogeneous wireless systems in [28].

In particular, authors think about the case when several diverse users, each using a distinct MAC protocol, share a number of separate channels. Achieving high throughput by using the unused channels is the purpose of the intelligent node's channel access strategy learning. The intelligent node has two major obstacles: (i) neither the spectrum environment nor the actions of other nodes are known in advance; and (ii) only a portion of the spectrum environment can be seen, and (iii) the temporal dynamics of the spectrum states are complicated. By including the long short-term memory layer into the deep Q-network, they may aggregate historical data and capture the underlying temporal characteristic in the heterogeneous networks, hence overcoming the aforementioned problems.

The use of deep reinforcement learning in cellular vehicle to vehicle heterogeneous networks is the subject of research in [29]. In that setup, time division duplex wireless networks house macro cell users, voice to data clusters (V-Clusters), and V2V nodes, all of which share the available spectrum via the cellular uplink. The access patterns that M-UEs and V-Clusters use are orthogonal. Due to a lack of knowledge about other users' access habits, V2V nodes have the challenge of obtaining an appropriate approach to reuse others users' resources for access while minimizing substantial disturbance. In order to improve spectrum access without previous knowledge, DRL technologies is used to train V2V nodes in an unsupervised manner. In order to maximize the total throughput of the HetNets, the V2V nodes are programmed to intelligently choose appropriate frames for spectrum access using a hybrid spectrum access algorithm called D2HSA, which is based on a double deep Q-network.

The authors of [30] suggest a method that uses dynamic spectrum access in conjunction with multi-hop forwarding via vehicles. The first protocol is G-hop, which stands for group based multi-hop broadcasting. G-hop uses the depth first search technique to group cars that have similar features, such speed and communication distance. Due to the priority-based message forwarding within and between groups, the number and range of relay vehicles (also known as channel competitors) are limited. In addition, they accomplish dynamic spectrum access by using deep reinforcement learning methods. Using deep reinforcement learning, they develop a GOEA-based global optimization method. To learn the time varying process, GOEA suggests a network structure that combines recurrent neural networks and deep Q-networks. To maximize the global utility, it applies a reward technique.

Optimal throughput and fair spectrum access may be achieved in [31]. In addition, they bring forth a learning based approach for dynamic spectrum access that lets secondary users tweak their settings to choose the best access strategy for making the most of the network's throughput. With the use of a recurrent neural network and prioritized experience replay, the Dueling Deep Q-Network may increase the rate of convergence. When compared to the current Dueling DQN and DQN systems, the suggested RDRL strategy achieves better convergence speeds and channel throughputs, according to extensive testing data. Investigates Deep Recurrent Q-Networks as a kind of DRL for Dynamic Spectrum Access under incomplete observations in [32]. Here, they zero in on a situation where there are a number of distinct channels and a wide variety of Primary Users (PUs). Two major obstacles in their issue formulation are the assumption that their DRQN node is unaware of the other nodes' behavior patterns and the use of past observations to anticipate the future state of the channel. Learning a channel access method that maximizes channel use while minimizing collisions is the objective of the DRQN. They demonstrate via comprehensive simulation findings that a DRQN based strategy can manage a range of communication settings, including dynamic ones, with suitable state, action, and reward definitions.

Explores a novel DRL based medium access control protocol for multi-channel heterogeneous networks (HetNets), called multi-channel deep-reinforcement learning multiple access, in [33]. Their focus here is on HetNets, in which many radio networks use distinct media access control protocols to communicate with a shared access point via separate wireless channels. The MC-DLMA node has three main obstacles: (i) lack of prior information about the environment; (ii) channel allocation in HetNets is based on multiple MAC protocols; and (iii) channel capabilities might vary. In order to facilitate faster and more effective spectrum usage, MC-DLMA seeks to identify the best access strategy for transmitting on those pre-allocated channels. Conventional DRL methods, such as the initial deep Q-network method, are inapplicable to their issue because of the intricate temporal connection of spectrum states in HetNets.

For cognitive radio networks operating in mobile environments, the DSA issue is the primary emphasis of [34]. They model the random mobility of users by designing the exact positions of SUs and main users to be generated at random within the network's reach in each time slot. Since a user's mobility might alter their position, they provide two distance thresholds for each SU. That way, SUs can investigate potential chances for spectrum sharing among users located at various points in time. Next, they provide a DSA method for mobility (MD3SA) that utilizes deep reinforcement learning and relies on a Double Deep Q-Network. The suggested approach outperforms the random access technique and the classic DDQN-based DSA method without mobility consideration in simulation results.

In heterogeneous wireless networks, the emphasis is on spectrum sharing [35]. In these networks, individual nodes use distinct Media Access Control algorithms to send data packets to a shared wireless access point. Various access protocols based on Deep Reinforcement Learning have been developed for use in heterogeneous wireless networks in the past. But there is a vast variety of coexisting situations, with various MAC protocols and varied numbers of nodes. When faced with new, unexpected situations, current methods need starting from scratch to train new models, which takes a long time. To solve that problem, we provide Generalized Multiple Access, a new MAC protocol that uses the meta-RL algorithm. In order to solve all the problems or gaps identified in this section, we proposed a model to fulfill all these problems discussed in section.4.

3. Proposed method

To measure the efficacy of spectrum usage, we provide a new network-aware spectrum efficiency metric in this study. Considerations like network topology, device capabilities, connection quality, user priority, and node density are included into the suggested measure. The suggested measure incorporates area efficiency implicitly. A unique transmit power, which is dependent on network factors, maximizes efficiency for a point-to-point connection. Our study and numerical findings demonstrate that our measure provides a fresh viewpoint on optimizing wireless systems at the network level and helps make better use of spectrum, which is particularly useful for future networks that will likely include diverse, dynamic spectrum access systems, and Multi-RAT technology. Also covered were the benefits and drawbacks of a two-user scenario, which shed light on potential efficient use of spectrum in new areas. We conclude that the suggested measure has the ability to provide a new perspective on measuring spectrum usage and might be helpful in squeezing the most out of each accessible hertz of spectrum in dynamic and diverse radio settings. The determination of a threshold power for interfering tolerance, taking into account the total interference power generated by the transmitters in a channel, is an area that needs more investigation in future study. For wireless systems of the future, it would be helpful to evaluate the suggested measure for hierarchical network designs and fading channels.

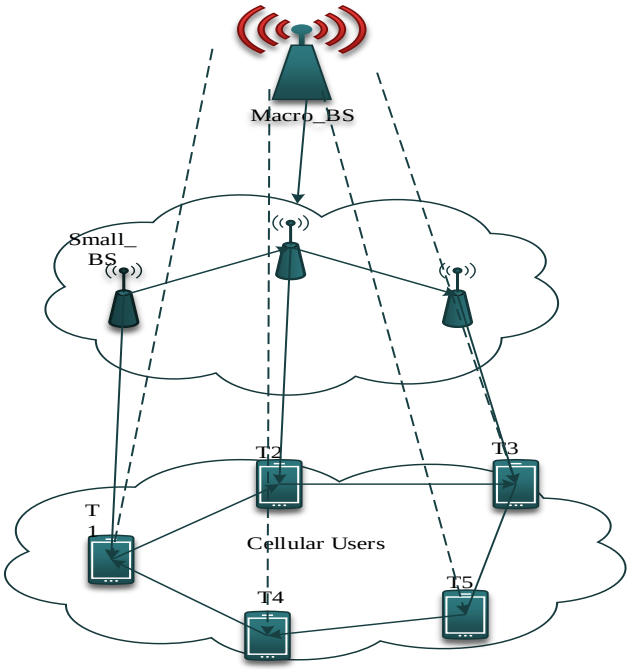

Fig. 1: Heterogeneous Networks Model.

3.1. Network assumptions

The evolution of the suggested approach is detailed here. A paradigm for heterogeneous networks with dynamic spectrum access is shown in Fig.1. A heterogeneous network is defined in this study as one that makes use of a wide variety of radio access methods. The sender (Tx_k) performs remote user service d_k, with a power P_k on frequency f_k, on a point-to-point link. Let P_{th} be the minimum power required before other users are unaffected. Power that was obtained from $\mathrm{T}_{\mathrm{x}_k}$ would weaken to P_{th} at a distance $d_{th_k}$from the transmitter.

Thus every user outside the radiusd_k, but within the radius d_{th_k} face interference from $\mathrm{T}_{\mathrm{x}_k}$. In figure 1, the region known as the "area of affected users" is barred from reusing frequency f_k.

Taking into account all of the inherent fluctuations and uncertainties of the wireless environment, we investigate the RSA issue in a 5G small-cell network. Without limiting ourselves, let's pretend that the cellular network has $\mathcal{N}$ shared users with $\mathcal{M}$ accessible channels. Then, to provide light on how users compete for spectrum resources, we establish $\mathcal{N} > \mathcal{M}$. Use the notation "shared users" for display purposes. $\mathcal{N}$, i.e. $\mathcal{N} = \{1,2, \dots, N\}$and the collection of accessible channel resources such $\mathcal{M}$, i.e. $\mathcal{M} = \{1,2, \dots, M\}$.

Time is divided up into frames, with F time periods in each frame. The packet comes at the beginning of each frame randomly, i.e. following a Poisson distribution, and each user has a limited buffer. To communicate with the access point, the user node uses a pre-assigned channel and changes its media access control (MAC) protocol. An acknowledge character (ACK) signal is returned to the sender user by the AP when it successfully receives the packet. A collision happens when several users try to access an identical channel at once, and as a result, none of them are able to transmit the current time. No user has access to any other user's protocol details. In this system, you may find the following kinds of users:

- TDMA: delivers data packets within a predetermined time limit.
- Q-ALOHA: transmit data packets using the Q-ALOHA protocol with a predetermined probability q in every single slot.
- Generate a value at random using fixed-window ALOHA (FW-ALOHA) $w \in [W_1, W_2]$ after a data packet has been sent, and then wait for w time frames before sending another one.
- CU: use the TDMA regulations, it keeps track of the channel's status (BUSY/IDLE) for a certain amount of time and uses the deep Q network to choose which channel to transmit data packets.

Everyone else can only send on a single channel, but CU has the ability to transmit on several channels. The idea is for the CU to be able to access any channel that isn't being used, even if there are a few of them.

There can be a lot of spectrum squandered by the communication network. The three-channel heterogeneous system is shown in figure 1 as a channel state diagram for 20 consecutive time periods. If the channel is white in the graphic, it means it is idle; if it is black, it means it is active. The spectrum capabilities are going unused due to the abundance of white regions. This is why we're aiming to optimize CU in order to boost throughput and prevent transmit collision. The CU must deduce the PUs' future behavior from previous observations, such as ACK and channel status, as it lacks any prior knowledge about PUs.

The indeterminate locations of users [26], which may be caused by their stochastic motions, are a particular reason producing the nonnegligible uncertainty in a small-cell network, as illustrated in figure 1. In order to fix the missing location data and provide a thorough description of this uncertainty, for every user $n \in \mathcal{N}$, Assuming the precise whereabouts of the unknown $l_n = \{x_n, y_n\}$ is situated inside a circular region with the approximate but inaccurate position $\hat{l}_n = \{\hat{x}_n, \hat{y}_n\}$ as center and φ_n as radius, i.e.,

$$\begin{aligned} l_n &= \hat{l}_n + \Delta l_n, \\ \Delta l_n \in \mathcal{L}_n &= \{\Delta x_n^2 + \Delta y_n^2 \leqslant \varphi_n^2\} \end{aligned} \tag{1}$$

where $\Delta \boldsymbol{l}_n = \{\Delta x_n, \Delta y_n\}$ standing for the uncertainty zone and denoting the estimate inaccuracy. It is worth mentioning that users stochastic movement may cause the uncertain zone $\mathcal{L}_n$ to be irregular. However, be assured that there will always be a circle that contains this irregular area. Therefore, alternative irregular shapes may be conservatively estimated using the formed uncertain circular area $\mathcal{L}_n$.

The distances $d_{n,n'}$ and $d_{n,S}$ between users n and SBS and n' and n because the small-cell base station (SBS) are obviously not finite, and they are given by:

$$\begin{aligned} d_{n,S} &= |\boldsymbol{l}_n - \boldsymbol{l}_S| = \left|(\hat{\boldsymbol{l}}_n + \Delta \boldsymbol{l}_n) - \boldsymbol{l}_S\right|, \\ d_{n,n'} &= |\boldsymbol{l}_n - \boldsymbol{l}_{n'}| = \left|(\hat{\boldsymbol{l}}_n + \Delta \boldsymbol{l}_n) - (\hat{\boldsymbol{l}}_{n'} + \Delta \boldsymbol{l}_{n'})\right|, \end{aligned} \tag{2}$$

where $\boldsymbol{l}_S$ is the position of the SBS and $|\cdot|$ is the Euclidean norm, respectively.

To illustrate the signal's power gain propagation and the desired (interference) power gain, this work used the free space path-loss (PL) framework with rayleigh fading $h_{n,S}^m$ $\left(h_{n,n'}^m\right)$ may be provided among node n and the SBS (a different node n') via channel m:

$$h_{n,*}^m = \begin{cases} h_{n,S}^m = d_{n,}^{-\alpha_m} \times \vartheta_m = |\boldsymbol{l}_n - \boldsymbol{l}_S|^{-\alpha_m} \times \vartheta_m, \\ h_{n,n'}^m = d_{n,n'}^{-\alpha_m} \times \vartheta_m = |\boldsymbol{l}_n - \boldsymbol{l}_{n'}|^{-\alpha_m} \times \vartheta_m, \end{cases} \tag{3}$$

where α_m is the PL exponent, and ϑ_m does the PL on channel have an immediate random component $m(m \in \mathcal{M})$.

Finally, the location uncertainties are converted to the channel state. Many interpretations of the location uncertainty assumption are therefore possible. In particular, it defines the imperfect location estimate at the outset. However, it may also show how the ever-changing wireless environment contributes to transmission/interference link uncertainty.

3.2. Spectrum efficiency metric

Two parameters are currently being introduced δ and β_k. δ stands for user density, or the amount of user devices in relation to the total area β_k is a value that indicates the importance of node k. Assume that K is the sum of all network users, as index by $k \in (1, K)$. For expressing the spectrum effectiveness, η_k for user k, we suggest the following formulation, which is obtained heuristically.

$$\eta_k = \alpha_k R_k \tag{4}$$

where,

$$\alpha_k = \begin{cases} \frac{1}{\exp\left[\left(d_{th_k}^2 - d_k^k\right)\beta_k \delta\right]} & \text{if } d_{th_k} \geq d_k \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

$$R_k = \log_2(1 + \gamma_k)$$

R_k relies on the transmit/receive power and user k's information rate per unit bandwidth; γ_k is user k's Signal-to-Interference-Noise-Ratio (SINR); α_k is a weighting factor that is directly related to the number of nodes in the impacted ring (as illustrated in figure 1) and the relevance of k; d_{th_k} is the distance at which power of transmitter $\mathrm{Tx_k}$ to attenuate to P_{th}; d_k δ represents the density of nodes, which is the distance among the transmitter and receiver, and β_k determines how important or high-priority a user is $k, 0 \leq \beta_k \leq 1$
P_k is the transmit power by receiver k; $P_{R_k} = P_k d_k^{-n}$ where n is the constant for propagation loss and is the received power.
N_0 is the noise power as a function of bandwidth. Here are some features of the weighting factor that we have noticed:α_k : (a) $0 \leq \alpha_k \leq 1$; (b) $d_{th_k} \to d_k, \alpha_k \to 1$, No users experience interference when the region of impacted users decreases to zero. Consequently, it carries more weight; and (c) $d_{th_k} \to \infty, \alpha_k \to 0$, during when interference affects all users. Therefore, less importance. To make the analysis more manageable, the weighting factor makes use of the exponential function. Because of its built-in capability, it also enables heuristics to make soft judgments when rewarding or punishing. In addition, for inputs of any size, the function as it is now produces limited results. A greater reuse factor and reduced spectrum consumption are hallmarks of a high-quality connection. The capabilities of the receiver and the state of the channel determine this. The performance of a faulty connection will be penalized by a scaling coefficient α_k since it will impact more users in the system compared to a good link. Users are penalized according to the number of nodes impacted by the scaling factor α_k. How much are $\left(d_{ih_\star}^2 - d_k^2\right)\delta$ is directly proportional to k, the user's impact on the network nodes. A density function that depends on time, frequency band, and location may be used to probabilistically represent the density of nodes δ. An estimate of the predicted spectrum efficiency might be obtained from this. Not all bits are the same. This metric $\beta_k \in (0,1)$ used to indicate that one user is more important than another. Its relationship to the real priority is inversely proportional. That is, $\beta_k = 0$indicates uppermost priority and $\beta_k = 1$represents the lowest priority. The dependability and accessibility of spectrum for customers whose services need higher priority (such as public safety), may be maintained unaffected by the total amount of nodes impacted by setting β_k to zero. Bits per second per unit bandwidth, as defined by Shannon's capacity, is the traditional reduction of the measure.
Each user's spectrum efficiency in the network may be found using (1). On the other hand, optimizing the spectrum management platform across all network users is essential for optimal spectrum usage. The interaction and compromise between them must be taken into consideration. Hence, a spectrum efficiency measure at the network level, denoted as η, may be helpfully defined as

$$\eta = \sum_{k=1}^{K} \eta_k = \sum_{k=1}^{K} \alpha_k R_k \quad (6)$$

Where α_k and R_k are assumed by (3) correspondingly.

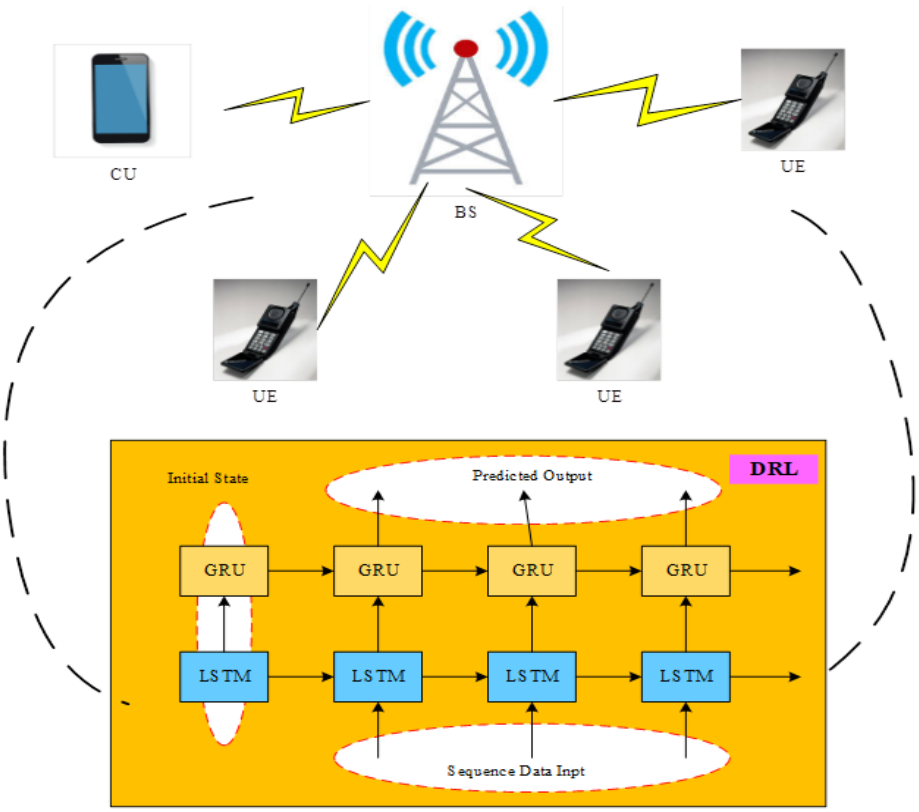

Fig. 2: Proposed Hybrid RNN Model.

Fig.2 Illustrate the heterogeneous wireless network system integrated with a DRL-based learning framework. The upper part shows the network topology with a central base station (BS) communicating with multiple user equipments (UEs) and a control unit (CU). The DRL module in the lower section employs a hybrid LSTM-GRU structure to learn temporal features and predict optimal transmission strategies from sequential data inputs.

3.3. Environment learning for spectrum access

Here we lay out the specifics of the suggested approach, which comprises the deep neural network, the design of the actions, states, and rewards. DQN is a refined version of the Q-learning algorithm. Directly feeding high-dimensional perceptual vector into the deep Q network for policy learning is how it employs end-to-end reinforcement learning. When compared to Q-learning, which relies on a table, this method approximates the Q function using deep neural networks. In Q-learning, $Q(s_t, a_t)$ is the predictable reward, i.e., Q value, of enchanting action a_t in the state s_t. In DQN, $Q(s_t, a_t)$ becomes$Q(s_t, a_t; \theta)$, where $\theta$the feature that the DLNN uses is. The data set (s_t, a_t, r_t, s_{t+1}) D, a first-in, first-out (FIFO) experience pool, stores the created time step. After that, a mini-batch is formed by randomly extracting N pairs of data from D. By using neural network back-propagation, the loss function L(θ) is reduced as stated by

$$L(\theta) = \frac{1}{N}\sum_{t=1}^{N} \left[r_t^t + \gamma \max_a Q(s_{t+1}^t, a; \theta) - Q(s_t^t, a_i^t; \theta)\right]^2 \quad (7)$$

The following is an expanded version of the dueling deep Q network (D3QN) built on top of DQN:

- Resolve the overestimation issue in DQN with Double DQN. By using the target network (θ'), it modifies the loss function along with updates it accordingly.

$$L(\theta) = \frac{1}{N}\sum_{t=1}^{N} \left[r_t^t + \gamma Q\left(s_{t+1}^t, \arg\max_a Q(s_t^t, a, \theta); \theta'\right) - Q(s_t^t, a_t^t; \theta)\right] \quad (8)$$

- In DQN's dueling mode, you may alter the network's topology by removing branches [34]. In order to evaluate the benefit value of every action and the value of the existing state, a value layer and an advantage layer are added.

Action

Prior to the beginning of each time slot, CU decides on an action. Just for the sake of argument, let's say there are K potential channels. In time slot t, the CU's job is to choose $n \in \{0,1,2, \dots, K\}$ paths for transmitting the packet of data. We identify $a_t^k \in \{0,1\}$ (wait or transmit) as the action of the k th channel of the CU at time slot t, where $k \in \{1,2, \dots, K\}$. The CU joint action is signified by $a_t = [a_t^1, a_t^2, \cdots, a_t^K]$ for every K channel.

State

Decisions made by the CU are based on the state. In order to aid in decision making, CU keeps an eye on all channels. The k^{th} channels state is represented by during time slot t b_t^k, where $b_t^k = 0$ means that the k^{th} network is idle, and $b_t^k = 1$ implies that the channel is in use. While collecting the action of PUs, CU must avoid colliding with them. The data that has been gathered is referred to as $o_t = [o_t^1, o_t^2, \cdots, o_t^k, \cdots, o_t^K]$, where o_t^k is the opinion of the k^{th} at time t via the channel by CU. For example, in the case when PU transmits data but CU does not, $b_t^k = 1, a_t^k = 0$, we have $\delta_t^k = 1$ and $d_t^k = 0$ otherwise, assumed through

$$\sigma_t^k = \begin{cases} 1, & b_t^k = 1, a_t^k = 0 \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Given the time-dependent nature of the channel states, we may calculate the current state s_t by integrating the channel observations over the previous M time slots.

$$\begin{aligned} s_t &= [o_{t-M}^{\mathrm{T}}, o_{t-M+1}^{\mathrm{T}}, \cdots, o_{t-1}^{\mathrm{T}}]^{\mathrm{T}} \\ &= \begin{bmatrix} o_{t-M}^1 & o_{t-M}^2 & \cdots & o_{t-M}^K \\ o_{t-M+1}^1 & o_{t-M+1}^2 & \cdots & o_{t-M+1}^K \\ \vdots & \vdots & \ddots & \vdots \\ o_{t-1}^1 & o_{t-1}^2 & \cdots & o_{t-1}^K \end{bmatrix} \end{aligned} \quad (10)$$

Reward

At time slot t, we signify the ACK signal $z_t = [z_t^1, z_t^2, \cdots, z_t^k, \cdots, z_t^K]$, where $z_t^k \in \{0,1\}$ indicates that the k^{th} channel has received an ACK signal. At time t, d_t is the total amount of packets in CU's buffer. Utilizing idle channels and maximizing network throughput are the goals of CU. Afterwards, we represent the incentive as $r_t = [r_t^1, r_t^2, \dots, r_t^k, \dots, r_t^K]$, where r_t^k is the reward for the k^{th} channel at time slot t, assumed through

$$r_t^k = \begin{cases} 1, & z_t^k = 1, a_t^k = 1, d_t > 0 \\ -1.5, & z_t^k = 0, a_t^k = 1, d_t > 0 \\ -0.5, & z_t^k = 0, a_t^k = 0, d_t > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

A penalty of -1.5 for collisions, -0.5 for waste, as well as 1 for successful transmitting are the consequences of various channel utilizations.

3.4. Neural network

Given that each channel is orthogonal to the others, CU is able to assess the predicted reward for a particular channel independently of the others. Furthermore, when K grows, the action space size, which is 2^k, will expand exponentially. The outcome is a very large neural network and extremely expensive storage. The neural network simply cannot provide the Q value of every action. In order to solve this problem, we partition the current state S_t into separate states s_t^k of each channel, and we define

$$\begin{aligned} s_t &= \begin{bmatrix} o_{t-M}^1 & o_{t-M}^2 & \cdots & o_{t-M}^K \\ o_{t-M+1}^1 & o_{t-M+1}^2 & \cdots & o_{t-M+1}^K \\ \vdots & \vdots & \ddots & \vdots \\ o_{t-1}^1 & o_{t-1}^2 & \cdots & o_{t-1}^K \end{bmatrix} \\ &= [s_t^1 \quad s_t^2 \quad \cdots \quad s_t^k \quad \cdots \quad s_t^K]. \end{aligned} \tag{12}$$

Where

$$s_t^k = [o_{t-M}^k \quad o_{t-M+1}^k \quad \cdots \quad o_{t-1}^K]^{\mathrm{T}} \tag{13}$$

The suggested Dueling DQN-based neural network architecture is shown graphically in Figure 2. Our internal neural network is a basic fully connected (FC) model. An advantage layer and a value layer are the two primary components of a neural network. The output of the 128-unit and 2-unit FC layers that make up the advantage layer is the advantage value A(s,a). Output: state value V(s) from the value layer's 128-and 1-unit FC layers. You can get the Q-value Q(s,a) by putting A(s,a) and V(s) together. In the neural network, s_t^k is fed into the advantage and value layers at time slot s_t^k. The value layer produces the state value V(s;θ) as its output, whereas the advantage layer produces the advantage value A(s,a;θ). The value of the action-state Q(s,a;θ) is obtained by adding the two values. For channel k, we use the action-state value notation as $Q_t^k = \left[Q_{t,0}^k, Q_{t,1}^k\right]$, where $Q_{t,0}^k$ is the Q value representing not transfer, and $Q_{t,1}^k$ is the Q-value that indicates transmission. At last, the bigger Q value is obtained by CU from $Q_{t,0}^k$ and $Q_{t,1}^k$ in order to determine the conduct a_t^K of channels k. The combined action may be obtained by repeating this technique for each channel.

$$a_t = [a_t^1, a_t^2, \cdots, a_t^K] \tag{14}$$

Our approach to addressing problems is based on a D3QN algorithm. In order to enhance the access policy, CU interacts alongside the environment to gather historical information. By decreasing the loss function L(θ), as shown in equation (6), the optimization strategy modifies the neural network's parameters. The knowledge pool D stores the history data as (s_t, a_t, r_t, s_{t+1}).

During the training phase, data is randomly pulled from D in order to train the neural network and compute the loss function. Data in D will have a lower reference value in a dynamic environment due to channel changes. So, once the environment changes, it's important to extract the historical data from D as soon as feasible. This indicates that there need to be a limit on the scope of the D-pool of experiences. The reward reduction rate γ often approaches 1 in DRL tasks. In the long run, the agent may reap the most benefits because of their enhanced vision. The long-term benefit is irrelevant in an ever-changing environment, thus it's more vital to focus on the short-term gains while trying to adapt to it. Therefore, the discount rate γ for rewards must not be very high.

The conditions in the surroundings where S_t is measured serve as input to the models that have been suggested. One step toward spectrum allocation is DDQN's final product. A post-process transforms this action choice into the mapping matrix $[X,Z]^T$. Direct application of the processed action to the allocation vectors is possible as an output of the Hybrid RNNs. Therefore, the size of the output action an is (2L+1)×M, with the frequency distribution X for DBS being the first (L+1)×M and the allocation Z among the heterogeneous nodes being the following LÓM.

3.5. Dynamic spectrum allocation using hybrid RNNs

One kind of artificial neural network that may simulate the changing behavior of a time series is the recurrent neural network (RNN). Recurrent Neural Networks (RNNs) work by making use of sequential data. All inputs and outputs are considered to be completely separate in a conventional neural network. But the network has to know the previous observations in order to forecast the upcoming state. Another way to look of RNNs is as having "memory" that stores data on all the calculations that have been done up to this point. While recurrent neural networks (RNNs) have endless potential in principle, in fact they can only look back a certain number of steps. Recurrent neural networks (RNNs) are so-called because they execute the same calculation for each element in a sequence, having the output depending on the results of the computations that came before.

The network that handles spectrum auctions is a hybrid RNN, meaning it combines LSTM and GRU algorithms. The prediction model relies heavily on establishing the Hybrid RNNs model's structure. In most cases, the size of the eigenvector dictates how many nodes are used in the input layer of a hybrid RNN. There are three input nodes because representative indicators reflecting the three parts of the information the bidders' interference, experience, and economic ability are chosen during the auction.

It is possible to get a near approximation of the true value by means of continuous data testing. To find the hidden layer's neuron count, we may utilize the tried-and-true trial-and-error technique, which relies on the user's past actions. We can find out that four neurons is the minimum variation by starting with a certain amount of neurons in the training network's hidden layer, adjusting it up or down over time while using the same training data. What kind of output data is often used to establish the dimensionality of the output vector. The fitting ability is determined by the number of hidden layers. It takes a certain amount of hidden layers to adequately portray a complicated relationship. Having said that, adding more hidden layers will make training more time-consuming and difficult. The output value allows for the realization of broad judgments about evaluation indices. Fig. shows the construction of a simulation of Hybrid RNNs.

3.5.1. LSTM

An example of a recurrent neural network (RNN) is a Long Short-Term Memory (LSTM) network. In many cases, RNN is not a good choice since learning lengthy data sequences causes gradients to be lost. By selecting which pieces of data to utilize, LSTM eliminates this issue. As the input sequence lengthens, the gradients at the beginning of the input diminish and become zero, making it more difficult to capture the influence of the early phases. The limitations of the RNN are addressed by the LSTM framework, which consists of an input gate, an output gate, and a forget gate.

Long short-term memory (LSTM) is organized with the help of its three gates. All of the cells are controlled by these three gates. Memory cells are hidden units in the LSTM architecture that are used for long-term dependency. It uses these memory units to store information that has to be remembered for a long time. In order to retrieve or contribute data to the cell state, the LSTM may delicately regulate structures known as gates. Making a decision on which cell state data to ignore is the first step in long short-term memory (LSTM). It is the forget gate that makes this determination. The input gate is then used to determine whether the cell state should be updated with the new context. A sigmoid layer is used for the purpose of determining which pieces of information need updating, while a *h* layer is employed to generate a vector of potential new pieces of information. At last, the output gate decides what data is sent out. In order to avoid data loss, one of the most crucial aspects of the LSTM design is its ability to retain inputs without forgetting them.

Set the input array $x = (x_1, x_2, \dots x_4)$, a mapping is computed by the network and applied to the output $y = (y_1, y, \dots y_t)$. We may find the activations of the units using the following equations. "ғ" stands for the logistic sigmoid functional. The input gate is symbolised by I, the forget gate by f, the output gate by o, and the cell activation vector by c. The dimension of the hidden vector h is shared by all of these vectors. The weight matrix from cell to gate vector are denoted by the (W) terms. The output of the activation process is Tanh, and that's it. Sigmoid and tangent hyperbolic activation functions are often used by LSTM networks.

$$f_t = \sigma\left(W_f^*[h_{t-1}, x_t] + b_f\right) \tag{15}$$

$$i_t = \sigma(W_i{}^*[h_{t-1}, x_t] + b_i) \tag{16}$$

$$\begin{aligned} \dot{\mathrm{C}}_t &= \tanh\left(W_C{}^*[h_{t-1}, x_t] + b_C\right) \\ C_t &= f_t^* C_{t-1} + i_t^* \mathrm{C}_t \\ o_t &= \sigma(W_o^*[h_{t-1}, x_t] + b_o) \\ h_t &= o_t^* \tanh\left(C_t\right) \end{aligned} \tag{17}$$

Repeatedly running the aforementioned procedure via the LSTM keeps iterating. So that the LSTM output values are as close to the training data as possible, the model learns the weight variables 0 and bias parameters ().

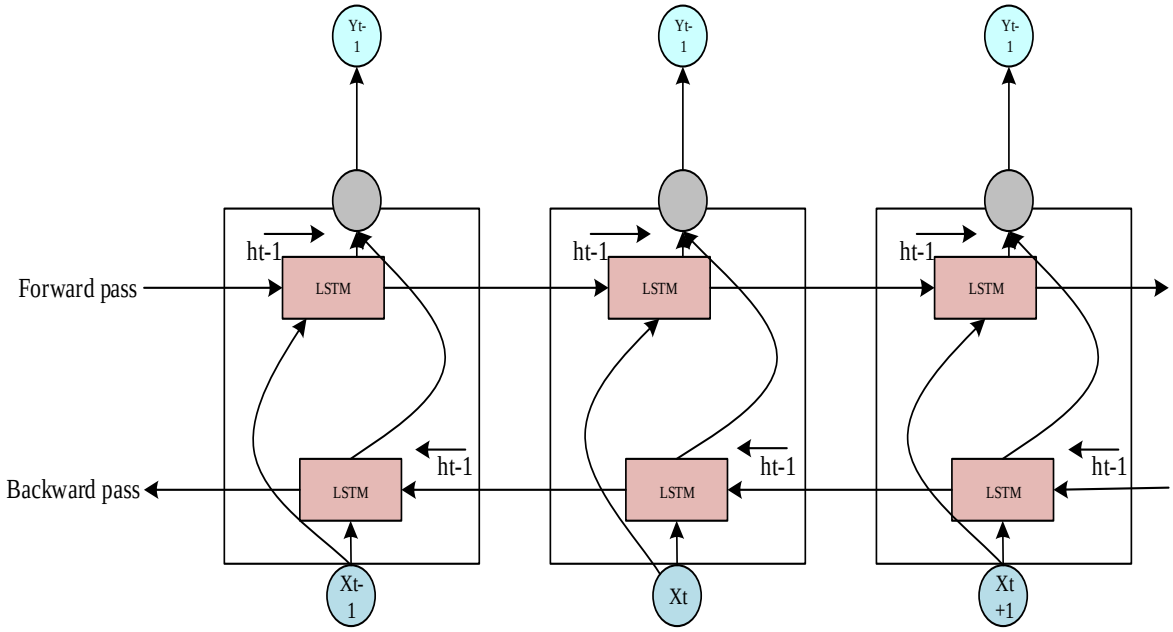

Fig. 3: LSTM Model.

Fig.3 illustrates Long Short-Term Memory (LSTM) architecture used for sequence modeling. The model processes input data in both forward and backward directions through parallel LSTM layers, capturing contextual dependencies from past (forward pass) and future (backward pass) time steps. Each LSTM unit receives the hidden state from the previous time step and outputs predictions at each time instance, enhancing temporal feature learning.

3.5.2. GRU

GRU is a design that solves the vanishing gradient issue that occurs in LSTM and other iterative neural networks. In the long run, GRU can figure out what depends on what. One design that relies on RNNs is the GRU structure, which stands for Gated Recurrent Unit. By

verifying the updated data in the concealed state, we may eliminate the long-term dependencies that arise in RNN-based approaches. Inside the GRU, you'll find an update gate and a reset gate. Like an LSTM's forget and input gates, the update gate determines what data to retain, what to discard, and how to add fresh data. The amount of data that should be erased is determined by the reset gate. The GRU is gaining popularity daily due to the fact that it produces a simpler model compared to the classic LSTM model. In comparison to LSTM, GRU's training time and accuracy in forecasting are both significantly improved by its very simple structure.
To carry out these procedures, the following equations are used:

$$r_t = \sigma(W_r h_{s-1} + U_r x_t) \tag{18}$$

$$\hat{\mathrm{h}}_{\mathrm{t}} = \tanh\left(\mathrm{W}(r_t * h_{t-1}) + U x_t\right) \tag{19}$$

$$z_4 = \sigma(W_z h_{t-1} + U_x x_t) \tag{20}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \hat{h}_t \tag{21}$$

h_4 and h_{4-1} stand for the results of the present and past situations, correspondingly. You can see the update gate (z_7) and the reset gate (r_t) in action. W_r and U are the weight matrices, while an is the logistic sigmoid function.

3.5.3. LSTM-GRU

In order to make predictions, several research have used models that include linear and non-linear components. This research proposes a hybrid model that combines LSTM and GRU algorithms. An LSTM using 128 hidden neurons makes up the first layer, while a GRU using 64 hidden neurons makes up the second. A thick layer containing a single neuron makes up the third layer. In order to create a single, fully linked layer, the LSTM model's output is passed into the GRU model.
The first level of prediction is generated by the LSTM network. The GRU layer then uses the LSTM layer's output to create the final prediction. In a full network design, the LSTM, GRU, and dense layers are all present.
The initial stage of the model involves feeding serialized spectrum data into the LSTM layer. The input data is sent along every LSTM neuron across the route, and each one then produces a weighted value. Following the LSTM layer, the data is passed on to the GRU layer. Along the way from the layer of LSTM to the GRU layer, a weighted average is created. After the GRU layer sends input, the output neuron calculates the weight based on that data. After that, we compare the result to the starting values to get the cost function. The weights are adjusted to reflect the discrepancy between the actual and anticipated values when the cost function reaches its lowest point. Future estimations may be derived from the stored weights.

4. Results and discussion

4.1. Simulation environment

The writers ran a number of computer simulations to verify and validate the previously stated concepts. Table 1 displays the parameters that were utilized for the simulation. The model took into account a segment of the network with access points distributed uniformly throughout a regular mesh. There are three data sources used here, as previously mentioned. To begin, the regular grid of base stations keeps tabs on the total number of people online. We then turn our attention to the city surveillance systems, where we make the assumption that the accuracy of human identification decreases exponentially with distance. In the examined region, the cameras are distributed uniformly and put at random.

Table 1: Single Simulation Run Parameters

Parameter Name	Unit	Value
Area size	Km	21.0 × 18.0
Base station spacing	Km	3.0
Map grid spacing	Km	1.0
Number of Master base stations	-	30
Max number of UEs	-	100
Max number of persons	-	300
Number of cameras	-	50
Number of Small Aps	-	100
Number of time stamps	-	6
Frequency	MHz	2400
Transmit power	dBm	20.0
Max bandwidth	MHz	20.0
Interpolation method	-	nearest neighbor

After that, we'll go over the hyper-parameters that make our DDQN as well as Hybrid RNNs approaches work. The network architecture of DDQN and critic of Hybrid RNNS is same, consisting of three hidden layers of fully connected neural networks having 500,1000,500 neurons. This allows for a more accurate comparison of the two techniques. In Hybrid RNNS, the actor network is likewise structured in this way.

Our research shows that heterogeneous networks with small sizes and few sub-channels may achieve comparable rate performance by using varying numbers of neurons at every hidden layer. On the other hand, when using a small number of neurons, stability drastically decreases as the action space grows. A customized activation function is used in the third hidden layer of the Hybrid RNNS to meet the constraint requirement in issue (P1). Each column of the DBS allocation matrix X is processed using the softmax function since the cumulative probability restriction for assigning one typical channel is one. As a corollary, this is to choose the top "candidate" for every channel in order to improve their rating on review networks. In order to determine the optimal "combination" of smaller channels for allocation,

the sigmoid function has been applied to each rows of Z for the heterogeneous node. With a value of 0.9, the discount rate and the investigation rate are $\epsilon = 0.9$ subject to a rate of degradation of 0.9995. Both algorithms have their learning rates set to the same value as $\beta = 10^{-4}$. The Hybrid RNNS model's soft update rate is $\tau = 0.01$.

All of these stages make up one simulation iteration. It all starts with finding the user-to-access-point distance in terms of the standard geometric formula. In the next step, the power that each user receives is determined using the free-space loss concept and their proximity to the access points. When the access point's power value is determined, the users are then allocated to that one. Following this, all users inside a single access point have an equal chance of using the radio resources. At last, the user's bitrate is determined (using Shannon's algorithm) by calculating the signal-to-noise ratio with interference value.

4.2. Throughput analysis

The performance assessment parameter used in this research is throughput, which is the average amount of data packets successfully sent for every timeslot, averaged across N timeslots. The following normalization algorithm is used to compute the throughput of nodes:

$$T = \sum_{\tau=t-N+1}^{t} \frac{n_\tau}{N} \quad (22)$$

where $\sum_{\tau=t-N+1}^{t} n_\tau$ denotes the quantity of time slots that have been successfully communicated out of a total of N time slots, with N standing for the overall amount of time slots.

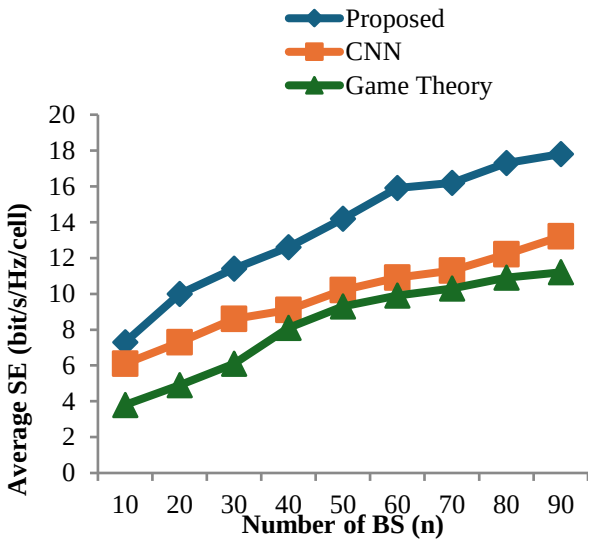

Fig. 4: Average Spectral Efficiency Analysis.

Fig.4 illustrate the Comparison of average spectral efficiency (SE) versus the number of base stations (BS) for three different approaches Game Theory model, CNN model and proposed model. The results show that the proposed method consistently outperforms the baseline models, achieving higher SE as the number of BS increases, indicating superior scalability and efficiency in dense network deployments.

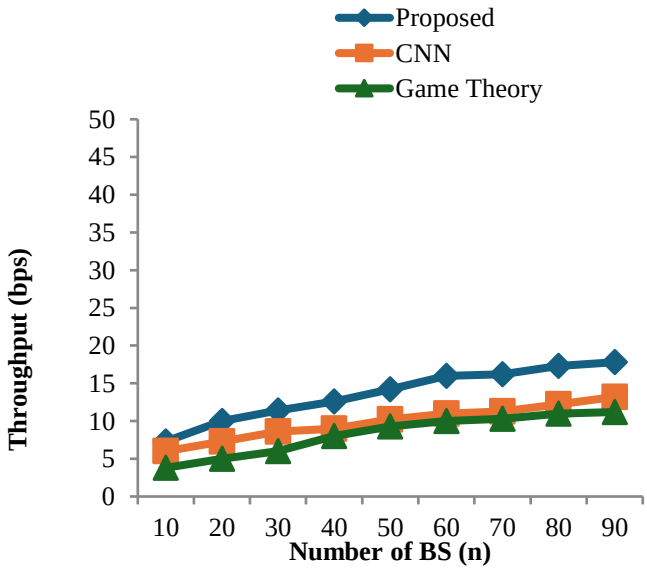

Fig. 5: Throughput Analysis.

Fig.5 illustrates Throughput performance comparison with varying numbers of base stations (BS) for the proposed method, CNN-based model, and Game Theory-based approach. The proposed model achieves consistently higher throughput across all BS configurations, demonstrating its effectiveness in optimizing data transmission in dense network scenarios.

We set the total amount of time slots N to 1000 for the computation of short-term throughput, with each restart slot lasting 1 millisecond. With this, we can see how well the nodes in the network have been doing during the last one second in terms of short-term throughput. We can get the average throughput across t time slots by setting N to t, which allows us to compute long-term throughput.

5. Conclusion

By defining the issue of optimization to maximize the cumulative log-rate while assuring the system needs of UEs, we examine the spectrum distribution in heterogeneous networks. To address this optimization issue, we provide two techniques that are based on DRL. In the presence of dynamic network conditions, we demonstrate that both algorithms are capable of learning to adjust their allocation strategies for access transmission and backhaul. Additionally, as the scale of heterogeneous networks increases, the suggested DRL based Hybrid RNNs outperform static strategies in terms of system sum log-rate and convergence time, outperforming DDQNs. In comparison to the conventional full-spectrum reuse technique, the simulation results demonstrate the suggested methods' superior sum log-rate performance and their encouraging potential.

Acknowledgement

I would like to thank my guides Tapaswini Samant and Swati Swayamsiddha for their support in my research work.

References

[1] Khadem Mina, Ansarifard, M., Mokari, N., Javan, M.R., Saeedi, H., & Jorswieck E.A. "Dynamic Fairness-Aware Spectrum Auction for Enhanced Licensed Shared Access in 6G Networks." IEEE Transactions on Communications, vol. 46, no. 1, 2023,pp 1-13, https://doi.org/10.1109/TCOMM.2024.3486985.

[2] Poshattiwar, Sadhana D., and S. B. "Dynamic Spectrum Sensing for 5G Cognitive Radio Networks Using Optimization Technique." Journal of Electrical Systems, 2024.

[3] Kopyto, David Jonas, Lindner, S., Schulz, L., Stolpmann, D., Bauch, G., & Timm-Giel, A. "Deep Learning-Based Dynamic Spectrum Access for Coexistence of Aeronautical Communication Systems." Proceedings of the 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall), 2022, pp. 1–5. https://doi.org/10.1109/VTC2022-Fall57202.2022.10012932.

[4] Zhang, Xu, Chen, P., Yu, G., & Wang, S. "Deep Reinforcement Learning Heterogeneous Channels for Poisson Multiple Access." Mathematics, vol. 11, no. 4, 2023, p. 992. MDPI, https://doi.org/10.3390/math11040992.

[5] Atimati, E., Crawford, D. H., and Stewart, R. W. "Intelligent Shared Spectrum Coordination in Heterogeneous Networks." Proceedings of the 2023 IEEE Virtual Conference on Communications (VCC), 2023, pp. 252–257. https://doi.org/10.1109/VCC60689.2023.10474686.

[6] Xu, Y Lou, J., Wang, T., Shi, J., Zhang, T., Paul, A., & Wu, Z. "Multiple Access for Heterogeneous Wireless Networks with Imperfect Channels Based on Deep Reinforcement Learning." Electronics, vol. 12, no. 23, 2023, p. 4845. MDPI, https://doi.org/10.3390/electronics12234845.

[7] Park, J, Jin, H., Joo, J., Choi, G., & Kim, S.C. "Double Deep Q-Learning Based Backhaul Spectrum Allocation in Integrated Access and Backhaul Network." Proceedings of the 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2023, pp. 706–708 , https://doi.org/10.1109/ICAIIC57133.2023.10067029.

[8] Gnanaselvam, R., and Vasanthi, M. S. "Dynamic Spectrum Access-Based Augmenting Coverage in Narrow Band Internet of Things." International Journal of Communication Systems, vol. 37, 2023. https://doi.org/10.1002/dac.5629.

[9] Devipriya, S., Leo Manickam, J. M., and Jasmine Mystica, K. "A Deep-Learning Based Approach to Resource Allocation in NOMA Based Cognitive Radio Network with Heterogeneous IoT Users." Proceedings of the 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2022, pp. 1–6. https://doi.org/10.1109/ICDCECE53908.2022.9793269.

[10] Zhang, Y., Li, X., Ding, H., & Fang, Y. "A Joint Scheme on Spectrum Sensing and Access with Partial Observation: A Multi-Agent Deep Reinforcement Learning Approach". In Proceedings of the 2023 IEEE/CIC International Conference on Communications in China (ICCC), pp. 1-6. https://doi.org/10.1109/ICCC57788.2023.10233366.

[11] Kumar, Ratnesh, Singh, C.K., Upadhyay, P.K., Salhab, A.M., Nasir, A.A., & Masood, M. "IoT-Inspired Cooperative Spectrum Sharing with Energy Harvesting in UAV-Assisted NOMA Networks: Deep Learning Assessment." IEEE Internet of Things Journal, vol. 10, no. 24, 2023, pp. 22182–22196. IEEE, https://doi.org/10.1109/JIOT.2023.3304126.

[12] Yang, Tao, Zhang, W., Bo, Y., Sun, J., & Wang, C. "Dynamic Spectrum Sharing Based on Federated Learning and Multi-Agent Actor-Critic Reinforcement Learning." Proceedings of the 2023 International Wireless Communications and Mobile Computing Conference (IWCMC), 2023, pp. 947–952. IEEE, https://doi.org/10.1109/IWCMC58020.2023.10182572.

[13] Zhang, Liang, and Ying-Chang Liang. "Deep Reinforcement Learning for Multi-Agent Power Control in Heterogeneous Networks." IEEE Transactions on Wireless Communications, vol. 20, no. 4, 2021, pp. 2551–2564. IEEE, https://doi.org/10.1109/TWC.2020.3043009.

[14] Zhang, Hao, Leng, S., Zhao, P., & He, J. "A Dynamic Consensus Scheme for Unlicensed Spectrum Sharing in Heterogeneous Networks." Proceedings of the 2023 IEEE 23rd International Conference on Communication Technology (ICCT), 2023, pp. 803–808. IEEE, https://doi.org/10.1109/ICCT59356.2023.10419465.

[15] Chew, Daniel, and Andrew Cooper. "Spectrum Sensing in Interference and Noise Using Deep Learning." Proceedings of the 2020 54th Annual Conference on Information Sciences and Systems (CISS), 2020, pp. 1–6. IEEE, https://doi.org/10.1109/CISS48834.2020.1570617443.

[16] Ahmed, Ramsha, Chen, Y., Hassan, B., Du, Du, L., Hassan, T., & Dias, J"Hybrid Machine-Learning-Based Spectrum Sensing and Allocation with Adaptive Congestion-Aware Modeling in CR-Assisted IoV Networks." IEEE Internet of Things Journal, vol. 9, no. 22, 2022, pp. 25100–25116. IEEE, https://doi.org/10.1109/JIOT.2022.3195425.

[17] Ali, Hussen Yesuf, Sun Goulin, and Abegaz Mohammed Seid. "Autonomous RACH Resource Slicing for Heterogeneous IoT Devices Communication Using Deep Reinforcement Learning." 2021 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), 2021, pp. 125–130. IEEE, https://doi.org/10.1109/ICT4DA53266.2021.9672226.

[18] Kim, Donghyun, Kwon, S.S., Jung, H., & Lee, I. "Deep Learning-Based Resource Allocation Scheme for Heterogeneous NOMA Networks." IEEE Access, vol. 11, 2023, pp. 89423–89432. IEEE, https://doi.org/10.1109/ACCESS.2023.3307407.

[19] S. Mirbolouk, M. Valizadeh, M. C. Amirani, and S. Ali, "Relay selection and power allocation for energy efficiency maximization in hybrid satellite-UAV networks with CoMP-NOMA transmission," IEEE Transactions on Vehicular Technology, vol. 71,2022, no. 5, pp. 5087–5100, https://doi.org/10.1109/TVT.2022.3152048.

[20] Yu, Yifan, Soung Chang Liew, and Tao Wang. "Multi-Agent Deep Reinforcement Learning Multiple Access for Heterogeneous Wireless Networks With Imperfect Channels." IEEE Transactions on Mobile Computing, vol. 21, no. 11, 2022, pp. 3718–3730. IEEE, https://doi.org/10.1109/TMC.2021.3057826.

[21] Lu, Lan, Gong, X., Ai, B., Wang, N., & Chen, W. "Deep Reinforcement Learning for Multiple Access in Dynamic IoT Networks Using Bi-GRU." Proceedings of the 2022 IEEE International Conference on Communications (ICC), 2022, pp. 3196–3201. IEEE, https://doi.org/10.1109/ICC45855.2022.9838614.

[22] Allahham, Mhd Saria, Abdellatif, A.A., Mhaisen, N., Mohamed, A., Erbad, A.M., & Guizani, M. "Multi-Agent Reinforcement Learning for Network Selection and Resource Allocation in Heterogeneous Multi-RAT Networks." IEEE Transactions on Cognitive Communications and Networking, vol. 8, no. 2, 2022, pp. 1287–1300. IEEE, https://doi.org/10.1109/TCCN.2022.3155727.

[23] Janiar, S. B., and V. Pourahmadi. "Deep-Reinforcement Learning for Fair Distributed Dynamic Spectrum Access in Priority Buffered Heterogeneous Wireless Networks." IET Communications, vol. 15, no. 6, 2021, pp. 674–682. IET, https://doi.org/10.1049/cmu2.12098.

[24] Liston, M. J., and K. R. Dandekar. "Entropy-Based Exploration in Cognitive Radio Networks Using Deep Reinforcement Learning for Dynamic Spectrum Access." Proceedings of the 2021 IEEE 21st Annual Wireless and Microwave Technology Conference (WAMICON), 2021, pp. 1–5. IEEE, https://doi.org/10.1109/WAMICON47156.2021.9444294.

[25] Li, Y., Wang, Y., Li, Y., & Shen, B "Multi-User Dynamic Spectrum Access Based on LRQ Deep Reinforcement Learning Network." Proceedings of the 2023 25th International Conference on Advanced Communication Technology (ICACT), 2023, pp. 79–84. IEEE, https://doi.org/10.23919/ICACT56868.2023.10079537.

[26] Kaur, Amandeep, Thakur, J., Thakur, M., Kumar, K., Prakash, A., & Tripathi, R.K. "Deep Recurrent Reinforcement Learning-Based Distributed Dynamic Spectrum Access in Multichannel Wireless Networks with Imperfect Feedback." IEEE Transactions on Cognitive Communications and Networking, vol. 9, no. 1, 2023, pp. 281–292. IEEE, https://doi.org/10.1109/TCCN.2023.3234276.

[27] Schulz, Lennart, Kopyto, D., Stolpmann, D., Lindner, S., Bauch, G., & Timm-Giel, A. "Hidden Node-Aware Dynamic Spectrum Access Using Deep Learning for Coexisting Aeronautical Communication Systems." Proceedings of the 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall) , 2023, pp. 1–5. IEEE, https://doi.org/10.1109/VTC2023-Fall60731.2023.10333681.

[28] Chen, H., Zhao, H., Zhou, L., Zhang, J., Liu, Y., Pan, X., Liu, X., & Wei, J. "A Dueling Deep Recurrent Q-Network Framework for Dynamic Multichannel Access in Heterogeneous Wireless Networks." Wireless Communications and Mobile Computing, vol. 2022, 2022, pp. 1–13.Hindawi, https://doi.org/10.1155/2022/9446418.

[29] Huang, J., Peng, J., Xiang, H., Li, L., & Yang, Y. "Hybrid Spectrum Access for V2V Heterogeneous Networks with Deep Reinforcement Learning." Proceedings of the 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), 2022, pp. 1091–1095. IEEE, https://doi.org/10.1109/WCSP55476.2022.10039356.

[30] Wang, Y., Li, X., Wan, P., & Shao, R. "Intelligent Dynamic Spectrum Access Using Deep Reinforcement Learning for VANETs." IEEE Sensors Journal, vol. 21, no. 15, 2021, pp. 15554–15563. IEEE, https://doi.org/10.1109/JSEN.2021.3056463.

[31] Chen, M., Liu, A., Liu, W., Ota, K., Dong, M., & Xiong, N. "RDRL: A Recurrent Deep Reinforcement Learning Scheme for Dynamic Spectrum Access in Reconfigurable Wireless Networks." IEEE Transactions on Network Science and Engineering, vol. 9, no. 1, 2022, pp. 364–376. IEEE, https://doi.org/10.1109/TNSE.2021.3117565.

[32] Xu, Y., Yu, J., and Buehrer, R. "The Application of Deep Reinforcement Learning to Distributed Spectrum Access in Dynamic Heterogeneous Environments with Partial Observations." IEEE Transactions on Wireless Communications, vol. 19, no. 7, 2020, pp. 4494–4506. IEEE, https://doi.org/10.1109/TWC.2020.2984227.

[33] Ye, X., Yu, Y., and Fu, L. "Multi-Channel Opportunistic Access for Heterogeneous Networks Based on Deep Reinforcement Learning." IEEE Transactions on Wireless Communications, vol. 21, no. 3, 2021, pp. 794–807. IEEE, https://doi.org/10.1109/TWC.2021.3099495.

[34] Jiang, Z., and Han, L. "DRL-Based Dynamic Spectrum Access for Cognitive Radio Networks in Mobility Scenarios." Proceedings of the 2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST), 2023, pp. 1–5. IEEE, https://doi.org/10.1109/IAECST60924.2023.10503376.

[35] Liu, Z., Wang, X., Zhang, Y., & Chen, X. "Meta Reinforcement Learning for Generalized Multiple Access in Heterogeneous Wireless Networks." Proceedings of the 21st International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2023, pp. 570–577. IEEE, https://doi.org/10.23919/WiOpt58741.2023.10349896.