5GhNet: an intelligent QoE aware RAT selection framework for 5G-enabled healthcare network

Bhanu Priya¹ · Jyoteesh Malhotra¹

Received: 22 November 2020 / Accepted: 16 November 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The COVID-19 outbreak has stimulated the digital transformation of antiquated healthcare system to a smart hospital, enabling the personalised and remote healthcare services. To augment the functionalities of these intelligent healthcare systems, 5G & B5G heterogeneous network has emerged as a robust and reliable solution. But the pivotal challenge for 5G & B5G connectivity solutions is to ensure flexible and agile service orchestration with acknowledged Quality of Experience (QoE). However, the existing radio access technology (RAT) selection strategies are incapacitated in terms of QoE provisioning and Quality of Service (QoS) maintenance. Therefore, an intelligent QoE aware RAT selection architecture based on software-defined wireless networking (SDWN) and edge computing has been proposed for 5G-enabled healthcare network. The proposed model leverages the principles of invalid action masking and multi-agent reinforcement learning to allow faster convergence to QoE optimised RAT selection policy. The analytical evaluation validates that the proposed scheme outperforms the other existing schemes in terms of enhancing personalised user-experience with efficient resource utilisation.

Keywords Artificial Intelligence · Double deep reinforcement learning · Software-defined wireless networking · Edge computing · QoE · Resource utilisation factor

1 Introduction

The COVID-19 pandemic has caused a massive acceleration in the adoption of telehealth services due to its ability to offer virtual care at zero transmission risk. In accordance with the report (Ugalmugale et al. 2020), the global telemedicine industry was expected to witness compound annual growth rate of around 15% per year by the middle of the decade but now it is expected to grow by 19.3% and 175.5 billion dollars over the same period after this global health emergency. This substantial growth in the telehealth industry has led to the exploitation of wireless communication technologies to support wide range of next-generation healthcare use cases with strict QoS provision. But harmonising all these use-cases to solely associate with one RAT is impractical. In essence to this, 5G heterogeneous networks (HetNets) seem to be a reliable solution. 5G HetNet accommodate the advanced requirements of characteristic

e-health applications and offer holistic personalised services to patient through the symbiotic integration of various radio access technologies. In addition to this, it offers multi-faceted benefits such as efficient resource utilization, upgraded scalability and seamless connectivity. The distinction among various RATs in terms of frequency bands, protocols, physical and MAC layer multiple access technologies creates an issue in the effective exploitation of 5G HetNets, one of which is context-aware RAT selection.

Therefore, the non-trivial aspect for 5G HetNets is the connectivity to the suitable RAT in accordance with the user preference to efficiently improve the quality of experience. In light of above said, numerous attempts has been made in the existing literature for suitable RAT selection. Among them, Multi-Attribute Decision Making (MADM) is a basic approach which considers multiple network parameters for RAT selection. For instance in Van et al. (2017), authors proposed an effective handover approach based on improved TOPSIS method integrated with content-centric networking that allows seamless connectivity with QoS guarantees. Yadav et al. (2018) presented a context aware network selection strategy based on MADM method that allows seamless connectivity for the transmission of a patient's physiological

✉ Bhanu Priya
bpriya812@gmail.com

¹ Department of Engineering and Technology, GNDU RC, Jalandhar, India

data to the clinicians with reduction in unnecessary switching. Zhong et al. (2020) presented a cross-layer architecture based on cognitive cycle and a cognitive MADM approach that considers network parameters along with the user's QoE to select optimal network. Authors in (Desogus et al. 2019), proposed a network selection algorithm named as TYDER that calculates network reputation on the basis of QoS parameters in accordance with the service type to offer better user experience. Bhatia et al. (2019) compared the efficiency of various MADM scheme for optimal RAT selection in accordance with various cognitive wireless body area network (WBAN) data traffics. However, the inability of MADM method to handle the imprecise and uncertain data resulted in the integration of fuzzy approach with it that derives exact weights for the efficient decision-making. Skondras et al. (2019) presented a novel VHO scheme that exploits the pentagonal interval-valued fuzzy TOPSIS algorithm to select a suitable network that satisfies the QoS requirements of demanded service in a 5G-vehicular cloud computing system. However, the uncertainty in the information gathered from metrics makes MADM method unsuitable for the network selection problem. Therefore, the fuzzy set theory and fuzzy linguistic assessment have been adopted in the literature (Al-Janabi and Alkaim 2020; Krishankumar et al. 2021) to attain flexible decision approach for random environment. For instance, Authors in (Priya et al. 2020), presented a hybrid scheme that exploits the benefits of both fuzzy logic and MADM method to avoid ranking abnormality and unnecessary handovers in 5G-enabled Industry 4.0 communication scenarios. Zhu et al. (2019) presented a novel adaptive multiservice network selection scheme that hybridises the fuzzy logic with MADM technique for the context aware optimal network selection in MEC enabled 5G HetNet. Barmounakis et al. (2017) presented a framework that exploits fuzzy inference system for optimizing the RAT selection and traffic steering as per-traffic flow and user demands in 5G network environments. But in some of the problems, lack of consensus in the criterion evaluation hinders the application of these methods and the performance gets worsen with the increase in evaluation criterion and network of interdependencies.

Another dynamic approach employed for RAT selection is game theory as discussed in (Ning et al. 2020), which leverages both cooperative and decentralised non-cooperative game theory based network selection scheme respectively for two telehealth sub-networks i.e., intra-WBANs and beyond-WBANs. Salih et al. (2016) modelled the network selection problem as a non-cooperative game to optimise the vertical handover decision in a heterogeneous network. Authors in (Rajesh et al. 2017), adopted a game theory based network selection scheme that guarantees QoS at reasonable cost and high revenues for access network. Authors in (Goyal et al. 2020), proposed a network selection scheme based

on non-cooperative game theory for heterogeneous network that maximises user QoE and network revenues. Arabi et al. (2019) proposed a matching game theoretic approach for a user-RAT association that ensures high efficiency and throughput in an autonomic IoT environment. Nevertheless, the RAT selection schemes based on game theory lacks stability and adaptability in correspondence with the environment resulting in unguaranteed convergence.

To overcome the complex RAT selection decision in highly dynamic and uncertain environment, the machine learning algorithm emerged as a powerful tool that ensures optimal network selection on the basis of experience samples. Therefore, Nguyen et al. (2017) developed a network feedback framework that employs reinforcement learning algorithm to learn an optimal RAT selection policy that converges faster to a set of correlated equilibrium and incurs low signalling overhead. In (Wang et al. 2019), authors adopted a distributed algorithm which exploits both machine learning and game theory at user side to ensure gainful switching and better resource utilisation. Authors in (Sandoval et al. 2019), leveraged deep reinforcement learning (DRL) framework for the selection of a RAT that maximises throughput while reducing power consumption and operational costs in accordance with the alerts generated in the smart city scenario. Mollet et al. (2020) presented an offline network selection scheme based on the double deep reinforcement learning (DDRL) approach to reduce number of handovers and alleviate the adverse QoS in 5G mm-wave networks. A distributed optimisation method based on multi-agent reinforcement learning (MARL) is developed in Kumar et al. (2019) to guarantee user specific requirements and maximum long term network utility in 5G heterogeneous network. Authors in (Ding et al. 2019), leveraged an energy efficient algorithm based on multi-agent deep Q network (DQN) network to ensure user satisfaction and maximum network utility in OFDMA based uplink HetNets. Authors in (Wang et al. 2019), employed the MARL approach with enhanced reward function to learn an optimal RAT selection policy that ensures better network load balancing and reduction in overall power consumption.

Nonetheless, the emergence of telehealth diverse use-cases in more autonomic and interactive manner has laid emphasis on the better personalised experience but the existing literature turns out to be incompetent in addressing the aforementioned issue. As a consequence, the mobility management solution focusing on user QoE in 5G & B5G enabled network has become a crucial issue to be addressed. Within this paradigm, a scheme must be designed with the following goals: (i) to promote the end-user personalised experience, (ii) to adapt the diversity of service requirements and real-time dynamic information, (iii) to effectively optimise the network resources. To comprehend the aforementioned goals, a MARL based approach is introduced to

realise intelligent RAT selection scheme that authenticates the QoE provisioning with service requirement maintenance. Moreover, a novel SDWN-Edge powered dynamic framework with intrinsic flexibility is delineated to ensure effective real-time decision making for a smart healthcare environment. The following contributions has been made in this paper to achieve an optimal RAT selection policy:

- (i) A novel data-driven layered architecture has been proposed for application agnostic RAT selection in smart healthcare systems.
- (ii) A service aware quantised QoE model has been developed to ensure QoE provisioning with optimal resource utilisation.
- (iii) A generalised and flexible multi-agent reinforcement learning approach has been leveraged with regard to users and services.
- (iv) The exhaustive simulations have been conducted to validate the performance of the proposed scheme.

The innovation in the proposed approach resides in the intelligent framework that employs a novel context-aware quantised QoE model and invalid action masking scheme to facilitate fine-grained RAT selection. The context-aware quantised QoE model allows better user experience while guaranteeing efficient resource utilisation whereas the invalid action masking scheme eliminates the invalid RATs to achieve faster convergence to an optimal solution. Moreover, the proposed framework leverages the concept of SDWN and edge computing to ensure efficient service and network orchestration. The exploitation of SDWN allows centralised computing management to add flexibility, adaptability and manageability in the proposed framework whereas the edge computing ensures storage capabilities and faster processing of real-time applications that are critical to the smart healthcare network. In addition to it, each agent invokes double deep Q policy network to conduct a reasonable context aware network selection with the comprehension of patient's preferences. Therefore, the proposed scheme ensures QoE aware RAT selection with guaranteed network resource optimisation.

The rest of the paper is organised as follows. A background scrutinizing the characteristic 5G-enabled healthcare use-cases is presented in Sect. 2. Section 3 illustrates the proposed architecture and elaborates the communication flow between various entities in the proposed framework. Subsequently, Sect. 4 formulates the QoE optimised RAT selection problem and scrutinizes the proposed approach for its optimisation. The Sect. 5 discusses the empirical evaluation of the proposed scheme and its comparative scheme with other instinctive schemes. Finally, a concise conclusion has been presented in the last section.

2 Background

The overwhelming proliferation of healthcare service requests due to COVID-19 has created an unprecedented stress on the communication networks. To ensure essential levels of connectivity for smart health at substantial scale, 5G emerged as a sophisticated connectivity solution which powers healthcare providers and empowers smart healthcare delivery models at more convenient care. Therefore, in accordance with the most neoteric scientific literature; three major representative 5G powered e-health applications along with their stringent communication requirements has been identified and synthesised in this section. Each e-health application has been articulated into different representative scenarios and the same are discussed below in detail:

- (i) **Pervasive Monitoring:** It refers to the transmission of physiological and biovital signals of the patient to the clinicians and medical staff for the continuous monitoring of the patient health. The heterogeneous bio-signals collected through the various wearable and unobtrusive sensors constituting the WBAN, is uploaded to the electronic health record in the medical network for the continuous health monitoring (Malasinghe et al. 2017). The telemonitoring in the patient care centre generates the traffic (including physiological parameters) that demands data rate upto 300 Mbps, latency and jitter of the order of 250 ms and 25 ms, respectively and $1-10^{-3}$ reliability (Cisotto et al. 2020; Thuemmler et al. 2016).
- (ii) **Video Consultation:** This e-consultation aspect of telemedicine delivers diagnostic and therapeutic services through the establishment of a video interaction communication link between the patient and specialists. For teleconsultation, it is necessary to provide seamless connection with suitable audio and video capabilities between patient and clinicians along with the transmission of the heterogeneous biovital signals. For the real-time interactivity and low-delay communication, these applications tolerate some packet loss as a tradeoff. Therefore, to deliver these services, data rate of about 1 Gbps, along with the latency of the order of 20 ms, small jitter (10 ms) and reliability higher than 10^{-3} is required (Cisotto et al. 2020; Kim et al. 2019).
- (iii) **Wireless service robots:** Intelligent robots and autonomous agents delivers AI based diagnosis and telesurgery for the management of patients. Furthermore, it reproduces Ultrasonography, Magnetic Resonance Images (MRI), Computed Tomography (CT), X-ray scan and offers seamless assistance to the patients (Zhang et al. 2018). With the objective of handling system immanent latency and security,

this class of e-health requires highly reliable and low latency connectivity. To process real-time data including massive collections of images, videos and sensory data, about 1 Gbps of data rate, end-to-end (E2E) latency of the order of 1 ms and approximately $1-10^{-7}$ reliability is recommended (Cisotto et al. 2020; Simsek et al. 2016; Imran et al. 2020).

The characteristic telehealth use cases and their stringent QoS requisites discussed above are summarised in Fig. 1. The services and system aspect group of the third generation partnership project (3GPP) has described the key performance indicators required to characterise the next-generation healthcare services which are discussed in Table 1.

3 5GhNet:an intelligent QoE aware RAT Selection framework for 5G-enabled healthcare network

In 5G & B5G enabled healthcare network, QoE management will be a daunting task, as QoE is anticipated to be competently and autonomously managed for each patient and corresponding demanded service. Therefore, a data-driven

Fig. 1 5G-enabled healthcare scenarios communication scenario and their stringent requirements

Table 1 List of considered attributes and their description

Attributes	Description
User-experienced	Signifies the minimum achiev
Data rate (Mbps)	Able data rate for a user in real network for the satisfaction of specific service
End-to-end latency (ms)	One-way delay that accounts the time taken by the packet to travel from source to destination
Jitter (ms)	Measures the uncertainty in latency due to non-idealities and errors in communication
Packet loss rate (10^{-x})	Measures the probability of successfully delivering network layer packets within the given time constraints in accordance with the demanded services

architecture has been proposed for personalised QoE management in the next-generation healthcare network. Initially, the prerequisites to be fulfilled by the proposed architecture to capture the subjective characteristic of the user are discussed below:

- The architecture must incorporate a definite cognitive computing capability.
- The architecture must leverage a data-driven scheme that can predict the user demand.
- The architecture must efficiently manage the communication resources based on the QoS requirements and the predicted user-centric service to maintain a satisfactory QoE.

Specifically, an autonomic architecture presented in Fig. 2 is logically partitioned into five layers namely (i) Infrastructure layer, (ii) RAT abstraction layer, (iii) Edge computing layer, (iv) Software-defined wireless networking layer and (v) e-health Cloud layer. The proposed architecture works right from the physiological data collection phase to RAT selection decision-making. The detailed explanation about each layer is given below:

- Infrastructure layer:** For the continuous collection of the patient's biovital signal information, WBAN constituting smart clothing, body wearable sensors, data sensing and acquisition devices is deployed in the patient care unit. The real-time physiological data of each patient including blood oxygen saturation level (SpO₂), heart rate, temperature, respiration rate, systolic blood pressure (Mukherjee et al. 2021) collected from the WBAN is transmitted to the network through patient-trusted gateway i.e. android smartphone and smart tablets (Rahmani et al. 2017). These smart terminals comprise two submodule namely Sensor Data Collector and Network Context Discovery Component (Varga et al. 2015). The former collects and store the real-time physiological data of patient from the WBAN whereas the latter furnish dynamic information of the available RATs

Fig. 2 5GhNet: an intelligent QoE aware RAT Selection framework for 5G-enabled healthcare network

including data rate, E2E latency, jitter and packet loss rate (PLR) in accordance with the patient. As the data collected by both the submodules is private in nature. Therefore, the sensor data collector message is protected through a uniquely identifiable Patient ID scheme (Tartarini et al. 2017; Al-Janabi et al. 2018; Chen et al. 2018) as presented in Figs. 3 and 4 respectively.

With the differentiated Patient ID, the Data Processing Component merges and synchronises the data received from both the sensor data collector and network context discovery component for each patient at regular interval as depicted in Fig. 5. Meanwhile, the healthcare professional will receive the health analysis results from edge computing layer.

Fig. 3 Structure of Sensor Data Collector message

- (ii) **Radio Node Layer:** The management of RATs in multiple RAT system is controlled by a set of different entities which leads to suboptimal utilization of the overall network resources. Therefore, the RAT Abstraction layer acts as a unified single entity that handles the RAT specific functionality within the network. Moreover, it manages RAT specific control plane communication with users and possesses both

Fig. 4 Structure of the Network Context Discovery Component message

Fig. 5 Structure of the Data Processing Component message

Patient ID	Patient Trusted Gateway MAC address	Body Temperature (°C)	Heart Rate (/min)	Respiration Rate (/min)	Blood Oxygen Saturation Level (SpO2 %)	Systolic Blood Pressure (mmHg)	Patient network status vector
------------	-------------------------------------	-----------------------	-------------------	-------------------------	--	--------------------------------	-------------------------------

management and control functionality (Manjeshwar et al. 2019). Radio nodes, such as the gNodeB (gNB) in 5G New Radio (5G NR), or an evolved NodeB in Long-Term Evolution (LTE), Radio over Fiber (RoF) radio access points and LoRa access points provide RAN services to the end devices. Apart from 5G NR, LTE, RoF and LoRa, there exists a wide range of near-field wireless access technologies, including ZigBee, Bluetooth. All these radio nodes are controlled by the command received from the SDWN layer through southbound interface.

- (iii) Edge computing layer: The data generated in the infrastructure layer is volume and the user demand is heterogeneous and delay-sensitive. Therefore, local processing is realised through the edge computing layer to improve user experience for an intensive e-health applications. It comprises computing engine to perform fast and lightweight data processing for computation-sensitive, data-sensitive and delay-sensitive tasks (Chen et al. 2018; Lloret et al. 2017). It includes three submodules viz. real-time data collector, service cognitive engine and network cognitive engine (Hao et al. 2018). The main function of real-time data collector is to extract and relay the external data including real-time physiological data and inter-

nal data comprising network statistics to their respective conditioning units i.e. service cognitive engine and network cognitive engine. Both these engines employ data-driven scheme that allows cognition to external as well internal data for the sake of context-awareness and QoS maintenance for smart healthcare applications respectively.

Service Cognitive Engine: The service cognitive engine carries out the extensive big data analysis (Hadi et al. 2020) to execute smart healthcare as presented in Fig. 6. The core of the service cognitive engine is the multilayer perceptron layer (MLP) model trained with the database present in the e-health cloud that classifies the patient's real-time physiological data into health risk grades namely low, medium and high. These disease risk level are further mapped to the patient-centric services as mentioned in Table 2 (Chen et al. 2018).

Network Cognitive Engine: In case of ultra dense 5G HetNets, more than one RAT can probably get shortlisted to offer RAN services in accordance with the demanded service. Therefore, the network analysis component leverages the MLP classifier over the network status map to indicate the suitability of a RAT to offer a characteristic application

as mentioned in Table 3. This scheme of classifying RATs into valid and invalid actions, termed as invalid action masking (Priya et al. 2020), that eliminates the sub-optimal actions for the robustness and faster learning of the proposed model. Moreover, it calculates and return the reward *SQoE* to the SDWN layer. Furthermore, the service request generator combines this valid dynamic network status map with the service request relayed by the service cognitive engine

to generate a network cognitive engine message defined in Fig. 7 which is transmitted to the SDWN layer.

- (iv) Software-defined wireless networking layer: This layer simply implements a distributed control plane modelled as multi-agent system to address the extensibility issue present in the single controller system (Sun et al. 2020). Amidst the multiple controllers present in the SDWN layer, each controller manages a subset of RATs leading

Fig. 6 Big data analysis in edge computing layer

Table 2 Risk evaluation and corresponding service reference

Physiological parameters	Respiration rate (/min)	Heartbeat (/min)	Temperature (°C)	SpO2 (%)	Systolic pressure (mmHg)	Patient-centric services
Risk level						
Low risk	[12–20], [21–25] or [10–11]	[60–90], [91–100] or [51–59]	[36.0–37.0], [37.1–37.5] or [35.5–35.9]	≥95 or [90–94]	[89–139] or [140–159]	Pervasive monitoring
Medium risk	[26–35]	[101–140] or [41–50]	[37.6–38.9] or [35.0–35.5]	[85–90]	[160–179] or [85–89]	Video consultation
High risk	[36–50] or [6–9], >50, or ≤5	[141–160] or [30–40], >160 or <30	[39.0–39.9] or <35.0, ≥40.0	[80–85], <80	[180–189] or [80–84], ≥190 or <80	Wireless robotic care

Table 3 QoS requirements of considered service classes

Scenario	User-experienced data rate	E2E latency (ms)	Jitter	PLR
Pervasive monitoring	0.010–300 Mbps	<250 ms	<25	$1 - 10^{-3}$
Video consultation	0.137–1 Gbps	5–20 ms	<10	$1 - 10^{-3}$
Wireless robotic care	1–2 Gbps	0.5–1 ms	0.5–1	$1 - 10^{-7}$

Fig. 7 Structure of Network Cognitive Engine message

to efficient network resource management and better service request processing (Shantharama et al. 2018). As a solution to QoE aware RAT selection problem, each constituent controller invokes the DDRL policy network on the reception of a network cognitive engine message to achieve fine-grained RAT selection policy. After a successful training phase, the MARL system realised in the distributed control plane relay faster decision on the optimisation problem.

- (v) e-health Cloud: The e-health cloud runs critical applications to realise efficient and secure healthcare network, such as it offers visualization of medical data and records to the clients (including medical staff and patient) through a dedicated gateway (Patel et al. 2015; Al-Janabi et al. 2017, 2020). Moreover, the electronic health record comprising medical information of the patient, medical history, underlying disease and its

nature is utilised at the edge server to train the big data analytics engine for estimating the patient's health risk level.

The communication flow diagram presented in Fig. 8 elucidates the interactive process between user-side and rest of the layers. The real-time physiological data captured by the sensors deployed in patient-care unit is collected by the patient-trusted gateway through near-field communication technology like zigbee and bluetooth. Furthermore, these patient-trusted gateways offload the patient's real-time physiological data along with network statistics including data rate, E2E latency, jitter and PLR of each available RAT to the edge computing layer. The edge computing layer realises the cognitive intelligence over the collected physiological data of patient in the big data analytics engine to predict the disease risk namely low, medium, high, which is further mapped to characteristic telehealth service. Subsequently, the edge computing layer transmits this service request along with the valid network map processed by the network cognitive engine to the radio access network intelligent controller (RANC). In order to manage multiple telehealth services at a time, task monitoring module present in RANC distributes the services as a task among the multiple controllers. Each controller implements a double deep Q network which

Fig. 8 Communication flow between different entities of the proposed architecture

relays optimal network selection decision to the considered environment through southbound interface on the basis of real-time network status and patient demand.

4 Mathematical analysis

This section provides a comprehensive description of the mathematical approach primarily focussing on the service-aware QoE optimised RAT selection model. Subsequently, the RAT selection problem is formulated, followed by the investigation of MARL method to achieve near-optimal policy for the optimisation problem.

4.1 Service-Aware quantised QoE model

Most of the research on QoE calculation is based on the subjective evaluation focusing on the direct feedback from the user. But the feedback collection and data collection is restricted by the controlled environment. Therefore, the objective method is inferred to describe the subjective experience. In the light of above-mentioned, a QoE modelling approach is adopted that employs a weighted sum scheme defining the multi-objective function into a single aggregate objective function $SQoE$ (Wang et al. 2017; Hao et al. 2018). It includes all key performance indicators (KPIs) i.e. data rate, E2E latency, jitter and PLR associated with the distinct radio access network. Nevertheless, each KPI has different dimensions and measurement units. Therefore, these quantities are normalised and depicted as their respective QoS ratios and discussed below in QoE model:

Definition 1 Assume that $b(i,j)$ and $B(j)$ (where $i \in patient$, $j \in validRAT$) respectively denotes the minimum data rate requirement of the service and data rate offered by the chosen radio access technology. Then, the QoS ratio for data rate (QoS_d) is given as:

$$QoS_d = B(j)/b(i,j) \quad (1)$$

Definition 2 Assume that $d(i,j)$ and $D(j)$ (where $i \in patient$, $j \in validRAT$) respectively denotes the maximum tolerable delay requirement of the service and E2E latency offered by the chosen radio access technology. Then, the QoS ratio for E2E latency (QoS_d) is given as:

$$QoS_d = d(i,j)/D(j) \quad (2)$$

Definition 3 Assume that $g(i,j)$ and $G(j)$ (where $i \in patient$, $j \in validRAT$) respectively denotes the maximum tolerable jitter requirement of the service and jitter offered by the chosen radio access technology. Then, the QoS ratio for jitter (QoS_g) is given as:

$$QoS_g = g(i,j)/G(j) \quad (3)$$

Definition 4 Assume that $p(i,j)$ and $P(j)$ (where $i \in patient$, $j \in validRAT$) respectively denotes the maximum tolerable packet loss rate requirement of the service and packet loss rate offered by the chosen radio access technology. Then, the QoS ratio for PLR (QoS_p) is given as:

$$QoS_p = p(i,j)/P(j) \quad (4)$$

Definition 5 A user personalised QoE evaluation model (Ahmad et al. 2016; Hemmati et al. 2017; Zhang et al. 2019) is developed using the sigmoid function for the given QoS ratios namely QoS_d , QoS_l , QoS_g , QoS_p and is given as:

$$SQoE = \frac{\alpha}{1 + \exp(-\beta * (w_b * QoS_b + w_d * QoS_d + w_g * QoS_g + w_p * QoS_p))} \quad (5)$$

where α and β are the parameters constraining the quantization of $SQoE$ and w_b , w_d , w_g and w_p represents the weights decided on the basis of service correspondingly for data rate, E2E latency, jitter and PLR. These positive weights for each service class mentioned in Table 2. are calculated using Analytical Hierarchy Process (AHP). Pervasive monitoring demands high reliability and low delay, jitter requirements. Therefore, the correspondingly weights for it are defined as $w_{pm} = [0.16 \ 0.20 \ 0.10 \ 0.54]$. On contrast, video consultation service is characterised by strict QoS requirements in terms of delay and jitter to avoid impermissible level of quality and the weights for the same are $w_{vo} = [0.10 \ 0.42 \ 0.42 \ 0.06]$. Analogously, low E2E latency and high reliability is recommended for wireless robotic care and the same can be verified from the weights, given as $w_{wr} = [0.15 \ 0.45 \ 0.10 \ 0.30]$.

5 Problem formulation

The 5G & B5G enabled networks experiences heterogeneous service requirements and regards QoE as a prime measurement approach to corroborate the communication services. Therefore, the objective is to realise an optimal user-RAT association that maximises the personalised user-experience. In light of above said, the QoE optimised RAT generalised selection problem ζ is formulated as

$$\zeta : \max \sum_{j \in validRAT} \sum_{i \in patient} SQoE * y(i,j) \quad (6)$$

subject to :

$$C1 : \sum_{i \in patient} b(i,j) * y(i,j) < B(j), \forall j \in valid \ RAT \quad (7)$$

$$C2 : \sum_{i \in \text{patient}} d(i, j) * y(i, j) > D(j), \forall j \in \text{valid RAT} \quad (8)$$

$$C3 : \sum_{i \in \text{patient}} g(i, j) * y(i, j) > G(j), \forall j \in \text{valid RAT} \quad (9)$$

$$C4 : \sum_{i \in \text{patient}} p(i, j) * y(i, j) > P(j), \forall j \in \text{valid RAT} \quad (10)$$

$$C5 : \sum_{j \in \text{valid RAT}} y(i, j) \leq 1, \forall i \in \text{patient} \quad (11)$$

The optimisation problem formulated in Eq. (6) can be optimised through the proper user-RAT association in the considered network. The constraints defined in Eqs. (7)–(10) ensures the better QoS requirements maintenance of the services demanded by the patient and adjust the radio network utilisation in terms of data rate, E2E latency, jitter, PLR within certain physical capabilities whereas the Eq. (11) guarantees that only a single RAT must be assigned to a user at given t time.

6 QoE optimised RAT selection framework

To address the QoE optimised RAT selection problem (RSP) discussed in the previous section, an AI driven framework is presented. Furthermore, a brief introduction over MARL is discussed, followed by the designing of its key elements. Finally, the MARL based RAT selection algorithm is proposed. The symbols and their conventions considered through the mathematical modelling and constructing the algorithm is presented in Table 4.

The proposed framework mainly distributes its logic between edge computing and SDWN layer to achieve QoE optimised RAT selection scheme. The former layer implements MLP classifiers for the identification of service and valid RATs whereas latter employs the concept multi-agent

reinforcement learning to achieve fine-grain association policy. Therefore, an extensive description of the mathematical approaches implemented at both the layer is discussed in the upcoming subsections.

6.1 Edge computing layer

The fundamental parts of the edge computing layer are service cognitive engine and network cognitive engine which identifies the characteristic service class and valid RATs respectively. The service cognitive engine employs a MLP classifier M_b in the big data analytics engine which classifies the real-time physiological data PD_b into a risk level and the same is defined as

$$M_b = f_b(PD_b : u_b(v : \tau_b)) \quad (12)$$

where $f_b(\cdot)$ and $u_b(\cdot)$ correspondingly are the activation function and parameter set for the big data analytics engine. On the other hand, v and τ_b respectively denotes the layer size and the variable parameters. The input layer of big data analytics engine comprises five neurons followed by a hidden layer comprising 100 neurons with *tanh* activation function (Yamamoto et al. 2020), whereas the output layer comprising 3 neurons is setup with *softmax* activation function (Vinayakumar et al. 2019). All the layers are defined as $y = f_l(KgT + h)$, where f_l depicts activation function, K represents the weight in the layer g and h is the bias vector. The risk level classified by the big data analytics is further mapped into a characteristic service class as described in algorithm 1. On the other hand, the network cognitive engine implements a MLP classifier M_n in the network analysis component to identify valid and invalid RATs on the basis of NS_n network status map of n RATs and the same is defined as

$$M_n = f_n(NS_n : u_n(\xi : \tau_n)) \quad (13)$$

here $f_n(\cdot)$ and $u_n(\cdot)$ respectively are the activation function and parameter set for the network analysis component whereas ξ and τ_n denotes the layer size and the variable parameters for the same. The input layer of the network analysis component includes four neurons followed by a hidden layer which comprises 100 neurons with *tanh* activation function. Subsequently, the output layer is set up with two neurons and *softmax* activation function (Rustam F et al. 2020). Subsequently, the network analysis component calculates the *SQoE* on the selection of corresponding valid RATs and returns it as a reward to the respective local controller as discussed in algorithm 2.

Table 4 Symbols and their conventions

Symbols	Description
s_t	State
a	Action
r	Reward
ϵ	Exploration rate
γ	Discounted factor
θ	Online network weights
θ'	Target network weights
b_m	Mini-batch size
D	Experience buffer

Algorithm 1 Big Data Analytics Engine

Input: real-time physiological data PD_b
Output: User-specific service S

```

1: for every decision point do
2:    $\forall b \in [1, i]; M_{PD_b} = f_b(PD_b : u_b(\nu : \tau_b))$   $\triangleright$  Identification of risk level using MLP
   classifier trained with the e-health database
3:   if  $M_{PD_b} == Low\ Risk$  then
4:      $S == Pervasive\ Monitoring$ 
5:   else
6:     if  $M_{PD_b} == Medium\ Risk$  then
7:        $S == Video\ consultation$ 
8:     else
9:       if  $M_{PD_b} == High\ Risk$  then
10:         $S == Robotic\ Care$ 
11:      end if
12:    end if
13:  end if
14: end for

```

Algorithm 2 Network Analysis Component

Input: User-specific service S , patient's network status NS_n
Output: reward $SQoE$

```

1: for every decision point do
2:    $\forall n \in [1, RAT_n]; M_{NS_n} = f_n(NS_n : u_n(\xi : \tau_n))$   $\triangleright$  classification of each RAT into a
   suitable service class on the basis of patient's network status
3:   if  $M_{NS_n} == S$  then
4:      $A = RAT_j$   $\triangleright$  valid action
5:   else
6:      $I = RAT_q$   $\triangleright$  invalid action
7:   end if
8:   for every decision point do
9:      $\forall a \in A; SQoE = \alpha / (1 + \exp(-\beta * (w_b * QoS_b + w_d * QoS_d + w_g * QoS_g + w_p * QoS_p)))$ 
     $\triangleright$  Service-aware QoE calculation
10:    return  $SQoE$  as a reward to the respective local controller
11:   end for
12: end for

```

Fig. 9 System model

Fig. 10 Decentralised partially observable Markov decision process

6.2 Software-defined wireless networking layer

The system model illustrated in Fig. 9, exhibits a SDWN based RAT association solution highlighting the two-tier heterogeneous controller approach. The SDWN layer is logically distributed into k local controllers supervised by a global controller which augments the better radio resource management and authorization of k distributed local controllers (Xu et al. 2019; Manjeshwar et al. 2019). Each local controller governs a set of RATs, forming a part of dynamic network. As a consequence of distributed property, the QoE optimised RAT selection problem is formulated as a Decentralised Partially Observable Markov Decision Process (DPOMDP) which is depicted in Fig. 10 and this genre of problem can be comprehended with the assistance of reinforcement learning. Therefore, the investigation in this paper considers the SDWN layer as a multi-agent system that models each local controller as a DDRL agent to tackle the RSP in a distributed fashion. In MARL, each DDRL agent interacts with DPOMDP modelled as $M=(s,a,r,\gamma)$ where s,a,r,γ respectively denotes state, action, reward and discount factor. Each DDRL agent observes state s , based on which action a is selected. Subsequently, the effect of action a is translated into a reward r which is disclosed to the agent to further improve its optimal policy and the definition for the same is defined as:

- (i) **State Space:** The state perceived by each DDRL agent from the edge computing layer at time t represents the well-defined service S and valid network status NS_{vm} with respect to a specific patient. The valid real-time network status is the collection of network parameters that includes data rate $B(j)$, E2E latency $D(j)$, jitter $G(j)$ and PLR $P(j)$ associated with each valid RAT in the network intensive environment.

Hence, the state space s defined for a DDRL agent at time t is

$$s_t^j = \{S, B(j), D(j), G(j), P(j)\}, j \in \text{valid RAT} \quad (14)$$

and s_t denotes the overall state of the system (Bhattacharya et al. 2019),

$$s_t = \left[\bigcup_{j \in \text{valid RAT}^j} s_t^j \right] \quad (15)$$

- (ii) **Action Space:** In obedience to every state $s \in s_t$, each DDRL agent implements an action $a \in A$, denoted as

$$A = \{RAT_1, RAT_2, \dots, RAT_p, \dots, RAT_{j-1}, RAT_j\} \quad (16)$$

It can be noticed that actions available at each step is the list of valid RATs shortlisted by the network cognitive engine in the dynamic healthcare environment.

- (iii) **Reward Design:** After implementing each valid action a in each step, the network environment feeds back a reward r to the agent in certain state s at time t . As model free MDP method is assumed in which the deciding agent always aim to maximise the accumulated rewards without considering the transition probabilities. Each state-action pair will have a value $Q(s, a)$ i.e. the expected discounted reward for state s and action a . Specifically, the reward function must be related to the objective function. The main objective of the proposed work is to maximise the personalised QoE of the user with efficient radio resource utilisation. To achieve these multiple objectives, the reward function is designed as follows:

$$QoS_{Benefitcriterion} = \begin{cases} -1, b_{min} \geq B, b_{max} < B \\ B/b_{min}, b_{min} < B \leq b_{max} \end{cases} \quad (17)$$

$$QoS_{Costcriterion} = \begin{cases} -1, c_{min} > C, c_{max} \leq C \\ c_{max}/C, c_{min} \leq C < c_{max} \end{cases}, \quad (18)$$

here benefit criterion corresponds to data rate whereas cost criterion denotes E2E latency, jitter, PLR and the reward r is given as:

$$r = SQoE = \frac{\alpha}{1 + \exp(-\beta * (w_b * QoS_b + w_d * QoS_d + w_g * QoS_g + w_p * QoS_p))} \quad (19)$$

The QoS ratio defined for benefit criterion in Eq. (17) clearly elucidates that when the minimal requirement of user specific service exceeds the available benefit criteria such as data rate of the considered RAT, then the user fails to access the given RAT leading to poor user experience. Similarly, when the benefit criterion of the RAT is greater than the maximum demand of user, the user experience doesn't grow due to consistent user satisfaction. But better user experience is only guaranteed when the available benefit criterion of RAT lies between the maximum and minimum demand of service (Du et al. 2020). On the contrast, the QoS ratio defined in Eq. (18) clearly defines that the user experience declines when the cost criterion such as E2E latency, jitter and PLR of RAT exceeds the maximum requirement of the user-specific service. Moreover, the constraints defined in Eqs. (17) and (18) are clearly favourable to the rational utilisation of network resources. Therefore, in this way the reward design analyse the user QoS demand for optimising QoE, successively leading to efficient resource management. For the sake of more clarity, resource utilisation factor (RUF) has been introduced to quantify the efficient resource utilisation and is discussed below:

$$RU_{Benefitcriterion} = \begin{cases} -1, b_{min} \geq B, b_{max} < B \\ 1, b_{min} < B \leq b_{max} \end{cases} \quad (20)$$

$$RU_{Costcriterion} = \begin{cases} -1, c_{min} > C, c_{max} \leq C \\ 1, c_{min} \leq C < c_{max} \end{cases}, \quad (21)$$

here $RU_{Benefitcriterion}$ and $RU_{Costcriterion}$ denotes resource utilisation for benefit criterion and cost criterion respectively. To optimise the overall resource utilisation in 5G-enabled healthcare network, the resource utilisation factor (RUF) is defined as

$$RUF = w_b * RU_b + w_d * RU_d + w_g * RU_g + w_p * RU_p, \quad (22)$$

here RU_b, RU_d, RU_g, RU_p corresponds to resource utilisation factor related to bandwidth, delay, jitter and PLR. Therefore, in this way $SQoE$ defined as reward in Eq. (19) turns out to be a comprehensive metric that ensures personalised user experience through better network resource management in such an elaborative healthcare environment.

Multi-agent reinforcement learning is a distributed version of a single-agent reinforcement learning and expertise at taking dynamic actions in multi-task system (Zhang et al. 2019). The distributed processing nature of MARL empowers it for interactive process and network selection decision in distributed control plane of SDWN layer. After sufficient training, the multiple agents in control plane ensure faster decision on the problem discussed in section 5. A DRL algorithm has been a well adopted method in comparison to RL algorithms due to its ability to handle large space size (François-Lavet et al. 2019). The DQN-based model takes advantage of target and online network exploiting deep neural network (DNN) to stabilize the whole performance. The input to DNN is the state described in section 6.2 and the output comprises Q-values $Q(s, a : \theta)$ for all viable actions a where θ signifies the DNN weights. To derive the approximates value $Q^*(s, a)$, DNN is trained through experiences defined as (s, a, r, s') . Specifically, the DQL algorithm updates its DNN's weight θ to reduce the loss function, represented as

$$F(\theta) = E \left[\left(y_i^{DQN} - Q_i(s, a_i; \theta) \right)^2 \right] \quad (23)$$

where $y_i^{DQN} = r + \gamma \max_{a' \in A} Q_i(s', a'; \theta')$ and θ' denotes the target network weights. The action a_i is selected from the online network $Q_i(s, a_i : \theta)$ through ϵ -greedy policy (El Helou et al. 2015). Even though, the target network is a replica of $Q_i(s, a_i : \theta)$, its weights are constant whereas the online network weights are updated for a number of iterations. DQN leverages the experience replay strategy which breaks the correlation between the consecutive samples to avoid learning instability. The sampling of mini-batches b_m from experience buffer D is utilised to train the deep neural network. The selection of mini-batch size is crucial in tuning deep learning systems as the small batch size show faster convergence to good solutions and low generalisation error in comparison to higher batch size (Zhao et al. 2019). The over-optimistic estimation is clearly elucidated in Eq. (23) where the max operator applies on the same Q-value for action selection and action evaluation leading to inaccurate derived policy. Therefore, Double DQN (DDQN) algorithm (Hasselt et al. 2016) is exploited which employs two separate DQN networks to decouple the action selection and action evaluation. The target function for DDQN is defined as

$$y_i^{DDQN} = r + \gamma Q(s', \arg \max_{a' \in A} Q_i(s', a'; \theta); \theta') \quad (24)$$

Explicitly, both the online and target network utilises the next state s' to calculate the optimal $Q_i(s'; a'_i; \theta)$ value. In the proposed work, each controller employs a DDQN that comprises three layers namely an input, hidden and an output layer. Correspondingly, the DDQN's mapping function with input ρ is defined as

$$Y = f_{dd}(\rho, a_i : u_{dd}(\mu : \tau_{dd})) \quad (25)$$

a_i is the action whose Q-values has to be computed, $u_{dd}(\cdot)$ signifies parameter set while μ is associated to the model scale and τ_{dd} is the varying parameters. Finally, with the given values of discount factor γ and reward r , target value y_i^{DDQN} is calculated. The choice of discount factor plays a crucial role in learning as it signifies the importance given by the agent to the future rewards in comparison to the instantaneous future. Specifically, $\gamma=0$ makes the agent myopic by

only considering current rewards, whereas $\gamma=1$ enables the agent to consider future rewards with more weightage making the policy to converge slowly (Zhao et al. 2019). Then, Adam optimisation is employed to update online network weights θ on the basis of $F(\theta)$ i.e. loss function and defined as

$$F(\theta) = E\left(y_i^{DDQN} - Q_i(s, a_i : \theta)\right)^2 \quad (26)$$

Adam optimisation is preferred over stochastic gradient descent algorithm in case of sparse gradient problems as it calculates independent adaptive learning rate (Saraiva et al. 2020; Bhattacharya et al. 2019) and efficiently reduces the loss function described in Eq. (26). The overall algorithm of multi-agent DDQN implemented in RANC is presented below:

Algorithm 3 Multi-agent DDQN Algorithm for the selection of appropriate RAT

Input: valid network state s_t , mini-batch size b_m
Output: association to suitable RAT_j
Initialise: network parameters, experience replay memory D , learning rate α , discount factor γ , exploration rate ϵ , online network weights θ , target network weights θ' , online action-value function and target action-value function.

- 1: **for** $e = 1$ to ζ **do**
- 2: Initialise the network state to s_0
- 3: **for** iteration = 1:T steps in episodes ζ **do**
- 4: **if** exploration **then**
- 5: Each agent choose a random action $a \in A$ with probability ϵ
- 6: **else**
- 7: choose $a = \operatorname{argmax}_a Q(s_t, a; \theta)$
- 8: **end if**
- 9: Each agent implements action a_t , receive an immediate reward r_t and execute next state s_{t+1}
- 10: Each agent stores the memory segment comprising (s_t, a_t, r_t, s_{t+1}) into experience buffer D
- 11: Each agent samples the mini-batch of b_m tuples (s_i, a_i, r_i, s_{i+1}) from experience buffer D
- 12: Each agent sets $y_i^{DDQN} = r + \gamma Q(s', \operatorname{argmax}_{a' \in A} Q_i(s', a'; \theta); \theta')$
- 13: Each agent updates θ by performing gradient descent step on $(y_i^{DDQN} - Q(s_i, a_i, \theta))^2$
- 14: Each agent replaces target parameters $\theta' = \theta$ at every C iterations
- 15: **end for**
- 16: Update experience replay of each agent
- 17: **end for**

The proposed method is composed of three major parts namely: service identification, RAT classification and association policy learning. The service identification and RAT classification is respectively carried out by the service cognitive engine and network cognitive engine present in the edge

computing layer. On the other hand the distributed DDRL agents present in SDWN layer are responsible for the association policy learning. The service cognitive engine comprises big data analytics engine which implements a MLP classifier M_b to classify the real-time physiological data into

Table 5 Network system model parameters

Parameter	Chosen values
Considered RATs	[5G NR, LTE, RoF, LoRa]
b_{min}	[300 Mbps, 75 Mbps, 1000 Mbps, 0.02 Mbps]
b_{max}	[700 Mbps, 300 Mbps, 100,000 Mbps, 0.05 Mbps]
d_{min}	[10 ms, 80 ms, 0.01 ms, 101 ms]
d_{max}	[30 ms, 100 ms, 0.1 ms, 150 ms]
g_{min}	[1 ms, 1 ms, 10^{-14} ms, 1 ms]
g_{max}	[20 ms, 20 ms, 10^{-9} ms, 20 ms]
p_{min}	[10, 10, 10^{-20} , 15]
p_{max}	[15, 15, 10^{-9} , 20]
Requested applications	[Pervasive monitoring, video consultation and robotic care]

Table 6 Selected hyperparameter for DDRL agent

Hyperparameter	Chosen values
Possible state space	3
Mini-batch size	32
Initial exploration rate	1.0
Final exploration rate	0.1
Discounted factor	0.90
Learning rate	0.01
Double learning	On
Control policy	ϵ -greedy policy
Optimiser	Adam

a disease risk namely low, medium, high, which is further mapped to a characteristic telehealth service. On the other hand, the MLP classifier M_n implemented in network cognitive engine classifies the network status map into valid and invalid RATs. The characteristic service class along with the valid network status classified by the edge computing layer serve as a state to the DDRL agent employed in the logically distributed SDWN layer. Consequently, the DDRL agent selects an action among the valid RATs and accordingly, receives a reward in terms of $SQoE$. The reward $SQoE$ serve as comprehensive metric that concentrates on the QoE optimisation which involves network resource management while guaranteeing satisfactory QoE levels. On the basis of the received reward, DDRL agent learns an QoE optimised RAT association policy.

7 Results and discussion

To estimate the performance of the proposed scheme, the simulation results are discussed in this section. Initially the considered simulation environment is presented for the evaluation of proposed mathematical framework and its solution through DDRL. Consequently, the comparison of the proposed algorithm with a DQN based, random, and greedy

Fig. 11 Comparison of proposed scheme with other schemes on the basis of $SQoE$ vs episodes

scheme is analysed on the basis of $SQoE$ and resource utilisation factor (RUF). Lastly, the impact of hyperparameters on the performance of the proposed scheme is scrutinised.

7.1 Simulation environment

In this section, the proposed mathematical framework and its solution is analysed by considering a 5G UDN environment equipped with numerous RAT access points such as 5G NR, LTE, RoF and LoRa. It is considered that the decision making for network selection is affected by the requested application namely pervasive monitoring, video consultation and robotic care. Based on the stringent requirements of these applications, the network parameters considered for the RAT selection are data rate (Mbps), E2E latency (ms), jitter (ms), packet loss rate (per 10^{-x}). To describe the data rate, E2E latency, jitter and PLR dynamics in network, a discrete model is leveraged in which these network parameters are approximated to several levels. In every slot, the combined data rate-E2E latency-jitter-PLR state for each RAT can be

one of all possible combinations with the same probability (Du et al. 2020). The user is assumed to be stationary, when the statistical distribution of the data rate-E2E latency-jitter-PLR state remains constant in independent slots. For the four RATs, the maximum and minimum value of these network parameters is described in Table 5. On the other hand, MIMIC III database (Johnson et al. 2019) is used to obtain time series of clinical measurements such as blood oxygen saturation level (%), heart rate (/min), respiration rate (/min), temperature ($^{\circ}\text{C}$), systolic blood pressure (mmHg). Consequently, hyperparameters chosen for the agent is presented in Table 6.

7.2 Convergence analysis

To evaluate the effectiveness of the proposed scheme, it has been compared with other RAT selection schemes namely DQN based scheme, random and greedy scheme on the basis of convergence statistics. In a DQN based scheme, each local controller is modeled as a DQN agent whose state space comprises the service request along with the network status related to all the RATs whereas the action space includes all the RATs present in the considered environment. Consequently, random scheme selects each RAT with an equal probability in comparison to the greedy scheme which selects the nearest RAT for each service request irrespective of its characteristics requirements.

The comparison among the introduced schemes in terms of SQoE and convergence speed is presented in Fig. 11. Specifically, at the initiation of the learning process both the proposed and DQN based scheme explores to acquire considerable reward information for each well-defined state which results in low value of SQoE. But with the

Fig. 12 Comparison of proposed scheme with other conventional schemes on the basis of Resource Utilisation Factor vs episodes

Fig. 13 The CDF of SQoE by traversing the considered 5G HetNet

increase in training episodes, both the schemes tend to follow the exploitation policy to select actions returning higher rewards. The proposed scheme demonstrates clear advantage over the other schemes in terms of convergence rate as it only requires 30 episodes to reach convergence while DQN based scheme converges in 60 episodes. The reason behind the faster convergence of the proposed scheme is the symbiotic integration of invalid action masking and DDRL approach. The former allows the function approximation to acquire simple mapping (i.e. only the Q-values associated to valid RATs) whereas latter reduces the probable over-estimation errors by implementing the argmax operation only over the valid actions for faster convergence. On the other hand, DQN based approach suffers from a catastrophic interference phenomenon that adversely affects its learning

Fig. 14 Comparison of proposed scheme with other conventional schemes on the basis of average SQoE

Fig. 15 The CDF of Resource Utilisation Factor traversing considered 5G HetNet

Fig. 16 Comparison of proposed scheme with other conventional schemes on the basis of Average Resource Utilisation Factor

stability resulting in slower convergence rate. Moreover, the proposed approach allows more accurate value estimates i.e. higher $SQoE$ and better policy due to the existence of

DDQL algorithm that intelligently traverse the considered parameter space to attain the maximum $SQoE$ in less training episodes. On the contrast, DQN based approach exhibit unstable performance due to the maximum overoptimistic estimates that adversely affects its quality of learning. In case of random scheme, the RATs are randomly selected for a specified service request which minimises the number of suitable RAT selection leading to reduced $SQoE$ whereas the greedy algorithm always selects a RAT whose benefit and cost criterion exceeds the maximum and minimum demand of the service request resulting in unacceptable performance.

To further analyse the behaviour of four considered schemes, the resource utilization factor for each of them is examined. The convergence curve presented in Fig. 12 clearly elucidates that the average resource utilization factor for the proposed scheme rises quickly at the beginning which proves that the agent has been trained perfectly as a resource scheduler. Although the curve has levelled off but some fluctuations can be noticed due to the selection of maximum value of greedy factor as 0.9 in training phase, which means that the agent may select a suboptimal action with a probability of 0.1 in each episode. Furthermore, the random and greedy scheme showcase worst performance as the former selects the RAT without QoE awareness whereas the latter exhibits greedy nature with respect to the action-value estimates.

7.3 Empirical evaluation

This section investigates the efficiency of the proposed algorithm on the basis cumulative distribution function (CDF) calculated through extensive simulation. It is concluded from Fig. 13 that when the CDF is about 0.5, the proposed method outperforms the DQN based scheme by 0.05 units in terms of $SQoE$. Furthermore, the Fig. 13 also elucidates that the proposed scheme incurs $SQoE$ gain of 0.53 and 0.62 respectively in comparison to random and greedy scheme. In order to get better comprehension in the $SQoE$ gain, Fig. 14 clearly depicts that the proposed scheme has an obvious supremacy over other instinctive schemes with regard to $SQoE$ gains, 7.7% improvements to DQN based scheme, 121% and 171% to random and greedy scheme, respectively. In brief, the proposed scheme employs invalid action

Table 7 Comparison of the proposed scheme with other existing schemes

	Proposed scheme	DQN based scheme	Random scheme	Greedy scheme
Convergence rate	Fast	Medium	Constant	Constant
Accuracy	High	Medium	Low	Low
Better learning policy	Yes	No	No	No
Double deep learning	Yes	No	No	No
$SQoE$	Highest	High	Low	Lowest
RUF	Highest	High	Medium	Medium

masking scheme and double deep learning approach which produces more accurate value estimates, and better policies because of higher return and stable learning throughout the training process.

It can be concluded from the CDF of RUF represented in Fig. 15 that the proposed approach has a clear dominance over the DQN based scheme by 3.76%, when CDF is approximately 0.5. Furthermore, the Fig. 15 also depicts that the proposed scheme is more proficient than the random and greedy scheme as it leads them by 0.056 and 0.099 units respectively. The performance of proposed scheme is assessed by comparing it with other benchmark solutions in terms of average resource utilisation factor. Therefore, Fig. 16 clearly demonstrates that the proposed scheme has an obvious superiority over DQN based scheme, random and greedy scheme respectively by 4.44%, 6.31% and 10.58%. The proposed scheme attains more accurate value estimates as compared to DQN based scheme in terms of SQoE that ensures efficient personalised user experience through better network resource management whereas the random and greedy algorithm doesn't follow the constraints defined for the benefit and cost criterion leading to worst resource utilisation. The proposed scheme outperforms other conventional schemes in terms of convergence rate, accuracy and learning ability and the same has been summarised in Table 7.

7.4 Complexity analysis

This section provides a detailed evaluation of the time complexity of introduced algorithms namely proposed, DQN based, random and greedy approach. The complexity analysis for the proposed algorithm is discussed individually for the edge computing layer and SDWN layer. The edge computing layer employs two MLP classifier namely M_b and M_n in the big data analytics engine and network analysis component respectively. Hence, the time complexity analysed at the big data analytics engine is given as $tc_b = O(\zeta * TD_b * v_{bi} * v_{bh} * v_{bo})$, where ζ , TD_b , v_{bi} , v_{bh} , v_{bo} respectively represents the number of episodes, training data, input, hidden and output layer neurons related to the M_b classifier (Pedregosa F et al. 2011). Similarly the time complexity of network analysis component is given as $tc_n = O(\zeta * TD_n * \xi_{ni} * \xi_{nh} * \xi_{no})$, provided TD_n , ξ_{ni} , ξ_{nh} , ξ_{no} respectively is the training data, input, hidden and output layer neurons related to the M_n classifier. As the number of neurons related to input, hidden and output layer are approximately equivalent in case of M_b & M_n , the time complexity at edge computing layer can be expressed as: $tc_{edge} = tc_b + tc_n = O(\zeta * v_h * v_i * v_o * (TD_b + TD_n))$ (Serpén et al. 2014). The proposed scheme employs a distributed DDRL based association mechanism to select suitable network for each i patient. Each DDQN is a fully connected

deep neural network comprising three layers namely input, hidden and output layer. The input layer neurons corresponds to the number of elements in the state space i.e. $1 + j$ whereas the output layer neurons is the number of elements in the action space i.e. j valid RATs. Moreover, it has been considered that Y_{hn} fully connected hidden layers of each DDQN contains x_{hn}^y neurons. Therefore, the total number of weights that are required to be updated are $(1 + j) * x_{hn}^y + j * x_{hn}^{Y_{hn}} + \sum_{y=2}^{Y_{hn}} x_{hn}^{y-1} * x_{hn}^y$ (Zhang et al. 2020). Moreover, the time complexity of acquiring the corresponding experience (s_t, a_t, r_t, s_{t+1}) of each i patient in ζ episodes (comprising T time steps) is given as $O(\zeta * T * i)$. Assuming the complexity of updating a weight of the neuron is Z , the time complexity of DDQN network is given as $tc_{DDQN} = O(\zeta * T * Z * ((1 + j) * x_{hn}^y + j * x_{hn}^{Y_{hn}} + \sum_{y=2}^{Y_{hn}} x_{hn}^{y-1} * x_{hn}^y)) + O(\zeta * T * i)$ (Zhang et al. 2018). Therefore the total time complexity of the proposed scheme can be expressed as $tc_{DDQN} = O(\zeta * T * Z * ((1 + j) * x_{hn}^y + j * x_{hn}^{Y_{hn}} + \sum_{y=2}^{Y_{hn}} x_{hn}^{y-1} * x_{hn}^y)) + O(\zeta * T * (i + Z * ((1 + j) * x_{hn}^y + j * x_{hn}^{Y_{hn}} + \sum_{y=2}^{Y_{hn}} x_{hn}^{y-1} * x_{hn}^y)))$.

On the other hand, the total time complexity of DQN based approach accounts the state and action space of valid as well as invalid RATs (n) and the same is given as $tc_{DDQN} = O(\zeta * T * Z * ((1 + j) * x_{hn}^y + j * x_{hn}^{Y_{hn}} + \sum_{y=2}^{Y_{hn}} x_{hn}^{y-1} * x_{hn}^y)) + n * x_{hn}^{Y_{hn}} + \sum_{y=2}^{Y_{hn}} x_{hn}^{y-1} * x_{hn}^y$ (Mismar et al. 2018), where the former term corresponds to the time complexity related to the M_b MLP classifier that identifies the service and the latter term defines the time complexity of the DQN network. The time complexity of the proposed and DQN based approach mainly grows with the number of agents and dimension of the state and action space (Nasir et al. 2019). As the number of agents is same in both the approaches, the proposed scheme ensures reduced complexity as it considers the state and action space of only j valid RATs in comparison to DQN based approach which considers the state and action space of both valid and invalid RATs (n). The time complexity of random scheme is defined as $O(\zeta)$ (Mismar et al. 2018), as the action a is randomly sampled from a given set of actions whereas the time complexity for greedy scheme is given as $O(\zeta * T * (1 + n) * n)$ (Efroni et al. 2019).

7.5 Training efficiency with different learning Hyperparameter

The optimal selection of hyperparameter plays a substantial role in the successful functioning of proposed multi-agent deep neural networks but it depends upon the considered problem. Consequently, its selection requires extensive trial and error procedure to achieve convergence over reasonable time (Zhao et al. 2019; Bhattacharya et al. 2019). Therefore, Fig. 17 showcases the impact of hyperparameters on the performance of the proposed scheme. One of the

Fig. 17 Convergence performance under different **a** learning rate, **b** discount factor, **c** optimisation strategy, **d** mini-batch sizes

hyperparameter is learning rate which quantifies the amount by which neural network weights are updated during training. In regard to this, it can be concluded from Fig. 17a that the convergence performance with learning rate of 0.01 is better than the 0.1 and 0.001 as a larger learning rate can force the model to converge faster to insignificant solution, while a less learning rate can turn the process slower. Hence, learning rate need to be carefully tuned. Another hyperparameter which has a significant effect on the performance of trained model is the discount factor that signifies the importance of future rewards for the current state. Therefore, Fig. 17b showcase the performance of discount factor on the convergence statistics and it is inferred that low discount factor allows faster training dynamics whereas higher discount factor promotes inaccuracy and instabilities. On the other hand, Fig. 17c clearly elucidates faster convergence to the global minimum with the Adam optimisation strategy as it combines the heuristics of both RMSProp and Stochastic Gradient Descent with Momentum (SGDM). Adam optimization algorithm leverages square gradients to ascend the learning rate as employed in RMSprop and exploits the

momentum through the rolling average of the gradient as in SGDM. Lastly, the effect of mini-batch size on the convergence statistics of the proposed model is illustrated in Fig. 17d. It is clearly inferred that the low mini-batch size (i.e. 32) allows faster convergence to best solutions in contrast to higher batch sizes as it permits the faster updation of neural network weights. Moreover, it also offers a low generalisation error and regularizing effect due to presence of steeper gradient descent direction.

8 Conclusion

The advent of telehealth technology accentuated on the QoE provisioned network connectivity to realise an adaptive and user-centric medical infrastructure. Within this paradigm, numerous RAT selection solutions has been proposed in the existing literature but they failed to address the aforementioned issue. Therefore, a data-driven SDWN-Edge enabled architecture has been proposed to authenticate the personalised user experience and conduct efficient network

resource utilisation. The proposed intelligent access network selection model leverages an invalid action masking scheme and multi-agent reinforcement learning that ensures faster convergence to fine-grained QoE optimised RAT selection policy. The *SQoE* gain improvement has been achieved of the order of 7.7%, 121% and 171% respectively in comparison to DQN based, random and greedy schemes that has been corroborated through substantial simulations. It has been found that resource utilisation factor has been enhanced by 4.44%, 6.31% and 10.58% in comparison to existing DQN based, random and greedy schemes respectively. The results obtained have indicated faster convergence to optimal RAT selection policy and better generalisation ability. The proposed RAT selection scheme only respects the user interests and preferences. Therefore, augmentation of the proposed approach to a balanced RAT selection solution that considers the preferences of both RAT and user is envisioned. Lastly, the proposed architecture will be beneficial for the researchers and engineers working in the area of smart healthcare network design.

Acknowledgements Author would like to thank University Grant Commission, New Delhi for Junior Research Fellowship.

Availability of data and material This manuscript has no associated data file.

Declarations

Conflict of interest The authors declare no conflict of interest, financial or otherwise.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Ahmad A, Floris A, Atzori L (2016) QoE-centric service delivery: a collaborative approach among OTTs and ISPs. *Comput Netw* 110:168–179. <https://doi.org/10.1016/j.comnet.2016.09.022>
- Al-Janabi S (2018) Smart system to create an optimal higher education environment using IDA and IOTs. *Int J Comput Appl* 42:244–259. <https://doi.org/10.1080/1206212X.2018.1512460>
- Al-Janabi S, Alkaim AF (2020) A nifty collaborative analysis to predicting a novel tool (DRFLLS) for missing values estimation. *Soft Comput* 24:555–569. <https://doi.org/10.1007/s00500-019-03972-x>
- Al-Janabi S, Hussein NY (2020) The reality and future of the secure mobile cloud computing (SMCC): survey. In: Farhaoui Y (ed) *Big data and networks technologies*. BDNT 2019. Lecture notes in networks and systems. Springer, Cham, pp 231–261
- Al-Janabi S, Al-Shourbaji I, Shojafar M, Abdelhag M (2017) Mobile cloud computing: challenges and future research directions. In: *Proceedings of 10th International Conference on Developments in eSystems Engineering (DeSE)*, IEEE, Paris, pp 62–67. <https://doi.org/10.1109/DeSE.2017.21>
- Arabi S, Hammouti HE, Sabir E, Elbiaze H, Sadik M (2019) RAT association for autonomic IoT systems. *IEEE Network* 33(6):1–8. <https://doi.org/10.1109/mnet.2019.1800513>
- Barmounakis S, Kaloxylas A, Spapis P, Alonistioti (2017) Context-aware, user-driven, network-controlled RAT selection for 5G networks. *Comput Netw* 113:124–147
- Bhatia M, Kumar K (2019) Network selection in cognitive radio enabled wireless body area networks. *Digit Commun Netw* 6:75–85. <https://doi.org/10.1016/j.dcan.2018.03.003>
- Bhattacharya R et al (2019) QFlow: a reinforcement learning approach to high QoE video streaming over wireless networks. In: *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, Catania, pp 251–260
- Chen X, Li Z, Zhang Y, Long R, Yu H, Du X, Guizani M (2018) Reinforcement learning-based QoS/QoE-aware service function chaining in software-driven 5G slices. *Trans Emerg Telecommun Technol* e3477:1–18. <https://doi.org/10.1002/ett.3477>
- Chen M, Li W, Hao Y, Qian Y, Humar I (2018) Edge cognitive computing based smart healthcare system. *Futur Gener Comput Syst* 86:403–411. <https://doi.org/10.1016/j.future.2018.03.054>
- Cisotto G, Casarin E, Tomasin S (2020) Requirements and enablers of advanced healthcare services over future cellular systems. *IEEE Commun Mag* 58(3):76–81. <https://doi.org/10.1109/MCOM.001.1900349>
- Desogus C, Anedda M, Murrioni M, Muntean GM (2019) A traffic type-based differentiated reputation algorithm for radio resource allocation during multi-service content delivery in 5G heterogeneous scenarios. *IEEE Access* 7:27720–27735
- Ding H, Zhao F, Tian J, Li D, Zhang H (2019) A deep reinforcement learning for user association and power control in heterogeneous networks. *Ad Hoc Netw* 102:1–18
- Du Z, Jiang B, Wu Q, Xu Y, Xu K (2020) Exploiting user demand diversity: QoE game and MARL based network selection. In: Du Z (ed) *Towards user-centric intelligent network selection in 5G heterogeneous wireless networks*. Springer, Singapore, pp 101–130
- Efroni Y, Merlis N, Ghavamzadeh M, Mannor S (2019) Tight regret bounds for model-based reinforcement learning with greedy policies. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, pp 12224–12234
- El Helou M, Ibrahim M, Lahoud S, Khawam K, Mezher D, Cousin B (2015) A network-assisted approach for RAT selection in heterogeneous cellular networks. *IEEE J Sel Areas Commun* 33(6):1055–1067
- François-Lavet V, Henderson P, Islam R, Bellemare MG, Pineau J (2018) An introduction to deep reinforcement learning. *Found Trends Mach Learn* 11(3–4):219–354. <https://doi.org/10.1561/22000000071>
- Goyal P, Lobiyal DK, Katti CP (2018) Game theory for vertical hand-off decisions in heterogeneous wireless networks: a tutorial. In: Bhattacharyya S, Gandhi T, Sharma K, Dutta P (eds) *Advanced computational and communication paradigms*. Lecture notes in electrical engineering. Springer, Singapore, pp 422–430
- Hadi MS, Lawey AQ, El-Gorashi TEH, Elmighani JMH (2020) Patient-centric HetNets powered by machine learning and big data analytics for 6G networks. *IEEE Access* 1:1–17. <https://doi.org/10.1109/access.2020.2992555>
- Hao Y, Jiang Y, Hossain MS, Ghoneim A, Yang J, Humar I (2018) Data-driven resource management in a 5G wearable network using network slicing technology. *IEEE Sens J* 19(19):8379–8386. <https://doi.org/10.1109/jsen.2018.2883976>
- Hasselt HV, Guez A, Silver D (2016) Deep reinforcement learning with double Q-learning. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, pp 2094–2100
- Hemmati M, McCormick B, Shirmohammadi S (2017) QoE-aware bandwidth allocation for video traffic using sigmoidal

- programming. *IEEE Multimedia* 24(4):80–90. <https://doi.org/10.1109/MMUL.2017.4031305>
- Imran MA, Abdulrahman Sambo Y, Abbasi QH, Soldani D, Innocenti M (2020) 5G Communication systems and connected healthcare. In: Imran MA, Abdulrahman Sambo Y and Abbasi QH (eds) Enabling 5G Communication systems to support vertical industries. <https://doi.org/10.1002/9781119515579.ch7>
- Johnson A, Pollard T, Mark R (2019) MIMIC-III clinical database demo (version 1.4). *PhysioNet*. <https://doi.org/10.13026/C2HM2Q>
- Kim KS et al (2019) Ultrareliable and low-latency communication techniques for tactile internet services. *Proc IEEE* 107(2):376–393. <https://doi.org/10.1109/JPROC.2018.2868995>
- Krishankumar R, Arun K, Kumar A et al (2021) Double-hierarchy hesitant fuzzy linguistic information-based framework for green supplier selection with partial weight information. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-021-06123-2>
- Kumar B, Sharma L, Wu SL (2019) Online distributed user association for heterogeneous radio access network. *Sensors* 19(6):1–23. <https://doi.org/10.3390/s19061412>
- Lloret J, Parra L, Taha M, Tomás J (2017) An architecture and protocol for smart continuous eHealth monitoring using 5G. *Comput Netw* 129:340–351. <https://doi.org/10.1016/j.comnet.2017.05.018>
- Malasinghe LP, Ramzan N, Dahal K (2017) Remote patient monitoring: a comprehensive study. *J Ambient Intell Humaniz Comput* 10(1):57–76. <https://doi.org/10.1007/s12652-017-0598-x>
- Manjeshwar AN, Roy A, Jha P & Karandikar (2019) A control and management of multiple RATs in wireless networks: an SDN approach. In: *Proceedings of the 2nd 5G World Forum (5GWF)*. IEEE, Dresden, pp 596–601. <https://doi.org/10.1109/5GWF.2019.8911703>
- Mismar FB, Evans BL (2018) Deep Q-learning for self-organizing networks fault management and radio performance improvement. In: *Proceedings of 52nd Asilomar conference on signals, systems, and computers - Pacific Grove, IEEE, CA*, pp 1457–1461
- Mollet MS, Abubakar AI, Ozturk M, Kaijage S, Kisangiri M, Zoha A, Abbasi QH (2020) Intelligent handover decision scheme using double deep reinforcement learning. *Phys Commun* 42(2020):1–12. <https://doi.org/10.1016/j.phycom.2020.101133>
- Mukherjee A, Ghosh S, Behere A, Ghosh SK, Buyya R (2021) Internet of Health Things (IoHT) for personalized health care using integrated edge-fog-cloud network. *J Ambient Intell Humaniz Comput* 12:943–959. <https://doi.org/10.1007/s12652-020-02113-9>
- Nasir YS, Guo D (2019) Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. *IEEE J Sel Areas Commun* 37(10):2239–2250. <https://doi.org/10.1109/JSAAC.2019.2933973>
- Nguyen DD, Nguyen HX, White LB (2017) Reinforcement learning with network-assisted feedback for heterogeneous RAT selection. *IEEE Trans Wireless Commun* 16(9):6062–6076
- Ning Z, Dong P, Wang X, Hu X, Guo L, Hu B, Guo Y, Qiu T, Kwok RYK (2020) Mobile edge computing enabled 5G health monitoring for internet of medical things: a decentralized game theoretic approach. *IEEE J Select Area Commun* 39(2):463–478
- Patel A, Al-Janabi S, AlShourbaji I, Pedersen J (2015) A novel methodology towards a trusted environment in mashup web applications. *Comput Secur* 49:107–122
- Pedregosa F et al. (2011) Neural network models (supervised). *scikit-learn*. https://scikit-learn.org/stable/modules/neural_networks_supervised.html. Accessed 15 Aug 2021
- Priya B, Malhotra J (2020) 5GAuNetS: an autonomous 5G network selection framework for Industry 4.0. *Soft Comput* 24:9507–9523. <https://doi.org/10.1007/s00500-019-04460-y>
- Priya B, Malhotra J (2020) QAAs: QoS provisioned artificial intelligence framework for AP selection in next-generation wireless networks. *Telecommun Syst*. <https://doi.org/10.1007/s11235-020-00710-9>
- Rahmani Amir M, Gia Tuan N, Negash B, Anzanpour A, Azimi I, Jiang M, Liljeberg P (2017) Exploiting smart e-health gateways at the edge of healthcare Internet-of-Things: a fog computing approach. *Futur Gener Comput Syst* 78:641–658. <https://doi.org/10.1016/j.future.2017.02.014>
- Rajesh L, Boopathybagan K, Ramesh B (2017) User demand wireless network selection using game theory. In: Nath V (ed) *Proceedings of the international conference on nano-electronics, circuits and communication systems*. Lecture notes in electrical engineering. Springer, Singapore, pp 39–53
- Rustam F et al (2020) Sensor-based human activity recognition using deep stacked multilayered perceptron model. *IEEE Access* 8:218898–218910. <https://doi.org/10.1109/ACCESS.2020.3041822>
- Salih YK, See OH, Ibrahim RW (2016) An intelligent selection method based on game theory in heterogeneous wireless networks. *Trans Emerg Telecommun Technol* 27(12):1641–1652. <https://doi.org/10.1002/ett.3102>
- Sandoval RM, Canovas-Carrasco S, Garcia-Sanchez A, Garcia-Haro J (2019) A reinforcement learning-based framework for the exploitation of multiple RATs in the IoT. *IEEE Access* 7:123341–123354. <https://doi.org/10.1109/ACCESS.2019.2938084>
- Saraiva J, Braga IM, Monteiro VF, Lima FRM, Maciel T, Freitas W, Cavalcanti FRP (2020) Deep reinforcement learning for QoS-constrained resource allocation in multiservice networks. *J Commun Inf Syst* 35(1):66–76
- Serpen G, Gao Z (2014) Complexity analysis of multilayer perceptron neural network embedded into a wireless sensor network. *Procedia Comput Sci* 36:192–197. <https://doi.org/10.1016/j.procs.2014.09.078>
- Shantharama P, Thyagaturu A, Karakoc N, Ferrari L, Reisslein M, Scaglione A (2018) LayBack: SDN management of multi-access edge computing (MEC) for network access services and radio resource sharing. *IEEE Access* 6:57545–57561. <https://doi.org/10.1109/ACCESS.2018.2873984>
- Simsek M, Aijaz A, Dohler M, Sachs J, Fettweis G (2016) 5G-Enabled tactile internet. *IEEE J Sel Areas Commun* 34(3):460–473. <https://doi.org/10.1109/jsac.2016.2525398>
- Skondras E, Michalas A, Vergados DD (2019) Mobility management on 5G vehicular cloud computing systems. *Veh Commun* 16(2019):15–44. <https://doi.org/10.1016/j.vehcom.2019.01.001>
- Sun P, Guo Z, Wang G, Lan J, Hu Y (2020) MARVEL: enabling controller load balancing in software-defined networks with multi-agent reinforcement learning. *Comput Netw* 177:1–10. <https://doi.org/10.1016/j.comnet.2020.107230>
- Tartarini L, Marotta MA, Cerqueira E, Rochol J, Both CB, Gerla M, Bellavista P (2017) Software-defined handover decision engine for heterogeneous cloud radio access networks. *Comput Commun* 115:21–34. <https://doi.org/10.1016/j.comcom.2017.10.018>
- Thuemmler C, Paulin A, Lim AK (2016) Determinants of next generation e-health network and architecture specifications. In: *Proceedings of IEEE 18th Int. Conf. on e-Health Networking, Applications and Services (Healthcom)*. IEEE, Munich, pp 1–6
- Ugalmugale S, Swain R (2020) *Telemedicine Market Size By Service (Tele-consulting, Tele-monitoring, Tele-education/training), By Type (Telehospital, Telehome), By Specialty (Cardiology, Gynecology, Neurology, Orthopedics, Dermatology, Mental Health), By Delivery Mode (Web/Mobile Telephonic, Visualized, Call Centers), Industry Analysis Report, Regional Outlook, Growth Potential, Price Trends, Competitive Market Share & Forecast, 2020–2026*. Global Market Insights. <https://www.gminsights.com/industry-analysis/telemedicine-market>. Accessed 26 Sep 2020
- Van D, Ai Q, Liu Q (2017) Vertical handover algorithm for WBANs in ubiquitous healthcare with quality of service guarantees. *Information* 8(1):1–16. <https://doi.org/10.3390/info8010034>

- Varga N, Piri E, Bokor L (2015) Network-assisted smart access point selection for pervasive real-time mHealth applications. *Procedia Comput Sci* 63:317–324. <https://doi.org/10.1016/j.procs.2015.08.349>
- Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Al-Nemrat A, Venkatraman S (2019) Deep learning approach for intelligent intrusion detection system. *IEEE Access* 7:41525–41550. <https://doi.org/10.1109/ACCESS.2019.2895334>
- Wang Y et al (2017) A data-driven architecture for personalized QoE management in 5G wireless networks. *IEEE Wirel Commun* 24(1):102–110. <https://doi.org/10.1109/MWC.2016.1500184WC>
- Wang X, Li J, Wang L, Yang C, Han Z (2019) Intelligent user-centric network selection: a model-driven reinforcement learning framework. *IEEE Access* 7:21645–21661. <https://doi.org/10.1109/ACCESS.2019.2898205>
- Wang X, Su X & Liu B (2019) A novel network selection approach in 5G heterogeneous networks using Q-learning. In: *Proceedings of the 26th International Conference on Telecommunications (ICT)*. IEEE, Hanoi, pp 309–313. <https://doi.org/10.1109/ICT.2019.8798797>
- Xu F, Ye H, Yang F, Zhao C (2019) Software defined mission-critical wireless sensor network: architecture and edge offloading strategy. *IEEE Access* 7:10383–10391. <https://doi.org/10.1109/access.2019.2890854>
- Yadav P, Agrawal R, Kashish K (2018) Heterogeneous network access for seamless data transmission in remote healthcare. *Int J Grid Distrib Comput* 11(8):69–86
- Yamamoto H et al (2020) Forecasting crypto-asset price using influencer tweets. In: Barolli L, Takizawa M, Xhafa F, Enokido T (eds) *Advanced information networking and applications*. AINA 2019. *Advances in intelligent systems and computing*. Springer, Cham, pp 940–951
- Zhang Q, Lin M, Yang LT, Chen Z, Khan SU, Li P (2018) A double deep Q-learning model for energy-efficient edge scheduling. *IEEE Trans Serv Comput* 12(5):739–749
- Zhang Q, Liang Y-C, Poor HV (2020) Intelligent user association for symbiotic radio networks using deep reinforcement learning. *IEEE Trans Wireless Commun* 19(7):4535–4548. <https://doi.org/10.1109/TWC.2020.2984758>
- Zhang Q, Liu J & Zhao G (2018) Towards 5G enabled tactile robotic telesurgery. *arXiv:1803.03586* [cs.NI]. <https://arxiv.org/pdf/1803.03586.pdf>. Accessed 15 July 2020
- Zhang X, Sen S, Kurniawan D, Gunawi H, Jiang J (2019) E2E: embracing user heterogeneity to improve quality of experience on the web. In: *Proceedings of the ACM Special Interest Group on Data Communication - SIGCOMM '19*, ACM, Beijing, pp 289–302. <https://doi.org/10.1145/3341302.3342089>
- Zhang K, Yang Z & Basar T (2019) Multi-agent reinforcement learning: a selective overview of theories and algorithms. *arXiv:1911.10635*. <https://arxiv.org/pdf/1911.10635.pdf>. Accessed 27 July 2020
- Zhao N, Liang YC, Niyato D, Pei Y, Wu M, Jiang Y (2019) Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks. *IEEE Trans Wireless Commun* 18(11):5141–5152
- Zhong Y, Wang H, Lv H (2020) A cognitive wireless networks access selection algorithm based on MADM. *Ad Hoc Netw* 109(2020):1–9. <https://doi.org/10.1016/j.adhoc.2020.102286>
- Zhu A, Guo S, Liu B, Ma M, Feng H, Su X (2019) Adaptive multi-service heterogeneous network selection scheme in mobile edge computing. *IEEE Internet Things J* 6(4):6862–6875. <https://doi.org/10.1109/jiot.2019.2912155>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.