



# Reinforcement learning for safe autonomous two-device navigation of cerebral vessels in mechanical thrombectomy

Harry Robertshaw<sup>1</sup> · Benjamin Jackson<sup>1</sup> · Jiaheng Wang<sup>1</sup> · Hadi Sadati<sup>1</sup> · Lennart Karstensen<sup>2</sup> · Alejandro Granados<sup>1</sup> · Thomas C. Booth<sup>1,3</sup>

Received: 22 January 2025 / Accepted: 6 March 2025 / Published online: 3 April 2025  
© The Author(s) 2025

## Abstract

**Purpose** Autonomous systems in mechanical thrombectomy (MT) hold promise for reducing procedure times, minimizing radiation exposure, and enhancing patient safety. However, current reinforcement learning (RL) methods only reach the carotid arteries, are not generalizable to other patient vasculatures, and do not consider safety. We propose a safe dual-device RL algorithm that can navigate beyond the carotid arteries to cerebral vessels.

**Methods** We used the Simulation Open Framework Architecture to represent the intricacies of cerebral vessels, and a modified Soft Actor-Critic RL algorithm to learn, for the first time, the navigation of micro-catheters and micro-guidewires. We incorporate patient safety metrics into our reward function by integrating guidewire tip forces. Inverse RL is used with demonstrator data on 12 patient-specific vascular cases.

**Results** Our simulation demonstrates successful autonomous navigation within unseen cerebral vessels, achieving a 96% success rate, 7.0 s procedure time, and 0.24 N mean forces, well below the proposed 1.5 N vessel rupture threshold.

**Conclusion** To the best of our knowledge, our proposed autonomous system for MT two-device navigation reaches cerebral vessels, considers safety, and is generalizable to unseen patient-specific cases for the first time. We envisage future work will extend the validation to vasculatures of different complexity and on in vitro models. While our contributions pave the way toward deploying agents in clinical settings, safety and trustworthiness will be crucial elements to consider when proposing new methodology.

**Keywords** Reinforcement learning · Mechanical thrombectomy · Machine learning · Artificial intelligence · Autonomous navigation · Endovascular intervention

## Introduction

Ischemic stroke causes 3.48 million global deaths annually and accounts for \$36.5 billion in direct medical costs in the USA alone [1, 2]. Mechanical thrombectomy (MT) has emerged as a standard treatment for acute ischemic stroke resulting from large vessel occlusion, providing improved

functional outcomes and lower mortality rates compared to medical treatment alone [3, 4]. In MT, a guidewire is used to navigate a guide catheter from the femoral or radial artery to the internal carotid artery (ICA). An ‘access catheter’ is then advanced ahead of the guide catheter for vessel branch access. Once the access catheter is within the ICA, the guide catheter is advanced to make a stable platform. The guidewire and access catheter are then retracted, and a micro-guidewire within a micro-catheter is passed through the stable guide catheter and navigated to the target thrombus site which is typically within the M1 segment of the middle cerebral artery (MCA). The final step is to remove the micro-guidewire and exchange it for a stent retriever to remove the thrombus, thereby restoring blood flow to the brain.

MT is most effective when performed early, ideally within 24 h of stroke onset [4]. However, only 3.1% of stroke patients in the UK receive MT, despite 10% being eligible [5, 6]. Lim-

✉ Thomas C. Booth  
thomas.booth@kcl.ac.uk

<sup>1</sup> Surgical and Interventional Engineering, School of Biomedical Engineering and Imaging Sciences, Kings College London, London, UK

<sup>2</sup> AIBE, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

<sup>3</sup> Department of Neuroradiology, Kings College Hospital, London, UK

ited access to MT centers, long travel times, and procedural risks like vessel damage and radiation exposure pose significant challenges for the delivery of MT [7, 8]. Robotic surgical systems offer a potential solution by increasing access to MT and reducing operator risks. Tele-operated MT could allow centralized, highly trained specialists to perform the procedure remotely, improving access in underserved regions. Alternatively, assistive robots in peripheral hospitals could help less experienced operators perform MT safely. In either scenario, integrating AI into robotic systems may further enhance procedural efficiency and safety, especially for less experienced users (e.g., interventional radiologists who are not used to performing neurointerventional procedures) [9, 10].

A recent review on the autonomous navigation of endovascular interventions found that 71% (10/14) of studies performed experiments in the blood vessels within or around the heart [11]. More recently, a combined reward model using demonstrator data in inverse reinforcement learning (IRL) has been shown to provide improved success rates over a traditional dense reward function in the autonomous navigation of devices from the common iliac artery to the ICA; however, this work was limited to training and testing on exactly the same vasculature [12]. While reinforcement learning (RL) models have been tested on unseen vessel structures previously (for aortic arch navigation), these structures are typically simplified and computer generated [13, 14]. Furthermore, while 29% (4/14) of studies reviewed in [11] measure the forces during navigation, there is no evidence of excessive force penalization. This is important to consider because (1) excessive vessel wall contact forces can induce vasoconstriction and damage, leading to reactive intimal proliferation chronically, or more importantly, dissection-induced distal embolization and infarction in the acute setting [15]; (2) it is known that robotic endovascular manipulators can reduce vessel wall contact forces [16]; and (3) patient and public involvement exercises on robotic MT showed concern about micro-guidewire safety—AI assistance was a potential solution for safety.

The aim of this study was to train an RL model capable of autonomously navigating a micro-guidewire and micro-catheter from the ICA to the MCA in unseen patient vasculatures while incorporating device tip forces. The primary objective was to demonstrate in an MT environment that RL can successfully perform this navigation in complex, unseen patient vasculatures. The secondary objectives were: (1) to assess whether adding force feedback in the reward function improves surrogate measures of patient safety and/or affects navigation performance and (2) to validate previous findings that a combined reward model enhances navigation performance. This study used two-dimensional Cartesian tracking coordinates as inputs to the RL agent. Using this proposed method, it is plausible that effective clinical translation would

be enacted using image-tracking to obtain the guidewire and catheter tip points from live fluoroscopic images during MT [17]. This would then provide the inputs (i.e., two-dimensional coordinates) needed for the RL agent to navigate the vasculature. Furthermore, as tip forces are only required during training as part of the reward function, there would be no need to measure forces during the intervention. As the trained RL agent does not need a reward input, no additional equipment or software would be required in the clinic.

Our contributions are as follows: (1) We applied RL to train autonomous navigation models for micro-guidewire and micro-catheter in cerebral vessels, specifically targeting MT, for the first time, (2) we tested the model on realistic, unseen patient vasculatures, and (3) we introduced penalization for instrument forces in the RL model to enhance patient safety.

## Methods

### Navigation task

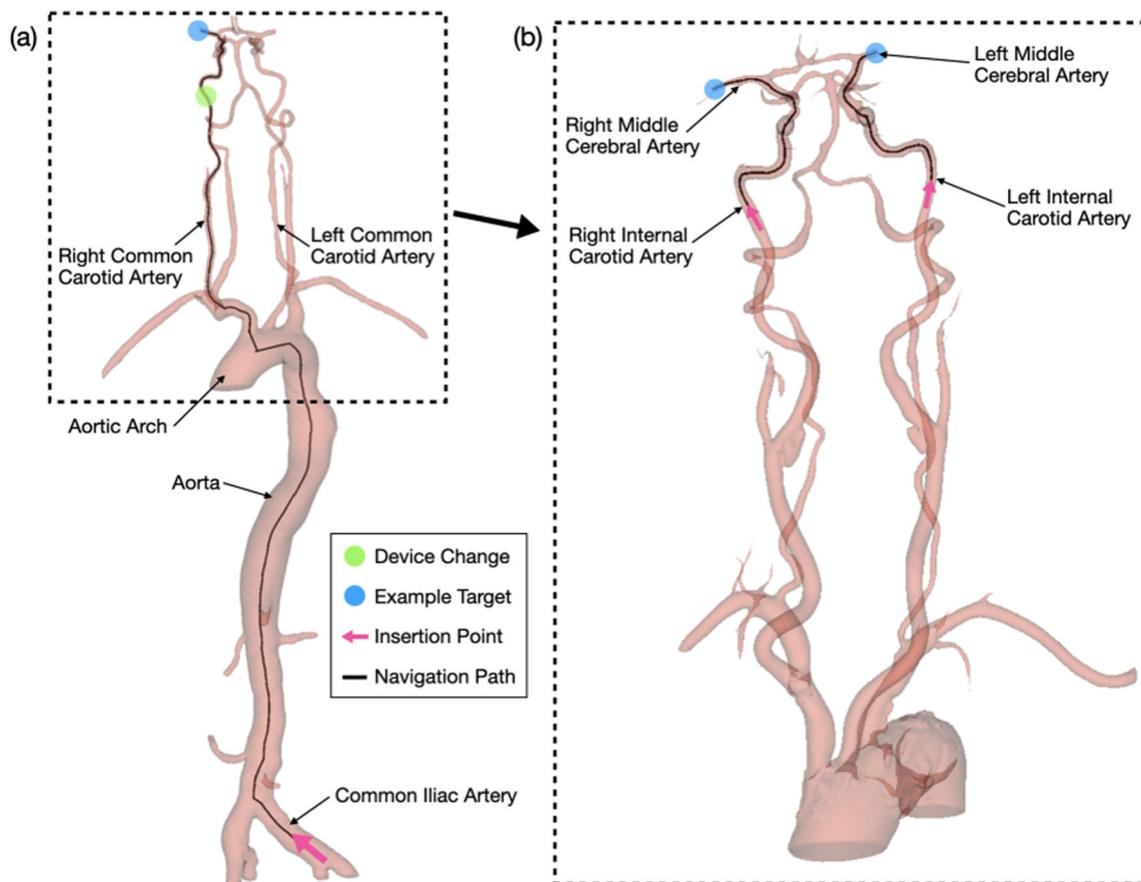
The first MT stage of navigating a large guide catheter to form a stable platform has recently been achieved autonomously using RL [12]. This study simulates the second stage, navigating the micro-guidewire and micro-catheter to the target site (Fig. 1). When these devices are manipulated toward the MCA, navigation challenges include overcoming tortuous anatomy of the ICA and avoiding catheterization of incorrect branches such as the ipsilateral anterior cerebral artery. Rotation of the micro-guidewire (whose tip is curved) is required to avoid catheterization of incorrect branches, while both the tortuous anatomy and rotation cause the devices to interact with one another.

A target location was randomly sampled from ten centerline points in the right MCA (RMCA) or the left MCA (LMCA), starting from a stable guide catheter platform in the right ICA (RICA) or the left ICA (LICA) cervical segments [18]. Targets in the RMCA could only be navigated from the RICA and vice-versa for the left side. For each navigation attempt, the target location and the vasculature were changed.

The device position was described by three points equally spaced 2 mm apart along the device. These were denoted as  $(x', y')_{i=1,2,3}$ , with  $(x', y')_1$  representing the Euclidean instrument tip coordinates. The target location was specified by the target's  $(x', y')$  coordinates. Observations comprised current and previous device positions, target location, and the previous action taken.

### Simulation environment

The in silico environment for the navigation task builds on previous work from [10, 12, 13] and utilized the stEVE



**Fig. 1** Example navigation path taken for **a** entire MT navigation path and **b** micro-guidewire and micro-catheter component of navigation path

framework [19]. The BeamAdapter plugin for Simulation Open Framework Architecture (SOFA) was used to model the Echelon 10 (1.7 F) micro-catheter (Medtronic, Minnesota, USA) and the Synchro 14 micro-guidewire (Stryker NeuroVascular, Cork, Ireland), characterized by Young's moduli of 47 and 43 MPa, respectively [20, 21]. To represent the rigid distal ICA and the inflexible M1, the simulation assumed rigid vessel walls (with an empty lumen). The simulation's fidelity to real-world guidewire behavior was ensured by utilizing a tensile testing machine to measure the tensile strength of the devices, which facilitated the calculation of its stiffness, while friction between wall and guidewire has been iteratively tuned to mimic guidewire behavior in a test-bench setup. This is based on previous methodology which has previously allowed in silico to ex vivo translation of autonomous endovascular navigation designs using models of porcine liver vasculature [22].

Input parameters, including guidewire rotation and translation speed, were applied at the proximal device end. The rotational and translational speed was constrained to a maximum of  $180^\circ\text{s}^{-1}$  and  $40\text{mm}\text{s}^{-1}$ , respectively. Similar to a clinical scenario with fluoroscopy, feedback during the navigation was given as two-dimensional ( $x'$ ,  $y'$ ) track-

ing coordinates of three points along each device's tip; no visual information showing the geometry of the patient vasculature was given. Experiments were performed using a dual-tracking (combined wire and catheter tracking) method [12].

## Dataset

Twelve computed tomography angiography (CTA) scans which encompassed the aortic arch and extended to the cerebral vessels (obtained with UK Research Ethics Committee 20/ES/0005) were processed into three-dimensional surface meshes using the vascular modeling toolkit and segmentation tools in 3D Slicer [23, 24]. Characteristics of each vasculature can be found in Table 3, and CTA scan parameter settings can be found in Appendix B. These were loaded into SOFA, which allowed augmentation to be applied during training via random scaling (0.7 to 1.3 for height and width) to enhance generalization.

Demonstrator data were collected in silico for two random targets in the RMCA and LMCA across ten patient vasculatures, with four navigation's per vasculature using inputs from a keyboard. The demonstrator was a robotics

engineer with 2-years' experience using in silico navigation environments and in vitro endovascular robots, with 2 years of tuition from an expert interventional neuroradiologist (UK consultant; US attending equivalent). Data collected included actions (rotation and translation), device tracking, and target location. Data were split by left or right target navigation.

## Controller architecture

All RL models in this study were trained using a Soft Actor-Critic (SAC) controller, implemented using PyTorch [25]. This is shown in Fig. 2 and is adapted from previous work [12, 13]. The architecture includes a Long Short-Term Memory layer for learning trajectory-dependent state representations, which has been shown to allow probing of the correct vessel when the target branch is not unambiguously located from the target coordinates [13]. The architecture also includes feedforward layers for controlling the guidewire. The controller takes observations as input, and a Gaussian policy network outputs mean ( $\mu$ ) and standard deviation ( $\sigma$ ) for expected actions, representing the micro-catheter's and micro-guidewire's rotation and translation. During training, actions are sampled from the  $\sigma$ , but for evaluation,  $\mu$  is used directly for deterministic behavior.

Maximum entropy IRL was employed to learn reward functions from demonstrations for navigating vascular branches, aiming to replicate expert behavior while maximizing entropy to account for variability [26]. This method, suited for complex environments like endovascular navigation, can outperform methods such as behavioral cloning and standard RL, which may overfit or struggle with sparse rewards and has provided improvements over the state of the art in autonomous endovascular interventions [12]. A feed-forward neural network with four fully connected layers was used in the IRL model, trained with expert trajectories. Training of the IRL model involved 1 million iterations with a learning rate of  $1 \times 10^{-2}$ , entropy regularization ( $\alpha = 0.1$ ), and a 0.01 length penalty. Separate IRL models were trained for each vascular branch, with the ipsilateral model returning reward values for the current observations.

## Reward functions

Six reward functions were used to assess whether adding micro-guidewire tip force feedback improved in silico results. Each reward function was calculated in every simulation step based on the agent's actions. Therefore, actions led to different reward quantities across reward functions and, hence, a variation in the final model. The first three reward functions followed prior work [12]:

- $R_1$  (Eq. 1): Dense reward, where *pathlength* is the distance between the guidewire tip and the target, and  $\Delta\text{pathlength}$  is the change from the previous step.

$$R_1 = -0.005 - 0.001 \cdot \Delta\text{pathlength} + \begin{cases} 1.0 & \text{if target reached} \\ 0 & \text{else} \end{cases} \quad (1)$$

- $R_2$  (Eq. 2): IRL-derived reward, where  $R_{RICA}$  or  $R_{LICA}$  is calculated based on the carotid artery the target is in.

$$R_2 = \begin{cases} R_{RICA} & \text{if target in RICA} \\ R_{LICA} & \text{if target in LICA} \end{cases} \quad (2)$$

- $R_3$  (Eq. 3): A combined reward model of  $R_1$  and  $R_2$ , scaled by  $\alpha = 0.001$ .

$$R_3 = R_1 + \alpha R_2 \quad (3)$$

Three additional reward functions ( $R_4$ ,  $R_5$ , and  $R_6$ ) were created by adding force feedback to  $R_1$ ,  $R_2$ , and  $R_3$ , achieved by implementing a collision monitor on the micro-guidewire tip, to give  $P_{x,y,z}$  at each simulation step. Analysis of demonstrator data guidewire tip forces showed that  $\|P\|$  values remained at 0.80 N or less; therefore, if  $\|P\|$  was greater than a threshold value of 0.85 N, a negative reward was given. This was added to  $R_1$ ,  $R_2$ , and  $R_3$ , to give  $R_4$ ,  $R_5$ , and  $R_6$ , as shown in Eq. 4. The reward functions were fine-tuned iteratively through RL training by systematically adjusting the reward weights and  $\alpha$ , with each adjustment followed by evaluation on key metrics of success rate, procedure time, and tip force reduction. As guidewire tip forces are only used in the reward function during training, the agent does not require a force input during in silico evaluation or any experimental stages beyond this (e.g., in vitro evaluation).

$$R_j = R_{j-3} - 0.01 \cdot \begin{cases} P - 0.85 & \text{if } \|P\| > 0.85 \\ 0 & \text{if } \|P\| \leq 0.85 \end{cases} \quad \text{for } j = 4, 5, 6 \quad (4)$$

## Implementation

Training consisted of  $1 \times 10^7$  exploration steps, with each navigation task (or episode) considered complete when the target was reached within a 5 mm threshold, which is suitable to cover the entire cross section of the vessel. A timeout of 200 steps (approximately 27 s) was set for computational efficiency. Training was performed on an NVIDIA DGX A100 node (Santa Clara, California, USA) with 8 GPUs and took approximately 48 h per model.

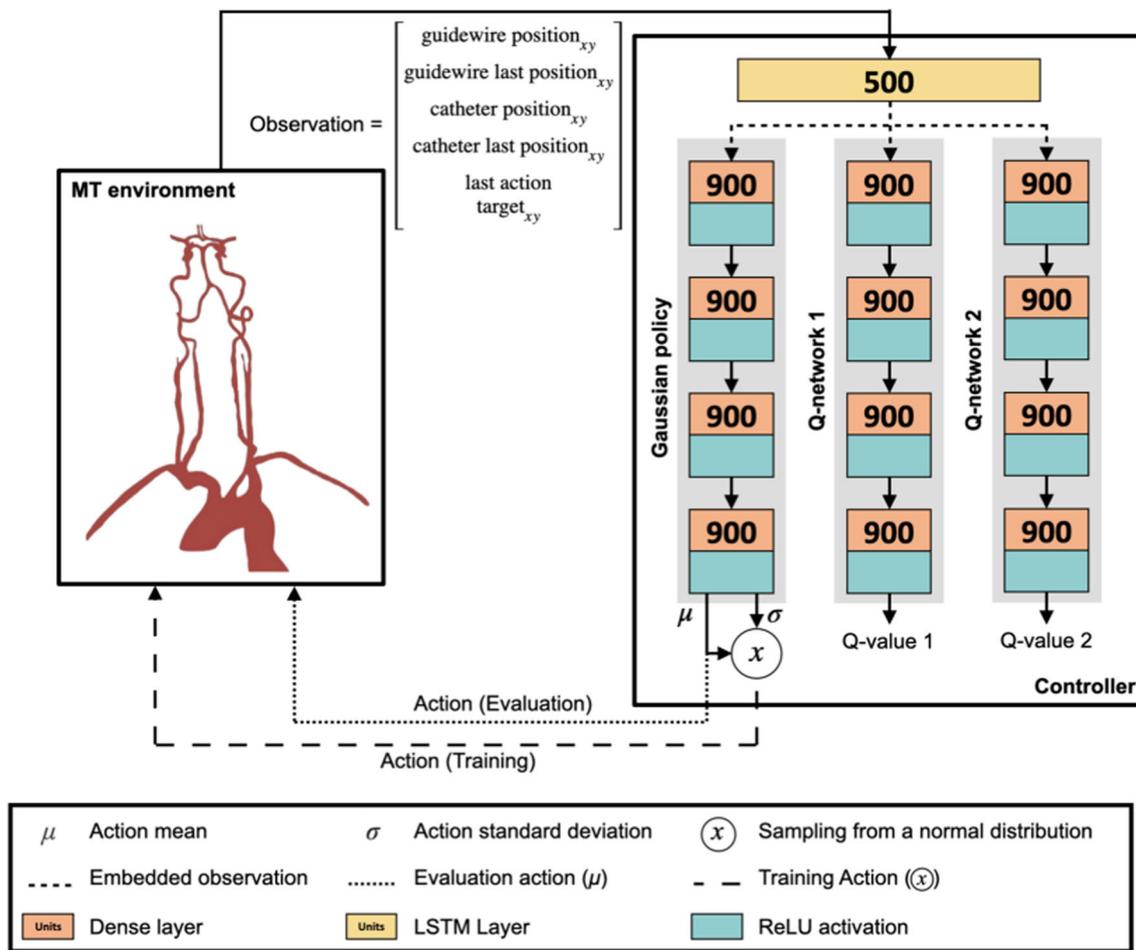


Fig. 2 Architecture of the proposed controller based on a SAC design. Reward and action input for the Q-networks are not included in this diagram. Adapted from [13]

### Experimental design and evaluation

Two experiments were conducted. First, a single patient vasculature was used for both training and testing to replicate and confirm that a combined reward model provides improved success rates over a traditional dense reward function, as in [12], for the second stage of MT. Second, ten different patient vasculatures were used during training with augmentation applied, while two were reserved for testing only (Fig. 3). This created a train/test split of approximately 80/20 [27].

For the first experiment, five targets per branch were used for training and five different targets for evaluation. For the second experiment, all ten targets in each branch were utilized for evaluation. Evaluations were conducted every  $2.5 \times 10^5$  exploration steps for 80 episodes, recording the success rate, procedure time, and path ratio [12]. The models with the highest success rate for each reward function were compared against one another for path ratio, procedure time, mean forces, and exploration steps. Comparative statistical analyses were conducted using two-tailed paired Student’s t

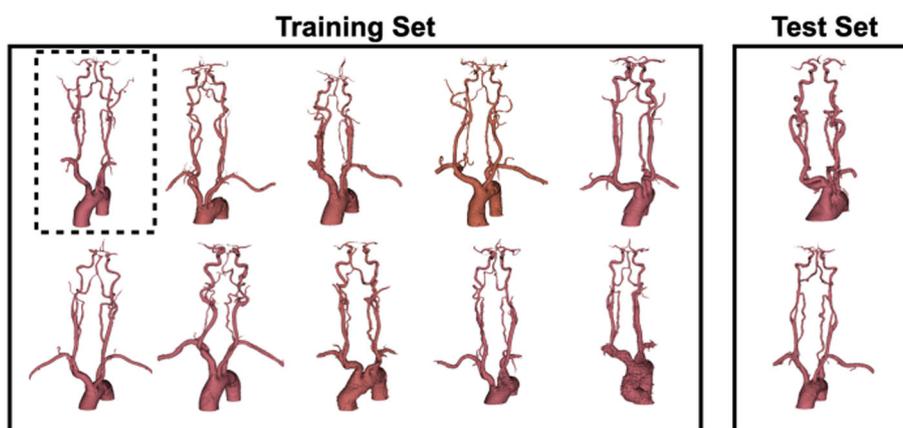
tests and analysis of variance, with a predetermined significance threshold set at  $p = 0.05$ .

### Results

#### Training and testing on the same patient vasculature

Table 1 presents the results of training and testing with reward functions  $R_{1-6}$  on the same patient vasculature. Both  $R_1$  (dense without force feedback) and  $R_3$  (combined model without force feedback) reach success rates of 100%, while  $R_4$  (dense with force feedback) and  $R_6$  (combined model with force feedback) reach 97% and 96%, respectively ( $R_{1,4}$ :  $p = 0.157$ ,  $R_{3,6}$ :  $p = 0.081$ ). Both  $R_2$  (IRL without force feedback) and  $R_5$  (IRL with force feedback) had a 44% success rate. The addition of force feedback was associated with significantly longer procedure times for  $R_2$  to  $R_5$  (8.1 s to 15.7 s [ $p = 0.001$ ]) and  $R_3$  to  $R_6$  (5.0 s to 6.9 s [ $p < 0.001$ ]) and significantly lower mean forces for  $R_1$  to  $R_4$  (0.29 N

**Fig. 3** Patient vasculatures used for training (10 cases) and testing (2 cases). The dashed box represents the single vasculature used for training and testing in the first experiment (Sect. 2.7)



**Table 1** Results of testing and training on the same patient vasculature

Reward function	Success rate (%)	Procedure time (s)	Path ratio (%)	Mean force (N)	Exploration steps
$R_1$	100 ± 0	3.8 ± 0.3	100 ± 0	0.29 ± 0.37	325 × 10 <sup>6</sup>
$R_2$	44 ± 50	8.1 ± 8.6	75.1 ± 23.8	0.30 ± 0.37	125 × 10 <sup>6</sup>
$R_3$	100 ± 0	5.0 ± 2.1	100 ± 0	0.25 ± 0.36	300 × 10 <sup>6</sup>
$R_4$	97 ± 16	3.8 ± 1.0	96 ± 4.5	0.25 ± 0.36	175 × 10 <sup>6</sup>
$R_5$	44 ± 50	15.6 ± 9.4	80.0 ± 18.9	0.28 ± 0.36	75 × 10 <sup>6</sup>
$R_6$	96 ± 19	6.9 ± 2.4	94.6 ± 7.3	0.24 ± 0.35	350 × 10 <sup>6</sup>

Values are reported as mean ± standard deviation. *Success Rate*: percentage of evaluation episodes where target is reached. *Path Ratio*: percentage distance navigated to target in unsuccessful episodes, calculated by dividing the total distance navigated toward the target by the initial distance. *Procedure Time*: time from the start of navigation to the target location for successful episodes. *Mean Force*: mean magnitude of tip forces during evaluation. *Exploration Steps*: number of training steps taken to reach the point at which the results are provided. The reported values are mean ± standard deviation (standard deviation values may exceed logical bounds (0–100%) due to the statistical calculation)

to 0.25 N [ $p < 0.001$ ]) and  $R_2$  to  $R_5$  (0.30 N to 0.28 N [ $p = 0.044$ ]). Procedural time for  $R_1$  to  $R_4$  (3.8 s to 3.8 s [ $p = 0.794$ ]) and mean force for  $R_3$  to  $R_6$  (0.25 N to 0.24 N [ $p = 0.249$ ]) were both not significantly different. The maximum force reported across all evaluations was 1.0 N or less.

### Testing on unseen patient vasculatures

Table 2 shows the results of testing reward functions  $R_1$ ,  $R_3$ ,  $R_4$ , and  $R_6$  on two unseen patient vasculatures. Overall, there was slight degradation in outcomes except for mean force when compared to the initial ‘single vasculature training/testing’ outcomes. A notable exception was  $R_6$ , which showed that when testing on unseen vasculatures was compared to the initial ‘single vasculature training/testing’ scenario, results remained consistent across success rate (96% to 96% [ $p = 0.987$ ]), procedure time (6.9 s to 7.0 s [ $p = 0.754$ ]), path ratio (94.6% to 93.5% [ $p = 0.463$ ]), and mean force (0.24 N to 0.24 N [ $p = 0.947$ ]).  $R_6$  achieved the highest success rate of 96%, followed by 93% for  $R_1$  ( $p = 0.306$ ). In contrast to the initial ‘single vasculature training/testing’ scenario, force feedback led to decreased procedure times for both  $R_1$  to  $R_4$  (8.1 s to 4.5 s,  $p = 0.004$ )

and  $R_3$  to  $R_6$  (13.6 s to 7.0 s,  $p < 0.001$ ). Mean forces decreased significantly when force feedback was included for  $R_1$  to  $R_4$  (0.29 N to 0.24 N [ $p < 0.001$ ]) and  $R_3$  to  $R_6$  (0.27 N to 0.24 N [ $p = 0.017$ ]). Illustrative differences in tip forces during navigation for  $R_1$  and  $R_6$  can be seen in Fig. 4. The maximum force reported across all evaluations was 1.0 N or less.

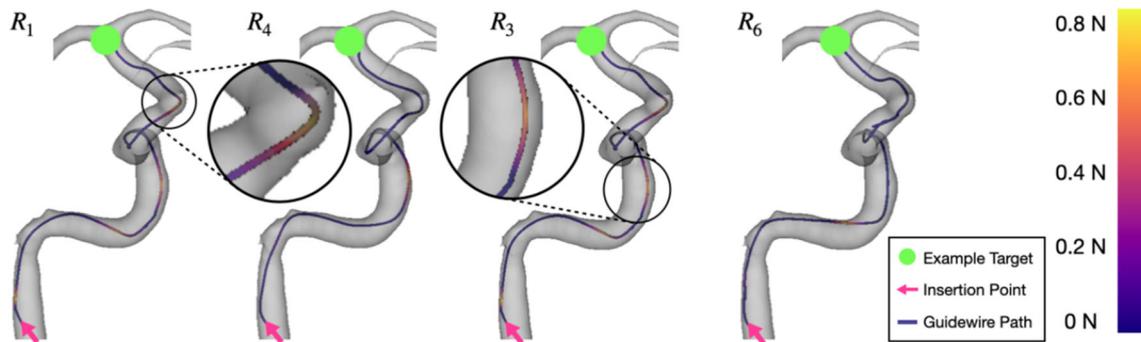
### Discussion

This study is the first to achieve autonomous navigation of both micro-guidewire and micro-catheter during the second stage of MT, from the ICA to the MCA. It builds upon previous research by evaluating the reward function used by [12], while testing on unseen patient vasculatures. This proof-of-concept in silico study also demonstrated that incorporating tip force feedback into the reward function is feasible in the MT second stage scenario shown here and may reduce mean forces on vasculature. Furthermore, when force feedback is incorporated into a combined reward model and tested on unseen vasculatures, improved success rates and reduced procedure times are seen. The same model appeared

**Table 2** Results of testing on two unseen vasculatures after training on ten different vasculatures

Reward function	Success rate (%)	Procedure time (s)	Path ratio (%)	Mean force (N)	Exploration steps
$R_1$	93 ± 27	8.1 ± 8.8	89.6 ± 23.8	0.29 ± 0.37	400 × 10 <sup>6</sup>
$R_3$	88 ± 33	13.6 ± 10.4	88.1 ± 23.6	0.27 ± 0.36	625 × 10 <sup>6</sup>
$R_4$	78 ± 42	4.5 ± 3.0	84.1 ± 28.7	0.24 ± 0.35	200 × 10 <sup>6</sup>
$R_6$	96 ± 19	7.0 ± 3.9	93.5 ± 15.1	0.24 ± 0.35	700 × 10 <sup>6</sup>

Table 1 concludes IRL's ineffectiveness compared to other reward functions, and hence,  $R_{2,5}$  were omitted when testing on unseen vasculatures



**Fig. 4** Navigation path to RMCA for  $R_1$  (dense without force feedback),  $R_3$  (combined model without force feedback),  $R_4$  (dense with force feedback), and  $R_6$  (combined model with force feedback), with example regions of higher tip forces indicated

generalizable with no domain shift when transitioning from training/testing on a single vasculature to training on multiple vasculatures and testing on unseen ones. While the current autonomous MT navigation system is at technology readiness level (TRL) 3 [28], the first successful autonomous navigation of the second stage of MT on unseen real patient anatomies might contribute toward increasing the TRL, a step forward to realizing the benefits of fully autonomous MT.

### IRL reward function

Results from training and testing on the same vasculature aligned with previous work [12] and showed high success rates for  $R_{1,4}$  and  $R_{3,6}$ , while  $R_{2,5}$  had a lower success rate. These initial results concluded IRL's ineffectiveness compared to the other reward functions, leading to the omission of  $R_2$  and  $R_5$  in subsequent testing on unseen vasculatures. This ineffectiveness may stem from the inherent difficulty in capturing the nuances of expert decision making in highly dynamic environments like endovascular navigation, where factors such as subtle anatomical differences and force feedback cues play a role.

Training RL models, particularly in IRL settings, remains a considerable challenge. The dual task of learning a reward function from demonstrations and optimizing a policy introduces additional complexity. Our findings highlight the sensitivity of RL approaches to the reward function, which impact convergence speed as exploration steps increased with

the addition of IRL when moving from  $R_1$  to  $R_3$  ( $400 \times 10^6$  to  $625 \times 10^6$ ) and from  $R_4$  to  $R_6$  ( $200 \times 10^6$  to  $700 \times 10^6$ ).

Despite this, expert demonstrator data could plausibly still be useful in enhancing autonomous endovascular navigation, whether this is through a different form of IRL or alternative RL algorithms. Nevertheless, the addition of demonstrator data has been shown to improve RL in different scenarios, while enhancing sample efficiency, helping with sparse rewards, and potentially enabling safer exploration [29], and has effectively been utilized in previous autonomous endovascular navigation work [30].

### Force feedback

A decrease in success rate, and path ratio, as well as an increase in procedure time, and training time, was observed when moving from testing and training on the same vasculature to testing on unseen vasculatures. While this overall slight degradation in outcomes is expected due to the difficult task of navigating a completely unseen vasculature, there were two exceptions. First, mean forces were lower demonstrating safety. Second, a maximum success rate of 96% is promising for the combined model. While the addition of guidewire tip force feedback in the reward function decreased success rate for the dense reward function (93% to 78%), the addition improved the success rate for the combined model (88% to 96%). As there are multiple reward signals for all reward functions examined (tip force, steps taken, path length), the reason for this difference may be that

$R_6$  provides the optimal balance of these signals. By fine-tuning the scaling factors for each reward function, it may be possible to further improve training results. Force feedback also significantly reduced procedure time for unseen vasculatures ( $R_{1,4}$ : 8.1s to 4.5s,  $R_{3,6}$ : 13.6s to 7.0s)—a result not seen when testing and training on the same vasculature. Plausibly, the addition of forces may discourage the tip from applying significant force to the vessel wall which can occur when the tip is caught within the lumen preventing it from moving; the net result might be an increase in overall velocity due to decreased friction and subsequently decrease in procedure time.

Our motivation for including force feedback was to decrease mean forces as a patient safety feature. While its addition decreased the mean forces on the guidewire tip, we acknowledge that the mean forces across all experiments in unseen vasculature were low in any case with a maximum of 0.29 N, well below the vessel rupture threshold of 1.5 N proposed by [10]. Although no study has analyzed instrument forces during RL training as a method for increasing safety during the procedure, previous studies have measured instrument and vessel wall forces as part of their experiments [31], which suggests that our approach could be translated to in vitro environments.

## Limitations

Similar to previous methods [10, 12, 13], in silico work cannot measure fluoroscopy time (a surrogate of radiation exposure) which has been included in some clinical studies [32]. Additionally, using a keypad-operated controller for data collection may not fully reflect the actions of an experienced neurointerventional radiologist. The small dataset of patient vasculatures used for training and testing limits generalizability. Although augmentation introduced a degree of variability, this dataset is not able to replicate the different configurations of the Circle of Willis or branching patterns of the cerebral arteries found in the population, and a larger, more diverse, dataset may be required to ensure effective in vitro translation. As the current work focuses entirely in silico, physical characteristics of guidewires, especially force transmission, may not be fully captured. Hence, future work should aim to understand the impact that training with in silico force feedback has on in vitro models, while providing in vitro validation of the proposed methods. While this work incorporated real-world mechanical data to enhance the accuracy of force models, future work could improve simulation-to-real transfer by modeling vessels as elastic bodies to reflect the behavior of real vessels—however, any added benefit may be imperceptible during the second MT stage used in the current scenario, as the distal ICA is rigid and the M1 is somewhat inflexible. Future work may also compare tracking- and visual-based RL algorithms, to inves-

tigate whether the same benefits can be gained from providing the entire image frame at each simulation step.

## Conclusion

In conclusion, this study proposes an offline RL algorithm that autonomously navigates (micro-guidewires and micro-catheters) to cerebral vessels in unseen patients while considering the amount of force applied to the device tip, representing the first demonstration of autonomous navigation in both cerebral vessels and unseen patient vasculatures. We evaluated various reward functions and confirmed the superior performance of a combined reward model with force feedback ( $R_6$ ), which maintained a 96% success rate in patient vasculatures the model had not seen previously. The addition of force feedback not only reduced procedure times but also decreased the mean forces exerted, plausibly enhancing patient safety by decreasing both time to treatment and the likelihood of iatrogenic vessel damage. Future work should focus on translating the proposed method to in vitro experiments and extending the dataset to integrate more diverse patient vasculatures.

## Appendix A Vasculature characteristics

**Table 3** Characteristics of vasculatures used during the study [33]. Tortuosity is computed as the ratio between two distances: (1) the geodesic length along the centerline of the artery and (2) the Euclidean distance between the two end-points of the artery segment. This index lies in  $[1, +\infty)$ : the closer to 1 the less tortuous. Radius was calculated as the origin

Vasculature no.	Train/test	RICA tortuosity	LICA tortuosity	RICA radius (mm)	RMCA radius (mm)	LICA radius (mm)	LMCA radius (mm)
1	Train	1.16	1.14	2.05	1.21	1.51	2.18
2	Train	1.25	1.18	2.13	1.28	2.48	1.15
3	Train	1.35	1.42	3.46	1.00	3.30	0.77
4	Train	1.16	1.13	0.99	1.19	3.66	1.19
5	Train	1.10	1.24	3.80	1.66	2.49	1.58
6	Train	1.19	1.27	3.10	1.40	1.33	1.02
7	Train	1.47	1.60	2.87	1.78	3.56	1.59
8	Train	1.41	1.33	3.27	1.13	2.83	1.10
9	Train	1.39	1.23	2.36	1.61	4.26	1.51
10	Train	1.27	1.42	3.51	1.64	3.52	1.44
11	Test	1.15	1.11	5.80	0.98	5.35	0.93
12	Test	1.21	1.21	3.33	0.98	3.20	0.84

## Appendix B CTA scan parameter settings

### B.1 Scan Protocol

- *Scanner Type*: Multi-detector CT (Optima 660 GE Healthcare, 64 section)

### B.2 Acquisition parameters

- *Slice Thickness*: 0.625 mm
- *Detector Collimation*: 40.0 mm
- *Pitch*: 0.984 mm
- *Rotation Time*: 0.3–0.5 s (faster for reducing motion artifacts)
- *Field of View (FOV)*: 500 mm

### B.3 Contrast injection

- The acute stroke imaging protocol consisted of craniocervical arterial phase acquisition after intravenous injection of 50 mL of iohexol, 647 mgmL<sup>-1</sup> (5mLs<sup>-1</sup>) (Omnipaque 300; GE Healthcare)
- *Saline Flush*: 40 mL (saline bolus)
- *Bolus Tracking*: Self-triggered—observe contrast and scan when suitable contrast in ascending aorta and carotid artery plus auto-minimum delay setting (15–20 s post-injection for optimal arterial phase)

### B.4 Radiation dose parameters

- *Tube Voltage (kVp)*: 100 kVp
- *Tube Current (mA)*: 315 mA with dose reduction on 40%
- *DLP (Dose Length Product)*: Toward 600 mGy cm

### B.5 Reconstruction parameters

- *Reconstruction Algorithm*: Standard kernel for arteries
- *Post-processing*: 2/5 Maximum intensity projection (MIP)

### B.6 Patient positioning

- Supine position, head immobilized, and neck slightly extended to avoid venous overlap with arteries

**Acknowledgements** The authors would like to thank Han-Ru Wu (Nuffield Department of Primary Care Health Sciences, University of Oxford, UK) for her work on computing the characteristics for the vasculatures used.

**Funding** Partial financial support was received from the WELLCOME TRUST (Grant Agreement No 203148/A/16/Z) and the Engineering and Physical Sciences Research Council Doctoral Training Partnership (Grant Agreement No EP/R513064/1). For the purpose of Open Access,

the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

## Declarations

**Conflict of interest** The authors have no conflict of interest to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, Abbasi-Kangevari M, Abbastabar H, Abd-Allah F, Abdelalim A, Abdollahi M, Abdollahpour I, Abolhassani H, Aboyans V, Abrams EM, Abreu LG, Abrigo MRM, Abu-Raddad LJ, Abushouk AI, Acebedo A et al (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet* vol 396, pp 1204–1222, 10 [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0140673620309259>
2. Tsao CW, Aday AW, Almarzooq ZI, Anderson CA, Arora P, Avery CL, Baker-Smith CM, Beaton AZ, Boehme AK, Buxton AE, Commodore-Mensah Y, Elkind MS, Evenson KR, Eze-Nliam C, Fugar S, Generoso G, Heard DG, Hiremath S, Ho JE, Kalani R et al (2023) “Heart disease and stroke statistics - 2023 update: a report from the American heart association. E93–E621, 2
3. Bendszus M, Fiehler J, Subtil F, Bonekamp S, Aamodt AH, Fuentes B, Gizewski ER, Hill MD, Krajina A, Pierot L, Simonsen CZ, Zelenáček K, Blauenfeldt RA, Cheng B, Denis A, Deutschmann H, Dorn F, Flottmann F, Gellißen S, Gerber JC, Goyal M, Haring J, Herweh C, Hopf-Jensen S, Hua VT, Jensen M, Kastrup A, Keil CF, Klepanec A, Kurča E, Mikkelsen R, Möhlenbruch M, Müller-Hülsbeck S, Münnich N, Pagano P, Papanagiotou P, Petzold GC, Pham M, Puetz V, Raupach J, Reimann G, Ringleb PA, Schell M, Schlemm E, Schönenberger S, Tennøe B, Ulfert C, Vališ K, Vítková E, Vollherbst DF, Wick W, Thomalla G; TENSION Investigators (2023) Endovascular thrombectomy for acute ischaemic stroke with established large infarct: multicentre, open-label, randomised trial. *Lancet* 402(10414):1753–1763. [https://doi.org/10.1016/S0140-6736\(23\)02032-9](https://doi.org/10.1016/S0140-6736(23)02032-9)
4. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, Yavagal DR, Ribo M, Cognard C, Hanel RA, Sila CA, Hassan AE, Millan M, Levy EI, Mitchell P, Chen M, English JD, Shah QA, Silver FL, Pereira VM et al (2018) Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med* 378:11–21
5. McMeekin P, White P, James MA, Price CI, Flynn D, Ford GA (2017) Estimating the number of UK stroke patients eligible for endovascular thrombectomy. *Eur Stroke J* 2:319–326
6. (2023) Sentinel Stroke National Audit Programme, Ssnap annual report 2023,” Healthcare Quality Improvement Partnership, 11

- accessed 23 October 2024. (online). Available: <https://www.hqip.org.uk/resource/ssnap-nov-2023/>
7. Hausegger KA, Schedlbauer P, Deutschmann HA, Tiesenhäuser K (2001) Complications in endoluminal repair of abdominal aortic aneurysms. *Eur J Radiol* 39(1):22–33
  8. Klein LW, Miller DL, Balter S, Laskey W, Haines D, Norbash A, Mauro MA, Goldstein JA (2009) Occupational health hazards in the interventional laboratory: time for a safer environment. *Soc Interv Radiol* 250:538–544
  9. Crinnion W, Jackson B, Sood A, Lynch J, Bergeles C, Liu H, Rhode K, Pereira VM, Booth TC (2022) Robotics in neurointerventional surgery: a systematic review of the literature. *J Neurointerv Surg* 14:539–545
  10. Jackson B, Crinnion W, De Iturrate Reyzaabal M, Robertshaw H, Bergeles C, Rhode K, Booth T (2023) Comparative verification of control methodology for robotic interventional neuroradiology procedures. *Int J Comput Assist Radiol Surg* 18(11):1977–1986. <https://doi.org/10.1007/s11548-023-02991-2>
  11. Robertshaw H, Karstensen L, Jackson B, Sadati H, Rhode K, Ourselin S, Granados A, Booth TC (2023) Artificial intelligence in the autonomous navigation of endovascular interventions: a systematic review. *Front Hum Neurosci* 17 (online). Available: <https://www.frontiersin.org/articles/10.3389/fnhum.2023.1239374/full>
  12. Robertshaw H, Karstensen L, Jackson B, Granados A, Booth TC (2024) Autonomous navigation of catheters and guidewires in mechanical thrombectomy using inverse reinforcement learning. *Int J Comput Assist Radiol Surg* 6 (online). Available: <https://link.springer.com/10.1007/s11548-024-03208-w>
  13. Karstensen L, Ritter J, Hatzl J, Ernst F, Langejürgen J, Uhl C, Mathis-Ullrich F (2023) Recurrent neural networks for generalization towards the vessel geometry in autonomous endovascular guidewire navigation in the aortic arch. *Int J Comput Assist Radiol Surg* 18(9):1735–1744. <https://doi.org/10.1007/s11548-023-02938-7>
  14. Scarponi V, Duprez M, Nageotte F et al (2024) A zero-shot reinforcement learning strategy for autonomous guidewire navigation. *Int J CARS* 19:1185–1192. <https://doi.org/10.1007/s11548-024-03092-4>
  15. Takashima K, Shimomura R, Kitou T, Terada H, Yoshinaka K, Ikeuchi K (2007) Contact and friction between catheter and blood vessel. *Tribol Int* 40:319–328
  16. Rafii-Tari H, Riga CV, Payne CJ, Hamady MS, Cheshire NJ, Bicknell CD, Yang GZ (2016) Reducing contact forces in the arch and supra-aortic vessels using the Magellan robot. *J Vasc Surg* 11:1422–1432
  17. Eyberg C, Karstensen L, Pusch T, Horsch J, Langejürgen J (2022) A ros2-based testbed environment for endovascular robotic systems. *Curr Direct Biomed Eng* 7:89–92
  18. Shapiro M, Becks T, Riina HA, Raz E, Zumofen D, Jafar JJ, Huang PP, Nelson PK (2014) Toward an endovascular internal carotid artery classification system. *Am J Neuroradiol* 35:230–236
  19. Karstensen L, Robertshaw H, Hatzl J, Jackson B, Langejürgen J, Breininger K, Uhl C, Sadati SMH, Booth T, Bergeles C, Mathis-Ullrich F (2024) Learning-based autonomous navigation, benchmark environments and simulation framework for endovascular interventions. *arXiv preprint*, 10 (online). Available: <http://arxiv.org/abs/2410.01956>
  20. Duriez C, Cotin S, Lenoir J, Neumann P (2006) New approaches to catheter navigation for interventional radiology simulation. *Comput Aided Surg* 11:300–308
  21. Faure F, Duriez C, Delingette H, Allard J, Gilles B, Marchesseau S, Talbot H, Courtecuisse H, Bousquet G, Peterlik I, Cotin S (2012) SOFA, a multi-model framework for interactive physical simulation. Springer, Berlin (online). Available: [https://doi.org/10.1007/8415\\_2012\\_125](https://doi.org/10.1007/8415_2012_125)
  22. Karstensen L, Ritter J, Hatzl J, Pätz T, Langejürgen J, Uhl C, Mathis-Ullrich F (2022) Learning-based autonomous vascular guidewire navigation without human demonstration in the venous system of a porcine liver. *Int J Comput Assist Radiol Surg* 17:2033–2040
  23. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R (2012) 3d slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 30:1323–1341
  24. Piccinelli M, Veneziani A, Steinman DA, Remuzzi A, Antiga L (2009) A framework for geometric analysis of vascular structures: application to cerebral aneurysms. *IEEE Trans Med Imaging* 28:1141–1155
  25. Paszke A, Gross S, Massa F, Lerer A, Google JB, Chanan G, Killeen T, Lin Z, Gimeshain N, Antiga L, Desmaison A, Xamla AK, Yang E, Devito Z, Nabla MR, Tejani A, Chilamkurthy S, Ai Q, Steiner B, Facebook LF, Facebook JB, Chintala S (2019) Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 32:8024–8035
  26. Ziebart BD, Maas A, Bagnell JA, Dey AK (2008) Maximum entropy inverse reinforcement learning. In: Proceedings of the twenty-third AAAI conference on artificial intelligence, pp 1433–1438
  27. Gholamy A, Kreinovich V, Kosheleva O (2018) A pedagogical explanation of a pedagogical explanation part of the computer sciences commons. Departmental Technical Reports (CS), 1209
  28. Mankins J (1995) Technology readiness level—a white paper. NASA, Office of Space Access and Technology, Advanced Concepts Office 4 [Online]. Available: <https://www.researchgate.net/publication/247705707>
  29. Ramírez J, Yu W, Perrusquía A (2022) Model-free reinforcement learning from expert demonstrations: a survey. *Artif Intell Rev* 55:3213–3241
  30. Kweon J, Kim K, Lee C, Kwon H, Park J, Song K, Kim YI, Park J, Back I, Roh JH, Moon Y, Choi J, Kim YH (2021) Deep reinforcement learning for guidewire navigation in coronary artery phantom. *IEEE Access* 9 166 409–166, 422
  31. Chi W, Dagnino G, Kwok TMY, Nguyen A, Kundrat D, Abdelaziz MEMK, Riga C, Bicknell C, Yang G-Z (2020) Collaborative robot-assisted endovascular catheterization with generative adversarial imitation learning. *IEEE International Conference on Robotics and Automation (ICRA)* 2414–2420
  32. Beaman C, Gautam A, Peterson C, Kaneko N, Ponce L, Saber H, Khatibi K, Morales J, Kimball D, Lipovac JR, Narsinh KH, Baker A, Caton MT, Smith ER, Nour M, Szeder V, Jahan R, Colby GP, Cord BJ, Cooke DL, Tateshima S, Duckwiler G, Waldau B (2023) Robotic diagnostic cerebral angiography: a multicenter experience of 113 patients. *J NeuroInterv Surg*
  33. Lahlouh M, Chenoune Y, Blanc R, Piotin M, Escalard S, Fahed R, Szweczyk J, Passat N (2023) Automated aortic anatomy analysis: from image to clinical indicators. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS. Institute of Electrical and Electronics Engineers Inc,

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