On Task-specific Autonomy in Robotic Interventions: A Multimodal Learning-based Approach for Multi-level Skill Assessment during Cyborg Catheterization

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY-NC-SA 4.0

SUBMISSION DATE / POSTED DATE

30-10-2021 / 06-11-2021

CITATION

Omisore, Olatunji; Duan, Wenke; Du, Wenjing; Han, Shipeng; Akinyemi, Toluwanimi; Wang, Lei (2021): On Task-specific Autonomy in Robotic Interventions: A Multimodal Learning-based Approach for Multi-level Skill Assessment during Cyborg Catheterization. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.16908403.v1>

DOI

[10.36227/techrxiv.16908403.v1](https://doi.org/10.36227/techrxiv.16908403.v1)

On Task-specific Autonomy in Robotic Interventions: A Multimodal Learning-based Approach for Multi-level Skill Assessment during Cyborg Catheterization

Olatunji Mumini Omisore, *Member, IEEE*, Wenke Duan, Wenjing Du, Yifa Li, Shipeng Han, Toluwanimi Akinyemi, *Graduate Student Member, IEEE*, Lei Wang, *Senior Member, IEEE*

Abstract—Recent insights from human-robot intelligence and deep learning raise hope towards task-specific autonomy in robotic intravascular interventions. This can gradually eliminate the operational hazards faced during intravascular interventions. However, lack of learning-based methods for characterizing the multimodal data generated during cyborg catheterization hinders the drive towards autonomous robotic control. Also, multiplexing salient features from multiple data-sources can enhance effective assessment and classification of domain skills for apt intelligent surgeon-robot (cyborg) catheterization during intravascular interventions. In this study, task-specific autonomous intervention is envisioned upon an isomorphic master-slave robotic catheter system that exhibit hand defter techniques used in Cath Labs. To drive cyborg catheterization, stacking-based deep neural network is developed for three-level skill assessment. Network training and evaluation were performed on multimodal temporal and spatial datasets recorded during in-vivo cyborg catheterization in pig and rabbits. Based on the dataset, the model achieved classification performance of $96.43 \pm 0.50\%$, $86.77 \pm 1.55\%$, and $95.50 \pm 0.49\%$ for the three-level skill assessments on the multimodal data from intra-animal (rabbit and pig separately) and inter-animal (mixture) procedures. Skill assessment results from multimodal datasets were compared with unimodal and bimodal datasets, and the model shows better performance with the multimodal features compared to the cases of unimodal ($87.72 \pm 1.06\%$, $69.81 \pm 3.73\%$ and $69.76 \pm 2.76\%$) and bimodal ($94.15 \pm 0.75\%$, $73.67 \pm 1.71\%$, and $80.78 \pm 2.51\%$) cases in the three study categories. Thus, effectiveness of the proposed learning model is demonstrated with capability of handling large multimodal cyborg catheterization dataset for specified procedures. This study is envisioned as a step towards task-specific autonomy in robotic intravascular interventions. Cases of autonomous intravascular catheterization and tool deployment can be adapted with the cyborg synergy demonstrated in this study.

Index Terms— multimodal learning, skill assessment, cyborg catheterization, autonomous navigation, robotic interventions

Manuscript received June 1, 2021; revised xxx xx, xxxx; accepted xxx xx, xxxx. Date of publication xxxxxx xxx, xxxx. This paper was recommended by Associate Editor X.X Xxxx. (Olatunji Mumini Omisore, Wenke Duan, and Wenjing Du contributed equally to this work; Correspondence: Lei Wang). This work was supported by National Key Research and Development program of China (#2019YFB1311700); National Natural Science Foundation of China (#U1713219 and #61950410618); Shenzhen Natural Science Foundation (#JCYJ20190812173205538), and CAS President's Postdoctoral Fellowship.

O. M. Omisore, W. Duan, W. Du, Y. Li, S. Han, T. Akinyemi, and L. Wang are with Research Centre for Medical Robotics and Minimally Invasive Surgical Devices, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. (e-mails: omisore@siat.ac.cn, wk.duan@siat.ac.cn, wj.du@siat.ac.cn, yf.li1@siat.ac.cn, sp.han@siat.ac.cn, tolu@siat.ac.cn, and wang.lei@siat.ac.cn). Authors are also with the CAS Key Laboratory for Health Informatics, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.

I. STUDY BACKGROUND

In addition to faster recovery and better cosmesis, intravascular interventions have shown good capabilities in reducing the trauma associated with the conventional approaches used for interventions of ischemic abdominal and heart diseases [1-3]. Purposive learning interventional intravascular surgery is presumed as conglomerate of advances in robotic systems and data science [4]. As emerging interdisciplinary fields, both areas of discourse are aimed to achieve sufficient data extraction, refined feature engineering and qualitative knowledge processing; which are imperative for surgical skill evaluation during intravascular interventions. Tool catheterization and deployment are vital parts of intravascular interventions. Catheterization involves navigation of endovascular tools viz. guidewire or catheters, from an insertion point such as the radial, femoral, or brachial ports; into an intervention site. While difficulties arise from tool hysteresis, lack of useful feedback, and beating-heart (in percutaneous coronary interventions); tool catheterization is a more challenging procedure due to vascular bifurcation, blood opaqueness, vessel fragility, and other dynamic factors. These show patients' safety could be at risk, and are deemed as reasons hindering the advancement of research on autonomous robotic interventions. Hence, interventionists are fully engaged in catheterization procedures meanwhile they are unable to implement other procedures with full mindset, simultaneously.

A. Motivation

Interventionists trained with skillful integration of visual, tactile, and intuitional stimuli achieve high success rates in performing intravascular catheterization. Safety concerns of interventionists and patients during percutaneous intravascular interventions motivated the development and commercialization of robotic catheter systems [5]. To ensure patients' safety, autonomous navigation can be adapted in intravascular catheterization to facilitate interventional decision making and actions, for instance when judgment and expertise of the interventionists are most crucial [6]. Adaptation robotic autonomy in intravascular interventions requires building useful cognition and intuition into new or existing robotic catheter systems for intelligent navigation. While trained interventionists can achieve high success rates with skillful interplay stimuli during interventions, the RCSs require total control dependence of interventionists for robotic catheterization. Also, humans execute specific tasks involved during robotic interventions but they are mostly evaluated with metrics that do not account for cyborg's quality and skillfulness

towards the task's success [7]. Furthermore, the existing approaches used for assessment of interventionists' skill levels for safe interventions lack qualitative and quantitative measures. Thus, effective skill assessment systems are vital to facilitate training of novice-level interventionists with steep learning curves. There exists a high cognitive demand on physicians that human-robot collaborative tool manipulation is a foreseeable step towards autonomous robotic catheterization in intravascular interventions. To this end, purposive learning of tool manipulation skills in human-robot cohort is herein investigated for intravascular interventions.

Recently, the systems are increasingly embraced to eliminate operational hazards interventionists face in *Cath Labs*. Clinical safety and feasibility of robotic intravascular interventions have been reported [5, 8]. Motivated with the increasing maturity in surgical robotic systems and deep learning methods, autonomous robotic catheterization can be launched base on the taxonomy in Yang *et al.* [1]. Initial step requires apt evaluation of interventionists' skills for unified cyborg intravascular manipulation to relieve surgeons' dependence and ensure robots intuition in performing specific interventional tasks.

B. Literature Review

Tool catheterization is a major task that takes during percutaneous intravascular interventions. Possibility of autonomous catheter navigation was investigated for intracardiac interventions in Fagogenis *et al.* [9]. The study was promoted by possibility of over-the-wire distal sensing, while its navigation control was based on thigmotaxis led by haptics feedback. Also, Chen *et al.* [10] developed a deep learning system for autonomous robotic navigation of needle and catheter for drawing blood in the peripheral vessels. Recently, the use of electromagnetic trackers, fiber bragg grating, and multilayer soft electronic is been embraced for cardiac interventions [11-14], while backlash compensation during robotic intravascular catheterization were investigated in [14]. Despite adapting motion and force feedbacks for tool navigation in the robotic prototypes, replicating the sensorimotor skill levels of surgical experts for intravascular manipulation is still a problem. Maor *et al.* [15] indicates that cyborg (surgeon-robot cohort) for intravascular catheterization is imminent towards autonomous robotic catheterization. As existing robotic systems lack definite ways of simultaneously engaging interventionists' cognition and robotic intuition, development on task-specific autonomy of robotic intravascular catheterizations is speculated and evaluated in this study.

Recognition of manipulation skills is important for training and assessment of PCI interventionists. In some studies, interventionists' kinematics and kinesthetics data were used for catheterization skill analysis and hand motion recognition. Studies in the field of signal processing established that hand movements can be characterized with muscle activities during robotic PCI. Thus, Surface EMG (sEMG) signals have been used for technical skills assessment in PCIs [4]. sEMG is an electrophysiological signal modality that describes interventionists' muscular activities during PCI procedures. Application of sEMG-based interface shows high potentials for human-robot synergies. Recently, Meattini *et al.* [16] proposed the use of sEMG-based interface for robotic hands using machine learning and human-robot synergies. In PCI domain,

conventional machine learning methods have been adopted for multimodal data coupling. T. Zhou *et al.* [17] showed that sEMG-based pattern analysis could aid characterization of interventionists' catheterization skills during PCIs. Reiley *et al.* [18] developed a model for automatic recognition of the dexterous hand motions used to maneuver needles during surgical skill training and evaluation. Willems *et al.* [19] assessed endovascular skills of interventionists using the simulator-based testing and rating. In a closer field, X. Zhou *et al.* [20, 21] used hierarchical mixture method for assessing interventionists' skills during intravascular tool manipulations based on multimodal data. Nevertheless, skill training and evaluation approaches in all the methods are based on conventional learning-based techniques. While the two groups are solitary known for studies on sEMG-based hand motion classification for interventionists' evaluation and skill assessments, their reported studies were based on inappropriate methods of input-output data alignment and class labeling. This will not only give wrong assessment results [7], but also hinder application of sEMG-based control for cyborg navigation [21].

Cyborg catheterization and skill evaluation are new research areas in intravascular interventions. This research direction is posited of driving task-specific autonomy in intravascular characterization. This is still lacked in existing commercial robots, such as the top-tier *i.e.* CorPath GRX robotic system (Corindus Vascular Robotics, USA), used for intravascular interventions. The state-of-the-art involved approaches of sub-millimeter navigation but without possibility of surgeon and robotic cohort during PCI. Azizian & Patel [22] developed a methodological concept of data fusion for catheter tracking in robot-assisted catheterization was reported. Recently, Du *et al.* [23] explored interventionists' technical manipulation skills for robot-assisted intravascular catheterization. The study shows that multimodal features from hand kinematics, kinesthesia, and haptics components could be enhanced for proper identification of surgeons' manipulation skills during PCI catheterization. Kim *et al.* [24] recently introduced methods to generate distal guidewire motions in tortuous vessels. He used adapted response of a two degree-of-freedom (2-DoF) robotic catheter system to demonstrate guidewires control.

C. Research Objective

Concerns of data representation, alignment, fusion, and co-learning are major issues deterring adaption of multi-modal learning-based approaches [25]. While representation and fusion modalities were previously addressed [17, 20-24], our focus herein is to design proper data alignment and co-learning, motivated on deep learning. However, the studies were focused on improving PCI paradigm from interventionists' hand stimuli perspectives without input from robotic systems. Hence, they serve no way of preventing interventionists from radiation exposure. Specifically, in this study, a deep learning model is developed for intelligent multi-level skill assessment during cyborg catheterization. This is envisaged to be promising for task-specific autonomy in robotic PCIs. Deep Learning is a subfield of machine learning that involves application of neural networks structured to work like humans' brains [26]. Thus, the hypothesis proposed in this study is evaluated close to how expert interventionists evaluate newbies in the field. The rest of this paper is organized as follows. The methodology developed

to validate the hypothesis proposed is presented in Section 2. This includes details of a robotic catheter system, multimodal data acquisition, feature extraction and pre-processing, fusion strategies, and a stacking-based deep learning model proposed for skill level assessment during cyborg catheterization. Details of the experiments carried out and the results are described in Section 3; Extensive discussion about the experimental results and performance validation are given in Section 4. Finally, conclusions of this study and future works are in Section 6.

II. MATERIALS AND METHODS

This study seeks to cover both research gaps that were identified and discussed above. Thus, we herein hypothesize that multiplexing *multimodal* cyborg data could illustrate interventionists' stimuli interplay for qualitative classification of catheterization skills during intravascular interventions better than using *unimodal* and *bimodal* cyborg data. For this purpose, we portray an analogy that cyborg manipulation skills (based on the workflow in Fig. 1), can be estimated from the interventionist's hand kinematic and kinesthetic data (Fig. 1b), robot's proximal feedback (Fig. 1c), and tool's poses in angiograms during intravascular interventions. These data modalities are pre-processed by extracting unique features from typical data sources on surgeon's and robot's-side; and finally multiplexed to assess the manipulation skill and outcomes of procedures performed with surgeon-robot cohort during intravascular catheterization procedures. Cyborg cognition and intuition are characterized with effective data multiplexing in a multi-layer deep neural network. For this, a stacking-based deep neural network (SDNN) model is developed and validated *in-vivo* to assess and classify the skill levels of interventionists during cyborg intravascular catheterization.

A. Design of Robotic Catheter System

For seamless cyborg integration, a robotic system with master

and slave devices is designed with isomorphic resemblance of how interventionists defter their hands for tool grasping and manipulation during intravascular characterization. As shown in Fig. 2, the interventional robot system is composed of a slave-end actuator with four degrees of freedom and a size of $56 \times 22 \times 16$ cm and the master-end operation with two degrees of freedom and a size of $45 \times 19 \times 13$ cm. Mechanism of the latter is composed with rotary encoders, magnetostrictive sensors, magnetic powder brakes, and TCP communication components that convert motion an interventionist triggered at the master-end device into control signals that is transmitted to the slave device. The slave device is capable of dexterous 4-DoF motion with millinewton axial-feed and grasps feedback force resolutions. The slave robot has a smart grasper for changing orientation of clamped tools (i.e. guidewire) and sensing robot-tool grasp force.

With these devices, intravascular cannulation can be achieved with endovascular tools such as guidewire, while interventionists' hand movement and tool motion data can be fed back through a control system. Control signals are decoded from the potentiometer sensor-bar and transmitted to slave robot closed to real-time (transmission delay < 0.1 s). The minimal-delay control protocol was implemented on sampled data signals for sub-millimeter intermittent and continuous motion, such that the slave device can navigate endovascular tools with both low axial (≤ 0.5 mm) and radial ($\leq 0.5^\circ$) feed accuracies. Operators, namely interventionists, sense proximal axial feed resistance of the guidewire through the manipulator in real time. The master-slave communication adopts TCP/IP to realize signal intercommunication, which can realize wired and wireless remote communication, and the wired communication delay is less than 200ms; the master-slave mechanism has built-in 8000mah lithium battery, which can be used for more than two hours of hardware at a time.

Fig. 1: detailed view of proposed multimodal learning-based cyborg catheterization (a) an artistic view to advocate cyborg intravascular catheterization; (b) sensor placement for multimodal data: includes EM sensors, sEMG electrodes and fiber optic ultra-glove used for acquisition of kinematic and kinesthetic data at master-side; (c) master and slave robotic devices with designated isomorphism that aids Cath Lab manipulation similarities; (d) workflow of the experimental design used for hypothesis validation.

Fig. 2: CAD model of the robotic catheter system for cyborg catheterization; The interventionist is seated at (a) the control station guiding the procedure and able to plan other interventional tasks; (b) the master device is manipulated to trigger navigation commands with which the (c) slave device performs indeed catheterization in the (d) operational room. Interventionists get visual feedback of procedures from the (e) intraoperative display generated with the (f) X-ray imaging machine located in the interventional room.

The robotic catheter system is capable of axial, radial, and hybrid navigation commands for intravascular interventions, and it can execute the commands to mimic natural hand manipulation conventions of interventionists as applied in the *Cath Labs*. In this study, analysis of cyborg tool manipulation skills were decoded from muscular activity from sEMG electrodes, hand motion data from electromagnetic (EM) sensors, and finger flexion signals obtained from fiber optic signals in ultra-glove (Figs. 1b-d).

B. Signal Acquisition and Processing

While guidewire manipulations are basically done with axial translation and radial rotation hand motions, existing studies have considered interventionists' skill analytics as a 2-DoF motion with push or pull in the axial direction and clockwise or counter-clockwise rotation in radial direction [27]. However, interventionist's requires six types of movements in which their hand and fingers are dexterously applied for endovascular tool manipulation. The previous studies on interventionists' skills analysis were based on motion pattern identification [17, 23]; meanwhile, interventionists manipulate the endovascular tools with master-slave robotic system for intravascular catheterization, and tool navigation feedback data are acquired from multiple sources. The multimodal data are adopted in this study to capture unique contribution of the interventionists and robotic system during typical *in-vivo* cyborg catheterization. Based on this objective, data from the multimodal sources are processed for intelligent multi-level skill assessment during the catheterization procedures. The data acquisition, pre-processing techniques, and feature extraction procedures applied on the spatiotemporal data are presented in the following subsections.

1) sEMG Signal Acquisition

A commercial configurable electromyography system (BioNomadix, BIOPAC Systems, Inc., Goleta, CA, USA), integrated with the AcqKnowledge software was used for sEMG signal acquisition. sEMG data are acquired with four channels of the BIOPAC Systems. As demonstrated in Fig. S1 of the supplementary materials, the electromyography data were acquired at the surface of abductor pollicis brevis (APB), flexor carpi radialis (FCR), opponens pollicis (OP), and

extensor carpi radialis (ECR), which are the four muscles of interest (Fig. 3a). Muscle selection was done by a random trial study to find the most appropriate location for acquiring interventionists' kinesthetic signals during catheterization procedures [23]. The acquisition system has been customized to accept connection of loads close to these four channels for calibration purpose, and this was unloaded after the calibration procedure. Calibration based on maximum voluntary contraction amplitude for each interventionist. This is important for fidelity, accuracy and reliability when reading sEMG signals. The calibration procedures that were used in this study are described in the supplementary materials (Fig. S2).

To ensure high signal quality, skin of the operator was properly cleaned with alcohol prior to each experimental session, and the electrodes were placed at the best performing points on each targeted muscles. The signals were acquired from each subject and processed offline using standard procedures. The recorded signals were passed to a 10-500 Hz band-pass filter to ensure that all components of the signals do not have frequencies outside the bandwidth of a typical range for each signal types. For sEMG signals, a 50 Hz notch filter was designed in the ACQKnowledge software to eliminate power frequency disturbances that could be present during the signal acquisition. Both filters were designed with the signal processing toolbox in MATLAB 2019a. Sequel to filtering noise components in the acquired signal, sEMG normalization procedure was done in order to ensure a fair comparison between signals acquired from all the subjects. For this purpose, the maximum value (MV) was used to normalize the sEMG signal. This involves determining the peak value in a certain sEMG data for a distinct movement recorded from a certain subject, and normalizing it to a unique range in which the remaining parts of the signal are expressed as a percentage of the peak value (%MV). This normalization method is used to ensure that a common ground is established when comparing signal from different operators irrespective of their individual differences. Finally, 14 features of sEMG signals from each channel were extracted (that is, a total of 56 features) and used for the skill evaluation procedure. Details of the 14 features are given in Table S1 of the supplementary materials.

2) Finger Motion Signal Acquisition

Dexterity of the subjects' hands was observed as movement of their finger during each during the intravascular catheterization trial. For this purpose, a 14-sensor 5DT Data Glove Ultra (Fig. 3b) was worn on the dominant hand of the subjects while finger flexion signals were recorded as the subjects manipulate his hand to operate the master robot.

Fig. 3: Multiple sensor location for kinesthetic and kinematic signal acquisition with (a) sEMG electrodes; (b) ultra-glove fibers and EM sensors.

The fingers' flexion data is measured with 2 sensors per finger, and as well with one sensor for abduction between every two consecutive fingers. Use of Data Glove Ultra is not always guaranteed as response of the fiber-optic sensors varies with respect to size of subjects' hands; and also, sensors can have non-linear relationships with the fingers' joints anatomical angles. The latter is due to position or influence of other joint movements during hand gesticulations. Thus, subjects' finger flexion data are recorded based on a pre-designed calibration system. The idea followed is all users make specific movement with massive hand dexterity and the data is saved. Other subsequent finger flexion recorded data are calibrated with the former to avoid sensor nonlinearities effects when similar types of data were recorded from the same subject but over different time periods. Validation of the calibrated signals confirms that the recorded data are not affected by experimental procedures. Finally, normalization of the calibration data is set to a range of 0 and 1, and these values are exported for further processing. Thus, finger flexion from recorded from 14 channels that is tip, base, and abduction (between consecutive fingers) of each finger were acquired and scaled to analyze interventionists' finger dexterity during a typical catheterization. Scaled signals were calibrated and the inter-point flexions of each finger were computed as independent features.

3) Hand Motion Signal Acquisition

Kinematics of interventionists' hand motion signals were obtained with two Aurora 6-DoF Flex Tube, $\Phi 1.3$ mm, 40 Hz electromagnetic (EM) position sensors (Northern Digital Inc., Canada). The sensors were fixed with rubber glues on the 5DT Data Glove Ultra which is worn by the interventionists during the intravascular catheterization procedures. The EM sensors were positioned to be fitted on tips of the surgeon's thumb and forefinger fingers, as presented in Fig. 3b. Each EM sensor can record three-dimensional (x, y, z) position information and orientation (pitch, yaw, roll) angle data with a field generator having a measurement volume of $20 \text{ cm} \times 20 \text{ cm} \times 7 \text{ cm}$. The EM field is employed to transmit electromagnetic signals to make the EM sensors work. It has position and orientation tracking accuracy of 0.9 mm and 0.8° , respectively; and approximately 20 ms latency for stable closed-loop control. In this study, only the position data is used for further processing. The 3-D position of hand motion is depicted by the Cartesian coordinates, and intra-point displacement of hand motion is obtained based on plane reflection of the coordinates. To eliminate interference between the magnetostrictive sensor on master device of the intravascular robot and the EM and fiber optic-based sensors worn on subjects hand, the latter sensors were position a bit far away from the finger tips. This served to avoid direct contact with the magnetostrictive sensor, while the field generator was position closer to subjects who were also advised to ensure limiting their fingers on the knob of the master device. These were done to guarantee little-to-no mutualism between the different sensors. Displacement of tool grasper in the robotic slave device was recorded between two adjacent points during catheterization procedures.

4) Guidewire Motion Acquisition

Proximal motion of the endovascular tools are obtained with the encoder of the slave device. The distal motion variables of the endovascular tools are obtained from the X-ray images that

were acquired during the catheterization procedures. Details of the medical imaging procedures and segmentation are presented in the next sub-section. For coherence of this write-up and readers' clarity, only the procedures used for feature acquisition from the X-ray angiograms are discussed here. Tip centroid of the flexible endovascular tool is evaluated by computing the translation displacement and rotation angle between two adjacent points during catheterization procedures. These features are used to characterize tool's response to slave robot's motion command during the tool navigation process.

C. Image Acquisition and Processing

Distal motion and force sensing in robotic intravascular catheterization are usually limited with lack of miniaturized sensors and instability of over-the-wire devices. As an alternative in this study, we utilized image-based approach in determining the distal motion of the guidewire during cyborg catheterization. For this purpose, we acquired and processed fluoroscopic x-ray images that were acquired with our self-developed [28] and commercial X-ray machines during the *in-vivo* studies, respectively. The former is a proprietary five degrees of freedom C-arm machine we developed for real-time medical imaging. In every trial, the X-ray machine uses an X-ray tube powered under 70kVolts and 5 mA to produce the X-rays images of the bone and vascular structure of the rabbit. Iohexol Omnipaque was used as contrast agent for visibility of vascular structure at the start of the studies. During the cyborg catheterization, the contrast dye was injected to aid acquisition of 2D angiograms with navigation views of the guidewire and possibly the blood vessels. For clear interventionists' visualization, 2D X-ray images were acquired at high resolutions without angiography subtraction. While the X-ray machine produces sequences of fluoroscopic images at an average of 10 frames-per-second, the commercial C-arm machine was set to output its image sequences at an average of 15 frames-per-second.

Custom Matlab codes were written to read image content from the native format output gotten from the C-arm machines, and the automated tool segmentation and tracking framework in Fig. 4 was proposed and developed for sensitive extraction of guidewire pixels in the X-ray angiograms without specificity loss [29]. The method was based on vesselness measurement (Eq. 1) and ridge detection (Eq. 2) at pixel-level analysis. For the former, intra-frame pixel subtraction was performed based on the first image frame (background image), followed by top-hat filtering which was applied to extract morphological information up to every small elements and details from the X-ray image frames. Next, large speckle noises were search in a threshold range to remove all background pixels. Median filter was used for pixel denoising to remove noise speckle and background pixels that overlaps guidewire pixels' intensities.

Fig. 4: Block diagram of the proposed segmentation and tracking method

Contrast enhancement is used to increase the difference between bright and dark regions of the angiogram and smoothing is also performed with median filtering to reduce noise effects. While Frangi filter [29] was applied to detect vessel-like structures in the frames, intensity values returned for pixels in each frame are mapped into a range of 0 to 1, probability of being guidewire or background pixel.

$$\tilde{X}_{en}(\sigma) = \begin{cases} 0 & \text{if } \lambda_2 > 0 \\ \exp\left(-\frac{\mathcal{R}_B^2}{2\beta^2}\right) \times \left(1 - \exp\left(-\frac{\mathcal{S}^2}{2c^2}\right)\right) & \text{else} \end{cases} \quad (1)$$

$$\tilde{X}_v^d = \begin{cases} \lambda_{n-1}(p_0) > \mu_v & \text{if } p_0 \text{ is valley} \\ \lambda_{n-1}(p_0) < \mu_v & \text{if } p_0 \text{ is ridge} \end{cases} \quad (2)$$

Where $\mathcal{R}_B$ is a ratio denoting the largest cross section area that accounts for deviation from blob-like structures and distinguishes between line and plate patterns in second order structure of the image $\left(\frac{\partial}{\partial x} \tilde{X}_{en}\right)$; $\mathcal{S}$ is a local degree measure of the cross sectional blobness of the image. The factors are parameterized as $\mathcal{R}_B \triangleq |\lambda_1|/|\lambda_2|$, and $\mathcal{S} = \sqrt{\sum_k \lambda_k^2}$; where β and c are factors used to regulate sensitivity of structure contrast with respect to noise and background pixels. λ_1 and λ_2 are eigenvalues. The vesselness index from different values of σ in Eq. 2 are analysed to obtain a final estimate of vesselness.

$$\tilde{X}_v(p_0) \oplus SE = \bigcup_{e \in SE} \tilde{X}_v(p_0)_e \quad (3)$$

Sequel to that, the guidewire structure is tracked and extracted with the ridge and valley detection method and joining interleaved pixels that have intensity values in a defined range as the guidewire pixels (Eq. 3). Tip centroid location of the guidewire is defined as the top-most coordinates in the guidewire pixels. Displacement of the tool is determined as the norm of a vector defined on points $\{\tilde{X}_v(p_i), \tilde{X}_v(p_j)\}$, while its orientation is computed as direction of the vector using a dot product. For compatibility with other data modalities, the spatial pose data tuples are converted to a time series based on the frame rates the imaging machines operate.

D. Multi-level Skill Assessment

A stacking-based deep neural network (SDNN) model is proposed as part of the skill assessment and procedure evaluation system workflow presented in Fig. 1. The model operates on pluralities of basic learning modules, stacked one after another, to synthesize an assessment system for evaluating interventionists' catheterization skills and trial outcomes. The SDNN model has component that are developed to perform the following three basic operational functions.

1) Data Fusion Module

Analysis of interventionists' skills is envisaged on information obtained from variety of data sources namely the EM sensors, sEMG electrodes, fiber optic ultra-glove, slave tool grasper, x-ray angiograms. By fusing information from the multiple modalities, manipulation skills are represented with the kinematics and kinesthetics of interventionists' hand and endovascular tool during the cyborg catheterization.

Resolution-based dynamic time warping was done for proper alignment of multimodal data. Given n time series data $X := \{x_a, x_b, \dots, x_n\} \forall x_a \cong x_b \cong x_n \text{ s.t. } x_i = [x_i^1, x_i^2, \dots, x_i^m] \in \mathbb{R}^{d \times m_i}$, the time series data can be converted to a similar feature space $\mathcal{F}$ to optimize Eq. 4, where matrices of binary alignment paths P_a and P_b for windows in the time series, respectively.

$$\begin{aligned} \arg \min_{P_a, P_b} & \|x_a P_a - x_b P_b\|_{\mathcal{F}}^2 \\ \text{s.t. } P_a & \in \{0,1\}^{m_a \times m} \\ P_b & \in \{0,1\}^{m_b \times m} \end{aligned} \quad (4)$$

The time-based explicit data alignment procedure only regards the overlapping portions of all modalities to achieve unsupervised temporal alignment of time series with different start, stop, and sampling times. Thus, useless lagging and leading segments of the data are determined with respect to the analysis in Fig. 5. Next the useful signals from the data sources are processed based on modalities sampling time to cohere information from the different data sources. Similarities between segments of the multimodal data are ensured by constructing fixed-length temporal orderings with window sequence of each data modality. A key understanding of the unsupervised alignment is all data modalities have the same structure despite recapitulated from different systems [31].

2) Multimodal Feature Extraction:

Data segments in each record represent unit hand movements of the interventionists and arranged to carry out multi-level skill assessment in the SDNN model. The network is designed to couple salient features describing interventionists' stimuli interplay, and classify the catheterization trials based on interventionists' manipulation skills and trial outcomes. Each experimental trial is recorded as a pre-processed matrix data with a dimension of $1 \times F_t \times S_t$, where F_t is the number of features and S_t is the length of data obtained for a catheterization trial (t). The data acquisition, pre-processing techniques, and feature extraction applied on the spatiotemporal data are available in the *Materials and Methods* section. Based on the procedures, a total of 76 features ($F_t = 76$) were extracted. These include 56 features from sEMG signals, 14 features from finger motion, 2 features from hand motion, from interventionists (*i.e. subject-side*); displacement from proximal part of the slave device (*i.e. slave-side*), along with tool axial displacements and angular change from the guidewire tip centroid (*i.e. tool-side*) at distal slave-side in the robotic system.

3) Deep Network Modeling:

The model has stacked layers which enhance self-learnable features that can be decisively trained down to minutiae within a simplified network structure [32]. The SDNN model shown in Fig. 6 is developed for learning the interventionists' tool manipulation skills during a cyborg catheterization procedure, and predicting the catheterization trial outcomes into different success levels. The deep learning architecture has an encoding-decoding structure with five blocks stacked next to each other. In this purported structure, a sequential constructor with two linear modules is adapted to exploit temporal and spatial details of the extracted features and utilize the information for prediction purposes.

The encoding segment includes three blocks stacked with similar composition of layers. As the basic building units of each block, the layers consist of *tensor-in tensor-out* computation function and states holding the weights and biases of trainable parameters. The blocks use convolution layers to analyze the temporal and spatial invariants from the multimodal features based on their shared-weights and translation invariance characteristics. As shown in Fig. 6, the convolution layers apply 64 kernels (at size of 3, stride of 1, and same length padding), which is followed by batch normalization to ensure the network processes tensors faster and stable. For this, input tensors at the input layer are re-centered and re-scaled. This applies a transformation that maintains mean and standard deviation outputs close to 0 and 1, respectively. Furthermore, a rectified linear unit (ReLU) is used to allow small gradient for inactive cases. Feature multiplexing is done by transforming summed weighted input in each block with a ReLU activation function. These salient details encode the skills an interventionist applied in a catheterization trial.

In the decoding segment, two major blocks are added. Sizes of the feature maps are reduced to small-dimension feature maps vector by means of convolutional and max-pooling operations in the first three blocks. Although feature multiplexing provides details into minutiae that reveals intricacies needed to classify a catheterization trial. However, effects of feature redundancy are reduced with a follow-up dropout layer in each block. The second module includes two blocks stacked with different composition of layers. The primary operation in both blocks is done in the prediction layer that has fully connected linear nodes between the input and output nodes. While the first block transforms its input by a linear equation into a tensor with output sample size of 128; the second block does the final skill-level prediction with respect to subject's local and global profile at a sample size of 3. Local profile is based on analyzing values of each trial with respect to all trials made by an interventionist, while the global profile involves all the trials made towards this study procedure. In the final block, each catheterization trials are classified into three different levels base on the success prediction level of the trial. These are Level-A (High-level), Level-B (Low-level) success, or Level-C (Unsuccessful).

Fig. 5: Unsupervised alignment of multimodal data signal processing

III. EXPERIMENTAL RESULTS

In this section, the experimental procedures designed to validate the proposed system, details of the data preprocessing and feature extraction procedures, and the evaluation studies carried out are presented in this section.

Fig. 6: SDNN model developed with stacked convolutional and dense nodes to sufficiently multiplex and learn data statistics in the *multimodal* features

A. In-vivo Experiment and Datasets

To test our hypothesis, in-vivo experimental studies were performed for cyborg intravascular catheterization of three vascular paths in two mammalian class animals. Validity of the hypothesis was verified and evaluated based on different target vessels with varying difficulty of guidewire delivery. The paths are auricle-to-coronary artery in rabbits, femoral-to-coronary vein, and femoral-to-renal artery in a pig. Tool manipulations are analyzed by multiplexing temporal and spatial signals extracted from interventionists' catheterization and underlying robot's contribution. The hypothesis was validated with a nine-day in-vivo experimental study in which several subjects were guided to carry out cyborg catheterization per day. The study included a total of nine subjects with varying intravascular tool navigation skills recruited from both the Leading Medical Service Company and Shenzhen Institute of Advanced Technology. The study was designed to follow approved Institutions' ethical review, and subjects' informed consents were obtained accordingly. Animal care followed the procedures advocated by the Institutions' Ethics committee.

Daily studies started with path creation in which an indwelling needle was inserted via auricle vessel in rabbits, and introducer sheath via femoral vessel in pig. Mammalia (six rabbits 2.21 ± 0.29 kg; and one 35.1 kg pig) were pre-medicated intravenously with 2.5 mg/kg of isoflurane for rabbits and intramuscularly with 40 μ g/kg of anticholinergics for pig. In the nine-day studies, 60 cyborg guidewire catheterization trials (rabbit: 26, pig: 34) were performed as recorded in **Movies S1** and **S2**. Subjects were pre-informed about the chosen vascular pathways by using visual analyses made from MRI and fluoroscopic imaging of animals. Catheterization assessments were implemented based on how subjects catheterize the designated vascular pathways. For objective assessment, dataset recorded from the 60 trials were labeled using a 3-level scale, where Level-A presented those successful trials that were completed in less than the median time (82.5 secs for rabbit, and 92 secs for pig) and from the subject with more than 60% successful rate, the remained successful trials are represented as Level-B; and Level-C as unsuccessful trials.

B. Model Implementation and Training

Three study categories were analyzed by examining the three different features modalities i) for rabbit trials only; ii) for pig trials only; and iii) for both rabbit and pig trials combined. The SDNN model is implemented with PyTorch® framework. Network training was done with an adaptive moment (adam) optimizer with dynamic learning rate initialized at 0.002, a weight decay value of 10^{-6} , and 10% dropout rate which was added for regularization. Performance of the final model were obtained under training epochs of 200, and batch size = 16. The supervised training was done in a batch mode at a size of 20. During training, the learning rate was gradually decreased with an applied momentum of 0.9, while regularization with weight decay rate of 10^{-4} is utilized to avoid over-fitting, and the adam regularizer was used for optimizing the SDNN model. Moreover, mini-batches with batch size of 16 were employed to maximize GPU usage and avoid GPU-CPU memory transfers in each training iteration. The optimizer is run for 4 iterations per mini-batch and 200 epochs for each layer.

During the in-vivo studies, a total of 60 catheterization trials (23 in rabbit and 37 in pig) were performed, and statistics of the recorded procedures shows a mean data length ($\bar{S}_t$) of 132 ± 87 . The obtained catheterization data were analyzed for the 3 categories described above; which include inter-animal (i and ii) and intra-animal (iii) data analyses. For each category, the dataset is partitioned into distinctive 60% and 40% sets which are used for training and evaluation of the SDNN. Model parameterization described above was utilized for network implementation in the three categories. Then, the network was trained on 10 different random partitioning of training dataset with a validation split of 20%. Next, our proposition is validated by illustrate interventionists' stimuli interplay within the stated hypothesize. While multimodal feature characterization involves multiplexing features from more than two data modalities at subject's side, the unimodal and bimodal feature characterization were implemented as characterization of cyborg's catheterization skills based on features from one and two data modalities at the subject's side, respectively.

C. Classification Results

In the three categories, robot's cognition data were acquired as guidewire motion data (i.e. translation displacement and rotation angle between two adjacent points) and the slave device motion; these were combined with data modalities obtained from interventionists' side. Random partitioning of datasets with the multimodal features was observed 10 times, and the skill assessment results were obtained as average classification accuracies for each study category. Average classification accuracy values of 96.43 ± 0.50 %, 86.77 ± 1.55 %, and 95.50 ± 0.49 % were obtained for the three study categories as shown in Fig. 7a-c, respectively. These indicate the model's ability to assess the catheterization skills users demonstrated with the cyborg setup in the three study categories, respectively. It can be interpreted that the proposed SDNN model achieves very high classification accuracies in the three case scenarios, with the highest accuracy (98.2%) value observed on data from study with rabbit only, while 84.5%, the least accuracy value, was observed from the study that involved only pig. The third study category involves

dataset from both animals, and the classification accuracies obtained are relatively higher than that of pig only study. These performances demonstrate the potentials of using the stacked network for characterization and assessment of cyborg catheterization skill and possible aid the development of autonomous robotic catheterization during the interventions.

To evaluate the usefulness of multiplexing the features that characterizes interventionist's cognition and robot's intuition, cases of unimodal and bimodal feature combinations were also observed for skill evaluation in the three categories. Compared with the accuracies achieved with the multimodal features, the SDNN provided classification accuracies of 87.72 ± 1.06 %, 69.81 ± 3.73 % and 69.76 ± 2.76 % for assessment of skill levels based on unimodal feature learning; and accuracy values of 94.15 ± 0.75 %, 73.67 ± 1.71 %, and 80.78 ± 2.51 % for assessment based on bimodal feature multiplexing in the three study categories, respectively. In the unimodal case, the highest accuracies were achieved when only the features of "Muscular Activity" signals, whereas the case of features from "Muscular Activity" and "Finger Flexion" signals produced the highest accuracies in the case of bimodal feature evaluation. In all cases, signals from subject-side were combined with the slave-side and tool-side features. However, other feature groupings employed to validate skill levels assessment with both unimodal and bimodal feature sets yielded lower accuracies, as presented in Fig. 7. In conventional skill assessment, navigation analysis of endovascular tools at proximal and distal points is often used but in subjective manners. To improve this practice, features of the slave grasper and guidewire motions are applied for qualitative and quantitative skill assessment in the second study category. The different feature combinations show that different skill levels can be analyzed and detected however, the performances of the SDNN varied for each distinct feature multiplexing, as presented in Fig. 7.

Furthermore, intra-level performance of the model is analyzed by measuring how it accurately predicted skill level users displayed for each trial recorded during the cyborg catheterization. The predictive analysis was obtained for a random partition of the datasets using datasets with multimodal feature sets only. The results obtained for each study category are shown in Fig. 8a-c, respectively. Comparative analyses of the results were made and SDNN model was provided the best results when applied on rabbit dataset while the least performance accuracy was obtained from the pig dataset. This can be attributed to number of subjects and total trials carried out for each animal studies. Nevertheless, intermediate classification results were obtained when the 60 trials (dataset from both rabbit and pig) were used to validate the proposed SDNN model in study category (iii). Compared with the mean accuracy values, each randomly partitioned data shows accuracy differences ranging between 2-5% which typifies the dynamic ability of subjects within each skill level. In addition, the results of these validation studies direct the possibility of surgeon-robot cohort and skill level assessment in intravascular catheterization. Moreover, it was realized that better evaluation results were obtained by combining features that reflect both interventionists' cognition and robot's intuition, extracted from multiple dataset. These show that skill assessment during cyborg catheterization can be automated to replace the

traditional scoring system currently used for evaluating procedures made by interventionists, and this could aid autonomous robotic catheterization.

Comparative studies were carried out to explicate the proposed stacked deep learning network on datasets from each individual animal and a merger of both. For this purpose, four well known machine learning methods, namely support vector machine (SVM), linear discriminant analysis (LDA), multilayer perceptron (MLP), random forest (RF), were implemented for skill level assessment based on the in-vivo datasets. The different data modalities characterizing the cyborg skills (i.e. interventionists' cognition and robot's intuition) were labeled and used to validate the four machine learning methods in the unique study categories. To begin with, classification accuracies of the stacked learning network were compared with the results obtained from the conventional learning networks in the three study groups. The five learning-based methods were used to assess the skill levels achieved by subjects in the cyborg catheterization trials; this involved classifying success-level outcome of each trial as a proxy subjects' skill levels. For instance, Fig. 9 shows the

assessment accuracies of each method on different unimodal, bimodal, and multimodal feature combinations in the rabbit trial. The radar charts shows that the proposed SDNN has classification accuracy values of 87.00%, 67.38% and 46.95%, respectively for skill level assessment during the cyborg catheterization in the rabbit trials. These are quite lesser than the accuracy values obtained with MLP and SVM in the study category. A similar trend was observed with bimodal feature combinations in the rabbit trials. However, as shown in Fig. 10, assessment of skill levels on the feature modalities was accessed in the pig dataset. The results show the proposed SDNN performed slightly better than the five methods except MLP. Just like existing learning-based methods, performance of the proposed SDNN mode gets higher with the addition of features from more data modalities. For instance, bimodal features gave a better classification accuracy values compared to performances achieved with unimodal feature combinations. While the maximum classification accuracies that were obtained with unimodal and bimodal features are 66.33% and 82.36% respectively, better classification performances were achieved with the *multimodal* feature combinations.

Fig. 7: Experimental results: classification accuracies obtained with proposed SDNN for skill level assessment based on trials in (a) rabbit trials only; (b) pig trials only; and (c) rabbit and pig trials, combined.

Fig. 8: Intra-level classification results in the three study categories. Results are from a random partitioning of the datasets from (a) rabbit trials only, (b) pig trials only data, and (c) rabbit and pig trials combined.

Fig. 9: Performance evaluation; comparison of SDNN vs. MLP, RF, SVM, and LDA model with only rabbit data

It can be seen in Fig. 10 that a combination of “MA + HM + FF + GM + SM” yielded a classification accuracy of 86.77% while a maximum accuracy of 84.34% was achieved with features from four-modal set. Analysis of the evaluation results shows that some of the conventional models, viz. MLP and SVM, performed a bit more than the SDNN using the different feature modalities combined from the cyborg dataset. It is worthy to note that acquisition of slave motion data was a later development during the study implementation. Basically, it was done to actualize increase autonomy in cyborg catheterization; thus, it was not acquired during the rabbit study. Similarly, comparative assessment of the five methods was based on features obtained in both animals, simultaneously. This time, classification accuracies were also examined and, as shown in Fig. 11, assessment of skill levels on different unimodal and bimodal feature combinations in the dataset shows the proposed SDNN performed slightly better than the five methods except MLP. While the maximum classification accuracies that were obtained with unimodal and bimodal features are 66.33% and 82.36% respectively, a better performance, proxy as classification accuracy, was achieved with the *multimodal* feature combinations. It can be seen in Fig. 10 that a combination of “MA + HM + FF + GM + SM” yielded a classification accuracy of 86.77% while a maximum accuracy of 84.34% was achieved with features from four-modal set. Analysis of the evaluation results shows that some of the conventional models, viz. MLP and SVM, performed a bit more than the SDNN using the different feature modalities combined from the cyborg dataset. It is worthy to note that acquisition of slave motion data was a later development during the study implementation. Basically, it was done to actualize increase autonomy in cyborg catheterization; thus, it was not acquired during the rabbit study. Similarly, comparative assessment of the five methods was based on features obtained in both animals, simultaneously. This time, classification accuracies were also examined and, as shown in Fig. 11, assessment of skill levels on different unimodal and bimodal feature combinations in the dataset shows the proposed SDNN performed slightly better than the five methods except MLP. In fact, it had a better classification performance than SVM when multimodal features: “MA + FF + HM + GM” was accessed, and similarly that MLP when bimodal features: “MA + HM +

FF” was assessed. In summary, the proposed SDNN shows a similar trend for other methods when different data modalities are applied for skill level and task outcome classification.

IV. DISCUSSION

Recent success of machine intelligence is attributed to models that learn with internal data representations [31]. As reflected in the validation results, the proposed SDNN demonstrated significantly high classification accuracies for the catheterization trials in the three study categories. Thus, it is vital to establish the performance of the SDNN within the context of machine learning. The recent progress in machine intelligence is attributed to the availability of exploded data, low-cost computation and storage capacities, and newly developed learning algorithms. Thus, deep learning was reported to be suboptimal when trained with small-sized data [31-34]. The *in-vivo* experiments produced a multimodal dataset of 54,430 samples from both animals which were pre-processed as discussed in *Materials and Methods* section, and used to characterize subjects’ manipulation skills. Comparatively, MLP demonstrates mean accuracies of $98.52 \pm 0.60\%$, $94.79 \pm 0.75\%$, and $91.40 \pm 1.01\%$ when used on the *multimodal* feature combination in the three study categories, respectively. These values are higher than the mean accuracy values of $96.43 \pm 0.50\%$, $86.77 \pm 1.55\%$, and $90.52 \pm 0.49\%$ obtained with the proposed SDNN model. A similar trend was observed for *bimodal* and *unimodal* feature combinations.

Unlike in MLP model, the proposed SDNN model guaranteed global minima, using early stopping, during each training session with the ten random data partitions. A similar status could be achieved by deciding the number of hidden neurons and optimal validation check in the MLP model. However, the dynamics to this is very high as imperfect model tuning may result in underfitting or overfitting problems in the MLP model. Furthermore, SDNN shows nearly similar accuracy values with conventional SVM model in several multimodal feature combinations. This could be understood as the proposed SDNN model has similar performance with SVM when assessing tool manipulation skills and trial outcomes in dataset from small- and big-sized animals. This is backed up with literature saying SVM model is a suitable learning-based method with reliable performance in this domain [35].

Furthermore, emphases are placed on stability of proposed SDNN model in the current study while further improvements are desired. It is noteworthy to elucidate that the proposed model performed better than several other conventional machine learning models. For instance, the LDA and RF models which, for most of the feature combinations in the cyborg data, demonstrated lower classification accuracies compared to the values that were obtained with the SDNN model. A few exceptions were found in comparing the classification results for several *unimodal* feature combinations in the second study category. As presented in Fig. 7, LDA shows higher mean accuracy values of 71.36 ± 2.79 , 38.11 ± 0.78 , and 54.76 ± 1.71 compared to the performance attained by the SDNN (43.27 ± 1.52 , 34.66 ± 0.89 , and 66.32 ± 2.43) for “MA+GM”, “HM+SM”, and “FF+SM”, respectively. These performances can be mainly attributed to structure of the

classification models on small-sized dataset with fewer parameters. For instance, highest accuracy values were obtained in cases of the highest possible features combination for each study category. The analyses of Figs. 7-10 show that the learning-based models could perform better when multiple data types from the interventionist and robot were used to characterize the cyborg catheterization procedures. The models automatically learn more feature details from best feature maps which enable discovery of apt representations needed for skill level classification in the cyborg dataset. The conventional learning models benefit from careful data engineering with considerably high domain expertise used to achieve the feature extraction and transformation from the raw data [29]. Thus, we advocate that only the SDNN and related deep learning models might be proper for online skill evaluation under task-specific autonomous robotic intravascular interventions.

Fig. 10: Performance evaluation; comparison of SDNN vs. MLP, RF, SVM, and LDA model with only pig data.

Fig. 11: Performance evaluation; comparison of five learning-based models with combined rabbit and pig data.

To stimulate discussion, a roadmap for task-specific autonomy in robotic intravascular interventions is opinionated. We herein proposed a stacking-based deep neural network for assessment and classification of cyborg skill levels during intravascular catheterization. Experimental *in-vivo* studies were observed for proper data acquisition, and stacked layers of the deep neural network were utilized for intelligent extraction and coupling of salient spatiotemporal features obtained from a robot-assisted catheterization. The *in-vivo* studies show that the isomorphic robotic platform could enhance cyborg manipulation, as both master and slave devices performed stably during all the 60 trials. Thus, the unsuccessful trials are as result of subjects' manipulation cognitive skills in intravascular catheterization. The proposed SDNN has a stable performance with very high classification accuracies (96.43%, 86.77%, and 95.50%) for three-level skill assessments base on features from multimodal data sources. These accuracy values are above those of unimodal and bimodal features from the same setup. Thus, utilization of the learning-based model with the *multimodal* features extracted is a plausible way towards task-specific autonomous robotic interventions. Level of data modality used to model the network has a direct relationship with network's performance.

V. CONCLUSION

Robotic systems are gradually embraced to eliminate the operational hazards interventionists face during intravascular interventions; however, current trend requires total control dependence and intuition of human experts. Meanwhile, analytical learning of interventionists' tool manipulation skills from multiple data-sources could enhance appropriate transfer for intelligent surgeon-robot tool navigation during intravascular interventions. In this study, a master-slave robotic system exhibiting defter isomorphism close to how interventionists dexterously manipulate endovascular tools in the Cath Lab is proposed. As a vital step towards autonomy in robotic catheterization, a purposive learning of guidewire catheterization skills in human experts is investigated. This involved extraction of intelligent features and representation of human manipulation knowledge using a stacking-based deep learning model. Classification of trial outcomes based on the

data that were extracted from slave device indicates that the robotic system has some contributions in cyborg manipulations. With this, task-specific autonomous catheterization can be implemented for specific intravascular paths, whereas surgeons can take over in other specific tasks or if the robot's autonomy is deemed unreliable. Using multimodal features provided a reliable classification model; however, multiplexing motion and force control inputs with resultant proximal and distal feedbacks shall be investigated to extend the paradigm for task-specific autonomous control during robotic intravascular catheterization.

REFERENCES

- [1] Yang G., Cambias J., Cleary K., Daimler E., Drake J., Dupont P., Hata N., Kazanzides P., Martel S., Patel R., Santos V., and Taylor R., Medical Robotics-Regulatory, Ethical, and Legal Considerations for Increasing Levels of Autonomy, *Science Robotics* 2(4), (2017).
- [2] Ali Z., Nef H., Escaned J., Werner N., Banning A., Hill J., Bruyne B., Monotorfano M., Lefevre T., Stone G., Crowley A., Matsumura M., Maehara A., Lansky A., Fajadet J., and Mario C., Safety and Effectiveness of Coronary Intravascular Lithotripsy for Treatment of Severely Calcified Coronary Stenoses—The Disrupt CAD II Study, *Circulation: Cardiovascular Interventions*, 12(10):1-10, (2019).
- [3] Reardon M., Van Mieghem N., and Popma J., et. al., Surgical or Transcatheter Aortic-Valve Replacement in Intermediate-Risk Patients, *New England Journal of Medicine*, 376(14), 1321-1331 (2017).
- [4] Maier-Hein L., Vedula S., Speidel S., et.al., Surgical Data Science for Next-generation Interventions, *Nat Biomed Eng.*, 1:691-696 (2017).
- [5] Mahmud E., Naghi J., Ang L., Harrison J., Behnamfar O., Pourjabbar A., Reeves R., and Patel M., Demonstration of the Safety and Feasibility of Robotically Assisted Percutaneous Coronary Intervention in Complex Coronary Lesions Results of the CORA-PCI Study (Complex Robotically Assisted Percutaneous Coronary Intervention), *JACC: Cardiovascular Interventions*, 10, 1320-7, (2017).
- [6] M. Thai P. Phan, T. Hoang, S. Wong, N. Lovell, and T. Do, "Advanced Intelligent Systems for Surgical Robotics", *Advanced Intelligent Systems*, 2(8), 1-33 2020.
- [7] Ortenzi, V., Controzzi, M., Cini, Leitner J., Bianchi M., Roa M. and Corke P. Robotic Manipulation and the Role of the Task in the Metric of Success. *Nature Machine Intelligence*, 1:340-346, (2019).
- [8] Hirai T., Kearney K., Kataruka A., Gosch K., Brandt H., Nicholson W., Salisbury A., and Grantham J., TCT-221 Initial Report of Safety and Feasibility of Robotic-Assisted Chronic Total Occlusion Coronary Intervention, *Journal of the American College of Cardiology*, 74(13), S1, (2019).
- [9] Fagogenis G., Mencattelli, Machaidze Z., Rosa B., Price K., Wu F., Weixler V., Saeed M., Mayer J., Dupont P., "Autonomous Robotic Intracardiac Catheter Navigation using Haptic Vision", *Science Robotics*, 4(29), 1-12, (2019).

- [10] Chen A., Balter M., Maguire T. and Yarmush M., Deep Learning Robotic Guidance for Autonomous Vascular Access, *Nature Machine Intelligence*, 2:104–115, (2020).
- [11] Jäckle S., Eixmann T., Schulz-Hildebrandt H., Hüttmann G., Pätz T., “Fiber Optical Shape Sensing of Flexible Instruments for Endovascular Navigation”, *International Journal of Computer Assisted Radiology and Surgery*, 14, 2137–2145, (2019).
- [12] Han M., Chen L., Aras K., Liang C., and Chen X. *et al.*, “Catheter-integrated Soft Multilayer Electronic Arrays for Multiplexed Sensing and Actuation during Cardiac Surgery”, *Nature Biomedical Engineering*, 3:PP, (2020).
- [13] Cerenelli L., Bortolani B., and Marcelli E., “CathROB: A Highly Compact and Versatile Remote Catheter Navigation System”, *Applied Bionics and Biomechanics*, 1-13 (2017).
- [14] Omisore O. Han S., Ren L., Wang G., Ou F., Hui L., Wang L., Towards Characterization and Adaptive Compensation of Backlash in a Novel Robotic Catheter System for Cardiovascular Intervention, *IEEE Transactions on Biomedical Circuits and Systems*, 12:824-838, (2018).
- [15] Maor E., Eleid M., Gulati R., Lerman A., and Sandhu G., “Current and Future Use of Robotic Devices to Perform Percutaneous Coronary Interventions: A Review”, *Journal of American Heart Association*, 117:1-8 (2017).
- [16] R. Meattini, S. Benatti, U. Scarcia, D. De Gregorio, L. Benini, and C. Melchiorri, “A sEMG-Based Human-Robot Interface for Robotic Hands Using Machine Learning and Synergies”, *IEEE Trans. on Components, Packaging and Manufacturing Technology*, 8(7): 1149-58, Feb. 2018.
- [17] T. Zhou, O. M. Omisore, W. J. Du, W.K. Duan, Y. Zhang and L. Wang, “A Preliminary Study on Surface Electromyography Signal Analysis for Motion Characterization during Catheterization”, proceedings of 12th International Conference on Intelligent Robotics and Applications, pp. 617–628, (2019).
- [18] C. Reiley, B. Varadarajan, B. Vagvolgyi, and H. Lin, “Automatic recognition of surgical motions using statistical modeling for capturing variability”, *Studies in Health Technology and Informatics*, 132:396-401, 2008.
- [19] M. Willems, J. Vliet, V. Willams, L. Kool, D. Bergqvist, and J. Blankenstein, “Assessing endovascular skills using the simulator for testing and rating endovascular skills (STRESS) machine,” *European Journal of Vascular Endovascular Surgery*, 37(4):431–436, 2009.
- [20] X. Zhou, G. Bian, and Z. Hou, An Interventionalist-Behavior-Based Data Fusion Framework for Guidewire Tracking in Percutaneous Coronary Intervention. *IEEE Trans. Syst. Man Cybern. Syst.* 50(11):4836-49, 2020.
- [21] R. Li, X. Zhou, G. Bian, X. Xie and Z. Hou, “Recognition of Endovascular Manipulations using Recurrent Neural Networks”, **41st IEEE International Conference on Engineering in Medicine and Biology Society**, Berlin, Germany, July 23-27 2019.
- [22] M. Azizian and R. Patel “Data fusion for catheter tracking using Kalman filtering: applications in robot-assisted catheter insertion”, *Proceedings of SPIE Medical Imaging 2011: Visualization, Image-Guided Procedures and Modeling*, 796413(2011):1-11.
- [23] W. J. Du, O. M. Omisore, W. K. Duan, T. Zhou, X. D. Lv, Y. F. Li, S. P. Han, Y. Al-handarish, Q. H. Liu, L. Wang, “Exploration of Interventionists’ Technical Manipulation Skills for Robot-Assisted Intravascular PCI Catheterization”, *IEEE Access*, 2020(8): 53750-65.
- [24] Y. Kim, A. Kapoor, R. Finocchi, and E. Girard, “An Experimental Validation of Behavior-Based Motions for Robotic Coronary Guidewire Crossing Techniques”, **International Symposium on Experimental Robotics**, Buenos Aires, Argentina, November 5-8, 2020.
- [25] T. Baltrusaitis, C. Ahuja, L. Morency, “Multimodal Machine Learning: A Survey and Taxonomy”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423-443, 2019.
- [26] O. Omisore, S. Han, L. Ren, A. Elazab, N. A. Azeez, T. A. Talaat, H. Li, and L. Wang, “Deeply-Learnt Damped Least-Squares Method for Inverse Kinematics of Snake-Like Robots”, *Neural Networks*, 107 (2018): 34-47, 2018.
- [27] A. Rolls, C. Riga, C. Bicknell, D. Stoyanov, C. Shah, I. Herzele, M. Harmdy, and N. Cheshire, “A pilot study of video-motion analysis in endovascular surgery: Development of real-time discriminatory skill metrics”, *European Journal of Vascular Endovascular Surgery*, 45(5):509–515, 2013.
- [28] H. Liang, J. Xiong, Y. Liu, Z. Gan, and Z. Xia, Vibration Analysis and Optimization of Mobile C-arm CT System, **7th IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems**, July 31 – August 4, 2017, Hawaii, USA
- [29] Omisore O., W. Du, W. Duan, H. Wang, Y. Zheng, Y. Al-Handarish, T. O. Akinyemi, Y. Liu, J. Xiong, and L. Wang, “Automatic Tool Segmentation and Tracking during Robotic Intravascular Catheterization for Cardiac Interventions”, *Quantitative Imaging in Medicine and Surgery*, 11(6):2688-2710, (2021).
- [30] Roads B. and Love B., “Learning as the unsupervised alignment of conceptual systems”, *Nature Machine Intelligence*, 2:76–82 (2020), <https://doi.org/10.1038/s42256-019-0132-2>
- [31] Low C., Park J., Teoh A., Stacking-Based Deep Neural Network: Deep Analytic Network for Pattern Classification, *IEEE Transactions on Cybernetics*, PP:1-14, doi: 10.1109/TCYB.2019.2908387, (2019).
- [32] Jordan M. and Mitchell M., Machine learning: Trends, perspectives, and prospects, *Science*, 349(6245), 255-260, 2015
- [33] LeCun Y., Bengio Y., and Hinton G., Deep learning, *Nature*, 521(7553), 436-444, (2015).
- [34] X. Zhou, X. Xie, Z. Feng, Z. Hou, G. Bian, R. Li, Z. Ni, S. Liu, Y. Zhou, “A Multilayer and Multimodal-Fusion Architecture for Simultaneous Recognition of Endovascular Manipulations and Assessment of Technical Skills”, *IEEE Transactions on Cybernetics*, PP. 1-13.