Robot-Assisted Minimally Invasive Surgical Skill Assessment—Manual and Automated Platforms

Renáta Nagyné Elek[1] and Tamás Haidegger[1,2]

[1]Antal Bejczy Center for Intelligent Robotics, University Research, Innovation and Service Center, Óbuda University, Bécsi út 96/b, 1034 Budapest, Hungary, renata.elek@irob.uni-obuda.hu
[2]Austrian Center for Medical Innovation and Technology, Viktor Kaplan-Straße 2/1, 2700 Wiener Neustadt, Austria, haidegger@irob.uni-obuda.hu

Abstract: The practice of Robot-Assisted Minimally Invasive Surgery (RAMIS) requires extensive skills from the human surgeons due to the special input device control, such as moving the surgical instruments, use of buttons, knobs, foot pedals and so. The global popularity of RAMIS created the need to objectively assess surgical skills, not just for quality assurance reasons, but for training feedback as well. Nowadays, there is still no routine surgical skill assessment happening during RAMIS training and education in the clinical practice. In this paper, a review of the manual and automated RAMIS skill assessment techniques is provided, focusing on their general applicability, robustness and clinical relevance.

Keywords: Robot-Assisted Minimally Invasive Surgery; surgical robotics; surgical skill training; surgical skill assessment

1   Introduction

Minimally Invasive Surgery (MIS) has shown to improve the outcome of specific types of surgeries, due to fact that the operator reaches the organs in interest through small skin incisions. This results in less pain, quicker recovery time and smaller scars on the patient. While the benefits of MIS for the patient are clear, this technique is definitely hard to master for the clinician. To perform traditional MIS, surgeons have to learn the handling of the specific surgical instruments, the manipulation of the endoscopic camera (or coordination on that with the assistance), they have to operate in ergonomically sub-optimal postures [1–4].

To answer these challenges, the concept of Robot-Assisted Minimally Invasive Surgery (RAMIS) was introduced almost four decades ago. To increase ergonomy, robotic systems typically offer a 3D vision system, and their instruments are easier to control than traditional MIS tools. Furthermore, due to the instruments' rescaled movements or special design, RAMIS can be more accurate than

traditional MIS. During the relatively short history of RAMIS, da Vinci Surgical System (Intuitive Surgical Inc., Sunnyvale, CA) emerged to be the dominating surgical robot on the market. The da Vinci is a teleoperated system, where the surgeon sits at a master console, and the patient-side robot copies the motions of the surgeon within the patient. There are more than 5500 da Vinci Surgical System in clinical practice at the moment, and around a million procedures performed in the world yearly [3, 5].

While the development of RAMIS was a bold step forward in modern medicine to help surgeons to realize MIS, it is still a complicated, evolving technique to learn. In the early years, there has been strong criticism that the da Vinci is not providing the overall benefit, claimed [6–8]. The lack of training of robotic surgeons had a great impact in this opinion. Intuitive and the whole research community developed new training platforms to answer these challenges. These have become the first authentic source of data to develop and validate skill assessment methods.

In the research of RAMIS skill assessment, da Vinci Application Programming Interface (da Vinci API, Intuitive Surgical Inc.) was the first source of surgical data, but it was read-only and not accessible widely. With the development of the da Vinci Research Kit (DVRK), the data collection from the da Vinci Surgical System became available for the researchers as well [9]. More recently, Intuitive teamed up with InTouch Health to create a safe telecommunication network for its robot fleet deployed at US hospitals [10]. They extended the cooperation under the concept of Internet of Medical Things [11]. With this collaboration Intuitive is creating the technical possibility to see and assess the performance of its robots and their users.

RAMIS can be learned by surgeons, which process is often represented by learning curves. Learning curve is a graph, where the experience is represented graphically (e.g., time to complete compared to training times). Basically, there are two main approaches of surgical robotics training: patient side and master console training. Patient side training contains the patient positioning and port placement and basic laparoscopic skills (such as creation of pneumoperitoneum, application of clips etc.). Console training involves the handling of the master arms, the camera and the pedals, and cognitive tasks as well. There are lots of console training methods for RAMIS, which can provide the required practice for the surgeon [12]:

- virtual reality simulators;
- dry lab training;
- wet lab training;
- training in the operating room with a mentor.

Each has its own advantage and disadvantage, but from the clinical applicability point of view, the most important question is how fairly do these assess surgical skills. Nowadays, there is still no objective surgical skill assessment method used in the operating room (OR) beyond board examination more experienced

surgeons may provide some feedback, but rarely quantify the skills of their colleagues.

It may be important to evaluate surgical skills for quality assurance reasons, when that becomes part of the hospital's quality management system. More commonly, only the proof of participation at theoretical and practical training is required. Arguably, objective feedback could assist trainees and practicing surgeons as well in improving their skills along the carrier. The fundamental challenge with skill assessment is that traditionally, the patient outcome used to be the only objective metric, and given the amazing variety and individual characteristic of each procedure, it has been really hard to derive distinguishing skill parameters. The subjective evaluation provided by other experts did not make it easy to compare results and metrics, therefore more generally agreed, standardized evaluation practices and training platforms had to be developed. A good example for this is the Fundamentals of Laparoscopic Surgery (FLS), a training and assessment method developed by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) in 1997, and widely adapted: it measures the manual skills and dexterity of an MIS surgeon, and provides a comparable scoring [13]. A similar metric for RAMIS surgeons recently introduced, called Fundamentals of Robotic Surgery (FRS) [14].

In general, to understand the notions of 'skill' and 'skill assessment', let us consider the Dreyfus model [15]. The Dreyfus model refers to the evolution of the learning process, and it describes the typical features of the expertise levels at various phases (Fig. 1). For example, a novice (in general) can only follow simple instructions, but an expert can well react to previously unseen situations. In the literature, we can find other skill models, such as the classic Rasmussen model, which was created for modeling skill-, rule-, and knowledge-based performance levels [16]. An other approach for modeling skills is recently created by Azari et al., which is specifically made for surgical skills (Fig.2) [17]. RAMIS provides a unique platform to measure parameters which can help us in defining these skill levels objectively, since it makes low level motion data and spatial information available. Now, the problem is to find the proper parameters and algorithms that define the surgical skills [18].

In this paper, we review the main approaches to RAMIS skill assessment from manual to fully automated, focusing on the platforms aiming to achieve wider acceptance. Beyond the technical RAMIS skill assessment, we collect the existing approaches to non-technical RAMIS skill assessment as well. The main techniques employed are presented in every cited case, along with the estimated impact of them.

2   Methods

To find relevant publications in the field of manual and automated skill assessment in RAMIS, we used PubMed and Google Scholar databases. The last search performed on December in 2018. This paper is mainly focusing on automated ap-

Figure 1

Dreyfus model of skill acquisition. It defines 5 expertise levels and shows the differences between their qualities [19]
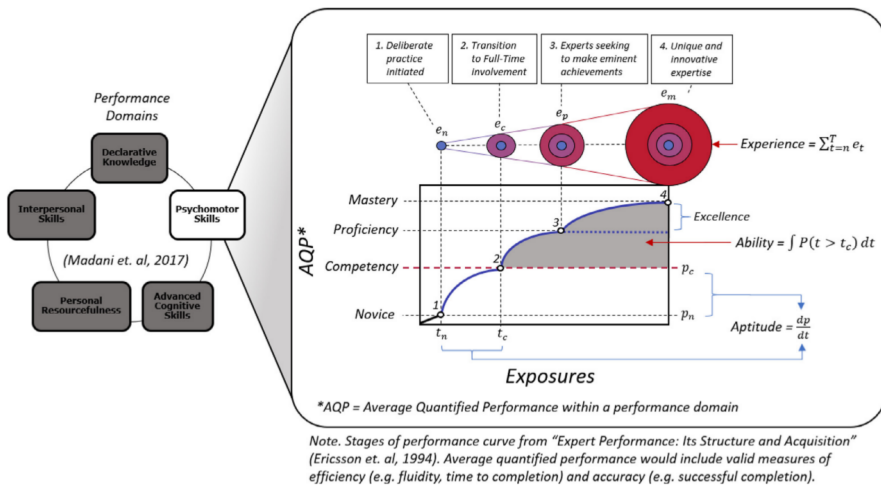

Figure 2

Quantified performance model for surgical skill performance. The model describes the terms of 'skill': experience, excellence, ability and aptitude [17]

proaches, thus training systems and manual techniques are only introduced. To find relevant publications for manual techniques, we used the keywords 'surgical robotics' and 'manual skill assessment' or 'manual skill evaluation'. From the identified publications, we chose 23 based on the relevance and citation index. In the case of virtual reality simulators, we use the keywords 'surgical robotics' and 'virtual reality' and 'training' or 'simulator'. We chose 8 publications to introduce this topic. To find publications for automated approaches and data collection, we used the keywords 'surgical robotics' and 'automated' and 'skill assessment' or 'skill evaluation', or in case of data collection 'surgical robotics' and 'data collection'. We found 47 relevant publications, and the automated techniques are summarized in Table 1. The table has the following columns:

- 'Aim': summarizes the goals of the cited paper;

- 'Input data': used type of data for the skill assessment; algorithm

- 'Data collection': sensor type, data collector device;

- 'Training task': suturing, knot-tying, etc.;

- 'Technique': used algorithms;

and the year of the publication with the reference. Finally, we introduce non-technical skill assessment techniques. For this, we used 12 relevant publications based on the keywords 'surgical robotics' and 'non-technical skill', or 'physiological symptoms' and 'stress'.

3　Manual assessment

In the case of manual RAMIS skill assessment, just like with traditional MIS, a team of expert surgeons in the OR (or post-operatively) evaluates the execution of the intervention based on their knowledge, the specific OR workflow and the expected outcome. This approach is easy to implement, yet very costly (in terms of human resource effort). It may be accurate averaged over multiple reviewers, but each individual assessment is quite subjective across boards, and it may be heavily distorted by personal opinions and influenced by the level of expertise of that particular domain. The types of objective manual surgical skill evaluation in the case of RAMIS are *generic, procedure-specific* and *error-based* [20]. The simplest approach is the error-based manual assessment, because it only requires a typical error detection during the procedures. Procedure specific techniques examine the skills what needed in specific interventions. Generic manual skill assessment is the most complex approach; it evaluate the global skills of the surgeons.

A typical approach of manual RAMIS skill assessment is not to quantify the overall skills, just to evaluate particular skills needed in specific procedures, or only measure the errors made during the execution. In many cases, procedure-specific assessment is required, where the assessment metric is created for a specific surgical procedure (such as cholecystectomy, radical prostatectomy, etc.). Prostatectomy Assessment and Competence Evaluation (PACE) scoring is created for robot-assisted radical prostatectomy skill assessment. PACE metric includes the following evaluation points [21]:

- bladder drop;

- preparation of the prostate;

- bladder neck dissection;

- dissection of the seminal vesicles;

- preparation of the neuro-vascular bundle;

- apical dissection, anastomosis.

Cystectomy Assessment and Surgical Evaluation (CASE) is for robot-assisted radical cystectomy procedures. CASE evaluates the skills based on eight main domains [22]:

- pelvic lymph node dissection;

- development of the peri-ureteral space;

- lateral pelvic space;

- anterior rectal space;

- control of the vascular pedicle;

- anterior vesical space;

- control of the dorsal venous complex;

- apical dissection.

In the case of PACE and CASE, surgical proficiency was represented in every domain on a 5-point Likert scale, where 1 means the lowest and 5 means the highest performance (the score meaning is defined in every domain, such as injuries). Beyond these two specific methods, we can find further scoring metrics for other interventions in the literature [23, 24].

For the above scoring methods refer to the execution of the procedure. In most of the cases, any damage caused reflects the skills of the surgeons retrospectively: such as blood loss, tissue damage, etc. Generic Error Rating Tool (GERT) is a framework to measure technical errors during MIS; it was specifically created for gynecologic laparoscopy [25]. The validation tests showed promising results for the usability of GERT for objective skill assessment (its correlation to OSATS was examined) [26].

Generic manual assessment techniques evaluate the skills, based on the whole procedure/training technique, considering several points of the surgery, but not considering a specific technique. Global Evaluative Assessment of Robotic Skills (GEARS) was particularly created for robotic surgery, where expert surgeons assess the operator's robotic surgical skills manually. GEARS metric involves the assessment of the followings [12]:

- depth perception (from overshooting target to accurate directions to the right plane);

- bimanual dexterity (one from hand usage to using both hands in a complementary way);

- efficiency (from inefficient efforts to fluid and efficient progression);

- force sensitivity (from injuring nearby structures to negligible injuries);

- robotic control skills (based on camera and hand positions).

The surgical experts score the performance on a five scale score system. GEARS is a well-studied metric: we can find validity tests and comparisons with GEARS in the literature [12, 27–37]. The original paper of GEARS showed results for the clinical usability (the experts' scores were significantly higher than novice surgeons' based on 29 subjects), and later publications provided construct validity as well.

There exist several modifications to the basic scoring skill assessment techniques. Takeshita et al. specified GEARS for endoluminal surgical platforms, called 'Global Evaluative Assessment of Robotic Skills in Endoscopy' (GEARS-E) [38]. GEARS-E is similar to GEARS, it measures depth perception, bimanual dexterity, efficiency, tissue handling, autonomy and endoscope control, but it was created for Master and Slave Transluminal Endoscopic Robot (MASTER) surgeries. GEARS-E is not yet widespread because it was developed in 2018, but the pilot study showed correlations to surgical expertise when using the MASTER.

Objective Structured Assessment of Technical Skills (OSATS) was originally created for evaluating traditional MIS skills along with FLS in 1997. OSATS involves the following evaluation points [39, 40]:

- respect for tissue (used forces, caused damage);

- time and motion (efficiency of time and motion);

- instrument handling (movements fluidity);

- knowledge of instruments (types and names);

- flow of operations (stops frequency);

- use of assistants (proper strategy);

- knowledge of specific procedure (familiarity of the aspect of the operation).

OSATS has an adaptation to robotic surgery: the Robotic Objective Structured Assessments of Technical Skills (R-OSATS) [41, 42]. R-OSATS metric evaluate the skills of the surgeon based on the depth perception/accuracy, force/tissue handling, dexterity and efficiency. R-OSATS was tested typically with gynecology students, it has construct validity, and in the tests, both the interrater and intrarater reliability were high [43].

4   Virtual Reality simulators

While Virtual Reality (VR) surgical robot simulators primarily support training, they can also be a great tool to measure surgical skills objectively in a well-defined environment, since all motions, contacts, errors, etc. can be computed in the VR environment. A typical RAMIS simulator involves a master side construction and the virtual surgical task simulation. The master side is responsible

for to study the usage of a teleoperation system (master arm handling, foot pedals, etc.), and to test the ergonomy. The simulation of the surgical task in case of a surgical robot simulator has to looking life-like and be clinically relevant. During the training, the VR simulators often estimate the skills based on manual skill assessment techniques (such as OSATS), but in an automated way.

Since the da Vinci dominating the global market, VR simulators are also focusing on da Vinci surgery. There are more than 2000 da Vinci simulators at the customer sites around the globe [44]. At the moment, there are six commercially available da Vinci surgical robot simulators: the da Vinci Skills Simulator (dVSS, Intuitive Surgical Inc.), dV-Trainer (Mimic Technologies Inc., Seattle, WA), Robotic Surgery Simulator (RoSS, Simulated Surgical Sciences LLC, Buffalo, NY), SEP Robot (SimSurgery, Norway), Robotix Mentor (3D systems (formerly Symbionix), Israel) and the Actaeon Robotic Surgery Training Console (BBZ Srl, University of Verona [45]). A novel surgical simulation program is the SimNow by da Vinci (Intuitive Surgical Inc.) [46]. SimNow involves surgical training using virtual instruments, guided and freehand procedure simulations and tracking skills and optimizing learning with management tools. In this section, the three most common types of VR simulators are reviewed: the DVSS, the dV-Trainer and the RoSS (Fig. 3).

DVSS can be attached to an actual da Vinci (da Vinci Xi, X or Si), with the main benefit that the surgeon can train on the actual robotic hardware, yet, it poses logistical problems, since while a trainee uses the simulator, the robot cannot be used for surgery. The dVSS contains the following surgical training categories [47]:

- EndoWrist manipulation;
- camera and clutching;
- energy and dissection;
- needle control;
- needle driving;
- suturing;
- additional games.

The dVSS is measures the skills based on the economy of motion, time to complete, instrument collisions, master workspace range, critical errors, instruments out of view, excessive force applied, missed targets drops, misapplied energy time. The simulator costs about $85000–585000 (the extra $500000 is for the console) [47–52].

The dV-Trainer emulates the da Vinci master console, thus it operates separated from the actual da Vinci robot. It contains additional training exercises to the dVSS [47]:

- troubleshooting;

- Research Training Network (virtual reality exercises to match physical devices in use by the research training network);

- Maestro AR (augmented reality; exercises that allow 3D interactions).

The dV-Trainer assesses skill with a very similar metric to the dVSS. In newer dV-Trainer versions, an alternative scoring system is available, called 'Proficiency Based System', which based on expert surgeon data, and the interpretation of the data is different, furthermore the user can customize the protocol. The dV-Trainer costs about $96000.

RoSS (as the dV-Trainer) is a stand-alone da Vinci simulator involving numerous modules:

- orientation module;

- motor skills module;

- basic surgical skills module;

- intermediate surgical skills module;

- blunt dissection and vessel dissection;

- hands-on surgical training module.

RoSS assesses the skills of the surgeon based on the camera usage, the number of left and right tool grasps, the distance while the left and right tool was out of view, the number of errors (collision or drop), the time to complete the task, the collisions of tools and tissue damage. RoSS costs about $126000.

In the literature, most papers dealing with surgical robot simulators are focused on the curriculum and the technical layout, yet, in this paper, the skill assessment and scoring part is crucial.

5   Automated assessment

Surgical robotics provides a unique platform to evaluate surgical skills automatically. RAMIS automated skill assessment does not need additional sensors to examine the surgeon's movements, camera handling, focusing on the image etc., because these events/errors/movements can be recorded straight with the robotic system. Automated assessment can be a powerful tool to evaluate surgical skills due to its objectivity, furthermore it does not require human resources, however, in some cases, it can be hard to implement these.

Two main types of automated skill assessment methods can be recognizable in the literature: *global information-based* and *language model-based* skill assessment. Global information-based automated skill assessment means that the surgical skill is evaualted based on the whole procedure, based on the data of the endolscopic video, kinematic data, or other additional sensor data. The other approach is to evaluate skills on the subtask level, called language-model based

Figure 3

Virtual reality simulators for the da Vinci Surgical System [47, 53, 54]. A) da Vinci Skills Simulator, b) dV-Trainer, c) Robotic Surgery Simulator, d) Robotix Mentor, e) SEP Robot, f) Actaeon Robotic Surgery Training Console

skill assessment. Here, the first challenge is to recognize the surgical subtasks (often called 'surgemes'), then create a model for the procedure, and compare the models for skill assessment. Global skill assessment is easier to implement compared to language model-based techniques, but language models can be more accurate, and they are closer to the natural training (an expert will teach to the novice what was wrong on the subtask level, such as the way to hold the needle in a suturing task).

5.1   Data collection for automated assessment

The development of automated RAMIS skill assessment methods requires solutions for surgical data collection. The data - which correlates the surgical skills - can be kinematic, video or additional sensor-based (e.g. force sensor). It is not trivial to access even to training data from RAMIS platforms. The da Vinci has a read-only research API (da Vinci Application Programmer's Interface, Intuitive Surgical Inc.), but it is only accessible to a very few chosen groups. The da Vinci API provides a robust motion data set and it can streams the motion vectors, including joint angles, Cartesian position and velocity, gripper angle, joint velocity and torque data from the master side of the da Vinci, furthermore events such as instrument changes [55].

To collect kinematic and sensory data from the da Vinci for research usage, the

Figure 4
JIGSAWS surgical tasks: knot-tying, suturing and needle passing (captured from the
video dataset)

da Vinci Research Kit (DVRK) is a more accessible tool. The DVRK (developed
by a consortium led by Johns Hopkins University and Worcester Polytechnic In-
stitute) is a research platform containing a set of open source software and hard-
ware elements, providing complete read and write access to the first generation
da Vinci [9]. DVRK is programmable via Robot Operating System (ROS) open
source library [56]. The DVRK community is relatively small, but growing with
only 35 DVRK sites [57].

While most of the da Vinci's have remote access and data storing enabled, due
to legal and liability causes, clinical datasets are not available widely. In this
case, annotated databases can provide input to RAMIS skill evaluation research.
JHU–ISI Gesture and Skill Assessment Working Set (JIGSAWS) (developed by
the LCSR lab at Hopkins and Intuitive) is an annotated database for surgical skill
assessment, collected over training sessions [58]. JIGSAWS contains kinematic
data (Cartesian positions, orientations, velocities, angular velocities and gripper
angle of the manipulators) and stereoscopic video data captured during dry lab
training (suturing, knot-tying and needle-passing). The dataset recorded on a da
Vinci involving surgeons with different expertise level (based on a manual eval-
uation technique). Beyond the manual skill annotations, JIGSAWS also includes
annotations about the gestures ('surgemes').

Another approach is to capture surgical data with an additional data collecting
device. A novel approach for da Vinci data collection, the dVLogger was devel-
oped in 2018 by Intuitive Surgical Inc. The dVLogger directly captures surgeons
motion data on the da Vinci Surgical System. DVLogger can be easily connected
to the da Vinci's vision tower with ethernet connection, and it records the data at
50 Hz. DVLogger provides the following informations from the da Vinci [59]:

- kinematic data (such as instrument travel time, path length, velocity);

- system events (frequency of master controller clutch use, camera move-
  ments, third arm swap, energy use);

- endoscopic video data.

DVLogger can be a powerful tool in surgical skill assessment studies, due to its
easy usage enables the data collection for everyone, during live surgeries as well,

however, it is a novel recording device, thus it is not well-known widely yet.

SurgTrak (created by the University of Minnesota and University of Washington) is an additional hardware and software set which can be used for the da Vinci as well [60, 61]. With SurgTrak, the endoscopic data can be captured from the DVI output of the da Vinci master side with an Epiphan DVI2USB device. The surgical instruments' position and orientation can be recorded with a 3D Guidance trakSTAR magnetic tracking system. Furthermore, grasper and wrist position is achievable with SurgTrak.

The above data collection techniques are useful for capturing kinematic and video data, but in some cases other devices/sensors are needed to evaluate surgical skills with specific algorithms. Force sensors are often used in the field of surgical skill assessment. It is possible to estimate the used forces during the training based on the motor currents, but due to the construction of the da Vinci, it can be very noisy. A more popular approach is when an additional force sensor is used, such as developed in U. Pennsylvania in [62]. In this case, accelerometers were placed on the da Vinci arms (which measured instrument vibrations), and a training board with a force sensor, which measured the forces during different types of training. They showed correlation between the measured data and the skill level.

5.2 Global information-based skill assessment

One approach for automated RAMIS skill assessment is to examine the whole procedure based on kinematic/video/additional sensor data. These methods are easier to implement than language model-based techniques, because they do not require the segmentation of the whole procedure (see details below). While global information-based methods are not sensitive to the performance quality of specific gestures, they can be as effective as language model-based techniques. There is an obvious correlation between the surgical skills and the kinematic data (Fig. 5), thus this is the most well-studied area in global information-based skill assessment [63–72], but we can find video, additional sensor-based [62, 73, 74], and the comparisons of several inputs [55, 75] automated techniques as well. Global information-based skill assessment is not as deeply studied as language model-based methods, in general.

For the global methods, the classification of the input data is needed. We can find a great summary of these in [68] (Fig. 6). The raw data (which can be any kind of data: endoscopic image, force, kinematic, etc – in the figure you can find a specific example for kinematic-data based assessment) have to be processed with some kind of feature extraction technique, and in some cases, dimensionality reduction is needed as well. The processed data can be classified, and the skill can be predicted based on the extracted features from the data.

In [68], we can find a motion-based automated skill assessment. Their input was the JIGSAWS dataset. They used 4 types of kinematic holistic features: sequential motion texture, discrete Fourier transform, discrete cosine transform and approximate entropy. After the feature extraction and dimensionality reduc-
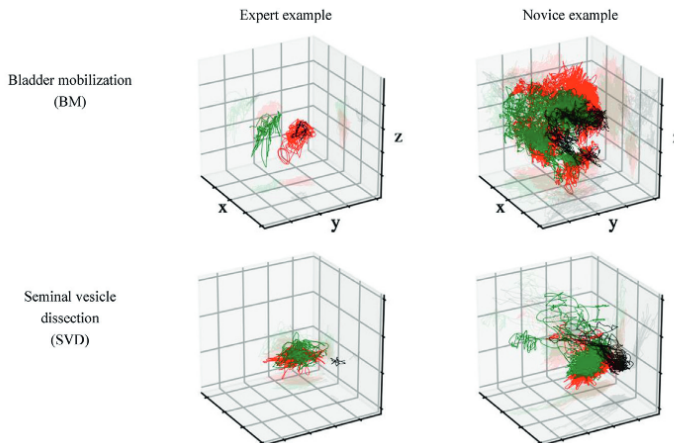
Figure 5

Robot trajectories in case of a novice and an expert surgeon during robot-assisted radical prostatectomy (red: dominant instrument, green: non-dominant instrument, black: camera) [59]
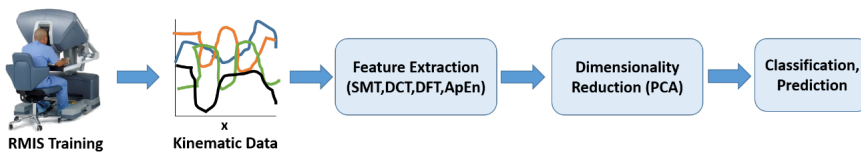

Figure 6

Flowdiagram for automated surgical skill assessment [68]

tion, they classified the data and predicted the skill score. The skill scoring was performed with a weighted holistic feature combination technique, which means that different prediction models were used to produce a final skill score. With this method a modified-OSATS score and a Global Rating Score was estimated. The results showed more accuracy than Hidden Markov Model-based solutions [68]. For more approaches, see Table 1.

5.3 Language model-based skill assessment

A surgical procedure model can be built with different motion granularity. A surgical procedure (such as *Laparoscopic cholecystectomy*) is built from tasks (e.g., *exposing Calot's triangle*), which is built from subtasks (*e.g., blunt dissection*), which is built from surgemes (*grasp*), which is built from dexemes (*motion primitives*) (Fig. 7). Global skill assessment methods approach the skill evaluation from the highest procedure/task level, thus not adverting the fact that surgical tasks are built from several, sometimes very different surgemes. These surgemes are not equally easy or complicated to execute, and even if a clinician believed

Level of granularity	Definition	Time span	Com-plexity	Example
Operation	The entire invasive part of the procedure.	20–200 min	very high	Laparoscopic cholecystectomy
Task	Well delimited surgical activity with a given high-level target/goal to achieve.	1–5 min	high	Pneumo-peritoneum → Exposing-Calot's triangle → …
Subtask	Circumscribed activity segments that accomplish specific minor landmarks in completing the surgical task.	0.1–2 min	moderate	Retraction of the gallbladder → Blunt dissection at the Cystic duct → Blunt dissection at the Cystic art. → …
Surgeme	An atomic unit of intentional surgical activity resulting in a perceivable and meaningful outcome.	0.1–0.5 min	low	Approach the tissue → Perform dissecting motion → …
Motion primitive	General elements of motion patterns, that can be directly translated into robot commands.	1–5 sec	very low	Penetrate connective tissue → Open the dissector → Remove the dissector

Figure 7

A surgical procedure built from different levels [101]. Language model-based RAMIS skill assessment techniques typically evaluate the skills on the surgeme level.

to have intermediate skills based on a global skill assessment technique, he/she can be excellent/poor in just one, but very important surgeme and vice versa. Language model-based surgical skill assessment aims to assess surgical skills on the surgeme level, thus it requires three main steps: task segmentation, gesture recognition and gesture-based skill assessment. This approach has the further advantage that with the models defined, we can study the transitions between the surgemes, and benchmark those as well. This approach has been considered to be a cornerstone of the emerging field of Surgical Data Science (SDS) [76].

It was the Hopkins group who first proposed surgeme-based skill assessment [77], discrete Hidden Markov Models (HMM) were built for task and for surgeme level as well to assess skill. In the practice, skill evaluation was based on a model built from annotated data (known expertise level), and this model tested against the new user. To create a model for user motions, they had to identified the surgemes with feature extraction, dimensionality reduction and classifier representation techniques. After that, the two models were compared. To train the discrete HMMs they used vector quantization. Their method worked with 100% accuracy using task level models and known gesture segmentation, at 95% with task level models and unknown gesture segmentation, and at 100% with the surgeme level models in correctly identifying the skill level.

The input of language model-based skill assessment methods can be kinematic data [77–86], video data [87] or both [88–92]. In the literature, we can find surgical activity/workflow segmentation as well [93–100]. For the details of the state-of-the-art see Table 1.

6 Non-technical surgical skill assessment

Surgical robotic interventions can put extra cognitive load on the surgeon, especially in the case of risky, high-complexity tasks, or in emergency. Furthermore,

surgical robotic operations require teamwork, thus excellent communication and problem solving skills are needed from the surgeon (and from all of the operators as well). For all the above reasons, non-technical surgical skills are also important in case of surgical robotics, however, it is not a well-studied area. Non-technical skills involves cognitive skills (such as decision making, memory, reaction time) and social skills (such as communication skills, working ability in a team and as a leader) [20, 102].

The NASA Task Load Index (NASA-TLX) was not originally created for surgery, but has been used in this field successfully [102]. NASA-TLX is a subjective scoring tool, including questions about mental, physical and temporal demand, furthermore performance, effort and frustration [20], with the advantage to quantify subjective parameters, and making them comparable to other experiments. To conform to the needs of surgical skill assessment, Surgery Task Load Index (SURG-TLX) was derived from NASA-TLX, but this technique is not yet used for robotic surgery, just for traditional MIS [103]. SURG-TLX examines the impact of different types of stress (such as task complexity, situational stress, distractions) in case of surgeons. Non-technical Skills for Surgeons (NOTSS) was created specifically for non-technical surgical skill assessment. NOTSS metric includes the examination of situation awareness, decision making, task management, communication and teamwork, and leadership [104]. NOTSS was recently used in surgical robotics non-technical skill assessment as well [105].

The Interpersonal and Cognitive Assessment for Robotic Surgery (ICARS) was the first objective method for RAMIS non-technical skill assessment. For ICARS, 28 non-technical behaviours were identified by expert surgeons based on the Delphi method [102, 106]. In the ICARS metrics, there are four main types of non-technical skills: checklist and equipment, interpersonal, cognitive and resource skills.

Nowadays, there are not any kind of automated skill assessment method for non-technical skills. Electroencephalography (EEG) could be employed to estimate non-technical skills during RAMIS, but due to the complexity of an EEG, it is not a well-known method for surgical skill assessment [102]. There are some limited studies in this field [107]. Guru et al. used EEG signals (nine-channel EEG recording with a neuro-headset) for cognitive skill assessment during RAMIS training. They placed the sensor on the frontal, central, parietal and occipital regions. The statistical analysis showed that with cognitive metrics, there were significant differences between the groups for the basic, intermediate and expert skills based on the data of 10 surgeons.

On the other hand, there are several methods aimed at measuring physiological signals, which can refer to the stress level, however, these are not used in RAMIS widely yet. Stress directly influences the performance of a surgeon, thus the measurement of the sress level can be a tool for non-techninal surgical skill assessment [108]. In the literature, we can find examples to stress-related signals of the human body: skin temperature [109, 110], temperature of the nose [111], heart rate, skin conductance, blood pressure, respiratory period [112] etc. In case of surgical performance, tremor is the most studied physiological signal, but it

did not refer to the stress level in all cases [113].

7   Conclusion

Surgical skill assessment is an essential component to improve the level of training, and for providing quality assurance in primary care. Robotic surgery provides a unique platform to evaluate surgical skills objectively, since it inherently collects a wide range data. Nowadays, in the clinical practice, there is no routinely employed objective skill evaluation method. In the literature of Robot-Assisted Minimally Invasive Surgery, there are two main approaches for technical skill assessment: manual and automated. There are several validated manual evaluation methods, such as GEARS and R-OSATS, which are relatively easy to implement, but require an expert panel, prone to subjective bias. Automated RAMIS skill assessment is also a heavily studied area: there are global and language model-based methods. These are harder to implement, but in the near future, these can become an extremely powerful tool to objectively evaluate surgical skills, until we see a gradual takeover of robotic execution [114]. With the help of surgical robotics, data can be easily captured with automated tools. The input can range from kinematic data produced by the motion of the surgeon (which is the most studied approach), to endoscopic video data and force signals, etc. Automated methods can predict skills score without using human resources, and permit personalized skill training. With the novel training techniques, we hypothesize significantly improved surgical performance, therefore better patient outcome in the clinical practice.

References

[1]   R. M. Satava. Surgical Robotics: The Early Chronicles: A Personal Historical Perspective. *Surgical Laparoscopy Endoscopy & Percutaneous Techniques*, 12(1):6–16, 2002.

[2]   K. H. Fuchs. Minimally Invasive Surgery. *Endoscopy*, 34(2):154–159, 2002.

[3]   A. Takács, D. A. Nagy, I. Rudas, and T. Haidegger. Origins of Surgical Robotics: From Space to the Operating Room. *Acta Polytechnica Hungarica*, 13.1:13–30, 2016.

[4]   K. Cleary and C. Nguyen. State of the Art in Surgical Robotics: Clinical Applications and Technology Challenges. *Computer Aided Surgery*, 6(6):312–328, 2001.

[5]   S. Maeso, M. Reza, J. A. Mayol, J. A. Blasco, M. Guerra, E. Andradas, and M. N. Plana. Efficacy of the Da Vinci Surgical System in Abdominal Surgery Compared With That of Laparoscopy: A Systematic Review and Meta-Analysis. *Annals of Surgery*, 252(2):254–262, 2010.

[6]   A. Paczuski and S. M. Krishnan. Analyzing Product Failures and Improving Design : A Case Study in Medical Robotics, access date: 2018-12-20.

[7]   S. Tsuda, D. Oleynikov, J. Gould, D. Azagury, B. Sandler, M. Hutter, S. Ross, E. Haas, F. Brody, and R. Satava. SAGES TAVAC safety and effectiveness analysis: Da Vinci ℝ Surgical System (Intuitive Surgical, Sunnyvale, CA). *Surg Endosc*, 29(10):2873–2884, 2015.

[8]   H. Alemzadeh, J. Raman, N. Leveson, Z. Kalbarczyk, and R. K. Iyer. Adverse Events in Robotic Surgery: A Retrospective Study of 14 Years of FDA Data. *PLoS ONE*, 11(4):e0151470, 2016.

[9]   P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio. An open-source research kit for the da Vinci ℝ Surgical System. pages 6434–6439. IEEE, 2014.

[10]   InTouch Health Announces Strategic Collaboration With Intuitive Surgical. https://intouchhealth.com/strategic-collaboration-with-intuitive-surgical/. Access date: 2018-12-20., 2016.

[11]   A. Pedersen. Intuitive Surgical Could Help Usher in a New Era for Medtech. https://www.mddionline.com/intuitive-surgical-could-help-usher-new-era-medtech. Access date: 2018-12-20., 2018.

[12]   A. N. Sridhar, T. P. Briggs, J. D. Kelly, and S. Nathan. Training in Robotic Surgery—an Overview. *Curr Urol Rep*, 18(8), 2017.

[13]   J. Sándor, B. Lengyel, T. Haidegger, G. Saftics, G. Papp, A. Nagy, and G. Wéber. Minimally invasive surgical technologies: Challenges in education and training. *Asian J. of Endoscopic Surgery*, 3(3):101–108, 2010.

[14]   R. Smith, V. Patel, and R. Satava. Fundamentals of robotic surgery: A course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development. *The international journal of medical robotics + computer assisted surgery: MRCAS*, 10(3):379–384, September 2014.

[15]   A. Peña. The Dreyfus model of clinical problem-solving skills acquisition: A critical perspective. *Med Educ Online*, 15, 2010.

[16]   J. Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):257–266, May 1983.

[17]   D. Azari, C. Greenberg, C. Pugh, D. Wiegmann, and R. Radwin. In Search of Characterizing Surgical Skill. *Journal of Surgical Education*, March 2019.

[18]   T. M. Kowalewski and T. S. Lendvay. Performance Assessment. In D. Stefanidis, J. R. Korndorffer Jr., and R. Sweet, editors, *Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties*, Com-

prehensive Healthcare Simulation, pages 89–105. Springer Intl. Publishing, Cham, 2019.

[19] J. T. O'Donovan, B. Kang, and T. Höllerer. Competence Modeling in Twitter : Mapping Theory to Practice. 2015.

[20] J. Chen, N. Cheng, G. Cacciamani, P. Oh, M. Lin-Brande, D. Remulla, I. S. Gill, and A. J. Hung. Objective assessment of robotic surgical technical skill: A systemic review (accepted manuscript). *J. Urol.*, 2018.

[21] A. A. Hussein, K. R. Ghani, J. Peabody, R. Sarle, R. Abaza, D. Eun, J. Hu, M. Fumo, B. Lane, J. S. Montgomery, N. Hinata, D. Rooney, B. Comstock, H. K. Chan, S. S. Mane, J. L. Mohler, G. Wilding, D. Miller, K. A. Guru, and Michigan Urological Surgery Improvement Collaborative and Applied Technology Laboratory for Advanced Surgery Program. Development and Validation of an Objective Scoring Tool for Robot-Assisted Radical Prostatectomy: Prostatectomy Assessment and Competency Evaluation. *J. Urol.*, 197(5):1237–1244, 2017.

[22] A. A. Hussein, K. J. Sexton, P. R. May, M. V. Meng, A. Hosseini, D. D. Eun, S. Daneshmand, B. H. Bochner, J. O. Peabody, R. Abaza, E. C. Skinner, R. E. Hautmann, and K. A. Guru. Development and validation of surgical training tool: Cystectomy assessment and surgical evaluation (CASE) for robot-assisted radical cystectomy for men. *Surg Endosc*, 32(11):4458–4464, 2018.

[23] A. A. Hussein, N. Hinata, S. Dibaj, P. R. May, J. D. Kozlowski, H. Abol-Enein, R. Abaza, D. Eun, M. S. Khan, J. L. Mohler, P. Agarwal, K. Pohar, R. Sarle, R. Boris, S. S. Mane, A. Hutson, and K. A. Guru. Development, validation and clinical application of Pelvic Lymphadenectomy Assessment and Completion Evaluation: Intraoperative assessment of lymph node dissection after robot-assisted radical cystectomy for bladder cancer. *BJU International.*, 119(6):879–884, 2017.

[24] A. A. Hussein, R. Abaza, C. Rogers, R. Boris, J. Porter, M. Allaf, K. Badani, M. Stifelman, J. Kaouk, T. Terakawa, Y. Ahmed, E. Kauffman, Q. Li, K. Guru, and D. Eun. Development and validation of an objective scoring tool for minimally invasive partial nephrectomy: Scoring for partial nephrectomy (SPaN). *J. Urol*, 199(4):e159–e160, 2018.

[25] H. Husslein, L. Shirreff, E. M. Shore, G. G. Lefebvre, and T. P. Grantcharov. The Generic Error Rating Tool: A Novel Approach to Assessment of Performance and Surgical Education in Gynecologic Laparoscopy. *J Surg Educ*, 72(6):1259–1265, 2015 Nov-Dec.

[26] H. Husslein, E. Bonrath, T. Grantcharov, and G. Lefebvre. Validation of the Generic Error Rating Tool (GERT) in Gynecologic Laparoscopy (Preliminary Data). *Journal of Minimally Invasive Gynecology*, 20(6):S106, 2013.

[27] P. Ramos, J. Montez, A. Tripp, C. K. Ng, I. S. Gill, and A. J. Hung. Face, content, construct and concurrent validity of dry laboratory exercises for robotic training using a global assessment tool. *BJU International*, 113(5):836–842, 2014.

[28]   A. C. Goh, D. W. Goldfarb, J. C. Sander, B. J. Miles, and B. J. Dunkin. Global evaluative assessment of robotic skills: Validation of a clinical assessment tool to measure robotic surgical skills. *J. Urol.*, 187(1):247–252, 2012.

[29]   R. Sánchez, O. Rodríguez, J. Rosciano, L. Vegas, V. Bond, A. Rojas, and A. Sanchez-Ismayel. Robotic surgery training: Construct validity of Global Evaluative Assessment of Robotic Skills (GEARS). *J Robot Surg*, 10(3):227–231, 2016.

[30]   M. A. Aghazadeh, I. S. Jayaratna, A. J. Hung, M. M. Pan, M. M. Desai, I. S. Gill, and A. C. Goh. External validation of Global Evaluative Assessment of Robotic Skills (GEARS). *Surg Endosc*, 29(11):3261–3266, 2015.

[31]   M. Liu, S. Purohit, J. Mazanetz, W. Allen, U. S. Kreaden, and M. Curet. Assessment of Robotic Console Skills (ARCS): Construct validity of a novel global rating scale for technical skills in robotically assisted surgery. *Surg Endosc*, 32(1):526–535, 2018.

[32]   K. R. Ghani, D. C. Miller, S. Linsell, A. Brachulis, B. Lane, R. Sarle, D. Dalela, M. Menon, B. Comstock, T. S. Lendvay, J. Montie, J. O. Peabody, and Michigan Urological Surgery Improvement Collaborative. Measuring to Improve: Peer and Crowd-sourced Assessments of Technical Skill with Robot-assisted Radical Prostatectomy. *Eur. Urol.*, 69(4):547–550, 2016.

[33]   A. Guni, N. Raison, B. Challacombe, S. Khan, P. Dasgupta, and K. Ahmed. Development of a technical checklist for the assessment of suturing in robotic surgery. *Surg Endosc*, 32(11):4402–4407, 2018.

[34]   Q. Ballouhey, P. Clermidi, J. Cros, C. Grosos, C. Rosa-Arsène, C. Bahans, F. Caire, B. Longis, R. Compagnon, and L. Fourcade. Comparison of 8 and 5 mm robotic instruments in small cavities : 5 or 8 mm robotic instruments for small cavities? *Surg Endosc*, 32(2):1027–1034, 2018.

[35]   S. L. Vernez, V. Huynh, K. Osann, Z. Okhunov, J. Landman, and R. V. Clayman. C-SATS: Assessing Surgical Skills Among Urology Residency Applicants. *J. Endourol.*, 31(S1):S95–S100, 2017.

[36]   A. J. Hung, T. Bottyan, T. G. Clifford, S. Serang, Z. K. Nakhoda, S. H. Shah, H. Yokoi, M. Aron, and I. S. Gill. Structured learning for robotic surgery utilizing a proficiency score: A pilot study. *World J Urol*, 35(1):27–34, 2017.

[37]   A. Volpe, K. Ahmed, P. Dasgupta, V. Ficarra, G. Novara, H. van der Poel, and A. Mottrie. Pilot Validation Study of the European Association of Urology Robotic Training Curriculum. *Eur. Urol.*, 68(2):292–299, 2015.

[38]   N. Takeshita, S. J. Phee, P. W. Chiu, and K. Y. Ho. Global Evaluative Assessment of Robotic Skills in Endoscopy (GEARS-E): Objective assessment tool for master and slave transluminal endoscopic robot. *Endosc Int Open*, 6(8):E1065–E1069, 2018.

[39]   H. Niitsu, N. Hirabayashi, M. Yoshimitsu, T. Mimura, J. Taomoto, Y. Sugiyama, S. Murakami, S. Saeki, H. Mukaida, and W. Takiyama. Using the Objective Structured Assessment of Technical Skills (OSATS)

global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today*, 43(3):271–275, 2013.

[40] N. Y. Siddiqui, M. L. Galloway, E. J. Geller, I. C. Green, H.-C. Hur, K. Langston, M. C. Pitter, M. E. Tarr, and M. A. Martino. Validity and reliability of the robotic Objective Structured Assessment of Technical Skills. *Obstet Gynecol*, 123(6):1193–1199, 2014.

[41] M. R. Polin, N. Y. Siddiqui, B. A. Comstock, H. Hesham, C. Brown, T. S. Lendvay, and M. A. Martino. Crowdsourcing: A valid alternative to expert evaluation of robotic surgery skills. *Am. J. Obstet. Gynecol.*, 215(5):644.e1–644.e7, 2016.

[42] M. E. Tarr, C. Rivard, A. E. Petzel, S. Summers, E. R. Mueller, L. M. Rickey, M. A. Denman, R. Harders, R. Durazo-Arvizu, and K. Kenton. Robotic objective structured assessment of technical skills: A randomized multicenter dry laboratory training pilot study. *Female Pelvic Med Reconstr Surg*, 20(4):228–236, 2014 Jul-Aug.

[43] N. Y. Siddiqui, M. L. Galloway, E. J. Geller, I. C. Green, H.-C. Hur, K. Langston, M. C. Pitter, M. E. Tarr, and M. A. Martino. Validity and reliability of the robotic Objective Structured Assessment of Technical Skills. *Obstet Gynecol*, 123(6):1193–1199, 2014.

[44] Intuitive Surgical Investor Presentation 021218 — Surgery — Cardiothoracic Surgery. https://www.scribd.com/document/376731845/Intuitive-Surgical-Investor-Presentation-021218. Access date: 2018-12-20.

[45] F. Bovo, G. De Rossi, and F. Visentin. Surgical robot simulation with BBZ console. *J Vis Surg*, 3, 2017.

[46] Intuitive — Products Services — Education Training. https://www.intuitive.com/en/products-and-services/da-vinci/education. Access date: 2018-12-20.

[47] D. Julian, A. Tanaka, P. Mattingly, M. Truong, M. Perez, and R. Smith. A comparative analysis and guide to virtual reality robotic surgical simulators. *The Intl. Journal of Medical Robotics and Computer Assisted Surgery*, 14(1), 2018.

[48] Intuitive Surgical - da Vinci Si Surgical System - Skills Simulator. https://www.intuitivesurgical.com/products/skills_simulator/. Access date: 2018-12-20.

[49] A. Tanaka, C. Graddy, K. Simpson, M. Perez, M. Truong, and R. Smith. Robotic surgery simulation validity and usability comparative analysis. *Surg Endosc*, 30(9):3720–3729, 2016.

[50] H. Schreuder, R. Wolswijk, R. Zweemer, M. Schijven, and R. Verheijen. Training and learning robotic surgery, time for a more structured approach: A systematic review: Training and learning robotic surgery. *BJOG: An Intl. Journal of Obstetrics & Gynaecology*, 119(2):137–149, 2012.

[51] A. N. Sridhar, T. P. Briggs, J. D. Kelly, and S. Nathan. Training in Robotic Surgery—an Overview. *Curr Urol Rep*, 18(8), 2017.

[52] R. Smith, M. Truong, and M. Perez. Comparative analysis of the functionality of simulators of the da Vinci surgical robot. *Surg Endosc*, 29(4):972–983, 2015.

[53]   BBZ - Medical Technologies. http://www.bbzsrl.com/index.html. Access date: 2018-12-20.

[54]   SEP robot trainer. http://surgrob.blogspot.com/2013/10/sep-robot-trainer.html. Access date: 2018-12-20.

[55]   R. Kumar, A. Jog, B. Vagvolgyi, H. Nguyen, G. Hager, C. C. G. Chen, and D. Yuh. Objective measures for longitudinal assessment of robotic surgery training. *The Journal of Thoracic and Cardiovascular Surgery*, 143(3):528–534, 2012.

[56]   ROS.org — Powering the world's robots. http://www.ros.org/. Access date: 2018-12-20.

[57]   Cisst/SAW stack for the da Vinci Research Kit. Contribute to jhu-dvrk/sawIntuitiveResearchKit. https://github.com/jhu-dvrk/sawIntuitiveResearchKit. Access date: 2018-12-20.

[58]   Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Bejar, D. D. Yuh, C. C. G. Chen, R. Vidal, S. Khudanpur, and G. D. Hager. JHU–ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling. page 10.

[59]   A. J. Hung, J. Chen, A. Jarc, D. Hatcher, H. Djaladat, and I. S. Gill. Development and Validation of Objective Performance Metrics for Robot-Assisted Radical Prostatectomy: A Pilot Study. *J. Urol.*, 199(1):296–304, 2018.

[60]   K. Ruda, D. Beekman, L. W. White, T. S. Lendvay, and T. M. Kowalewski. SurgTrak — A Universal Platform for Quantitative Surgical Data Capture. *Journal of Medical Devices*, 7(3):030923–030923–2, July 2013.

[61]   SurgTrak: Affordable motion tracking & video capture for the da Vinci surgical robot - SAGES Abstract Archives. https://www.sages.org/meetings/annual-meeting/abstracts-archive/surgtrak-affordable-motion-tracking-and-video-capture-for-the-da-vinci-surgical-robot/.

[62]   E. D. Gomez, R. Aggarwal, W. McMahan, K. Bark, and K. J. Kuchenbecker. Objective assessment of robotic surgical skill using instrument contact vibrations. *Surg Endosc*, 30(4):1419–1431, 2016.

[63]   T. N. Judkins, D. Oleynikov, and N. Stergiou. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surg Endosc*, 23(3):590, 2009.

[64]   I. Nisky, M. H. Hsieh, and A. M. Okamura. The effect of a robot-assisted surgical system on the kinematics of user movements. *Conf Proc IEEE Eng Med Biol Soc*, 2013:6257–6260, 2013.

[65]   M. J. Fard, S. Ameri, R. B. Chinnam, A. K. Pandya, M. D. Klein, and R. D. Ellis. Machine Learning Approach for Skill Evaluation in Robotic-Assisted Surgery. *arXiv:1611.05136 [cs, stat]*, 2016.

[66]   Y. Sharon, T. S. Lendvay, and I. Nisky. Instrument Orientation-Based Metrics for Surgical Skill Evaluation in Robot-Assisted and Open Needle Driving. *arXiv:1709.09452 [cs]*, 2017.

[67] M. J. Fard, S. Ameri, R. D. Ellis, R. B. Chinnam, A. K. Pandya, and M. D. Klein. Automated robot-assisted surgical skill evaluation: Predictive analytics approach. *The Intl. Journal of Medical Robotics and Computer Assisted Surgery*, 14(1):e1850.

[68] A. Zia and I. Essa. Automated surgical skill assessment in RMIS training. *Int J Comput Assist Radiol Surg*, 13(5):731–739, 2018.

[69] Z. Wang and A. M. Fey. SATR-DL: Improving Surgical Skill Assessment and Task Recognition in Robot-assisted Surgery with Deep Neural Networks. *arXiv:1806.05798 [cs]*, 2018.

[70] Y. Sharon and I. Nisky. What Can Spatiotemporal Characteristics of Movements in RAMIS Tell Us? *Journal of Medical Robotics Research*, page 1841008, 2018.

[71] K. Liang, Y. Xing, J. Li, S. Wang, A. Li, and J. Li. Motion control skill assessment based on kinematic analysis of robotic end-effector movements. *Int J Med Robot*, 14(1), 2018.

[72] Z. Wang and A. Majewicz Fey. Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. *Int J Comput Assist Radiol Surg*, 2018.

[73] J. D. Brown, C. E. O Brien, S. C. Leung, K. R. Dumon, D. I. Lee, and K. J. Kuchenbecker. Using Contact Forces and Robot Arm Accelerations to Automatically Rate Surgeon Skill at Peg Transfer. *IEEE Trans Biomed Eng*, 64(9):2263–2275, 2017.

[74] M. Ershad, R. Rege, and A. M. Fey. Meaningful Assessment of Robotic Surgical Style using the Wisdom of Crowds. *Int J Comput Assist Radiol Surg*, 13(7):1037–1048, 2018.

[75] R. Kumar, A. Jog, A. Malpani, B. Vagvolgyi, D. Yuh, H. Nguyen, G. Hager, and C. C. Grace Chen. Assessing system operation skills in robotic surgery trainees. *Int J Med Robot*, 8(1):118–124, 2012.

[76] L. Maier-Hein, S. Vedula, S. Speidel, N. Navab, R. Kikinis, A. Park, M. Eisenmann, H. Feussner, G. Forestier, S. Giannarou, M. Hashizume, D. Katic, H. Kenngott, M. Kranzfelder, A. Malpani, K. März, T. Neumuth, N. Padoy, C. Pugh, N. Schoch, D. Stoyanov, R. Taylor, M. Wagner, G. D. Hager, and P. Jannin. Surgical Data Science: Enabling Next-Generation Surgery. *Nature Biomedical Engineering*, 1(9):691–696, 2017.

[77] C. E. Reiley and G. D. Hager. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. *Med Image Comput Comput Assist Interv*, 12(Pt 1):435–442, 2009.

[78] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager. Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. *Computer Aided Surgery*, 11(5):220–230, 2006.

[79] C. E. Reiley, H. C. Lin, B. Varadarajan, B. Vagvolgyi, S. Khudanpur, D. D. Yuh, and G. D. Hager. Automatic recognition of surgical motions using statistical modeling for capturing variability. In *Studies in Health Technology and Informatics*, pages 396–401, 2008.

[80] B. Varadarajan, C. Reiley, H. Lin, S. Khudanpur, and G. Hager. Data-Derived Models for Segmentation with Application to Surgical Assess-

ment and Training. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, Lecture Notes in Computer Science, pages 426–434. Springer, Berlin, Heidelberg, 2009.

[81] L. Tao, E. Elhamifar, S. Khudanpur, G. D. Hager, and R. Vidal. Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation. In *Information Processing in Computer-Assisted Interventions*, Lecture Notes in Computer Science, pages 167–177. Springer, Berlin, Heidelberg, 2012.

[82] N. Ahmidi, Y. Gao, B. Béjar, S. S. Vedula, S. Khudanpur, R. Vidal, and G. D. Hager. String motif-based description of tool motion for detecting skill and gestures in robotic surgery. *Med Image Comput Comput Assist Interv*, 16(Pt 1):26–33, 2013.

[83] S. Sefati, N. Cowan, and R. Vidal. Learning Shared, Discriminative Dictionaries for Surgical Gesture Segmentation and Classification. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, volume 4, 2015.

[84] F. Despinoy, D. Bouget, G. Forestier, C. Penet, N. Zemiti, P. Poignet, and P. Jannin. Unsupervised Trajectory Segmentation for Surgical Gesture Recognition in Robotic Training. *IEEE Transactions on Biomedical Engineering*, 63(6):1280–1291, 2016.

[85] S. Krishnan, A. Garg, S. Patil, C. Lea, G. Hager, P. Abbeel, and K. Goldberg. Transition State Clustering: Unsupervised Surgical Trajectory Segmentation for Robot Learning. In A. Bicchi and W. Burgard, editors, *Robotics Research: Volume 2*, Springer Proceedings in Advanced Robotics, pages 91–110. Springer Intl. Publishing, Cham, 2018.

[86] G. Forestier, F. Petitjean, P. Senin, F. Despinoy, A. Huaulmé, H. I. Fawaz, J. Weber, L. Idoumghar, P.-A. Muller, and P. Jannin. Surgical motion analysis using discriminative interpretable patterns. *Artif Intell Med*, (91):3–11, 2018.

[87] B. B. Haro, L. Zappella, and R. Vidal. Surgical gesture classification from video data. *Med Image Comput Comput Assist Interv*, 15(1):34–41, 2012.

[88] H. C. Lin and G. Hager. User-Independent Models of Manipulation Using Video Contextual Cues. *Workshop on Modeling and Monitoring of Computer Assisted Interventions*, 2009.

[89] L. Zappella, B. Béjar, G. Hager, and R. Vidal. Surgical gesture classification from video and kinematic data. *Medical Image Analysis*, 17(7):732–745, 2013.

[90] A. Malpani, S. S. Vedula, C. C. G. Chen, and G. D. Hager. Pairwise Comparison-Based Objective Score for Automated Skill Assessment of Segments in a Surgical Task. In D. Stoyanov, D. L. Collins, I. Sakuma, P. Abolmaesumi, and P. Jannin, editors, *Information Processing in Computer-Assisted Interventions*, Lecture Notes in Computer Science, pages 138–147. Springer Intl. Publishing, 2014.

[91] N. Ahmidi, L. Tao, S. Sefati, Y. Gao, C. Lea, B. B. Haro, L. Zappella, S. Khudanpur, R. Vidal, and G. D. Hager. A Dataset and Benchmarks

for Segmentation and Recognition of Gestures in Robotic Surgery. *IEEE Transactions on Biomedical Engineering*, 64(9):2025–2041, 2017.

[92] S. Jun, M. S. Narayanan, P. Agarwal, A. Eddib, P. Singhal, S. Garimella, and V. Krovi. Robotic Minimally Invasive Surgical skill assessment based on automated video-analysis motion studies. In *2012 4th IEEE RAS EMBS Intl. Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 25–31, 2012.

[93] C. Lea, G. D. Hager, and R. Vidal. An Improved Model for Segmentation and Recognition of Fine-Grained Activities with Application to Surgical Training Tasks. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 1123–1129, 2015.

[94] Automated skill assessment for individualized training in robotic surgery — —Science of Learning. http://scienceoflearning.jhu.edu/research/automated-skill-assessment-for-individualized-training-in-robotic-surgery. Access date: 2018-12-20.

[95] A. Malpani, S. S. Vedula, C. C. G. Chen, and G. D. Hager. A study of crowdsourced segment-level surgical skill assessment using pairwise rankings. *Int J CARS*, 10(9):1435–1447, 2015.

[96] S. Krishnan, A. Garg, S. Patil, C. Lea, G. D. Hager, P. Abbeel, and K. Goldberg. Unsupervised Surgical Task Segmentation with Milestone Learning. In *Proc. Intl Symp. on Robotics Research (ISRR)*, 2015.

[97] C. Lea, A. Reiter, R. Vidal, and G. D. Hager. Segmental Spatiotemporal CNNs for Fine-Grained Action Segmentation. In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, pages 36–52. Springer Intl. Publishing, 2016.

[98] R. DiPietro, C. Lea, A. Malpani, N. Ahmidi, S. S. Vedula, G. I. Lee, M. R. Lee, and G. D. Hager. Recognizing Surgical Activities with Recurrent Neural Networks. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Lecture Notes in Computer Science, pages 551–558. Springer Intl. Publishing, 2016.

[99] S. S. Vedula, A. O. Malpani, L. Tao, G. Chen, Y. Gao, P. Poddar, N. Ahmidi, C. Paxton, R. Vidal, S. Khudanpur, G. D. Hager, and C. C. G. Chen. Analysis of the Structure of Surgical Activity for a Suturing and Knot-Tying Task. *PLoS ONE*, 11(3):e0149174, 2016.

[100] A. Zia, C. Zhang, X. Xiong, and A. M. Jarc. Temporal clustering of surgical activities in robot-assisted surgery. *Int J Comput Assist Radiol Surg*, 12(7):1171–1178, 2017.

[101] T. D. Nagy and T. Haidegger. A DVRK-based Framework for Surgical Subtask Automation. *Acta Polytechnica Hungarica*, 14(Special Issue on Platforms for Medical Robotics Research (accepted manuscript)), 2019.

[102] Understanding and Assessing Nontechnical Skills in Robotic Urological Surgery: A Systematic Review and Synthesis of the Validity Evidence. *Journal of Surgical Education*, 2018.

[103] M. R. Wilson, J. M. Poolton, N. Malhotra, K. Ngo, E. Bright, and R. S. W. Masters. Development and Validation of a Surgical Workload Measure:

The Surgery Task Load Index (SURG-TLX). *World J Surg*, 35(9):1961–1969, 2011.

[104] S. Yule, R. Flin, N. Maran, D. Rowley, G. Youngson, and S. Paterson-Brown. Surgeons' Non-technical Skills in the Operating Room: Reliability Testing of the NOTSS Behavior Rating System. *World Journal of Surgery*, 32(4):548–556, April 2008.

[105] N. Raison, K. Ahmed, T. Abe, O. Brunckhorst, G. Novara, N. Buffi, C. McIlhenny, H. van der Poel, M. van Hemelrijck, A. Gavazzi, and P. Dasgupta. Cognitive training for technical and non-technical skills in robotic surgery: A randomised controlled trial. *BJU International*, 122(6):1075–1081, December 2018.

[106] N. Raison, T. Wood, O. Brunckhorst, T. Abe, T. Ross, B. Challacombe, M. S. Khan, G. Novara, N. Buffi, H. Van Der Poel, C. McIlhenny, P. Dasgupta, and K. Ahmed. Development and validation of a tool for non-technical skills evaluation in robotic surgery-the ICARS system. *Surg Endosc*, 31(12):5403–5410, 2017.

[107] K. A. Guru, E. T. Esfahani, S. J. Raza, R. Bhat, K. Wang, Y. Hammond, G. Wilding, J. O. Peabody, and A. J. Chowriappa. Cognitive skills assessment during robot-assisted surgery: Separating the wheat from the chaff. *BJU Intl.*, 115(1):166–174, 2015.

[108] C. M. Wetzel, R. L. Kneebone, M. Woloshynowych, D. Nestel, K. Moorthy, J. Kidd, and A. Darzi. The effects of stress on surgical performance. *The American Journal of Surgery*, 191(1):5–10, 2006.

[109] K. A. Herborn, J. L. Graves, P. Jerem, N. P. Evans, R. Nager, D. J. McCafferty, and D. E. McKeegan. Skin temperature reveals the intensity of acute stress. *Physiol Behav*, 152(Pt A):225–230, 2015.

[110] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddharaju, R. Joseph, A. Mandapati, B. Dunkin, and B. Bass. Fast by Nature - How Stress Patterns Define Human Experience and Performance in Dexterous Tasks. *Scientific Reports*, 2:305, 2012.

[111] How the temperature of your nose shows how much strain you are under - The University of Nottingham. https://www.nottingham.ac.uk/news/pressreleases/2018/january/how-the-temperature-of-your-nose-shows-how-much-strain-you-are-under.aspx. Access date: 2018-12-20.

[112] C. L. etitia Lisetti and F. Nasoz. Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *EURASIP J. Appl. Signal Process.*, 2004:1672–1687, 2004.

[113] G. G. Youngson. Nontechnical skills in pediatric surgery: Factors influencing operative performance. *Journal of Pediatric Surgery*, 51(2):226–230, 2016.

[114] T. Haidegger. Autonomy for surgical robots: Concepts and paradigms. *IEEE Trans. on Medical Robotics and Bionics*, 1(2):65–76, 2019.

Table 1

Automated surgical skill assessment techniques in RAMIS. Used abbreviations: HMM: Hidden Markov Model, LDA: Linear Discriminant Analysis, GMM: Gaussian Mixture Model, PCA: Principal Component Analysis, SVM: Support Vector Machines, LDS: Linear Dynamical System, NN: Neural Network.

Aim	Input data	Data collection	Training task	Technique	Year	Ref.
kinematic data-based skill assessment	completion time, total distance traveled, speed, curvature, relative phase	da Vinci API	dry lab (bimanual carrying, needle passing, suture tying)	dependent and independent t-tests	2009	[63]
framework for skill assessment of RAMIS training	stereo instrument video, hand and instrument motion, buttons and pedal events	da Vinci API	dry lab (manipulation, suturing, transection, dissection)	PCA, SVM	2012	[55]
examine the effect of teleoperation and expertise on kinematic aspects of simple movements	position, velocity, acceleration, time, initial jerk, peak speed, peak acceleration, deceleration	magnetic pose tracker	dry lab (reach, reversal)	2-way ANOVA	2013	[64]
longitudinal study tracking robotic surgery trainees	basic kinematic data, torque data, events from pedals, buttons and arms, video data	da Vinci API	dry lab (suturing, manipulation, transection, dissection)	SVM	2013	[75]
generate an objective score for assessing skill in gestures	basic kinematic and video data	JIGSAWS	dry lab (suturing, knot tying)	SVM	2014	[90]
discriminate expert and novice surgeons based on kinematic data	completion time, path length, depth perception, speed, smoothness, curvature	da Vinci API	dry lab (suturing)	logistic regression, SVM	2016	[65]
instrument vibrations-based skill assessment	completion time, instrument vibrations, applied forces	da Vinci API	dry lab (peg transfer, needle pass, intracorporeal suturing)	stepwise regression	2016	[62]
automatic skill evaluation based on the contact force	contact forces, robot arm accelerations, time	da Vinci and Smart Task Board	peg transfer	regression and classification	2017	[73]
skill assessment based on instrument orientation	time, path length, angular displacement, rate of orientation change	da Vinci Research Kit	dry lab (needle driving)	2-way ANOVA	2017	[66]
discriminate expert and novice surgeons based on kinematic data	completion time, path length, depth perception, speed, smoothness, curvature, turning angle, tortuosity	da Vinci API	dry lab (suturing, knot-tying)	k-Nearest Neighbor, logistic regression, SVM	2018	[67]
skill score prediction	sequential motion texture, discrete Fourier transform, discrete cosine transform and approximate entropy	JIGSAWS	dry lab (suturing, knot tying, needle passing)	nearest neighbor classifier, support vector regression	2018	[68]

Aim	Input data	Data collection	Training task	Technique	Year	Ref.
objective skill level assessment based on metrics associated with stylistic behavior	basic kinematic and physiological data	limb inertial measurments unit, electromagnetic joint position tracker, EMG, GSR, IMU, cameras	da Vinci Skills Simulator tasks (ring and rail, suture sponge)	crowd sourced analysis	2018	[74]
characterization of open and teleoperated suturing movement	speed, curvature, torsion of movement trajectories	da Vinci Research Kit, JIGSAWS	dry lab (suturing)	fitting the one-sixth power law, types of ANOVA	2018	[70]
assess expertise and recognize surgical training activity	basic kinematic data	JIGSAWS	dry lab (suturing, knot-tying, needle-passing)	multi-output deep learning architecture	2018	[69]
evaluate skills based on kinematic data	time, errors, movement speed, jerkiness, trajectory redundancy, target scoring, trajectory volatility, max deviation	MicroHand S, magnetic sensor	dry lab (pick and place, ring threading)	one-way ANOVA	2018	[71]
evaluate skills based on a deep learning model	basic kinematic data	JIGSAWS	dry lab (suturing, knot tying, needle passing)	deep convolutional NN	2018	[72]
gesture classification	basic kinematic data	da Vinci API	dry lab (suturing)	local feature extraction, LDA, Bayes classifier	2006	[78]
gesture classification and recognition	basic kinematic data	da Vinci API	dry lab (suturing)	LDA, strawman GMM, 3-state HMM	2008	[79]
compare task versus subtask	basic kinematic data	da Vinci API	dry lab (suturing)	vector quantization, HMM	2009	[77]
gesture classification	basic kinematic data and video contextual cues (suture line deformations)	da Vinci API	dry lab (suturing)	HMM, high-order polynomial fitting to the extracted suturing line	2012	[88]
gesture classification and recognition	basic kinematic data	da Vinci API	dry lab (suturing)	LDA, HMM	2009	[80]
gesture classification	basic kinematic data	JIGSAWS	dry lab (suturing, knot-tying, needle passing)	sparse HMM	2012	[81]
gesture classification	video features (image intensities, image gradients, optical flow)	JIGSAWS	dry lab (suturing, needle passing, knot tying)	LDS, bag-of-features, multiple kernel learning	2012	[87]
gesture classification	basic kinematic data and video features (Space-Time Interest Points)	JIGSAWS	dry lab (suturing, needle passing, knot tying)	LDS, bag of features, multiple kernel learning	2018	[89]
gesture classification	basic kinematic data	JIGSAWS, da Vinci Surgical System	dry lab (suturing)	descriptive curve coding, common string model, SVM	2013	[82]
gesture classification	basic kinematic data	JIGSAWS	dry lab (suturing, needle passing, knot tying)	Shared Discriminative Sparse Dictionary Learning, SVM, HMM	2015	[83]

Aim	Input data	Data collection	Training task	Technique	Year	Ref.
providing individualized feedback to surgical trainees	basic kinematic data	n/a	dry lab (suturing, knot tying)	automatic identification of motifs in the tool motion signal	2015	[94]
segmentation of surgical tasks into smaller phases	basic kinematic and video data	JIGSAWS	dry lab (suturing, knot tying)	binary classifier, crowd-sourced segment ratings	2015	[95]
unsupervised segmentation of surgical tasks into smaller phases	basic kinematic (position) and video (object grasp events and surface penetration) data	da Vinci Research Kit	dry lab (pattern cutting, suturing, needle passing)	milestone learning with Dirichlet Process Mixture Models	2015	[96]
recognizing surgical activities	basic kinematic data	JIGSAWS	dry lab (suturing)	Recurrent NN	2016	[98]
gesture classification and recognition	basic kinematic data	Raven-II, Sigma 7	peg transfer	unsupervised trajectory segmentation, k-Nearest Neighbors, SVM	2016	[84]
describe differences in task flow	basic kinematic and video data	da Vinci API	dry lab (suturing, knot tying)	hierarchical semantic vocabulary	2016	[99]
gesture classification	basic kinematic and video data	JIGSAWS	dry lab (suturing, knot tying, needle passing)	HMM, Sparse HMM, Markov semi-Markov Conditional Random Field, Skip-Chain CRF, Bag of spatiotemporal Features, LDS	2017	[91]
temporal clustering of surgical activities	basic kinematic and video data	n/a	live surgery (two-handed robotic suturing, uterine horn dissection, suspensary ligament dissection, running robotic suturing, rectal artery skeletonization and clipping)	Hierarchical Aligned Cluster Analysis, Aligned Cluster Analysis, Spectral Clustering, GMM	2017	[100]
gesture classification and recognition	basic kinematic and video data	da Vinci SKILLS Simulator, SIMIMotion motion capture system	simulated tasks (peg tranfer, pick and place)	Decision Tree Algorithm Model	2012	[92]
gesture classification	basic kinematic data	JIGSAWS	dry lab (suturing, needle passing)	Transition State Clustering, uses hierarchical Dirichlet Process GMM	2018	[85]
gesture classification	basic kinematic data	JIGSAWS/RAVEN-II, Sigma.7, leap motion device/dataset of micro-surgical suturing tasks captured using a dedicated robot	dry lab (suturing, needle passing, knot tying/peg transfers/micro-surgical suturing)	Symbolic Aggregate approXimation, Bag of Words, vector space model	2018	[86]

Aim	Input data	Data collection	Training task	Technique	Year	Ref.
action segmentation and recognition	kinematic (end effector positions, velocity, gripper state, skip-length features) and video (distance to the closest object part from each tool, relative position of each tool to the closest object part) data	JIGSAWS	dry lab (suturing, needle passing, knot tying)	Skip-Chain Conditional Random Field, Deformable Part Model	2015	[93]
action segmentation	basic kinematic and video data	JIGSAWS	dry lab (suturing)	Segmental Spatiotemporal Convolutional NN	2016	[97]