


RESEARCH ARTICLE OPEN ACCESS

A Five-Dimensional Three-Layer Digital Twin to Train a Reinforcement Learning Agent for Interaction Control of a Robotic Exoskeleton in Adolescent Idiopathic Scoliosis Rehabilitation

Farhad Farhadiyadkuri  | Xuping Zhang

Department of Mechanical and Production Engineering, Aarhus University, Aarhus, Denmark

Correspondence: Xuping Zhang (xuzh@mpe.au.dk)

Received: 24 November 2024 | **Revised:** 5 February 2025 | **Accepted:** 19 March 2025

Funding: The authors received no specific funding for this work.

Keywords: adolescent idiopathic scoliosis | deep reinforcement learning | digital twin | reinforcement learning-based impedance control | robotic rehabilitation

ABSTRACT

Adolescent idiopathic scoliosis (AIS) is a sideways curvature of the spinal column combined with a vertebral rotation that usually occurs in adolescents without any known causes. Bracing, the most common conservative treatment of AIS, has not fully exploited the benefits of the active control approaches powered by artificial intelligence (AI), although AI has entered a wide range of applications. The correction forces exerted by the brace are controlled passively by regulating the tightness of the brace's strap. Besides, training the learning-based control methods using a virtual model is of high importance in the AIS brace treatment, since training using trial and error on human subjects may result in unexpected pressure and injuries on the patient's torso. However, digital twin (DT) modeling, an emerging technology, has not been implemented into the AIS brace treatment yet. In this paper, reinforcement learning-based position-based impedance control (RLPIC) is proposed to enable a robotic brace to learn the desired physical interaction between the robotic brace and the human torso. A five-dimensional (5D) three-layer DT is also developed to be used for training the RLPIC in a simulated environment. The 5D three-layer DT consists of a physical system, a three-layer digital model of the physical system, including the robotic brace, human torso, and the physical human-robot interaction (HRI), a bidirectional connection between them, and an optimization dimension. A neural network-based regression model is also proposed to estimate the unknown parameters of the digital model. Numerical simulations and real-time experiments are performed to validate the 5D three-layer DT model. The proposed RLPIC trained using the 5D three-layer DT is verified using numerical simulations in terms of position tracking, velocity tracking, and HRI control. It is concluded that the proposed learning-based interaction control approach can improve the HRI control by learning the desired interaction in a simulated environment.

Abbreviations: 3D, three dimensional; 5D, five dimensional; 5D three-layer DT, five-dimensional three-layer digital twin; AI, artificial intelligence; AIS, adolescent idiopathic scoliosis; DOF, degrees of freedom; DS, digital system; DT, digital twin; HRI, human-robot interaction; IC, impedance control; IDC, inverse dynamic control; IIOT, industrial internet of things; MB, multi body; MB-FE, multi body-finite element; MRAIC, model reference adaptive impedance control; NN, neural network; NPIC, novel position-based impedance control; P2V, physical to virtual; PIC, position-based impedance control; PPO, proximal policy optimization; PS, physical system; RL, reinforcement learning; RLPIC, reinforcement learning-based position-based impedance control; RMSE, root mean square error; SGP, Stewart-Gough platform; V2P, virtual to physical; VPIC, variable position-based impedance control.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *International Journal of Mechanical System Dynamics* published by John Wiley & Sons Australia, Ltd on behalf of Nanjing University of Science and Technology.

1 | Introduction

Adolescent idiopathic scoliosis (AIS) is an abnormal curvature with three-dimensional (3D) deformation of the human spine that occurs in adolescents without any known reason. Bracing, as the most common nonsurgical treatment, is prescribed to halt or mitigate curvature progression and avoid surgery. Three common types of braces, including full-time rigid, night-time rigid, and soft braces, have been developed to date [1–7]. One of the most important challenges in the AIS brace treatment is to fit the brace on the patient's trunk and adjust the in-brace correction pressure, which has a direct impact on the brace's effectiveness. The brace is currently adjusted passively by regulating the tightness of the brace's strap using follow-up and X-ray checks with trial and error on the patient. Although only a few robotic braces that use typical motion control and force control strategies, for example, PID, have been developed [8–11], the advantages of active closed-loop control methods have not been fully exploited in the AIS brace treatment yet. Impedance control (IC) has been widely implemented in robotic rehabilitation [12–16], but IC has not been implemented into the AIS brace treatment yet. The IC is a better solution than typical motion or force control for human–robot interaction (HRI) control, where the robot has direct contact with the human subject as the environment, because it regulates the dynamic relationship between motion and interaction force to provide smooth physical HRI in contrast to typical motion control and force control strategies that control only motion and force independently. In our previous works [17, 18], a robotic brace was developed, and novel IC strategies were proposed to control the biomechanical interaction in the AIS brace treatment. However, the major issue with the typical IC is how to define the desired impedance parameters. What is the desired impedance model for the AIS brace treatment?

On the other hand, reinforcement learning (RL) is a widely studied topic in learning and control communities. The concept of RL policy, that is, taking the states of the environment (observations) and generating actions to perform a task in an optimal manner, is similar to the role of a controller in a control system. In RL, the first step is to determine the environment, which can be a simulated environment or the physical one. The environment in RL includes the plant, for example, the robot, its surrounding workplace, for example, the human body, and any other elements outside the RL agent. The observations for the RL agent are defined using sensory measurements. The action variables can be the control input, for example, the actuator torques. The action can also be any other variables, for example, the reference position that is used as the desired input to the low-level controller. The next step is to define the reward function. Since the RL algorithm is trying to take the best actions for performing the task in an optimal manner by maximizing the reward function, the reward function should be defined such that it will be increased if the task is performing well, and the desired goal is approaching. After that, a policy should be chosen to enable the agent to map the observations into the optimal actions by maximizing the reward function. Finally, a training algorithm is defined to train the policy and find the optimal solutions. Sutton and Barto [19] presented a good introduction to RL, and the specific RL problems in robotics were described [20–22]. Chatzilygeroudis et al. [23] presented recent advancements in RL,

with a focus on data-efficient algorithms, and deep learning advanced solutions for RL problems were reviewed by Arulkumaran et al. [24]. In brief, RL is an interesting topic for control engineers because a typical control problem is mapped to an RL problem without any need for concerns about designing the control law and tuning the controller gains. Since there is no research work concerning learning-based active control in the AIS brace treatment, the RL-based IC is an open research topic for interaction control of the AIS brace treatment so that the robotic brace can learn the desired impedance parameters required to provide the desired physical HRI for the AIS brace treatment. The RL agent learns from the observations obtained from trial and error by applying good or bad actions to the environment. However, in the AIS brace treatment, training the RL agent using trial and error on a human subject may cause serious challenges, e.g., injuries in the human torso and new spinal deformities because of the wrong actions that the RL agent might take. Besides, the long-time process and harmful radiation exposure due to repeated radiographic examinations are the other disadvantages of trial and error with human subjects. Therefore, computational modeling can be adopted to develop a simulated environment for training the RL agent without trial and error on human subjects. However, the output of the computational model might be different from its physical counterpart because of the model uncertainty. Therefore, the significant issue to be addressed is how to improve the computational model and increase the model's reliability to mirror the biomechanical behavior of the AIS brace treatment. On the other hand, digital twin (DT) is an emerging technology that entered our lives in the field of production and engineering, and it is also going to lead to a revolution in healthcare [25]. A DT is a digital replica of a physical system (PS) that is updated using data from the PS and allows simulation and prediction of the state of the system in a virtual environment. The concept of DT was first implemented in the NASA Apollo 13 program in the 1960s [26, 27]. The first documented definition of DT was provided by Michael Grieves in his presentation in the context of product life cycle management in 2003 and later in a white paper [28]. A comprehensive classification of the research articles in the field of DT was presented in a recently published two-part review paper [29, 30]. It is indicated that the DT has been implemented in different application domains, including manufacturing, mechanics, civil, structure, aerospace, energy, and healthcare. Although the review paper [29] shows a leading role of manufacturing in the DT field, healthcare has the least number of published papers in DT. An automatic gait data control system for fully actuated lower limb exoskeleton digital twinning was proposed by improving the integration of DT with the medical rehabilitation field and analyzing the patient's gait data through a simulation experiment [31]. A robot-assisted telerehabilitation system with integrated industrial internet of things (IIOT) and DT was developed for post-stroke patients with upper limb dysfunction [32]. Although some attempts were made to develop DT in robotic rehabilitation, there are no research articles focusing on creating a DT for the AIS brace treatment. Therefore, developing a DT of the AIS brace treatment will overcome the above-mentioned limitations and it can be used for training the RL agent in a virtual environment.

In brief, various DT models, including 3D and 5D, have been proposed in the past decade [28, 29, 33]. Grieves [28] proposed a 3D DT composed of a PS, a digital system (DS), and the

connections between them. A 5D DT that consists of PS, DS, a service component, DT data, and connections was proposed by Tao et al. [33]. A new 5D DT of a lithium-ion cell composed of PS, DS, physical to virtual (P2V), virtual to physical (V2P), and optimization was proposed in the two-part review paper [29, 30] and this is adopted to develop a 5D three-layer DT of the AIS brace treatment in this paper. Besides, a RL-based position-based impedance control (RLPIC) algorithm is proposed for active closed-loop interaction control of the AIS brace treatment. The RL agent is trained using the 5D three-layer DT. The P2V data flow is established by estimating the unknown parameters of the 5D three-layer DT using a Neural Network (NN)-based regressor according to the data set collected from real-time experiments. Using the NN-based regression model, the time needed for manually tuning the contact parameters is reduced. Furthermore, to establish a V2P data flow, the desired impedance parameters for a given desired displacement are predicted using the RLPIC trained by the 5D three-layer DT. The RLPIC can improve the safety and compliance of the patient by learning the desired HRI for AIS treatment in the simulated environment (DT). The developed DT can also potentially improve the brace adjustment process without trial and error on the human subject. Note that, in this paper, the real-time experiments refer to the practical tests that are performed on the PS. Figure 1 presents the main contribution of this paper.

The rest of the paper is organized as follows: The 5D three-layer DT is created in Section 2. The proposed RLPIC algorithm is

formulated in Section 3. Section 4 presents the numerical simulations and real-time experiments. Section 5 concludes the paper.

2 | 5D Three-Layer DT of the AIS Brace Treatment

A DT of a PS mimics the behavior of its physical counterpart in a virtual environment. In this paper, the DT technology is exploited to improve and optimize the brace treatment of AIS by modeling and predicting its biomechanical behavior in advance. First, an overview of our preliminary works [17, 18] on the modeling of the AIS brace treatment is explained in Section 2.1. Second, the 5D three-layer DT model of the AIS brace treatment is presented in Section 2.2. It consists of five dimensions, including the PS, its three-layer digital model, P2V data flow, V2P data flow, and optimization. The physical model consists of the robotic brace developed in our previous work [17], the human torso, and the physical HRI. The three-layer digital model is created by adding the solid body model of the human torso to the multi body (MB) model of the robotic brace and modeling the HRI as a mass–spring–damper system. To develop a P2V data flow, the unknown parameters of the HRI model are identified by training an NN-based regressor using data sets from real-time experiments. The identified values of the unknown parameters are optimized to develop the fifth dimension. The V2P data flow is established by training the RLPIC algorithm proposed in Section 3 using the 5D three-layer DT to predict the desired impedance model required for AIS treatment. The NN-based regression model used for

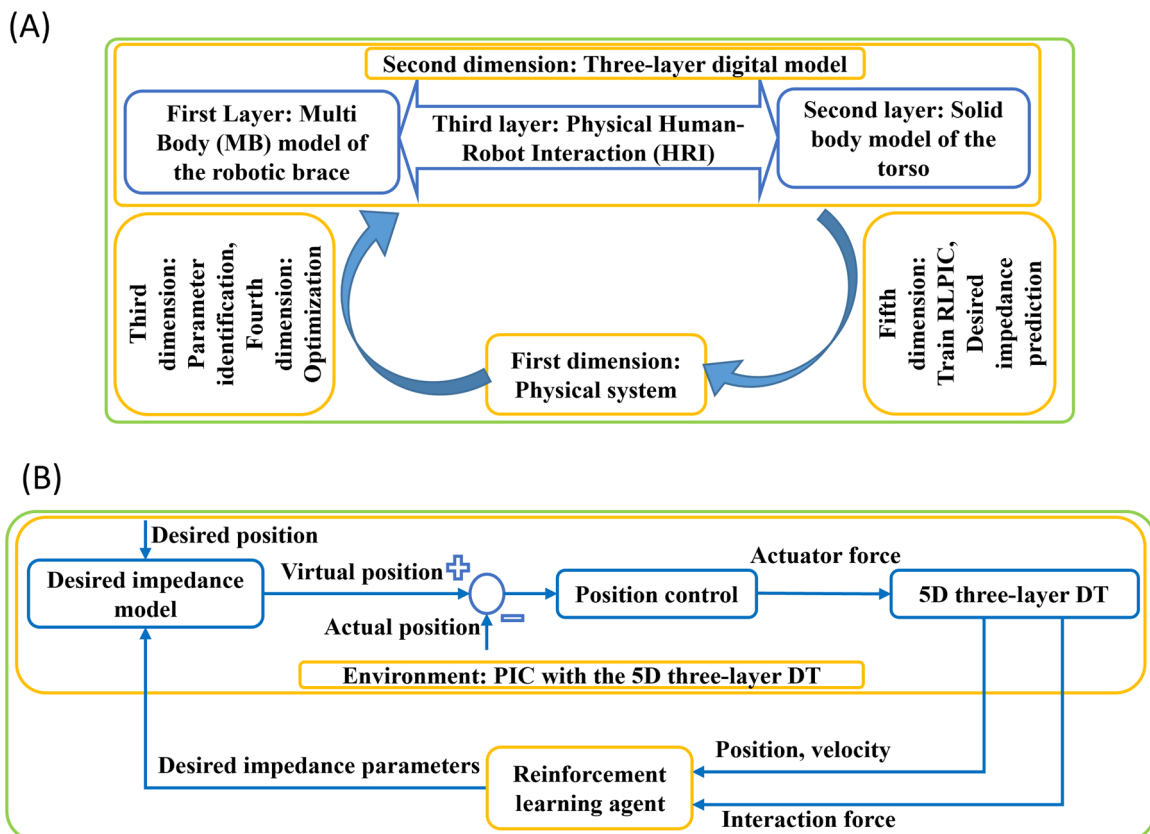


FIGURE 1 | The main contribution of this paper: (A) Five-dimensional (5D) three-layer digital twin (DT) model. (B) Reinforcement learning-based position-based impedance control (RLPIC) for the AIS brace treatment trained using the 5D three-layer DT.

identifying the unknown parameters of the digital model is finally presented in Section 2.3.

2.1 | Preliminaries

An overview of our preliminary works [17, 18] in the field of the AIS brace treatment modeling is presented in this subsection. A robotic brace was developed using a 3D scan of a scoliosis patient [17]. It consists of three Stewart–Gough platforms (SGPs) and 18 degrees of freedom (DOF). The analytical model of each SGP of the robotic brace was independently derived using the Lagrangian formulation [17]. Besides, a multi body-finite element (MB-FE) Simscape model along with an analytical model of the robotic brace were developed [18]. The analytical model of the robotic brace was developed by deriving the dynamics model of three SGPs as a single unit.

2.2 | The DT Model

The 5D three-layer DT model of the AIS brace treatment, shown in Figure 2, is presented. It consists of five dimensions, and each

dimension is highlighted with orange in Figure 2A. The PS, as the first dimension of the 5D three-layer DT model, consists of the robotic brace [17], the patient’s trunk, and the physical HRI between them. The robotic brace has one fixed ring placed around the hip and three moving rings that have direct contact with the ribs connected to vertebrae T4, T7, and T11.

The second dimension is the three-layer digital model of the PS, shown in Figure 2A. Three layers of the digital model include (a) the MB model of the robotic brace, (b) the solid body model of the human torso, and (c) HRI modeling. The MB model of the robotic brace developed based on its SolidWorks model [18] is added to the solid body model of the human torso, and the physical interaction between the moving rings of the MB model and the solid body model of the torso is defined as a virtual mass–spring–damper system using the “Spatial Contact Force” block from Simscape Multibody Library. It models the physical contact between a pair of 3D geometries and uses a built-in penalty method to compute the normal and friction forces. The implementation detail of the three-layer digital model is presented in Section 4.1.

The third, fourth, and fifth dimensions of the 5D three-layer DT model, shown in Figure 2A, are P2V data flow, optimization,

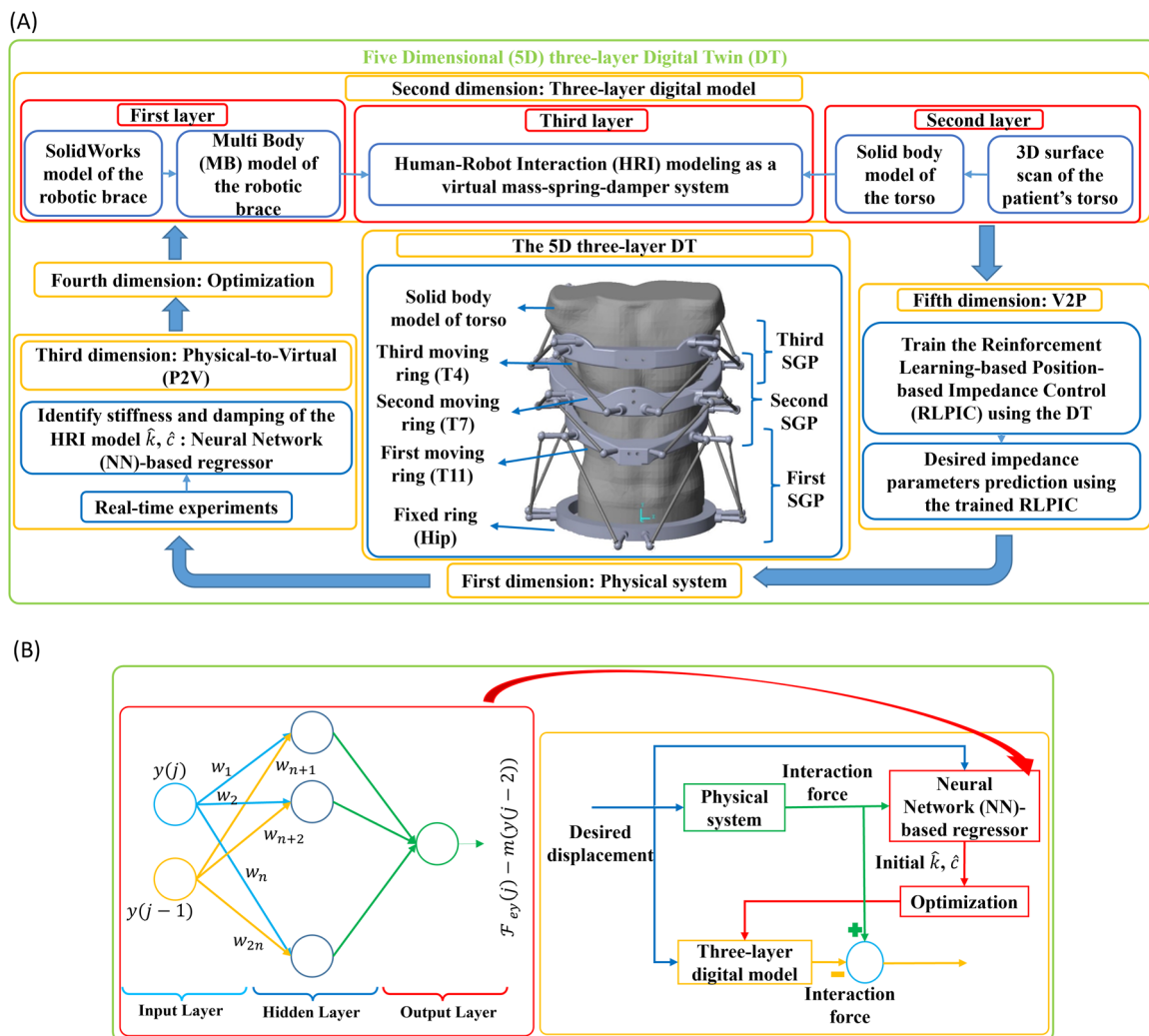


FIGURE 2 | Five-dimensional (5D) three-layer digital twin (DT): (A) Five dimensions and three layers of the DT are shown. (B) The third and fourth dimensions of the DT are presented in detail.

and V2P data flow, respectively. These three dimensions are established to twin the three-layer digital model with its physical counterpart by establishing bidirectional data flow (P2V, V2P) between the PS and the digital model and optimizing the unknown parameters of the digital model. P2V data flow is a connection from the PS to the digital model in which the data from PS are used to update the state of the digital model [29]. To establish the P2V data flow (third dimension), real-time experiments are conducted and the unknown parameters of the HRI model, including stiffness and damping, are estimated using an NN-based regression model and the experimental data. Numerical simulations using the three-layer digital model indicate that the magnitude of the output interaction force varies on changing the unknown parameters of the “Spatial Contact Force” block. The unknown parameters of the “Spatial Contact Force” block must be chosen such that the output interaction force of the three-layer digital model fits the output interaction force of the PS. Therefore, real-time experiments are carried out and the interaction force–motion data are collected. The interaction force–motion data set is used to identify the unknown parameters of the HRI model using the NN-based regression model, which is described in detail in Section 2.3.

In the fourth dimension, the estimated parameters are optimized using the ‘Parameter Estimator’ App from the Simulink Design Optimization Toolbox. The P2V data flow is a connection between the digital model and PS, in which the digital model is used to predict the state of the PS [29]. The RLPIC algorithm is trained using the three-layer digital model. The trained RLPIC is then utilized to predict the desired impedance model required for HRI control of the PS and establish the V2P data flow. In other words, the RLPIC algorithm trained using the digital model acts as a predictor to create data flow from DS to PS. In addition, the trained RLPIC algorithm can be implemented in the PS to provide the desired impedance model for HRI control. The details of the V2P data flow are presented in Sections 3 and 4.2.

2.3 | Parameter Identification

To identify the unknown parameters of the digital model, the NN-based regression model is explained in detail, as shown in Figure 2B. The linear mass–spring–damper model of the HRI is defined as follows:

$$m\ddot{y} + c\dot{y} + ky = \mathcal{F}_{ey}, \quad (1)$$

where m , c , and k denote the torso mass, virtual damping, and virtual stiffness of the physical HRI model, respectively. y and \mathcal{F}_{ey} also represent the displacement of the torso and the interaction force along the y -direction, respectively. Note that the y -direction is defined along the intersection line of the coronal and transverse planes with the positive direction to the right side of the coronal plane seen from the front view. Besides, \dot{y} and \ddot{y} indicate the velocity and acceleration of the torso at the interaction point, respectively. The x -direction is also defined along the intersection line of the sagittal and transverse planes with the positive direction to the front side of the sagittal plane seen from the front view, and the z -axis is perpendicular to both x - and y -axes with the positive direction to the top of the torso. In addition, the origin of the reference frame is located at the

center of the fixed ring. The torso mass m is also assumed to be 40kg according to the torso mass of a typical human subject. Note that the formulation is written for only one DOF and the same procedure can be implemented for the other DOFs as well. The main goal is to identify the virtual stiffness and damping parameters of the HRI model k and c (the stiffness and damping parameters of the “Spatial Contact Force” block) using the interaction force \mathcal{F}_{ey} and torso’s displacement y measurements collected from the real-time experiments. For this purpose, Equation (1) is discretized and rewritten as follows:

$$\begin{aligned} ay(j) + by(j-1) &= \mathcal{F}_{ey} - my(j-2), \quad a = m + c + k * dt, \quad b \\ &= -2m - c, \end{aligned} \quad (2)$$

where j and dt denote the time step and the sampling time, respectively. To identify the unknown parameters k and c , the parameters a and b should be estimated first. The regression method can be used to estimate the parameters a and b using real-time measurements. Since an NN with one hidden layer can act as a regression model, an NN-based regression model is designed to estimate the parameters a and b , as shown in Figure 2B. The output of the NN-based regression model is computed as follows:

$$\begin{aligned} ReLU((w_1 + w_2 + \dots + w_n)y(j) + (w_{n+1} + w_{n+2} \\ + \dots + w_{2n})y(j-1)) \\ = F_{ey} - my(j-2) \end{aligned} \quad (3)$$

The activation function of the NN-based regression model is assumed to be *ReLU*. The size of the hidden layer is considered n . w_1 to w_n and w_{n+1} to w_{2n} denote the weights of the NN-based regression model, respectively. Comparison of Equations (2) and (3) shows that $a = w_1 + w_2 + \dots + w_n$ and $b = w_{n+1} + w_{n+2} + \dots + w_{2n}$. The NN is trained using real-time measurements and the weights, and parameters a and b are obtained. Finally, the unknown parameters k and c are computed by replacing a and b into Equation (2).

The estimated k and c are then considered as an initial guess and they are optimized using the “Parameter Estimator” App from the Simulink Design Optimization Toolbox, such that the interaction force obtained from DT and the PS for a given displacement fit each other and the DT is validated. The details of the third and fourth dimensions are shown on the right-hand side of Figure 2B.

It should be mentioned that the 5D-three-layer DT model is patient-specific, meaning that it can be personalized for any individual patient. Three layers of the DT model can be personalized for each individual patient. The MB model of the robotic brace can be personalized by performing a 3D surface scan of a new scoliosis patient and updating the CAD model of the robotic brace by designing and adding new rings to the inner part of the current moving rings such that the outer part of the new rings fits the surface of the new patient’s torso. Then, the solid body model of the torso can be personalized using the new 3D surface scan of the new patient. Finally, the same procedure as that described in Sections 2.2 and 2.3 is utilized to identify the unknown parameters of the HRI model according

to the data set collected from real-time experiments on the new patient. Hereby, the DT model is personalized for each individual patient.

3 | The Proposed RLPIC Algorithm

First, an overview of our previous works [17, 18] on the HRI control of the AIS brace treatment is presented. Two novel controllers from the IC family, including model reference adaptive impedance control (MRAIC) [17] and novel position-based impedance control (NPIC) [18], were proposed. In MRAIC, an adaptive law was designed to regulate the impedance parameters such that the actual impedance model follows the reference impedance model. In NPIC, the actual impedance parameters are estimated using interaction force–motion measurements and the impedance parameters' errors were then used to regulate the measured interaction force, which is the primary input of the PIC, such that the desired impedance model is achieved.

In this section, an RLPIC algorithm is proposed to improve the performance of the typical PIC in terms of pose tracking, velocity tracking, and HRI control. The schematic diagram of the RLPIC is shown in Figure 3. First, the typical PIC algorithm is described. In PIC, the virtual pose of the robot X_v is computed using the desired impedance model and the measured interaction force. An internal motion control loop is then utilized to provide the virtual pose so that the desired pose is achieved in the presence of the physical HRI. The desired impedance model is considered as follows:

$$M_d \ddot{e} + C_d \dot{e} + K_d e = \mathcal{F}_e, \quad (4)$$

where M_d , C_d , and K_d are 18×18 diagonal matrices and denote the desired mass, damping, and stiffness parameters, respectively. Note that the desired impedance parameters of each DOF are defined independently and the correlations between the impedance parameters of different DOFs are assumed to be zero [34]. This is why the desired impedance matrices M_d , C_d , and K_d have a diagonal form with the size of 18×18 . $e = X_d - X$, \dot{e} , and \ddot{e} are 18×1 vectors representing the pose error, velocity error, and acceleration error, respectively. X_d and X denote the 18×1 desired pose and actual pose vectors of the robotic brace, respectively. The actual pose of the robotic brace is defined as $X = [X_1^T \ O_{10}^T \ X_2^T \ O_{20}^T \ X_3^T \ O_{30}^T]^T$, where

$X_i = [x_i \ y_i \ z_i]^T$, $i = 1, 2, 3$ denote the position vector of the i th moving ring with respect to the world frame. Besides, $O_{i0} = \theta_i [s_{xi} \ s_{yi} \ s_{zi}]^T$, $i = 1, 2, 3$ represents the orientation of the i th moving ring expressed in the screw representation form with respect to the world frame. It should also be mentioned that the x -, y -, and z -directions are defined along the normal vector to the coronal, sagittal, and transverse planes, respectively. Note that the 18×1 interaction force vector \mathcal{F}_e includes both contact forces and contact torques. The virtual pose of the robotic brace X_v required to achieve the desired pose X_d in the presence of the measured interaction force \mathcal{F}_e is computed by rewriting Equation (4) as follows:

$$M_d \ddot{X}_v + C_d \dot{X}_v + K_d X_v = M_d \ddot{X}_d + C_d \dot{X}_d + K_d X_d - \mathcal{F}_e, \quad (5)$$

where \dot{X}_d and \ddot{X}_d represent the desired velocity and acceleration of the robotic brace. \dot{X}_v and \ddot{X}_v also denote the virtual velocity and acceleration of the robotic brace, respectively. The inverse dynamic control (IDC) is then utilized for the pose control of the robotic brace, while the virtual pose X_v is applied as a reference input to the IDC. Therefore, the control law of the IDC is formulated as follows:

$$\mathcal{F} = M(\ddot{X}_v + K_d(\dot{X}_v - \dot{X}) + K_p(X_v - X)) + C\dot{X} + G + \mathcal{F}_e, \quad (6)$$

where K_p and K_d are 18×18 diagonal matrices and denote the proportional and derivative gains of the IDC. M , C , and G denote the 18×18 mass matrix, the 18×18 Coriolis and the centrifugal matrix, and the 18×1 gravity vector of the robotic brace, respectively [17, 18]. \dot{X} represents the actual velocity of the robotic brace and \mathcal{F} is also an 18×1 vector representing the projection of the actuator forces τ in the task space.

The main issue in implementing IC into the AIS brace treatment lies in finding a way to determine the desired impedance parameters. In other words, finding the desired impedance parameters required for providing the desired HRI for the AIS brace treatment is challenging. Although there is no research article focusing on implementing IC into the AIS brace treatment, the importance of the in-brace correction pressure/force (interaction force \mathcal{F}_e) has been widely studied in the literature [35–38]. In the AIS brace treatment, it is crucial for orthotists to know how much interaction force \mathcal{F}_e is exerted on the patient's torso by the brace, because the interaction force \mathcal{F}_e has a direct impact on the AIS curvature correction and the effectiveness of

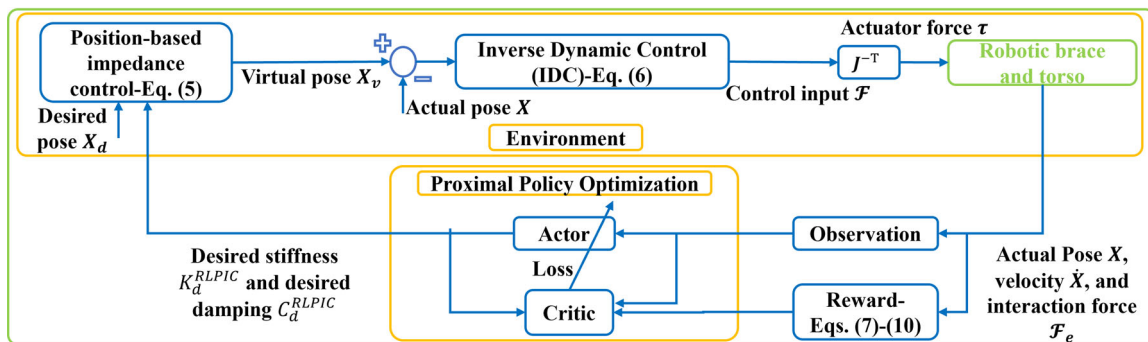


FIGURE 3 | Schematic diagram of the reinforcement learning-based position-based impedance control (RLPIC).

the AIS brace treatment. The interaction force \mathcal{F}_e reported by the above-mentioned in vivo studies has an extensive range, and it is still unclear [35–38]. On the other hand, the impedance parameters directly impact the interaction force \mathcal{F}_e and motion variables. This is why determining the desired impedance parameters required for providing the desired interaction force \mathcal{F}_e and motion for the AIS brace treatment is challenging. In this section, an RL-based approach is proposed to predict the desired impedance parameters of the PIC (\mathbf{K}_d and \mathbf{M}_d) according to the interaction force and motion measurements. For this purpose, the environment, actions, observations, the reward function, the RL algorithm, and the architecture of the NN used for training the RL algorithm are defined. The robotic brace, human torso, and the PIC approach presented in Equations (4)–(6) are assumed to be the environment for the RL agent, as shown in Figure 3. Note that the 5D three-layer DT is used to train the RL agent because the RL agent learns from both good and bad actions and training the RL algorithm using the human subject may cause unexpected motion/forces and consequently injuries on the patient's torso. Therefore, instead of using the PS as the environment, the 5D three-layer DT along with the PIC is assumed to be the environment for the RL agent. The desired stiffness \mathbf{K}_d and damping \mathbf{C}_d are chosen as the action variables to allow the RL agent to control the physical HRI. Predicting the desired impedance parameters by the RL agent while following the desired pose will allow the robot to provide the desired HRI. The observation space consists of the robot's pose \mathbf{X} , desired pose \mathbf{X}_d , velocity $\dot{\mathbf{X}}$, desired velocity $\dot{\mathbf{X}}_d$, and the interaction force \mathcal{F}_e .

The design of the reward function plays a vital role in learning a task using an RL agent. The main goal of the reward function is to quantify the goodness of a selected action, and it is used for updating the current policy to take the best action in the next time step. Design of the reward function depends on the type of the task and it should be defined such that the robot gets rewards to incentivize such a behavior. In this paper, the reward function is defined such that the robotic brace follows the desired trajectory \mathbf{X}_d , keeps tracking a reference velocity in the task space $\dot{\mathbf{X}}_d$, and avoids exerting a large interaction force on the patient's torso. The reward function r is defined as follows: [39]

$$\mathbf{r} = \mathbf{w}_p \mathbf{r}_p + \mathbf{w}_f \mathbf{r}_f + \mathbf{w}_v \mathbf{r}_v, \quad (7)$$

where \mathbf{r}_p , \mathbf{r}_f , and \mathbf{r}_v are 18×1 column vectors denoting the pose reward, the interaction force reward, and the velocity reward in 18 DOFs, respectively. The weights \mathbf{w}_p , \mathbf{w}_f , and \mathbf{w}_v are 1×18 row

vectors representing the weights of the pose, the interaction force, and velocity rewards in 18 DOFs, respectively. The weights are tuned manually according to the importance of the corresponding reward. The reward terms for i^{th} DOF are defined as follows:

$$\mathbf{r}_p(i) = \exp(-[c_{pi}(\mathbf{X}_d(i) - \mathbf{X}(i))]^2), \quad (8)$$

$$\mathbf{r}_f(i) = \exp(-[c_{fi}(\mathcal{F}_{e,\max}(i) - \mathcal{F}_e(i))]^2), \quad (9)$$

$$\mathbf{r}_v(i) = \exp(-[c_{vi}(\dot{\mathbf{X}}_d(i) - \dot{\mathbf{X}}(i))]^2), \quad (10)$$

where c_{pi} , c_{fi} , and c_{vi} denote the error coefficients of the corresponding reward term. $\mathcal{F}_{e,\max}$ represents the upper boundary for the interaction force to ensure safe physical HRI. The error coefficients should be chosen such that each error has reasonable sensitivity [39].

The proximal policy optimization (PPO) algorithm is chosen as the RL algorithm [40, 41]. It is based on an actor-critic structure, combined with the trust region method. The architecture of the actor's NN consists of two fully connected layers of 256 neurons with ReLU as the activation function, followed by a fully connected layer with a size of one neuron and tanh as the activation function, which is in parallel with one fully connected layer with one neuron. The architecture of the NN for the critic in the PPO algorithm consists of two fully connected layers with 256 neurons. The activation function for each layer is ReLU. A third fully connected layer is also used to map the output of the previous layer into proper dimensions.

4 | Numerical Simulations and Experimental Results

In Section 4.1, the 5D three-layer DT is validated through numerical simulations and real-time experiments. The proposed RLPIC is trained and verified using the 5D three-layer DT in Section 4.2. The workflow of this section is shown in Figure 4. First, real-time experiments are conducted, and the interaction force–displacement measurements are collected. The same displacement is then applied as the primary input to the 5D three-layer DT. The NN regression model proposed in Section 2.3 is trained using data collected from real-time experiments and the unknown parameters of the “Spatial Contact Force” block are estimated. The estimated unknown parameters are finally optimized

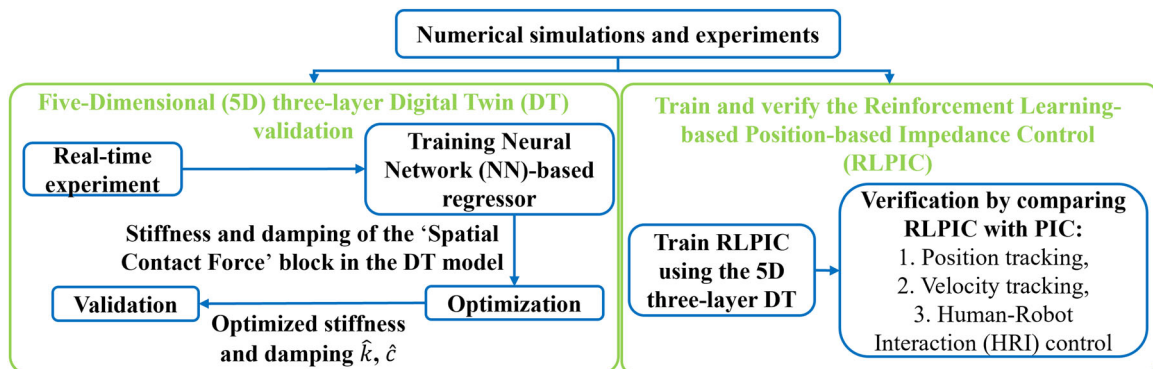


FIGURE 4 | Numerical simulations and experiments.

using the “Parameter Estimator” App from the Simulink Design Optimization Toolbox such that the output interaction force obtained from numerical simulations fits the interaction force collected through real-time experiments. Second, the RL agent is trained using the 5D three-layer DT and the RLPIC is verified in terms of position tracking, velocity tracking, and HRI control.

4.1 | Validation of the 5D Three-Layer DT Model

The 5D three-layer DT of the AIS brace treatment is validated through numerical simulations and real-time experiments. The implementation details of the three-layer digital model are explained first, as shown in Figure 5A. The MB model of the robotic brace is created using its SolidWorks model. The ‘Simscape Multibody Link’ plugin installed on SolidWorks is used to create an “.xml” file from the SolidWorks model of the robotic brace. The MB model of the robotic brace is then created in Simscape by importing the “.xml” file into Simscape using the MATLAB command “smimport.” “smimport” is a MATLAB command that is used for importing the CAD or URDF files into MATLAB. Note that the geometry and configuration of the MB model of the robotic brace fit the physical model of the robotic brace.

The solid body model of the human torso is created. One “File Solid” block from the Simscape Multibody Library is used to create the solid body model of the torso. The ‘File Solid’ block is used to model a solid body, with its geometry, inertial, color, and reference frame taken from a CAD file. The geometry of the torso model is taken from the 3D surface scan of a scoliosis patient [17]. The density of the torso, 1000 kg/m^3 , is taken from [42]. The solid body model of the torso is connected to the World Frame in the MB model of the robotic brace with a “6 DOF” joint available in the Simscape Multibody Library.

To create the HRI model, “Spatial Contact Force” blocks are used. Since direct connection of the moving rings in the MB model and the solid body model of the torso to the “Spatial

Contact Force” blocks is not allowed in Simscape, six proxies are connected to the moving rings and six proxies are also connected to the solid body model of the torso using the “Rigid Transform” block. The proxies are defined using the “Brick Solid” block. The physical contact between the moving rings and the torso model is defined as the physical contact between the proxies connected to the moving rings and the torso model. The proxies from the solid body model of the torso are finally connected to the base frame of the “Spatial Contact Force” blocks and the proxies from the moving rings are connected to the follower frame of the “Spatial Contact Force” block to form the three-layer digital model. Note that the “Rigid Transform” block defines and maintains a translational and rotational relationship between two frames during simulations. Besides, the “Brick Solid” block creates a prismatic shape with its geometry, inertia, and color.

The proxies are not directly dictating the exact geometry of the contacting surfaces. The proxies are mainly used as a simplified representation of the physical contact behavior, for example, stiffness, damping, and friction, of the complex bodies. The exact geometry (shape, size, and surface) of the bodies is not represented by the proxies. This is why six proxies at some specific points of the torso and moving rings’ CAD models were used to represent an approximate shape and geometry of the torso and moving rings. Besides, the mass of the proxies is defined as zero to avoid any effects on dynamics.

Second, real-time experiments are conducted. In the real-time experiment, shown in Figure 5B, a healthy human subject from our research group wears the robotic brace, and a 10 mm displacement along the $-y$ -direction (left side in the coronal plane seen from the front view) is implemented using the robotic brace equipped with a closed-loop control system. Note that only the second SGP of the robotic brace (two rings located in the middle of the robotic brace) is used for real-time experiments. The MRAIC strategy is utilized to provide the desired HRI control. National Instruments (NI) compact RIO 9082 is

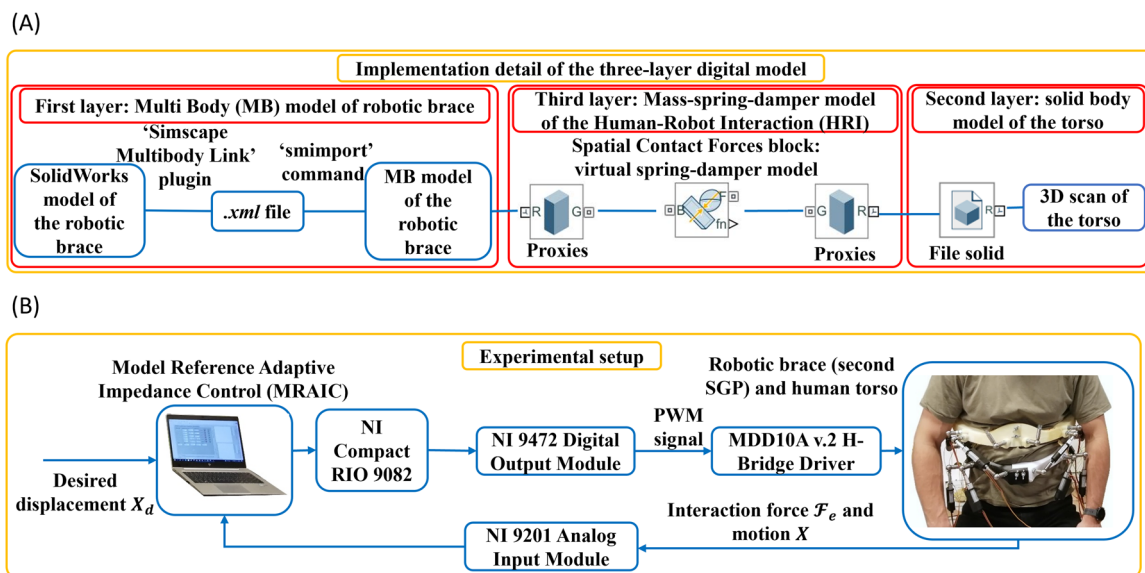


FIGURE 5 | Implementation details and experimental setup: (A) Implementation details of the three-layer digital model. (B) Experimental setup by which the interaction force–motion data are collected to be used for the five-dimensional (5D) three-layer digital twin (DT) validation.

used for implementing the real-time control system. The NI 9201 input modules and NI 9472 output modules are used for measuring the interaction force–motion data and implementing the PWM signal to the linear actuators. The interaction force–motion data are collected to be used for the unknown parameters' identification, optimization, and DT validation.

Third, the NN regression model presented in Section 2.3 is trained using the collected data. The interaction forces obtained from the experiments and predicted by the NN regression model along with the error between them are shown in Figure 6. It can be seen that the interaction forces obtained from the experiments and the NN regression model fit each other. The optimal size of the hidden layer of the NN regression model is $n = 10$. The root mean square error (RMSE) is 0.031343. The unknown parameters of the “Spatial Contact Force” block k and c are estimated using the NN regression model in Equations (2) and (3) as 10,000 N/m and 100 Ns/m, respectively. The estimated unknown parameters are optimized using the “Parameter Estimator” App from the Simulink Design Optimization Toolbox. In this simulation, the gradient descent method and the sequential quadratic programming algorithm [43] are selected for solving the optimization problem and the sum squared errors are defined as the cost function. The optimal values for the parameters k and c are 6400 N/m and 150, respectively. Note that the performance of the MRAIC controller used in the real-time experiments, which depends on the controller gains, has direct effects on the measured interaction force and displacement data set. Therefore, the MRAIC performance can act as a source of the error that can be seen in Figure 6. In addition, the size of collected data set is the other source of the error that can be seen in Figure 6. More experimental tests need to be conducted to create a huge source of data set. The huge source of data set will improve the training process and increase the accuracy of the output interaction force predicted by the NN regressor. Besides, the noisy measurements, experimental system setup, and fabrication errors are the sources of error in the experiments.

Figure 7 shows the interaction force obtained from the 5D three-layer DT in comparison with the measured interaction force

obtained from the real-time experiments. The error between the interaction force of the DT and the PS is reasonable and the DT is validated. Besides, it can be seen that the physical HRI needs approx. 3 s to be settled down and stabilized. One reason behind the large discrepancy that is shown in Figure 7 is that the MRAIC will provide a smooth interaction force in the experiments, while the interaction force obtained from the numerical simulations of the DT model is not smooth because the former is open loop. The other reason is that the MRAIC needs transient time to provide the desired behavior and the steady-state response. However, the numerical simulations of the DT model are open loop, and the initial transient that is seen in the experiments is not simulated with the numerical simulations of the DT model. Note that the frequency of the data acquisition in the real-time experiments and the sample time in the numerical simulations is 50 ms and the total running time is 12 s.

Note that since the X-ray images and biomechanical data of the patient's torso are not available, it is not possible to create a biomechanical model of the patient's torso. This is why it is assumed that the desired displacement of the torso required for AIS curvature correction X_d is given and the AIS curvature evolution is not studied in this paper. The desired displacement of the patient's torso can be obtained using the X-ray images and the biomechanical characteristics of the patient's torso. In other words, in this paper, the main purpose is to implement the desired displacement to the patient's torso using the trained RLPIC, while providing a smooth interaction force. Besides, this assumption does not risk invalidating the adaptability of the RLPIC algorithm for the brace, because the proposed trained RL agent has the capability to adapt to the changes made in the environment, that is, torso model. In other words, the RL agent learns the desired impedance parameters required for providing the desired displacement and smooth interaction force according to the interaction force and motion data. Therefore, a complete biomechanical model of the torso instead of using a solid body model of the torso will not have any impact on the adaptability of the proposed RLPIC algorithm. In addition, the 5D three-layer DT model is patient-specific. Although the torso is modeled as a single solid body, the mass of the solid body of the torso is different from patient to patient and the stiffness

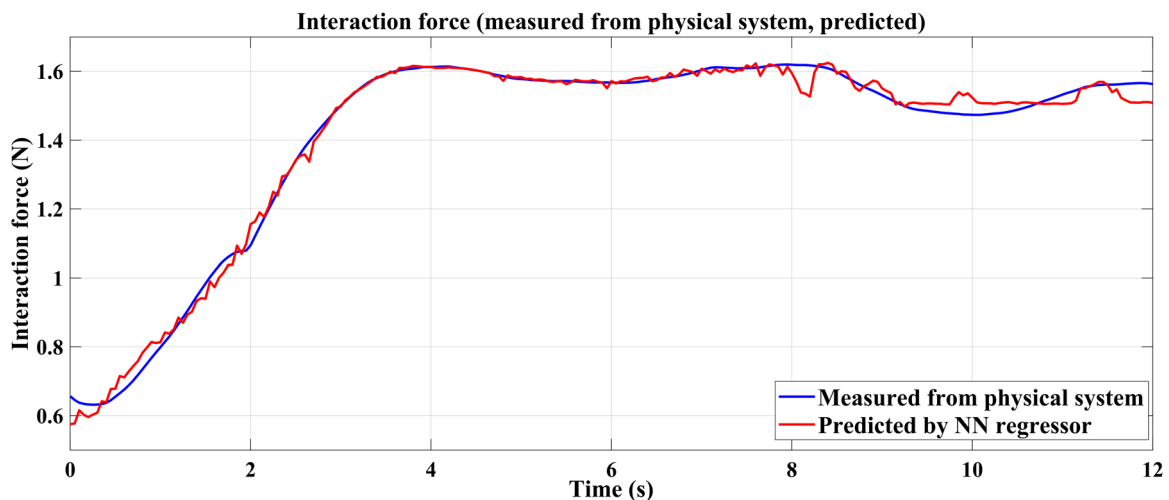


FIGURE 6 | Interaction force in the y -direction obtained from the real-time experiments, the interaction force predicted by the neural network (NN)-based regression model, and the error between them.

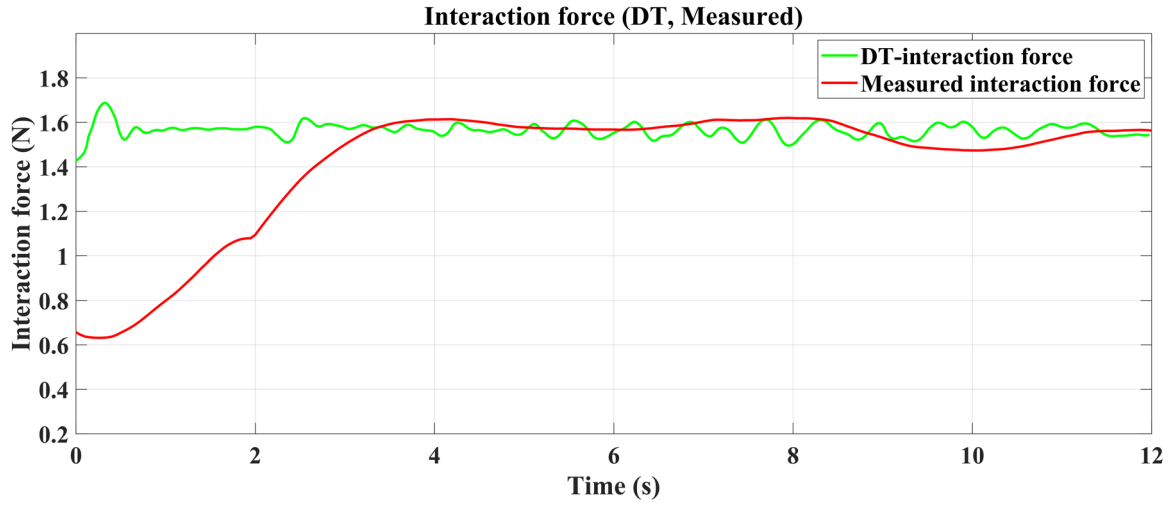


FIGURE 7 | Interaction force in the y -direction obtained from the real-time experiments and the output interaction force from the five-dimensional (5D) three-layer digital twin (DT).

and damping parameters \mathbf{k} and \mathbf{c} of the physical HRI between the robotic brace and the torso model are also patient-specific.

Since the proposed RLPIC controller is trained using a patient-specific 5D three-layer DT model, the adaptability of the RLPIC is not affected by modeling the torso as a solid body.

4.2 | Verify the Proposed RLPIC

In this subsection, the proposed RLPIC is trained using the 5D three-layer DT and the performance of the RLPIC is verified by comparing the RLPIC with typical PIC in terms of position tracking, velocity tracking, and HRI control. For this purpose, the RLPIC and PIC are applied to the second SGP in the y -direction and the IDC is utilized for pose control of the second SGP in the other five DOF. In other words, it is assumed that the physical interaction occurs only in the y -direction to mimic the scoliosis curvature correction, which is also along the y -direction (intersection of the coronal plane and transverse plane). The desired trajectory in the y -direction is defined as a cubic polynomial with -10 mm displacement, while the desired translational and rotational displacements in the other five DOFs are assumed to be zero. Two Simulink models are created by implementing the PIC and RLPIC to the 5D three-layer DT model. The “RL agent” block from the Simulink Library is used to implement the RLPIC algorithm formulated in Section 3 to the 5D three-layer DT model. The “RL agent” block is a Simulink block that is used to simulate and train an RL agent. The observation space is considered a 7×1 vector as $[y_{2d} \ \dot{y}_{2d} \ y_2 \ \dot{y}_2 \ \mathcal{F}_{ey} \ x_2 \ z_2]^T$, where y_{2d} , \dot{y}_{2d} , y_2 , \dot{y}_2 , \mathcal{F}_{ey} , x_2 , and z_2 represent the desired position in the y -direction, the desired velocity in the y -direction, the actual position in the y -direction, the actual velocity in the y -direction, the interaction force in the y -direction, the actual position in the x -direction, and the actual position in the z -direction, respectively. Note that all these variables belong to the second SGP. The desired stiffness \mathbf{K}_d is defined as the action space for the RL agent.

The desired stiffness parameter is a dominant factor for low-velocity motion like AIS treatment using a robotic brace [44].

Therefore, the desired stiffness parameter in RLPIC (K_d^{RLPIC}) is defined as the action variable to be regulated using the RL agent and a constraint is imposed on the desired damping of the RLPIC (C_d^{RLPIC}) to ensure system stability [44–47]. The desired mass for both PIC and RLPIC is defined as a constant value ($M_d^{\text{RLPIC}} = M_d^{\text{PIC}} = 2$ Kg). The desired damping of the RLPIC is assumed to be $C_d^{\text{RLPIC}} = 2\sqrt{M_d^{\text{RLPIC}} \times K_d^{\text{RLPIC}}} = 2\sqrt{2 \times K_d^{\text{RLPIC}}}$ to obtain a critically damped response and ensure system stability [44–47]. The desired stiffness for PIC (K_d^{PIC}) is assumed to be 6000 N/m and the desired damping of the PIC is defined as $C_d^{\text{PIC}} = 2\sqrt{2 \times K_d^{\text{PIC}}}$ to ensure stability. The proportional and derivative gains for the IDC are considered to be 6500 and 1000 for other five DOF.

In this simulation, the reward function is defined using Equation (7)–(10) as follows:

$$r = 2r_y + 2r_{\dot{y}} + 2r_{\mathcal{F}_{ey}} + 2r_x + 2r_z, \quad (11)$$

$$r_y = \exp(-[80(y_{2d} - y_2)]^2), \quad (12)$$

$$r_{\dot{y}} = \exp(-[80(\dot{y}_{2d} - \dot{y}_2)]^2), \quad (13)$$

$$r_{\mathcal{F}_{ey}} = \exp(-[5(\mathcal{F}_{ey} - 4)]^2) \quad (14)$$

$$r_x = \exp(-[80(x_2 - x_2(t=0))]^2), \quad (15)$$

$$r_z = \exp(-[80(z_2 - z_2(t=0))]^2), \quad (16)$$

where r_y , $r_{\dot{y}}$, $r_{\mathcal{F}_{ey}}$, r_x , and r_z represent the position reward in the y -direction, the velocity reward in the y -direction, the interaction force reward in the y -direction, the position reward in the x -direction, and the position reward in the z -direction, respectively. The initial guesses for the error coefficients c_{pi} , $c_{\dot{p}_i}$, and c_{vi} are chosen to be in the same range as that chosen by Anand AS [39], and the final values for the error coefficients are obtained by trial and error. The reward terms are defined such that the RL agent is rewarded if the robot approaches the desired position and velocity in the y -direction while having no displacements in the x and z -directions and maintaining the interaction

force in the y -direction less than $4N$. The conditions for episode termination are defined as follows:

$$y > 0.006 \text{ m}, y < -0.4 \text{ m}, |y'| > 0.6 \frac{\text{m}}{\text{s}}, \quad (17)$$

$$|\mathcal{F}_{ey}| > 50 \text{ N}, \quad (18)$$

$$|x| > 0.02 \text{ m}, |z| > 0.02 \text{ m}. \quad (19)$$

The PPO algorithm is utilized for training the RL agent, and the architecture of the actor-critic NN is defined as explained in Section 3. The hyperparameters of the PPO agent are as follows: sample time: 0.050s, discounted factor: 0.99, batch size 64, experience horizon 512, entropy loss weight: 0.01, clip factor: 0.2, number of epochs 10, advantage estimation method *GAE*, and the *GAE* factor 0.95. The hyperparameters of the actor and critic are assumed to be as follows: learning rate 0.0003, gradient threshold 1, optimizer *adam*, denominator offset 10^{-8} , gradient decay 0.9, squared gradient decay 0.999, and gradient threshold method *L2 norm*. The L2 regularization parameters for actor and critic networks are defined as 10^{-5} and 0.0002, respectively. The PPO agent is trained for 1000

episodes and the maximum length of the episode is set to 240, while the total simulation time for each episode is 12 s.

The results obtained from the trained RLPIC are compared with those of the PIC in terms of position tracking, velocity tracking, and HRI control. The desired position in the y -direction and the actual position in the y -direction obtained from the RLPIC and PIC are shown in Figure 8. It can be seen that the position tracking error for RLPIC is less than that of the PIC. In other words, the RLPIC enables the robot to follow the desired position while it has physical interaction with the torso. Figure 9 shows the desired velocity in the y -direction compared to the velocity obtained from the RLPIC and PIC. As can be seen in Figure 9, the desired velocity is followed in both RLPIC and PIC, while the velocity obtained from RLPIC is more stable than that of the PIC at the beginning of the simulation. The velocity of the PIC shows a sudden change of approx. -2.8 mm/s at the beginning of the simulations, while the velocity of RLPIC experiences only -0.6 mm/s sudden changes at the beginning.

The positions in the x - and z -directions obtained from the IDC while the RLPIC and PIC are applied to the y -direction are also

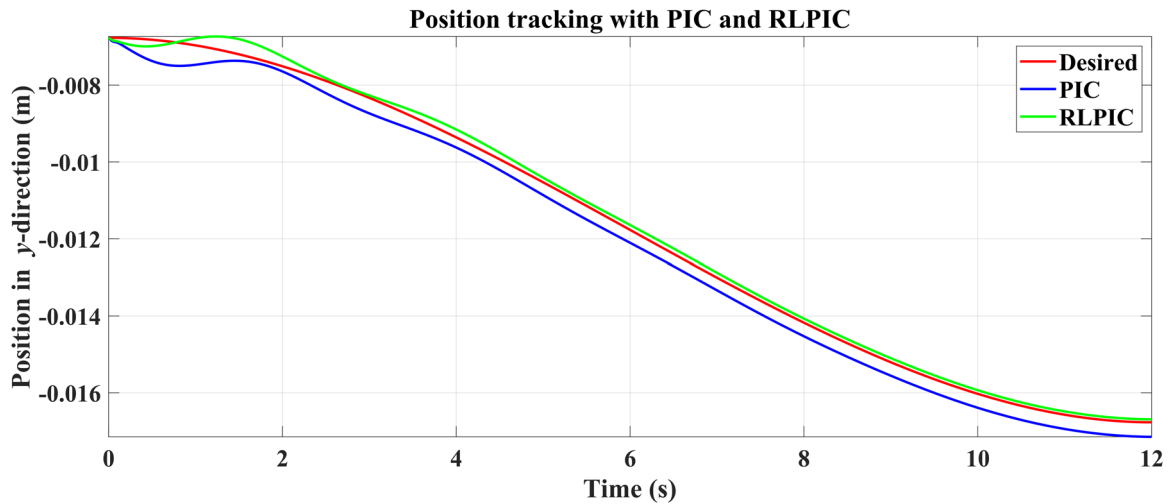


FIGURE 8 | Comparison of the position in the y -direction obtained from the RLPIC and PIC with the desired position in the y -direction.

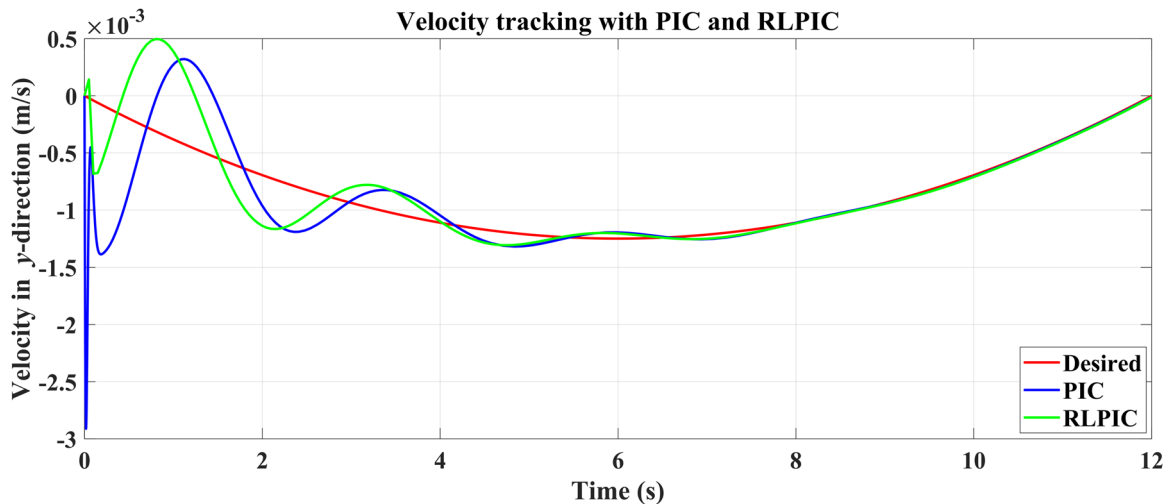


FIGURE 9 | Velocities in the y -direction obtained from the RLPIC and PIC compared to the desired velocity in the y -direction.

compared with the desired positions in the x - and z -directions in Figures 10 and 11, respectively. It is indicated that the IDC enables the robot to maintain its initial position in the x - and z -directions. The maximum deviations from the desired position in the x - and z -directions in the case where the RLPIC is applied to the y -direction are less than that of the case where the PIC is applied to the y -direction. The interaction forces in the y -direction for the RLPIC and PIC are presented in Figure 12. It is concluded that the interaction force obtained from RLPIC is more stable than that of the PIC at the beginning of the physical HRI. After approx. 4 seconds, the interaction force in both the RLPIC and PIC settles down and smooth interaction forces are produced. As can be seen in Figure 12, the interaction force from the RLPIC is smaller than the maximum interaction force defined in the reward function. It is concluded that the RL algorithm can provide a safe and compliant interaction force by selecting the proper reward and penalty terms in the reward function without designing any force control law. Figure 13 shows the desired stiffness in the y -direction (K_d^{RLPIC}) predicted by the RL agent. The desired damping is obtained

using $C_d^{\text{RLPIC}} = 2\sqrt{2} \times K_d^{\text{RLPIC}}$. As can be seen in Figure 13, the average stiffness value predicted using the RLPIC is 49,990 N/m and the variation range of the desired stiffness is 25–30 N/m, i.e., 0.05%–0.06% of its average value. It is concluded that the percentage of the stiffness variation range is too small, meaning that the desired stiffness is predicted well using the RL agent. Since the stiffness and damping parameters of the physical HRI model k and c are estimated as a constant value and constant desired impedance parameters in the RLPIC can consequently increase the smoothness of the interaction force \mathcal{F}_e , the desired impedance parameters in the RLPIC are assumed to be constant. Therefore, the deviation of stiffness from its average value can be a good measure of accuracy. Besides, the predicted desired impedance parameters satisfy three criteria that are defined for evaluating the performance of RLPIC. In other words, these predicted values for the impedance parameters lead to acceptable position tracking, velocity tracking, and HRI control and the moving ring of the robotic brace follows the desired trajectory and velocity, while providing a smooth interaction force and maintaining it smaller than a

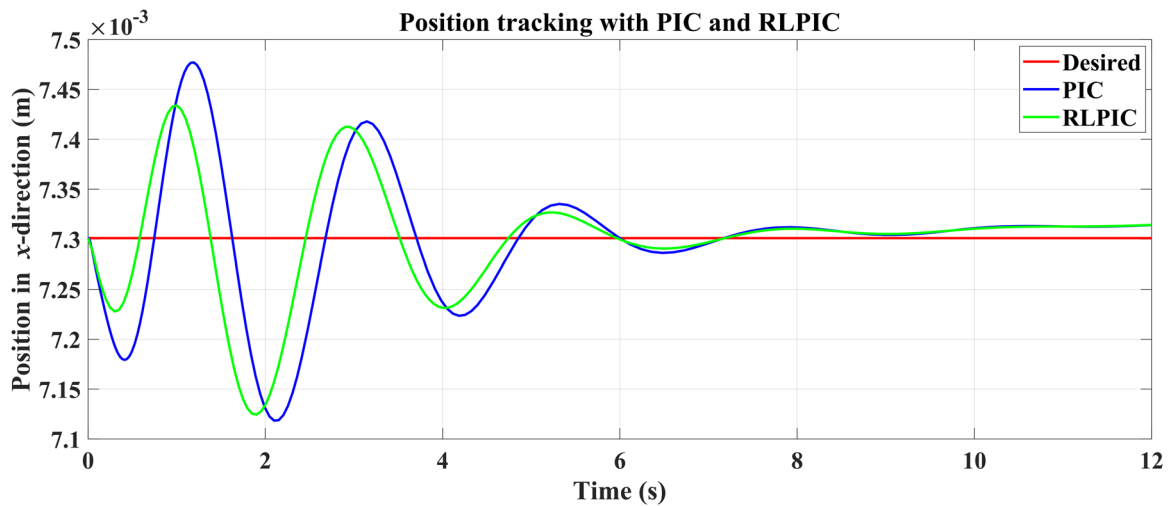


FIGURE 10 | The position in the x -direction obtained from the IDC while the RLPIC and PIC are applied in the y -direction is compared with the desired position in the x -direction.

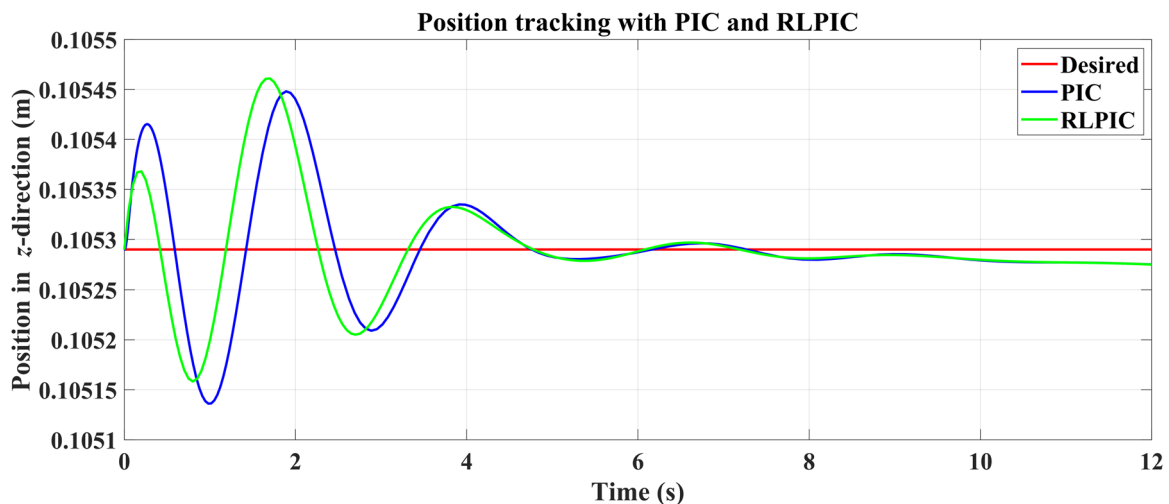


FIGURE 11 | The position in the z -direction obtained from the IDC while the RLPIC and PIC are applied in the y -direction is compared with the desired position in the z -direction.

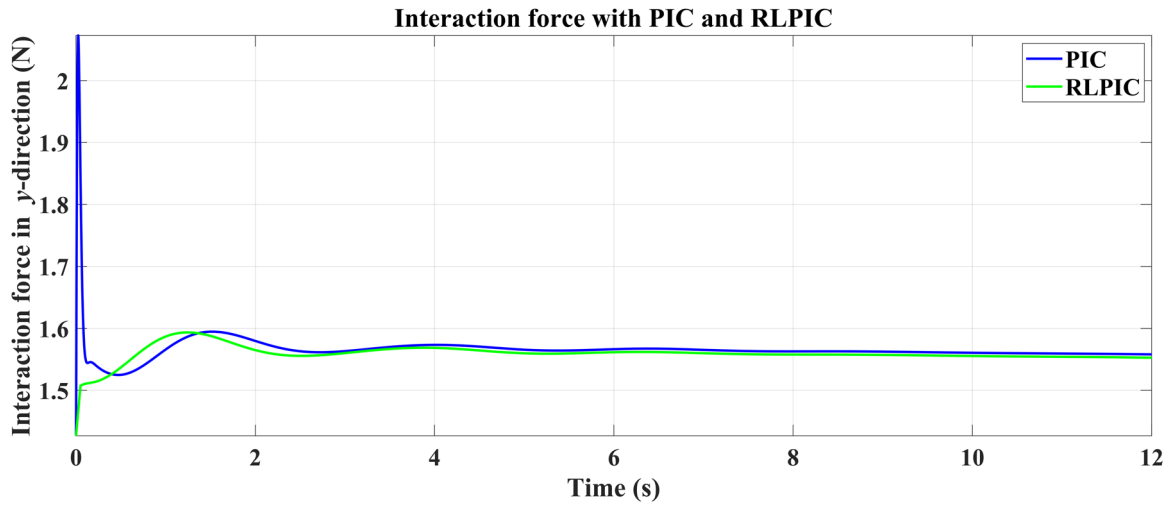


FIGURE 12 | Interaction force in the y -direction obtained from the RLPIC and PIC.

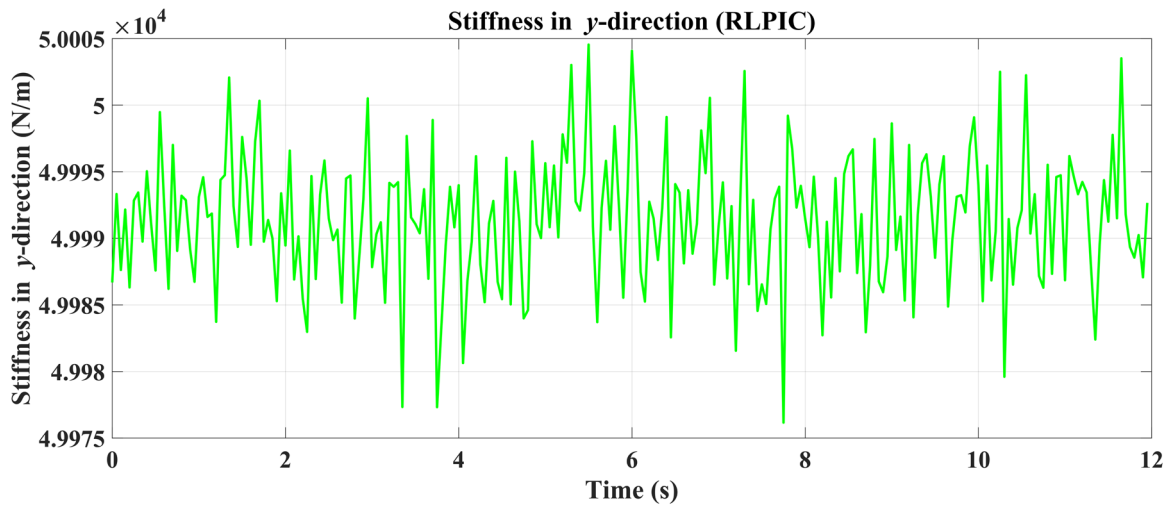


FIGURE 13 | Desired stiffness K_d^{RLPIC} of the RLPIC algorithm in the y -direction predicted by the reinforcement learning agent.

TABLE 1 | Root mean square error (RMSE) of positions in the y -, x -, and z -directions, velocity in the y -direction, and the interaction force in the y -direction.

	Root mean square error (RMSE)				
	Position in the y -direction (mm)	Velocity in the y -direction (mm/s)	Position in the x -direction (mm)	Position in the z -direction (mm)	Interaction force in the y -direction (N)
PIC	0.31054	0.95586	0.072539	0.067516	0.1526
RLPIC	0.15245	0.22340	0.054103	0.049757	0.0163

threshold value. This is why the stiffness value that the RLPIC algorithm converges to is the true value of stiffness. In other words, the desired stiffness is predicted by the RL agent and the robotic brace learns the desired physical HRI by trial and error in a virtual environment. However, the desired impedance model in PIC is constant and it is not adapted according to the instantaneous interaction force–motion measurements.

To quantify the performance of RLPIC, the RMSEs of the positions in y , x , and z -directions, velocity in the y -direction, and the interaction force in the y -direction for both RLPIC and PIC are presented in Table 1. The RMSE was computed as

$\sqrt{\text{mean}((\text{desiredvalue} - \text{actualvalue})^2)}$. As shown in Table 1, the RMSEs of the positions in the y -, x -, and z -directions and the velocity in the y -direction for RLPIC is less than those of the PIC. Therefore, the performance of the RLPIC is verified in terms of position tracking and velocity tracking. To verify the performance of the RLPIC in terms of HRI control, the desired value in the RMSE formula is replaced by the final value of the interaction force to quantify the deviation of the interaction force from its final (steady state) value and evaluate how smooth the HRI is. In Table 1, it can be seen that the deviation of the interaction force from its final value (RMSE of the interaction force) for RLPIC is less than that of the PIC. It is

concluded that the HRI in RLPIC is smoother than that of the PIC and the performance of RLPIC is verified in terms of HRI control as well. In brief, it is concluded that implementing the RL algorithm into the PIC can improve the performance of the typical PIC in terms of motion tracking and providing safe HRI by predicting the desired impedance parameters.

Although the RLPIC implies complications, the advantages of RLPIC in adapting the impedance parameters of RLPIC to the changes that occur in the environment and the capability of predicting the desired impedance parameters make the RLPIC controller a better solution than typical PIC in most cases. One of the most challenging points in a typical PIC is how to determine the desired impedance parameters. The constant impedance parameters of PIC should be chosen by trial and error, while the trained RLPIC has the capability to predict the desired impedance parameters according to the measured interaction force and motion data. Besides, the desired criteria for HRI control can be satisfied by considering the criteria in the reward function of RLPIC, while the PIC does not have such an advantage to consider all the HRI constraints. In addition, the impedance parameters of the environment are not constant and constant impedance parameters of PIC do not provide the desired HRI for these types of environments. However, the RLPIC has the capability to regulate the variable impedance parameters to adapt the RLPIC controller to the changes occurring in the environment. Although Variable Position-based Impedance Control (VPIC) can also be a solution, the RLPIC has the potential to show better performance than VPIC by only defining a proper reward function without any need to design an adaptive law to regulate the variable impedance parameters of typical VPIC.

5 | Conclusion

AIS brace treatment has not been integrated with DT technology and learning-based HRI control approaches. This paper proposed an RLPIC algorithm that enables the robotic brace to learn the desired HRI required for AIS rehabilitation. Besides, a 5D three-layer DT of the AIS brace treatment is developed to be used for training the RLPIC algorithm to avoid trial an error on human subjects. The 5D three-layer DT includes the PS, a digital model, bidirectional data flow (P2V and V2P), and optimization. The three-layer digital model is created by adding the MB model of the robotic brace to the solid body model of the human torso, while the physical HRI between the robotic brace and the torso is modeled as a mass–spring–damper system using the “Spatial Contact Force” block. Real-time experiments and numerical simulations are carried out to update the unknown parameters of the HRI model, establish the P2V data flow, and validate the DT. An NN-based regression model is proposed to estimate the unknown stiffness and damping parameters of the physical HRI model using the interaction force–motion data collected from the real-time experiments. The estimated parameters are optimized using the ‘Parameter Estimator’ App from the Simulink Design Optimization Toolbox such that the interaction force obtained from the real-time experiments fits the interaction force obtained from the 5D three-layer DT. Finally, the RLPIC algorithm trained using the DT model is utilized to predict the desired HRI and establish the V2P data flow such that the robotic brace follows a desired

trajectory, maintains the interaction force to less than a maximum value, and tracks the desired velocity.

Since the RLPIC algorithm is trained using the DT model, the more accurate the digital model, the more reliable the RLPIC algorithm is in the clinical rehabilitation therapy. On the other hand, the ML algorithms used for identifying the unknown parameters of the DT model improve the digital model accuracy and twin the digital model with its physical counterpart. In other words, ML algorithms improve the reliability of the DT model and consequently can increase the reliability and effectiveness of the RLPIC algorithm in the clinical rehabilitation therapy. Besides, ML algorithms can be useful in updating the unknown parameters of the DT model such that the DT model adapts to the biomechanical changes in the patient’s torso that occur during the period of growth spurt.

Although an RLPIC approach is proposed and a 5D three-layer DT is created for training the RL agent, further improvements are needed. For example, more real-time experiments with different types of motion need to be performed to collect a huge source of the data set that can be used for unknown parameter identification, DT validation, and training the RL agent for different variants of the environment. In addition, a more complete version of mass–spring–damper models, for example, nonlinear mass–spring–damper models, can also be used for HRI modeling as an interesting subject for future work. However, this represents preliminary research work in implementing RLPIC and DT modeling into the AIS brace treatment and no other research work has focused on implementing RLPIC and DT modeling into the AIS brace treatment. This is why the simplest mass–spring–damper model, that is, the linear model with constant coefficients, is utilized in this paper for HRI modeling. Besides, the DT model can also improve the brace design process and brace adjustments by predicting the biomechanical behavior of the brace performance in advance without any need for trial and error on human subjects. It can also play an important role in actual AIS rehabilitation by training the RLPIC algorithm, which has the potential to be used in the actual AIS rehabilitation to provide the desired HRI required for AIS treatment. Furthermore, improving the RL algorithm used in the RLPIC algorithm by adopting advanced RL-based approaches currently used in robotic applications will be an interesting topic for future work. Note that the trained RLPIC algorithm has not been validated through real-time experiments yet because of the lack of human subjects and limited time. As ongoing work, we will conduct more real-time experiments to collect more data and validate the trained RLPIC algorithm.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Farhad Farhadiyadkuri and Xuping Zhang. The first draft of the manuscript was written by Farhad Farhadiyadkuri and all authors commented on previous versions of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The research data are available from the corresponding author upon reasonable request.

References

1. T. Maruyama, K. Takesita, T. Kitagawa, and Y. Nakao, "Milwaukee Brace," *Physiotherapy Theory and Practice* 27, no. 1 (2011): 43–46, <https://doi.org/10.3109/09593985.2010.503992>.
2. M. J. Khan, V. M. Srinivasan, and A. H. Jea, "The History of Bracing for Scoliosis," *Clinical Pediatrics* 55, no. 4 (2016): 320–325, <https://doi.org/10.1177/0009922815615829>.
3. H. R. Weiss and M. Werkmann, "'Brace Technology' Thematic Series—The Scoliotic® Chêneau Light™ Brace in the Treatment of Scoliosis," *Scoliosis* 5, no. 1 (2010): 19, <https://doi.org/10.1186/1748-7161-5-19>.
4. C. T. Price, D. S. Scott, F. E. Reed, Jr., and M. Riddick, "Nighttime Bracing for Adolescent Idiopathic Scoliosis With the Charleston Bending Brace. Preliminary Report," *Spine* 15, no. 12 (1990): 1294–1299, <https://doi.org/10.1097/00007632-199012000-00011>.
5. R. S. Fayssoux, R. H. Cho, and M. J. Herman, "A History of Bracing for Idiopathic Scoliosis in North America," *Clinical Orthopaedics & Related Research* 468, no. 3 (2010): 654–664, <https://doi.org/10.1007/s11999-009-0888-5>.
6. A. G. Veldhuizen, J. Cheung, G. J. Bulthuis, and G. Nijebanning, "A New Orthotic Device in the Non-Operative Treatment of Idiopathic Scoliosis," *Medical Engineering & Physics* 24, no. 3 (2002): 209–218, [https://doi.org/10.1016/S1350-4533\(02\)00008-5](https://doi.org/10.1016/S1350-4533(02)00008-5).
7. M. S. Wong, J. C. Y. Cheng, T. P. Lam, et al., "The Effect of Rigid Versus Flexible Spinal Orthosis on the Clinical Efficacy and Acceptance of the Patients With Adolescent Idiopathic Scoliosis," *Spine* 33, no. 12 (2008): 1360–1365, <https://doi.org/10.1097/BRS.0b013e31817329d9>.
8. A. Ali, V. Fontanari, M. Fontana, and W. Schmölz, "Spinal Deformities and Advancement in Corrective Orthoses," *Bioengineering* 8, no. 1 (2021): 2, <https://doi.org/10.3390/bioengineering8010002>.
9. P. Joon-Hyuk, P. Stegall, and S. K. Agrawal, "Dynamic Brace for Correction of Abnormal Postures of the Human Spine." in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, (2015), 5922–5927, <https://doi.org/10.1109/ICRA.2015.7140029>.
10. A. Ali, V. Fontanari, W. Schmölz, and S. K. Agrawal, "Active Soft Brace for Scoliotic Spine: A Finite Element Study to Evaluate In-Brace Correction," *Robotics* 11, no. 2 (2022): 37, <https://doi.org/10.3390/robotics11020037>.
11. R. Ray, L. Nouaille, B. Colobert, L. Calistri, and G. Poisson, "Design and Position Control of a Robotic Brace Dedicated to the Treatment of Scoliosis," *Robotica* 41, no. 5 (2023): 1466–1482, <https://doi.org/10.1017/S0263574722001825>.
12. A. L. Jutinico, J. C. Jaimes, F. M. Escalante, J. C. Perez-Ibarra, M. H. Terra, and A. A. G. Siqueira, "Impedance Control for Robotic Rehabilitation: A Robust Markovian Approach," *Frontiers in Neurobotics* 11 (2017): 11–43, <https://doi.org/10.3389/fnbot.2017.00043>.
13. Y. Zhou, J. She, Z.-T. Liu, C. Xu, and Z. Yang, "Implementation of Impedance Control for Lower-Limb Rehabilitation Robots." in *2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*, Victoria, BC, Canada, (2021), 700–704, <https://doi.org/10.1109/ICPS49255.2021.9468210>.
14. J. Bai, A. Song, T. Wang, and H. Li, "A Novel Backstepping Adaptive Impedance Control for an Upper Limb Rehabilitation Robot," *Computers & Electrical Engineering* 80 (2019): 106465, <https://doi.org/10.1016/j.compeleceng.2019.106465>.
15. Z. Li, Z. Huang, W. He, and C. Y. Su, "Adaptive Impedance Control for an Upper Limb Robotic Exoskeleton Using Biological Signals," *IEEE Transactions on Industrial Electronics* 64, no. 2 (2017): 1664–1674, <https://doi.org/10.1109/TIE.2016.2538741>.
16. L. Luo, L. Peng, Z. Hou, and W. Wang, "An Adaptive Impedance Controller for Upper Limb Rehabilitation Based on Estimation of Patients' Stiffness." in *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Macau, Macao (2017), 532–537, <https://doi.org/10.1109/ROBIO.2017.8324471>.
17. F. Farhadiyadkuri, A. M. Popal, S. S. Paiwand, and X. Zhang, "Interaction Dynamics Modeling and Adaptive Impedance Control of Robotic Exoskeleton for Adolescent Idiopathic Scoliosis," *Computers in Biology and Medicine* 145 (2022): 105495, <https://doi.org/10.1016/j.combiomed.2022.105495>.
18. F. Farhadiyadkuri and X. Zhang, "Novel Interaction Control in Adolescent Idiopathic Scoliosis Treatment Using a Robotic Brace," *Journal of Intelligent & Robotic Systems* 109 (2023): 73, <https://doi.org/10.1007/s10846-023-02010-1>.
19. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (The MIT Press, 2018).
20. M. P. Deisenroth, "A Survey on Policy Search for Robotics," *Foundations and Trends in Robotics* 2, no. 1–2 (2013): 1–142, <https://doi.org/10.1561/23000000021>.
21. J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement Learning in Robotics: A Survey," *International Journal of Robotics Research* 32, no. 11 (2013): 1238–1274, <https://doi.org/10.1177/0278364913495721>.
22. P. Kormushev, S. Calinon, and D. Caldwell, "Reinforcement Learning in Robotics: Applications and Real-World Challenges," *Robotics* 2, no. 3 (2013): 122–148, <https://doi.org/10.3390/robotics2030122>.
23. K. Chatzilygeroudis, V. Vassiliades, F. Stulp, S. Calinon, and J. B. Mouret, "A Survey on Policy Search Algorithms for Learning Robot Controllers in a Handful of Trials," *IEEE Transactions on Robotics* 36, no. 2 (2020): 328–347, <https://doi.org/10.1109/TRO.2019.2958211>.
24. K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," *IEEE Signal Processing Magazine* 34, no. 6 (2017): 26–38, <https://doi.org/10.1109/MSP.2017.2743240>.
25. T. Erol, A. F. Mendi, and D. Doğan, "The Digital Twin Revolution in Healthcare," *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* 22–24, (2020): 1–7, <https://doi.org/10.1109/ISMSIT50672.2020.9255249>.
26. B. R. Barricelli, E. Casiraghi, and D. Fogli, "A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications," *IEEE Access* 7 (2019): 167653–167671, <https://doi.org/10.1109/ACCESS.2019.2953499>.
27. H. X. Nguyen, R. Trestian, D. To, and M. Tatipamula, "Digital Twin for 5G and Beyond," *IEEE Communications Magazine* 59, no. 2 (2021): 10–15, <https://doi.org/10.1109/MCOM.001.2000343>.
28. M. Grieves, "Digital Twin: Manufacturing Excellence Through Virtual Factory Replication," White Paper (2014): 1, <https://www.3ds.com/fileadmin/PRODUCTS-SERVICES/DELMIA/PDF/Whitepaper/DELMIA-APRISO-Digital-Twin-Whitepaper.pdf>.
29. A. Thelen, X. Zhang, O. Fink, et al., "A Comprehensive Review of Digital Twin—Part 1: Modeling and Twinning Enabling Technologies," *Structural and Multidisciplinary Optimization* 65, no. 12 (2022): 354, <https://doi.org/10.1007/s00158-022-03425-4>.
30. A. Thelen, X. Zhang, O. Fink, et al., "A Comprehensive Review of Digital Twin—Part 2: Roles of Uncertainty Quantification and Optimization, a Battery Digital Twin, and Perspectives," *Structural and Multidisciplinary Optimization* 66, no. 1 (2022): 1, <https://doi.org/10.1007/s00158-022-03410-x>.
31. W. Wang, Y. He, F. Li, J. Li, J. Liu, and X. Wu, "Digital Twin Rehabilitation System Based on Self-Balancing Lower Limb

Exoskeleton,” *Technology and Health Care* 31, no. 1 (2023): 103–115, <https://doi.org/10.3233/THC-220087>.

32. M. M. R. Khan, M. S. H. Sunny, T. Ahmed, et al., “Development of a Robot-Assisted Telerehabilitation System With Integrated IIoT and Digital Twin,” *IEEE Access* 11 (2023): 70174–70189, <https://doi.org/10.1109/ACCESS.2023.3291803>.

33. F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, “Digital Twin in Industry: State-of-the-Art,” *IEEE Transactions on Industrial Informatics* 15, no. 4 (2019): 2405–2415.

34. H. D. Taghirad, *Parallel Robots: Mechanics and Control* (CRC Press, 2013), <https://doi.org/10.1201/b16096>.

35. A. Ahmad, N. Abu Osman, H. Mokhtar, W. Mehmood, and N. A. Kadri, “Analysis of the Interface Pressure Exerted by the Chêneau Brace in Patients With Double-Curve Adolescent Idiopathic Scoliosis,” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 233, no. 9 (2019): 901–908, <https://doi.org/10.1177/0954411919856144>.

36. V. M. Pham, A. Houilliez, A. Schill, A. Carpentier, B. Herbaux, and A. Thevenon, “Study of the Pressures Applied by a Chêneau Brace for Correction of Adolescent Idiopathic Scoliosis,” *Prosthetics & Orthotics International* 32, no. 3 (2008): 345–355, <https://doi.org/10.1080/03093640802016092>.

37. J. C. Gesbert, B. Colobert, L. Rakotomanana, and P. Violas, “Idiopathic Scoliosis and Brace Treatment: An Innovative Device to Assess Corrective Pressure,” *Computer Methods in Biomechanics and Biomedical Engineering* 24, no. 2 (2021): 131–136, [1080/10255842.2020.1813729](https://doi.org/10.1080/10255842.2020.1813729).

38. F. K. Fuss, A. Ahmad, A. M. Tan, R. Razman, and Y. Weizman, “Pressure Sensor System for Customized Scoliosis Braces,” *Sensors* 21, no. 4 (2021): 1153, <https://doi.org/10.3390/s21041153>.

39. A. S. Anand, G. Zhao, H. Roth, and A. Seyfarth. “A Deep Reinforcement Learning Based Approach Towards Generating Human Walking Behavior With a Neuromuscular Model.” in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, Toronto, ON, Canada (2019), 537–543, <https://doi.org/10.1109/Humanoids43949.2019.9035034>.

40. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. “Proximal Policy Optimization Algorithms.” (2017), <https://doi.org/10.48550/arXiv.1707.06347>.

41. A. Raffin, A. Hill, M. Ernestus, A. Gleave, A. Kanervisto, and N. Dormann. *Stable Baselines3* (2023), <https://github.com/DLR-RM/stable-baselines3>.

42. A. Chawla, S. Mukherjee, and B. Karthikeyan, “Mechanical Properties of Soft Tissues in the Human Chest, Abdomen and Upper Extremities,” *Institution of Engineers, Journal of Mechanical Engineering, Technical Report* (2013): 1.

43. P. T. Boggs and J. W. Tolle, “Sequential Quadratic Programming,” *Acta Numerica* 4 (1995): 1–51, <https://doi.org/10.1017/S0962492900002518>.

44. H. F. N. Al-Shuka, S. Leonhardt, W. H. Zhu, R. Song, C. Ding, and Y. Li, “Active Impedance Control of Bioinspired Motion Robotic Manipulators: An Overview,” *Applied Bionics and Biomechanics* 2018 (2018): 8203054, <https://doi.org/10.1155/2018/8203054>.

45. A. Y. Mersha, S. Stramigioli, and R. Carloni. “Variable Impedance Control for Aerial Interaction,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2014): 3435–3440, <https://doi.org/10.1109/IROS.2014.6943041>.

46. X. Zhang, L. Sun, Z. Kuang, and M. Tomizuka, “Learning Variable Impedance Control via Inverse Reinforcement Learning for Force-Related Tasks,” *IEEE Robotics and Automation Letters* 6, no. 2 (2021): 2225–2232, <https://doi.org/10.1109/LRA.2021.3061374>.

47. L. Roveda, J. Maskani, P. Franceschi, et al., “Model-Based Reinforcement Learning Variable Impedance Control for Human-Robot Collaboration,” *Journal of Intelligent & Robotic Systems* 100 (2020): 417–433, <https://doi.org/10.1007/s10846-020-01183-3>.