A Natural-Behavior-Based Multilayer Learning Framework for Decoding Endovascular Manipulations

Xiao-Hu Zhou[1], Hong-Bin Liu[4], Xiao-Liang Xie[1], Zhen-Qiu Feng[1], Zeng-Guang Hou[1,2,3], *Fellow, IEEE*, Gui-Bin Bian[1], Shi-Qi Liu[1], Rui-Qi Li[1,3], Zhen-Liang Ni[1,3] and Yan-Jie Zhou[1,3]

[1]State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[2]CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing 100190, China

[3]University of Chinese Academy of Sciences, Beijing 100049, China

[4]Centre for Robotics Research, King's College London, London WC2R 2LS, U.K.

Email: xiaohu.zhou@ia.ac.cn

***Abstract*—The complexities of endovascular manipulations result in the difficulties developing human-robot interfaces (HRI) to maintain natural manipulations of interventionalists. In this study, a multilayer learning framework is proposed to decode six typical endovascular manipulations by fusing four types of natural behavior. Based on the characteristics of behavioral data, the framework is designed to three layers to decouple manipulations partly. Six classification-based and three rule-based fusion algorithms are evaluated for performance comparisons. Extensive experiments and statistical analysis demonstrate that the proposed framework can achieve the overall accuracy of 96.90% based on the best three-behavior fusion scheme, much higher than those on the best single-behavior scheme (92.67%) and the best two-behavior fusion scheme (95.50%). These hopeful results indicate the great potential of the framework to facilitate the future development of novel HRI for endovascular robotics.**

I. Introduction

Cardiovascular diseases (CVDs) are the leading cause of death all over the world [1]. Endovascular procedures (e.g., percutaneous coronary intervention, PCI) are currently primary therapies for the treatment of CVDs. Despite minor trauma and short recovery time to patients, exposure to high doses of X-ray radiation also results in an increased incidence of cancer, cataracts and other disease to medical staff, which limits their practical use.

In the past decade, endovascular robotics (Magellan$^{\mathrm{TM}}$ [2] and CorPath GRX [3]) have seen a growing interest in the treatment of CVDs. Most of them are designed as a master/slave control way, not only removing the operators from radiation source, but also increasing the precision and stability of tool motions [4]. Despite the increased application of robot-assisted procedures, the endovascular manipulations in manual procedures (called natural manipulations) have not been considered in the designs of existing robotic systems. Unlike conventional bedside techniques, robot-assisted procedures are implemented via multi-DoF joysticks or navigation buttons [5]. Therefore, endovascular tools are removed from interventionalists' hands, resulting in lack of force feedback and forgoing manipulation skill accumulated in conventional procedures. Although some studies have explored these issues, they are currently still in the stage of research [6]–[9]. Up to now, existing endovascular robots are still facing a lot of challenges. One of the major problems is that natural manipulations can be hardly learned by robots since it is difficult to implement the transformation from natural endovascular manipulations to the ones in robot-assisted procedures [10]. Consequently, it is very necessary to decode complicated manipulations of operators for robots to perform corresponding manipulations in time.

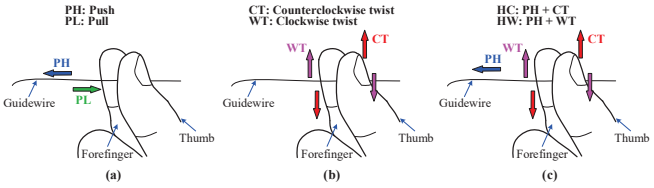

Fig. 1. The description of six endovascular manipulation patterns. (a) Pull (PH) and Pull (PL). (b) Counterclockwise twist (CT) and clockwise twist (WT). (c) Push and counterclockwise twist (HC), push and clockwise twist (HW)

PCI is a typical endovascular procedure implemented by using guiding catheter, guidewire, balloon/stent catheter etc. Firstly, a guiding catheter is inserted into radial or femoral artery and threaded to the appropriate coronary ostium. Then a guidewire is put into coronary artery through the catheter lumen until its tip threads past the stenosis. Next, a balloon/stent catheter is delivered along the positioned guidewire to the desired treatment site and inflated to keep the blocked artery open [11]. Among these steps, guidewire delivery is the most important and involves more types of manipulation mode than others. Generally, endovascular manipulations in guidewire delivery can be categorized to three modes: axial (AX), circumferential (CF) and combined (CB) [12], [13] (see Fig. 1). The AX mode includes two opposite-direction patterns: *Push*

(PH) and *Pull* (PL) to achieve guidewire advancement and retraction. By clamping the tool with the thumb and forefinger of right hand, the PH pattern is implemented via the hand motion from right to left, while the PL is achieved by the opposite motion. The CF mode can adjust the orientation of guidewire tip when encountering with vascular bifurcations. It also consists of two opposite patterns: *counterclockwise twist* (CT) and *clockwise twist* (WT), which are implemented via the twisting motion generated by the two fingers. Sometimes, simultaneous push and twist manipulations are used to adjust the tool's position and orientation dynamically. Similarly, *push and counterclockwise twist* (HC), *push and clockwise twist* (HW) are two patterns of this mode.

Recently, some researchers used natural behaviors (motion signals from surgeons' body during surgical procedures) to analyze surgical manipulations. In [14], the operator's hand motion profiles and velocity in endovascular surgery were collected with an electromagnetic (EM) sensor mounted on the thumb. Based on these information, typical manipulation patterns were determined by testing both *in vitro* phantoms and commercial simulators. Researchers at Nagoya University also used EM sensors to acquire hand and wrist motion in endovascular surgery simulation [15]. The collected data, together with the information from other sensors, were processed to extract related features for technical skill assessment. In [16], muscle activity was used to identify surgical manipulations automatically, as well as to distinguish abnormal ones in real time. With noninvasive registered electromyography (EMG) sensors, Villarruel *et al.* [17] used muscle activity to design a robotic surgical system controlled by for remote surgery. Li *et al.* [18] analyzed finger motion with 14 custom-made bend sensors to obtain a comprehensive information and reflect hand function clinically. Recent studies explored the application of accelerometers, and the acquired average and maximum acceleration were obtained for skill assessment [19].

Some others have explored the fusion of multimodal behaviors to describe surgical manipulations more accurately and completely. By using fiber-optic bend sensors and accelerometers, researchers at Imperial College London [20] developed a data glove to record the operator's finger motion and manipulation force in laparoscopic surgery. They analyzed these behavioral data for optimal sensor selection and surgical skill assessment. In [21], finger motion and muscle activity were collected by a data glove and electromyography (EMG) sensors, respectively. These natural behaviors were applied to the assessment of manipulation ergonomy in laparoendoscopic single-site surgery.

Thus far, there are many studies looked at the application of natural behaviors for skills assessment. We hope that such behaviors can also be used to decode endovascular manipulations for facilitating the development of novel HRI, but few studies focused on this issue. Moreover, existing fusion of two types of natural behaviors still only provides limited information, which is hardly used in real-time applications. For more complete information, it is necessary to explore an appropriate framework to integrate four types of natural behaviors mentioned above. Motivated by that, this research aims to develop a natural-behavior-based multilayer learning framework to learn endovascular manipulations of interventionalists for endovascular robots by fusing multimodal natural behaviors.

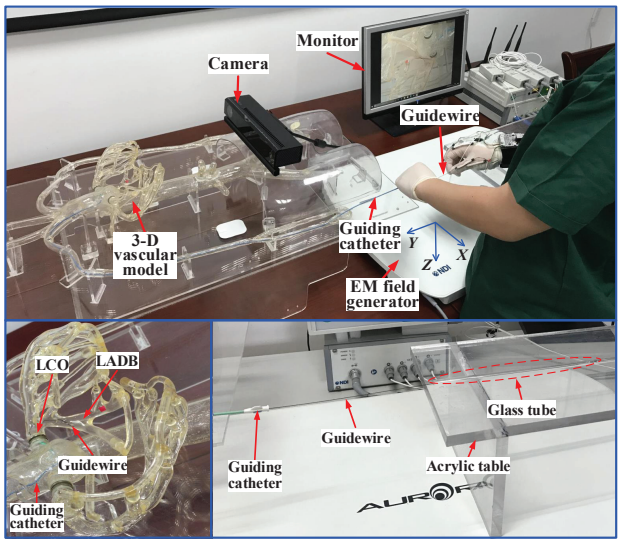

Fig. 2. Experimental setup. (a) Phantom-based simulator. (b) Coronary arteries. (c) Acrylic table and glass tube.

The remainder of this paper will introduce the method in Section II. In Section III, experimental results are presented and discussed. Finally, we conclude in Section IV.

II. Methods

A. Data Acquisition

1) Phantom-based simulator: To simulate clinical practice, a phantom-based simulator (see Fig. 2), including a guiding catheter, medical guidewire, high-definition camera, 3-D vascular phantom, monitor, acrylic table and glass tube, is developed for performing endovascular manipulations. The catheter tip is positioned at the left coronary ostium, and the guidewire is inserted into the coronary artery through the lumen of the catheter. The phantom, filled with specialized oil to substitute for blood, is used to simulate the vascular system of humans. The camera is placed near the phantom to simulate X-ray fluoroscopy and provide two-dimensional (2-D) navigation.

2) Sensor deployment: According to above analysis, human motion produced by the interventionalist's body can be considered as natural behaviors contributing to endovascular manipulations. In this study, we will mainly discuss hand motion, muscle activity, finger motion and proximal force. Table I shows sensor deployment for natural behavior acquisition. The three-dimensional (3-D) position (*X, Y, Z*) acquired from EM sensors are considered as hand motion of operators. From more specific Fig. 3, surface electrodes of EMG sensors are placed based on corresponding anatomical locations. Two fiber-optic bend sensors are mounted on thumb's metacarpophalangeal and forefinger's interphalangeal

TABLE I
SENSORS AND CORRESPONDING ACQUIRED NATURAL BEHAVIORS

Sensor	Location	Abbr.	Behavior
EM (40 Hz, NDI Inc.)	Thumb	EM_T	Hand motion (HM)
	Forefinger	EM_F	
EMG (1500 Hz, Noraxon Inc.)	Biceps brachii	EMG_{BB}	Muscle activity (MA)
	Triceps brachii	EMG_{TB}	
	Dorsal interossei	EMG_{DI}	
	Abductor pollicis brevis	EMG_{APB}	
Fiber-optic bend (40 Hz, 5DT)	Thumb	FOB_T	Finger motion (FM)
	Forefinger	FOB_F	
Accelerometer (1500 Hz, Noraxon Inc.)	Hand back	Accele.	Proximal force (PF)

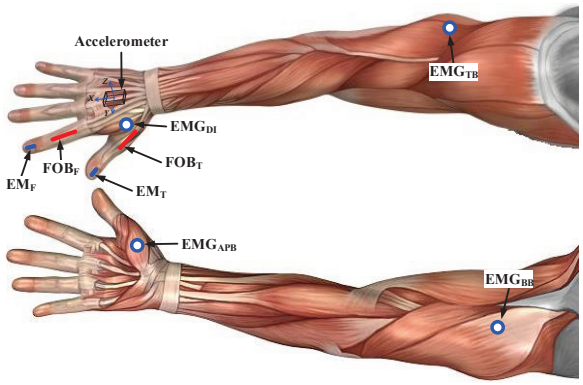

Fig. 3. Sensor deployment

joint, respectively [22]. The accelerometer's X-axis is along the middle finger and a Cartesian space is spun by three axes.

3) Experimental protocol: In data collection, ten interventionalists were recruited. All subjects have right dominant hands, guaranteeing the consistency between endovascular manipulations and the simulator. Before data collection, they were trained to familiar endovascular manipulations on the simulator. The axial mode (PH and PL) was achieved by advancing and retracting the guidewire between the left coronary ostium (LCO) and the left anterior descending branch (LADB); The circumferential mode (CT and WT) was implemented at a bifurcation of the LADB; The combined mode (HC and HW) was started at the LCO. Each subject's manipulation lasted for five seconds and repeated ten times for a specific pattern. A short rest (2-3 minutes) between two attempts was allowed to prevent muscle fatigue. All attempts of each pattern were started at the consistent posture/gesture of subjects and initial state of the guidewire, and digital clock signals were used to control synchronous acquisition of multimodal behaviors.

4) Preprocessing: For data filtering, a notch filter (50 Hz) and a band-pass filter (10-500 Hz) are used to remove the noise in muscle activity (EMG signals), while median filters are employed to remove spurious spikes and outliers for other behaviors. Due to the different sampling rates, the alignment between high sampling rate data and the low one is necessary.

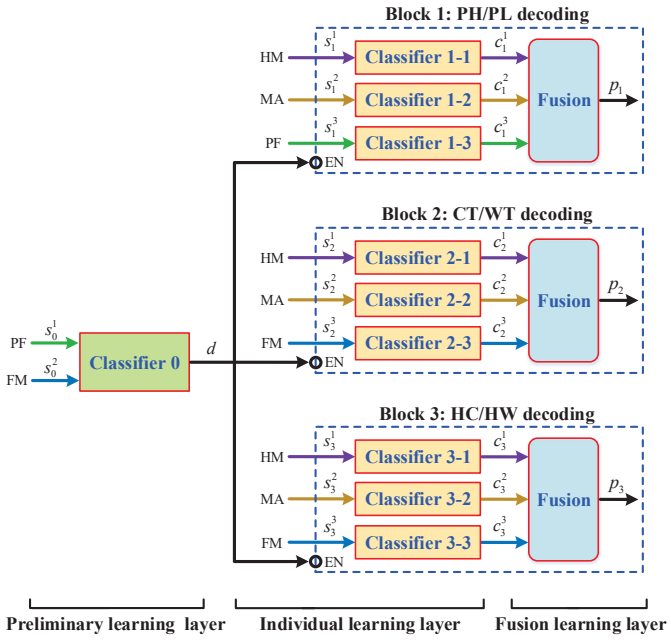

Fig. 4. The multilayer learning framework. "EN" enables a block when input the corresponding trigger signal.

To this end, the mean absolute value of muscle activity is calculated using a fixed-length EMG data. The sequence length is the ratio of the two sampling rates (1500 Hz/40 Hz). Similarly, the mean value of proximal force is extracted. For other behaviors, the filtered data is regarded as the corresponding feature. Besides, the windowing (200 ms) [23] and overlapping (75%) [24] technique are adopted. Then the segmented data are normalized with min-max scaling method, and further concatenated to a feature vector as an experimental sample. Thus 970 (each pattern) and 5,820 (six patterns) samples are obtained from one subject's manipulations.

B. Multilayer Learning Framework

Section I has mentioned that six endovascular manipulation patterns can be categorized into three modes. From the pre-experiments, we found that proximal force only contributes to the AX and CB modes, while finger motion is only involved in the CF and CB modes. Hence, manipulation patterns can be decoupled by making full use of these obvious characteristics. Specifically, the decoding task can be implemented by classifying three modes at first, and recognizing six patterns in the specific mode subsequently. To this end, a multilayer learning framework (see Fig. 4) is designed with three layers: preliminary learning layer, individual learning layer, and fusion learning layer. The first layer is used to classify AX/CF/CB for preliminary decoding, and the latter two layers consist of three blocks for PH/PL, CT/WT, and HC/HW decoding, respectively.

1) Preliminary learning layer: In this layer, three manipulation modes (AX, CF and CB) are preliminarily decoded by the classifier 0, whose input is the feature vector concatenated by proximal force (s_0^1) and finger motion (s_0^2). This procedure

TABLE II
LABELS OF ENDOVASCULAR MANIPULATION MODES AND PATTERNS

Mode	Label	Pattern	Label
AX	$(1, 0, 0)^T$	PH	$(1, 0, 0, 0, 0, 0)^T$
		PL	$(0, 1, 0, 0, 0, 0)^T$
CF	$(0, 1, 0)^T$	CT	$(0, 0, 1, 0, 0, 0)^T$
		WT	$(0, 0, 0, 1, 0, 0)^T$
CB	$(0, 0, 1)^T$	HC	$(0, 0, 0, 0, 1, 0)^T$
		HW	$(0, 0, 0, 0, 0, 1)^T$

can be denoted as

$$d = \Phi(s_0) \tag{1}$$

where d is the preliminary predicted result, a 3-D column vector representing AX, CF or CB in Table II, Φ denotes the classifier, s_0 is the concatenated feature vector. As a trigger signal, the predicted result is further used to activate the corresponding block in the next layer.

To select the most appropriate preliminary classifier for this task, six popular classification models, linear discriminant analysis (LDA), random forest (RF), support vector machine (SVM), extreme learning machine (ELM), generalized regression neural networks (GRNN), and back-propagation neural network (BPNN) are compared.

2) Individual learning layer: In a specific block, multimodal behavioral data are processed separately by different classifiers to obtain individual predictions, representing the corresponding behavior's semantic information. Specifically, this layer can decode PH/PL using classifiers 1-1 to 1-3, CT/WT using classifiers 2-1 to 2-3, and HC/HW using classifiers 3-1 to 3-3. This procedure can be denoted as

$$c_i^j = \Psi_i^j(s_i^j),\ i \in [1, N],\ j \in [1, M] \tag{2}$$

where c_i^j denotes the individual predicted probability vector produced by the individual classifier Ψ_i^j on the feature vector s_i^j. Both N and M are set to three in the framework.

Similarly, those classifiers adopted in the first layer are also used as the candidate individual classifiers. For a specific behavior, the one with the best decoding performance is selected as the final individual classifier.

3) Fusion learning layer: The individual predicted probability represents the corresponding behavior's semantic information, and describes the possibility that an endovascular manipulation belongs to a certain pattern. It is probably that the obtained probabilities may be dissimilar for different behaviors, and the complementarity among different semantic information should be fully utilized. By concatenating individual predicted results, a semantic information vector is established for this purpose. Subsequently, the vector is then processed by a fusion algorithm for further decoding. This procedure can be denoted as

$$p_i = \Gamma_i(c_i^1, ..., c_i^M),\ i \in [1, N] \tag{3}$$

where p_i is the final predicted result, a six-dimensional (6-D) column vector representing corresponding patterns in Table II, Γ_i is the corresponding fusion algorithm.

Above six models are also used in this layer as the classification-based fusion algorithms. Besides, average rule (AR), majority voting rule (MVR) and max rule (MR) are also considered as rule-based fusion algorithms [25]. Different from the individual learning layer, this layer integrates the individual predicted probabilities produced by corresponding individual classifiers.

Due to the individuality of subjects, subject-specific models are established independently. Firstly, each subject's data is divided into three non-overlapping parts: local-training dataset (40%), fusion-training dataset (40%) and testing dataset (20%). Then, six preliminary classifiers are trained with the local-training dataset, and the fusion-training dataset is used to test them to select the best one. After that, the preliminary classifier with the highest average accuracy on all subjects' data is the SVM model. Next, the local-training dataset, together with ground truth labels are utilized to build individual classifiers. For each case in different blocks, the best one is determined by comparing decoding performance on the fusion-training dataset. After that, the best individual classifiers are SVM for HM, RF for MA, GRNN for PF, and BPNN for FM. Finally, obtained semantic information from the best individual classifiers are employed to train six classification-based fusion algorithms (no need for training rule-based ones). The testing dataset is used to evaluate the decoding performance of the proposed framework. The hyperparameters in some classification models or fusion algorithms are determined by 4-fold cross validation.

III. RESULTS AND DISCUSSIONS

A. Preliminary Learning Schemes

Based on ten subjects' testing dataset, the average decoding accuracies obtained from different candidate classification models are shown in Fig. 5(a). By comparing the results of different classifiers, the SVM model yields an accuracy of 97.87%, indicating the best decoding performance, which is consistent with the previous training result. Therefore, this model is selected as the final preliminary classifier for the following decoding. Furthermore, the overall result on all testing data is presented as a confusion matrix [see Fig. 5(b)], further indicating the detail of manipulation decoding. From the figure, more than 3826 testing samples of AX are decoded accurately, achieving a recall of 98.61% and a precision of 98.13%, while only 97.06% of CB samples are decoded effectively.

B. Individual Learning Schemes

After the above procedure, the testing samples, which are classified to corresponding modes accurately, are further processed by candidate individual classifiers in corresponding blocks. This can further validate the rationality of the selection of individual classifiers on the fusion-training dataset. For each

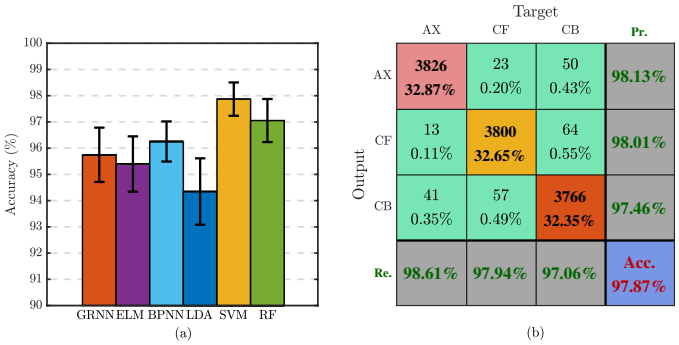

Fig. 5. The results of preliminary learning schemes. (a) The accuracy achieved by six preliminary classification models. (b) The confusion matrix based on the SVM model. (Acc.: accuracy, Re.: recall, Pr.: precision. The number in the top of a cell is the class count, and the bottom one is the percentage of the count to the total number of testing samples, similarly hereinafter.)

TABLE III
AVERAGE ACCURACY (%) OF INDIVIDUAL LEARNING SCHEMES

Block	Scheme	Best classifier	Accuracy
PH/PL	**HM**	**SVM**	**95.82%**
	MA	RF	91.74%
	PF	GRNN	93.78%
CT/WT	HM	SVM	94.76%
	MA	RF	92.16%
	FM	**BPNN**	**95.05%**
HC/HW	HM	SVM	92.91%
	MA	RF	91.32%
	FM	**BPNN**	**93.18%**

modality, the best individual classifier and achieved average accuracy are given in Table III. It can be seen that the best classifiers on the testing dataset are consistent with the ones on the fusion-training dataset, demonstrating the previous selection is reasonable. In each block, the classifier achieved the best decoding performance for a specific modality is selected for the next multi-behavior fusion. For PH/PL decoding, the best single-behavior scheme (BSBS) yields an accuracy of 95.82%, which is obtained by the SVM model using hand motion (HM). The FM-based BPNN model achieves the BSMS for both CT/WT decoding (95.05%) and HC/HW decoding (93.18%). The decoding results of BSBS are shown in yellow highlighted cells in Table III. By comparing the decoding results, HM indicates high potential in decoding PH/PL patterns, and FM outperforms other natural behaviors in classifying twist-involved manipulations, while MA shows the poor capability in recognizing endovascular manipulations.

C. Fusion Learning Schemes

In this part, multi-behavior fusion is explored, and nine candidate fusion algorithms are evaluated on the testing dataset to find the most advantageous method. The highest average accuracy of a specific multi-behavior fusion and the corresponding algorithm are shown in Table IV. It can be seen that the decoding accuracies of multi-behavior fusion schemes outrun those of single-behavior schemes. For PH/PL decoding, the highest accuracy (98.61%) is obtained by the GRNN fusion model based on PF-HM fusion scheme, which is determined as the best two-behavior fusion scheme (BWBFS). Similarly, FM-HM fusion scheme achieves the BWBFS for CT/WT decoding (97.13%) using the SVM model, and also for HC/HW decoding (96.97%) using the RF model. Form the table, the best three-behavior fusion schemes (BHBFS) outperform corresponding two-behavior fusion and single-behavior schemes. Specifically, the BHBFS are achieved by the SVM model for PH/PL decoding (99.63%), the BPNN model for CT/WT decoding (98.92%) and HC/HW decoding (98.46%), respectively. The decoding results of BWBFS and BHBFS are shown in cyan and magenta highlighted cells in Table IV, respectively.

TABLE IV
AVERAGE ACCURACY (%) OF FUSION LEARNING SCHEMES

Block	Scheme	Best classifier	Accuracy
PH/PL	HF-MA	SVM	97.57%
	MA-PF	GRNN	95.35%
	PF-HM	**GRNN**	**98.61%**
	HM-MA-PF	**SVM**	**99.63%**
CT/WT	HM-MA	SVM	95.24%
	MA-FM	BPNN	96.29%
	FM-HM	**SVM**	**97.13%**
	HM-MA-FM	**BPNN**	**98.92%**
HC/HW	HM-MA	SVM	95.03%
	MA-FM	BPNN	95.91%
	FM-HM	**RF**	**96.97%**
	HM-MA-FM	**BPNN**	**98.46%**

Fig. 6 further displays the decoding details of nine best schemes with confusion matrices. Similar to decoding accuracy, recall and precision also indicate a continuous upward trend with the increase in the number of used behaviors. Based on these confusion matrices, the final decoding accuracies under different best schemes can be calculated. They are 92.67% for the BSBS, 95.50% for the BWBFS, and 96.90% for the BHBFS, respectively. These results indicate the effectiveness of the multilayer learning framework on decoding endovascular manipulations.

Furthermore, the decoding abilities of different schemes are also evaluated with F_1-score, which is the harmonic average of recall and precision. In this study, F_1-score is calculated by the macro-average method. The recall, precision, F_1-score are used to draw the radar figures under different blocks in Fig. 7. The green, blue and red lines represent the BSBS, BWBFS and BHBFS, respectively. From the subfigures, the BHBFS covers the largest area than the others, and the BWBFS is better than the BSBS. These results further demonstrate that the decoding performance can be improved by the appropriate fusion of more natural behaviors.

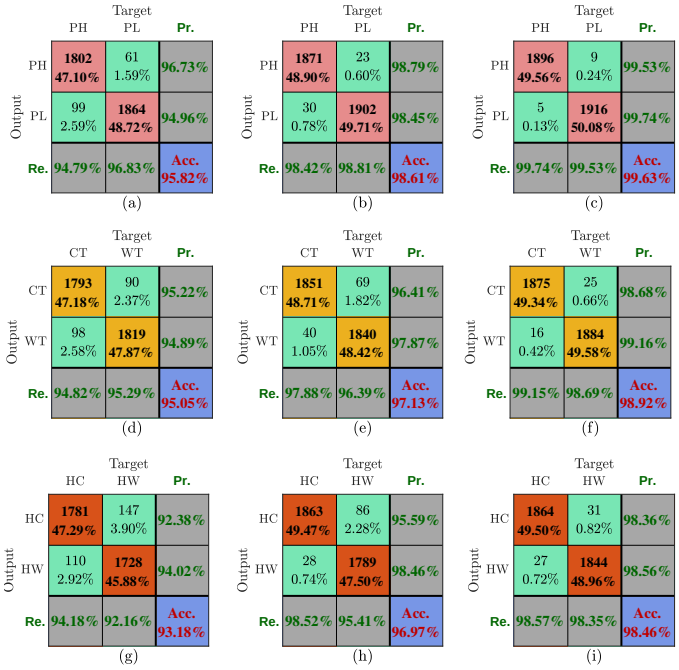

Fig. 6. The confusion matrices of nine best schemes. (a), (b) and (c) are the BSBS, BWBFS and BHBFS under PH/PL decoding; (d), (e) and (f) are corresponding schemes under CT/WT decoding; (g), (h) and (i) are corresponding schemes under HC/HW decoding.

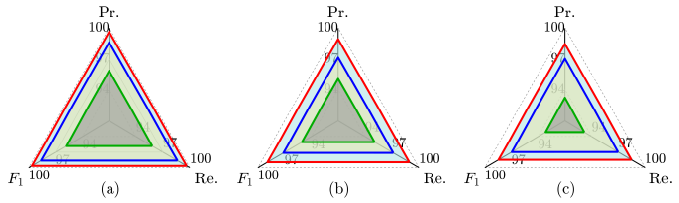

Fig. 7. Radar figures based on average Re., Pr., and F_1-score (%). (a) PH/PL decoding. (b) CT/WT decoding. (c) HC/HW decoding. The green, blue and red lines represent the BSBS, BWBFS and BHBFS, respectively.

D. Discussion

This paper mainly discusses decoding endovascular manipulations with natural behaviors of interventionalists. In clinical practice, a practical and feasible framework needs acceptable decoding performance, which is mainly affected by behaviors, classification models, and fusion algorithms.

For different behaviors, hand and finger motion are acquired with high-stability and low-noise data, demonstrating more competitive ability for manipulation decoding than muscle activity. Through appropriate fusion with others, muscle activity also indicates the potential to improve the decoding accuracy. The multi-behavior fusion can fully take advantages of not only modality-specific contents but also complementarity among multimodality to obtain more accurate decoding. By using more modalities, the relationship between endovascular manipulations and behavioral data can be also described more completely. For different manipulations, the proposed framework shows much more difficulties in decoding the combined manipulations since they are more complicated and involve more behaviors than others.

From the decoding results, some classification models, BPNN, SVM, GRNN and RF, indicate higher appropriateness to decode endovascular manipulations than others. In addition, it can be found that LDA shows poor decoding capability, which means that the dominant relationship between manipulations and behaviors is nonlinear. Moreover, the classification-based models outperform the rule-based ones in terms of decoding performance. This is because the formers hold high robustness to individual difference and high sensitivity to sample change. By making use of the correlation between different patterns, the proposed framework can also optimize the decoding structure through decoupling endovascular manipulations partly. It can also be applied in other situations involved more fusion models and more behaviors because of its high extensibility.

IV. Conclusion

This paper presents a natural-behavior-based multilayer learning framework for decoding endovascular manipulations. Compared with single-behavior schemes, multi-behavior fusion can bring considerable improvement in decoding performance. The multilayer structure can also be optimized by partly decoupling the relationship between manipulation patterns. In subsequent work, sensor miniaturization and integration will be considered for more convenient acquisition, and a novel HRI will be developed based on the proposed learning framework to maintain natural manipulations of interventionalists.

Acknowledgment

This research is supported by the IEEE CIS Graduate Student Research Grants. The authors want to acknowledge all subjects who participate in the experiments.

References

[1] S. Mendis, P. Puska, B. Norrving, W. H. Organization, *et al.*, *Global atlas on cardiovascular disease prevention and control*. Geneva: World Health Organization, 2011.

[2] C. V. Riga, C. D. Bicknell, A. Rolls, N. J. Cheshire, and M. S. Hamady, "Robot-assisted fenestrated endovascular aneurysm repair (fevar) using the Magellan system," *Journal of Vascular and Interventional Radiology*, vol. 24, no. 2, pp. 191–196, 2013.

[3] C. C. Smitson, L. Ang, A. Pourdjabbar, R. Reeves, M. Patel, and E. Mahmud, "Safety and feasibility of a novel, second-generation robotic-assisted system for percutaneous coronary intervention: first-in-human report," *Journal of Invasive Cardiology*, vol. 30, no. 4, pp. 152–156, 2018.

[4] C. V. Riga, C. D. Bicknell, M. S. Hamady, and N. J. Cheshire, "Evaluation of robotic endovascular catheters for arch vessel cannulation," *Journal of Vascular Surgery*, vol. 54, no. 3, pp. 799–809, 2011.

[5] W. Saliba, V. Y. Reddy, O. Wazni, J. E. Cummings, *et al.*, "Atrial fibrillation ablation using a robotic catheter remote control system: initial human experience and long-term follow-up results," *Journal of the American College of Cardiology*, vol. 51, no. 25, pp. 2407–2411, 2008.

[6] C. J. Payne, H. Rafii-Tari, and G.-Z. Yang, "A force feedback system for endovascular catheterisation," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1298–1304, 2012.

[7] Y. Thakur, J. S. Bax, D. W. Holdsworth, and M. Drangova, "Design and performance evaluation of a remote catheter navigation system," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 7, pp. 1901–1908, 2009.

[8] X. Ma, S. Guo, N. Xiao, S. Yoshida, and T. Tamiya, "Evaluating performance of a novel developed robotic catheter manipulating system," *Journal of Micro-Bio Robotics*, vol. 8, no. 3-4, pp. 133–143, 2013.

[9] J. Guo, S. Guo, T. Tamiya, H. Hirata, and H. Ishihara, "Design and performance evaluation of a master controller for endovascular catheterization," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 1, pp. 119–131, 2016.

[10] H. Rafiitari, C. J. Payne, and G.-Z. Yang, "Current and emerging robot-assisted endovascular catheterization technologies: a review," *Annals of Biomedical Engineering*, vol. 42, no. 4, pp. 697–715, 2014.

[11] A. Mendes, "Percutaneous coronary intervention (PCI)," *Nature Clinical Practice Cardiovascular Medicine*, vol. 10, no. 5, p. 257, 2015.

[12] E. D. Grech, "ABC of interventional cardiology: Percutaneous coronary intervention. II: the procedure," *British Medical Journal*, vol. 326, no. 7399, pp. 1137–1141, 2003.

[13] X.-H. Zhou, G.-B. Bian, X.-L. Xie, and Z.-G. Hou, "An interventionalist-behavior-based data fusion framework for guidewire tracking in percutaneous coronary intervention," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, DOI: 10.1109/TSMC.2018.2876465, 2018.

[14] G. Srimathveeravalli, T. Kesavadas, and X. Li, "Design and fabrication of a robotic mechanism for remote steering and positioning of interventional devices," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 6, no. 2, pp. 160–170, 2010.

[15] C. Tercero, H. Kodama, C. Shi, K. Ooe, S. Ikeda, T. Fukuda, F. Arai, M. Negoro, G. Kwon, and Z. Najdovski, "Technical skills measurement based on a cyber-physical system for endovascular surgery simulation," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 9, no. 3, pp. 25–33, 2013.

[16] Y. Nakaya, C. Ishii, T. Nakakuki, Y. Nishitani, and M. Hikita, "Distinction of abnormality of surgical operation on the basis of surface EMG signals," *IEEJ Transactions on Industry Applications*, vol. 132, pp. 241–249, 2012.

[17] J. E. G. Villarruel and B. T. Corona, "Proposal for a remote surgery system based on wireless communications, electromyography and robotics," in *Proceedings of Electronics, Robotics and Automotive Mechanics Conference*, pp. 93–98, 2008.

[18] X. Li, R. Wen, Z. Shen, Z. Wang, K. D. K. Luk, and Y. Hu, "A wearable detector for simultaneous finger joint motion measurement," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 3, pp. 644–654, 2018.

[19] A. Sánchez, O. Rodríguez, R. Sánchez, G. Benítez, R. Pena, O. Salamo, and V. Baez, "Laparoscopic surgery skills evaluation: analysis based on accelerometers," *JSLS: Journal of the Society of Laparoendoscopic Surgeons*, vol. 18, no. 4, p. e2014.00234, 2014.

[20] R. C. King, L. Atallah, B. P. Lo, and G.-Z. Yang, "Development of a wireless sensor glove for surgical skills assessment," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 673–679, 2009.

[21] F. Pérez-Duarte, M. Lucas-Hernández, A. Matos-Azevedo, J. Sánchez-Margallo, I. Díaz-Güemes, and F. Sánchez-Margallo, "Objective analysis of surgeons' ergonomy during laparoendoscopic single-site surgery through the use of surface electromyography and a motion capture data glove," *Surgical Endoscopy*, vol. 28, no. 4, pp. 1314–1320, 2014.

[22] A. Hollister, D. J. Giurintano, W. L. Buford, L. M. Myers, and A. Novick, "The axes of rotation of the thumb interphalangeal and metacarpophalangeal joints," *Clinical Orthopaedics and Related Research®*, vol. 320, pp. 188–193, 1995.

[23] P. Riley and M. Veloso, "On behavior classification in adversarial environments," in *Proceedings of the International Symposium on Distributed Autonomous Robotic Systems*, pp. 371–380, 2000.

[24] Y. Z. Arslan, M. A. Adli, A. Akan, and M. B. Baslo, "Prediction of externally applied forces to human hands using frequency content of surface EMG signals," *Computer Methods and Programs in Biomedicine*, vol. 98, no. 1, pp. 36–44, 2010.

[25] H. Yang, Q. Du, and B. Ma, "Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 875–879, 2010.