Abed Soleymani, Xingyu Li, Mahdi Tavakoli, Artificial Intelligence in Robot-Assisted Surgery: Applications on Surgical Skills Assessment and Transfer, In Olfa Boubaker (Ed.), Medical Robots and Devices: New Developments and Advances, Elsevier, 2023. ISBN 9780443184604.

Chapter 7

Artificial Intelligence in Robot-Assisted Surgery: Applications to Surgical Skills Assessment and Transfer

Abed Soleymani[a], Xingyu Li[a], and Mahdi Tavakoli[a]

[a] *Electrical and Computer Engineering Department, University of Alberta, Edmonton, AB, Canada. E-mail: zsoleymani, xingyu, mahdi.tavakoli@ualberta.ca*

ABSTRACT

Safety and high-quality education are very important in the learning procedure of surgical trainees before operation on real patients. In the conventional method of education and evaluation, residents are educated and evaluated through traditional methods that are cumbersome, qualitative, and subjective. Quantitative surgical skills assessment and transfer approaches can improve the quality and accuracy of the evaluation and education in surgical training programs. In this chapter, we propose a comprehensive review of AI-powered approaches that detect and incorporate the underlying skills-related features of surgical trajectories to classify and improve the levels of expertise of users in surgical training platforms. To this end, we investigate the functionality, advantages, and drawbacks of current skills evaluation and transfer methods with a focus on robot-assisted surgery applications.

KEYWORDS

Deep learning, Haptics, Machine learning, Robot-assisted surgery, Surgical skills assessment and transfer, Virtual fixtures

2.1 INTRODUCTION

Robot-assisted surgery (RAS) is becoming more popular in modern clinical practice. A surgeon must acquire a variety of skills to conduct RAS safely and effectively since inadequate preparation may negatively affect clinical outcomes Birkmeyer et al. (2013). To help surgical trainees, accurate and reliable methods of assessment and transfer of surgical skills should be available with informative and instructive feedback.

As a convention, RAS skills assessments are conducted through outcomes-based analyses, specially designed checklists, and specific scores Ahmidi et al. (2017). For instance, Martin et al. created "Objective Structured Assessment of Technical Skill (OSATS)", which incorporated operation-specific checklists for pass/fail judgments of the trainees Martin et al. (1997). Another conventional method for identifying levels of robotic surgery expertise is the "Global Evaluative Assessment of Robotic Skills (GEARS)", proposed by Goh et al. (2012) Variability in the human's interpretation of similar events makes such evaluation methods expensive, time-demanding, less efficient, and less reliable. In addition, such observational methods neglect small but potentially important changes in the trainee's skills, preventing them from providing insights and targeted feedback into the surgical outcomes.

Autonomous skills evaluation approaches, however, have the potential to resolve all of the above-mentioned limitations Funke et al. (2019). Surgical robotics technologies are making surgical procedure data more accessible, allowing artificial intelligence (AI) (e.g., machine learning (ML) and deep learning (DL) models) to be incorporated in a variety of RAS skills evaluation and transfer tasks. Recent advances in AI have opened the way for using highly complex surgical recordings to extract meaningful features and build a computerized model of users based on their performance during operations and use the model to classify users' level of expertise. The AI model can identify skills-associated features and then transfer them to the trainee's trajectory to better reflect skillful behavior. A robotic surgery platform may utilize the enhanced trajectory as a reference to generate a virtual fixture that serves as a skillful guide for the user's hand toward a better executive trajectory.

This chapter, in the beginning, will discuss categories of automated methods that extract salient skills-related features from surgical recordings to use them to rate the user performance. Additionally, the benefits and drawbacks of the proposed methods, as well as device regularity and patient safety issues will be discussed. We will also introduce a variety of haptic cue-based skills transfer methods to enhance the skillful behavior of less-experienced users using surgical robot platforms. Finally, in the Conclusions section, we provide several insights and promising research areas related to surgical skills assessment and transfer.

2.2 SURGICAL SKILLS ASSESSMENT

Inductive learning-based models and domain knowledge-based models are main AI categories used in autonomous RAS skills assessment Muralidhar et al. (2018). Inductive learning-based models use data-driven approaches with minimal field knowledge to avoid user bias in the learning process. Because the structure of such methods and even their hyperparameters are mostly determined by the input data (which is usually big data), they have the advantage of reduced training effort (see Fig. 2.1).

Conversely, domain knowledge-based models do not rely on statistical models (e.g., DL or ML models) to discover known features from the system dynamics or human experiences. Surgical robot platforms are very complex physical systems, which cannot be accurately modeled using limited training data. Such

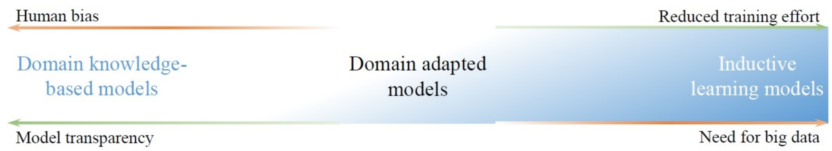

FIGURE 2.1 Different AI model training paradigms according to the strength of prior assumptions about the data, model structure, and task dynamics. As we incorporate more domain knowledge and human bias into the training procedure, we achieve better model transparency over fewer training data samples with the cost of having extra feature engineering effort.

limitations lead us to model uncertainties and unmodeled dynamics. Using field knowledge as a prior decreases uncertainty and makes it easier to solve modeling problems with fewer training data points von Rueden et al. (2019). As a result of incorporating field knowledge in the training stage, the model is often more transparent (i.e., can be clearly understood and explained in human terms) which in turn increases the reliability of the final solution in safety-critical applications such as robotic surgery von Rueden et al. (2019).

2.2.1 Inductive learning-based models

Traditional machine learning methods were used to build the first autonomous surgical skills assessment systems. As surgical trials are composed of a sequence of several predefined subtasks, Rosen et al. proposed the Markov structure of a given surgical task to reveal the user's skills level Rosen et al. (2001, 2002); MacKenzie et al. (2001). Various methods were later used to extend basic hidden Markov models (HMMs) by training a unique HMM for each skill level Reiley and Hager (2009); Tao et al. (2012). Specifically, these studies train separate HMMs for each user and assess their distance from an ideal HMM trained over the data of an expert user. A user's performance is measured by the distance between his or her model and the expert's model. Besides HMMs limited recognition rate and challenge for determining the true number of hidden states, they need manual annotations over trajectories, which is time-consuming. Furthermore, HMMs map a given trajectory to static descriptor space which makes it possible to lose important time-related information within the trajectory. Skills-related temporal features will be discussed in more depth in the following sections.

In recent years, deep learning models have become increasingly popular for RAS surgical skill evaluation applications. In some approaches, kinematic data (the translational and/or rotational trajectories of the robot end-effector) is fed into the convolutional neural networks (CNNs) and used to learn desired patterns for skill assessment on surgical training platforms Jian et al. (2020). Wang and Fey (2018) utilized a deep CNN to highlight the skills levels of individual users using the motion kinematics data of a given surgical operation. Moreover, Fawaz et al. developed a CNN architecture for identifying the surgical skills level of the user via latent pattern extraction of kinematics data of surgical

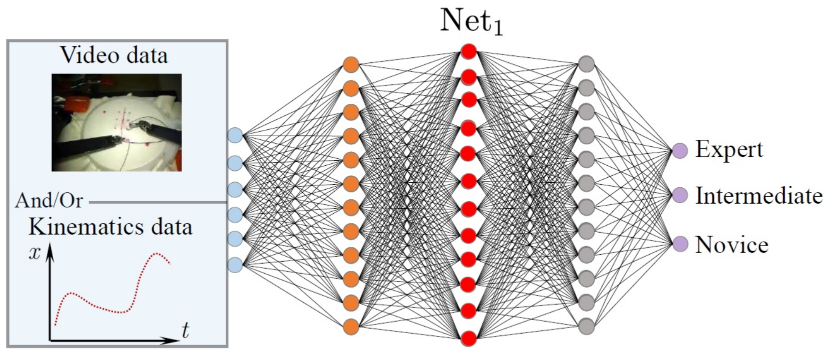

FIGURE 2.2 Inductive learning-based AI models for surgical skills assessment applications. In this paradigm, raw surgical data which can be in the form of video and/or kinematics data (i.e., translational and rotational trajectories of the end effector of the robotic platform) will be fed into the input of the model (mostly deep model). The model's depth, hyperparameters, and the structure of the network will be adjusted according to the complexity and expressiveness of the input data.

trainees performing basic robotic surgery tasks Fawaz et al. (2019). Nguyen et al. developed a classifier network via CNN and long short-term memory (LSTM) models with inertial measurement units (IMU) sensors to highlight user's skills level in a given surgical training data Nguyen et al. (2019).

There is ongoing extensive research on using easy-to-capture video data as the input for AI models which provides rich contextual details compared to previously-mentioned kinematic data. For instance, Kim et al. proposed a temporal CNN to evaluate the intraoperative skills level of capsulorhexis video trials Kim et al. (2019). Funke et al. proposed a DL model using a pre-trained 3D CNN as a temporal segmentation network on the sequence of video frames and optical flow fields for technical surgical skills evaluation tasks Funke et al. (2019). Liu et al. incorporated a supervised regression loss for video input as well as an unsupervised rank loss to train a DL model for RAS skills assessment Liu et al. (2019).

Recent techniques identify the relative variation in skills between pairs of surgeries by creating a pairwise ranking problem Jian et al. (2020). For instance, Doughty et al. incorporated an approach for predicting skills ranking based on video data sets. Using a novel loss function, they utilized both spatial and temporal features (i.e., visual features within each video frame and time-related features along the sequence of consecutive frames) to assess and rate skills Doughty et al. (2018). Doughty et al. (2019) introduced a novel model for long videos that determines relative skills level by learnable time-related attention modules. Li et al. introduced a spatial attention-based approach for skills assessment of video data. Authors introduced a new recurrent neural network (RNN) that incorporates high-level progress information of an ongoing task in addition to the stacked attention states from past frames Li et al. (2019).

Inductive methods discussed so far yield a global performance measure of the user for the entire task (i.e., expert, intermediate, or novice labels as illustrated in Fig. 2.2). To provide a more tailored and informative feedback to the users about their surgical performance and skills level (e.g., their mistakes and parts of the task they need to improve their skills), one effective and common approach is to decompose user's movements into blocks called surgemes (van Amsterdam et al. (2021)) and apply state of the art RAS skills evaluation approaches at the sub-task level of the operation. By using this approach, instead of having a global performance metric, a high-resolution surgical workflow will be analyzed, which returns more elaborate feedback about the performance of the different parts of the entire surgical task. Many publications have attempted to perform autonomous analysis of surgical activities in a fine-grained manner Lea et al. (2016); Menegozzo et al. (2019); DiPietro et al. (2016, 2019); Itzkovich et al. (2019); van Amsterdam et al. (2020) and reinforcement learning (RL) Liu and Jiang (2018). These approaches in addition to having the same limitations caused by their black-box nature also suffer from the over-segmentation problem (i.e., producing a large number of false action boundaries) and a poor prediction accuracy that prevents them from making accurate predictions, particularly for unpredicted events (e.g., sudden failures and restarts) van Amsterdam et al. (2021). The over-segmentation problem may arise since high-capacity inductive models (capacity refers to the model's ability to accommodate variations in input data, which is largely dependent upon how many parameters it can learn) mainly focus on the local variations of the data (i.e., small and unimportant details), instead of the global structure of the input trajectory. Furthermore, these methods heavily rely on hand-crafted gesture annotations to evaluate segmentation accuracy, which can take a long time and be subject to human bias. In addition, these approaches segment trajectories without providing any meaningful interpretation about the user's dexterity level and behavior at the sub-task level.

Inspired by the above mentioned limitations, Soleymani et al. proposed an intuitive, explainable, and unsupervised ML-based approach for approximate decomposition of structured surgical trajectories such in retrospective studies Soleymani et al. (2022b). The introduced dual-sparse dictionary learning algorithm decomposes each trajectory into dictionary atoms that captures the main variations of the data which are one general trend and several seasonal patterns. The proposed floating atoms concept is further utilized to accommodate temporal structures within trajectories and preserves information within the trajectory while mapping the data to embedding space. By reconstructing each trajectory according to the generated atoms of the training set (mainly expert trajectories as a benchmark), a vector of codes will be generated that is representative of the data in the low-dimentional embedding space. The code vector conveys important information about the skills level of the user and his/her abnormal behaviors within the task. The propsed approach does not need manual annotation and medicates oversegmentation problem since it captures main variations within the input trajectory and neglects local contents. On the other hand, segmentation borders in the proposed method is not as accurate as other related work with delicate annotated training data sets.

2.2.2 Domain knowledge-based models

While the end-to-end learning approaches presented in Section 2.2.1 have shown acceptable classification accuracies, they are black-box models with opaque decision-making procedures that are incomprehensible even for human experts. As a consequence, it is hard to provide meaningful feedback to the user's surgical performance or intuition about the contributing factors to the surgical outcome (here, intuitiveness and explainability mean how much the function or decision of a model are intuitive and explainable from the perspective of human logic, respectively). Moreover, DL models with large capacity require big data to prevent the final model to become overfitted. Since in the field of robotic surgery reliable, clean , and large data sets are very scarce, DL models have a tendency to overfit. This damages the model's generalization and results in models with poor performance in unpredictable situations (e.g., aborting and restarting a task).

Additionally, for safety-critical applications such as robotic surgery, the human user must understand whether the model is developed based on meaningful features or irrelevant clues and biases in the training set. Therefore, it is crucial to enhance the explainability and interpretability of the model to meet the ethical requirements of skills assessment methods for robotic surgery Molnar (2020). AI models that incorporate domain knowledge not only enhance interpretability and explainability, but they also improve learning performance especially when training data is not large Islam et al. (2021). Utilizing domain knowledge as a prior in data-scarce surgical tasks not only reduces uncertainty about the success of the operation but also makes the model easy-to-learn and more generalizable with smaller training data sets von Rueden et al. (2019). The final skills assessment model that integrates extracted manually engineered features, as shown in Fig. 2.3, provides further clarity to the explanation and interpretation of the calculations since the effect of each extracted feature to the final generated outcome is more transparent.

The domain knowledge-based approach to skills classification incorporates meaningful features as evaluation metrics including execution time (Judkins et al. (2009); Liang et al. (2018)), motion jerk (Liang et al. (2018)), total path length Judkins et al. (2009), etc., and run comparative statistical analysis on a single metric between different participants. Due to the statistical variations between and within participating users, there are significant overlaps between the extracted metrics. As a result, there is no reliable statistical difference amongt participants in terms of sugical skills level. It is primarily because some domain knowledge-based features including but not limited to motion jerk are very noisy for surgical tasks. Other features such as the total path length or task execution time are not informative enough to indicate the true level of skill of the user

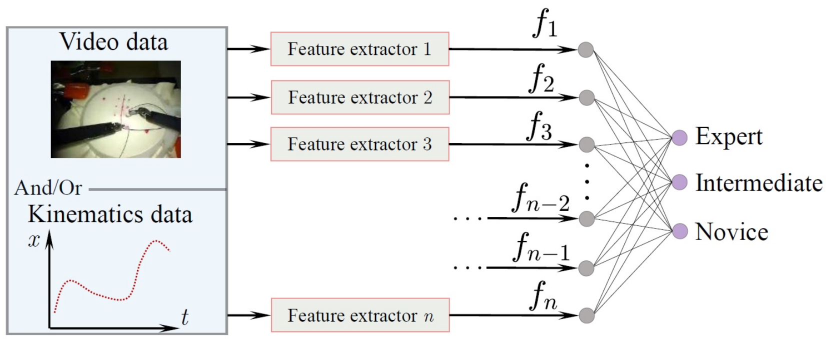

FIGURE 2.3 Domain knowledge-based AI models for surgical skills assessment applications. In this paradigm, the surgical raw data which can be in the format of video and/or kinematics data will be fed into the several manually engineered feature extraction blocks to extract clinically meaningful features f_i that reflect the skills level of a surgeon (e.g., total path length, smoothness, fluidity of the motion, etc.). The number of features (i.e., the value of n) and the structure of feature extraction functions totally depends on the task and the amount of available knowledge in that particular field.

as a single factor. As a result, the methods presented in the aforementioned papers remain case-specific and ungeneralizable for new tasks as they ignore time-related patterns and do not create a universal model for a variety of RAS skills assessment.

Furthermore, domain knowledge-based approaches do not consider detailed events within a given task of an operation; they mostly take into account general metrics over the whole process. Such domain knowledge-based studies, for example, ignore critical time-related features such as trajectory non-smoothness (i.e., presence of random movements including hand tremors or uncontrolled rapid motions) as an important factor to the skills assessment of a given trajectory. The smoothness of a trajectory appears to be one of the key features that can be extracted and utilized for skills assessment purposes. Smoothness evaluation is challenging because non-smoothness is a temporal characteristic that occurs frequently within all trajectories. The result is that the non-structured pattern of smoothness often gets indetectable by other dominant time-domain characteristics including general trend and seasonal patterns. In addition, there is no general and accurate domain knowledge-based approach for searching, detecting, and quantifying snoothness across the entire time series.

Various domain knowledge-based features can be concatenated to create a rich high-dimensional feature space to discover a more expressive and performant representation of RAS trajectories for the sake of skills assessment. Ensembling all of the above-mentioned clinically meaningful features returns an informative long feature vector that meaningfully highlights the skills level of the participant and shows subtle but important information within the surgical trajectory. As depicted in Fig. 2.3, all of these concatenated metrics can be fed into a downstream ML classifier (e.g., support vector machines (SVM) model (Cortes and

Vapnik (1995)), which is much simpler than the sophisticated classifier model introduced in Fig. 2.2. These features can also be fed into dimensionality reduction models (e.g., t-distributed stochastic neighbor embedding (tSNE) Van der Maaten and Hinton (2008)) to visualize the high-dimensional features in a two or three-dimensional map to let the user investigate the internal mechanism of the proposed model.

2.2.3 Domain adapted models

In the aforementioned papers, authors trained a network using an end-to-end learning paradigm or pure field knowledge-based approaches based on a popular dataset (e.g., JIGSAWS data set Gao et al. (2014)). A good approach is to combine inductive and knowledge-based models to develop a domain-adapted model that jointly incorporates both manually engineered metrics and data-driven end-to-end models for more efficient skills assessment purposes (see Fig. 2.4). In this way, thanks to informative features generated from domain knowledge, the end-to-end model does not need to be very complicated compared to inductive learning approaches (i.e., Net_2 in Fig. 2.4 is lighter than Net_1 in Fig. 2.3). Moreover, since a substantial number of informative features are extracted by end-to-end model, there is no need to expend too much time and effort for engineering informative domain knowledge-based features (i.e., the number of f_i in Fig. 2.4 in less than that of Fig. 2.3, or in other words $m < n$).

For example, Soleymani et al. extracted spatio-temporal features in the sequence of RAS video data by incorporating a pre-trained ResNet50 model He et al. (2016). Since the feature extraction network is not trained on the specific data set of surgical trials, it can extract the key features related to the skills level and surgical behaviors of the users. This approach at the crucial and prone-to-bias stage of feature extraction makes the model robust as well as generalizable to unseen test data. Moreover, fast Fourier transforms (FFTs) were used in the proposed method as an extra feature extraction layer to decompose the entire representation learning procedure into two phases: spatial-feature learning and temporal-feature extraction. This model uses FFT to represent a commonly accepted piece of domain knowledge that experts have dominant low-frequency components and negligible high-frequency activities (i.e., they have smooth movements). Compared with experts and intermediate users, novice users have smaller low-frequency coefficients (i.e., they show more hand tremors and unwanted random actions). The intermediate group falls somewhere in between expert and novice behavior. Extracted and manually manipulated features were fed into a downstream CNN to extract and learn features in an inductive learning paradigm to classify the skills level of each input trajectory. Due to their method's two-stage learning, they achieved a less complex model than other arts with sophisticated 3-D CNNs or complicated CNN+RNN models.

The core limitation of the mentioned approaches in this chapter is that the trajectory of each hand is treated as an independent set of data with no considera-

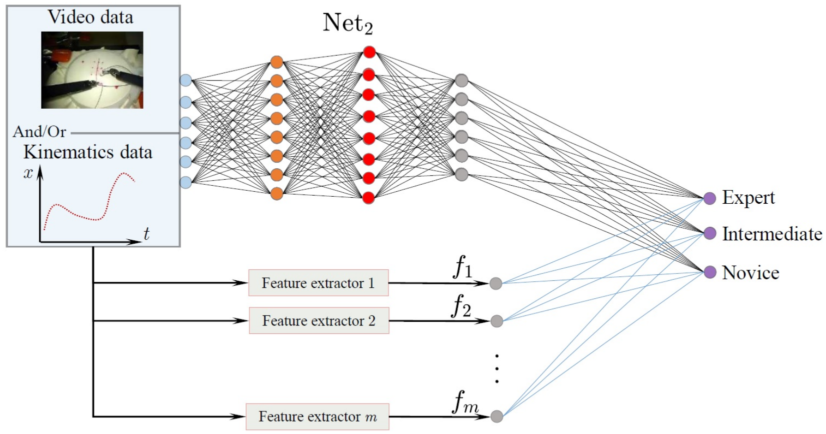

FIGURE 2.4 Domain adapted AI models for surgical skills assessment applications. In this paradigm, the surgical raw data which can be in the format of video and/or kinematics data will be simultaneously fed into a stack of manually engineered feature extraction blocks and inductive learning-based model Net_2 to extract clinically meaningful and data driven features that reflect the skills level of a surgeon. The advantages of such modelling approach are the reduced number of manually made features (i.e., $m < n$) compared to domain knowledge-based model presented in Section 2.2.2 which means less feature engineering effort and light-weighted structure of Net_2 relative to presented Net_1 in Section 2.2.1 which yields less training and implementation complications and low chance of overfitting.

tion given to possible collaboration between two hands (i.e., interaction between different data channels such as roll rotation of the right hand and yaw rotation of the left hand in knot tying task). However, when a surgeon performs sophisticated bimanual tasks (i.e., collaborative tasks requiring delicate coordination of both hands) what defines him/her as an expert is not just the independent performance of each hand but how he/she manages to synchronize the motions and rotations of both hands Vedula et al. (2016). The two hands pulling apart the two ends of the suture to tie a knot is a good example of hands collaboration in surgical operations. It seems considering and measuring the collaboration quality between two hands is a promising area in the field of surgical skills assessment for future work.

Another work which utilizes both data-driven and field knowledge-based models is Soleymani et al. (2022a). In the proposed approach, a data-driven learning process extracts smoothness features from the input data, and clinically approved features such as fluidity and economy of motion are used as well-known domain knowledge-based features to detect the true skill level of the given executive trajectory. The fluidity of movement quantifies how quickly and accurately a RAS task trajectory is executed in transnational or rotational space. Following is a method for calculating this metric which incorporates the time derivative of the input trajectory

$$f_{\text{fluid}} = \left(\frac{1}{T}\int_{t=0}^{T} |\dot{u}(t)| dt\right)^{-1}, \text{ or } \quad f_{\text{fluid}} = \left(\frac{1}{N}\sum_{t=0}^{N} |\dot{u}[t]|\right)^{-1} \tag{2.1}$$

where T is the trajectory execution time of the time series $u(t)$, dot specifies the time derivative, and N is the number of time stamps of discrete trajectory $u[t]$. High values of this metric are returned for quick and accurate trajectories, while low values will be generated for slow, non-accurate, and faulty trajectories and paths with abrupt temporal changes (i.e., task failures and human mistakes). It is generally accepted that the economy of motion contributes to the skills assessment of various activities by reflecting the total energy demand. It is also critical to note that human mistakes such as unintentional motions often have high velocity and large energy injection into the patient-side robot in surgical robotic applications, both of which can lead to dangerous and traumatizing outcomes. It was shown that the kinetic energy of a given trajectory approximates the critical factor of economy of motion. In different configurations of robotic platform in RAS, the total inertia of the patient-side robot remains quite the same, so the economy of motion metric will be calculated as follows:

$$f_{\text{econo}} = \left(\frac{1}{2}\int_{t=0}^{T} \dot{u}^2(t) dt\right)^{-1}, \text{ or } \quad f_{\text{econo}} = \left(\frac{1}{2}\sum_{t=0}^{N} \dot{u}^2[t]\right)^{-1}. \tag{2.2}$$

Smoothness is the most challenging metric. The reason for this is that non-smooth behaviors can occur at any moment in the trajectory and are relatively insignificant compared to the main variations of the data (e.g., general trends and seasonal patterns). The method that Soleymani et al. used in this paper is contrastive principal component analysis or cPCA in short. They created two fabricated data sets: one with smooth trajectories (background data set) and another with non-smooth trajectories (target data set). This study aims to uncover the most notable differences between these sets related to the smoothness of trajectories. At first, the covariance matrices of the target and background data sets are calculated as follows:

$$\boldsymbol{C}_t = \boldsymbol{X}_n \boldsymbol{X}_n^\top, \quad \boldsymbol{C}_b = \boldsymbol{X}_s \boldsymbol{X}_s^\top. \tag{2.3}$$

To highlight the non-smooth behaviors within the target set relative to the background set, the contrastive covariance matrix $\boldsymbol{C}_c$ and its singular value decomposition were calculated as follows

$$\boldsymbol{C}_c = \boldsymbol{C}_t - \alpha \boldsymbol{C}_b = \boldsymbol{W}_c \boldsymbol{\Lambda} \boldsymbol{W}_c^\top \tag{2.4}$$

where hyperparameter α denotes the contrastive strength parameter which represents the importance of target variances versus the irrelevant background variance. Now, the normal PCA can be applied on the projected data. This data-driven approach fully separates smooth and non-smooth trajectories and can be used together with two other knowledge-based methods for skills evaluation purposes. As highlighted in Soleymani et al. (2022a), such approach reveals label-free information within RAS trajectories and provides more reliable and tangible correction hints to the users. Unfortunately, such hints in inductive models even with high accuracies cannot be fully trusted due to lack of explainability of extracted features. One elegant instance of label-free information in this paper is classifying one intermediate surgeon close to expert surgeons in JIGSAWS data set. The high global rating score assigned to this surgeon proves the fact that this surgeon performs expertly compared to other participants.

2.3 SURGICAL SKILLS TRANSFER

Modern robotic surgery systems allow the application of haptic guidance forces to a trainee's hands in order to correct their motion with the aim of improving performance. In haptics and telerobotics, there is a rich literature dealing with therelationship between surgical mentor and the trainee from the perspective of *expert-in-the-loop* and haptics-enabled training Shahbazi et al. (2016); Sharifi et al. (2017); Tao et al. (2020); Rossa et al. (2021); Najafi et al. (2020); Zakerimanesh et al. (2019); Sharifi et al. (2020); Shahbazi et al. (2016); Atashzar et al. (2018). As an example, Shamaei et al. incorporated a trilateral shared control architecture between two users (one mentor and one trainee) for RAS skills training Shamaei et al. (2015). The platform is composed of one robot located at the patient-side that is simultaneously manipulated by two different user-side robots, one guided by the surgical mentor and one is used by the trainee. The authority level of the surgical mentor over the actions of the RAS trainee is determined by the dominance factor hyperparameter. This training program requires continuous supervision by an expert surgeon. Using a smart DL-based approach for RAS training can provide more opportunities and peace of mind to enable surgical trainees to practice surgery in a safe environment while receiving haptic feedback from a mentor.

In previous sections, it was discussed how incorporating rich skills-related knowledge into RAS training platforms enhances the training quality and reduces the need for the intervention of an expert surgical mentor throughout the entire process. AI can be used to overcome the challenges of human-robot interaction (HRI) and transfer mentorship experiences to trainees by engaging them in a collaborative action with the robot. This intelligent mentorship is even more crucial in complicated tasks which require long training procedure (more than one session) to become skillful Sigrist et al. (2013). A framework designed by Ershad et al., with inspiration from AI-powered surgical training technologies, detects trainees' flawful stylistic behaviors and provides haptic feedback for translational errors in a near-real-time manner Ershad et al. (2021). However, the study offers no suggestions for improving an individual user's performance.

Tan et al. (2019) proposed a laparoscopic robotic platform that utilizes both human demonstrations and reinforcement learning to teach surgical trainees to better manipulate the robotic tool. Tan et al. locally stored expert trajectories in a field-programmable gate array (FPGA) to replay and regenerate an agent for generative adversarial imitation learning. During the training stage, the novice trainee holds the surgical handle and feels the velocity and force patterns and memorizes the skillful translational and rotational trends for better performing the task in the future executions. The limitation of the proposed method is its lack of generalization to new trajectories that the user may want to execute in future applications.

The method presented by Zahedi et al. (2020) for a virtual kinesthetic teaching environment uses machine learning to aim the transfer of skills between mentor and trainee. In the training phase, mentor's demonstrations produce a map specifying the stiffness variations of various bone layers. An estimator model of motion similarity measures how similar a trainee's drilling motion pattern is to a mentor's pattern at different layers of the bone. Trainees are provided with a set of assisting and resisting forces to correct their deficient stylistic behavior while performing operational motions. The generated corrective force is proportional to the similarity of the novice trajectory and the recorded expert demonstration. It is beyond dispute that the resultant model for the given task cannot be generalized to other RAS tasks. Similar to Zahedi et al. (2020), another platform was developed by Fekri et al. (2021) to train novice users for orthopedic surgical drilling task. In Fekri et al. (2021), an ordinary RNN with a LSTM architecture was incorporated to create the model of an expert surgeon which generates a reference trajectory based on the captured demonstrations for guiding a novice trainee towards a better stylistic behavior.

Reinforcement learning, learning from demonstration, and imitation learning are commonly used approaches for transferring expert mentor skills to a robot. As an example, Chi et al. (2020) and Tan et al. (2019) used a model-free generative adversarial imitation learning approach in conjunction with a deep reinforcement learning model to learn and imitate the skillful behavior within a minimally invasive surgery task with unfamiliar dynamics. Some studies incorporate learning from demonstration (LfD) to adaptively mimic complicated surgical trajectories in various circumstances and then, plan a new online path for tracking in an environment with uncertain and unpredictable factors Osa et al. (2014). The goal of reinforcement learning is to capture the contributing skills-related features of a trajectory by using task-specific reward and/or regret functions. As a result, the implementation of these methods becomes difficult or even impossible for similar tasks. In addition, some reward functions such as the completion of a task are based on the completion of the whole process, making it impossible to provide the user with online and fine-grained feedback while the operation is being performed. In addition, it is difficult to design comprehensive, relevant, and meaningful reward/regret functions in complex tasks. Finally, there is always a risk of a distribution mismatch happening due to multimodal behaviors of the user's demonstrations since there are tons of possibilities involved in surgical tasks.

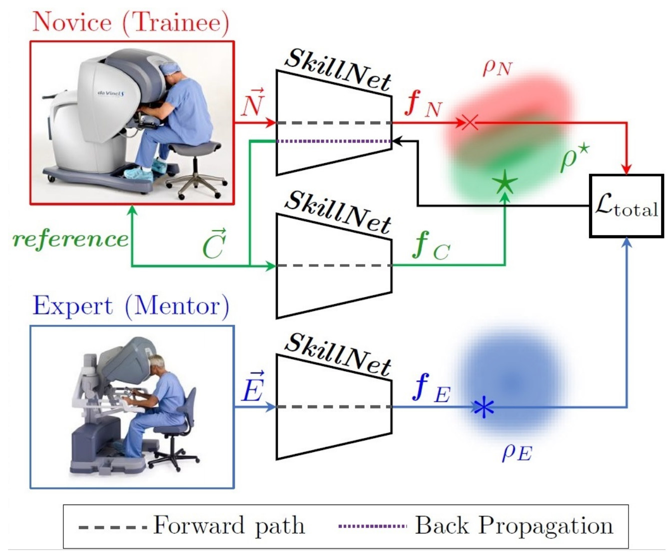

FIGURE 2.5 The schematic procedure of skills transfer algorithm presented in Soleymani et al. (2021) . First of all, the SkillNet extracts f_N and f_E that is sampled from novice and expert probability distributions ρ_N and ρ_E, respectively. After calculating intension and skill losses, the error is backpropagated through SkillNet to update the novice trajectory $\vec{N}$ to generate optimized trajectory $\vec{C}$ sampled from the probability distribution $\rho^\star$ which is similar to ρ_E than ρ_N. The generated enhanced combined trajectory $\vec{C}$ will be used as a reference for collaborative robots to apply corrective haptic forces to the trainee's hands. Used with permission of authors of Soleymani et al. (2021).

Taking into account the limitations outlined in the aforementioned papers and thanks to recent advances in DL, Soleymani et al. (2021) utilized artificial intelligence to detect the skillful behaviors of surgical experts and inject them into surgical trainees' activities. As a result of this, the human-robot collaboration will be controlled in a more skillful and dexterous manner (i.e., the novice user is performing the surgical operation on his/her own). The authors proposed a deep model named SkillNet to extract skills-related attributes from raw da Vinci kinematics data within each 20 seconds interval which allows SkillNet to operate in real-time skills evaluation and transfer tasks. The skills transfer algorithm constantly references the mentor's skillful features to generate the desired trajectory which the trainee user's trajectory should follow for the sake of better performance. This architecture design is partially inspired by image style transfer work in the field of computer vision Gatys et al. (2016). The final objective of the surgical skills transfer algorithm is constructing an optimized trajectory $\vec{C}$ initialized by the novice demonstration $\vec{N}$ such that $\vec{C}$ represents mentor's skillful behaviors with the lowest divergence compared to the initialization $\vec{N}$. In this way, the probability distribution of the novice trajectory will approach to that of the expert trajectory in feature space (see Fig. 2.5). In order

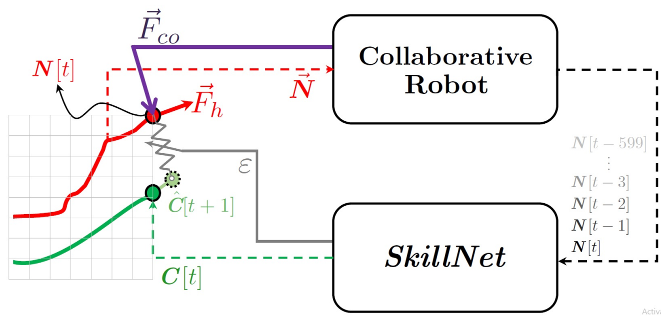

FIGURE 2.6 The architecture of RAS skill transfer. The participant's hand (or RAS tooltip) is virtually connected to one end of the spring and SkillNet exerts a corrective force on the other end to guide it to the enhanced trajectory (green solid line). With this method, the initialized trajectory (red solid line) is guided toward more dexterous stylistic behavior by providing mild and compliant guidance forces. The strength of the correction force $\boldsymbol{F_{co}}$ will be determined via the control gain $\boldsymbol{K}$ and the performance of the trainee in the past. Used with permission of authors of Soleymani et al. (2021).

to achieve this goal, the following losses need to be minimized: the *skill loss* (i.e., the norm-2 difference between the latent feature distributions of experts and given novice trajectory) and the *intention loss* (i.e., the reconstruction loss between initialization and final optimized trajectory) defined as below:

$$\mathcal{L}_{\text{skill}}(\vec{E},\vec{C}) = ||\mathcal{G}_{\boldsymbol{f_E}} - \mathcal{G}_{\boldsymbol{f_C}}||_2. \tag{2.5}$$

$$\mathcal{L}_{\text{intention}}(\vec{C},\vec{N}) = ||\vec{C} - \vec{N}||_2. \tag{2.6}$$

where for instance $\mathcal{G}_{\boldsymbol{F}}$ is the Gram matrix of the feature vector F. In this paradigm, skills transformation from an expert trajectory into that of novice trainee simply means generating an optimized trajectory based on minimizing the weighted linear combination of intention and skill losses, i.e., minimizing total loss

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{intention}}(\vec{C},\vec{N}) + \beta\mathcal{L}_{\text{skill}}(\vec{E},\vec{C}) \tag{2.7}$$

via gradient descent method where α and β are hyperparameters indicating relative importance of the intention loss versus skill loss in the optimization process.

The presented approach does not impose any restrictions on trainee activities or require field knowledge about the human user, robotic setup, or task. Due to the mentioned properties, the approach can be applied to a variety of robotic platforms and applications. SkillNet transfers skillful attributes to the novice demonstration $\vec{N}$ in real-time and returns $\vec{C}$ as well as the trainee's current performance, ε (see Fig. 2.6). $C[t]$ and $C[t-1]$ represent the last two points of $\vec{C}$ that are used to predict the next point

$$\hat{\boldsymbol{C}}[t+1] \approx 2\boldsymbol{C}[t] - \boldsymbol{C}[t-1] = \boldsymbol{C}[t] + \frac{\Delta T(\boldsymbol{C}[t] - \boldsymbol{C}[t-1])}{\Delta T}. \tag{2.8}$$

Lastly, as shown in Fig. 2.6, the collaborative robot generates a mild correction force F_{co} to user's hand to control him/her towards $\hat{C}[t+1]$ point using variable impedance control method

$$\vec{F}_{co} = \varepsilon \boldsymbol{K}(\hat{\boldsymbol{C}}[t+1] - \boldsymbol{N}[t]) \tag{2.9}$$

where $\boldsymbol{K} = \mathrm{diag}(k_x, k_y, k_z)$ is the virtual compliance coefficients matrix in the Cartesian coordinate system. The skill transfer algorithm presented in Soleymani et al. (2021) makes significant improvements over novice trajectories, makes them more predictable, reduces hand tremors, and cancels signal noise which are all clinically proven factors for RAS skills assessment.

2.4 FUTURE DIRECTIONS

The core limitation of the mentioned approaches in both skills assessment and transfer is that the trajectory of each hand is viewed as a set of independent data without taking into account potential collaboration between hands (i.e., interaction between different data channels such as vertical displacement of the right hand and that of the left hand in a bimanual lifting task). The true expert in executing complicated bimanual tasks (tasks that require delicate collaboration between both hands), such as surgical operations, is not the one who performs operations with each hand expertly, but the one who plans for future steps and executes a complicated series of correlated movements by using both hands together. The clinically proven relationship between hand coordination (cooperative relationship) and correspondence (lead-follower relationship) in bimanual task performance has not yet been quantitatively assessed in the evaluation of surgical skills. The topic seems to be an excellent direction for future research.

Another direction that still remains open is dexterous autonomous robotic surgery. The results of all mentioned surgical skills assessment and transfer methods can be incorporated into the reward-shaping procedure of training a smart agent developed by reinforcement learning methods or learning from demonstration paradigm for performing skillful surgical robotics tasks. Beyond ethics and regulatory problems, several issues including but not limited to the high dimensionality of demonstrated trajectories, inconsistency, and minor human error even in expert demonstrations are potential challenges that need to be addressed. It is beyond dispute that a proper trajectory encoding method benefits the tractability of the exploration-exploitation procedure of training a skillful agent.

2.5 CONCLUSIONS

In this chapter, a variety of approaches, their advantages, and limitations for extracting the skills-related features of a participant operating on RAS systems were presented. It also explained how one can use these features to classify the level of expertise of the participant and transfer them to a less skillful trajectory to help and teach less-experienced users to perform better in surgical operations. It also explained how to create an optimized trajectory with minimal reconstruction loss compared to the initialized novice trajectory while having more skillful features. The optimized trajectory can be used as a control reference command to generate a virtual corrective force on the RAS platform and guide the participant's hand toward more dexterous stylistic behavior. The enhancement metrics over the trainee's trajectory were introduced to measure the functionality and performance of the skills assessment algorithm. These metrics are including but are not limited to the motion predictability, reduction in hand tremor, and noise cancellation.

Ahmidi, N., Tao, L., Sefati, S., Gao, Y., Lea, C., Haro, B.B., Zappella, L., Khudanpur, S., Vidal, R., Hager, G.D., 2017. A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery. IEEE Transactions on Biomedical Engineering 64, 2025–2041.

van Amsterdam, B., Clarkson, M., Stoyanov, D., 2021. Gesture recognition in robotic surgery: a review. IEEE Transactions on Biomedical Engineering .

van Amsterdam, B., Clarkson, M.J., Stoyanov, D., 2020. Multi-task recurrent neural network for surgical gesture recognition and progress prediction, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 1380–1386.

Atashzar, S.F., Shahbazi, M., Tavakoli, M., Patel, R.V., 2018. A computational-model-based study of supervised haptics-enabled therapist-in-the-loop training for upper-limb poststroke robotic rehabilitation. IEEE/ASME Transactions on Mechatronics 23, 563–574.

Birkmeyer, J.D., Finks, J.F., O'Reilly, A., Oerline, M., Carlin, A.M., Nunn, A.R., Dimick, J., Banerjee, M., Birkmeyer, N.J., 2013. Surgical skill and complication rates after bariatric surgery. New England Journal of Medicine 369, 1434–1442.

Chi, W., Dagnino, G., Kwok, T.M., Nguyen, A., Kundrat, D., Abdelaziz, M.E., Riga, C., Bicknell, C., Yang, G.Z., 2020. Collaborative robot-assisted endovascular catheterization with generative adversarial imitation learning, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 2414–2420.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20, 273–297.

DiPietro, R., Ahmidi, N., Malpani, A., Waldram, M., Lee, G.I., Lee, M.R., Vedula, S.S., Hager, G.D., 2019. Segmenting and classifying activities in robot-assisted surgery with recurrent neural networks. International journal of computer assisted radiology and surgery 14, 2005–2020.

DiPietro, R., Lea, C., Malpani, A., Ahmidi, N., Vedula, S.S., Lee, G.I., Lee, M.R., Hager, G.D., 2016. Recognizing surgical activities with recurrent neural networks, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 551–558.

Doughty, H., Damen, D., Mayol-Cuevas, W., 2018. Who's better? who's best? pairwise deep ranking for skill determination, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6057–6066.

Doughty, H., Mayol-Cuevas, W., Damen, D., 2019. The pros and cons: Rank-aware temporal attention for skill determination in long videos, in: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7862–7871.

Ershad, M., Rege, R., Fey, A.M., 2021. Adaptive surgical robotic training using real-time stylistic behavior feedback through haptic cues. IEEE Transactions on Medical Robotics and Bionics 3, 959–969.

Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A., 2019. Accurate and interpretable evaluation of surgical skills from kinematic data using fully convolutional neural networks. International journal of computer assisted radiology and surgery 14, 1611–1617.

Fekri, P., Dargahi, J., Zadeh, M., 2021. Deep learning-based haptic guidance for surgical skills transfer. Frontiers in Robotics and AI 7, 586707.

Funke, I., Mees, S.T., Weitz, J., Speidel, S., 2019. Video-based surgical skill assessment using 3d convolutional neural networks. International journal of computer assisted radiology and surgery 14, 1217–1225.

Gao, Y., Vedula, S.S., Reiley, C.E., Ahmidi, N., Varadarajan, B., Lin, H.C., Tao, L., Zappella, L., Béjar, B., Yuh, D.D., et al., 2014. Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling, in: Miccai workshop: M2cai, p. 3.

Gatys, L.A., Ecker, A.S., Bethge, M., 2016. Image style transfer using convolutional neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2414–2423.

Goh, A.C., Goldfarb, D.W., Sander, J.C., Miles, B.J., Dunkin, B.J., 2012. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. The Journal of urology 187, 247–252.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Islam, S.R., Eberle, W., Ghafoor, S.K., Ahmed, M., 2021. Explainable artificial intelligence approaches: A survey. arXiv preprint arXiv:2101.09429 .

Itzkovich, D., Sharon, Y., Jarc, A., Refaely, Y., Nisky, I., 2019. Using augmentation to improve the robustness to rotation of deep learning segmentation in robotic-assisted surgical data, in: 2019 International Conference on Robotics and Automation (ICRA), IEEE. pp. 5068–5075.

Jian, Z., Yue, W., Wu, Q., Li, W., Wang, Z., Lam, V., 2020. Multitask learning for video-based surgical skill assessment, in: 2020 Digital Image Computing: Techniques and Applications (DICTA), IEEE. pp. 1–8.

Judkins, T.N., Oleynikov, D., Stergiou, N., 2009. Objective evaluation of expert and novice performance during robotic surgical training tasks. Surgical endoscopy 23, 590–597.

Kim, T.S., O'Brien, M., Zafar, S., Hager, G.D., Sikder, S., Vedula, S.S., 2019. Objective assessment of intraoperative technical skill in capsulorhexis using videos of cataract surgery. International journal of computer assisted radiology and surgery 14, 1097–1105.

Lea, C., Vidal, R., Reiter, A., Hager, G.D., 2016. Temporal convolutional networks: A unified approach to action segmentation, in: European Conference on Computer Vision, Springer. pp. 47–54.

Li, Z., Huang, Y., Cai, M., Sato, Y., 2019. Manipulation-skill assessment from videos with spatial attention network, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 0–0.

Liang, K., Xing, Y., Li, J., Wang, S., Li, A., Li, J., 2018. Motion control skill assessment based on kinematic analysis of robotic end-effector movements. The International Journal of Medical Robotics and Computer Assisted Surgery 14, e1845.

Liu, D., Jiang, T., 2018. Deep reinforcement learning for surgical gesture segmentation and classification, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 247–255.

Liu, D., Jiang, T., Wang, Y., Miao, R., Shan, F., Li, Z., 2019. Surgical skill assessment on in-vivo clinical data via the clearness of operating field, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 476–484.

Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. Journal of machine learning research 9.

MacKenzie, L., Ibbotson, J., Cao, C., Lomax, A., 2001. Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. Minimally Invasive Therapy & Allied Technologies 10, 121–127.

Martin, J., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C., Brown, M., 1997. Objective structured assessment of technical skill (osats) for surgical residents. British journal of surgery 84, 273–278.

Menegozzo, G., Dall'Alba, D., Zandonà, C., Fiorini, P., 2019. Surgical gesture recognition with time delay neural network based on kinematic data, in: 2019 International Symposium on Medical Robotics (ISMR), IEEE. pp. 1–7.

Molnar, C., 2020. Interpretable machine learning. Lulu. com.

Muralidhar, N., Islam, M.R., Marwah, M., Karpatne, A., Ramakrishnan, N., 2018. Incorporating prior domain knowledge into deep neural networks, in: 2018 IEEE international conference on big data (big data), IEEE. pp. 36–45.

Najafi, M., Rossa, C., Adams, K., Tavakoli, M., 2020. Using potential field function with a velocity field controller to learn and reproduce the therapist's assistance in robot-assisted rehabilitation. IEEE/ASME Transactions on Mechatronics 25, 1622–1633.

Nguyen, X.A., Ljuhar, D., Pacilli, M., Nataraja, R.M., Chauhan, S., 2019. Surgical skill levels: Classification and analysis using deep neural network model and motion signals. Computer methods and programs in biomedicine 177, 1–8.

Osa, T., Sugita, N., Mitsuishi, M., 2014. Online trajectory planning in dynamic environments for surgical task automation., in: Robotics: Science and Systems, pp. 1–9.

Reiley, C.E., Hager, G.D., 2009. Decomposition of robotic surgical tasks: an analysis of subtasks and their correlation to skill, in: M2CAI workshop. MICCAI, London.

Rosen, J., Hannaford, B., Richards, C.G., Sinanan, M.N., 2001. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE transactions on Biomedical Engineering 48, 579–591.

Rosen, J., Solazzo, M., Hannaford, B., Sinanan, M., 2002. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden markov model. Computer Aided Surgery 7, 49–61.

Rossa, C., Najafi, M., Tavakoli, M., Adams, K., 2021. Robotic rehabilitation and assistance for individuals with movement disorders based on a kinematic model of the upper limb. IEEE Transactions on Medical Robotics and Bionics 3, 190–203.

von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., et al., 2019. Informed machine learning–a taxonomy and survey of integrating knowledge into learning systems. arXiv preprint arXiv:1903.12394 .

Shahbazi, M., Atashzar, S.F., Tavakoli, M., Patel, R.V., 2016. Robotics-assisted mirror rehabilitation therapy: a therapist-in-the-loop assist-as-needed architecture. IEEE/ASME Transactions on Mechatronics 21, 1954–1965.

Shamaei, K., Kim, L.H., Okamura, A.M., 2015. Design and evaluation of a trilateral shared-control architecture for teleoperated training robots, in: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE. pp. 4887–4893.

Sharifi, I., Talebi, H.A., Patel, R.R., Tavakoli, M., 2020. Multi-lateral teleoperation based on multi-agent framework: Application to simultaneous training and therapy in telerehabilitation. Frontiers in Robotics and AI 7, 538347.

Sharifi, M., Salarieh, H., Behzadipour, S., Tavakoli, M., 2017. Tele-echography of moving organs using an impedance-controlled telerobotic system. Mechatronics 45, 60–70.

Sigrist, R., Rauter, G., Riener, R., Wolf, P., 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. Psychonomic bulletin & review 20, 21–53.

Soleymani, A., Li, X., Tavakoli, M., 2021. Deep neural skill assessment and transfer: Application to robotic surgery training, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE. pp. 8822–8829.

Soleymani, A., Li, X., Tavakoli, M., 2022a. A domain-adapted machine learning approach for visual evaluation and interpretation of robot-assisted surgery skills. IEEE Robotics and Automation Letters 7, 8202–8208.

Soleymani, A., et al., 2022b. Surgical procedure understanding, evaluation, and interpretation: A dictionary factorization approach. IEEE Transactions on Medical Robotics and Bionics .

Tan, X., Chng, C.B., Su, Y., Lim, K.B., Chui, C.K., 2019. Robot-assisted training in laparoscopy using deep reinforcement learning. IEEE Robotics and Automation Letters 4, 485–492.

Tao, L., Elhamifar, E., Khudanpur, S., Hager, G.D., Vidal, R., 2012. Sparse hidden markov models for surgical gesture classification and skill evaluation, in: International conference on information processing in computer-assisted interventions, Springer. pp. 167–177.

Tao, R., Ocampo, R., Fong, J., Soleymani, A., Tavakoli, M., 2020. Modeling and emulating a physiotherapist's role in robot-assisted rehabilitation. Advanced Intelligent Systems 2, 1900181.

Vedula, S.S., Malpani, A.O., Tao, L., Chen, G., Gao, Y., Poddar, P., Ahmidi, N., Paxton, C., Vidal, R., Khudanpur, S., et al., 2016. Analysis of the structure of surgical activity for a suturing and knot-tying task. PloS one 11, e0149174.

Wang, Z., Fey, A.M., 2018. Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. International journal of computer assisted radiology and surgery 13, 1959–1970.

Zahedi, E., Khosravian, F., Wang, W., Armand, M., Dargahi, J., Zadeh, M., 2020. Towards skill transfer via learning-based guidance in human-robot interaction: an application to orthopaedic surgical drilling skill. Journal of Intelligent & Robotic Systems 98, 667–678.

Zakerimanesh, A., Hashemzadeh, F., Torabi, A., Tavakoli, M., 2019. A cooperative paradigm for task-space control of multilateral nonlinear teleoperation with bounded inputs and time-varying delays. Mechatronics 62, 102255.