IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Multi-Modal Sensing in Colonoscopy: A Data-Driven Approach

Viola Del Bono, *Graduate Student Member, IEEE*, Emma Capaldi, Anushka Kelshiker, Ayhan Aktas, Hiroyuki Aihara, and Sheila Russo, *Member, IEEE*

***Abstract*—Soft optical sensors hold potential for enhancing minimally invasive procedures like colonoscopy, yet their complex, multi-modal responses pose significant challenges. This work introduces a machine learning (ML) framework for real-time estimation of 3D shape and contact force in a soft robotic sleeve for colonoscopy. To overcome limitations of manual calibration and collect large datasets for ML, we developed an automated platform for collecting data across a range of orientations, curvatures, and contact forces. A cascaded ML architecture was implemented for sequential estimation of contact force and 3D shape, enabling an accuracy with errors of 4.7% for curvature, 2.37% for orientation, and 5.5% for force tracking. We also explored the potential of ML for contact localization by training a model to estimate contact intensity and location across 16 indenters distributed along the sleeve. The force intensity was estimated with an error ranging from 0.06 N to 0.31 N throughout the indenters. Despite the proximity of the contact points, the system achieved high localization performances, with 8 indenters reaching over 80% accuracy, demonstrating promising spatial resolution.**

***Index Terms*—Force and tactile sensing, soft sensors and actuators, modeling, control, and learning for soft robots.**

I. Introduction

SOFT robots and sensors, due to their highly deformable and compliant nature, often present significant challenges for traditional modeling and control [1], [2]. These systems inherently exhibit complex and non-linear behavior, with phenomena such as hysteresis, viscoelasticity, and gradual material degradation contributing to dynamic responses that are difficult to predict or generalize. As a result, conventional analytical or physics-based models often fail in capturing the full range of variability and non-linearity present in these category of robotic systems [3], [4].

Received 11 August 2025; accepted 8 December 2025. Date of publication 18 December 2025; date of current version 26 December 2025. This article was recommended for publication by Associate Editor T. M. Huh and Editor Y.-L. Park upon evaluation of the reviewers' comments. This work was supported by the National Institute of Biomedical Imaging and Bioengineering of the NIH under Award R21EB029154. *(Corresponding author: Sheila Russo.)*

Viola Del Bono, Emma Capaldi, Ayhan Aktas, and Sheila Russo are with the Department of Mechanical Engineering, Boston University, Boston, MA 02215 USA (e-mail: russos@bu.edu).

Anushka Kelshiker is with the Department of Biomedical Engineering, Boston University, Boston, MA 02215 USA.

Hiroyuki Aihara is with the Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215 USA.

This article has supplementary downloadable material available at https://doi.org/10.1109/LRA.2025.3645700, provided by the authors.

Digital Object Identifier 10.1109/LRA.2025.3645700

To bypass the need for creating complicated models that fall short in accurately capturing robots' response, machine learning (ML) approaches offer a powerful alternative, as they are data-driven, easy to implement, versatile, and robust [5], [6]. Various ML strategies have been used to implement space planning and control of soft manipulators [7], [8]. Moreover, data-driven methods have been recently proposed for soft sensing, including piezoelectric electronic skins [9], bioinspired piezoresistive soft fingers [10], textile proprioception [11], strain-based continuum robots [12], tactile sensors [13] and arrays [14], pneumatic-based sensors [15], [16], and unstructured soft capacitive sensors [17]. Among sensor types, soft optical sensors are particularly challenging to model, due to the complex and often multi-modal nature of how mechanical deformation affects light transmission [18], [19]. These deformations often affect multiple aspects of the optical signal simultaneously, resulting in spatially distributed, non-linear, and strongly coupled responses [19], making input-output relationships difficult to interpret and challenging to capture with analytical or physics-based models. ML-based soft optical sensors have been developed both for proprioception and exteroception [18], however most of the systems are based on commercial fibers [20] or Fiber Bragg grating sensors [21], [22], [23], introducing high cost, complexity, and limited flexibility [24].

While ML offers advantages in handling the complexity of soft sensing, its performance heavily depends on the availability of large, high-quality datasets [25], [26]. Collecting such data can be time consuming and labor intensive, particularly for soft sensors that lack intrinsic actuation and therefore rely on external or manual manipulation during calibration. As a result, the calibration phase often becomes a major bottleneck, limiting the scalability and practicality of ML-based sensing approaches [4].

In previous work [27], a soft robotic sleeve based on optical sensing was developed and extensively tested to detect contact forces during colonoscopy, sense its own shape, and deploy soft actuators when excessive force was detected, with the goal to minimize potential tissue damage and patient discomfort (Fig. 1(a)). The sleeve was manually calibrated and the optical data was interpreted by a polynomial fitting. Leveraging this study, we developed an automated platform capable of calibrating 3D sensor orientation, bending, and contact force. With the large datasets collected, we implemented an ML strategy to predict shape and contact force in real time, improving both accuracy and time efficiency compared to the prior model and fully manual calibration process. Additionally, to further explore

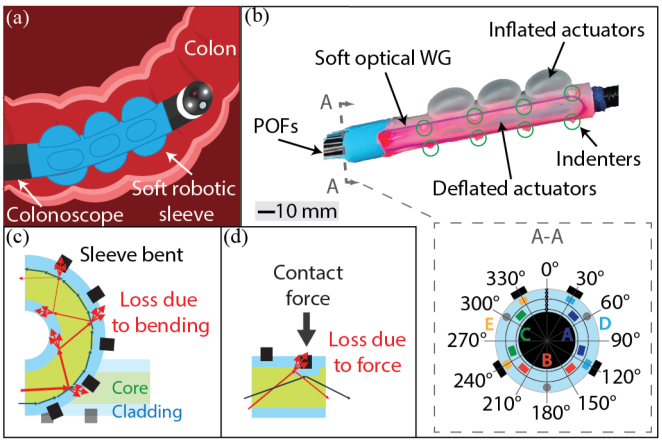

Fig. 1. (a) The soft robotic sleeve is navigating in the colon and continuously sensing contact forces. If an excessive force is detected, actuators automatically deploy to redistribute pressure and decrease patient discomfort. (b) Soft robot main components and cross section. (c) Optical loss caused by bending the robot. (d) Optical loss caused by contact force, amplified by the indenters located on the robot surface.

the potential of ML for multi-modal sensing, we conducted a preliminary study focused on contact force localization along the surface of the sleeve. Accurate force localization is critical in colonoscopy, where detection of contact locations can enhance navigation safety and reduce the risk of tissue damage [28], [29]. By enabling real-time, localized sensing of contact forces, ML-driven approaches can provide surgeons with richer feedback and improve the robot's interaction with the colon [30]. This study shows promising capabilities of the robot to simultaneously predict force intensity and location across 16 points on its surface, paving the way for integrating advanced sensing capabilities in colonoscopy procedures.

II. Materials and Methods

A. Soft Robotic Sleeve Overview

Fig. 1(b) shows the main components of the robot. The robot body is made of Ecoflex© 00–30 (Smooth-On, Inc.), with an outer diameter of 23 mm, and a 150 mm length. It contains five U-shaped soft optical waveguides (WGs -A, B, C, D, E in the cross sectional view) arranged in two layers. A set of eight indenters (circled in green) is integrated on the two external WGs, to amplify the contact force. All the WGs are connected to commercial plastic optical fibers (POFs -Industrial Fiber Optics, Inc.) connected to an external optical circuit. Three pneumatic actuation lines are spaced 120° apart and are connected to an external actuation unit. Additional details of the robot design can be found in [27]. As shown in the cross sectional view of Fig. 1(b), in this work, we approximate the robot's possible bending orientations into 12 directions spaced 30° apart, which will be later used to estimate shape.

The optical sensor working principle relies on total internal reflection, where the core material, made of optical adhesive (NOA 65, Norland Products Inc.), has a higher refractive index than the cladding, made of Ecoflex© 00–30, respectively 1.56 and 1.41 [27]. This results in a high light propagation throughout the length of the WGs. The latter are subjected to optical loss due to deformations, in particular 3D bending of the colonoscope (Fig. 1(c)) and contact force caused by the interaction of the colonoscope with the colon tissue, amplified by the 16 indenters on the sleeve surface (Fig. 1(d)). Relating these complex loss modes to the applied deformations is essential for our aim to create a multi-modal soft optical sensor able to sense its 3D shape, and external localized contact forces.

B. Calibration Platform Design

To minimize calibration time, reduce manual variability, and enable large-scale data collection for multi-modal sensor characterization, we developed a compact automated calibration platform, that is able to change the robot shape and apply external contact at different curvatures, while continuously recording the optical losses and axial forces. Fig. 2 shows the main components of the machine. The sleeve is placed on a mock colonoscope that is inserted and fixed on the machine in a straight position (see Fig. 2(b)). The colonoscope is held at two points with removable fixtures, connected to two 300° rotation stepper motors (2000 Series 5-Turn Dual Mode Servo 25–2 Torque, GoBuilda), at a 5:4 gear ratio to allow for 375° rotation. These correspond to motors 1 and 2in Fig. 2 and they control the sleeve orientation (axial rotation). Bending is achieved using counter-rotating arms that maintain a fixed center of rotation, with the right arm mounted on a 32 mm cylinder and the left on an 8 mm shaft running through its center, enabling independent rotation around a shared axis. Each arm is controlled individually by a servo motor (2000 Series Dual Mode Servo 25–2 Torque, GoBuilda), respectively motor 3 for the right arm and motor 4 for the left arm of Fig. 2. The machine can bend the sleeve up to a maximum curvature of 15 m^{-1} ($\approx 92°$). The platform can also apply external contact force by using a sliding arm, controlled by a 117 rpm DC motor (5203 Series Yellow Jacket Planetary Gear Motor, GoBuilda) at 20% maximum speed, which rotates a lead screw with a 1:4 gear ratio (see motor 5 and lead screw in Fig. 2(a)). This allows for force to be applied in a controlled and slow manner. A load cell (DYLY106 5 kg, CALT), coupled with an amplifier (HX711, Stemedu), is directly attached to the arm, and a 3D printed pressing fixture is connected to the cell. The pressing fixture is specific to the curvature of the sleeve, to best mimic contact forces experienced during colonoscopy. Since the curvature of the colonoscope changes during navigation, three different fixtures were printed with curvatures 0 m^{-1}, 5 m^{-1}, and 10 m^{-1} to evaluate three possible cases of contact. Two limit switches (Micro Limit Switch Bump Lever, GoBuilda) prevent the sliding arm to exceed the upped and lower bounds, avoiding undesired collisions. Lastly, the machine is powered by an external 20 V battery reduced to 12 V by a 6 A voltage regulator. An Arduino UNO controls both the machine motors, collects force data from the load cell, and optical signals from the five WGs embedded in the soft sleeve.

C. Data Collection

The designed machine can be used for shape and force calibration. All tests were conducted in an environment with controlled

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

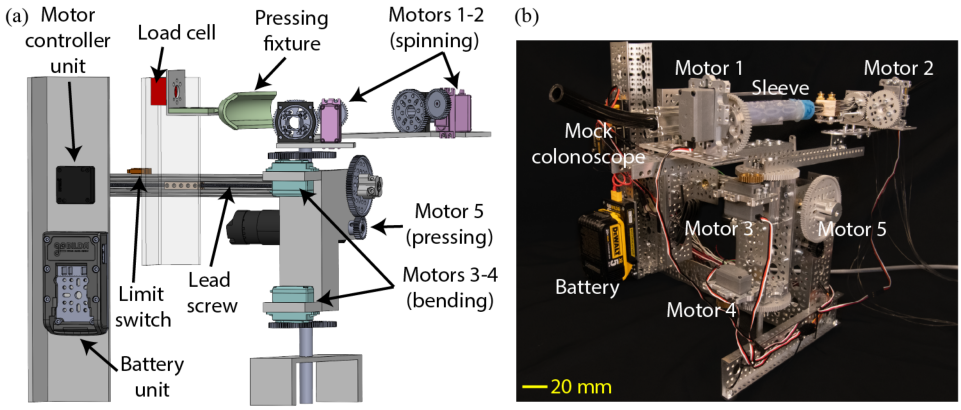

Fig. 2. (a) CAD model and (b) picture of the automated calibration machine, with main components labeled.

illumination and temperature. For the shape calibration, the sleeve is first mounted onto the machine at curvature 0 m^{-1} and axial orientation 0° (see Fig. 1(b).) When the calibration starts, the sleeve is gradually bent up to a maximum curvature of 15 m^{-1}, with 5° incremental steps, and the sensor losses are continuously recorded. The machine then returns to a straight position. This process is repeated five times for each of the 12 orientations (0° to 330° with steps of 30°, as schematized in Fig. 1(b)). The orientation discretization into 12 directions was chosen to achieve a fine sampling and resolution for shape reconstruction, while still limiting calibration time and algorithm computational cost. For the force calibration, the sleeve is again mounted onto the machine at its resting position. The user attaches the first pressing fixture (straight, curvature 0 m^{-1}) to the load cell and the test is started. The machine sliding arm moves forward until contact with the sleeve is detected by the load cell. The speed of the load applicator is set to 3 $\frac{mm}{s}$ in order to get fine sensor measurements, which are taken at 1 ms intervals. After contact, the sliding arm continues to advance until 8 N of force is detected by the load cell, then it unloads at the same speed until contact is no longer detected, while collecting sensor measurements throughout the whole procedure. This process is repeated five times for each axial orientation. After all orientations at curvature 0 m^{-1} are calibrated, the loading arm returns to its starting position at the upper limit switch. The pressing fixture is sequentially replaced with 5 m^{-1} and 10 m^{-1} curvatures, repeating the full orientation sequence each time with the sleeve bent accordingly. This calibration results in a total of 180 tests (5 loading cycles × 12 orientations × 3 curvatures). The above-described shape and force calibrations result in large datasets of 57000 data-points and 74310 data-points, respectively. The datasets are then combined and divided in three sub-datasets: training (60%), validation (20%), testing (20%), and shuffled using random initialisation, to access the data in a non-sequential manner.

D. Machine Learning Model Architecture

To leverage the large datasets collected by the automated platform for accurate sensor calibration and the development of a model for real-time prediction of shape and contact force, ML techniques were explored. The full architecture is schematized in Fig. 3 and consists of three models that estimate force, orientation, and curvature, respectively. A first model predicts contact force (Model 1) from the five live losses of the sensor. Then, two parallel models predict orientation (Model 2) and curvature (Model 3) of the robot based on the output of the force model and the 5 input losses. This concatenated approach was implemented to ensure that the application of contact would not influence shape prediction, as contact causes additional losses on the WGs. All the models were trained on datasets that included both zero-force and non-zero-force cases to capture the sleeve's behavior during pure bending as well as combined bending and contact. In order to choose the best ML model for each case, various algorithms were trained on the training dataset by using MATLAB Machine Learning Toolbox, and compared. The models evaluated were: Decision Trees, Support Vector Machine (SVM), Neural Networks, K-Nearest Neighbors (KNN), Ensemble methods, and Gaussian Process Regression.

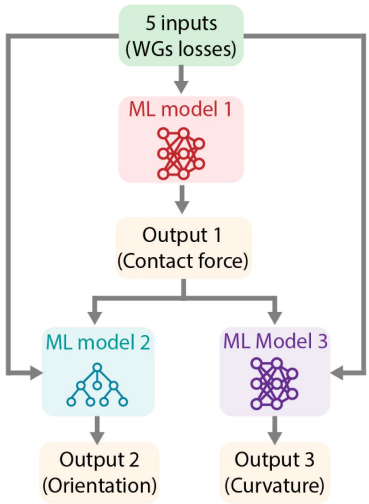

Fig. 3. ML cascaded architecture to live predict contact force and shape by using the 5 WGs inputs. Contact force is first estimated by Model 1 and then fed into Models 2 and 3 along with the WGs losses.

Given the calibration setup, regression models were used for force and curvature prediction, while orientation was treated as a classification task, since the sleeve was rotated through its 12 discrete orientations. The best model was chosen based on the lowest RMSE for regression and the highest accuracy for classification. A 5 fold cross-validation was implemented to prevent overfitting. Specifically, Model 1 consists of an Exponential Gaussian Process Regression (GPR), with a constant basis function. An ensemble subspace k-nearest neighbors (KNN) classifier was selected for Model 2, with 30 learners trained on a randomly selected subspace of 3 features. Similar to the force model, an Exponential Gaussian Process Regression with constant basis function was selected for Model 3.

III. Experiments

A. Offline Predictions

To validate separately the accuracy of each chosen ML model described in Section II-D, the model predictions were evaluated offline on the data collected during calibration (training, validation, and testing datasets). RMSE was calculated for the regression models of force (Model 1) and curvature (Model 3), while classification accuracy was calculated for the orientation model (Model 2).

B. Real-Time Shape Validation

We first evaluated in real-time the shape estimation of the implemented ML models with no contact force applied. Two experiments were conducted by predicting both orientation and curvature while varying one parameter and keeping the other constant. The robot was fixed in a straight position on the calibration machine. In the first experiment, its orientation was maintained at a constant value (respectively 0°, 120°, and 270°) while its curvature was adjusted sequentially to the following values: [0, 3, 6, 10, 12] m^{-1} for case 1 and [0, 4, 7, 11, 13] m^{-1} for cases 2 and 3. In the second experiment, the robot curvature was instead set to a fixed value (respectively 6 m^{-1}, 8 m^{-1}, and 10 m^{-1}) and the orientation was sequentially moved to the following values: [0°, 150°, 240°, 270°] for case 1, [0°, 120°, 210°, 270°] for case 2 and [0°, 150°, 240°, 270°] for case 3. In both experiment scenarios, the sleeve predictions of curvature and orientation were continuously recorded and compared with the true values, as well as the previous polynomial fit model of [27].

C. Real-Time Shape and Force Validation

A second set of experiments was carried out to evaluate the full algorithm cascade architecture, by simultaneously predicting shape and contact force. The goal of this test was to ultimately evaluate the accuracy of our approach, ensuring that shape estimation remains accurate throughout the pressing test, and therefore it is not influenced by additional sensor loss caused by contact force. The sleeve was placed at a set orientation and curvature, and the pressing fixture with corresponding curvature was moved until contact, and gently pressed on the sleeve body

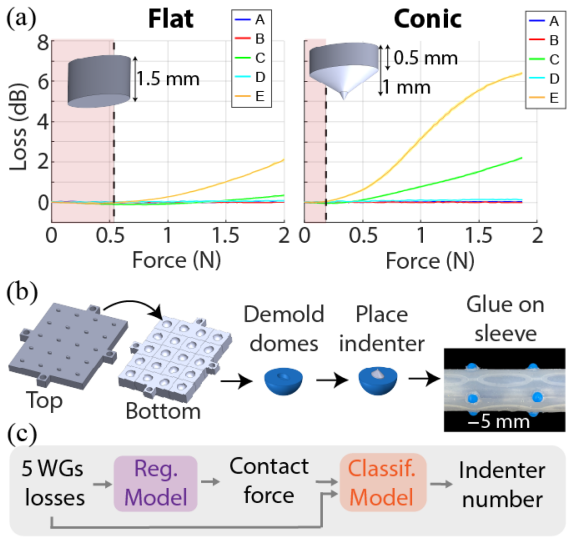

Fig. 4. (a) Conical indenters greatly improve sensitivity with respect to flat ones, when pressing on the same location (Indenter 14 showed). In fact, the optical responses start after 0.51 N for the flat design, and at 0.2 N for the conic design (see shaded areas in the plots). (b) Implementation of silicone domes and embedded indenters on the sleeve. (c) ML architecture to predict force intensity and indenter location.

up to 7 N, similar to the force calibration process described in Section II-C. Shape and force predictions were continuously recorded and compared with the true values, and the previous polynomial fit model [27]. The test was repeated three times by using the three calibrated curvatures of 0, 6, 10 m^{-1} and orientations 180°, 60°, 270°, respectively.

D. Contact Localization Study

To further investigate and push the boundaries of multi-modal optical sensing through machine learning, a preliminary study was carried out focusing on contact force localization. For this aim, each indenter should ideally produce a unique set of responses for the 5 WGs, when it is pressed individually. This distinction will enable accurate discrimination among the 16 contact locations corresponding to the 16 indenters embedded in the soft sleeve.

First, to increase sensitivity at low forces and signal differentiation, conically shaped indenters were implemented, with a sharper surface with respect to the previous flat design (see designs of Fig. 4(a)). We compared the sensitivity of these two shapes by manufacturing two sleeves integrating one shape or the other, and pressing on indenters on the same location with an Instron testing machine (5943, Instron), while recording the optical losses. Fig. 4(a) shows that for the same indenter location, the conical shape notably increases the sensitivity to low forces, as the optical responses start as early as 0.2 N, whereas the flat design causes a loss response only after 0.51 N. Additionally, the indenters were previously manually coated with glue in an uncontrolled manner [27], which limited consistency and repeatability. To ensure a more standard and repeatable fabrication, we embedded each indenter into a silicone dome, fabricated by casting MoldStar© 30 (Smooth-On Inc.) into 3D printed molds.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

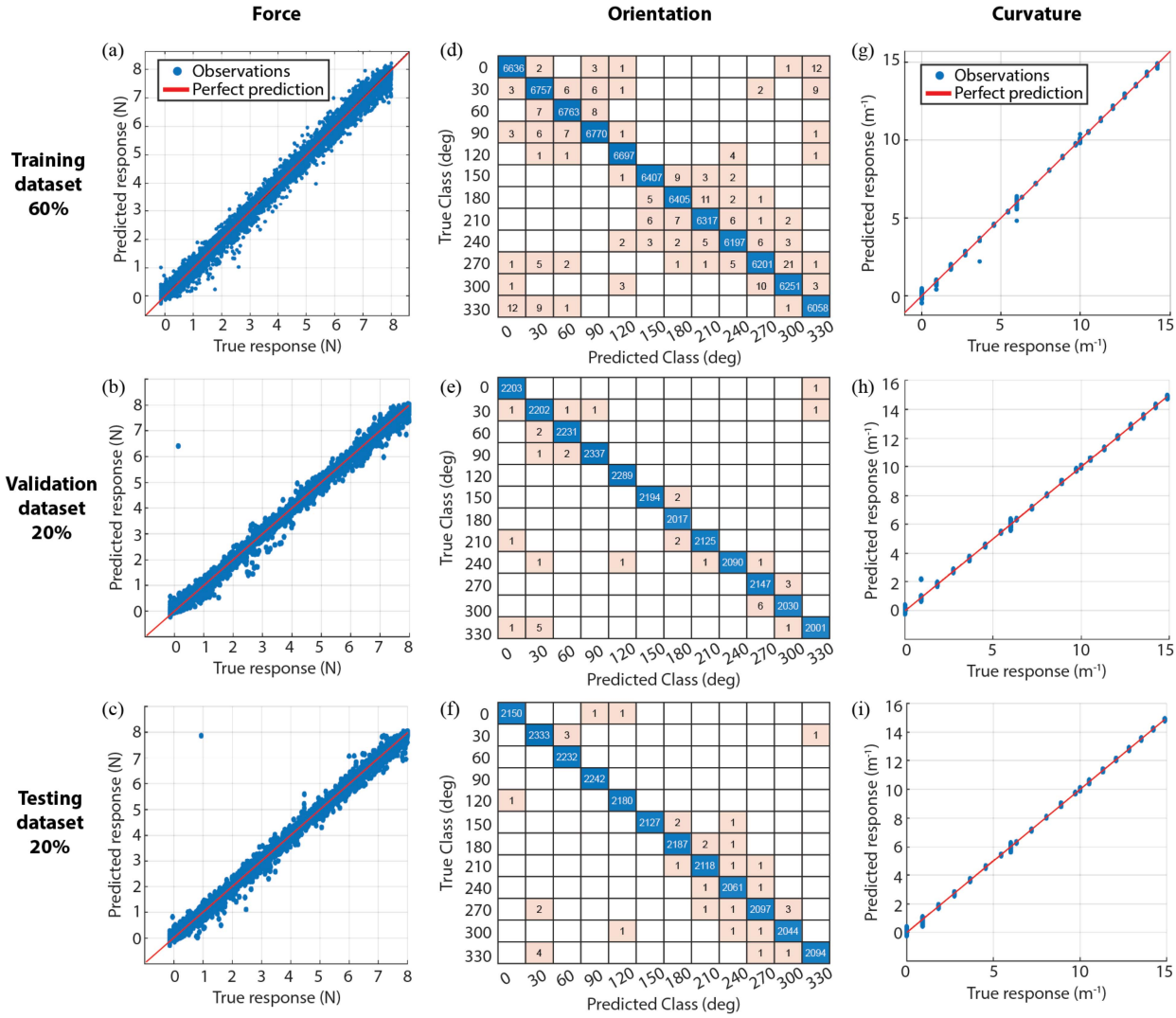

Fig. 5. Offline predictions of the ML models implemented for force (a), (b), (c), orientation (d), (e), (f), and curvature (g), (h), (i), evaluated on the three calibration sub-datasets. Force predictions achieved RMSEs of 0.076 N (train), 0.085 N (validation), and 0.084 N (test). Orientation accuracy reached 99.70%, 99.87%, and 99.86%, while curvature RMSEs were 0.023 m^{-1}, 0.021 m^{-1}, and 0.018 m^{-1}, respectively.

We finally glued the indenter-dome combined on the sleeve, as shown in the process of Fig. 4(b).

With the goal of being able to predict both force intensity and contact location, each indenter was individually calibrated by using the Instron machine with a flat fixture that gradually pressed on it up to 2 N, for three cycles. Given that there are four indenter along each side of a WG, this corresponds to a total pressing of 8 N, as performed in the general calibration of Section II-C. Indenter number, force applied, and the 5 WG losses were recorded. This process resulted in 7200 data-points collected. Two ML models were created: a regression model to predict force intensity, and a classification model to predict the indenter number that is being pressed. An Exponential Gaussian Process Regression with constant basis was selected for modeling force intensity, while a bagged trees ensemble method was chosen for the classifier, with 500 max number of splits, and 30 learners. Inspired by [21], a cascaded architecture similar to Section II-D was implemented, with the regression model receiving 5 WG signals as inputs, and the classifier receiving both the 5 WG signals and the corresponding predicted force as inputs (see Fig. 4(c)), allowing the classifier to benefit from additional information that may enhance contact localization performance. The models were evaluated on offline datasets (training-80%, validation-10%, testing-10%) and in real-time. 5 fold cross-validation was implemented to prevent overfitting.

IV. Results and Discussion

A. Offline Predictions

The force regression model resulted in an RSME of 0.076 N for the training dataset, 0.085 N for the validation dataset, and 0.084 N for the testing dataset (see Fig. 5(a), (b), (c)), respectively). The orientation classifier was able to estimate the 12 possible orientations with an accuracy of 99.70%, 99.87%, and 99.86% in the training, testing, and validation datasets, respectively (see Fig. 5(d), (e), (f)). Lastly, the curvature model

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

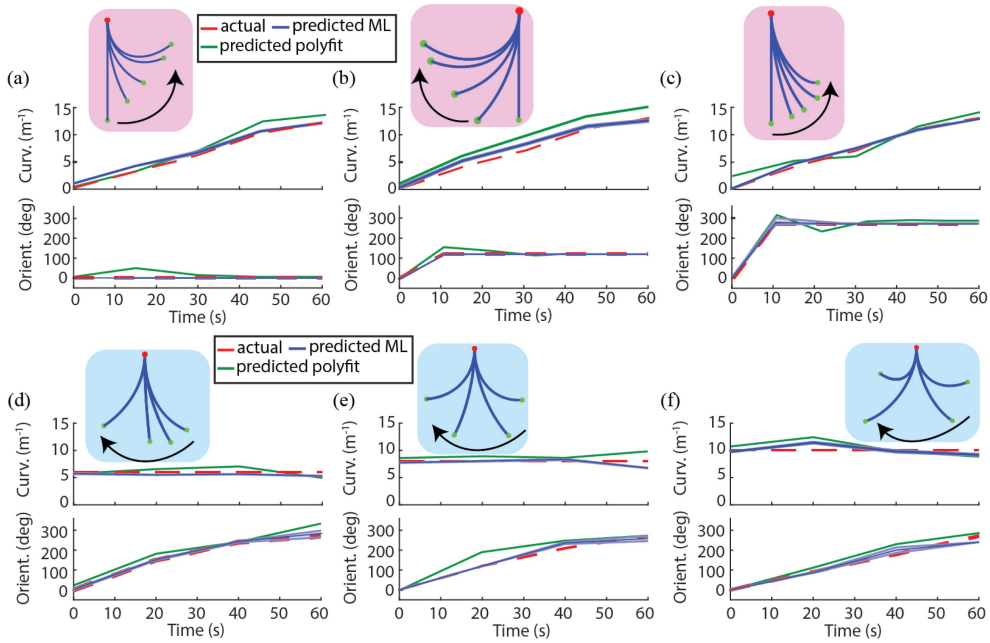

Fig. 6. Shape validation. (a), (b), (c) The sleeve curvature is gradually increased, while keeping the orientation constant, in three different cases [0, 120, 270]°. (d), (e), (f) The sleeve orientation is gradually changed, while keeping a constant curvature, in three different cases [6, 8, 10] m^{-1}. The green curves represent the benchmark polynomial model. See live demonstrations in the Supplementary video.

resulted in an RMSE of 0.023 m^{-1} for training, 0.021 m^{-1} for validation, and 0.018 m^{-1} for testing datasets (see Fig. 5(g), (h), (i), respectively). These results, with minimal deviations from the ground truth, demonstrate detection capabilities from the trained models and suggest that they are likely to generalize well on new data.

B. Real-Time Shape Validation

Fig. 6(a)(b), (c)shows the true and predicted curvature and orientation of the sleeve during the first experiment, where the curvature was varied while keeping a constant orientation. The ML approach was also benchmarked with the previous polynomial model (green curves in Fig. 6). Note that when the sleeve curvature is zero, the orientation is set to zero, which corresponds to the baseline configuration. For the ML approach, the curvature calculated mean absolute error (MAE) was 0.566 m^{-1} for case A, 0.697 m^{-1} for case B, and 0.278 m^{-1} for case C, corresponding to a mean relative error (MRE) of 4.72%, 5.37%, and 2.14%, respectively. The orientation prediction was correct throughout the whole duration of the first two cases (A-B), resulting in a 100% accuracy, while having an average accuracy of 98.07% for case C. For the polynomial method instead, curvature MAE resulted in 1.018 m^{-1}, 0.816 m^{-1}, 1.127 m^{-1} for A, B, and C, that is a MRE of 8.49%, 6.28%, 8.67%, respectively. Orientation accuracy resulted in 91%, 90.5%, and 79%. In Fig. 6(d), (e), (f), the results of the second tests are shown, where the curvature was kept constant and the sleeve was rotated in different orientations. For the ML approach, the curvature MAE resulted in 0.511 m^{-1} for case D, 0.581 m^{-1} for case E, and 0.712 m^{-1} for case F, corresponding to a MRE of 6.12%, 4.66%, and 5.33%, respectively. The orientation predictions resulted in an average accuracy of 98.07%, 95.79%, and 93.89% for case D, E, F, respectively. For the polynomial method, curvature MAE resulted in 0.819 m^{-1}, 1.014 m^{-1}, 0.976 m^{-1} for D, E, and F, that is a MRE of 11.9%, 9.75%, 7.65%, respectively. Orientation accuracy resulted in 73%, 76%, and 83.5%. By comparing the metric of ML approach with the previous polynomial method, the ML techniques enables a significantly higher accuracy in predictions, with prediction relative errors reduced in average from 8.79% to 4.7% for curvature and from 17.8% to 2.37% for orientation estimation.

C. Real-Time Shape and Force Validation

Fig. 7 shows the results of the real-time full ML model predictions of curvature, orientation, and contact force, in the three configurations described in Section III-C, as well as the previous polynomial method. Table I evaluates the real-time MAE, MRE, and accuracy of the two models. In the configuration A (straight sleeve), since the target curvature is zero (making standard relative error undefined) we instead report a normalized relative error, computed by dividing the MAE by the maximum absolute predicted curvature. Among the three cases, the calculated average tracking errors for the ML approach are 5.5% for force, 2.13% for orientation, and 4.67% for curvature. The predicted values accurately follow the ground truth, demonstrating the capabilities of the proposed workflow. Performance decreased slightly during real-time testing compared to offline evaluation. However, this is expected, as real-time conditions introduce additional noise, timing variability, and environmental factors that are less controlled than in offline datasets. Such differences

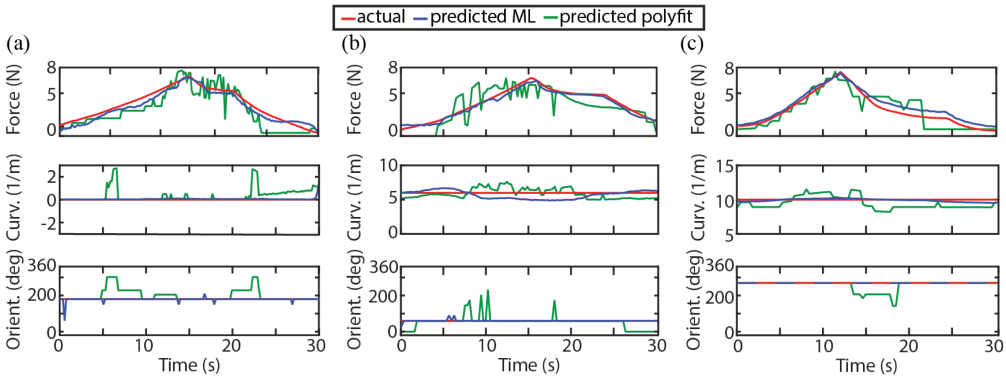

Fig. 7. Real-time predictions of shape and contact force in three scenarios: (a) sleeve is straight and oriented at 180°, (b) sleeve is bent at 6 m^{-1} and oriented at 60°, (c) sleeve is bent at 10 m^{-1} and oriented at 270°. The green curves represent the benchmark polynomial model. See live demonstrations in Supplementary video.

TABLE I
REAL-TIME VALIDATION: ML VS. POLYNOMIAL

Config.	Force	Orientation	Curvature
A (ML)	MAE = 0.45 N; MRE = 5.6%	Acc.=96.29%	MAE = 0.09 m^{-1}; MRE = 3.8%
A (Poly)	MAE = 1.13 N; MRE = 14.2%	Acc.=74.81%	MAE = 1.38m^{-1}; MRE = 12.27%
B (ML)	MAE = 0.40 N; MRE = 5.03%	Acc.=97.34%	MAE = 0.53 m^{-1}; MRE = 6.74%
B (Poly)	MAE = 1.34 N; MRE = 16.86%	Acc.=85.22%	MAE = 1.08 m^{-1}; MRE = 18.04%
C (ML)	MAE = 0.48 N; MRE = 6.02%	Acc.=100%	MAE = 0.25 m^{-1}; MRE = 3.49%
C (Poly)	MAE = 1.05 N; MRE = 13.13%	Acc.=82.95%	MAE = 2.09 m^{-1}; MRE = 20.9%

are common in systems involving real-world sensor data and do not necessarily reflect reduced model robustness. For the polynomial model instead, results show a worse performance, with average tracking errors of 14.73% for force, 19% for orientation, and 17.07% for curvature. This ultimately proves the overall sensing improvement of our ML approach.

D. Contact Localization Results

The offline evaluation of the force regression model resulted in an RMSE of 0.03 N for the training, 0.026 N for the validation and 0.022 N for the test datasets. The classifier had an accuracy of 94.9%, 98.47%, 99.03% for the three datasets, respectively. The slightly higher performance on the validation and test sets may reflect good generalization and relatively consistent patterns across the datasets. The training results are reported in Fig. 8(a). To validate simultaneously the predictions of force intensity and contact localization on real-time, a test was conducted by pressing on each of the 16 indenters respectively through the Instron machine, up to 2 N, similar to the calibration process (Section III-D, setup shown in Fig. 8(c)). Each test was repeated three times to evaluate repeatability. The MAE for force estimation along the 16 indenters is reported in Fig. 8(b). The largest error, 0.31 N, was observed for indenter 3, while the smallest, 0.06 N, occurred with indenter 6. Fig. 8(c) displays the predicted versus true force over time for the top three performing indenters (6, 8, and 14). The predicted force shows a match

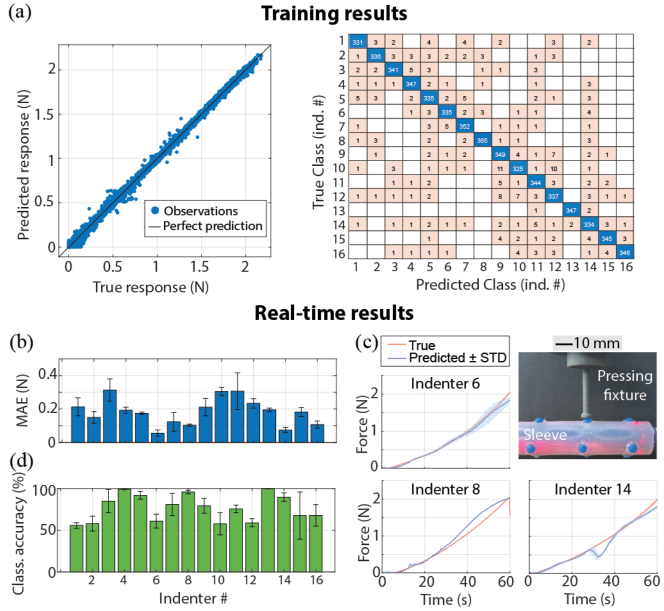

Fig. 8. (a) Training results on the force regression model (left) and indenter classification model (right). (b) Force MAE along the 16 indenters. The maximum error of 0.31 N was on indenter 3 and the minimum was 0.06 N on indenter 6. (c) True vs. predicted force for the top 3 performing indenters and test setup. (d) 8 out of 16 indenters reached a classification accuracy above 80%. See live demonstrations in the supplementary video.

with the ground truth, accurately capturing even small force variations. For classification instead (Fig. 8(d)), indenter 13 resulted in the highest accuracy of 100%, while indenter 1 had the lowest accuracy of 55.7%. The variability of the classification performance reflects minor indenter misalignments, causing diagonal contact and reduced signal quality. Overall, 8 out of 16 locations were predicted with an average accuracy greater than 80% throughout the whole 0–2 N range. As a preliminary study, without yet optimizing for complex indenter geometries or implementing targeted WG design for signal differentiation, these results represent a first step, highlighting the robot's potential for reliable contact localization. Moreover, in practical colonoscopy scenarios, such fine spatial resolution is not necessary, as contact points would be spaced farther apart along the endoscope. This would reduce the classification complexity and could lead

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

to improved performance, reinforcing the applicability of this approach despite some variability in accuracy.

V. Conclusion

This work proposes an ML strategy to exploit the multi-modal nature of soft optical sensors, implemented in a soft robotic sleeve for colonoscopy. An automated calibration platform was developed to collect large datasets on 3D bending and contact force application, eliminating manual error and reducing calibration time. A data-driven cascaded architecture enables accurate prediction of both contact force and 3D shape by a sequential structure which ensures that shape predictions are conditioned on contact occurrence, preventing force-related distortions from influencing shape estimation. Improved performance, compared to previous analytical approach, was achieved across varying robot configurations, with average errors of 4.7% for curvature, 2.37% for orientation, and 5.5% for force tracking. Furthermore, we introduced a preliminary contact force localization study on the sleeve surface. By implementing a sharper indenter shape, the sleeve increased its sensitivity to contact force up to 0.2 N. A cascaded ML architecture can accurately estimate both contact localization and intensity along the sleeve surface, even in a densely packed sensing layout where the 16 indenters are positioned in close proximity ($\sim$2.5 cm apart), with 8 aligned along the same optical WG. Force intensity was estimated with low error (0.06–0.31 N), and localization accuracy greater than 80% was achieved by 8 indenters. This level of spatial resolution is particularly promising given the overlapping and coupled optical signals, highlighting the robot's ability to interpret complex responses.

Future work will focus on improving ML models, evaluating simultaneous contact at multiple points, modeling complex shapes, and on integrating contact localization with shape sensing. Ultimately, the system is intended to evolve into a longer sleeve extending over the entire endoscope, capable of 3D shape reconstruction, real-time contact localization, and force estimation along its full length, in order to improve navigation, safety, and user feedback during colonoscopy.

References

[1] H. Wang, M. Totaro, and L. Beccai, "Toward perceptive soft robots: Progress and challenges," *Adv. Sci.*, vol. 5, no. 9, 2018, Art. no. 1800541.

[2] P. Polygerinos et al., "Soft robotics: Review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human-robot interaction," *Adv. Eng. Mater.*, vol. 19, no. 12, 2017, Art. no. 1700016.

[3] K. Chin, T. Hellebrekers, and C. Majidi, "Machine learning for soft robotic sensing and control," *Adv. Intell. Syst.*, vol. 2, no. 6, 2020, Art. no. 1900171.

[4] D. Kimet al., "Review of machine learning methods in soft robotics," *PLoS One*, vol. 16, no. 2, 2021, Art. no. e0246102.

[5] J. Liu, Y. Duo, X. Chen, Z. Zuo, Y. Liu, and L. Wen, "Data-driven methods for sensing, modeling and control of soft continuum robot: A review," *IEEE/ASME Trans. Mechatron.*, early access, May 23, 2025, doi: 10.1109/TMECH.2025.3566915.

[6] T. Čakurda, M. Trojanová, P. Pomin, and A. Hošovsk, "Deep learning methods in soft robotics: Architectures and applications," *Adv. Intell. Syst.*, vol. 7, no. 5, 2025, Art. no. 2400576.

[7] Z. Chen et al., "Data-driven methods applied to soft robot modeling and control: A review," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 2241–2256, 2025.

[8] J. Wang and A. Chortos, "Control strategies for soft robot systems," *Adv. Intell. Syst.*, vol. 4, no. 5, 2022, Art. no. 2100165.

[9] S. Shu, Z. Wang, P. Chen, J. Zhong, W. Tang, and Z. L. Wang, "Machine-learning assisted electronic skins capable of proprioception and exteroception in soft robotics," *Adv. Mater.*, vol. 35, no. 18, 2023, Art. no. 2211385.

[10] T. G. Thuruthel, B. Shih, C. Laschi, and M. T. Tolley, "Soft robot perception using embedded soft sensors and recurrent neural networks," *Sci. Robot.*, vol. 4, no. 26, 2019, Art. no. eaav1488.

[11] N. Pagliarani et al., "SoftTex: Soft robotic arm with learning-based textile proprioception," *IEEE Robot. Automat. Lett.*, vol. 10, no. 4, pp. 3779–3786, Apr. 2025.

[12] P. Wang et al., "Sensing expectation enables simultaneous proprioception and contact detection in an intelligent soft continuum robot," *Nature Commun.*, vol. 15, no. 1, 2024, Art. no. 9978.

[13] H. Jang, J. Bae, and K. Haninger, "Tactile sensor-based estimation of grasp force and contact state with soft fingers," *IEEE Robot. Automat. Lett.*, vol. 10, no. 7, pp. 7246–7253, Jul. 2025.

[14] S. Zhou et al., "Unlocking dynamic subtle stimuli tactile perception: A deep learning-enhanced super-resolution tactile sensor array with rapid response," *Adv. Intell. Syst.*, vol. 7, no. 5, 2025, Art. no. 2400913.

[15] L. Wang et al., "Soft robot proprioception using unified soft body encoding and recurrent neural network," *Soft Robot.*, vol. 10, no. 4, pp. 825–837, 2023.

[16] J. Wang, Z. Qiao, W. Zhang, and S. Li, "Proprioceptive and exteroceptive information perception in a fabric soft robotic arm via physical reservoir computing with minimal training data," *Adv. Intell. Syst.*, vol. 7, no. 4, 2025, Art. no. 2400534.

[17] S. Belk, S. Rosset, I. Anderson, and M. Hesam, "From single sensors to sensor arrays: Leveraging supervised machine learning to read multiple soft capacitive sensors with a single pair of wires," *Adv. Intell. Syst.*, vol. 7, no. 7, 2025, Art. no. 2400773.

[18] B. Shih et al., "Electronic skins and machine learning for intelligent soft robots," *Sci. Robot.*, vol. 5, no. 41, 2020, Art. no. eaaz9239.

[19] A. Gerald and S. Russo, "Soft sensing and haptics for medical procedures," *Nature Rev. Mater.*, vol. 9, no. 2, pp. 86–88, 2024.

[20] Y. Ni, S. Wakimoto, W. Tian, Y. Toda, T. Kanda, and D. Yamaguchi, "Length estimation of pneumatic artificial muscle with optical fiber sensor using machine learning," *Sensors*, vol. 25, no. 7, 2025, Art. no. 2221.

[21] T. Li et al., "An artificial intelligence-motivated skin-like optical fiber tactile sensor," *Adv. Intell. Syst.*, vol. 5, no. 8, 2023, Art. no. 2200460.

[22] N. Pagliarani, M. Filosa, R. M. A. Ayaz, C. J. Oton, C. M. Oddo, and M. Cianchetti, "Enriching contact information through fiber Bragg gratings-based exteroception in soft bending actuators," in *Proc. IEEE 7th Int. Conf. Soft Robot.*, 2024, pp. 1082–1087.

[23] L. Massari et al., "A machine-learning-based approach to solve both contact location and force in soft material tactile sensors," *Soft Robot.*, vol. 7, no. 4, pp. 409–420, 2020.

[24] J. Chen, B. Liu, and H. Zhang, "Review of fiber Bragg grating sensor technology," *Front. Optoelectron. China*, vol. 4, no. 2, pp. 204–212, 2011.

[25] D. Rajput, W.-J. Wang, and C.-C. Chen, "Evaluation of a decided sample size in machine learning applications," *BMC Bioinf.*, vol. 24, no. 1, 2023, Art. no. 48.

[26] Y. Gong, G. Liu, Y. Xue, R. Li, and L. Meng, "A survey on dataset quality in machine learning," *Inf. Softw. Technol.*, vol. 162, 2023, Art. no. 107268.

[27] V. Del Bono et al., "A soft robotic 'add-on' for colonoscopy: Increasing safety and comfort through force monitoring," *npj Robot.*, vol. 3, no. 1, pp. 1–17, 2025.

[28] W.-B. Cheng et al., "Modeling and in vitro experimental validation for kinetics of the colonoscope in colonoscopy," *Ann. Biomed. Eng.*, vol. 41, no. 5, pp. 1084–1093, 2013.

[29] L. Y. Korman et al., "Characterization of forces applied by endoscopists during colonoscopy by using a wireless colonoscopy force monitor," *Gastrointestinal Endoscopy*, vol. 71, no. 2, pp. 327–334, 2010.

[30] D. Wuet al., "A review on machine learning in flexible surgical and interventional robots: Where we are and where we are going," *Biomed. Signal Process. Control*, vol. 93, 2024, Art. no. 106179.