sustainability

MDPI

Article

Telepresence Robot with DRL Assisted Delay Compensation in IoT-Enabled Sustainable Healthcare Environment

Fawad Naseer [1,*], Muhammad Nasir Khan [1] and Ali Altalbe [2]

[1] Electrical Engineering Department, The University of Lahore, Lahore 54590, Pakistan
[2] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
* Correspondence: fawadn.84@gmail.com

Abstract: Telepresence robots have become popular during the COVID-19 era due to the quarantine measures and the requirement to interact less with other humans. Telepresence robots are helpful in different scenarios, such as healthcare, academia, or the exploration of certain unreachable territories. IoT provides a sensor-based environment wherein robots acquire more precise information about their surroundings. Remote telepresence robots are enabled with more efficient data from IoT sensors, which helps them to compute the data effectively. While navigating in a distant IoT-enabled healthcare environment, there is a possibility of delayed control signals from a teleoperator. We propose a human cooperative telecontrol robotics system in an IoT-sensed healthcare environment. The deep reinforcement learning (DRL)-based deep deterministic policy gradient (DDPG) offered improved control of the telepresence robot to provide assistance to the teleoperator during the delayed communication control signals. The proposed approach can stabilize the system in aid of the teleoperator by taking the delayed signal term out of the main controlling framework, along with the sensed IOT infrastructure. In a dynamic IoT-enabled healthcare context, our suggested approach to operating the telepresence robot can effectively manage the 30 s delayed signal. Simulations and physical experiments in a real-time healthcare environment with human teleoperators demonstrate the implementation of the proposed method.

Keywords: telepresence robot; IoT; healthcare environment; remote management

Citation: Naseer, F.; Khan, M.N.; Altalbe, A. Telepresence Robot with DRL Assisted Delay Compensation in IoT-Enabled Sustainable Healthcare Environment. *Sustainability* **2023**, *15*, 3585. https://doi.org/10.3390/su15043585

Academic Editor: Youseef Alotaibi

Received: 14 January 2023
Revised: 10 February 2023
Accepted: 10 February 2023
Published: 15 February 2023

1. Introduction

Telepresence robots are becoming an increasingly popular technology as they allow for remote communication and collaboration in various settings. They are often used in the healthcare, education, and business industries and are designed to promote sustainability by reducing the need for travel. By utilizing the Internet of Things (IoT), telepresence robots can be controlled remotely and integrated with other devices to enhance their capabilities. In particular, the use of telepresence robots helps to reduce carbon emissions and other environmental impacts related to travel, making it a more sustainable solution.

Recent improvements in the computer system with significant advancements in robotics have assisted individuals by providing agility and increased global awareness of the respective field. Autonomous unmanned robots were previously considered luxurious and had a small marketplace. General industrial robots were utilized extensively in the production field to improve manufacturing productivity. Recently, the importance of robots in our society has come to light. The use of telepresence technology in the community can enhance and improve social and cultural interchange. In particular, telepresence robots can connect individuals and create the impression that they are all collaborating in the same room, regardless of where they are located physically [1].

In developing nations such as Pakistan, the distribution of doctors is frequently an issue. This issue has impacted rural residents' access to quality healthcare. In 2021, the government disclosed that 21.2% of deaths in rural Pakistan were reported due to a lack of

medical attention by the skilled healthcare workforce. The "Doctor to the Rural" program is only one of the government initiatives devised to address the issue. Along with this program, the creation of a telepresence system is another workable alternative optimized intelligent framework for cyber-physical systems [2]. This research intends to create a cooperative control and navigation system for remote telepresence mobile robots.

The most important aspect of controlling a telepresence robot in an IoT-sensed environment is the connection cost for each IoT sensor while interacting with other resources. For telepresence robots to be able to carry out their respective jobs, a system design with sensory data gathering is necessary. The characteristics of sensors in IoT-enabled healthcare contexts include heterogeneity, many types of sensory data, and several average values. Data acquisition from IoT sensors and the telepresence robot's control platform is separated logically from that of other IoT sensors. Repeated sensors' connections with the components of software platforms can delay the process routines and the telepresence robot's activities.

The initial navigation system depends on the cooperation of human operational commands and the telepresence robot's independent movements [3]. The system's human operator chooses the system's ultimate goal. The telepresence robot determines the shortest route to its destination and proceeds independently. In the case of the human operator's control, the operating mode is changed to "manual mode" to direct the telepresence robot along paths that differ from those that the robot generates. In manual functionality mode, the telepresence robot is solely controlled, whereas autonomous robotic movements are not activated. The interaction of telepresence robots and human teleoperators in this framework is not fully realized.

In the case of the known time delay, the past state can be used for the backward prediction of the present state. Bar-Shalom [4] developed this idea with an imperfect algorithm. In [5], the authors describe the Energy-Aware Cluster-Based Routing Scheme for IoT-Assisted Networks, as well as challenges in the technical setup.

A non-linear system requires adjustments to estimate its state. A technique based on extrapolating a delayed measurement to the present utilizing a Kalman filter for past and present estimates was introduced by Larsen et al. [6]. In [7], an extension approach is described: interpolating a delayed measurement minimizes the computing time, even for considerable time delays.

In this study, we propose a novel DDPG-based framework to estimate the time delay by using DRL approaches while aiding the non-linear telepresence robot's navigation. However, a crucial communication time delay occurs when a filtering processor is linked to a sensor through a network. Additionally, extra post-processing time is needed when raw data from the sensor are processed after being acquired to update the dynamical system's states. This causes a delay between the measurement's acquisition and its availability to the filter. In such cases, general control approach methods cannot achieve telepresence robot state control. As a result, we suggest a novel method that uses DDPG to enhance prior states that consider the delays. However, the proposed method is to implement the deep deterministic policy-based control algorithm. This study first develops a telepresence robot state space along with the action space, which complies with the Markov model to create a separate reward function under the classification of different parameters. Then, the replay buffer is optimized by incorporating weighted training samples to control signal fluctuation in the low-speed sections.

This article also explains the telepresence robot tests using the anticipated midpoint curve. The outcomes demonstrate that the suggested algorithm performs better than traditional control techniques such as PID in effectively tracking the expected speed incorporated with human teleoperators. The telepresence robot only requires five or six episodes of fully automatic training before the network can be updated without expert tuning. In the proposed approach, human operators can choose the robot's control and navigation courses while viewing the robot's surroundings online. Additionally, the method offers the simultaneous integration of steady robot motion based on autonomous navigation using maps.

The rest of this paper is organized as follows. The literature review is described in Section 2. The telepresence robot proposed in this study is described in Section 3. The experimental setup and results are discussed in Section 4. The conclusions are explained in Section 5.

2. Literature Review

In this section, we analyze the previous research on the controllability and management of telepresence robots in IoT-sensed environments.

A system architecture based on the context-aware robot as a service (RAAS) [8] and sensory data and signals must be preoccupied and transformed into information that the robot can understand in order to manage the robot's action based on these data [9]. The processing of sensory data to address issues such as various IoT devices and value formats has been the subject of numerous studies. The methods used to transform the data into information include acquiring values, sensor access, and converting situational information.

Practical bandwidth usage and appropriate compression algorithms are crucial for real-time visual communication. The user experience with this gadget is more ergonomic and natural. The teleoperator's head movement enables biological stereoscopic control of the camera mounted on the robot, creating a three-dimensional vision [10]. Haptic feedback is also required to provide the teleoperator with a true-to-life experience. This can be accomplished by fusing the robotic perception of the external environment and transmitting it to the human user through external sensory input. Additionally, gesture control will enhance the social telepresence operation's remote motion control [11]. A telepresence robot with haptic feedback-enabled features can assist visually impaired patients [12].

In a virtual reality environment, Vlahovic et al. [13] compared four locomotion methods: controller movement, controller movement with tunnelling, teleportation, and a human joystick. For controller movement and the human joystick, which also caused the testers more physical discomfort, the overall quality of the experience was scored lower. Compared to perceived immersion, the authors discovered that comfort might significantly affect the quality of experience (QoE) for navigation in virtual reality environments.

Numerous telepresence robots have been developed recently for a variety of uses, including telemedicine [13], tele-education [14], and senior home care [15]. There are also several commercial telepresence robots available on the market with various applications, including impromptu talks at work, patient care in healthcare facilities, and remote education in schools. These robots include the Texal and PR2 from Willow Garage [16], the QB from Anybots [17], the VGo from VGo Communications [18], the RP-7i from InTouch Health [19], the MantaroBot from Vasteras Giraff [20,21], the Ava and RP-VITA from iRobot, and others. However, the widespread use of these telepresence robots remains to be achieved, and it faces numerous problems that require further research. It is critical to create telepresence robots compatible with new mobile devices, such as smartphones and tablets, so as to be used with them. Some telepresence robots, including the WU robots [22], can enable interactions with tablets. An IoT platform that uses a Modified Self-Adaptive Bayesian Algorithm (MSABA) to provide more precise assessments of HD [23] describes the IOT prediction system.

Research has been conducted to determine LQI, a channel prediction model, and RSSI in an outdoor setting concerning signaling [24]. The new intelligent mutation operator improves the security, privacy, integrity, and authenticity of the information system by identifying harmful requests and responses and helping to defend the system against assault [25]. Single-chip nodes for autonomous node programming over a USB model design were covered in another article [26]. Medical sensor networks (MSN), which took into account trust management in TelosB nodes, were used to collect tree protocols that were used and created for e-healthcare systems [27]. In [28], test case experiments were used to implement an experimental study on the ZigBee frequency agile (FA) scheme on the TelosB testbed. An accuracy-aware diffusion process mapping scheme using TelosB nodes, mobile Telos nodes, was developed in [29].

This study develops a trust-aware multi objective metaheuristic optimization-based secure clustering with route planning (TAMOMO-SCRP) technique for a cluster-based IIoT environment [30]. This is a feature-based approach, in which camera movement is detected using the inconsistency of image features in sequential image streams, as in PTAM [31] and ORB-SLAM [32]. The most recent approach, ORB-SLAM, has a reported 1% inaccuracy in the dimensions of the maps. DTAM [33] and CNN-RNN-LSTM [34] propose a direct method that uses all images as a single entity as a second option. Applications that leverage smart devices benefit significantly from the SLAM techniques. ORB-SLAM and LSD-SLAM need CPU, and DTAM needs GPU to become real-time. If the map is not too large, PTAM might perform in real time on smartphones [35].

The data amount has dramatically expanded with the development of IoT technologies, and data processing and transmission have become increasingly complex. Less delay, more throughput, and high confidentiality can be achieved by computing at the edges and storing data locally [36]. Wi-Fi routers are viewed as edge nodes in Para-Drop [37] that speak directly to users. Despite extensive research, very few edge computing applications regarding big social media data have been documented [38].

The objective of [39] was to create a controlled robot with only two wheels. The authors presented a detailed discussion on the use of Lego Mindstorm NXT [40] to design and test a robotic chassis that an AVR ATMega16 microprocessor would run. Their test demonstrated the need for a robot chassis to address mechanical and stability difficulties. Segway, a well-known robot that can balance itself while a person is standing on its platform, was created by [41]. To remain upright, it uses brushless DC electric motors with encoders and gyroscopic sensors in the wheels that are powered by lithium-ion batteries.

In [42], a new database approach is introduced using a hybrid meta-heuristic of the intrusion in robot communication with a teleoperator. However, a planner is still used to guide navigation. As in [43,44] selects an action toward either a defined or predicted objective while making a forecast about the map based on the RGB image. A planner is still employed to design a path towards the destination on the uncertain map. However, instantaneous actions are received from the neural network to follow the path. In [45], a robot is trained to travel in a novel area while free space in the environment is anticipated in a static virtual grid world. Meanwhile, [46] presents exploratory navigation using deep Q-learning, where a robot learns to avoid obstacles in unknown environments.

To solve the partial observation Markov decision process (POMDP), Ref. [47] treated the exploration as a direct policy search. Meanwhile, Ref. [48] demonstrated the development of the CSRO algorithm for route optimization, which was developed in-house. Due to DL's excellent perception capabilities, it has been widely used to create effective techniques in learning the mapping from sensor data to robot control. A perception network with RGB-D image inputs and a control network trained by the DQN was proposed by Tai and Liu in 2016 [49]. It should be noted that this technique guarantees that the robot will wander without colliding. A supervised learning problem was created by Bai et al. [50] to lower the map's Shannon entropy.

In [51], author describes the Bacteria for Aging Optimization Algorithm (BFOA), which finds the ideal hops in advancing the routing, and is utilized to offer trust-based, protected, and energy-efficient navigation in MANETs using a trust-based, protected, and energy-efficient navigation algorithm. Marroquin et al. offered a low-cost alternative for a mobile explorer robot with a camera and temperature/air humidity sensors. These sensors use open hardware and software, whose design is intended to allow inspection of the environment, in addition to being controlled remotely using the technology of the Internet of Things through a graphical user interface [52].

The IoT-based robot system proposed by Srividhya et al. is utilized for vigilante activities. This robot can walk in any direction and uses a Raspberry Pi to send live video to an Android device. The Raspberry Pi receives the signals produced by the Android application. The commands process the signals that have been received, and the robot is guided by the employed direction [53]. An IoT-based context-aware system presented by

Park, Choi, and Choi J. collects sensor data from the outside world and converts it to sensory data, and [54] proposes the Energy-Aware Cluster-based Routing (EACR-LEACH) protocol in WSN-based IoT. The Cluster Head (CH) selection is crucial in the clustering protocol. Wang et al. introduced an intelligent housekeeper, an Internet of Things (IoT)-based indoor mobility robot utilized for housekeeping services [55].

"Prabuwono et al. analysis depends on creating a visual methodology to control the semi-autonomous convoy [56]. Meanwhile, an intelligent adaptive method for IoT-enabled environments, and [57] presented a detailed review of multipath transport protocols to enhance communication.

3. The Proposed Model

Deep reinforcement learning (DRL) is chosen as the control method in this study because it offers a strong and adaptable framework for decision-making in complex, dynamic environments. In DRL, the agent discovers through trial and error how to act in a way that maximizes a reward signal. This makes it ideal for telepresence robots operating in a healthcare environment supported by the Internet of Things. The robot must navigate challenging environments and adjust to changing conditions while waiting for teleoperator communications to arrive. DDPG is an actor–critic method, meaning that it uses two neural networks: an actor network that outputs actions and a critic network that evaluates the actions. DDPG can handle continuous action spaces, essential in a telepresence robot control task, where the actions may involve continuous values such as velocity or direction.

The proposed algorithm seeks to increase the cumulative future reward R_t, which is defined as in Equation (1):

$$R_t = r_t + Y.r_{t+1} + Y^2.r_{t+2} + \ldots = \sum_{k=0}^{\infty} Y^k r_{t+k} \quad (1)$$

with gamma within the range of [0–1]. Under the state s_t and action a_t, the estimation of R_t is defined as the value function in Equation (2):

$$Q^{\pi}(s_t, a_t) = \mathbb{E}_{\pi}[R_t|s_t, a_t] = \mathbb{E}_{\pi}[\sum_{k=0}^{\infty} Y^k r_{t+k}|s_t, a_t] \quad (2)$$

To find the best action value $Q^*(s_t, a_t)$, it is usually the majority of all policies π. Afterwards, the optimal policy selects the action as in Equation (3) to train the optimal action value comprehensively.

$$a_t^* = \pi^*(s_t) = \text{argmax}_a[Q^*(s_t, a)] \quad (3)$$

The proposed algorithm structure to control the telepresence robot with DRL-assisted delay compensation in an IoT-enabled sustainable healthcare environment is depicted in Figure 1. Network *X* produces the action αX after receiving the state s. Following the Gaussian distribution and action constraint, the telepresence robot performs the action a to control the telepresence robot. The action value represented by network Q is used to assess the effectiveness of the action choice.

In network Q, there are also two hidden layers with 60 nodes, respectively. The output layer is linearly activated, while the hidden layers' activation functions are both ReLU. Table 1 displays more specific parameters regarding the scenario.

Figure 2 depicts the formation of networks X and Q. In network X, there are two hidden layers with 40 nodes in each of them, respectively. ReLU is used as the activation function for hidden layers, and Tanh is used for the output layer [58].

Repetitive parameter updating and training are required for network X to identify the best course of action accurately. Since network Q will be established first, network X will be updated in a manner closely associated with network Q.

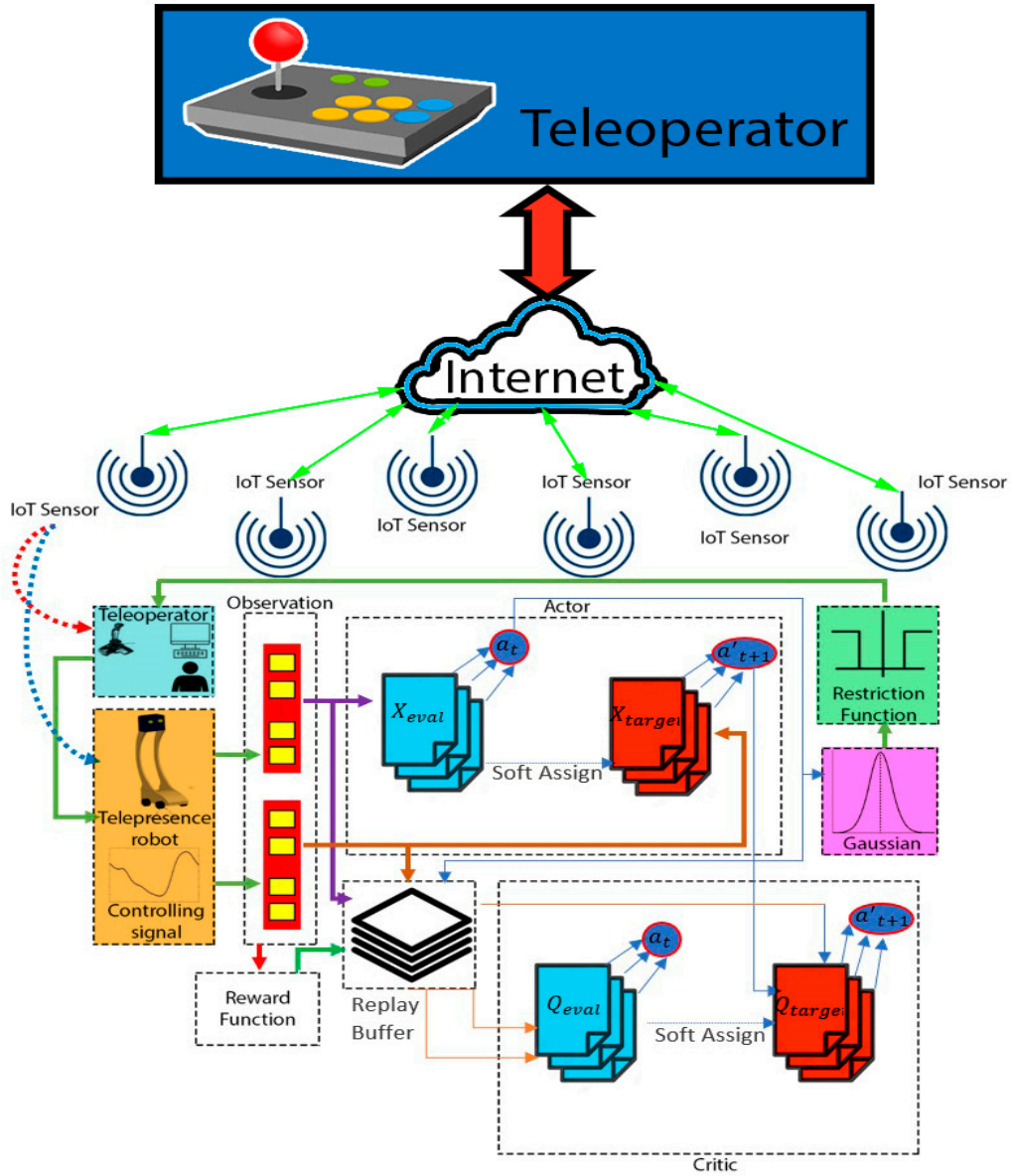

Figure 1. Proposed algorithm framework structure.

Table 1. List of parameters.

Symbol	Description	Value
ε	Soft assign rate	0.007
Υ	Discounting factor of reward	0.85
ξ	Decay rate	0.9996
σ^2	Initial variance of the exploration space	40

Unlike the Deep Q Network (DQN) and SPID, our network Q is unique. Instead of only state s, the state s and the action a are significant components of the input for network Q. Network Q produces a single-dimensional value Q rather than a multidimensional vector Q(a). Network X should be set up because it is challenging to acquire the action value entirely through network Q and the algorithm's freedom from the action space capacity is one of its benefits.

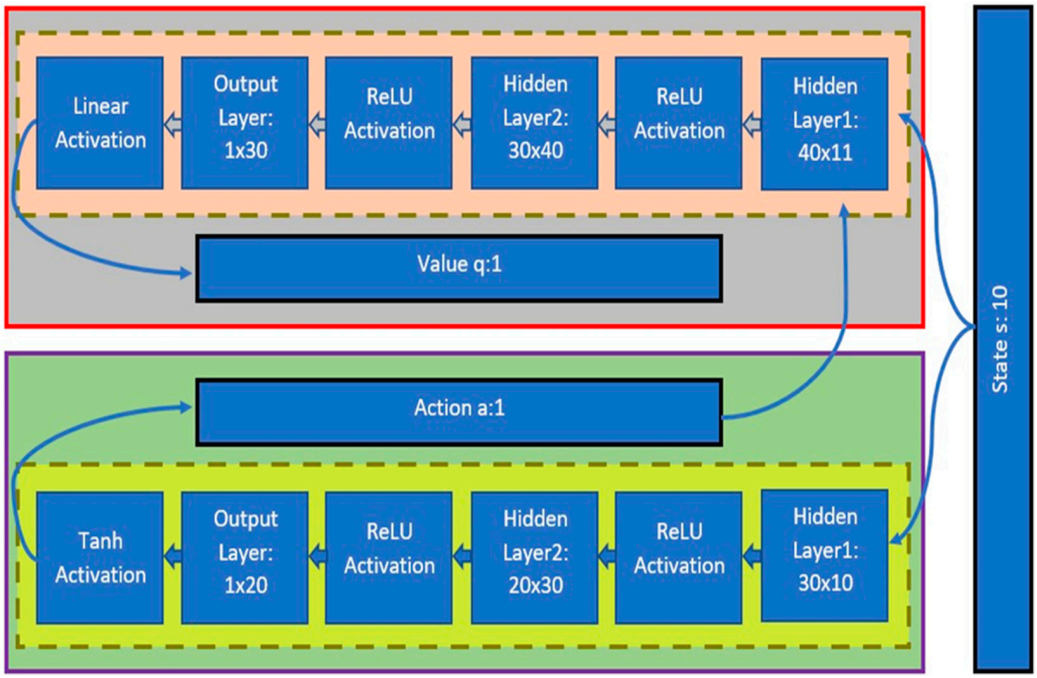

Figure 2. Q network and U network structure.

The proposed method mainly considers the DDPG algorithm and consists of the following:

1. Convolutional neural networks (CNN) are used to approximate networks Q and X, and they are updated to improve the algorithm's convergence.
2. To prevent the issue of sample correlation-related overfitting in neural networks, we use a replay buffer.
3. To enhance network convergence and stabilize the learning process, we develop the eval network and the target network.

The current action value y can be understood as the cumulative sum of the predicted future action value multiplied by the discount factor and the current reward value, according to the definition of the action value. As a result, the value of the current action y can be written as in Equation (4):

$$y = r_t + Y.Q_{target}\left(s_{t+1}, \acute{a}_{t+1} | \theta^{Q_{target}}\right)|_{\acute{a}_{t+1}} = X_{target}(s_{t+1}) \quad (4)$$

Significantly, the current action value y is more precise than the expected value by network Q with a calculated and actual reward constituent. The network parameters have been updated by the following formula Equation (5), whereas τ is the updating step, which is also called the learning rate.

$$\varsigma.\theta^{Q_{eval}} = \tau \,.\, \nabla.Loss\left(\theta^{Q_{eval}}\right) \quad (5)$$

Let us return to the X network: $X(s_t)$ is also divided into $X_{eval}(s_t)$ and the previous $X_{target}(s_t)$, with $X_{eval}(s_t)$ being used for network updating and action output and $U_{target}(s_t)$ is used to calculate $a\prime_{t+1}$. The updating equation for $X_{eval}(s_t)$ is established on the action value, which can be maximized by Equation (6):

$$\varsigma.\theta^{X_{eval}} = \tau.\nabla Q_{eval}\left(s_t, a_t | \theta^{X_{eval}}\right) \quad (6)$$

$$= \tau.\nabla_\theta X_{eval}(s_t).\, \nabla_a Q_{eval}\left(s_t, a_t | \theta^{X_{eval}}\right)\Big|_{a= X_{eval}(s_t)}$$

The actual target network parameter $\hat{\theta}^{target}$ can be stated as in Equation (7):

$$\hat{\theta}^{\text{target}} = (1 - \varepsilon) \cdot \theta^{\text{target}} + \varepsilon \cdot \theta^{\text{eval}} \tag{7}$$

We add Gaussian noise to the network output a_t, to control the network exploration rate, where the rate of ε is in the range of [0,1]. Thus, the actual modified network output $\hat{a}_t$ is as in Equation (8):

$$\hat{a}_t \widetilde{=} \mathrm{N}\left(a_t, \sigma^2\right) \tag{8}$$

In each training step, the decay rate ξ is decreased by the variance σ^2, as in Equation (9):

$$\sigma_{t+1} = \xi \cdot \sigma_t \tag{9}$$

Due to the high requirements for the tele-control of the telepresence robots' testing efficiency, dual networks and replay buffers have already strengthened their reliability. However, it is still necessary for such telepresence robot development to reduce the training time as much as feasible.

A typical technique to optimize the network is adding Gaussian noise to the output actions. However, the entire output from the neural network is typically selected when choosing an action. The proposed method can be used during the delay in the controlling signal from the teleoperator, considering that humans and neural networks excel in different areas.

The DRL-based techniques have a better capacity for self-adaptation during the lag of controlling signals from the teleoperator and will only require further training during lag compensation. A replay buffer is optimized to collect training samples to decrease the overall training time. Weighted training samples are added to the buffer to ensure the training concentration during a lack of signal. For each training step, the number of samples is as defined in Equation (10):

$$m_t = \sqrt{\frac{125}{v_t + 10}} \tag{10}$$

Algorithm 1 elaborates on the essential training algorithm. Based on the input current telepresence robot status s_t, the DRL-based process will be utilized to compute the output action a_t.

Algorithm 1. Training of Telepresence Robot Control Agent

Initialize : telepresence robot state, teleoperator command state.
Initialize : network Q_{eval}, X_{eval}, Q_{target}, X_{target} with random values
Initialize : replay buffer $\mathcal{R}$
for each episode, **do**
    observe the current state of the telepresence robot s_t
    for each step in the environment, **do**
        select action a_t from X_{eval} the network, according to the (s_t)
        wait 1 s to observe the telepresence robot's status s_{t+1}
        observe reward $r_t = R(s_t, a_t)$
        update current state $m_t = floor\left(sqrt\left(\frac{175}{v_t+12}\right)\right)$
        store (s_t, a_t, s_{t+1}, r_t) in replay buffer $\mathcal{R}$
        $Update : X_{eval}$ *and* Q_{eval}
        $Assign : \hat{\theta}^{target} = (1 - \varepsilon) \cdot \theta^{target} + \varepsilon \cdot \theta^{eval}$
    end for
end for

4. Results and Discussion

Several trials on physically manufactured telepresence robots are carried out to validate the proposed method. The telepresence robot prototype is trained using the real-time

practical implementation of the telepresence robot. The parameters are then fine-tuned using several experiments. Third, comparison experiments are used to validate the effects of network exploration intervention and replay buffer adjustment. In the end, various experiments are conducted for comparison with the prior autonomous technique.

The 3D model of our physically manufactured telepresence robot and the prototype that we developed for this study is shown in Figure 3. The telepresence robot is composed of two DC geared motors, one LCD screen, one high-resolution camera, and another circuit to control the kinematics of the telepresence robot. The experimental environment for the training and testing of the proposed method is shown in Figure 4. The location is a cardiology ward in Ghulam Muhammad Abad, the Faisalabad government's general hospital. A telepresence robot must travel 95 m to move from the starting point (a doctor's office) to the destination (an admitted patient ward). We reconstruct the hospital environment using Lidar. In a hospital setting, the main entrance of the reception area is where the telepresence robot begins its journey from point A. It travels through the main lobby, avoiding various static obstructions until it reaches destination point B, the patient admission ward. As illustrated in the figure, the multiple experiments follow a specific path according to the teleoperator-simulated natural control profiles, which will load data from the healthcare environment experiments for reproduction.

Figure 3. The 3D model and real working prototype of the telepresence robot.

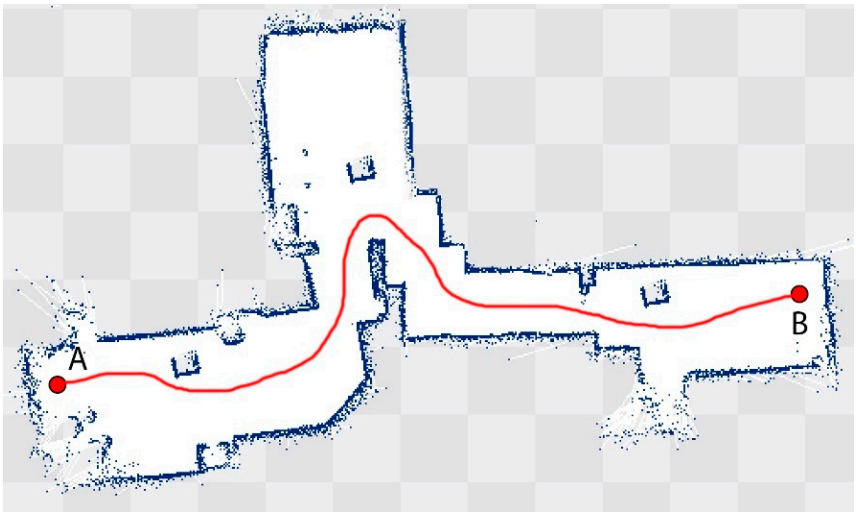

Figure 4. Moving of telepresence robot in healthcare environment experiment scenario from point A to point B.

Teleoperators can efficiently complete the telepresence robot's training and testing by implementing the DDPG algorithm. The DDPG installed on the telepresence robot performs a number of functions, including providing a controlling signal to the telepresence robot, receiving a control signal from the teleoperator, and retrieving the telepresence robot's current status.

The average difference equation $\text{Mean}_{|\text{diff}|}$ in Equation (11), the average action value, and the average reward will make up the evaluation indices, as well as the variance in the control difference Std_{diff} in Equation (12), and the average number of control errors $\overline{\text{N}}_{\text{error}}$ in Equation (13).

$$\text{Mean}_{|\text{diff}|} = \frac{\sum_{i=0}^{3600} \nabla v(0.5 \text{ x i})}{3600} \tag{11}$$

$$\text{Std}_{|\text{diff}|} = \frac{\sum_{i=0}^{3600} \nabla v(0.5 \text{ x i}) - \text{Mean}_{|\text{diff}|})^2}{3600} \tag{12}$$

$$\overline{\text{N}}_{\text{error}} = \frac{\sum_{i=1}^{N_{exp}} N_{error}^{i}}{N_{exp}} \tag{13}$$

The entire training curve, defined by each episode, consists of 80 cycles, because the control cycle of the algorithm lasts 1.25 s and the midway rule curve is 45 s. The results demonstrate that the telepresence robot can successfully control it after 60 to 80 training episodes. However, after 20 episodes, the network has progressively begun to grasp the control strategy of the controlling phases. According to the statistics, there were still variances in the controllability of the telepresence robot, even though the network could control it more precisely in episodes 40 and 50. The actual controlling behavior very closely resembled the desired controlling behavior after 80 to 100 episodes. The results of the experiments show that the target controllability will have less errors after more training sessions.

Figure 5 displays the midpoint curves considering the action values and average reward over time. After some training, the Q value and the average reward tend to stabilize when experiencing considerable initial fluctuations. They are discovered to vibrate more intensely in the midpoint rule curves. This is due to the similarity of each episode in the midpoint curves, as shown in Figure 6.

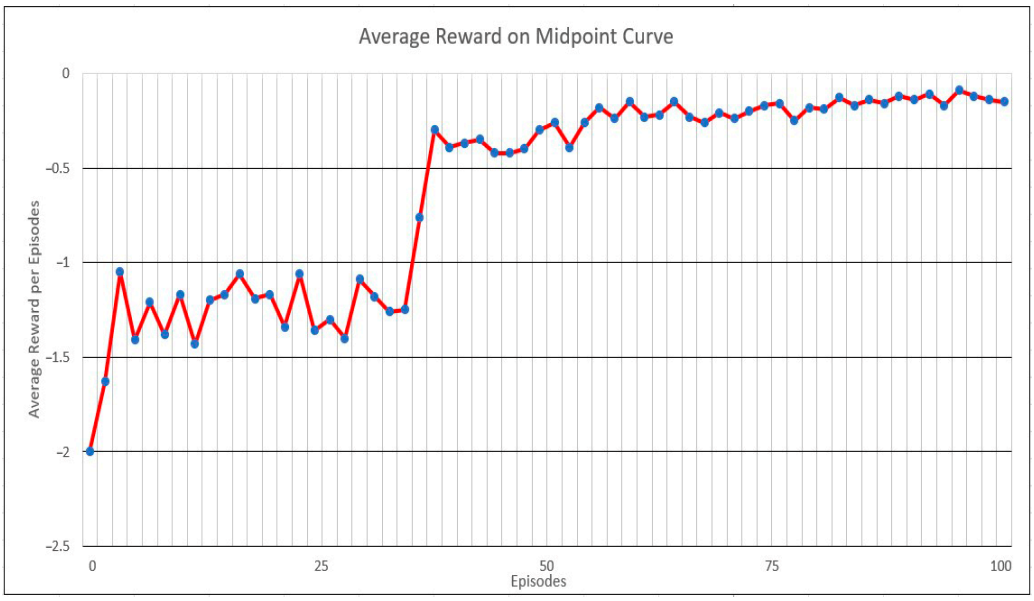

Figure 5. Average value of reward per episode.

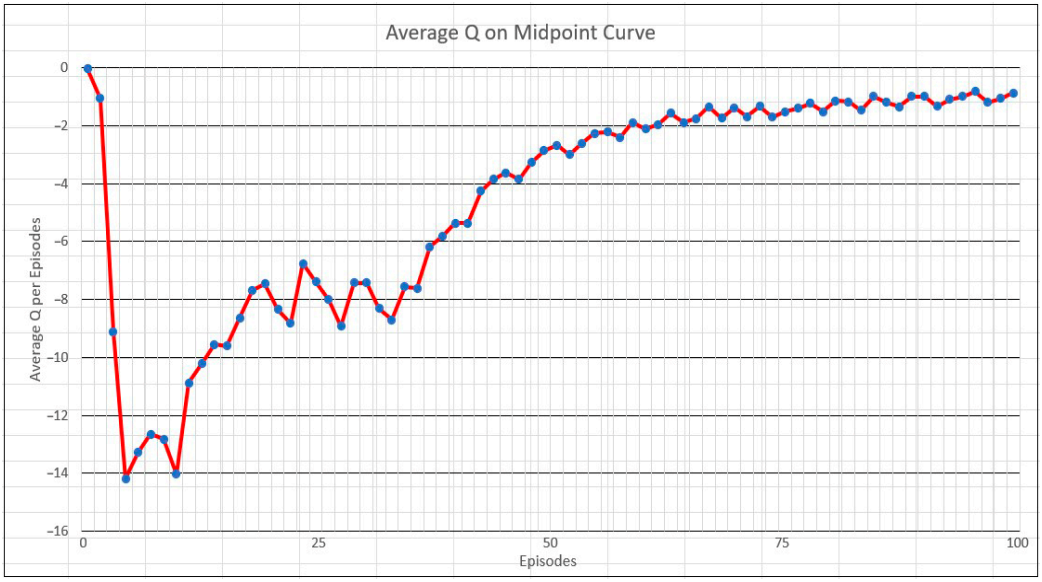

Figure 6. Average value of Q per episode.

Gaussian noise is attached to the actions of the network's output to fully understand the environment of the model during training. The size of the exploration space σ^2 is determined by the variance of the Gaussian noise. The network eventually picks up the environment model as the training continues. The decay rate influences the network's stability, and convergence will occur more slowly as it rises. When the decay rate ξ is more significant than 0.9999, the network cannot converge in four training cycles. The network has good stability when ξ is between 0.999 and 0.9999, as shown in Figure 7.

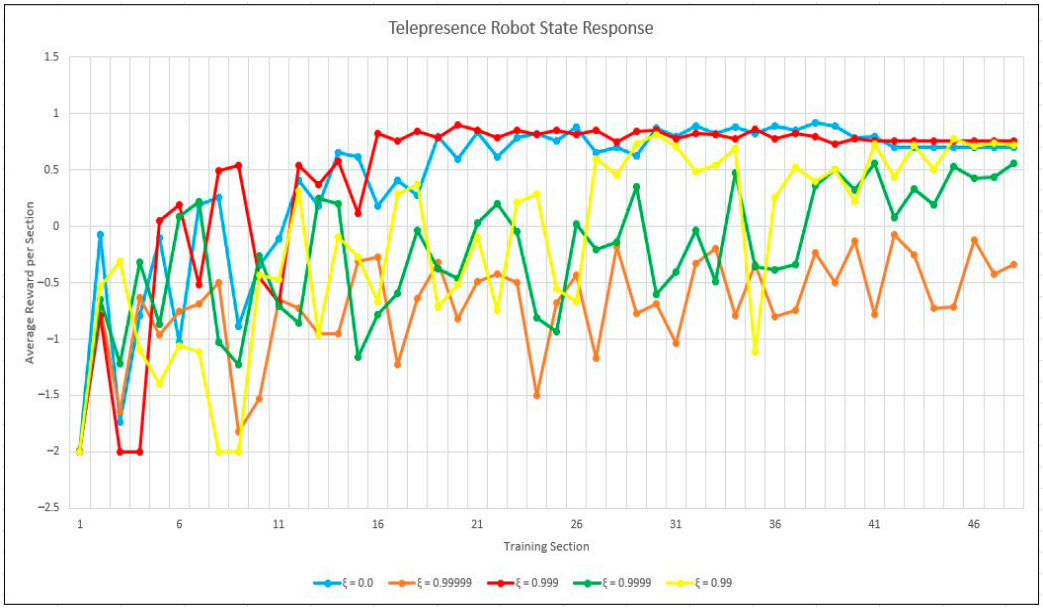

Figure 7. Telepresence robot state response.

The proposed method's most crucial assessment metric for the telepresence robot's control is the average number of controlling errors N_{error}. The results show that N_{error} may be more than five and usually emerges in low-speed parts after four training sessions using the current technique.

The target speed curve can be changed to enhance the replay buffer so as to increase the number of training samples added to the low-speed parts. A few comparative experiments

are carried out to confirm the effect of the optimization. N_{error}, $Mean_{diff}$, and Std_{diff} in the low-speed parts are decreased following optimization, whereas those in the high-speed sections are not significantly impacted, as shown in Table 2.

Table 2. Effect of optimization.

Parameters	Non-Optimized Mean	Optimized Mean
N_{exp}	2	15
$N_{error,\ v\leq30}$	6.500	0.867
$N_{error,\ v>30}$	0.500	0.200
$Mean_{\lvert diff\rvert,\ v\leq30}$	0.684	0.730
$Mean_{\lvert diff\rvert,\ v>30}$	0.544	0.602
$Std_{\lvert diff\rvert,\ v\leq30}$	0.852	0.839
$Std_{\lvert diff\rvert,\ v>30}$	0.830	0.910

Instead of using DDPG, many studies have used the segment proportion integration differentiation (SPID) approach. The fundamental concept of SPID is to categorize the environmental conditions into many cases and then to use PID with various proportional, integral, and differential parameters to regulate the speed in each situation. DDPG outperformed SPID when more training samples were added at the start or stop phases. While DDPG employs the current and the following three-second goal speeds, SPID uses the current target speed. After training, DDPG performs better than SPID in fully utilizing the environmental data. Tables 3 and 4 display the SPID and DDPG experimental findings for fifteen different trials. The outcomes demonstrate that DDPG moves faster and with fewer speed mistakes. The N_{error} average speed error count has dropped from 9.467 to 1.067.

Table 3. Comparison of mean value of DDPG with other algorithms.

Parameters	SPID Mean	DDPG Mean	SPID Best	DDPG Best
N_{exp}	15	15	-	-
N_{error}	9.467	1.067	3	0
$Mean_{\lvert diff\rvert}$	0.790	0.679	0.639	0.470
$Std_{\lvert diff\rvert}$	1.091	0.897	0.845	0.650

Table 4. Comparison of DDPG error, mean, and standard deviation with other algorithms.

	SPID	DDPG	SPID	DDPG	SPID	DDPG
	N_{error}	N_{error}	$Mean_{\lvert diff\rvert}$	$Mean_{\lvert diff\rvert}$	$Std_{\lvert diff\rvert}$	$Std_{\lvert diff\rvert}$
	21	2	0.872	0.734	1.200	1.015
	10	1	0.806	0.676	1.136	0.950
	8	0	0.800	0.615	1.072	0.736
	10	2	0.774	0.898	1.069	1.146
	6	0	0.838	0.877	1.190	0.964
	9	1	0.683	0.819	0.918	1.020
	7	2	0.813	0.631	1.118	0.794
	3	0	0.794	0.585	1.084	0.858
	3	3	0.639	0.674	0.845	0.956
	5	1	0.794	0.470	1.084	0.650
	9	0	0.792	0.626	1.072	0.854
	9	0	0.806	0.601	1.103	0.820
	14	1	0.737	0.839	1.041	1.130
	19	1	0.890	0.509	1.265	0.659
	9	2	0.807	0.631	1.164	0.904
Average	9.467	1.067	0.790	0.679	1.091	0.897

5. Conclusions

This paper proposes a DRL-based collaborative control framework for remote telepresence robots to compensate for delayed control signals. The DDPG algorithm optimizes the relationship between telepresence robot control and the teleoperator's delayed command signal, which could significantly reduce the total training time and improve the teleoperator control behavior in IoT-enabled healthcare environments. The replay buffer is expanded with the weighted training samples to improve and minimize the control instability. The suggested approach can reduce the control tracking errors and increase the training efficacy so as to better control the telepresence robot in the scenario of delayed signals. Compared to the other approaches, the DDPG-based method cooperates with the teleoperator. It allows smoother control in the case of 30-s delayed reception of the controlling signal from the teleoperator, with fewer control faults.

Moreover, this paper proposes a suitable method to control the telepresence robot in a more appropriate way in a dynamic unknown environment. Future work may be to develop a swarm-based approach to telepresence robots that will enable cooperation between different telepresence robots. The method can be designed so that multiple telepresence robots can work together in a coordinated and synchronized manner to achieve a common goal during a communication delay.

Author Contributions: Conceptualization, F.N. and M.N.K.; methodology, F.N.; software, A.A.; validation, F.N., M.N.K. and A.A.; formal analysis, F.N.; investigation, F.N. and M.N.K.; resources, F.N., M.N.K. and A.A.; data curation, F.N.; writing—original draft preparation, F.N.; writing—review and editing, F.N. and M.N.K.; visualization, F.N., M.N.K. and A.A.; supervision, M.N.K. and A.A.; project administration, M.N.K.; funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This Research work was funded by Institutional Fund Projects under grant no. (IFPIP: 210-611-1443). The authors gratefully acknowledge technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Araba.

Institutional Review Board Statement: This article does not contain any studies with human participants performed by any of the authors.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated during the study.

Conflicts of Interest: The authors declare no conflict of interest. The manuscript was written with the contribution of all authors.

References

1. Edwards, J. Telepresence: Virtual Reality in the Real World [Special Reports]. *IEEE Signal Process. Mag.* **2011**, *28*, 9–142. [CrossRef]
2. Alotaibi, Y. Automated Business Process Modelling for Analyzing Sustainable System Requirements Engineering. In Proceedings of the 2020 6th International Conference on Information Management (ICIM), London, UK, 27–29 March 2020. [CrossRef]
3. Soda, K.; Morioka, K. 2A1-E08 a Remote Navigation System Based on Human-Robot Cooperation (Cooperation between Human and Machine). In *Proceedings of the JSME Annual Conference on Robotics and Mechatronics (Robomec)*; Japan Science and Technology Agency: Kawaguchi, Japan, 2013. [CrossRef]
4. Bar-Shalom, Y. Update with Out-of-Sequence Measurements in Tracking: Exact Solution. *IEEE Trans. Aerosp. Electron. Syst.* **2002**, *38*, 769–777. [CrossRef]
5. Lakshmanna, K.; Subramani, N.; Alotaibi, Y.; Alghamdi, S.; Khalafand, O.I.; Nanda, A.K. Improved Me-taheuristic-Driven Energy-Aware Cluster-Based Routing Scheme for IoT-Assisted Wireless Sensor Net-works. *Sustainability* **2022**, *14*, 7712. [CrossRef]
6. Larsen, T.D.; Andersen, N.A.; Ravn, O.; Poulsen, N.K. Incorporation of Time Delayed Measurements in a Discrete-Time Kalman Filter. In Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No.98CH36171), Tampa, FL, USA, 18 December 1998. [CrossRef]
7. Bak, M.; Larsen, T.D.; Norgaard, M.; Andersen, N.A.; Poulsen, N.K.; Ravn, O. Location Estimation Using Delayed Measurements. AMC'98-Coimbra. In Proceedings of the 1998 5th International Workshop on Advanced Motion Control, Proceedings (Cat. No.98TH8354), Coimbra, Portugal, 29 June–1 July 1998. [CrossRef]
8. Chen, Y.; Hu, H. Internet of Intelligent Things and Robot as a Service. *Simul. Model. Pract. Theory* **2013**, *34*, 159–171. [CrossRef]

9. Loke, S.W. Context-Aware Artifacts: Two Development Approaches. *IEEE Pervasive Comput.* **2006**, *5*, 48–53. [CrossRef]
10. Cardenas, I.S.; Kim, J.-H. Advanced Technique for Tele-Operated Surgery Using an Intelligent Head-Mount Display System. In Proceedings of the 2013 29th Southern Biomedical Engineering Conference, Miami, FL, USA, 3–5 May 2013. [CrossRef]
11. Abdul Jalil, S.B.; Osburn, B.; Huang, J.; Barley, M.; Markovich, M.; Amor, R. Avatars at a Meeting. In Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction-CHINZ '12, New York, NY, USA, 2–3 July 2012. [CrossRef]
12. Park, C.H.; Howard, A.M. Real World Haptic Exploration for Telepresence of the Visually Impaired. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, New York, NY, USA, 5–8 March 2012. [CrossRef]
13. Vlahovic, S.; Suznjevic, M.; Kapov, L.S. Subjective Assessment of Different Locomotion Techniques in Virtual Reality Environments. In Proceedings of the 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), Cagliari, Italy, 29 May–1 June 2018. [CrossRef]
14. Shih, C.-F.; Chang, C.-W.; Chen, G.-D. Robot as a Storytelling Partner in the English Classroom-Preliminary Discussion. In Proceedings of the Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), Nigata, Japan, 18–20 July 2007. [CrossRef]
15. Gross, H.-M.; Schroeter, C.; Mueller, S.; Volkhardt, M.; Einhorn, E.; Bley, A.; Martin, C.; Langner, T.; Merten, M. Progress in developing a socially assistive mobile home robot companion for the elderly with mild cognitive impairment. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011), San Francisco, CA, USA, 25–30 September 2011. [CrossRef]
16. Wikipedia. Willow Garage. Available online: https://en.wikipedia.org/wiki/Willow_Garage#Robots (accessed on 1 November 2022).
17. Anybots. Anybots-Your Personal Avatar. Available online: http://anybots.com/#qbLaunch (accessed on 12 August 2022).
18. VGo Communications. VGo Robot. Available online: http://www.vgocom.com (accessed on 21 September 2022).
19. TelaDoc, I. InTouch Health | Home. Available online: http://www.intouchhealth.com/products-remote-presence-endpoint- (accessed on 14 December 2022).
20. Mantarobot Inc. Mantarobot. Available online: http://mantarobot.com/ (accessed on 13 August 2022).
21. Giraff Inc. Giraff Robot for Caregivers. Available online: http://www.giraff.org/professional-caregivers/?lang=en/ (accessed on 13 September 2022).
22. Lazewatsky, D.A.; Smart, W.D. An inexpensive robot platform for teleoperation and experimentation. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011. [CrossRef]
23. Subahi, A.F.; Khalaf, O.I.; Alotaibi, Y.; Natarajan, R.; Mahadev, N.; Ramesh, T. Modified Self-Adaptive Bayesian Algorithm for Smart Heart Disease Prediction in IoT System. *Sustainability* **2022**, *14*, 14208. [CrossRef]
24. Feng, Y.; Liu, L.; Shu, J. A Link Quality Prediction Method for Wireless Sensor Networks Based on XGBoost. *IEEE Access* **2019**, *7*, 155229–155241. [CrossRef]
25. Singh, S.P.; Alotaibi, Y.; Kumar, G.; Rawat, S.S. Intelligent Adaptive Optimisation Method for Enhance-ment of Information Security in IoT-Enabled Environments. *Sustainability* **2022**, *14*, 13635. [CrossRef]
26. Smeets, H.; Meurer, T.; Shih, C.-Y.; Marron, P.J. Demonstration abstract: A lightweight, portable device with integrated USB-Host support for reprogramming wireless sensor nodes. In Proceedings of the 2014 13th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Berlin, Germany, 15–17 April 2014. [CrossRef]
27. He, D.; Chen, C.; Chan, S.; Bu, J.; Vasilakos, A.V. A Distributed Trust Evaluation Model and Its Application Scenarios for Medical Sensor Networks. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 1164–1175. [CrossRef]
28. Rashid, R.A.; Rezan Bin Resat, M.; Sarijari, M.A.; Mahalin, N.H.; Abdullah, M.S.; Hamid, A.H.F.A. Performance investigations of frequency agile enabled TelosB testbed in Home area network. In Proceedings of the 2014 IEEE 2nd International Symposium on Telecommunication Technologies (ISTT), Langkawi, Malaysia, 24–26 November 2014. [CrossRef]
29. Wang, Y.; Tan, R.; Xing, G.; Wang, J.; Tan, X. Profiling Aquatic Diffusion Process Using Robotic Sensor Networks. *IEEE Trans. Mob. Comput.* **2014**, *13*, 880–893. [CrossRef]
30. Nagappan, K.; Rajendran, S.; Alotaibi, Y. Trust Aware Multi-Objective Metaheuristic Optimization Based Secure Route Planning Technique for Cluster Based IIoT Environment. *IEEE Access* **2022**, *10*, 112686–112694. [CrossRef]
31. Klein, G.; Murray, D. Parallel Tracking and Mapping for Small AR Workspaces. In Proceedings of the 2007 6th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Nara, Japan, 13–16 November 2007. [CrossRef]
32. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [CrossRef]
33. Newcombe, R.A.; Lovegrove, S.J.; Davison, A.J. DTAM: Dense tracking and mapping in real-time. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011. [CrossRef]
34. Singh Gill, H.; Ibrahim Khalaf, O.; Alotaibi, Y.; Alghamdi, S.; Alassery, F. Multi-Model CNN-RNN-LSTM Based Fruit Recognition and Classification. *Intell. Autom. Soft Comput.* **2022**, *33*, 637–650. [CrossRef]
35. Klein, G.; Murray, D. Parallel Tracking and Mapping on a camera phone. In Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Orlando, FL, USA, 19–22 October 2009. [CrossRef]
36. Taleb, T.; Dutta, S.; Ksentini, A.; Iqbal, M.; Flinck, H. Mobile Edge Computing Potential in Making Cities Smarter. *IEEE Commun. Mag.* **2017**, *55*, 38–43. [CrossRef]

37. Liu, P.; Willis, D.; Banerjee, S. ParaDrop: Enabling Lightweight Multi-tenancy at the Network's Extreme Edge. In Proceedings of the 2016 IEEE/ACM Symposium on Edge Computing (SEC), Washington, DC, USA, 27–28 October 2016. [CrossRef]
38. Alotaibi, Y.; Noman Malik, M.; Hayat Khan, H.; Batool, A.; ul Islam, S.; Alsufyani, A.; Alghamdi, S. Suggestion Mining from Opinionated Text of Big Social Media Data. *Comput. Mater. Contin.* **2021**, *68*, 3323–3338. [CrossRef]
39. Tsai, P.-S.; Wang, L.-S.; Chang, F.-R. Modeling and Hierarchical Tracking Control of Tri-Wheeled Mobile Robots. *IEEE Trans. Robot.* **2006**, *22*, 1055–1062. [CrossRef]
40. Yong, Q.; Yanlong, L.; Xizhe, Z.; Ji, L. Balance control of two-wheeled self-balancing mobile robot based on TS fuzzy model. In Proceedings of the 2011 6th International Forum on Strategic Technology (IFOST), Harbin, China, 22–24 August 2011. [CrossRef]
41. Salerno, A.; Angeles, J. A New Family of Two-Wheeled Mobile Robots: Modeling and Controllability. *IEEE Trans. Robot.* **2007**, *23*, 169–173. [CrossRef]
42. Alotaibi, Y. A New Meta-Heuristics Data Clustering Algorithm Based on Tabu Search and Adaptive Search Memory. *Symmetry* **2022**, *14*, 623. [CrossRef]
43. Chaplot, D.S.; Gandhi, D.; Gupta, S.; Gupta, A.; Salakhutdinov, R. Learning to Explore Using Active Neural SLAM. *arXiv* **2020**. [CrossRef]
44. Ramakrishnan, S.K.; Al-Halah, Z.; Grauman, K. Occupancy Anticipation for Efficient Exploration and Navigation. In Proceedings of the European Conference on Computer Vision ECCV, Glasgow, UK, 23–28 August 2020. [CrossRef]
45. Gupta, S.; Davidson, J.; Levine, S.; Sukthankar, R.; Malik, J. Cognitive Mapping and Planning for Visual Navigation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
46. Tai, L.; Liu, M. A robot exploration strategy based on Q-learning network. In Proceedings of the 2016 IEEE International Conference on Real-Time Computing and Robotics (RCAR), Angkor Wat, Cambodia, 6–10 June 2016. [CrossRef]
47. Monahan, G.E. State of the Art—A Survey of Partially Observable Markov Decision Processes: Theory, Models, and Algorithms. *Manag. Sci.* **1982**, *28*, 1–16. [CrossRef]
48. Anuradha, D.; Subramani, N.; Khalaf, O.I.; Alotaibi, Y.; Alghamdi, S.; Rajagopal, M. Chaotic Search-and-Rescue-Optimization-Based Multi-Hop Data Transmission Protocol for Underwater Wireless Sensor Networks. *Sensors* **2022**, *22*, 2867. [CrossRef]
49. Tai, L.; Liu, M. Mobile Robots Exploration through Cnn-Based Reinforcement Learning. *Robot. Biomim.* **2016**, *3*, 24. [CrossRef] [PubMed]
50. Bai, S.; Chen, F.; Englot, B. Toward autonomous mapping and exploration for mobile robots through deep supervised learning. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017. [CrossRef]
51. Srilakshmi, U.; Alghamdi, S.A.; Vuyyuru, V.A.; Veeraiah, N.; Alotaibi, Y. A Secure Optimization Routing Algorithm for Mobile Ad Hoc Networks. *IEEE Access* **2022**, *10*, 14260–14269. [CrossRef]
52. Marroquin, A.; Gomez, A.; Paz, A. Design and implementation of explorer mobile robot controlled remotely using IoT technology. In Proceedings of the 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Pucon, Chile, 18–20 October 2017. [CrossRef]
53. Srividhya, S.; Kumar, G.D.; Manivannan, J.; Mohamed Wasif Rihfath, V.A.; Ragunathan, K. IoT Based Vigilance Robot using Gesture Control. In Proceedings of the 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 15–16 February 2018. [CrossRef]
54. Sennan, S.; Kirubasri; Alotaibi, Y.; Pandey, D.; Alghamdi, S. EACR-LEACH: Energy-Aware Cluster-based Routing Protocol for WSN Based IoT. *Comput. Mater. Contin.* **2022**, *72*, 2159–2174. [CrossRef]
55. Wang, S.; Zhao, H.; Hao, X. Design of an intelligent housekeeping robot based on IOT. In Proceedings of the 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Okinawa, Japan, 28–30 November 2015. [CrossRef]
56. Lam, M.C.; Prabuwono, A.S.; Arshad, H.; Chan, C.S. A Real-Time Vision-Based Framework for Human-Robot Interaction. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 257–267. [CrossRef]
57. Tomar, P.; Kumar, G.; Verma, L.P.; Sharma, V.K.; Kanellopoulos, D.; Rawat, S.S.; Alotaibi, Y. CMT-SCTP and MPTCP Multipath Transport Protocols: A Comprehensive Review. *Electronics* **2022**, *11*, 2384. [CrossRef]
58. Zhuang, C.; Yamins, D. Using Multiple Optimization Tasks to Improve Deep Neural Network Models of Higher Ventral Cortex. *J. Vis.* **2018**, *18*, 905. [CrossRef]