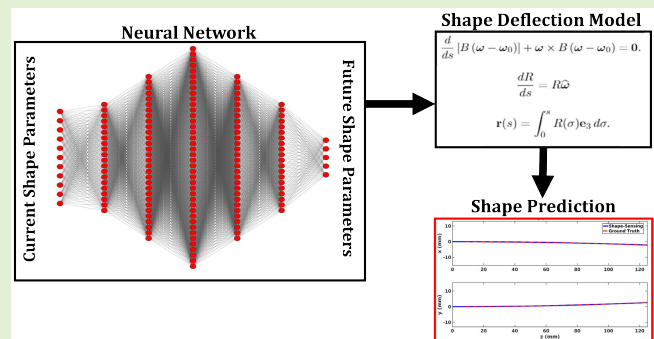


Hybrid Deep Learning and Model-Based Needle Shape Prediction

Dimitri A. Lezcano¹, Yernar Zhetpissov¹, Mariana C. Bernardes¹, *Member, IEEE*,
Pedro Moreira¹, *Member, IEEE*, Junichi Tokuda, *Member, IEEE*, Jin Seob Kim¹, *Member, IEEE*,
and Iulian I. Iordachita¹, *Senior Member, IEEE*

Abstract—Needle insertion using flexible bevel-tip needles is a common minimally invasive surgical technique for prostate cancer interventions. Flexible, asymmetric bevel-tip needles enable physicians for complex needle steering techniques to avoid sensitive anatomical structures during needle insertion. For accurate placement of the needle, predicting the trajectory of these needles intraoperatively would greatly reduce the need for frequently needle reinsertions, thus improving patient comfort and positive outcomes. However, predicting the trajectory of the needle during insertion is a complex task that has yet to be solved due to random needle–tissue interactions. In this article, we present and validate, for the first time, a hybrid deep learning and model-based approach to handle the intraoperative needle shape prediction problem through leveraging a validated Lie-group theoretic model for needle shape representation. Furthermore, we present a novel self-supervised learning (SSL) and method in conjunction with the Lie-group shape model for training these networks in the absence of data, enabling further refinement of these networks with transfer learning (TL). Needle shape prediction was performed in single-layer and double-layer homogeneous phantom tissue for C- and S-shaped needle insertions. Our method demonstrates an average root-mean-square prediction error of 1.03 mm over a dataset containing approximately 3000 prediction samples with the maximum prediction steps of 110 mm.

Index Terms—Deep learning, flexible needle, machine learning, medical device, model-based, shape prediction.



I. INTRODUCTION

PROSTATE cancer is a potentially fatal disease affecting every 1 in 8 men and is the second leading cause of cancer death for men in the United States of America killing every 1 in 41 men as of 2023 [1], [2]. To diagnose and treat prostate cancer, needle insertion procedures are

Manuscript received 18 March 2024; revised 27 March 2024; accepted 27 March 2024. Date of publication 12 April 2024; date of current version 31 May 2024. This work was supported in part by the National Institutes of Health under Grant R01CA235134 and in part by the Johns Hopkins University Internal Funds. The associate editor coordinating the review of this article and approving it for publication was Dr. Rahul Bhattacharyya. (Jin Seob Kim and Iulian I. Iordachita are co-senior authors.) (Corresponding author: Dimitri A. Lezcano.)

Dimitri A. Lezcano, Jin Seob Kim, and Iulian I. Iordachita are with the Mechanical Engineering Department, Johns Hopkins University, Baltimore, MD 21201 USA (e-mail: dlezcan1@jhu.edu; jkim115@jhu.edu; iordachita@jhu.edu).

Yernar Zhetpissov was with the Mechanical Engineering Department, Johns Hopkins University, Baltimore, MD 21201 USA (e-mail: yzhetpi1@jhu.edu).

Mariana C. Bernardes, Pedro Moreira, and Junichi Tokuda were with the Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115 USA (e-mail: mcostabernardesmatias@bwh.harvard.edu; plopesdafrotamoreira@bwh.harvard.edu; tokuda@bwh.harvard.edu).

Digital Object Identifier 10.1109/JSEN.2024.3386120

common minimally invasive surgical techniques used for biopsying potentially cancerous tissue as well as for prostate cryoablation [2], [3], [4]. Due to challenges during needle insertion, reinsertions of the needle are frequently performed to adjust needle positioning and obtain many samples, damaging surrounding tissue and nearby sensitive anatomical structures resulting in increased postoperative discomfort for the patient [5]. To improve patient outcomes, needle insertion targeting error must be mitigated, thus requiring guidance solutions to provide needle placement feedback to the physician [6]. If placement feedback is combined with an asymmetric bevel-tip needle, the physician can actively compensate for the needle deviation by changing the orientation of the bevel during insertion. Typically, real-time surgical guidance to track the needle's trajectory intraoperatively includes ultrasound imaging [7], [8], [9], [10], [11], [12], [13], [14], [15], [16] and magnetic resonance imaging (MRI) [17], [18], [19]. Computational tomography is another typical modality for surgical guidance, but requires large doses of radiation to the patient for near-to-real-time feedback [20], [21], [22]. Furthermore, intraoperative real-time imaging typically is 2-D, making it difficult for

accurate tracking of the needle's position. Alternatively, an up-and-coming nonimage-based solution uses bevel-tip needles embedded with fiber-Bragg grating (FBG) sensors to provide real-time 3-D needle shape feedback [8], [23], [24], [25], [26], [27], [28]. FBG sensors are MRI-compatible optical fibers that measure local strain induced in the optical cables at sensing locations along the fiber. In [29] and [30], we demonstrated that using these local curvature estimates from the embedded FBGs in conjunction with a Lie-group theoretic model we can achieve sub-millimeter accuracy in phantom and real tissue for various needle and tissue configurations.

For accurate placement of bevel-tip needles, multiple different methods were studied to predict the trajectory of needles during needle insertion. Typical methods for predicting the trajectory of bevel-tip needles include basic kinematic or mechanics-based models and finite-element modeling. Kinematic models of the needle usually describe the curvature of the needle through a constant curvature parameterization using either the unicycle or bicycle model [31], [32], [33], [34], [35], [36], [37], [38]. These models are lightweight and well defined; however, these do not characterize the nonconstant curvature the needle experiences inhibiting their ability to predict the needle shape upon longer insertions. Mechanics-based methods include beam mechanics modeled through virtual springs and modeling the needle-tissue interaction through parameterizing the tissue characteristics [3], [6], [39], [40], [41], [42], [43]. These models are complex and capable of capturing the needle's nonconstant curvature over longer insertions, but require preoperative information that would vary from patient to patient and are difficult to adapt on-the-fly. In our previous work, we demonstrated a sensor-based method in conjunction with a Lie-group theoretic model that leverages beam mechanics to flexibly optimize mechanics-based parameters on-the-fly using FBG sensors [29], [30], [44]. Furthermore, we performed a preliminary study in using data-driven methods to flexibly predict the trajectory of flexible bevel-tip needles in single- and double-layer tissue in [45].

Data-driven and machine learning-based methods in sensorized flexible needles, catheters, and continuum manipulators are on the rise in recent years. Data-driven models, specifically deep learning models, enable complex, nonlinear modeling grounded in experimental evidence for guided needle insertion; however, this typically requires large experimental datasets, including innumerable variations of needle insertion scenarios for accurate and precise performance over all insertion scenarios. The current state-of-art of machine learning in flexible needles is contained within methods for shape sensing, force sensing, autonomous control, and needle-tip prediction. The potential for deep learning methods in needle deflection was evaluated in a comparison study for phantom tissue insertions between Eulerian beam theory and multilayer perceptron [46]. Needle deflection was estimated using a local, just-in-time model from the robotic insertion platform inputs [47]. Under ultrasound guidance, a flexible needle is tracked using Kalman filtering, and the tracked shape is used to fit a mechanics-based needle-tissue interaction to predict needle-tip deflection and predicts needle deflection purely from a mechanics-based model and does not provide on-the-

fly prediction updates [16]. A dynamic linear autoregressive method was used to model the needle-tissue dynamics from a fixed window of 2-D needle coordinates and robot inputs in conjunction with a model-predictive controller for needle placement [48]. Deep learning was used in catheters for shape sensing, mapping FBG sensor feedback to the curvature of the catheter [49]. Convolutional neural networks were employed with optical CT fiber-embedded needle for needle-tip force estimation [50]. Continuum manipulators embedded with FBGs used linear regression (LR) to map the FBG sensor feedback to the estimated tip position of the manipulator [51]. Furthermore, gradient boosted random forests were used to detect manipulator collision based on image and FBG feedback [52]. Universally distributed Q learning was used to steer a flexible radio frequency ablative needle in phantom liver tissue from CT-based needle tracking [53]. In [54], a two-degree-of-freedom (DoF) continuum robot demonstrated shape sensing using neural networks based on FBG sensor readings. Near-infrared and ultrasound imaging were combined in conjunction with a robotic platform for autonomous cannulation in vitro using recurrent fully convolutional neural networks for image guidance [55]. Time-batched spectra from edge FBGs were used for extracting the sensor's 3-D shape with convolutional neural networks [56]. A 3-D UNet in conjunction with actor-critic deep reinforcement learning demonstrated end-to-end 3-D needle shape trajectory prediction in CT by predicting the future CT mask of the needle at a defined insertion step [57]. An on-the-fly adaptive Jacobian estimation algorithm was used to predict needle-tip pose upon lateral deflection of the needle base for robot-assisted needle insertion [58].

In this article, we present, for the first time, a hybrid deep learning and model-based method for needle shape prediction extending upon our previous work in [45]. A simple exponential fit is used to predict that the future shape parameters of the needle are provided in [45] in conjunction with a shape model. However, this method is limited to small insertion step sizes, 10-mm maximum prediction range, and C-shaped configurations only. Thus, to improve the accuracy of shape prediction over a wider range of insertion scenarios and larger prediction ranges, we use deep networks to improve future shape parameter prediction. We also present a novel self-supervised learning (SSL) approach in order to train neural networks for 3-D needle shape prediction without experimental training data. This hybrid method leverages the flexible, proven, Lie-group theoretic model presented in [59], [60], and [29] using FBG sensor curvature information to estimate the model's shape-sensing parameters. Deep learning methods, including neural networks, transfer learning (TL), and SSL, are used collectively to estimate shape parameters for 3-D shape prediction. Pure model-based methods for needle shape prediction are limited either to their understanding of the needle-tissue interactions or to oversimplified kinematics-based assumptions. Pure data-driven methods can suffer from the lack of regularization of the model's output resulting in solutions that are not physically feasible. Thus, the necessity for a hybrid data-driven and model-based approach conjoins the benefits of providing physically feasible solutions with a model-based method and the complexity of a data-driven

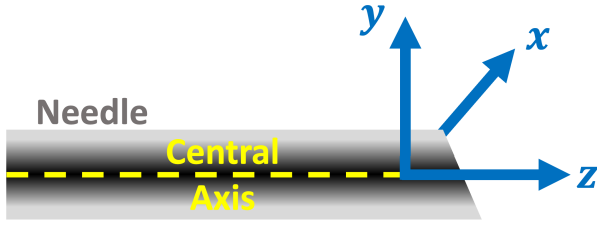


Fig. 1. Body-fixed axis orientation of the needle defined to the Lie-group theoretic shape model.

method, warranting the need for the presented hybrid approach to increase accuracy of needle shape prediction. The contributions of this article include the following:

- 1) the presentation of, for the first time, a hybrid deep learning and model-based method for needle shape prediction leveraging a sensor-based Lie-group theoretic needle shape model;
- 2) a novel SSL approach to train neural networks to achieve 3-D needle shape prediction in the absence of experimental data;
- 3) a TL approach to combine the presented supervised learning (SL) and SSL methods to improve training of deep networks in the presented hybrid approach to shape prediction with limited data.

In contrast, hybrid deep learning and model-based methods leverage the learning complexity of deep networks in conjunction with constraining the network's output to physically feasible solutions with a model. Training and test data are generated from [29] for needle insertions in phantom tissue for an 18G FBG-sensorized needle for varying needle insertion configurations. Prediction data include C-shape (nonrotation), insertions into single- and double-layer homogeneous tissue, and S-shape (180° rotation), insertions into single-layer homogeneous tissue. In Section II, we present our Lie-group theoretic shape and data-driven models for needle shape prediction. Section III-B demonstrates the experiments conducted in this study to evaluate our shape prediction framework. Results are presented in Section IV and discussed in Section V. Finally, Section VI concludes our paper with limitations of this framework and future research directions.

II. MODELS AND METHODS

A. Needle Shape Sensing

Our Lie-group theoretic model for shape sensing [29], [59] describes the curvature (ω_1 and ω_2 along the needle's local x - and y -axes, respectively) and torsion (ω_3 along the local z -axis) of the needle as follows:

$$\boldsymbol{\omega}(s) = [\omega_1 \ \omega_2 \ \omega_3]^T = \left(R^T(s) \frac{dR(s)}{ds} \right)^\vee \quad (1)$$

where $R(s) \in SO(3)$ denotes the rotation matrix in 3-D space, shown in Fig. 1, describing the orientation of the body-fixed frame attached at each point along the needle. $s \in [0, L]$ denotes the arclength of the needle with total insertion depth L . The \vee operation defines a 3-D vector associated with a 3×3 skew-symmetric matrix ($R^T(dR/ds)$) [61].

In tissue, the bevel-tipped needle experiences distributed loads in the direction of the bevel and can be modeled as an

inextensible elastic rod. As in homogeneous tissue, we assume that the uniform loads are distributed along the needle. Ideally, the needle will deflect in a single plane due to the needle's bevel tip. The intrinsic curvature of the needle, $\kappa_0(s)$, is used to capture this interaction in modeling the needle's natural bending curvature.

In this work, we explore the single-layer C-shaped, double-layer C-shaped, and single-layer S-shaped insertion scenarios. Each of these different insertion scenarios requires a different intrinsic curvature model with different κ_c parameters, which are described in our previous work [29]. For the single-layer C-shaped case

$$\kappa_0(s) = \kappa_c \left(1 - \frac{s}{L} \right)^2. \quad (2)$$

For the double-layer C-shaped case, we define two intrinsic curvature coefficients as $\kappa_{c,1}$ and $\kappa_{c,2}$ for the first and second layers, respectively, and the length of the needle in the first layer to be s^* . Thus, $\kappa_0(s)$ for the double-layer C-shaped case is given by

$$\kappa_0(s) = \begin{cases} \kappa_{c,1} \left(\frac{s^* - s}{L} \right)^2 + \kappa_{c,2} \left(1 - \frac{s^*}{L} \right) \left(1 + \frac{s^*}{L} - \frac{2s}{L} \right) & (0 \leq s \leq s^*) \\ \kappa_{c,2} \left(1 - \frac{s}{L} \right)^2 & (s^* \leq s \leq L). \end{cases} \quad (3)$$

For the single-layer S-shaped case, the needle is rotated 180° about its central axis, at some insertion depth, s^* , thus inducing a point of inflection. Therefore, similar to the single-layer C-shaped insertion, we define $\kappa_0(s)$ for the single-layer S-shaped insertion to be

$$\kappa_0(s) = \begin{cases} \kappa_c \left(\frac{s^*}{L} \right)^{2/3} \left(1 - \frac{s}{L} \right)^2, & 0 < s < s^* \\ \frac{\kappa_c}{2} \left(\left(\frac{s^*}{L} \right)^{2/3} - \left(\frac{L - s^*}{L} \right)^{2/3} \right) \left(1 - \frac{s}{L} \right)^2, & s = s^* \\ -\kappa_c \left(\frac{L - s^*}{L} \right)^{2/3} \left(1 - \frac{s}{L} \right)^2, & s^* < s \leq L. \end{cases} \quad (4)$$

The factor of $2/3$ in (4) was used as an approximative model to balance the tissue's interaction with the needle based upon the length of the needle preinsertion and postinsertion, as seen in [45].

Needle deformation, $\boldsymbol{\omega}(s)$, and the body-fixed orientation, $R(s)$, along the needle shaft can be determined through minimizing the elastic potential energy of the rod needle deformation. As seen in [62], [63], and [29], the Euler–Poincaré equation is a solution to minimize the elastic potential energy

$$\frac{d}{ds} [B(\boldsymbol{\omega} - \boldsymbol{\omega}_0)] + \boldsymbol{\omega} \times B(\boldsymbol{\omega} - \boldsymbol{\omega}_0) = \mathbf{0} \quad (5)$$

which will be solved together with (1). Here, B is stiffness matrix of the needle characterizing the bending and torsional stiffnesses, and the 3-D intrinsic curvature $\boldsymbol{\omega}_0 = [\kappa_0 \ 0 \ 0]^T$.

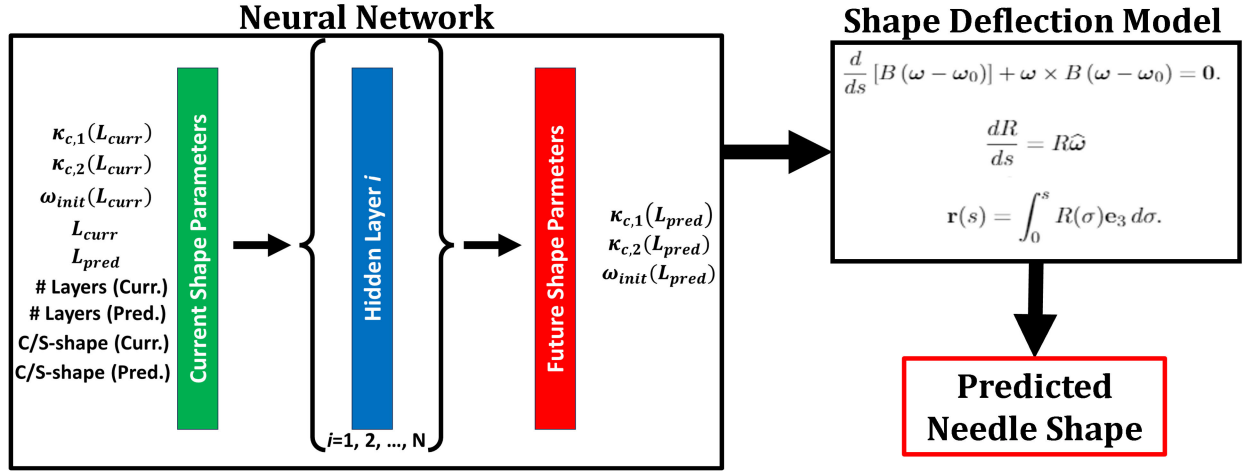


Fig. 2. Hybrid approach's framework to needle shape prediction. A fully connected neural network is used to predict the future shape parameters used in the presented Lie-group theoretic shape deflection model, based on the current shape parameters. Using the future parameters at a desired insertion depth, a predicted needle shape at the desired insertion depth is achieved. The definition of "(Current/Future) Shape Parameters" are provided in Section II-B.

In integrating (5), the initial angular deviation parameter, ω_{init} , is introduced to provide as the initial condition. Finally, through integrating the solution to (5), the needle shape, $\mathbf{r}(s)$, can be determined

$$\mathbf{r}(s) = \int_0^s R(\sigma) \mathbf{e}_3 d\sigma \quad (6)$$

where $\mathbf{e}_3 = [0 \ 0 \ 1]^T$.

We will define the function that solves Euler-Poincaré equation, (5), in conjunction with (1) and (6) using the intrinsic curvature equations in [29] as $f_{shape} : \mathbb{R}^5 \rightarrow SE(3) \times \mathbb{R}$. The input of this model is $(\kappa_{c,1}, \kappa_{c,2}, \omega_{init}^T)$. In the implementation of these differential equations, the finite-difference method was used with tensorized representations of the shape parameters, enabling backpropagated derivatives to be passed for machine learning methods based on gradient descent.

B. Needle Shape Prediction

In order to leverage the proven Lie-group theoretic shape model, machine learning is used to predict the future parameters for shape-sensing model. Through altering the shape model, we enable machine learning frameworks to learn the shape prediction through a simplified and information-rich space of shape parameters, as illustrated in Fig. 2. The input space we train machine learning models with is the following:

- 1) insertion depth for current and target prediction (L_{curr} and L_{pred});
- 2) whether the needle is currently and will be in an S-shaped configuration;
- 3) whether the needle is currently and will be in two layers;
- 4) the current κ_c estimates for layers 1 and 2 from (3) (if single layer, $\kappa_{c,2} = -1$), $\kappa_{c,1,curr}$ and $\kappa_{c,2,curr}$;
- 5) the current estimate for ω_{init} and $\omega_{init,curr}$.

The output space we require for solving the needle shape prediction problem is the following:

- 1) the predicted estimate for ω_{init} at L_{pred} and $\omega_{init,pred}$;
- 2) the predicted estimates for $\kappa_{c,1}$ and $\kappa_{c,2}$ at L_{pred} , $\kappa_{c,1,pred}$, and $\kappa_{c,2,pred}$.

Experimental insertion results are split based on the type of insertion scenario performed in each insertion experiment.

Machine learning models are trained using soft and hard one-layer tissue C-shaped insertions, soft-hard two-layer tissue C-shaped insertions, and hard one-layer S-shaped insertions. Tissue stiffness classifications are taken from [29]. These left hard-soft two-layer tissue C-shaped insertions and soft one-layer S-shaped insertions. We use input data normalization to improve learning results, such as the following.

- 1) insertion depth normalization (divide by needle's total length);
- 2) zero-centering and unit normalization of κ_c and ω_{init} values, based on experimental data observed;
- 3) binarization of logical input values, such as the S-shaped configuration and two-layer configuration.

1) **Linear Regression:** We first attempt to use a simple method to predict the needle shape parameters. Simply, we use LR to predict the output model parameters using the input variables listed in Section II-B with a single matrix. LR models assume that the general solution forms for an input, $\mathbf{x} \in \mathbb{R}^n$, and output $\mathbf{y} \in \mathbb{R}^m$ looks like

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad (7)$$

where $\mathbf{A} \in \mathbb{R}^{n \times m}$. In general, these models are simple to solve and, however, are unable to model inherent nonlinear relationships between the input dimensions, without the use of feature engineering [64].

2) **Fully Connected Neural Network:** We use a fully connected network to predict the parameters directly from gradient descent. Nonlinear activations are used in between the hidden layers of the network. The ReLU activation function from (8) is used as the nonlinear activation here. We test various network layer depths and sizes in our parameter exploration

$$\text{ReLU}(x) = \max(0, x). \quad (8)$$

The ReLU activation function (8) provides a nonlinear activation and mitigates the vanishing gradient problem found for common deep learning solutions, improving the network's ability to learn [65].

In the shape-sensing model, optimization bounds for sensor feedback when determining shape parameters are defined as

$\kappa_c \in [0, 0.01]$ and $\omega_{\text{init}} \in [-0.01, 0.01]^3$. Therefore, for the output layer, we use a concatenation of the sigmoid, $\sigma(x)$, and tanh nonlinear activation functions scaled by the 0.01 for the κ_c and ω_{init} predicted outputs, respectively,

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (9)$$

The sigmoid activation (9) is used for κ_c to constrain the output space to all nonnegative values upper-bounded. tanh activation is used for ω_{init} predictions to symmetrically bound the output space of our model. By providing these bounds to the model, we reduce the model's effort in learning, allowing for more of the focus to be on learning the relationship between the input and output spaces. Since the only use of sigmoid and tanh activations is at the classification section of the network, the vanishing gradient problem is not a concern, since ReLU nonlinear activations are used through the rest of the network.

C. Loss Functions

In this work, we use various different loss functions, experimented with training data-driven models for needle shape prediction.

1) **Mean-Squared Error:** We define a commonly used loss function mean-squared error (MSE) as follows:

$$\text{MSE}(\{y_i\}, \{\hat{y}_i\}) = \frac{1}{N} \sum_{i=1}^N \|\|_2[2] y_i - \hat{y}_i\|^2 \quad (10)$$

where $\|\|_2$ is the L_2 norm.

2) **Shape MSE:** Using the above loss functions provides a functional way to learn the parameters from given data; however, using a differentiable shape-sensing model, we enable our models to directly learn to reduce shape-prediction error. We define the shape MSE ($\text{MSE}_{\text{shape}}$) as the MSE from the predicted shape and the determined shape, using the predicted and target shape parameters, directly. Using f_{shape} from Section II-A, we define $\text{MSE}_{\text{shape}}$ as follows:

$$\begin{aligned} \text{MSE}_{\text{shape}}(\{y_i\}, \{\hat{y}_i\}) &= \frac{1}{N} \sum_{i=1}^N \\ &\times \frac{1}{L_i} \int_0^{L_i} \|\|_2[2] f_{\text{shape}}(y_i)(s) - f_{\text{shape}}(\hat{y}_i)(s)\|^2 ds \end{aligned} \quad (11)$$

where L_i is the insertion depth for the i th prediction.

3) **Self-Supervised Shape Mean-Squared Error:** Finally, extending upon (11), we can extend this method for an SSL approach where we compare projected needle shapes at varying insertion depths. We subscript f_{shape} by L as $f_{\text{shape}}(y; L)$ for computing the needle shape using shape parameters at an insertion depth of L . Thus, we define the SSL loss function of the shape MSE ($\text{MSE}_{\text{ss,shape}}$) to compare parameters $\{y_i\}$ and $\{z_i\}$ predicted at insertion depths $\{L_{y,i}\}$ and $\{L_{z,i}\}$ as follows:

$$\begin{aligned} \text{MSE}_{\text{ss,shape}}(\{y_i\}, \{z_i\}) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\min(L_{y,i}, L_{z,i})} \\ &\times \int_0^{\min(L_{y,i}, L_{z,i})} \|\|_2[2] f_{\text{shape}}(y_i; L_{y,i})(s) - f_{\text{shape}}(z_i; L_{z,i})(s)\|^2 ds. \end{aligned} \quad (12)$$

D. Deep Learning Techniques

To train the deep networks mentioned in Section II-B2, we explore a variety of learning techniques, including SL and SSL, and combining supervised and SSL with TL.

1) **Supervised Learning:** SL optimizes the network to best fit the output with labeled experimental data. Labeled experimental data provide real-world examples to the network to regress upon. SL takes an initialized network, usually randomly initialized, and samples the labeled dataset to iteratively minimize a loss function mentioned in Section II-C until the model's learning converges. The benefit of SL is that the networks learn to directly solve the problem at hand with real-world experience. However, SL restricts the network's learned domain to the domain of the dataset, and networks will not usually generalize to examples taken outside of the trained dataset.

2) **Self-Supervised Learning:** Instead of using a randomly weighted initialized deep model for parameter prediction, we leverage a deep learning techniques, SSL, to better train deep models with SL. SSL trains models on pretext tasks leveraging domain knowledge on unlabeled data [66]. SSL is typically used for image-based problems, leveraging the simple task of training a network to convert a colored image to grayscale [67]. The SSL approach presented here relies upon the assumption that upon further insertion, the needle will follow its currently carved trajectory [45]. Thus, given shapes $\mathbf{p}_1(s)$ and $\mathbf{p}_2(s)$, with arclengths L_1 and L_2 , respectively, then, ideally, we would observe, $\mathbf{p}_1(s) = \mathbf{p}_2(s) \forall s \in [0, \min(L_1, L_2)]$. Using this, we generate random reference shape parameters at some random insertion depth, compute the predicted parameters at varying insertion depths to forecast the needle shape at these varying arclengths, and compare the overlapping sections of the reference and predicted shapes using (12).

The benefit of SSL is that the model can teach itself sensible solutions to situations it has yet to observe, mitigating any potential issues with limited dataset amount, since it is able to generate infinite insertion scenarios. The purpose of using this method is to teach the model an initial, sensible estimate of what solutions to needle shape prediction should look like, rather than starting from random. More specifically, when training any neural network, typically, model parameters are initialized with random weights, which may not provide physically feasible needle shape predictions. The application of SSL in this work is to take a randomly weighted initialized neural network and train the network without experimental insertion data with a loss function to inform the model of physically feasible solutions to needle shape prediction. Furthermore, this method exposes the network to insertion scenarios that are not observed in the training dataset, better simulating difficult shape-prediction tasks.

3) **Transfer Learning:** To further refine self-supervised learned models, we use another deep learning technique, TL. TL is a tool to leverage networks with knowledge in a set of tasks that are different than the problem at hand [68]. In deep learning, a typical TL approach is to use a model trained on either different data and/or a different loss function than one that applies to the current data and task. We combine SL and

SSL in order to leverage the benefits of SSL mentioned in Section II-D2 and the experience of experimentally observed insertions from SL. For TL, we first train a model with SSL method and then transfer the model's training to the SL technique for further refinement of the model's experience in solving the shape-prediction problem. After the model is trained using SSL, we further train the deep model with our presented SL technique, via TL, to better refine the model's understanding of solving the shape-prediction problem using experimentally observed results. Namely, we take a model that was trained using SSL with the SSL loss function in (12) and train that model on experimental data with the SL loss function (11). Since SL requires a large dataset to efficiently train deep networks, SSL provides the model a baseline understanding of how to perform shape prediction followed by SL providing the model real-world experience for shape prediction. The SL refinement is necessary, because the aforementioned assumption was shown in [29] to not be completely accurate. Thus, we hypothesize that using SSL in conjunction with SL, we can better train our networks and handle the overfitting problem with this infinite-dimensional training.

III. EXPERIMENTS

Experimental data are generated from [29] for insertions into phantom tissue for single-layer homogeneous tissue for C- and S-shaped insertions and double-layer homogeneous tissue for C-shaped insertions. Insertions were performed into phantom tissue in 10-mm increments from 15 up to 125 mm in insertion depth. Soft and hard homogeneous phantom tissues were made with the hard tissue to have approximately a 50% increase in stiffness. Ground truth was provided from 3-D stereo reconstruction in [29], and shape-sensing errors were reported to average 0.160 ± 0.055 mm for over 650 needle insertions for needle insertions with C-shape in one- and two-layer tissue, and S-shape in one-layer tissue.

To create the machine learning dataset, we pair all shape-sensing results for each insertion depth with all of the shallower depths (i.e., a reference insertion depth of 35 mm predicts the needle shape at 45 mm, 55 mm, and so on up to 125 mm). Thus, we generate an experimental dataset size of 2944 prediction sampled pairs. Training and test splits are not split uniformly over the samples; instead, the training and test split was generated using the following needle insertion configurations.

- 1) Training (sample size = 2116|71.8% of dataset):
 - a) C-shape for soft and hard single-layer tissue;
 - b) S-shape for hard single-layer tissue;
 - c) C-shape for soft-hard double-layer tissue.
- 2) Test (sample size = 828|28.2% of dataset):
 - a) S-shape for soft single-layer tissue;
 - b) C-shape for hard-soft double-layer tissue.

The allocations of each of the training-testing data split configurations for the S-shape and two-layer C-shape were assigned randomly. By splitting the dataset this way, we remove any bias for the model to learn prediction with experience from all of the datasets, thus more rigorously testing

our methods ability to generalize to unforeseen insertion scenarios.

A. Data Preprocessing

Several data preprocessing steps are taken in order to improve learning results. Deep networks can experience exploding gradients with large parameter values; therefore, for label κ_c and ω_{init} values, the values are scaled between 0 and 1, and -1 and 1 , respectively. S-shaped and two-layer indicators are treated as a logical 0 and 1 floating-point value. The insertion depths are scaled by the needle's total length. To improve learning sensitivity to the κ_c and ω_{init} , z-score normalization is used on the input κ_c and ω_{init} values.

B. Deep Network Training

In training deep networks, we explore various configurations of network structures by modifying the hidden layer sizes of these networks. The 143 different network configurations were tested, including networks from two hidden layers to five hidden layers with varying layer sizes. Upon the beginning of this study, we started training networks with SL with small, two hidden layer networks with layer sizes of between two and ten neurons. Then, we expanded upon hidden network depth, increasing the number of layers of each of these networks to three, four, and five hidden layers. Networks with more than five hidden layers were found to be unstable to train on our data and are, thus, not further explored. Once we evaluated small-layered networks, we expanded hidden layer sizes up to the maximum layer size of 100 neurons. We also tested various network relative sizing structures where we some networks had more neurons in the innermost hidden layer than the input and output layers (e.g., 10-20-30-15-8), as well as the inverse of this (e.g., 10-5-4-5-6). The intuition with evaluating networks with larger innermost layers was to project the small input data size to a higher dimensional space for increased generality and for the inverse, to summarize our input data with fewer parameters from the smaller innermost layers. This parameter sweep procedure was performed heuristically through trial-and-error following general deep network parameter search; as we found networks were underfitting our experimental data, we would add layers and increase layer sizes, and if networks overfit our experimental data, we decreased the number of layers and the size of the hidden layers. We found that using SL only, the smaller networks with two to three hidden layers and smaller layer sizes between four and ten neurons outperformed the larger networks as we saw decaying training performance due to model overfit as we added parameters to our network structure. Furthermore, larger innermost layered networks did not train well on SL, since these models had too many parameters to fit on our limited experimental dataset. When we incorporated SSL and TL, we found that larger network performance drastically improved and became more consistent over various large networks inversely to smaller network performance, likely due to the increased dataset dimensionality. Smaller networks with fewer layers and smaller layer sizes did not generalize well to a larger dataset to fit, as they lacked sufficient number of parameters for learning.

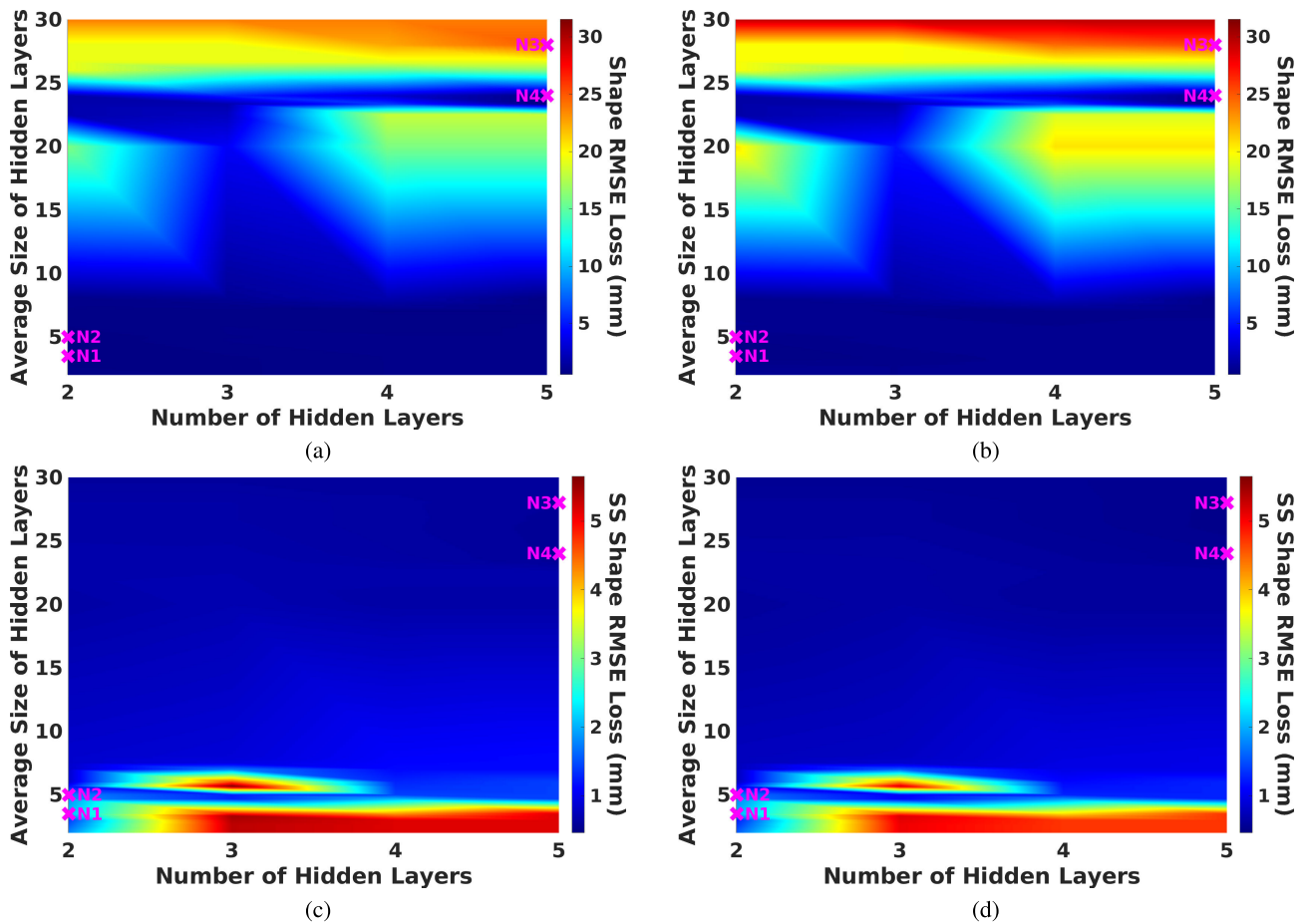


Fig. 3. Result of network configuration hyperparameter search for SL and SSL, comparing the number of hidden layers and the average size of the network's hidden layers to the training loss. The pink stars indicate network configurations presented in this work, N1–N4. (a) and (b) Networks trained with SL on the shape MSE loss function representing train and test loss, respectively, from (11). (c) and (d) Networks trained with SSL on the SS shape MSE loss function representing train and test loss, respectively, from (12).

Fig. 3 demonstrates this relationship illustrating training loss of the network performance over a summary of our deep network configuration search space. The average size of hidden layers summarizes the size of a hidden network, where larger networks will have larger average hidden layer sizes and smaller networks will have smaller average hidden layer sizes. Fig. 3 provides a summarization of the relationship we found while searching for an optimal network configuration between the size of a network and the depth of the network to the training and validation loss. The training and test loss plots were interpolated to provide information between evaluated network configurations. After training with supervised and SSL, the top-3 best-performing network configurations from supervised and SSL are then evaluated on TL. Thus, we concluded that we present the best-performing networks with the smallest and largest number of layers based on the best-performing training error to characterize the advantages for each of the supervised, SSL, and TL methods at the extremes of the tested network configurations. Through this exploration, we searched for the optimal configuration of these networks as well as studied the learning capabilities of small networks as compared to large networks with our aforementioned deep network techniques in Section II-D. Neural network configurations are denoted by a sequence

of hidden layer network sizes (e.g., 20-30-40 denotes a three hidden layer neural network with the layer sizes of 20, 30, and 40 neurons for the first, second, and third layers, respectively).

1) *Supervised Learning*: For training neural networks with SL, we use the Adam optimizer for optimizing the model's parameters, since it is robust to noisy and sparse gradients [69]. Models are trained for 3000 epochs and an initial learning rate of 0.05. Learning rate decay are used for decaying the learning rate by 10% every 1000 epoch and decay on test loss plateau with a 50% decay factor and a patience of updating for 50 epochs.

2) *Self-Supervised Learning*: SSL training is similar as referenced in Section III-B1, with a few modifications. First, we perform training for 10 000 epochs on batch sizes of 3000 per model. The learning rate is initialized to 0.005, and scheduled learning rate decay is used to decay the learning rate by 80% every 200 epochs. Learning rate decay upon plateau is not used, since dataset regeneration regularizes deep network learning.

One of the key aspects to using SSL is the requirement for data generation, since this method does not use experimental data for training. We uniformly sample the needle shape parameters, such as L_{ref} , κ_C , and ω_{init} , over their specified domains. L_{ref} is sampled in the ranges of 10% and 20 mm less

of the needle's total length. κ_c parameters are uniformly sampled over the range of 0–0.01. ω_{init} parameters are uniformly sampled over a 3-D sphere of radius 0.01. For ensuring that we train on all needle shape configurations, we generate needle prediction configurations with ratios of 2:1:1 for C-shaped single-layer, C-shaped double-layer, and S-shaped single-layer insertion configurations. We sample double-layer and S-shaped parameters uniformly from a distribution of 25%–75% of the needle's total length. We test multiple L_{pred} values over a set of insertion depths consisting of 10%, 25%, 50%, 75%, and 100% of the entire needle length. Every 100 epochs, we regenerate the sample dataset with the above configurations.

3) *Transfer Learning*: TL combines the SSL and SL methods together to train the deep networks we study in this article. After a model is trained with the SSL approach as mentioned in Section III-B2, the self-supervised trained model is then exposed to experimental data by further training via the SL approach demonstrated in Section III-B1. To ensure the self-supervised trained weights are not forgotten, we reduce the initialized learning rate to 0.0001 when training further with SL.

IV. RESULTS

Several metrics are used to compare the predicted needle shape, \mathbf{r} , to the ground-truth needle shape, \mathbf{r}_{gt} , discretized by their arclength s_i , $i = 1, \dots, N$. These metrics are listed below.

A. Root-Mean-Square Error

The overall root-mean-square error (RMSE) of the needle shape

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\mathbf{r}_{gt}(s_i) - \mathbf{r}(s_i)\|^2}. \quad (13)$$

B. In-Plane Error

The error measured in the natural bending plane of the needle

$$\text{IPE} = \frac{1}{N} \sum_{i=1}^N \|(0 \ 1 \ 1) \cdot (\mathbf{r}_{gt}(s_i) - \mathbf{r}(s_i))\|. \quad (14)$$

C. Out-of-Plane Error

The error measured in the plane orthogonal to the natural bending plane of the needle

$$\text{OPE} = \frac{1}{N} \sum_{i=1}^N \|(1 \ 0 \ 1) \cdot (\mathbf{r}_{gt}(s_i) - \mathbf{r}(s_i))\|. \quad (15)$$

D. Max Error (MAX)

The maximum error measured along the needle

$$\text{MAX} = \max_i (\|\mathbf{r}_{gt}(s_i) - \mathbf{r}(s_i)\|). \quad (16)$$

Note that \cdot above denotes matrix multiplication.

In Table I, we demonstrate the shape prediction results of various training methods and neural network configurations.

Notably, we observe that LR overfits the training data with a much larger test error than training error and does not perform well on the training dataset. The next improvement we notice is using SL on fully connected neural networks. SL on our limited size dataset trains small neural networks reasonably well, but these models observe to suffer from higher variance in training results, as models became much larger. The best-performing SL model on the test error is the smallest configuration with two hidden layers of sizes 5 and 2 neurons. Using SSL, larger models outperform their smaller counterparts on test error. While SSL alone is able to produce reasonable prediction errors, this method does not predict as well as the SL case. Finally, TL produces the best training shape prediction results using larger models and good test errors, as well. The smaller models notably does worse with TL than when using only SSL and SL.

Fig. 4 demonstrates the distribution error statistics for the best-performing models from SL and TL training methods. Both methods produce shape prediction RMSE under 3 mm over all insertion depths for all insertion scenarios with smaller variance in training errors seen using the TL method, illustrated in Fig. 4(a) and (b). The most difficult prediction for both methods appears to be from the S-shaped and two-layer insertion scenarios, as expected, since these situations have more factors affecting the needle shapes. Furthermore, seen in Figs. 4(b) and (d), both methods produce relatively consistent results independent of the prediction insertion step size (i.e., predicting the needle shape at 80 mm from 30 mm would be a 50-mm insertion step) while observing larger variances, as the insertion step size increases. The maximum evaluated prediction insertion step is a 110-mm insertion step. The transfer learned model keeps shape prediction maximum error for the train and test errors under 5 mm over most predictions with a single outlier in the C-shaped hard tissue insertion. The model trained using SL maximum error on the training dataset reaches 6 mm and experiences much larger variance than its counterpart trained with TL, but performs more consistently and accurately on the test dataset with maximum errors all under 4 mm. Since the transfer learned model has more consistent performance over the training dataset and relatively similar test error, we deem the 20-30-40-30-20 model as our model of choice for the rest of this discussion.

A 3-D ground-truth comparison with the needle shape predictions from our best-performing model demonstrates good correspondence in Fig. 5 for all insertion scenarios (C-shaped one-layer, C-shaped two-layer, and S-shaped one-layer). Predictions illustrated in Fig. 5 are example predictions of a 50-mm insertion step, predicting from a 75-mm insertion depth to the full length insertion depth of 125 mm. The strong correspondence between the ground-truth shapes and the predicted shapes over all of the insertion scenarios demonstrates the generalizability of the hybrid deep learning and model-based approach to needle shape prediction.

V. DISCUSSION

The shape prediction models presented here produced a spread of prediction errors, as reported in Table I. The linear model trained using least squares overfits the training data with

TABLE I

MEAN ERRORS OVER THE TRAINING, TEST, AND ENTIRE DATASET SPLITS FOR THE PREDICTED SHAPE RMSE WITH THE GROUND TRUTH FOR VARIOUS DEEP NETWORK MODELS AND TRAINING METHODS. THE BEST-PERFORMING MODELS IN EACH COLUMN ARE **BOLDED**

| Training Method | Hidden Layer Sizes | Mean Train RMSE (mm) | Mean Test RMSE (mm) | Mean Dataset RMSE (mm) |
|-----------------|---------------------|----------------------|---------------------|------------------------|
| LR | N/A | 9.27 | 23.49 | 13.85 |
| SL | 5-2 (N1) | 0.97 | 0.94 | 0.96 |
| | 5-5 (N2) | 1.48 | 1.54 | 1.50 |
| | 20-30-40-20-10 (N3) | 0.96 | 1.10 | 1.00 |
| | 20-30-40-30-20 (N4) | 23.80 | 23.21 | 23.63 |
| SSL | 5-2 (N1) | 2.20 | 1.94 | 2.13 |
| | 5-5 (N2) | 1.72 | 2.01 | 1.80 |
| | 20-30-40-20-10 (N3) | 1.84 | 2.10 | 1.87 |
| | 20-30-40-30-20 (N4) | 1.73 | 1.59 | 1.69 |
| TL (SSL+SL) | 5-2 (N1) | 2.80 | 2.62 | 2.75 |
| | 5-5 (N2) | 2.49 | 2.54 | 2.50 |
| | 20-30-40-20-10 (N3) | 0.86 | 1.17 | 1.06 |
| | 20-30-40-30-20 (N4) | 0.85 | 1.03 | 0.90 |

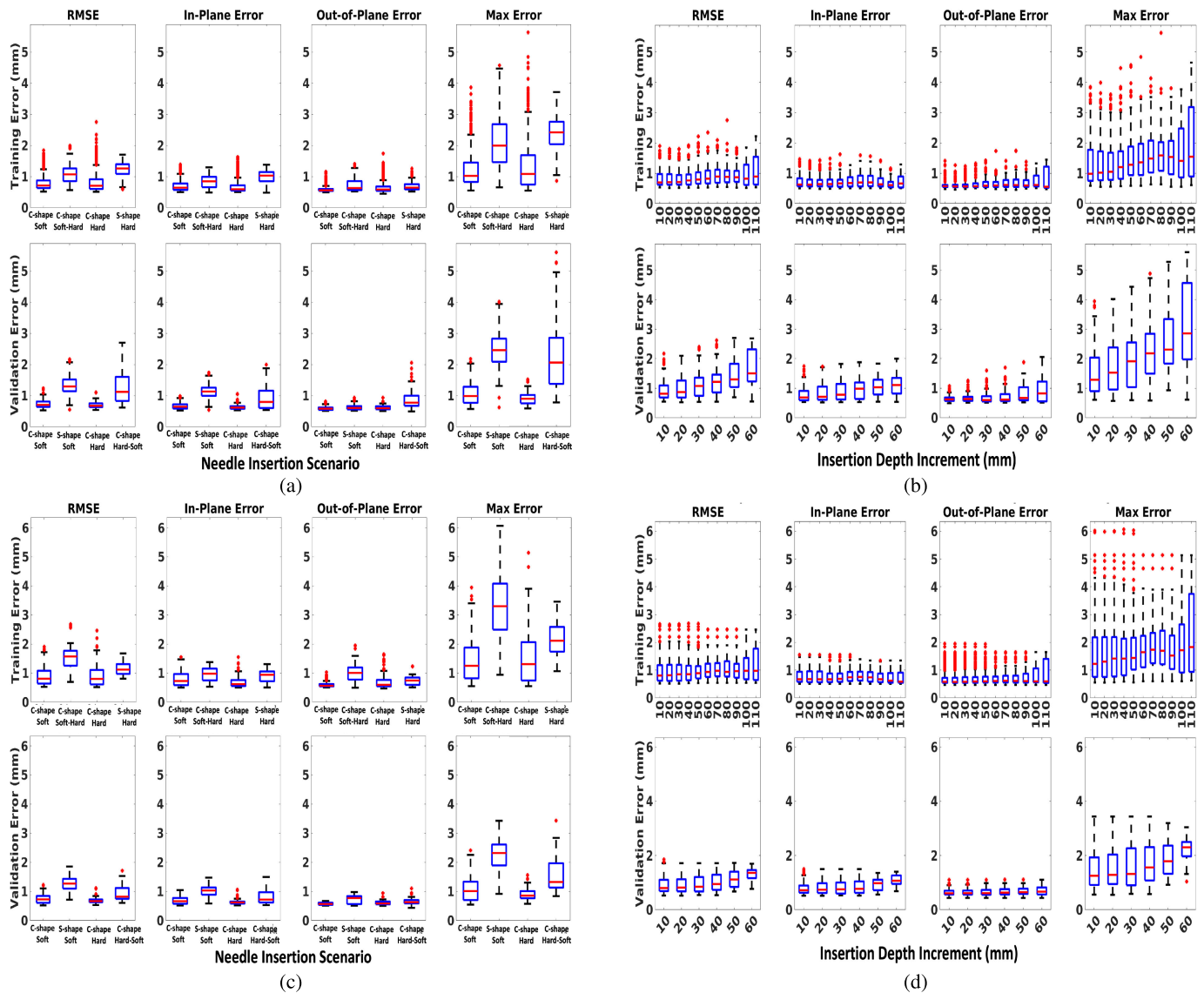


Fig. 4. Shape-prediction error statistics over each different insertion scenarios for (a) TL and (c) SL, as well as shape-prediction error statistics over various insertion prediction ranges for (b) TL and (d) SL. Note that since the training and test datasets are different, the x-axis labels for (a) and (c) do not directly correspond.

an average RMSE of 9.27 mm and an average test RMSE of 23.49 mm, larger than 100% increase in error for the unobserved data. Training fully connected neural networks

using SL improves these results dramatically, mostly for smaller networks. The best-performing model trained with SL is the smallest reported model with hidden layer configuration

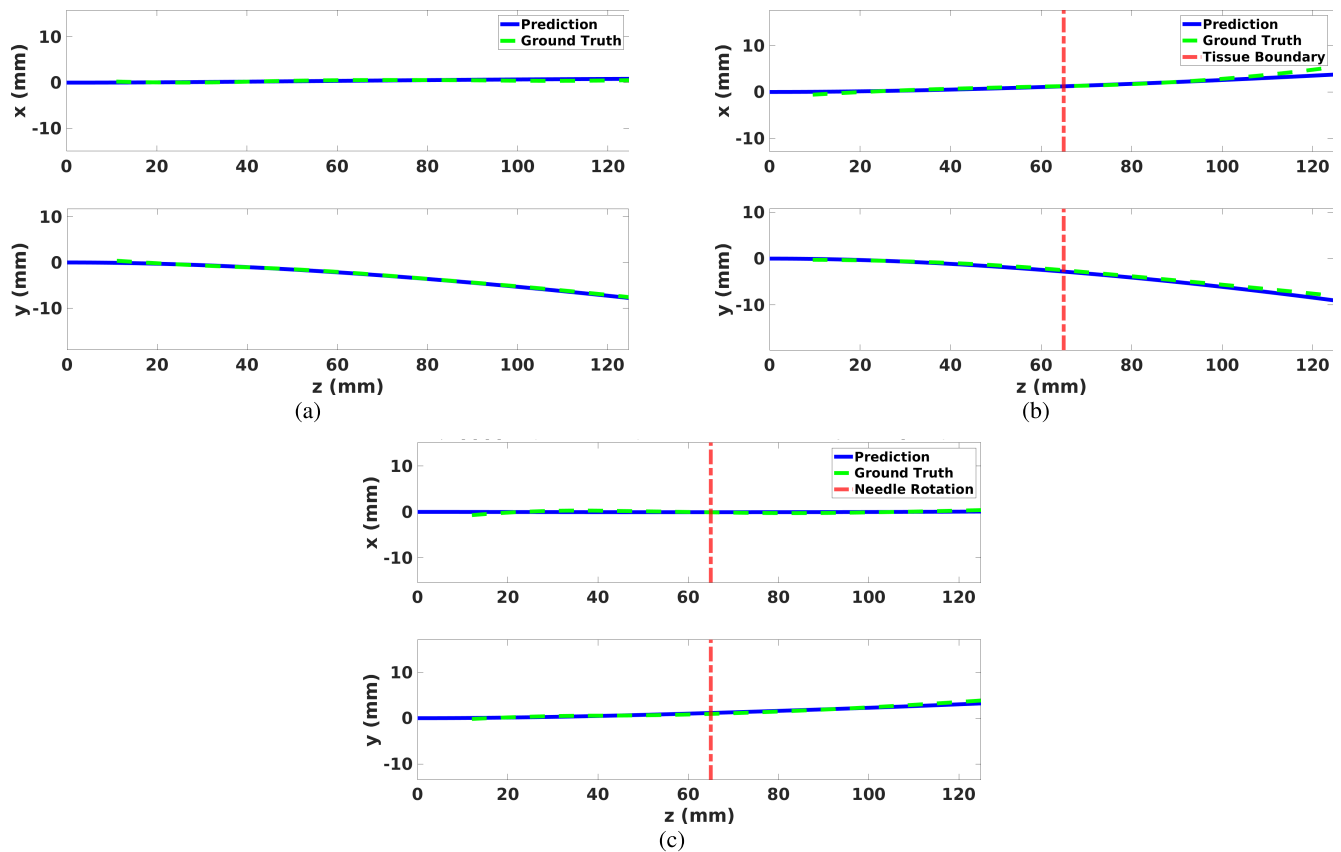


Fig. 5. Shape prediction results predicting the needle shape at 125 mm from 75-mm insertion depth using the best transfer learned model for (a) C-shaped one-layer shape, (b) C-shaped two-layer shape, and (c) S-shaped one-layer shape. The ground truth is shown in the green dashed line and the prediction in the blue line. The tissue boundary and the point of needle rotation locations are, respectively, denoted for the two-layer and S-shaped insertion scenarios with a red dashed line.

of neurons, 5-2. The deeper and larger network, 20-30-40-20-10, has the best train error, but does not perform as well on the test set as the 5-2 configuration. Furthermore, given the limited size of the training dataset, it appears that the training of larger networks with SL only is not stable, since the 20-30-40-30-20 configuration produced the largest train error of 23.80 mm. SSL is used to provide a baseline for these models to better understand preliminary features of the needle shape prediction problem with the assumption that the needle will follow its previous shapes. However, with only this assumption, average train and test errors of 1.73 and 1.59 mm, respectively, are achieved with the largest reported model with a configuration of 20-30-40-30-20. The smaller models do not perform well using SSL alone, due to underfitting the extensive dataset size of the simulated data, lending favorability to the larger models with more parameters to tune. Combining self-supervised and supervised with TL produces the best average training error of 0.85 mm and second-best test error for 1.03 mm for the largest model with the 20-30-40-30-20 configuration. Here, we observe how the improvement of providing an informed model initialization with SSL better enables the larger models to learn from the experimental data as seen from the drastic improvement of a 96% reduction in training error observed in 20-30-40-30-20 neural network configuration. This improvement is further justified, since the SSL loss function provides only an approximation of

predicted needle shapes, requiring further refinement with experimentally obtained data. Furthermore, the smaller models perform much worse with TL due to their lacking ability to generalize well with fewer parameters. Having exposure to a large dataset, these smaller models become biased and do not train well due to underfitting. While SL with small networks performs the best on the experimental test dataset, we believe the TL method with larger networks to outperform generally over more complicated needle insertion scenarios, as it is seen that the smaller networks underfit larger datasets.

Kinematics-based and mechanics-based bevel-tip needle shape prediction results are presented in [37]. Insertions into gelatin phantom were performed at an insertion depth of 115 mm for three trials, including an out-of-plane deflection from a 90° rotation of the needle during insertion. Reported errors for these results are presented in 2-D errors of the IPE and OPE. For the kinematics-based model, average maximum errors are reported as 3.77 mm for IPE and 2.20 for OPE. For the mechanics-based model, average maximum errors are reported as 2.20 mm for IPE and 1.72 for OPE. Our shape-prediction results for IPE and OPE perform just as good if not better than the average reported errors of the kinematics-based and mechanics-based models, as seen in Fig. 4(b). The mechanics-based model falls within our error distributions presented and, however, is specific to the needle and tissue mechanical parameters, while our model

is needle- and tissue-agnostic. The mechanics-based model required preinsertions into the gelatin phantom to calibrate the mechanical parameters of the model; however, preinsertions are not always clinically feasible. The kinematics-based model requires precise measurement of the bevel cut angle, which is clinically feasible; however, our method is needle-agnostic, since it only utilizes previous insertion information to predict the future needle shape at variable insertion depths. Our model has the capability to generalize to various needle shape prediction scenarios, since it uses experimental data from insertions from a single needle in conjunction with the presented SSL technique. Thus, our model provides a greater range of flexibility with at least as good results as the kinematics-based and mechanics-based models.

The only end-to-end, pure data-driven approach to needle insertion trajectory prediction found at the time of this work is presented in [57]. This data-driven approach uses a 3-D UNet and actor-critic deep reinforcement learning to generate 3-D medical image masks of the predicted needle shape from cryoablation procedures performed under interventional CT. Evaluation on needle-tip positioning was provided through a 2-D tip error from the forecast needle tip to the ground truth. The best-performing model achieved an average 2-D tip error of 24 mm, with a minimum error of 3 mm and a maximum of 44 mm. A qualitative analysis deduced that the 3-D UNet's predicted segmentation mask was frequently distorted or not feasible and that there was not enough data to properly train their network. In comparison with our hybrid approach, our model significantly outperformed the end-to-end learning-based approach, most likely due to the lack of model-based regularization provided in a pure learning-based method.

VI. CONCLUSION

In this article, we present and validate, for the first time, a hybrid deep learning and model-based method for flexible needle shape prediction in conjunction with a novel SSL approach to train deep networks to handle the needle insertion prediction problem in the absence of data. We experimentally validate our method in phantom gelatin insertions of an FBG-sensorized needle for needle insertion scenarios of single-layer homogeneous tissue of C- and S-shaped insertions, as well as double-layer homogeneous tissue for C-shaped insertions, resulting in a dataset size of almost 3000 samples. Insertion predictions are performed at various insertion depths up to the maximum insertion depth of 125 mm with predictions ranging as far as 110 mm ahead. Average reported errors for the best-performing model achieved are 0.85 mm for the training dataset and 1.03 mm for the test dataset. We also discuss in depth the benefits of various data-driven methods and neural network configurations with respect to their performance on our experimental dataset. Limitations of this work include lack of experiments in real tissue; validation with other needles, despite the model's flexibility to generalize to other models through the Lie-group theoretic shape model; and inefficient use of sequential data. Future work includes experimental validation of this approach in real tissue, experimental validation with multicore fiber FBG

sensors, neural network redesign to handle more than variable layered tissue and other shape configurations, and extension of this model to recurrent neural network or transformer network structures.

REFERENCES

- [1] *Key Statistics for Prostate Cancer*, Amer. Cancer Soc., Atlanta, GA, USA, 2023.
- [2] C. H. Pernar, E. M. Ebot, K. M. Wilson, and L. A. Mucci, "The epidemiology of prostate cancer," *Cold Spring Harbor Perspect. Med.*, vol. 8, no. 12, 2018, Art. no. a030361.
- [3] S. Okazawa, R. Ebrahimi, J. Chuang, S. E. Salcudean, and R. Rohling, "Hand-held steerable needle device," *IEEE/ASME Trans. Mechatronics*, vol. 10, no. 3, pp. 285–296, Jun. 2005.
- [4] B. R. Matlaga, L. A. Eskew, and D. L. McCullough, "Prostate biopsy: Indications and technique," *J. Urol.*, vol. 169, pp. 12–19, Jan. 2003.
- [5] S. Loeb, H. B. Carter, S. I. Berndt, W. Ricker, and E. M. Schaeffer, "Complications after prostate biopsy: Data from SEER-medicare," *J. Urol.*, vol. 186, no. 5, pp. 1830–1834, Nov. 2011, doi: [10.1016/j.juro.2011.06.057](https://doi.org/10.1016/j.juro.2011.06.057).
- [6] D. Glozman and M. Shoham, "Image-guided robotic flexible needle steering," *IEEE Trans. Robot.*, vol. 23, no. 3, pp. 459–467, Jun. 2007.
- [7] T. K. Adebare, A. E. Fletcher, and A. M. Okamura, "3-D ultrasound-guided robotic needle steering in biological tissue," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2899–2910, Dec. 2014.
- [8] N. Shahriari, R. J. Roesthuis, N. J. V. D. Berg, J. J. V. D. Dobbela, and S. Misra, "Steering an actuated-tip needle in biological tissue: Fusing FBG-sensor data and ultrasound images," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Jun. 2016, pp. 4443–4449.
- [9] S. Chen, F. Wang, Y. Lin, Q. Shi, and Y. Wang, "Ultrasound-guided needle insertion robotic system for percutaneous puncture," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, pp. 475–484, Mar. 2021.
- [10] S. J. Mathews et al., "Ultrasonic needle tracking with dynamic electronic focusing," *Ultrasound Med. Biol.*, vol. 48, no. 3, pp. 520–529, Mar. 2022.
- [11] M. Rabiei and B. Konh, "A portable robot to perform prostate brachytherapy with active needle steering and robot-assisted ultrasound tracking," in *Proc. Design Med. Devices Conf.*, Apr. 2022, Paper V001T11A005.
- [12] M. Waite, C. Rossa, R. Sloboda, N. Usmani, and M. Tavakoli, "3D shape visualization of curved needles in tissue from 2D ultrasound images using RANSAC," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 4723–4728.
- [13] H. R. S. Neshat and R. V. Patel, "Real-time parametric curved needle segmentation in 3D ultrasound images," in *Proc. 2nd IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechanics*, Oct. 2008, pp. 670–675.
- [14] B. Konh, B. Padasdao, Z. Batsaikhan, and S. Y. Ko, "Integrating robot-assisted ultrasound tracking and 3D needle shape prediction for real-time tracking of the needle tip in needle steering procedures," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 17, no. 4, p. e2272, Aug. 2021.
- [15] Y. Zhao, C. Cachard, and H. Liebgott, "Automatic needle detection and tracking in 3D ultrasound using an ROI-based RANSAC and Kalman method," *Ultrason. Imag.*, vol. 35, no. 4, pp. 283–306, Oct. 2013.
- [16] M. Waite, C. Rossa, R. Sloboda, N. Usmani, and M. Tavakoli, "Needle tracking and deflection prediction for robot-assisted needle insertion using 2D ultrasound images," *J. Med. Robot. Res.*, vol. 1, no. 1, Mar. 2016, Art. no. 1640001.
- [17] H. Su, M. Zervas, G. A. Cole, C. Furlong, and G. S. Fischer, "Real-time MRI-guided needle placement robot with integrated fiber optic force sensing," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, May 2011, pp. 1583–1588.
- [18] X. Li et al., "Automatic needle tracking using mask R-CNN for MRI-guided percutaneous interventions," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 10, pp. 1673–1684, Oct. 2020, doi: [10.1007/s11548-020-02226-8](https://doi.org/10.1007/s11548-020-02226-8).
- [19] M. Al-Maatoq, M. Facht, R. Rao, and C. Hoeschen, "Artifacts' detection for MRI non-metallic needles: Comparative analysis for artifact evaluation using K-means and manual quantification," *Magnetochemistry*, vol. 9, no. 3, p. 79, Mar. 2023.
- [20] G. Medan and L. Joskowicz, "Flexible needle and patient tracking using fractional scanning in interventional CT procedures," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 6, pp. 1039–1047, Jun. 2019.
- [21] J. Reisenauer et al., "Ion: Technology and techniques for shape-sensing robotic-assisted bronchoscopy," *Ann. Thoracic Surg.*, vol. 113, no. 1, pp. 308–315, Jan. 2022.

- [22] Y. Wei and S. Z. Zhou, "Multi-camera AR navigation system for CT-guided needle insertion task," in *Proc. IEEE Conf. Virtual Reality 3D User Interface Abstr. Workshops (VRW)*, Mar. 2023, pp. 919–920.
- [23] R. J. Roesthuis, N. J. Van De Berg, J. J. Van Den Dobbelen, and S. Misra, "Modeling and steering of a novel actuated-tip needle through a soft-tissue simulant using Fiber Bragg Grating sensors," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Mar. 2015, pp. 2283–2289.
- [24] L. Zhang et al., "A new method for fiber Bragg grating based needle shape sensing calibration," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2019, pp. 1953–1958.
- [25] M. Li, G. Li, B. Gonenc, X. Duan, and I. Iordachita, "Towards human-controlled, real-time shape sensing based flexible needle steering for MRI-guided percutaneous therapies," *Int. J. Med. Robot. Comput. Assist. Surgery*, vol. 13, no. 2, pp. 211–215, Jun. 2017.
- [26] N. J. van de Berg, J. Dankelman, and J. J. van den Dobbelen, "Design of an actively controlled steerable needle with tendon actuation and FBG-based shape sensing," *Med. Eng. Phys.*, vol. 37, no. 6, pp. 617–622, 2015, doi: [10.1016/j.medengphys.2015.03.016](https://doi.org/10.1016/j.medengphys.2015.03.016).
- [27] A. Donder and F. R. Y. Baena, "3-D path-following control for steerable needles with fiber Bragg gratings in multi-core fibers," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 3, pp. 1072–1085, Mar. 2023.
- [28] L. Zhang et al., "Fiber Bragg grating-based sensor system for sensing the shape of flexible needles," *Measurement*, vol. 206, Jan. 2023, Art. no. 112251.
- [29] D. A. Lezcano, I. I. Iordachita, and J. S. Kim, "Lie-group theoretic approach to shape-sensing using FBG-sensorized needles including double-layer tissue and S-shape insertions," *IEEE Sensors J.*, vol. 22, no. 22, pp. 22232–22243, Nov. 2022.
- [30] D. A. Lezcano, Y. Zhetpissov, A. Cheng, J. Seob Kim, and I. I. Iordachita, "Optical fiber-based needle shape sensing in real tissue: Single core vs. multicore approaches," 2023, *arXiv:2309.04407*.
- [31] W. Park, J. S. Kim, Y. Zhou, N. J. Cowan, A. M. Okamura, and G. S. Chirikjian, "Diffusion-based motion planning for a nonholonomic flexible needle model," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Jul. 2005, pp. 4600–4605.
- [32] R. J. Webster, III, J. S. Kim, N. J. Cowan, G. S. Chirikjian, and A. M. Okamura, "Nonholonomic modeling of needle steering," *Int. J. Robot. Res.*, vol. 25, nos. 5–6, pp. 509–525, May 2006.
- [33] V. Duindam, R. Alterovitz, S. Sastry, and K. Goldberg, "Screw-based motion planning for bevel-tip flexible needles in 3D environments with obstacles," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2008, pp. 2483–2488.
- [34] N. J. Cowan et al., "Robotic needle steering: Design, modeling, planning, and image guidance," in *Surgical Robotics: Systems Applications and Visions*. New York, NY, USA: Springer, 2011, pp. 557–582.
- [35] K. Henken, D. Van Gerwen, J. Dankelman, and J. Van Den Dobbelen, "Accuracy of needle position measurements using fiber Bragg gratings," *Minimally Invasive Therapy Allied Technol.*, vol. 21, no. 6, pp. 408–414, Nov. 2012.
- [36] R. Secoli and F. Rodriguez y Baena, "Closed-loop 3D motion modeling and control of a steerable needle for soft tissue surgery," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 5831–5836.
- [37] R. J. Roesthuis, M. Kemp, J. J. van den Dobbelen, and S. Misra, "Three-dimensional needle shape reconstruction using an array of fiber Bragg grating sensors," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 4, pp. 1115–1126, Aug. 2014.
- [38] J. Carriere, C. Rossa, R. Sloboda, N. Usmani, and M. Tavakoli, "Real-time needle shape prediction in soft-tissue based on image segmentation and particle filtering," in *Proc. IEEE Int. Conf. Adv. Intell. Mechatronics (AIM)*, Jul. 2016, pp. 1204–1209.
- [39] D. Glozman and M. Shoham, "Flexible needle steering and optimal trajectory planning for percutaneous therapies," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science, vol. 3217, 2004, pp. 137–144.
- [40] M. Khadem, B. Fallahi, C. Rossa, R. S. Sloboda, N. Usmani, and M. Tavakoli, "A mechanics-based model for simulation and control of flexible needle insertion in soft tissue," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 2264–2269.
- [41] M. Khadem, C. Rossa, R. S. Sloboda, N. Usmani, and M. Tavakoli, "Mechanics of tissue cutting during needle insertion in biological tissue," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 800–807, Jul. 2016.
- [42] S. Jiang and X. Wang, "Mechanics-based interactive modeling for medical flexible needle insertion in consideration of nonlinear factors," *J. Comput. Nonlinear Dyn.*, vol. 11, no. 1, Jan. 2016, Art. no. 011004.
- [43] B. Zhao et al., "Needle deflection modeling and preoperative trajectory planning during insertion into multilayered tissues," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 2, pp. 943–954, Apr. 2021.
- [44] A. Cheng, D. A. Lezcano, J. S. Kim, and I. I. Iordachita, "Optical fiber-based needle shape sensing: Three-channel single core vs. multicore approaches," in *Proc. Int. Symp. Med. Robot. (ISMR)*, Apr. 2023, pp. 1–7.
- [45] D. A. Lezcano, I. I. Iordachita, and J. S. Kim, "Trajectory generation of FBG-sensorized needles for insertions into multi-layer tissue," in *Proc. IEEE Sensors*, Jun. 2020, pp. 1–4.
- [46] C. Avila-Carrasco, M. Ruppel, R. Persad, A. Bahl, and S. Dogramadzi, "Analytical versus data-driven approach of modelling Brachytherapy needle deflection," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 4, pp. 519–528, Nov. 2020.
- [47] C. Rossa, T. Lehmann, R. Sloboda, N. Usmani, and M. Tavakoli, "A data-driven soft sensor for needle deflection in heterogeneous tissue using just-in-time modelling," *Med. Bio. Eng. Comput.*, vol. 55, no. 8, pp. 1401–1414, 2017.
- [48] W. Edwards et al., "Data-driven modelling and control for robot needle insertion in deep anterior lamellar keratoplasty," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 1526–1533, Apr. 2022.
- [49] F. Han, Y. He, H. Zhu, and K. Zhou, "A novel catheter shape-sensing method based on deep learning with a multi-core optical fiber," *Sensors*, vol. 23, no. 16, p. 7243, Aug. 2023.
- [50] N. Gessert et al., "Spatio-temporal deep learning models for tip force estimation during needle insertion," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 14, no. 9, pp. 1485–1493, Sep. 2019, doi: [10.1007/s11548-019-02006-z](https://doi.org/10.1007/s11548-019-02006-z).
- [51] S. Sefati, R. Hegeman, F. Alambeigi, I. Iordachita, and M. Armand, "FBG-based position estimation of highly deformable continuum manipulators: Model-dependent vs. data-driven approaches," in *Proc. Int. Symp. Med. Robot. (ISMR)*, Apr. 2019, pp. 1–6.
- [52] S. Sefati, S. Sefati, I. Iordachita, R. H. Taylor, and M. Armand, "Learning to detect collisions for continuum manipulators without a prior model," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11768, 2019, pp. 182–190.
- [53] X. Tan, Y. Lee, C.-B. Chng, K.-B. Lim, and C.-K. Chui, "Robot-assisted flexible needle insertion using universal distributional deep reinforcement learning," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 2, pp. 341–349, Feb. 2020.
- [54] X. T. Ha et al., "Shape sensing of flexible robots based on deep learning," *IEEE Trans. Robot.*, vol. 39, no. 2, pp. 1580–1593, Apr. 2023.
- [55] A. I. Chen, M. L. Balter, T. J. Maguire, and M. L. Yarmush, "Deep learning robotic guidance for autonomous vascular access," *Nat. Mach. Intell.*, vol. 2, pp. 104–115, Feb. 2020, doi: [10.1038/s42256-020-0148-7](https://doi.org/10.1038/s42256-020-0148-7).
- [56] S. Manavi Roodsari et al., "Shape sensing of optical fiber Bragg gratings based on deep learning," *Mach. Learn., Sci. Technol.*, vol. 4, no. 2, Jun. 2023, Art. no. 025037, doi: [10.1088/2632-2153/acda10](https://doi.org/10.1088/2632-2153/acda10).
- [57] M. A. Pincock, Y. Hu, S. Bandula, and D. C. Barratt, "End-to-end forecasting of needle trajectory in percutaneous ablation," *Proc. SPIE*, vol. 11598, pp. 585–592, Jan. 2021.
- [58] M. C. Bernardes et al., "Data-driven adaptive needle insertion assist for transperineal prostate interventions," *Phys. Med. Biol.*, vol. 68, no. 10, May 2023, Art. no. 105016.
- [59] J. S. Kim, M. Chatrasingh, S. Kim, J. Suthakorn, and I. I. Iordachita, "Fiber Bragg grating based needle shape sensing for needle steering system: Evaluation in inhomogeneous tissue," in *Proc. IEEE Sensors*, Oct. 2017, pp. 1–3.
- [60] J. S. Kim, J. Guo, M. Chatrasingh, S. Kim, and I. Iordachita, "Shape determination during needle insertion with curvature measurements," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 201–208.
- [61] G. S. Chirikjian and A. B. Kyatkin, *Harmonic Analysis for Engineers and Applied Scientists*. New York, NY, USA: Dover, 2016.
- [62] D. D. Holm, J. E. Marsden, and T. S. Ratiu, "The Euler–Poincaré equations and semidirect products with applications to continuum theories," *Adv. Math.*, vol. 137, no. 1, pp. 1–81, Jul. 1998.
- [63] J. S. Kim and G. S. Chirikjian, "Conformational analysis of stiff chiral polymers with end-constraints," *Mol. Simul.*, vol. 32, no. 14, pp. 1139–1154, Dec. 2006.
- [64] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. Hoboken, NJ, USA: Wiley, 2021.

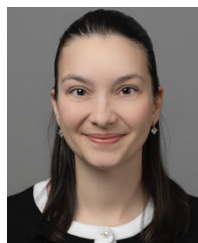
- [65] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2684–2691.
- [66] X. Zhai, A. Oliver, A. Kolesnikov, and L. Beyer, "S4L: Self-supervised semi-supervised learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1476–1485.
- [67] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.*, 2016.
- [68] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Hershey, PA, USA: IGI Global, 2010, pp. 242–264.
- [69] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.



Dimitri A. Lezcano received the M.S.E. degree in robotics from Johns Hopkins University, Baltimore, MD, USA, in 2019, and the B.A. degree in physics and mathematics from McDaniel College, Westminster, MD, in 2021. He is currently pursuing the Ph.D. degree in mechanical engineering from Johns Hopkins University.

His research interests are in computer-integrated surgical technologies, including flexible needle shape sensing, needle trajectory generation, and flexible needle steering.

Yernar Zhetpissov, photograph and biography not available at the time of publication.



Mariana C. Bernardes (Member, IEEE) received the B.S. degree in mechatronics engineering, the M.S. degree in electrical engineering from the University of Brasilia, Brasilia, Brazil, in 2006 and 2009, respectively, and the Ph.D. degree in automation and microelectronic systems from the University of Montpellier, Montpellier, France, in 2012.

She was an Assistant Professor at the University of Brasilia for seven years before joining the Brigham and Women's Hospital research team in 2022. She is a Research Fellow in Radiology at Brigham, Women's Hospital, and Harvard Medical School. She specializes in medical robotics, focusing on robot-assisted minimally invasive procedures, specifically in laparoscopy and image-guided percutaneous interventions.



Pedro Moreira (Member, IEEE) received the master's degree in electrical engineering from the Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, and the Ph.D. degree in automation and microelectronic systems from the University of Montpellier, Montpellier, France, in 2012.

He is an Instructor of Radiology with the Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. His research's primary focus has been the development of robotic systems for medical applications, particularly in robot-assisted percutaneous procedures and magnetic resonance imaging (MRI)-guided procedures.



Junichi Tokuda (Member, IEEE) received the B.S. degree in engineering, the M.S. degree in information science and technology, and the Ph.D. degree in information science and technology from the University of Tokyo, Tokyo, Japan, in 2002, 2004, and 2007, respectively.

He joined Brigham and Women's Hospital (BWH), Harvard Medical School (HMS), Boston, MA, USA, in 2007, and has been serving as an Associate Professor of Radiology at HMS since 2020. He is a Computer Scientist specializing in biomedical engineering at BWH, and a teaching affiliate of HMS. He is currently leading several NIH-funded medical robotics projects as a Principal Investigator (R01EB020667, R01CA235134, and R01EB034359). His research interests include computer-assisted interventions, particularly MRI-based monitoring and guidance of biopsies and ablations. He has extensive experience in developing, integrating, and clinically evaluating guidance software and devices to support clinical procedures.



Jin Seob Kim (Member, IEEE) received the M.S. degree in mechanical engineering from Seoul National University, Seoul, South Korea, and the M.S. and Ph.D. degrees in mechanical engineering from Johns Hopkins University, Baltimore, MD, USA, in 2000, 2004, and 2006, respectively.

He was an Assistant Research Professor of Mechanical Engineering and Robotics with Johns Hopkins University, where he did postdoctoral work at the Johns Hopkins Institute for NanoBioTechnology, Johns Hopkins Bloomberg School of Public Health and is currently a Senior Lecturer of Mechanical Engineering and Robotics. His research interests include medical robotics, in particular, sensor-based mathematical modeling of continuum robots, computational modeling of biological systems, and application of non-commutative harmonic analysis.



Iulian I. Iordachita (Senior Member, IEEE) received the M.Eng. degree in industrial robotics and the Ph.D. degree in mechanical engineering from the University of Craiova, Craiova, Romania, in 1989 and 1996, respectively.

In 2000, he was a Postdoctoral Fellow with the Brady Urological Institute, School of Medicine, Johns Hopkins University, Baltimore, MD, USA, and from 2002 to 2003, he was a Research Fellow with the Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. He is currently a Research Professor of Mechanical Engineering and Robotics with Johns Hopkins University. His research interests include medical robotics, image guided surgery with a specific focus on microsurgery, interventional magnetic resonance imaging (MRI), smart surgical tools, and medical instrumentation.