Data-Efficient Modeling of Hysteresis and Crosstalk for Inverse Kinematics of Soft Manipulators

Korn Borvorntanajanya*, *Student Member, IEEE*, Fung Flora Leung*, Jialei Shi, Enrico Franco, Philip Wai Yan Chiu, Yeung Yam, *Life Senior Member, IEEE*, and Ferdinando Rodriguez y Baena, *Member, IEEE*

***Abstract*—Nonlinearities in soft continuum manipulators, arising from material hysteresis and intersegmental coupling in multi-segment robots, present significant challenges for accurate open-loop inverse kinematics (IK) control. In particular, morphable pneumatic chambers adjust their shape and stiffness with internal pressure, increasing force output but also introducing nonlinearities that complicate control. This paper introduces a sequence-based machine learning approach that is data-efficient, modeling and compensating for both hysteresis and crosstalk in systems with morphable chambers. Through systematic comparison of Long Short-Term Memory (LSTM) and Transformer architectures under data-limited conditions, we demonstrate the effectiveness of sequence-based models in capturing temporal dependencies. We then propose a Recursive Segment-wise Crosstalk Compensation (RSCC) pipeline that decomposes control of multi-segment robots into independent single-segment subproblems, with each constituent model trained using 500 samples. Applied to a two-segment morphable-chamber manipulator, RSCC achieves approximately 11% normalized positional error and outperforms a monolithic multi-segment LSTM baseline trained on 2000 samples within the same workspace, highlighting its potential for precise open-loop control in minimally invasive surgical applications.**

***Index Terms*—Soft Robotic Systems; Hysteresis Modeling; Recursive Machine Learning; Open-Loop Control**

I. Introduction

SOFT robotic manipulators offer compliance, making them well suited for surgical applications that require safe tissue interaction [1]. Fluidic elastomer actuators (FEAs) achieve motion through internal pressurization and a cylindrical soft body with three or more chambers enables omnidirectional bending and elongation in compact spaces [2].

Our long-term goal is to develop a hybrid manipulator for endoscopic submucosal dissection (ESD). It consists of two soft segments, each containing three morphable chambers, and a distal tendon-driven segment for precise tissue manipulation [3]. During ESD, the proximal soft segments improve triangulation, often operate outside the visible field [4], and their compliance helps regulate contact forces and reduce perforation risk as shown in Fig. 1.

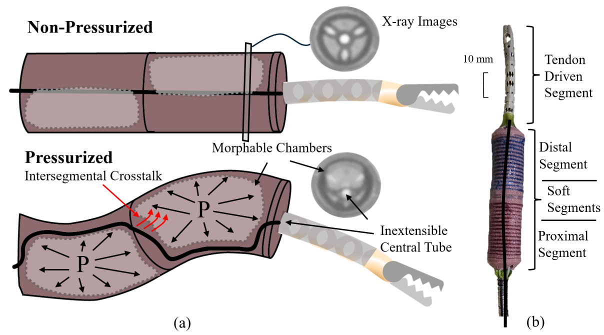

Fig. 1: (a) Illustration of the hybrid soft manipulator in non-pressurized (top) and pressurized (bottom) states. Intersegmental crosstalk (red arrows) results from asymmetric chamber actuation, with X-ray images showing the cross-sectional chamber morphing. (b) Image shows the assembled prototype.

A key advantage of this morphable FEA design is its high payload-size ratio. Soft surgical manipulators must remain compact to access internal anatomy yet deliver sufficient force for effective tissue manipulation [5]. Morphable FEAs can tune stiffness through pressurization-induced strain, which enhances lateral rigidity without additional mechanisms [6]–[8]. These actuators remain soft when unactuated yet can achieve high payload capacity in compact form [9].

However, this pressurization-based stiffening effect introduces substantial nonlinearities. Large chamber deformations generate pronounced hysteresis due to the hyperelastic material properties [7], [8]. Multi-segment design with a shared central lumen further introduces intersegmental crosstalk [2], [3], [5]. When one segment is actuated, its chamber expansion can deform adjacent chambers directly and via the central lumen, inducing unintended offset in neighboring segments.

To address these nonlinearities, closed-loop control methods are often adopted [3]. However, embedding sensors adds complexity that is undesirable in surgical contexts, motivating research into effective open-loop alternatives. Conversely, open-loop control suffers from cumulative errors due to hysteresis and crosstalk, which degrade accuracy and repeatability unless model parameters are exactly known. While parametric models (e.g., Preisach, Bouc–Wen) can capture elastic hysteresis, they are difficult to calibrate for multi-chamber, coupled morphologies [10]. As demonstrated by Liu et al., chamber coupling creates complex, sequence-dependent dynamics that are challenging to model accurately using analytical or finite element methods [11]. Data-driven techniques such as neural

Manuscript received: November, 25, 2025; Revised: March, 16, 2026; Accepted: May, 2, 2026.

This paper was recommended for publication by Editor Barbara Mazzolai upon evaluation of the Associate Editor and Reviewers comments.

This work has been supported by the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government, Multi-Scale Medical Robotics Centre, The Chinese University of Hong Kong. (*Corresponding author: Ferdinando Rodriguez y Baena*)

*Two authors contributed equally to this work.

Korn Borvorntanajanya, Jialei Shi, Enrico Franco, and Ferdinando Rodriguez y Baena are with The Hamlyn Centre for Robotic Surgery, Imperial College London, London, United Kingdom (email: f.rodriguez@imperial.ac.uk).

Fung Flora Leung, Philip Wai Yan Chiu and Yeung Yam are with the Multi-scale Medical Robotics Center, The Chinese University of Hong Kong, Hong Kong SAR, China (email: yyam@mae.cuhk.edu.hk).

Supplementary downloadable material for this letter is available at https://github.com/KornB/IK_soft_recursive, provided by the authors.

Digital Object Identifier (DOI): see top of this page.

networks have achieved up to 80% accuracy in quasi-static home-to-target tasks [12], but prior work has generally not accounted explicitly for hysteresis and crosstalk.

Sequence-based learning models such as Long Short-Term Memory (LSTM) [13] and Transformer networks [14] have shown strong capability in modeling temporal dependencies in nonlinear systems. In soft surgical manipulators, multi-segment architectures increase redundancy and data requirements, while material degradation limits the amount of data that can be collected [15]. This motivates the development of data-efficient learning frameworks that can accurately model complex hysteresis and intersegmental coupling with limited training data.

To this end, we propose a feedforward inverse kinematics framework for multi-segment soft manipulators with morphable pneumatic chambers. Our main contributions are:

- A modular inverse kinematics pipeline (RSCC) that decomposes multi-segment control into separate single-segment models with learned intersegmental crosstalk compensation, enabling accurate open-loop tracking with substantially fewer training samples than a monolithic multi-segment LSTM on the same robot and workspace.
- A sequence-based learning formulation for soft manipulators with morphable chambers, showing that modest history windows are sufficient to capture nonlinear hysteresis and crosstalk.

II. State of the Art

A. Soft Robotics Manipulators for GI Applications

In gastrointestinal (GI) applications, manipulators must remain below approximately 17 mm in diameter to pass through the narrowest regions of the colon [16]. This size constraint makes miniaturized fluidic elastomer actuators (FEAs) attractive, but often limits lateral force output and bending precision [5]. Radial fiber reinforcement has been used to suppress chamber expansion and improve bending accuracy [17], while variable-stiffness mechanisms have been introduced to improve payload capacity [18]. However, these strategies can reduce chamber deformability or require additional mechanisms. An alternative is the morphable-chamber design, where reinforcement is relocated externally so that the chambers can expand radially and increase rigidity during pressurization [7], [8], [19]. As summarized in Table I, most sub-15 mm soft manipulators provide limited tip force, below the level desirable for ESD tissue traction, which can require up to 0.4 N [20]. In contrast, morphable-chamber manipulators have demonstrated distal forces up to 2 N, providing a compact route toward clinically useful soft robotic instruments [6].

TABLE I: Comparison of State-of-the-Art Soft Manipulators d $\leq$ 15 mm

Designs	Actuation	DOFs	Diameter (mm)	Force(Segment/Device)(N)
FR-C [18]	P	2	14.5	0.36/N/A
FR-C [5]	P	6	9.5	$< 0.1/< 0.03$
M-C [3], [12],*	P	6	12	2/0.4

* This work, P: Pneumatic, FR: Fiber Reinforcement, M: Morphable, C: Chamber

B. Machine Learning Model for Hysteresis

Nonlinear and hysteretic FEA behavior depends strongly on actuator geometry, material properties, and pressurization history, making analytical kinematic models difficult to generalize across designs. A prior model-based approach for fiber-reinforced soft robots achieved sub-4% tip-position RMSE [2]. However, for morphable-chamber manipulators, where chamber deformation, hysteresis, and intersegmental crosstalk are more pronounced, extending such analytical models becomes challenging.

TABLE II: Comparison of State-of-the-Art Inverse Kinematics (IK) ML in Soft Continuum Manipulators (Only Open-Loop Control)

Robot Designs	Actuation	DOFs/Segments	Model	Gravity	n	%RMSE
PneuNet [21]	P	1/1	LSTM	Y	19635	1.95
McKibben Continuum [22]	P	2/1	FNN	N	2000	9.13
Tendon-driven Robot [13]	T	2/1	LSTM	N	3000	4.9
Tendon-driven Robot [14]	T	2/1	Tf	N	12000	3.7
McKibben Continuum [23]	P	4/2	RNN	N	12000	5.4
Soft Continuum [12]	P	4/2	FNN	N	1500	21
Soft Continuum *	P	4/2	LSTM	Y	500 E	11

* This work, P: Pneumatic, T: Tendon, n: Training Samples, E: each segment, Tf: Transformers

Machine learning (ML) provides an alternative by learning inverse mappings directly from data. Sequence-based models, including RNNs, LSTMs, and Transformers, can capture temporal dependencies from hysteresis and input history. Wu et al. [21] used a 50-step LSTM with 19,635 data points for pneumatic artificial muscle control, achieving less than 2% RMSE. Bai et al. [13] applied a 25-step LSTM to a tendon-driven surgical robot, reporting 4.0–5.8% workspace RMSE with 3000 samples. A Transformer-based framework used 12,000 samples for a cable-driven parallel continuum robot and achieved 3.5–4.3% RMSE under planar bending [14]. As summarized in Table II, Thuruthel et al. [22], [23] showed that a single-segment McKibben continuum robot required approximately 2000 samples with an FNN, whereas a two-segment system required around 12,000 samples with an RNN.

III. Soft Manipulator Characterization

The challenges of open-loop control are clarified by experimentally characterizing two nonlinear behaviors of the soft segments: hysteresis and intersegmental crosstalk. The soft manipulator utilized in this study is a two-segment, pneumatically actuated soft manipulator [3]. Each soft segment is 22 mm long, with an outer diameter of 12 mm, and incorporates three morphable chambers arranged at 120-degree intervals. A non-extensible central tube constrains axial elongation and directs pressurization to induce lateral bending.

A. Hysteresis in Morphable Chambers

To characterize hysteresis, we conducted experiments under quasi-static conditions, where velocity-dependent effects were assumed negligible. Pressure–bending angle responses were recorded under cyclic actuation for single-chamber, symmetric dual-chamber, and asymmetric dual-chamber configurations. As shown in Fig. 2 (left), all cases exhibit clear hysteresis, with offsets between the loading and unloading paths.

We further examined input coupling in task space by commanding the same target pressure state, $P = (1.6, 1.6, 0)$, through two different input paths: (1) inflating chamber 1 to $P = (1.6, 0, 0)$ and then chamber 2 to $P = (1.6, 1.6, 0)$, and (2) inflating chamber 2 to $P = (0, 1.6, 0)$ and then chamber 1 to $P = (1.6, 1.6, 0)$. As shown in Fig. 2 (right), these two sequences produced measurably different final tip positions. These results show that the manipulator response depends on both pressure history and actuation path, motivating a data-driven model that can implicitly capture these coupled effects.

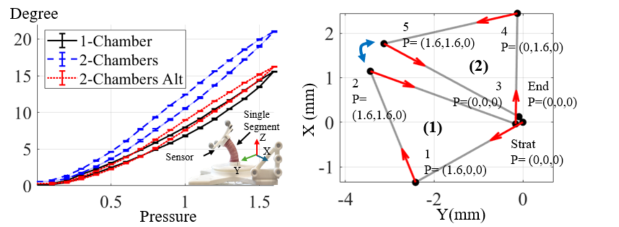

Fig. 2: Experimental characterization of hysteresis. (Left) Pressure-bending angle hysteresis loops for single and dual chamber actuation. Inset: experimental setup with optical tracking markers. (Right) Path-dependent spatial hysteresis in the task space, showing that different pressure input paths (gray lines) to the same target pressure P=(1.6, 1.6, 0) result in different final tip positions (points 2 and 5).

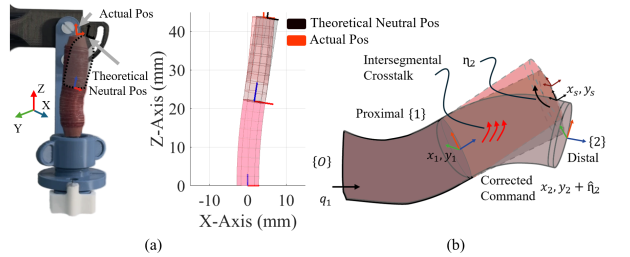

Fig. 3: Intersegmental crosstalk in a two-segment soft manipulator. (a) Experimental observation: the measured distal tip position (orange) deviates from the reference position (black) when the proximal segment is actuated, indicating coupling between segments. (b) Conceptual illustration of the crosstalk mechanism, where proximal actuation induces an unintended displacement η_2 that shifts the theoretical neutral position of the distal segment.

B. Intersegmental Crosstalk

Intersegmental crosstalk refers to the unintended motion of one segment induced by actuating its neighbouring segment. It is quantified as the displacement of the segment tip relative to a theoretical neutral position, defined as the tip location expected if the segment itself remained unpressurized and straight. This reference is constructed from EM-measured segment poses in the physical experimental configuration, so gravity and fixture effects are inherently included. For a two-segment manipulator, denoting the crosstalk offset of the j-th segment as η_j, two cases arise.

Case 1 (crosstalk on segment 2): The reference assumes that segment 2 remains unpressurized and straight, and is extended from the measured pose of segment 1:

$$\eta_2 = \mathbf{p}(T_2^O) - \left(\mathbf{p}(T_1^O) + l_2\,\mathbf{R}(T_1^O)\,\mathbf{e_z}\right),$$

where l_2 denotes the length of segment 2, T_j^O denotes the measured pose of the j-th segment tip in the base frame $\{O\}$, $\mathbf{p}(T)$ extracts the translation of T, $\mathbf{R}(T)$ extracts its rotation, and $\mathbf{e}_z = (0, 0, 1)$.

Case 2 (crosstalk on segment 1): The reference is the measured baseline pose of segment 1 in the unpressurized state:

$$\eta_1 = \mathbf{p}(T_1^O) - \mathbf{p}(\bar{T}_1^O),$$

where $\bar{T}_1^O$ is the measured unpressurized pose of segment 1. To illustrate this effect on the distal segment, the proximal segment was actuated with randomized pressure inputs while the distal segment remained unpressurized. The transforms T_1^O and T_2^O were obtained from electromagnetic (EM) measurements. A representative deviation of -1.02 mm in x and 0.27 mm in y was observed at 1.2 bar in proximal segment chamber 1, clearly illustrating the crosstalk effect (Fig. 3(a)). This displacement offset is used as the system state in the crosstalk prediction model described in Sec. IV-B2. In practice, the estimated offset can be compensated through input position correction, as illustrated in Fig. 3(b).

IV. Model Formulation & Control Strategy

This section presents RSCC, an end-to-end open-loop control pipeline that combines task-space IK with learned inverse models for multi-segment soft manipulators. Per-segment targets are computed via optimization, mapped to chamber pressures by LSTMs, and iteratively refined using predicted intersegmental crosstalk as shown in Fig. 4.

A. Inverse Kinematics for Single-Segment Robot

The IK of each segment is implemented using sequence-based LSTM and Transformer models. Both architectures use historical inputs over a sliding time window, allowing comparison of their effectiveness in capturing segment dynamics and sensitivity to time window length.

Each segment i is actuated by pressurizing the three internal chambers, with control input denoted as $q_i = (P_{i,1}, P_{i,2}, P_{i,3})$. Due to the non-extensible central lumen, segment motion is constrained to omnidirectional bending. Therefore, we define the task space as the projection of the tip position onto the xy-plane, (x_i, y_i), and augment it with the bending-plane angle as $\phi_i = \operatorname{atan2}(y_i, x_i)$. To avoid angular discontinuities, we embed this angle using its sine and cosine components, forming the state vector X_i:

$$X_{i,k} = (x_i, y_i, \cos\phi_i, \sin\phi_i)_k, \tag{1}$$

where k denotes the current discrete time step. To capture temporal dynamics and hysteresis, we define the augmented state vector $U_{i,k}$ as:

$$U_{i,k} = \Big(X_{i,k-1},\ \Delta(X_{i,k})\Big),\ \Delta(X_{i,k}) = X_{i,k} - X_{i,k-1}. \tag{2}$$

Here, $\Delta X_{i,k}$ explicitly captures the segment's dynamic change, facilitating hysteresis learning. The actuator pressures $\hat{q}_i$ are predicted from a sequence of these input vectors over a time window of length m:

$$\hat{q}_{i,k} = \mathcal{L}_i(U_{i,k}, U_{i,k-1}, \ldots, U_{i,k-m}). \tag{3}$$

where, $\mathcal{L}_i \colon \mathbb{R}^{8(m+1)} \to \mathbb{R}^3$ denotes either the LSTM or Transformer model.

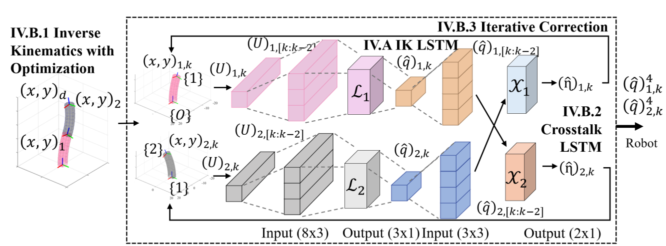

Fig. 4: RSCC pipeline and validation setup. (Left) CC-based IK optimization computes per-segment target tip positions (see IV-B1). (Center) IK LSTM predicts pressures (see IV-A), crosstalk LSTM estimates disturbances (see IV-B2), and an iterative correction loop (see IV-B3) refines commands.

B. Inverse Kinematics for Multi-Segment Robots

The proposed RSCC pipeline addresses the redundancy in inverse kinematics for multi-segment soft manipulators by mapping a desired end-effector position to appropriate segment-level commands. This is achieved by sequentially decomposing the global IK problem into segment-specific subproblems, enabling modular and precise control of the entire manipulator.

1) Redundant IK via Optimization: We formulate the IK of a two-segment soft manipulator as a constrained optimization problem. The objective is to determine segment-wise configurations that realize a distal tip target (x_d, y_d). Although each segment undergoes 3D bending, we constrain control to the tip's projection in the xy-plane due to limited mobility along the z-axis imposed by the non-extensible central lumen.

The optimization variables consist of the proximal segment's tip position (x_1, y_1) relative to the base frame and the distal segment's tip position (x_2, y_2) relative to the proximal segment's local frame as $\mathbf{Z} = (x_1, y_1, x_2, y_2)$. The cost function minimizes the squared Euclidean distance between the computed distal tip (x_t, y_t) and the target (x_d, y_d):

$$J(\mathbf{Z}) = (x_t - x_d)^2 + (y_t - y_d)^2. \tag{4}$$

The distal tip position (x_t, y_t) is computed by sequentially transforming segment frames under the constant-curvature (CC) assumption:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = P_{xy}\left(\sum_{i=1}^{2} R_i \begin{bmatrix} x_i \\ y_i \\ l_i \end{bmatrix}\right), \tag{5}$$

where $P_{xy}(\mathbf{p}) = (x, y)$ denotes planar projection of $\mathbf{p}$ onto the xy-plane. Each segment rotation R_i is obtained from its bending angle θ_i and plane angle ϕ_i using the standard CC formulation based on Rodrigues' rotation [24]. The complete IK optimization problem with constraints is formulated as:

$$\mathbf{Z}^* = \arg\min_{\mathbf{Z}} J(\mathbf{Z}) \text{ s.t. } \begin{cases} \theta_{[1,2]} \leq \theta_{\max}, \\ \phi_2 = \phi_1 + \pi, \\ (x_s - x_d)(-x_d) + (y_s - y_d)(-y_d) > 0, \\ x_1 \cdot x_d + y_1 \cdot y_d > 0, \end{cases} \tag{6}$$

these constraints ensure physically feasible and safe manipulator motions. The first constraint limits each segment's bending angle to θ_{max}. The second, third and fourth constraints define the S-shaped posture used in this study. Additionally, the second constraint sets a fixed bending-plane relationship between segments. The third constraint defines a reachable boundary, where (x_s, y_s) is the reference tip position if the distal segment were unbent (i.e. $(x_2, y_2) = 0$). The fourth constraint ensures that the proximal segment bends in the general direction of the target. This planar S-shaped configuration is commonly employed in single-port surgical manipulators to enhance triangulation and workspace coverage.

2) Intersegmental Crosstalk: To model the intersegmental crosstalk, we projected the crosstalk offset defined in (Sec. III-B) onto the xy-plane and implement an LSTM network that predicts the induced displacement $\hat{\eta}_{j,k}$ for segment j based on the recent actuation history of its neighbor, segment i:

$$\hat{\eta}_{j,k} = \mathcal{X}_j(q_{i,k}, q_{i,k-1}, \ldots, q_{i,k-m}), \quad i \neq j, \tag{7}$$

where $\mathcal{X}_j : \mathbb{R}^{3(m+1)} \to \mathbb{R}^2$ denotes the learned mapping from pressure history to the crosstalk-induced planar displacement. The training data for $\eta_{j,k}$ are obtained from EM pose measurements and corresponding reference positions as reported in (Sec. V-A2).

3) Iterative Correction Process: We introduce an iterative correction approach that refines segment-level commands by compensating for predicted intersegmental crosstalk. First, the optimization defined in (6) computes the desired segment tip configurations $\mathbf{Z}^*$, specifying both intermediate and distal segment positions in task space. These target points are then input into the inverse models $\mathcal{L}_i$ to predict corresponding actuator commands $\hat{q}_{i,k}$ as formulated in (3). Next, the predicted pressures are passed into the crosstalk model $\mathcal{X}_i$, defined in (7), to estimate the planar disturbances $\hat{\eta}_{i,k}$, which capture the effect of intersegmental coupling. To mitigate these disturbances, at each correction iteration h, the segment-level position is updated by subtracting the previously predicted crosstalk disturbance:

$$(x, y)^{(h)}_{i,k} = (x, y)^{(0)}_{i,k} - \hat{\eta}^{(h-1)}_{i,k}, \quad h = 1, 2, \ldots, H, \tag{8}$$

where $(x, y)^{(0)}_{i,k}$ denotes the initial target configuration, and H is the total number of correction steps. Each updated position $(x, y)^{(h)}_{i,k}$ is reprocessed through the inverse models $\mathcal{L}_i$, followed by a new crosstalk prediction $\hat{\eta}^{(h)}_{i,k}$ via $\mathcal{X}_i$. This feedback loop continues until H iterations are completed. The final corrected actuation commands $\hat{q}^{H}_{[1,2],k}$ are then deployed to the robot. In this work, we set $H = 4$ because command updates after the third iteration fell below 5%, indicating empirical convergence.

V. Model Training

The manipulator is actuated by digital pressure regulators (Tecno Basic, Hoerbiger) controlled via a microcontroller (Arduino Uno Rev4). The microcontroller communicates with Python scripts over a serial interface. Tip positions and orientations of each segment are tracked using EM sensors (NDI).

A. Data Collection Protocol

Data were collected under quasi-static conditions, allowing the system to settle for 2 seconds before each measurement. Pressure patterns were randomly generated by actuation 0, 1, or 2 chambers out of the 3 chambers per segment with probabilities 0.05, 0.25, and 0.7, respectively. Pressures for the selected chambers were randomly assigned within the safe operating range of 0 to 1.6 bar, as constrained by the robot's hardware. The resultant workspace reached was approximately 20 mm × 25 mm for the distal segment and 6 mm × 6 mm for the proximal segment. The robot was oriented horizontally to account for gravitational effects. Segment poses were recorded in a global reference frame O, with relative transformations computed as $T_i^{i-1} = \left(T_{i-1}^{O}\right)^{-1} T_i^{O}$, where T_{i-1}^{O} and T_i^{O} are the measured poses of segments $i-1$ and i respectively.

1) Single-Segment Data: Each proximal and distal segment was actuated independently using 500 randomized pressure commands. These data were used to train the individual IK models (Sec. IV-A).

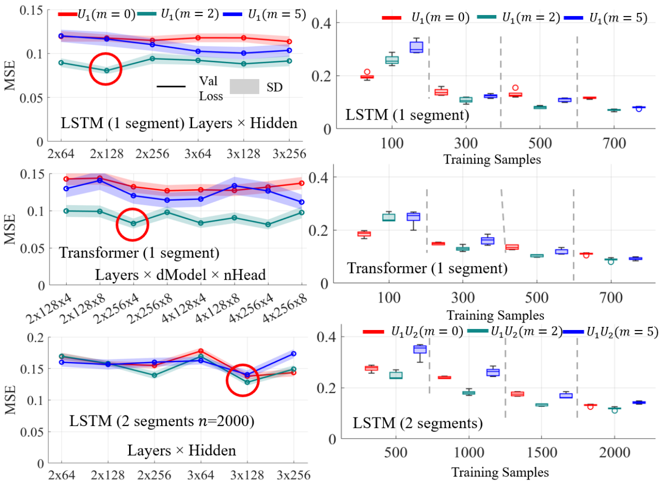

Fig. 5: Validation loss across different hyperparameters and time windows (m) for LSTM and Transformer models (left), and corresponding training-sample studies (right). Each point on the left represents the mean and standard deviation of validation loss across 5 random seeds. Red circles mark the best-performing configurations, which were selected for further evaluation with varying training samples in both one- and two-segment models.

2) Crosstalk Data: To construct the crosstalk dataset for each segment, we actuated only its neighbouring segment, corresponding to **Case 1** and **Case 2** in Section III-B. 500 samples are obtained for each case via randomized pressure commands to train each crosstalk predictor.

3) Two-Segment Data: A two-segment dataset was constructed to train a multi-segment LSTM baseline for benchmarking the proposed RSCC approach. The single-segment data (Sec. V-A1) was merged with a multi-segment dataset comprising 1000 randomized actuation patterns with simultaneous pressurization of both segments. This resulting dataset, totaling 2000 samples, contains paired input states $[U_{1,k}, U_{2,k}]$. The multi-segment IK LSTM, denoted as $\mathcal{L}_{1,2}\colon \mathbb{R}^{16(m+1)} \rightarrow \mathbb{R}^{6}$, was then trained using this dataset to predict the actuating pressures for $q_{1,k}$ and $q_{2,k}$.

B. Training Model Parameters

Datasets were divided by trajectory. Time-windowed samples from a given trajectory are assigned to train or validation (90:10), ensuring no overlapping windows cross splits. We performed a grid search on hyperparameters for **LSTM** (hidden size $\{64, 128, 256\}$, layers $\{2, 3\}$) and **Transformer** ($d_{\text{model}} \in \{128, 256\}$, layers $\{2, 4\}$, heads $\{4, 8\}$). Each configuration was evaluated at $m \in \{0, 2, 5\}$ using validation MSE (means $\pm$ SD over 5 seeds). All runs used Adam, batch size 32, dropout 0.15, learning rate 1×10^{-4}, and weight decay 1×10^{-5}. We selected the best-in-class models for subsequent experiments: LSTM $=$ 2 layers, hidden size 128; Transformer $=$ 2 layers, $d_{\text{model}} = 256$, 4 heads. The multi-segment LSTM baseline was configured with 3 layers and hidden size 128.

Across 500 training samples, tuning hyperparameters reduced the validation loss by approximately 12–18% for both architectures, indicating moderate sensitivity to model capacity as shown in Fig. 5. However, changing the time window m from 0 to 2 yielded a larger improvement of about 35–40%, confirming that temporal context contributes more to learning performance than network depth or width. For both LSTM and Transformer models, gains saturated when $m = 5$, suggesting diminishing returns for longer histories under limited data.

C. RSCC Pipeline Benchmarking

The RSCC pipeline is not directly comparable to a single LSTM, as it integrates multiple segment-wise networks with iterative crosstalk subtraction. Nevertheless, we compare the learning behavior and data efficiency between individual LSTMs within RSCC and the monolithic two-segment model. As shown in Fig. 5, RSCC exhibits improved sample efficiency at the submodel level. Increasing single-segment training data from 100 to 500 samples reduced validation loss by 45–50%, whereas expanding the monolithic two-segment dataset from 500 to 2000 samples reduced validation loss by only 25–30%. This comparison indicates that decomposing the problem into segment-wise mappings is easier to learn than fitting a single joint multi-segment mapping. For single-segment models, both LSTM and Transformer improved with data, but at 700 samples, the Transformer with $m = 5$ achieved about 10–15% lower loss than the LSTM, suggesting an advantage when more data are available.

VI. Experimental Validation

This section presents the experimental validation of the sequence-based models. Single-segment experiments evaluate model accuracy across different time windows (m). The time window length that achieved the lowest RMSE in Fig. 7 was used for both the crosstalk model and LSTM model in two-segment experiments. The two-segment robot experiments evaluate the performance of the RSCC pipeline in comparison to a multi-segment LSTM model, and under load-bearing conditions. Model details are listed in the Supplementary Materials.

A. Validation of a Single-Segment Robot

The single-segment study compared sequence models across time windows. We evaluated models with window lengths ($m \in \{0, .., 3\}$) using the augmented input $U_{i,k}$ from (3) and $m = 0$ using the input vector $X_{i,k}$. The models were evaluated using sequence point-to-point, circular, and rectangular trajectories. To enlarge the reachable set during testing, we used distal tip sensing while actuating the proximal segment.

1) Point-to-Point: Twenty point-to-point commands within the training workspace were used to evaluate the sequence-based models across different time windows. The distal tip position errors were recorded. The results are presented in Fig. 6. The configuration with $m = 2$ achieved the lowest RMSE for the LSTM, while Transformer performance is comparable across m. This aligns with our findings in (Sec. V-B). The Euclidean RMSE was reported in Table III.

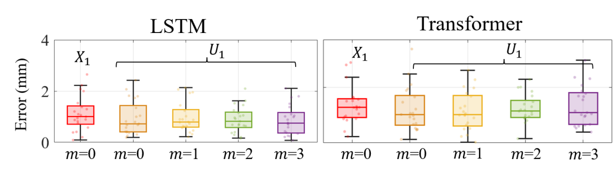

Fig. 6: Single-segment point-to-point results. Boxplots show tip-position errors between robot motion driven by predicted pressures and ground-truth tip positions for each model and input window m.

TABLE III: Euclidean RMSE (mean [95% CI] in mm)

	P2P $n=20$ **points**		**Trajectory** $n=50$ **points**		
m	LSTM	Transformer	LSTM	Transformer	Models (m=0) [Ref.]
0 (X)	1.23 [0.77, 1.69]	1.51 [1.02, 2.00]	1.92 [1.30, 2.54]	1.98 [1.39, 2.57]	2.41 [1.68, 3.14] [2]
0 (U)	0.96 [0.40, 1.52]	1.34 [0.84, 1.84]	1.43 [1.00, 1.86]	1.42 [0.98, 1.86]	2.01 [1.51, 2.51] [25]
1 (U)	1.08 [0.68, 1.48]	1.35 [0.77, 1.93]	1.17 [0.72, 1.62]	1.39 [0.95, 1.83]	2.12 [1.32, 2.92] [12]
2 (U)	**0.93 [0.62, 1.24]**	1.37 [0.91, 1.83]	**0.80 [0.49, 1.19]**	1.28 [0.86, 1.70]	
3 (U)	0.95 [0.52, 1.38]	1.34 [0.89, 1.89]	1.03 [0.68, 1.38]	1.92 [1.24, 2.60]	

2) Trajectory Control: Predefined shapes, including 8 mm-radius circles executed in clockwise and anti-clockwise directions and a rectangular pattern of 10 mm width and 4 mm length (total 50 points), were generated with random origin in the x-y plane to evaluate model performance. These inputs were unseen by the training models. Additionally, comparisons with a geometric-based model derived from [2], data-driven models including Gaussian Process (GP) [25] and FNN [12] were conducted.

The performance outcomes are shown in Fig. 7, and summarized quantitatively in Table III. The LSTM at $m=2$ again yielded the best performance, with the lowest RMSE among all models tested. These results show that the LSTM outperformed the Transformer, GP, FNN, and geometric model-based approaches in this evaluation.

3) Effect of Sampling Density: We evaluated the effect of sampling density on inverse kinematics accuracy using the best-performing model (LSTM, $m=2$). The workspace was discretized into bins where point-to-point commands were given in a random sequence with each grid containing 3 commands. Local RMSE and the corresponding training sample counts were reported per bin. As shown in Fig. 8, regions with higher sample density generally exhibit lower RMSE. A Pearson correlation analysis confirmed a negative correlation between the sample count per bin and the resulting RMSE ($r=-0.94, p<3.08\times10^{-10}$). This indicates that local data coverage is strongly associated with prediction accuracy across the workspace. Importantly, bin-wise RMSE is computed on unseen point-to-point executions, while sample density is measured from the training set, so the correlation serves as a practical indicator of how workspace coverage is associated with generalization error across the workspace rather than merely reporting training fit. While minor deviations from this trend exist (e.g., bin $X\in[-5,0]$, $Y\in[0,5]$), these likely stem from sequence-dependent hysteresis effects from preceding commands rather than insufficient local data.

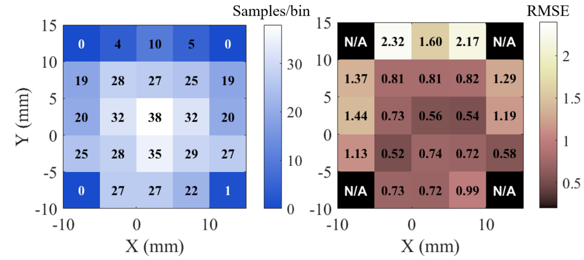

Fig. 8: Effect of sampling density on IK accuracy using the best LSTM model ($m=2$). (Left) Sample count per 5 mm bin. (Right) Tip RMSE from three point-to-point tests per bin, showing lower errors in densely sampled regions but deviations due to hysteresis.

4) Drift / temporal repeatability: To evaluate time-dependent drift under quasi-static conditions, the same 20-command point-to-point sequence was executed at T0 and repeated four times at 10-minute intervals. During each idle interval, the robot was held at the final commanded pressure and the sequence was re-executed without resetting. After the final run, the robot was returned to its neutral configuration and the sequence was repeated once more to assess recovery. Drift is assessed by comparing the tracked tip positions and the Euclidean error distributions across runs T0 to T40 and after reset T0_2. The Euclidean error exhibits a modest increase from T0 to T40, with the median remaining close to 1 mm, and returns near the initial level after reset, as shown in Fig. 9.

B. Validation of the Crosstalk Prediction

We validated the LSTM-based crosstalk prediction models described in Section IV-B2 using independent experimental data. For each segment, 50 randomized pressure commands were applied individually to the actuated segment i, while the adjacent segment j remained unpressurized.

The resulting tip displacement offset η_j was measured and compared with the predicted value $\hat{\eta}_j$ generated by model $\mathcal{X}_j$. The result is shown in Fig. 10. The RMSE of the Euclidean crosstalk offset was 0.230 mm (SD 0.128 mm) for proximal segment (from distal actuation) and 0.098 mm (SD 0.046 mm) for distal segment (from proximal actuation).

C. Validation of a Two-Segment Robot

In two-segment experiments, we compared the performance of our proposed RSCC pipeline ($m=2$) with a conventional

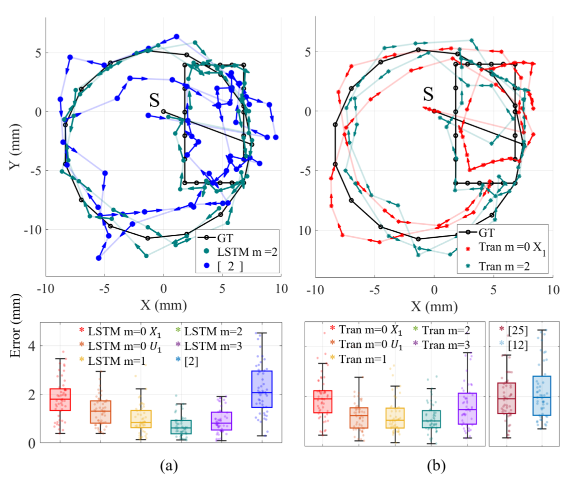

Fig. 7: Single-segment trajectory tracking on predefined circular and rectangular paths. Solid black lines denote ground-truth (GT) tip trajectories, while colored markers show robot motion driven by predicted pressure commands from different models. (a) LSTM-based comparison ($m=2$) using input U, benchmarked against the geometric model [2]. (b) Transformer-based comparison ($m=2$) with augmented inputs U, benchmarked against conventional input X. Boxplots indicate tip-position RMSE across trajectories with different time windows, Gaussian Process [25] and FNN [12].

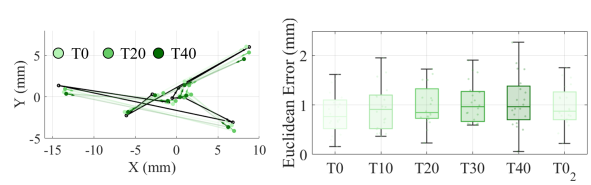

Fig. 9: Drift experiment. (Left) Tip trajectories at baseline (T0) and after 20 and 40 min (T20, T40), showing gradual deviation from the reference path (black). (Right) Boxplots of Euclidean position error over time (T0–T40), with $T0_2$ indicating the post-reset/repeat baseline.

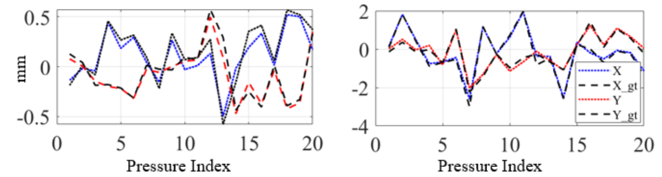

Fig. 10: Intersegment crosstalk displacement results. Predicted (solid line) vs. measured (dotted line) data for a representative subset of 20 randomized pressure commands from the 50-command validation set. (Left) Proximal-segment disturbance. (Right) Distal-segment disturbance.

LSTM model ($m = 2$) for a fair comparison on the hysteresis data. The commands for the proximal U_1 and distal segments U_2 were used as inputs to both the RSCC pipeline and the conventional LSTM model. We used a paired t-test to compare RMSE values from the RSCC and LSTM models on identical target points, testing the null hypothesis that their mean errors are equal. In this experiment, all data-driven controllers were executed at 20 Hz as the controller update and inference rate. The RSCC pipeline required 28 ms per command, whereas the monolithic LSTM required 48 ms per command (tested on an RTX 4070, CUDA 12.8).

1) Point-to-Point: Fifty random target points were selected within the distal workspace as (x_d, y_d). For each target, we computed per-segment commands using the optimizer without imposing an S-shaped constraint. Results are compared against an analytical baseline (*CC + geometric pressure mapping*) [2]. For a fair comparison, the analytical baseline uses the same target set and optimizer to compute the nominal segment configuration, but converts it to pressures via a static mapping without hysteresis memory or crosstalk compensation.

The RSCC pipeline reduced proximal RMSE from 0.93 mm with LSTM to 0.40 mm, and distal RMSE from 1.88 mm to 1.20 mm, as shown in Table IV (p $< 10^{-3}$) and Fig. 11. The *CC + geometric pressure mapping* baseline produced substantially larger errors, with proximal RMSE of 1.53 mm and distal RMSE of 4.98 mm. This gap is expected because the baseline does not capture path-dependent hysteresis or intersegmental crosstalk, and proximal errors propagate to the distal segment.

2) Trajectory Control with S-Shaped Posture: To evaluate the effectiveness of the RSCC pipeline in handling different geometric constraints, we executed experiments involving circular trajectories R as (x_d, y_d), with 30 sample points per circle, while varying the proximal segment radius r with defined $\theta_{1,\max}$ in (6).

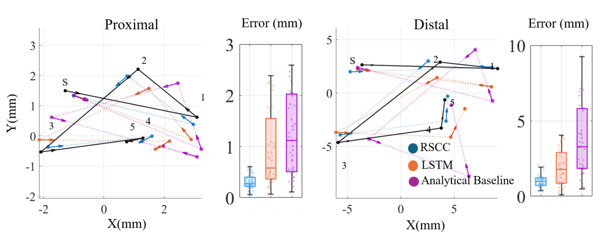

Fig. 11: Representative two-segment point-to-point trajectories for 5 targets, with Euclidean error boxplots summarizing all 50 targets. Left shows proximal tip positions, right shows distal. Ground truth is black circles, conventional LSTM is orange, RSCC is blue, and the analytical baseline is purple.

TABLE IV: Euclidean RMSE (mm), 95% CI, and p-value.

	P2P $n = 50$ points				**Circle S Shape** $n = 60$ points		
i	**RSCC**	LSTM	p	CC+ [2]	**RSCC**	LSTM	p
1	**0.40 [0.28, 0.52]**	0.93 [0.37, 1.49]	$< 10^{-4}$	1.53 [0.44, 2.62]	**0.47 [0.28, 0.66]**	0.65 [0.45, 0.85]	$< 10^{-5}$
2	**1.20 [0.96, 1.44]**	1.88 [0.76, 2.40]	$< 10^{-3}$	4.98 [1.21, 8.75]	**1.10 [0.86, 1.34]**	1.62 [1.36, 1.88]	$< 10^{-5}$

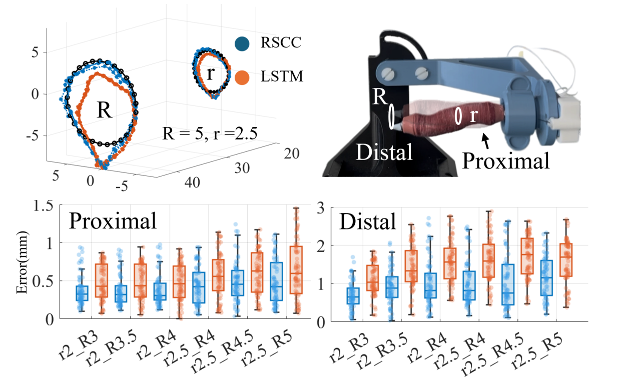

Fig. 12: Two-segment circular-trajectory results. Top: S-shaped posture proximal segment traces a circle of radius r (white), distal traces radius R. Left: ground-truth paths (black) with trajectories from RSCC (blue) and LSTM (orange). Bottom: boxplots of Euclidean tracking error for each r–R pair (left: proximal; right: distal).

The RSCC lowered proximal RMSE from 0.65 mm with LSTM to 0.47 mm, and distal RMSE from 1.62 mm to 1.10 mm (p $< 10^{-5}$). The RMSE and error distributions are reported in Table IV and Fig. 12.

3) S-Shaped Trajectory Under Payload (RSCC): The robustness of RSCC under external loading was evaluated by tracking a circular distal trajectory while maintaining the S-shaped configuration ($R = 4.5$, $r = \{2.0,\ 2.5\}$ mm). Discrete tip loads of 5, 10, and 20 g were attached to emulate tissue traction. An additional 10 g trial with distal tendon actuation was conducted to mimic traction-assisted ESD, using identical RSCC pressure commands.

Figure 13 shows distal trajectory tracking under different payloads. The median RMSE increased from approximately 1.0 mm (no load) to 1.1 mm (5 g), 1.6 mm (10 g), and 6.2 mm (20 g). For comparison, the 10 g tendon-actuated case achieved 2.2 mm. The larger error at 20 g was mainly caused by clamp displacement rather than degradation of the trajectory shape. The error pattern over waypoint index remains repeatable across cycles regardless of the applied payload. The first and second circles exhibit nearly identical error waveforms for all cases. This indicates that multi-step execution remains stable, with no clear progressive error accumulation across cycles, while tip loading primarily shifts the operating regime and

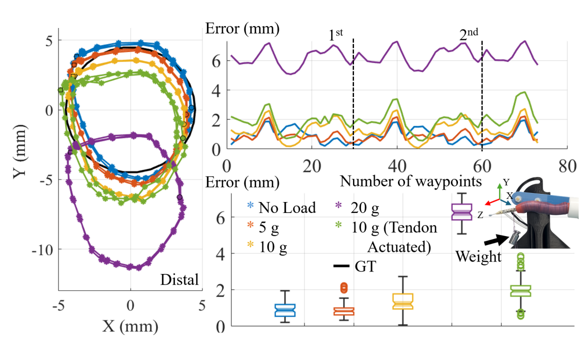

Fig. 13: Tracking performance of the distal segment under different payloads. Circular trajectories ($R = 4.5$ mm, $r = 2.5$ mm) were executed with 0–20 g tip loads and a 10 g tendon-actuated case.

introduces a repeatable bias.

VII. Discussion & Conclusion

This work presented RSCC, a modular inverse-kinematics framework for multi-segment soft continuum manipulators with morphable chambers. By combining sequence-based learning with recursive crosstalk compensation, RSCC captures hysteresis and intersegmental coupling using only 500 quasi-static samples per constituent model. Since the evaluation focuses on bending-dominant motion in a compact workspace, performance is interpreted through within-system learning curves and workspace-coverage analysis rather than absolute sample-count comparison, as shown in Fig. 8.

The proposed manipulator exhibits strong path-dependent hysteresis (Sec. III-A), making classical Bouc–Wen or Preisach models difficult to identify when extended to task-space history [10]. Sequence-based models captured these effects without explicit parameter tuning. For a single segment, augmenting the target state U with history reduced RMSE on unseen trajectories from 1.92 mm to 1.43 mm (p $<$ 10^{-3}). Under limited data, m=2 provided the best accuracy–generalization trade-off, improving circular-tracking RMSE from 1.43 mm to 0.8 mm, whereas shorter windows underfit and longer windows tended to overfit. Compared with previous sequence-based studies using larger datasets and fixed or arbitrary time windows [14], [21], these results highlight the importance of window selection in data-limited settings. LSTMs outperformed Transformers for $\leq$ 500 samples per segment, with Transformers becoming comparable only at 700 samples, suggesting that recurrent models are preferable when data are scarce. The final LSTM achieved about 6% RMSE relative to the command workspace, outperforming the 13%–15% RMSE of the baselines.

For two-segment control, RSCC improved data efficiency by decomposing the problem into smaller learned mappings, each trained on 500 samples, instead of relying on a monolithic two-segment LSTM trained on a larger simultaneously actuated dataset. Distal and proximal RMSEs reached 11% and 7.8% of their commanded workspaces, corresponding to 3.4% and 5.5% relative to the training workspace. The command workspace was intentionally constrained to preserve the S-shaped posture. RSCC also remained robust under loading, with distal error below 3 mm for payloads up to 10 g, both with and without tendon actuation. The larger 6.2 mm deviation at 20 g was mainly caused by clamp displacement.

Practical limitations arise mainly from the open-loop setting and physical changes in the robot. Since the controller does not receive physical feedback or update the model online, unmodeled drift or regime shifts can cause gradual tip-position deviations over extended operation, although this drift was largely reversible after venting and rest in the repeatability test. The repeatability and loading tests further indicate that payload-induced errors were largely repeatable within the tested range, with no clear progressive error accumulation across cycles. Long-term aging was not studied because the manipulator is intended for disposable or single-procedure use. Permanent changes, including material damage, leakage, or plastic deformation, can invalidate the learned mapping and lead to control failure. Additional limitations include operation in sparsely sampled workspace regions, non-quasi-static actuation, pressure saturation, and sensing disturbances.

In surgical telemanipulation, minimizing embedded sensing improves robustness, and open-loop accuracy can be clinically practical when the human remains in the loop. Future work will improve robustness through more uniform workspace coverage, targeted boundary-region sampling, and three-chamber pressurization for controllable stiffness modulation and higher payload capacity.

References

[1] L. Manfredi, "Endorobots for Colonoscopy: Design Challenges and Available Technologies," *Frontiers in Robotics and AI*, 2021.

[2] J. Shi *et al.*, "Position and orientation control for hyperelastic multisegment continuum robots," *IEEE/ASME T-MECH*, 2024.

[3] F. F. Leung *et al.*, "Hybrid manipulator with force estimation for lower gastrointestinal interventions," *IEEE R-AL*, pp. 9709–9716, 2024.

[4] Y. Cui *et al.*, "Robotics in therapeutic endoscopy (with video)," *Gastrointestinal Endoscopy*, vol. 96, pp. 402–410, Sept. 2022.

[5] J. Shi *et al.*, "Miniaturised soft manipulators with reinforced actuation chambers on the sub-centimetre scale," in *IEEE RoboSoft*, 2024.

[6] A. Garriga-Casanovas *et al.*, "Optimised design and performance comparison of soft robotic manipulators," in *2022 7th ICMERR*, 2022.

[7] V. Virdyawan *et al.*, "A variable stiffness pneumatic actuator for minimally invasive surgery devices," *IEEE T-MRB*, 2025.

[8] W. Gaozhang *et al.*, "A novel stiffness-controllable joint using antagonistic actuation principles," *Mechanism and Machine Theory*, 2024.

[9] F. Chen, Z. Song, S. Chen, G. Gu, and X. Zhu, "Morphological design for pneumatic soft actuators and robots with desired deformation behavior," *IEEE Transactions on Robotics*, vol. 39, no. 6, 2023.

[10] T. Wang *et al.*, "From model-driven to data-driven: A review of hysteresis modeling in structural and mechanical systems," *Mechanical Systems and Signal Processing*, vol. 204, p. 110785, 2023.

[11] Z. Liu *et al.*, "Modeling and analysis of soft pneumatic network bending actuators," *IEEE/ASME Transactions on Mechatronics*, 2021.

[12] A. Garriga-Casanovas, F. Shakib, V. Ferrandy, and E. Franco, "Hybrid control of soft robotic manipulator," *Actuators*, vol. 13, July 2024.

[13] W. Bai *et al.*, "Task-based lstm kinematic modeling for a tendon-driven flexible surgical robot," *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 2, pp. 339–342, 2022.

[14] X. Liu *et al.*, "Transformer-based segmented learning for kinematics modelling of a cable-driven parallel continuum robot," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 239, no. 8, 2025.

[15] M. Kasaei *et al.*, "Data-efficient non-parametric modelling and control of an extensible soft manipulator," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 2641–2647.

[16] A. Alazmani *et al.*, "Quantitative assessment of colorectal morphology: Implications for robotic colonoscopy," *Medical Engineering and Physics*, vol. 38, pp. 148–154, 2 2016.

[17] M. Cianchetti *et al.*, "Stiff-flop surgical manipulator: Mechanical design and experimental characterization of the single module," in *IEEE/RSJ IROS*, 2013, pp. 3576–3581.

[18] M. Brancadoro *et al.*, "Toward a variable stiffness surgical manipulator based on fiber jamming transition," *Frontiers Robotics AI*, vol. 6, 2019.

[19] S. Treratanakulchai *et al.*, "A novel soft robotic manipulator design with zig-zag chamber geometry," in *2024 10th International Conference on Automation, Robotics and Applications (ICARA)*, 2024, pp. 1–6.

[20] J. Gafford *et al.*, "Snap-on robotic wrist module for enhanced dexterity in endoscopic surgery," in *Proc. IEEE ICRA*, 2016, pp. 4398–4405.

[21] D. Wu *et al.*, "Deep-learning-based compliant motion control of a pneumatically-driven robotic catheter," *IEEE R-AL*, 2022.

[22] T. G. Thuruthel *et al.*, "Learning dynamic models for open loop predictive control of soft robotic manipulators," *Bioinspiration and Biomimetics*, vol. 12, 10 2017.

[23] T. G. Thuruthel, E. Falotico *et al.*, "Stable open loop control of soft robotic manipulators," *IEEE Robotics and Automation Letters*, 2018.

[24] R. J. WebsterIII and B. A. Jones, "Design and kinematic modeling of constant curvature continuum robots: A review," *The International Journal of Robotics Research*, vol. 29, no. 13, pp. 1661–1683, 2010.

[25] C. Relaño *et al.*, "Gaussian process regression for forward and inverse kinematics of a soft robotic arm," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 107174, 2023.