Robotics and Computer–Integrated Manufacturing 85 (2024) 102636

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

journal homepage: www.elsevier.com/locate/rcim

Sustainable Robotic Joints 4D Printing with Variable Stiffness Using Reinforcement Learning

Moslem Mohammadi [a], Abbas Z. Kouzani [a], Mahdi Bodaghi [b], John Long [a], Sui Yang Khoo [a], Yong Xiang [c], Ali Zolfagharian [a,*]

[a] School of Engineering, Deakin University, Geelong, Victoria 3216 Australia
[b] Department of Engineering, School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK
[c] School of Information Technology, Deakin University, Geelong, Victoria 3216 Australia

ARTICLE INFO

Keywords:
4D printing
Variable stiffness
Sustainable
Robots
Reinforcement learning

ABSTRACT

Nowadays, a wide range of robots are used in various fields, from car factories to assistant soft robots. In all these applications, effective control of the robot is vital to perform the tasks assigned to them. Soft robots and actuators have several advantages over traditional rigid manipulators, including lower power consumption, lighter weight, safer operation in contact with live tissues, inexpensive manufacturing costs, and quicker movements. However, controlling them is more challenging. This paper presents a three-dimensional (3D) printed structure combined with carbon fibres to provide a stimulus signal, known as four-dimensional (4D) printing. Depending on the application, the structure could provide various levels of stiffness to adapt to new conditions. A nonlinear controller based on reinforcement learning (RL) algorithms is also presented to control the stiffness of soft joints. The controller is tuned based on the mathematical model of the Simulink setup and then applied to the experimental setup. The results show that the RL controller has a high potential to adapt online to various unforeseen conditions. Additionally, this controller offers a significantly reduced lag for specific inputs, such as a sinusoidal signal, while considerably decreasing power consumption in contrast to a linear controller. This is a significant advantage of variable stiffness 4D-pritned soft joints for sustainable and circular robots manufacturing in portable medical and wearable sustainable robotic applications.

1. Introduction

Engineers and scientists have become very interested in Additive manufacturing (AM) because of its outstanding flexibility and capacity to print complex forms [1,2]. Recently, fabricating soft robots and actuators through Four-dimensional (4D) printing has become more common because of their ability to change their structure under specific environmental stimuli, such as temperature, electric field, magnetic field, and light [3,4]. In addition, 4D printing tools provide tremendous benefits for managing a variety of parameters, such as assembly time and total fabricating cost [5–7]. Researchers use various structures made of hydrogels, Liquid Crystal Elastomers, magnetic material, and Shape memory polymers (SMP) in the emerging 4D printing technology [3,8,9].

Even nowadays, a vast number of conventional robots have been fabricated via traditional manufacturing methods for a variety of industrial purposes. Some of these methods' most notable drawbacks are their high rates of energy consumption and high waste of material [10]. Researchers worldwide are exploring novel methods and mechanisms to develop soft robots to overcome the disadvantages of traditional rigid and stiff counterparts [11–14]. 4D printing could be considered noticeably beneficial for manufacturing soft robots compared to other manufacturing methods. 4D printing originated based on Three-Dimensional (3D) printing, so fabricating soft robots through 4D printing provides not only the fabrication of complex 3D structures based on smart material but also has the primary benefits of AM

Abbreviations: 3D, three-dimensional; 4D, four-dimensional; RL, reinforcement learning; AD, additive manufacturing; SMP, shape memory polymers; DoF, degrees of freedom; PID, proportional integral derivative; LQR, linear quadratic regulator; NN, neural networks; DRL, deep reinforcement learning; PLA, polylactic acid; AI, artificial intelligence; PPO, proximal policy optimization; A3C, asynchronous advantage actor-critical; TRPO, trust region policy optimization; TD3, twin-delayed deep deterministic policy gradient; DDPG, deep deterministic policy gradient.

* Corresponding author.
E-mail address: a.zolfagharian@deakin.edu.au (A. Zolfagharian).

https://doi.org/10.1016/j.rcim.2023.102636
Received 10 May 2023; Received in revised form 26 July 2023; Accepted 27 August 2023
Available online 31 August 2023

fabrication, such as controllability, reproducibility, and repeatability [15]. Robots frequently manipulate items in their environment as one of their numerous regular tasks [16]. These manipulators often have various Degrees of freedom (DoF) and imitate human limbs [17]. Robotic limbs can also be divided into soft and rigid types. Robotic manipulators with frames made of aluminium or steel are more popular to increase their tensile strength. Using those materials would result in rigid, heavy, and hard-to-move structures.

Investigation into the design, manufacturing and control of soft robotic manipulators has been sparked by advancements in material technology, robotics, and the growing need for lightweight, flexible, and portable robotic arms. The use of plastic, compliant mechanism, or carbon-fibre frames notably lower the cost of manufacture and the energy used by the gadget and provides flexibility. Also, compared to conventional rigid grippers, soft robotic grippers with compliance are far more effective in handling items with random shapes or fragility. The primary drawbacks of such systems are their low grabbing power and inability to perform various tasks in different environments due to the lack of variable stiffness in soft robotic grippers. These limits have led researchers to propose smart materials with variable stiffness and novel designs [18–20]. Shape memory material is one of the most well-known options for fabricating variable stiffness mechanisms that is a temperature-stimulated material [21–23]. SMPs could be 4D printed in a permanent structure (Fig. 1a). However, when its temperature reaches above the material's glass transition temperature, T_g, the structure can be temporarily distorted into another shape that shows variable stiffness [24](Fig. 1b). Then, by reducing the temperature to less than T_g, the temporary stiffness could be saved. Heating the sample again to the above T_g causes a recovery in the SMP to its initial printed geometry (Fig. 1c). Soft robots and actuators can exploit variable stiffness structures' benefits to gain the ability to operate in various situations.

The robotic arm's flexibility can result in different degrees of inaccuracy, such as slower settling time, particularly near the endpoint. When it comes to soft manipulators, rigid manipulator control algorithms fall short. Therefore, specific algorithms that are precise and compact must be designed to adapt to the flexibility of robotic arms. Proportional integral derivative (PID) and Linear quadratic regulator (LQR) are the most commonly used linear control techniques, although because of the nonlinearity of the system, they are not desired [25]. Akyuz et al. developed a PID controller with full-state feedback to control a robotic arm with a soft joint [26]. Also, other linear control methods are able to be coupled with the PID controller to evolve reliable controlling methods and increase their performance [27]. Another powerful and widely utilized method for linear control of robots' arms is the LQR. The LQR has become quite popular since this technique tries to minimize the quadratic cost. However, both LQR and PID responses produce comparable results in trajectory tracking [28]. Due to linear techniques' simplicity, it continues to be the industry's central controller [29].

Researchers have attempted various methods other than PID and LQR to design a controller for trajectory tracking of flexible joint robots. The fuzzy-PID controller is one of the most common controllers utilized for this purpose [30–33]. Neural networks (NN) are tested to gain better control over flexible joint manipulators [34,35]. A study has suggested a hybrid adaptive controller based on neuro-fuzzy logic [36]. Also, RL techniques are exploited to control soft links and joint mechanisms [37–40].

RL in the control field is a method to develop adaptive controllers which can be trained online (in real-time) and approach the optimum controller. RL employs a reward-penalty technique to adjust the control system's behaviour. The controller adapts its actions in a way to get more rewards. In the RL framework, this control system component is known as the "actor" [41] (Fig. 2). Another component to assess the system's effectiveness is called "critic" (Fig. 2). With the help of this method, a class of adaptive controllers can be implemented that learn optimum control strategies by solving Hamilton-Jacobi-Bellman equations online without requiring a thorough understanding of system dynamics [38]. This article applies this concept to control a stiff arm with a 4D-printed flexible joint, which potentially has a variety of applications in industrial procedures. Also, the stiffness of the joint can vary based on the external stimuli; therefore, the proposed controller is tailored to variable stiffness flexible joint robots.

The development of a tracking controller for robots equipped with variable stiffness structures subjected to significant uncertainties has remained a serious control issue despite the large amount of research conducted in this field [42–44]. The RL controller has a high potential to adapt online to various unforeseen conditions and offers robustness and adaptability compared to linear controllers [45,46] without requiring a thorough knowledge of the system. Therefore, the primary objective of this study is to create an RL controller for a soft robotic joint with variable stiffness qualities and evaluate the performance of the RL controller in adapting to various unforeseen conditions. Additionally, another project objective is presenting a 3D-printed structure combined with carbon fibres to fabricate a variable stiffness joint and assessing its stiffness characteristics, repeatability of changing its stiffness and required power and time to change its stiffness. The paper's objectives can contribute to the soft robotics field by addressing the gap in the research field, providing a novel approach to controlling variable stiffness flexible joints in 4D-printed structures, and leading to advancements in the control and performance of soft robots, enabling their

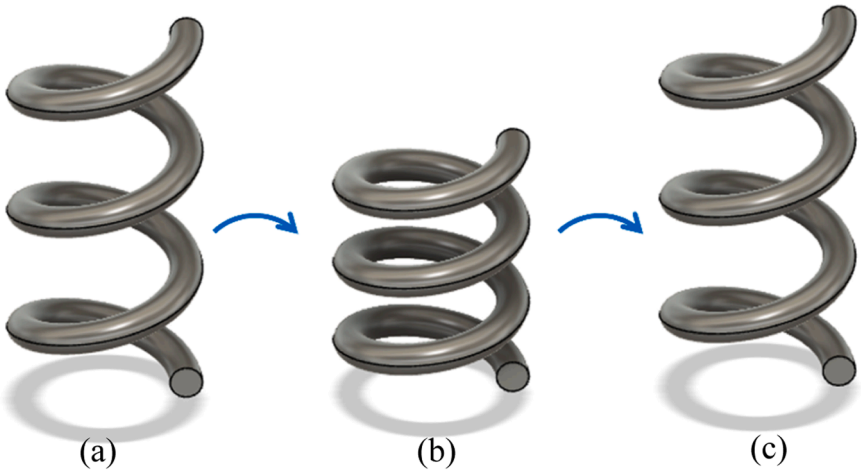

Fig. 1. The schematic design of 4D variable stiffness joint mechanism in this work, (a) 3D printed SMP structure with an initial stiffness, (b) heated structure and deformed to a new structure then cooled down to represent a new stiffness, (c) structure is heated again to recover its initial stiffness.

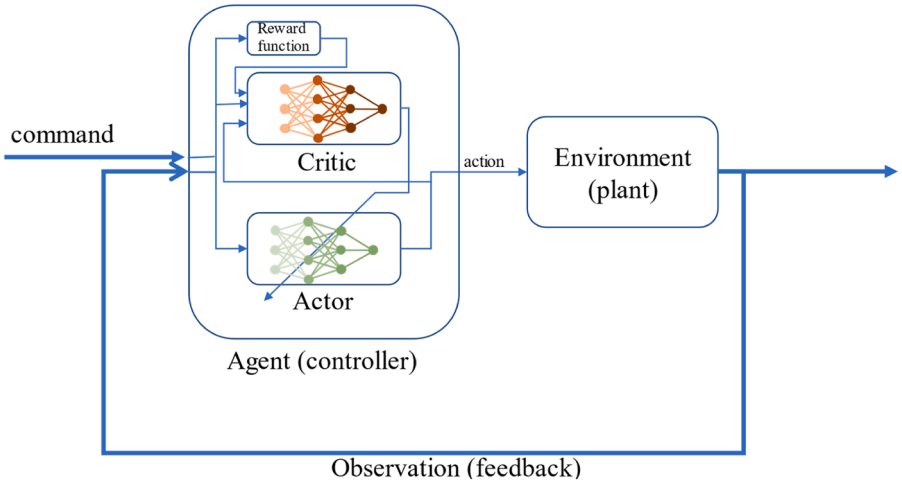

Fig. 2. Deep reinforcement learning controller.

practical use in various applications.

The paper is structured as follows: Section 2 describes the proposed methodology in the paper, the RL controller principles, and the mathematical model of the flexible joints; Section 3 presents the numerical simulations; Section 4 describes the experimental results, future directions, and applications of the proposed system; and the last section presents conclusions.

2. Methodology

We provide a method for developing a controller to increase the accuracy of robotic link actuation in following the input signal. At the same time, its joint is equipped with a variable stiffness 4D-printed joint by applying the Deep reinforcement learning (DRL) method. Moreover, the goal is to minimize the tracking error, vibration of the manipulator's arm, and power consumption. The experimental setup includes a motor and a rigid beam linked to a shaft. Two 4D-printed SMP structures used

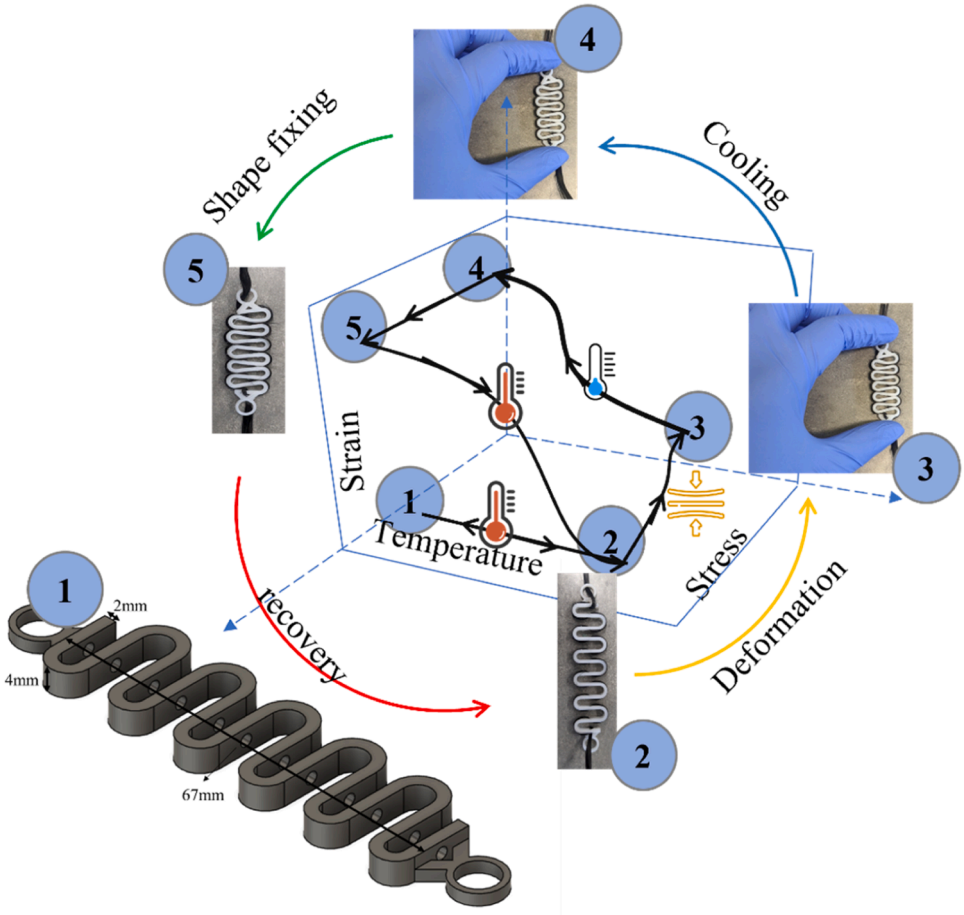

Fig. 3. 3D-printed spring designed to provide variable stiffness joint.

to connect the beam to the rotating gears provide the joint's flexibility. These structures give the joint more flexibility in terms of stiffness. The structure is 4D-printed based on SMP material with Polylactic acid (PLA) filament with a 1.75 mm diameter in the shape of an S-spring (Fig. 3-phase 1). In addition, the spring can be heated, deformed, and cooled down to present variable stiffness. A carbon fibre string is added to the spring manually by passing through several holes designed on the S-spring to provide a stimulus signal and to change the rigidness. The carbon fibre heating wire manufactured by Hefei Minco Heating Cable Co contains 12,000 threads, its resistance is 33 Ohm/m, and its diameter is 3 mm. The emitted thermal wavelength ranges from 8um to 18um, and the wire can withstand temperatures up to 200 °C.

Several designs with various thicknesses were 3D printed on the Flashforge printer to evaluate the potential of the PLA spring in recovering the initial shape. The recovery time was evaluated, and their durability on the flexible joint was checked. Based on the result, the final spring was designed, and Fig. 3 depicts the variable stiffness structure cycle. The design is heated to 70°C through the carbon fibre wire (Fig. 3-phase 1 to 2), powered by a direct current source set to 10 V. Then the spring is compressed (Fig. 3-phase 2 to 3), the power source is turned off to cool down the structure while the external force is still present (Fig. 3-phase 3 to 4). The spring's cooling procedure is conducted passively at ambient temperature (22 °C). While the temperature of the spring is below its T_g, the external force is gradually removed, and the spring maintains its new compressed shape that represents a higher stiffness (Fig. 3-phase 4 to 5). The structure's stiffness can be varied by varying the amount of compression in the shape of the 3D-printed structure. The spring recovers its initial form after the power source is connected, and the integrated carbon fibre provides temperatures higher than the T_g (Fig. 3-phase 5 to 2). Finally, the spring can retain its initial form by decreasing the temperature to lower than T_g (Fig. 3-phase 2 to 1). These diagrams depict an ideal condition, but in each cycle, the PLA spring undergoes slight deformation, resulting in a reduction in its recovery performance. A durability test was performed, and Fig. 4a illustrates the shape recovery of 10 springs with a thickness of 3mm over multiple cycles. As shown in the figure, the recovery performance (recovery to its original shape and length) of the springs decreases to 52% on average at the 25th cycle.

Moreover, the tensile test was performed on the variable stiffness spring to evaluate its stiffness, as depicted in Fig. 5. The figure shows that the spring has higher stiffness while it drops by increasing the spring's temperature. The compressed or expanded structure is not recovered fully to its initial shape after heating to its glass transition temperature. Different thicknesses of the structure (the length of the structure is fixed due to the Quanser robotic arm unit constraints) are tested to check their recovery efficiency. Fig. 4b shows that its recovery to initial length decreases by increasing the spring's thickness. On the other hand, the thinner printed structure becomes fragile and less stiff, which is not appropriate for the vibration damping test. As a trade-off, the structure with 2 mm thickness, 4 mm height, and 67 mm length (Fig. 3-phase 1) is affixed to the setup with 91% recovery performance.

In addition, the recovery time and the power consumption of the various designs are explored. Results are calculated based on the power consumption rate of the carbon fibre and the time consumed to recover the shape to 90% of its original length, based on Fig. 4b. The trend in the figure shows that thinner springs can recover faster than thicker designs. The structure length is equal; therefore, the power consumption rate must be equal while the voltage is kept constant. As a result, the thicker structure with a longer recovery time required more energy to reach its initial state. With the 10V power voltage and the carbon fibre length of 26 cm, the utilized electrical energy required for recovery is 100 J, 120 J, and 170 J for springs with 1.5, 2, and 3mm thickness, respectively.

2.1. Mathematical model

The system's different components, including mechanical and electrical parts, must be considered to develop a mathematical model of the flexible joint system. The system's input is the voltage applied to the DC motor, and its output is the angle of the motor's rotor and the rigid manipulator's angle; θ and α. A traditional mechanics-based analytical approach or the Lagrange method can be utilized to determine a mathematical model. The Lagrange equation is as follows:

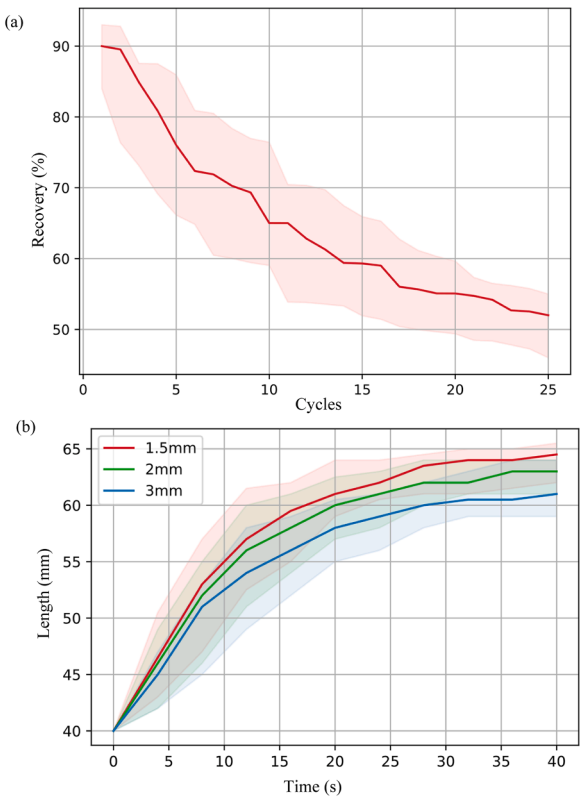

Fig. 4. a) Shape recovery of the springs with 3mm thickness in 25 cycles. The shaded area shows the 90% confidence range, while the solid line shows the average recovery. b) Shape recovery of the compressed 4D-printed joint with different design parameters. The solid line represents the average recovery, and the shade represents a 90% confidence interval.

$$L = T - V \tag{1}$$

V and T are elastic energies and rotational kinetic energies, expressed as:

$$T = \frac{1}{2}J\omega^2 = \frac{1}{2}J_{eq}\dot{\theta}^2 + \frac{1}{2}J_L(\dot{\theta}+\dot{\alpha})^2 \tag{2}$$

$$V = \frac{1}{2}k_s\alpha^2 \tag{3}$$

J_{eq} is the total manipulator's moment of inertia and J_L is the link's moment of inertia. Also, the Euler Lagrange equation is written in Eq. 4.

$$\frac{\mathrm{d}}{\mathrm{d}_t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = Q_i, \ while \ q_i = \begin{bmatrix}\theta \\ \alpha\end{bmatrix}, \ Q_i = \begin{bmatrix}\tau - B_{eq}\dot{\theta} \\ B_L\dot{\alpha}\end{bmatrix} \tag{4}$$

τ is the torque of the motor, and $B_{Eq.}$ and B_L are the viscous fraction coefficient of the manipulator and the link, respectively. By applying the partial derivative of the Lagrange respective to q_i after replacing Eqs. 2 and 3 in Eq. 1, we get to the following equations:

$$\begin{bmatrix}J_{eq}\ddot{\theta} + J_L\left(\ddot{\theta}+\ddot{\alpha}\right) \\ J_L\left(\ddot{\theta}+\ddot{\alpha}\right) + K_s\alpha\end{bmatrix} = \begin{bmatrix}\tau - B_{eq}\dot{\theta} \\ B_L\dot{\alpha}\end{bmatrix} \tag{5}$$

The load torque can be derived from Kirchhoff's laws:

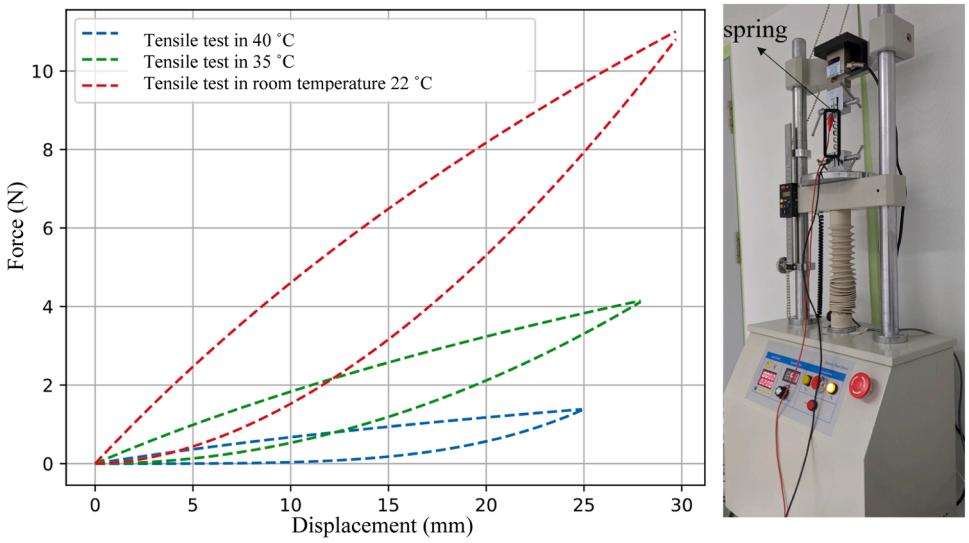

Fig. 5. a) Variable stiffness spring tensile test while the spring holds various temperatures. b) Tensile test setup.

$$\tau = \frac{k_g k_m V - k_m k_t k_g^2 \dot{\theta}}{R_m} \tag{6}$$

After replacing τ in Eq. 5 with its equivalent in Eq. 6, it can be rewritten as follows:

$$\begin{bmatrix} J_{eq}\ddot{\theta} + J_L\left(\ddot{\theta} + \ddot{\alpha}\right) \\ J_L\left(\ddot{\theta} + \ddot{\alpha}\right) + k_s\alpha \end{bmatrix} = \begin{bmatrix} \dfrac{k_g k_m V - k_m k_t k_g^2 \dot{\theta}}{R_m} - B_{eq}\dot{\theta} \\ B_L\dot{\alpha} \end{bmatrix} \tag{7}$$

K_s symbol in the equation represents the springs stiffness and motor back EMF constant, and its torque coefficient and high gear ratio of the motor are shown by k_m, k_t and k_g, respectively. Additionally, R_m represents the armature resistance of the motor. By analyzing Eq. 7, the state space equation of the system can be written as:

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{8}$$

With the following parameters while $x = [\theta \quad \alpha \quad \dot{\theta} \quad \dot{\alpha}]^T$

$$\begin{aligned} A &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\dfrac{K_s}{J_{eq}} & \dfrac{B_L}{J_{eq}} & \dfrac{k_s}{J_{eq}} & -\dfrac{B_L}{J_{eq}} \\ \dfrac{Ks}{J_L} & -\dfrac{B_L}{J_L} & -\dfrac{Ks}{J_L} & \dfrac{k_m k_t k_g^2}{R_m J_{eq}} - \dfrac{B_{eq} + B_L}{J_{eq}} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \dfrac{k_g k_m}{R_m J_{eq}} \end{bmatrix} \\ C &= [1 \quad 1 \quad 0 \quad 0], \quad D = 0 \end{aligned} \tag{9}$$

These matrices in Eq. 9 are utilized in the Simulink with the parameter provided in [47]. In Eq. 3, the assumption is made that the stiffness of the spring remains constant, despite the fact that it can be modified through a stimulating signal. However, the derived model remains valid, as it is assumed that the stiffness of the spring is controllable and can be adjusted to a new value in order to obtain a new model. This implies that the stiffness of the spring can be changed and fixed with the updated value.

2.2. RL algorithm

The creation of self-driving agents that interact with their surroundings to discover the most effective behaviours and enhance them over time via trial and error is one of the main objectives of the Artificial intelligence (AI) field. Making AI systems responsive and capable of successfully learning has been a long-lasting problem. RL provides a mathematical framework with clear rules for experience-driven, self-directed learning procedures. An actor and a critic are the two primary components of an RL-type controller in many applications (Fig. 2). The fundamental learning approach in autonomous controllers is RL, which uses feedback to enhance future actions. The actor conducts a sequence of acts on the environment (variable stiffness joint), and the critic is the one who watches the effects of these actions. The RL's significant components are the learning policy, reward, value function, and systems (environment) model. The system state, however, is a crucial component that must be known in order to assess the value function. The policy outlines the activities the actor will take to interact with the system. The system creates a single value (the reward) resulting from this interaction, which it then delivers to the critic.

The actor's actions are taken in order to maximize this reward. Because the value of this reward is based only on the present, it may sometimes result in an incorrect policy for the future. This is why the value function, which assesses the impact of the actor's actions over a long period, was included in the RL. In fact, even though the reward may initially drop for actions, the primary goal of RL is to maximize this function. In other words, the agent will sacrifice immediate rewards to gain future rewards. The estimation of this value function is the crucial element. If the system has a model, it may be utilized to simulate its operation and calculate the value function that should be maximized. Model-free techniques do not employ a model, while model-based approaches require a system model. The purpose of the critic NN is to estimate the value function. To update the critic network, the mean squared error between the estimated value and the actual observed value is minimized. This is performed by computing the gradient of the loss function with respect to the parameters of the critic network and adjusting the network's parameters.

On the other hand, the actor network is responsible for generating actions depending on the current state or observation of the system, and its gradient descent algorithm parameter will be updated if the action results in lesser rewards. The learning algorithm can be based on value iteration or policy iteration. In both ways, the learning algorithm updates the function to approach the optimum solution [48].

Some of the on-policy DRL methods, such as Proximal policy optimization (PPO) [49], Asynchronous advantage actor-critical (A3C) [50], and Trust region policy optimization (TRPO) [51], all avoid using past experiences. Therefore, learning directly from a real robot or adapting the controller to new conditions presents several challenges due to the considerable sample complexity. On the other hand, off-policy algorithms use prior knowledge to reduce sample complexity significantly.

Well-known examples are the Twin-delayed deep deterministic policy gradient (TD3) [52] and the Deep deterministic policy gradient (DDPG) [53]. Deep reinforcement learning and policy gradient approaches are combined in DDPG. It employs a deterministic strategy, meaning it outputs the action to be done directly rather than sampling from a probability distribution. This provides for more steady and predictable behaviour [53]. TD3 is a DDPG algorithm extension that overcomes some of its drawbacks. It employs a twin network design, with two distinct critics trained to estimate the value function of a given state-action combination. This contributes to a lower overestimation bias in value function estimations [52].

In this work, we investigate training RL methods, including TD3 and DDPG, in the simulation and then transfer them to the real robot to evaluate their efficiency in the experiment. Then, the joint stiffness will be changed to check the algorithms' reliability. Also, the algorithm learning procedure will be online; therefore, it will learn from the robot's experience and adapt to the new situation to track the input efficiently. Furthermore, as the most common linear controller, the PID controller will be implemented in simulation and on the robot to compare the RL technique with the classic controller.

2.3. Simulation

The state-space model of the flexible joint manipulator has been developed in the last section (Eq. 9). This formula simulates the model in Matlab Simulink R2020a with state space blocks. The Actor-critic agent controller is developed to control the system to track a step function. The RL agent's output is the DC motor's input voltage, which is bonded in the range of -10 to 10. In addition, the command is given to the critic to be able to modify the actor's choice for the following action.

At every timestep, we calculate the reward function based on two components:

$$R = R1 + R2 \tag{10}$$

The first component of the formula, Eq. 10, must be written to minimize the difference between the reference signal and the model response. The other component is calculating the amount of beam fluctuation to avoid providing vibration that is not desired. In RL learning, the agent tries to maximize the reward function, in contrast to LQR learning, that minimizing the cost function; Therefore, the amount of both components must be written in a way that maximizing the R function leads to minimizing the tracking error and the vibration of the joint. After trying different functions, the following reward function (Eq. 11) provides a better learning procedure regarding the number of experiments. This formula is the general form of the reward function, and the variables a_i b_i and c_i (positive real numbers) must be tuned. If the motor angle, θ, deviates from the reference angle, the first term will converge to -c_1 and when θ is equal to the reference signal, it will be equal to its maximum, a_1-c_1. Also, the second term follows the same scenario.

$$R = \left(a_1 e^{-b_1\left(\theta_{ref}-\theta\right)^2} - c_1\right) + \left(a_2 e^{-b2\alpha^2} - c_2\right) \; a_i, b_i, c_i \; \in \mathscr{R}^+ \tag{11}$$

The RL algorithm is applied to the real robot. Therefore, there must be limits on the applied force and the range of the system's variables, including θ, α and velocities, to ensure that the robot will not be damaged. Additionally, the agent's output can fluctuate at a high frequency, which can impair the system driver. Thus, a low-pass filter is applied to the action signal.

Using MATLAB/SIMULINK, the TD3, DDPG, and PID are numerically simulated, and a standard laptop is used to carry out the training with Intel® Core™ i5-10310 U Processor and 16GB RAM (without parallel processing). These controllers are applied to the state-space model to track a step function with a step time of 1s and an amplitude of 1 rad. The variables of the model are collected from the Quanser rotary flexible joint module datasheet. The motor and beam angle's initial values and velocities are set to zero. The PID controller is designed for the flexible joint with 1.3 N/m stiffness. The DDPG and TD3 are trained to control the same system. The training phase is finished after the controller obtains more rewards than the PID reward (the PID reward is calculated with the same reward function). After training the controller, the agent is exploited to control the flexible joint system compared to the traditional controller.

Fig. 6 shows the motor angle response to the step function with the RL agents (DDPG and TD3) and the PID method. The TD3 controller shows a lower rise time compared to DDPG and PID controllers (1.21 s, 1.27 s, and 1.24, respectively). However, the settling time of the PID (1.38 s) is less than that of other controllers (DDPG: 1.44 s, TD3: 1.62 s). Fig. 7 shows the manipulator fluctuation angle, α. Even though the RL controllers' fluctuation lasts longer, their maximum fluctuation is less than the PID-controlled manipulator (PID:0.18 rad, DDPG: 0.10 rad, TD3: 0.08 rad). In addition, the action (motor input voltage) consumed by RL algorithms is less than that of the PID controller. The amount of power that TD3 and DDPG use is proportional to (summation of the squared of the voltage value) 1 and 1.25, respectively. At the same time, the PID consume almost two times more energy than the TD3 controller.

After exploring various controllers' efficiency on the system, the stiffness of the joint is decreased to test the controllers' behaviour in another condition. The responses of the controllers are depicted in Fig. 8. Even though the controllers are tuned to different conditions, they show an acceptable result. However, the response of the TD3 is better compared to the PID controller. The linear controller has an 11.8% overshoot, while the agent has an 8.2%. Moreover, TD3′s rise time is 1.01s, which is better compared to the PID method, which is equal to 1.20s. Furthermore, the boundary of the fluctuation of the manipulator for all the controllers is increased.

The RL controller is able to learn and adapt to different situations. Therefore, the agents are set to retrain. The RL agents are capable of adapting to the new system in a few steps. After retraining, the TD3 is tested to compare with the PID controller. Fig. 9a illustrates the response of the system after applying the step function. According to Fig. 9a, the performance of the DRL controller is enhanced compared to Fig. 8a. the PID controller is not changed. Hence, the result is similar to the previous experiment, but the rise time of the RL agent is improved. Additionally, the oscillation of the link is enhanced because the arm's oscillation decreases faster than in the older version.

The experiment is repeated with sinusoidal signals with an amplitude of one radian and a frequency of pi. Results from the experiment that tracks the sinusoidal input as the desired trajectory are shown in Fig. 10. The figure depicts that the RL controller has less lag compared to the linear controller. Even though the RL response produces a little distorted signal at the output, the PID response amplitude is more affected and decreased.

3. Experimental results and discussion

An experimental setup that includes the following components was exploited to evaluate the controller system based on the RL agent controller. 1- Power Amplifier: Quanser VoltPaq-X1 2- Data Acquisition system 3- Quanser servo motor system equipped with soft joint 4- controller (Matlab/Simulink software) (Fig. 11). The spring is replaced with the 4D printed variable stiffness spring that has been integrated with carbon fibre to provide thermal energy to change the spring's stiffness. The sampling rate in the experiment is 40 samples/s (0.025 s). The incremental encoder sensor included in the Quanser unit is used to determine the rigid arm's position and is able to produce up to 4096 pulses per cycle.

The flexibility of the Quanser unit is provided by two springs that attach the rigid manipulator to the servo motor (Fig. 11). Instead of the constant stiffness spring, a 4D-printed structure is used in this experiment.

The provided feedback signals for the RL agent are the displacement

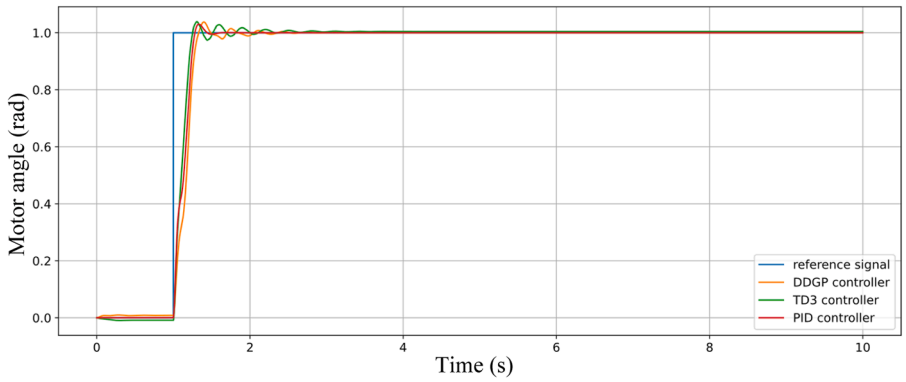

Fig. 6. The angle of the motor's rotor in response to step signal with various controllers.

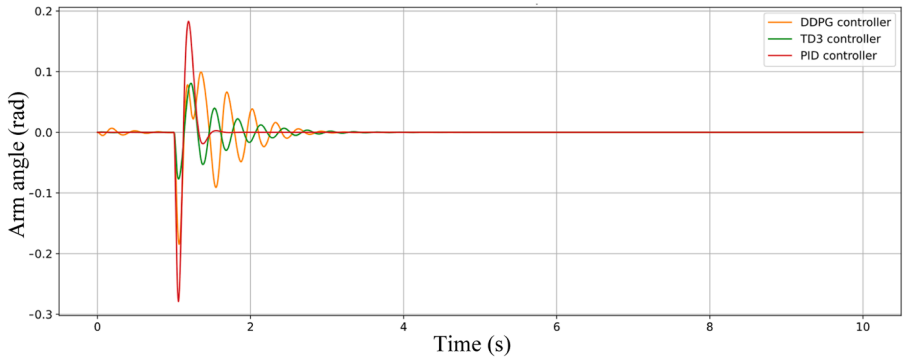

Fig. 7. Rigid arm's oscillation in response to step signal with various controllers.

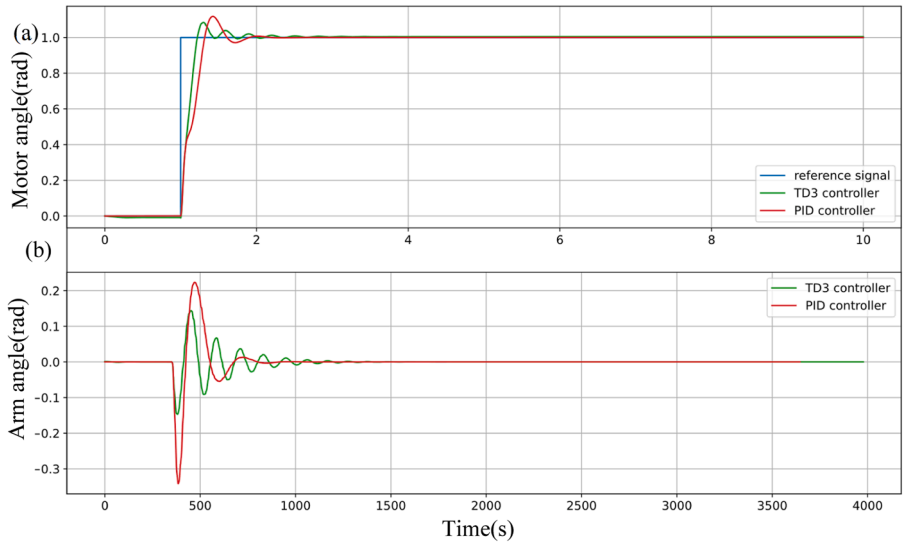

Fig. 8. Effectiveness of various methods after changing the joint's stiffness. a) Motor response to the step function while the stiffness is decreased. b) Manipulator oscillation while the stiffness is decreased.

and velocity of the motor's rotator and manipulator. Actor and critic networks have four layers (input, hidden layers, output) that are able to present a proper controller for the mentioned structure.

Fig. 12 depicts the angle of the motor while the controller aims to follow the input (the step function in the experiment was to rotate the manipulator 45 degrees, or $\pi/4$ rad). The figure shows that the PID controller's rise time is 0.26s, and the TD3 method signal shows the same amount. DDPG controller is slower than the other controllers, and its rise time is 0.33s. Also, other than the TD3 controller, whose overshoot is about 2%, overshoot for other controllers is negligible.

The fluctuation of the manipulator is shown in Fig. 13. Similar to the simulation, it depicts that the PID method oscillation is damped faster than the RL methods. Nonetheless, the RL methods' joint fluctuation magnitude, especially the DDPG controller, is smaller than the result of the linear controller. After evaluating controllers in the constant state, the test is conducted again while voltage is applied to the carbon fibre to change the spring stiffness to evaluate the performance of controllers in different circumstances. (Fig. 14). After warming the 4D-printed spring,

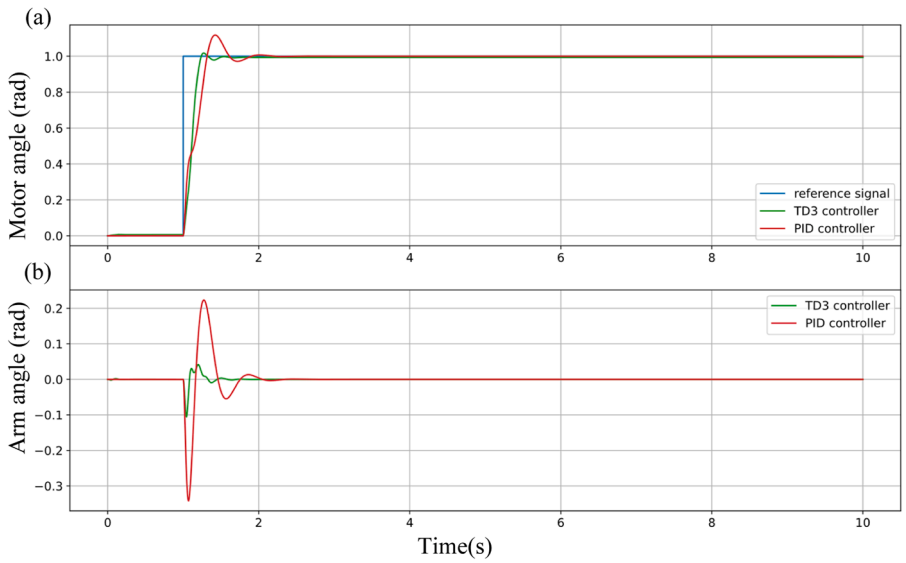

Fig. 9. Effectiveness of PID and TD3 controller after retraining the RL agent to the stiffness of the joint. a) Motor response to the step function. b) Manipulator oscillation.

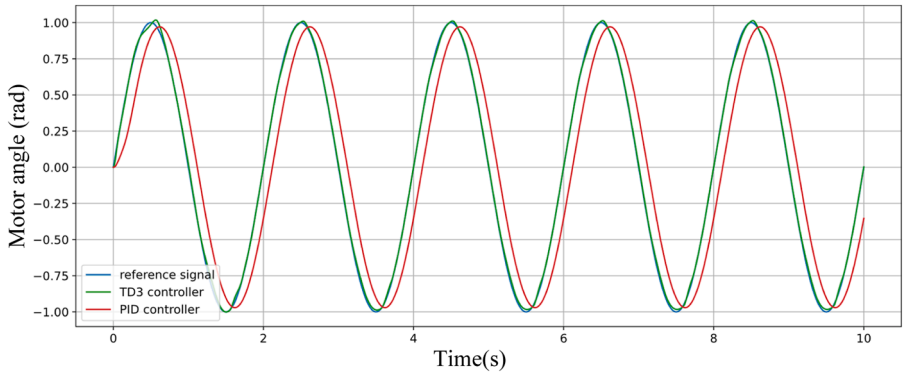

Fig. 10. Motor's angle while the system is activated with a sinusoidal signal under PID and TD3 controller.

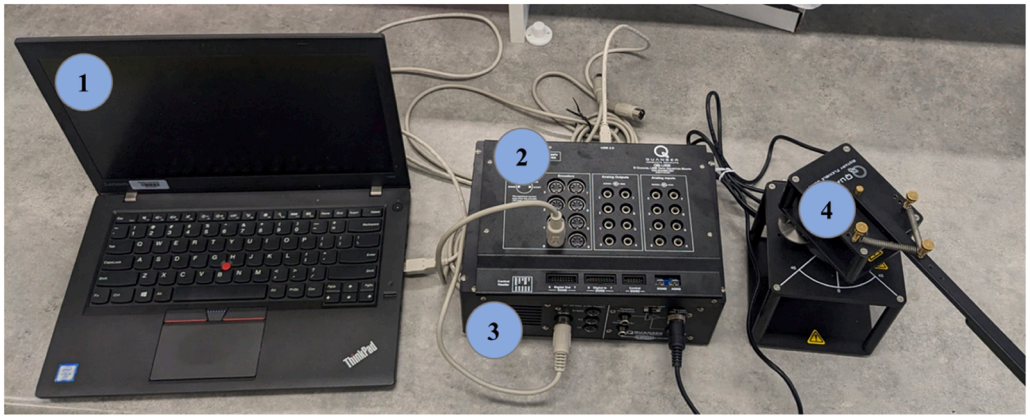

Fig. 11. The setup of experiments 1- Computer (controller algorithm) 2- Data (8 channel) Acquisition system 3- The Amplifier 4- Soft joint cubic Quanser system.

it is compressed and cooled down to maintain the new stiffness. Throughout this test, the TD3 controller is trained online and adjusts its parameters to control the system properly.

Fig. 15 shows that TD3′s parameters match the system and can control the system better than the PID. The TD3 algorithm is faster, and its rise time is less than that of the PID controller (0.27 and 0.33 s, respectively). Furthermore, the overshoot of the PID controller is about three times bigger than the RL agent (5.6 % and 2%, respectively). The stiffness of the joint is less than its initial condition, which leads to a higher fluctuation of the joint controlled by the PID controller. However, it still fades away faster than the oscillation associated with the RL controller. Once more, the magnitude of the fluctuation of the link is less

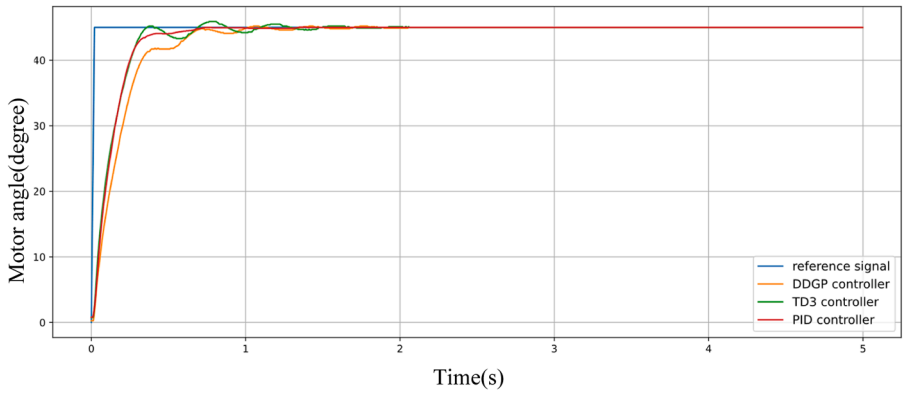

Fig. 12. The angle of the motor's rotor in response to step signal with various controllers in the experimental test.

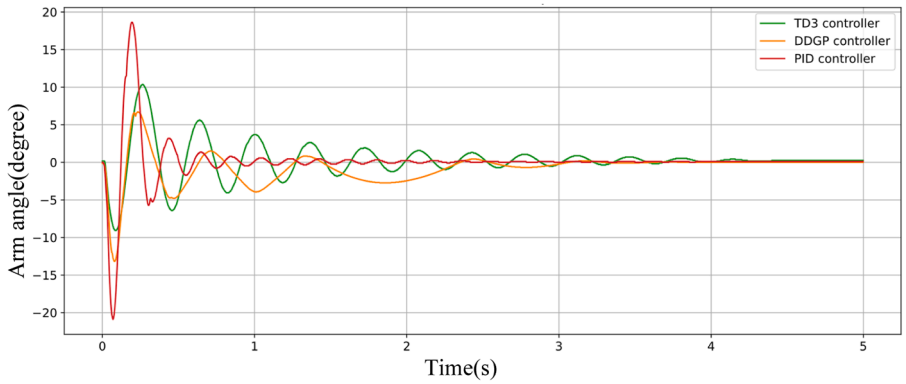

Fig. 13. Rigid arm's oscillation in response to step signal with various controllers in the experimental test.

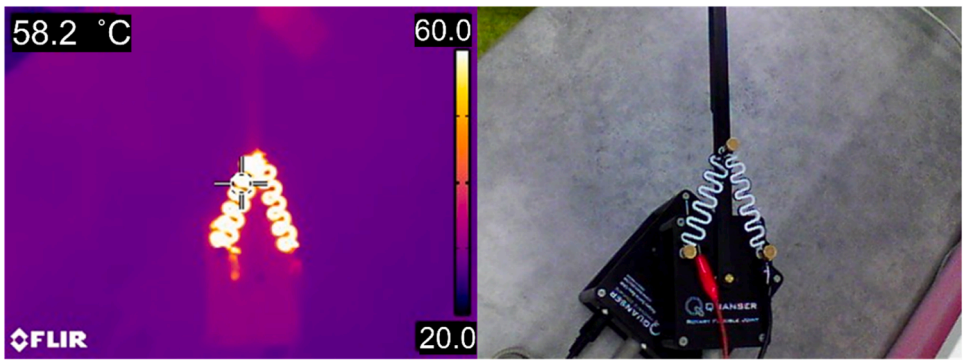

Fig. 14. a) The thermal camera snapshot and (b) the experimental setup while the stimulus signal is activated to change the stiffness of the joint.

when the RL controller is operating the system. In the experiment, the power consumption was derived from the voltage applied to the motor. Based on the mentioned signal, the power usage of the PID method is 1.73 times more compared to the TD3 controller.

According to the experiments, the 4D-printed spring provides variable stiffness for our setup without disassembling the arrangement and installing new elements. Regardless of its advantages, some aspects could be addressed in future works for further development. Integrating the carbon fibre and designing a new geometrical structure to change the stiffness without human interference is the next step of the 4D-printing process. In addition, decreasing the glass temperature of the material could be another challenge to overcome for including the variable stiffness structure in wearable devices. Better performance was obtained by developing the DRL controller to control the flexible joint. The system could adapt to different parameters without labour intervention. Furthermore, the controller decreases power consumption significantly, which is a considerable advantage for utilizing it in portable and lightweight devices.

The outcome of combining the time variable system with the adaptive and robust controller can be used for a variety of purposes. Providing a variable stiffness mechanism is crucial in most wearable devices, including assistant devices, rehabilitation devices, and prostheses. The subject that is wearing the device changes his/her affected organ. Therefore, the device must be time-variant too. Additionally, the power consumption of wearable devices is another challenge [54]. With the newly developed nonlinear controller proposed in this paper, the device could last longer with the same amount of connected battery compared to a traditional controller.

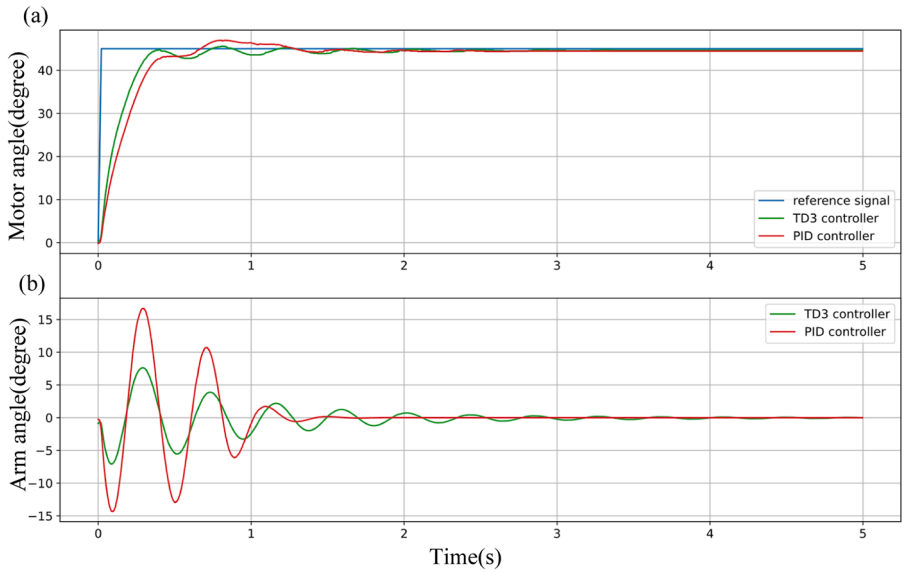

Fig. 15. The system's response after changing the stiffness of the joint under linear and retrained nonlinear controller. a) Motor angle. b) Joint fluctuation.

Fig. 16 depicts an application of the variable stiffness 4D-printed structure in human limb prostheses. The structure is placed on the joints of human limbs prosthesis, while the rigid arm can be activated by servo motors and controlled by DRL agents. The variability of the stiffness makes the device more friendly to the user. Its variable stiffness feature can be altered in favour of the user. For example, if the user tries to lift heavy objects, the joint stiffness can be increased to hold the heavy object easily. On the other side, the joint can have low stiffness. Therefore, the user can reach a broader range of motion with less required force. Also, the device can operate more efficiently in various situations simultaneously. For instance, the stiffness of the arm's joints can change related to the tool the subject is handling, or the ankle and knee stiffness can change depending on the walking pace. The weight of wearable devices is a severe drawback [54]. By decreasing the device's power consumption, the battery's mass can be decreased, resulting in a lightweight wearable device. Furthermore, this design can be integrated into the astronauts' bulky suit, which hinders them in their daily activities, to support their activity muscles. The inner temperature of the suit remains low for the convenience of the user, and the temperature of the variable stiffness joint can be controlled precisely to obtain desired stiffness.

Soft robots still need to be more developed to be utilized as ubiquitous technology. In the coming years, exploration will aid in revealing new potential in soft robotics fields using 4D printing, which is seen as a revolutionary technology. Because of their high degrees of freedom, soft actuators are more complex than rigid actuators. Therefore, new controlling paradigms should be explored alongside the soft robot's development to bring the soft robots into various applications.

4. Conclusion

This article discusses one of the most common techniques for autonomous controls using machine learning technology to control a 4D-printed variable stiffness structure in soft robotics. The following are the primary objectives of this paper:

- The variable stiffness flexible joints are introduced by equipping rigid robots with 4D-printed structures.
- The variable sizes of the variable stiffness structure (PLA S-shape springs), integrated with carbon fibre (to provide the electro-thermal stimulus signal), are additively manufactured.
- The efficiency of the structure's recovery and the required time to recover to its initial shape were explored. Additionally, its power consumption is evaluated.

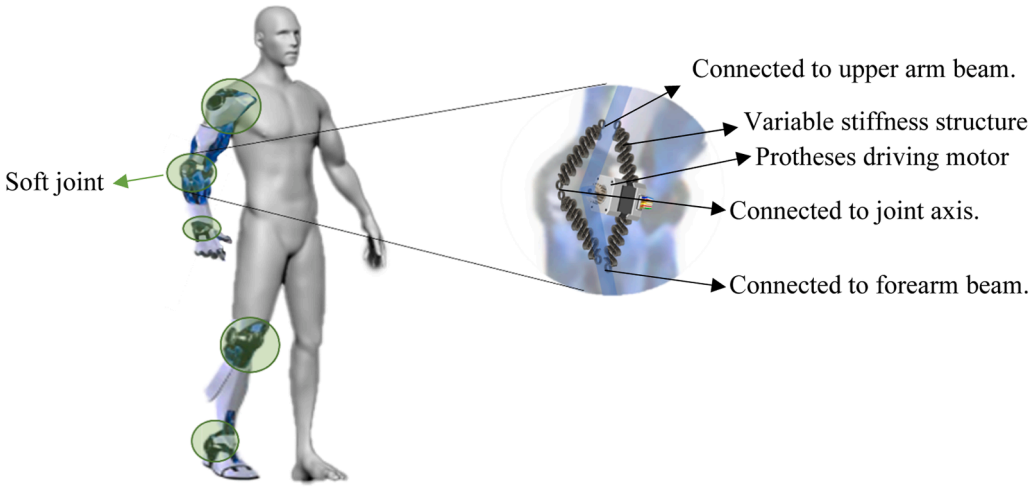

Fig. 16. The 4D-printed variable stiffness joint is controlled by the DRL controller in the assistant/rehabilitation wearable device on the elbow and knee.

- The reinforcement learning technique is utilized to control the system based on a reward function optimization principle.
- Set up the experimental system for evaluating the controller of a stiff arm with a variable stiffness soft joint. The experiment system is a standard robot manipulator that includes a rigid arm powered by a motor (the controller provides input voltage).
- The RL controller is trained to control the system based on the reward function in MATLAB Simulink.
- The controller is evaluated, and the simulation and experiment results show that it is more reliable and can learn to act in various conditions online.
- The efficiency of variable stiffness 4D-printed soft joints, along with their significantly reduced power consumption using reinforcement learning controller, highlights their significant advantage for sustainable and circular robot manufacturing.

In future work, the variable stiffness structures can be exploited in human-interactive applications such as wearable or rehabilitation devices. Advanced and creative future applications can get a good result with a low energy expense by linking soft and hard actuators with adaptive and reliable controlling algorithms like DRL methods.

CRediT authorship contribution statement

Moslem Mohammadi: Software, Methodology, Investigation, Writing – original draft, Validation. **Abbas Z. Kouzani:** Supervision, Methodology, Validation, Project administration. **Mahdi Bodaghi:** Supervision, Methodology, Investigation, Project administration, Writing – review & editing. **John Long:** Supervision, Writing – review & editing. **Sui Yang Khoo:** Supervision, Writing – review & editing. **Yong Xiang:** Supervision, Funding acquisition, Methodology, Writing – review & editing. **Ali Zolfagharian:** Conceptualization, Funding acquisition, Formal analysis, Investigation, Supervision, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research will be supported by ARC under the Discovery Early Career Award (project number DE240100960), which is scheduled to commence in 2024 and will be funded by the Australian Government.

References

[1] S.K. Sood, K.S. Rawat, G. Sharma, 3-D printing technologies from infancy to recent times: a scientometric review, IEEE Trans. Eng. Manage. (2022).
[2] A.J. Arockiam, K. Subramanian, R. Padmanabhan, R. Selvaraj, D.K. Bagal, S. Rajesh, A review on PLA with different fillers used as a filament in 3D printing, Mater. Today: Proc. (2021).
[3] J. Choi, O.-C. Kwon, W. Jo, H.J. Lee, M.-W. Moon, 4D printing technology: a review, 3D Print. Add. Manufact. 2 (4) (2015) 159–167.
[4] X. Sun, et al., Machine learning-evolutionary algorithm enabled design for 4D-printed active composite structures, Adv. Funct. Mater. 32 (10) (2022), 2109805.
[5] S.R. Kumar, S. Sridhar, R. Venkatraman, M. Venkatesan, Polymer additive manufacturing of ASA structure: influence of printing parameters on mechanical properties, Mater. Today: Proc. 39 (2021) 1316–1319.
[6] X. Huang, M. Panahi-Sarmad, K. Dong, R. Li, T. Chen, X. Xiao, Tracing evolutions in electro-activated shape memory polymer composites with 4D printing strategies: a systematic review, Composit., Part A 147 (2021), 106444.
[7] A. Sharma, A. Rai, Fused deposition modelling (FDM) based 3D & 4D printing: a state of art review, Mater. Today: Proc. (2022).
[8] S. Mallakpour, F. Tabesh, C.M. Hussain, 3D and 4D printing: From innovation to evolution, Adv. Colloid Interface Sci. 294 (2021), 102482.
[9] F. Momeni, X. Liu, J. Ni, A review of 4D printing, Mater. Des. 122 (2017) 42–79.
[10] S.Y. Hann, H. Cui, M. Nowicki, L.G. Zhang, 4D printing soft robotics for biomedical applications, Add. Manufact. 36 (2020), 101567.
[11] A. Zolfagharian, A. Kaynak, A. Kouzani, Closed-loop 4D-printed soft robots, Mater. Des. 188 (2020), 108411.
[12] M.N.I. Shiblee, K. Ahmed, M. Kawakami, H. Furukawa, 4D printing of shape-memory hydrogels for soft-robotic functions, Adv. Mater. Technolog. 4 (8) (2019), 1900071.
[13] M. López-Valdeolivas, D. Liu, D.J. Broer, C. Sánchez-Somolinos, 4D printed actuators with soft-robotic functions, Macromol. Rapid Commun. 39 (5) (2018), 1700710.
[14] C. de Marco, S. Pané, B.J. Nelson, 4D printing and robotics, Sci. Robot. 3 (18) (2018) eaau0449.
[15] S.Y. Hann, et al., Recent advances in 3D printing: vascular network for tissue and organ regeneration, Translat. Res. 211 (2019) 46–63.
[16] S.A. Ajwad, J. Iqbal, Emerging trends in robotics–a review from applications perspective, in: International Conference on Engineering and Emerging Technologies (ICEET), 2015.
[17] S. Ajwad, M. Ullah, B. Khelifa, J. Iqbal, A comprehensive state-of-the-art on control of industrial articulated robots, J. Balkan Tribolog. Assoc. 20 (4) (2014) 499–521.
[18] J. Zhou, et al., Adaptive variable stiffness particle phalange for robust and durable robotic grasping, Soft Robot. 7 (6) (2020) 743–757.
[19] M. Brancadoro, M. Manti, F. Grani, S. Tognarelli, A. Menciassi, M. Cianchetti, Toward a variable stiffness surgical manipulator based on fiber jamming transition, Front. Robot. AI 6 (2019) 12.
[20] M. Runciman, J. Avery, M. Zhao, A. Darzi, G.P. Mylonas, Deployable, variable stiffness, cable driven robot for minimally invasive surgery, Front. Robot. AI 6 (2020) 141.
[21] H. Sugiura, et al., Characterization of the variable stiffness actuator fabricated of SMA/SMP and MWCNT/IL: PDMS strain-sensitive heater electrode, IEEE Robot. Automat. Lett. (2022).
[22] H. Sugiura, S. Amaya, F. Arai, Miniaturized Sma/Smp variable stiffness actuator driven by the stretchable Mwcnt/Il/Pdms heater electrode for micromanipulation robots, in: 2022 IEEE 35th International Conference on Micro Electro Mechanical Systems Conference (MEMS), IEEE, 2022, pp. 5–8.
[23] W. Wang, C.Y. Yu, P.A. Abrego Serrano, S.-H. Ahn, Shape memory alloy-based soft finger with changeable bending length using targeted variable stiffness, Soft Robot. 7 (3) (2020) 283–291.
[24] Q. Ji, X.V. Wang, L. Wang, L. Feng, Customized protective visors enabled by closed loop controlled 4D printing, Sci. Rep. 12 (1) (2022) 1–12.
[25] C. Tawk, G. Alici, 4D-printed pneumatic soft actuators modeling, fabrication, and control. Smart Materials in Additive Manufacturing, Volume 2: 4D Printing Mechanics, Modeling, and Advanced Engineering Applications, Elsevier, 2022, pp. 103–140.
[26] I.H. Akyuz, E. Yolacan, H.M. Ertunc, Z. Bingul, PID and state feedback control of a single-link flexible joint robot manipulator, in: 2011 IEEE International Conference on Mechatronics, IEEE, 2011, pp. 409–414.
[27] M. Ahmad, M. Suid, M. Ramli, M. Zawawi, R.R. Ismail, PD fuzzy logic with non-collocated PID approach for vibration control of flexible joint manipulator, in: 2010 6th International Colloquium on Signal Processing & its Applications, IEEE, 2010, pp. 1–5.
[28] M.A. Ahmad, Vibration and input tracking control of flexible manipulator using LQR with non-collocated PID controller, in: 2008 Second UKSIM European Symposium on Computer Modeling and Simulation, IEEE, 2008, pp. 40–45.
[29] C. Chitu, J. Lackner, M. Horn, H. Waser, M. Kohlböck, A robust and optimal LQR controller design for Electric Power Steering system, in: Proceedings of the Joint INDS'11 & ISTET'11, IEEE, 2011, pp. 1–5.
[30] P. Sarkhel, N. Banerjee, N.B. Hui, Fuzzy logic-based tuning of PID controller to control flexible manipulators, SN Appl. Sci. 2 (6) (2020) 1–11.
[31] H. Ahmadian, H.A. Talebi, I. Sharifi, Adaptive controller design for single-link flexible joint manipulator with fuzzy-PID filter, in: 2020 28th Iranian Conference on Electrical Engineering (ICEE), IEEE, 2020, pp. 1–6.
[32] J. Ju, Y. Zhao, C. Zhang, Y. Liu, Vibration suppression of a flexible-joint robot based on parameter identification and fuzzy PID control, Algorithms 11 (11) (2018) 189.
[33] S.M. Ahmadi, M.M. Fateh, Composite direct adaptive Taylor series–fuzzy controller for the robust asymptotic tracking control of flexible-joint robots, Trans. Inst. Meas. Control 41 (14) (2019) 4023–4034.
[34] S.S. Ge, I. Postlethwaite, Adaptive neural network controller design for flexible joint robots using singular perturbation technique, Trans. Inst. Meas. Control 17 (3) (1995) 120–131.
[35] W. He, Z. Yan, Y. Sun, Y. Ou, C. Sun, Neural-learning-based control for a constrained robotic manipulator with flexible joints, IEEE Transact. Neur. Netw. Learn. Syst. 29 (12) (2018) 5993–6003.
[36] B. Subudhi, A.S. Morris, Soft computing methods applied to the control of a flexible robot manipulator, Appl. Soft Comput. 9 (1) (2009) 149–158.
[37] D. Pavlichenko, S. Behnke. Real-Robot Deep Reinforcement Learning: Improving Trajectory Tracking of Flexible-Joint Manipulator with Reference Correction, IEEE, 2022, pp. 2671–2677.
[38] D. Sendrescu, G. Bujgoi, D. Chintescu, Control of a rotary flexible joint experiment based on reinforcement learning, in: 2020 21th International Carpathian Control Conference (ICCC), IEEE, 2020, pp. 1–5.

[39] T. Long, et al., A vibration control method for hybrid-structured flexible manipulator based on sliding mode control and reinforcement learning, IEEE Transact. Neur. Netw. Learn. Syst. 32 (2) (2020) 841–852.
[40] S.K. Pradhan, B. Subudhi, Real-time adaptive control of a flexible manipulator using reinforcement learning, IEEE Trans. Autom. Sci. Eng. 9 (2) (2012) 237–249.
[41] A. Al-Tamimi, F.L. Lewis, M. Abu-Khalaf, Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof, IEEE Transact. Syst. Man Cybernet. Part B (Cybernet.) 38 (4) (2008) 943–949.
[42] J. Li, K. Ma, Z. Wu, Tracking control via switching and learning for a class of uncertain flexible joint robots with variable stiffness actuators, Neurocomputing 469 (2022) 130–137.
[43] J. Guo, Robust tracking control of variable stiffness joint based on feedback linearization and disturbance observer with estimation error compensation, IEEE Access 8 (2020) 173732–173754.
[44] T. Sun, Y. Chen, T. Han, C. Jiao, B. Lian, Y. Song, A soft gripper with variable stiffness inspired by pangolin scales, toothed pneumatic actuator and autonomous controller, Rob. Comput. Integr. Manuf. 61 (2020), 101848.
[45] W. Bai, Q. Zhou, T. Li, H. Li, Adaptive reinforcement learning neural network control for uncertain nonlinear system with input saturation, IEEE Transact. Cybernet. 50 (8) (2019) 3433–3443.
[46] H. Hua, Y. Fang, A novel reinforcement learning-based robust control strategy for a quadrotor, IEEE Trans. Ind. Electron. 70 (3) (2022) 2812–2821.
[47] Y. Xin, Z.-C. Qin, J.-Q. Sun, Robust experimental study of data-driven optimal control for an underactuated rotary flexible joint, Int. J. Control Autom. Syst. 18 (5) (2020) 1202–1214.
[48] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath, Deep reinforcement learning: a brief survey, IEEE Signal Process. Mag. 34 (6) (2017) 26–38.
[49] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347, 2017.
[50] V. Mnih, et al., Asynchronous methods for deep reinforcement learning, in: International Conference on Machine Learning, PMLR, 2016, pp. 1928–1937.
[51] J. Schulman, S. Levine, P. Abbeel, M. Jordan, P. Moritz, Trust region policy optimization, in: International Conference on Machine Learning, PMLR, 2015, pp. 1889–1897.
[52] S. Fujimoto, H. Hoof, D. Meger, Addressing function approximation error in actor-critic methods, in: International Conference on Machine Learning, PMLR, 2018, pp. 1587–1596.
[53] T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning." arXiv preprint arXiv:1509.02971, 2015.
[54] M. Mohammadi, A. Zolfagharian, M. Bodaghi, Y. Xiang, A.Z. Kouzani, 4D printing of soft orthoses for tremor suppression, Bio-Des. Manufact. (2022) 1–22.