This is a post-peer-review, pre-copyedit version of an article published in the International Journal of Computer Assisted Radiology and Surgery. The final authenticated version is available online at: https://doi.org/10.1007/s11548-018-1713-y.

Surgical skills: Can learning curves be computed from recordings of surgical activities?

Germain Forestier[1,3] · **Laurent Riffaud**[2,4] · **François Petitjean**[3] · **Pierre-Louis Henaux**[2,4] · **Pierre Jannin**[4]

Received: date / Accepted: date

Abstract Purpose Surgery is one of the riskiest and most important medical acts that are performed today. The need to improve patient outcomes and surgeon training, and to reduce the costs of surgery, has motivated the equipment of operating rooms with sensors that record surgical interventions. The richness and complexity of the data that are collected call for new methods to support computer assisted surgery. The aim of this paper is to support the monitoring of junior surgeons learning their surgical skill sets.

Methods Our method is fully automatic and takes as input a series of surgical interventions each represented by a low-level recording of all activities performed by the surgeon during the intervention (*e.g.*, cut the skin with a scalpel). Our method produces a curve describing the process of standardization of the behavior of junior surgeons. Given the fact that junior surgeons receive constant feedback from senior surgeons during surgery, these curves can be directly interpreted as learning curves.

Results Our method is assessed using the behavior of a junior surgeon in anterior cervical discectomy and fusion surgery over his first three years after residency. They revealed the ability of the method to accurately represent the surgical skill evolution. We also showed that the learning curves can be computed by phases allowing a finer evaluation of the skill progression.

Conclusion: Preliminary results suggest that our approach constitutes a useful addition to surgical training monitoring.

Keywords: Surgical data science, Surgical process model, DTW, Learning curves

1 Introduction

A learning curve (or experience curve) is a graphical representation allowing the visual assessment of the increase in skills with time and experience [44]. Learning curve (LC) definition, calculation and analysis have been at the center of many research in social sciences [34], psychology [27] and cognitive sciences [31]. The aim is to better understand how the Human learn how to perform a task. The first person to describe LCs was Hermann Ebbinghaus in 1885. His tests involved memorizing series of nonsense syllables, and recording the success over a number of trials. LC was also studied in economy, such as in the aircraft industry where the amount of man-hours needed to produce a unit decreases as the production increases [43]. LCs generally follow a power law, from which it is often said that they conform to *"the power law of practice"* [34]. As learning curves have been used to evaluate skill acquisition in multiple fields, many researchers started working toward defining it in the medical field [15,13].

For surgical training, it is generally accepted that practical skills are improving with time after an initial period of difficulty followed by an improvement and stabilization of performance [18,22]. This paradigm is following the standard Halstedian system "see one, do one, teach one" [35]. Monitoring the progression in develop-

✉ Germain Forestier
germain.forestier@uha.Fr

[1]IRIMAS, University of Haute-Alsace, Mulhouse, France
[2]Department of Neurosurgery, Univ. Hospital, Univ Rennes, Inserm, LTSI (Laboratoire Traitement du Signal et de l'Image) - UMR_S 1099, F-35000 Rennes, France
[3]Faculty of Information Technology, Monash University, Melbourne, Australia
[4]Univ Rennes, Inserm, LTSI (Laboratoire Traitement du Signal et de l'Image) - UMR_S 1099, F-35000 Rennes, France

ing surgical skills is of major interest from the teaching perspective. However, evaluating progress in surgical training is still very challenging because of how complex surgical processes are, and because of the high degree of specialization that is required (an average of eight to nine years after medical school). For example, while the relative importance of the different factors that cause surgical error is unknown [36], technical skill acquisition has been shown to correlate with a reduction in patient complications [8,4]: Performing the right action at the right time in surgery can greatly influence patient outcome. Training new surgeons is critical for the quality of care and is an important issue from the economical perspective. The training is often provided in a one-on-one scheme between a junior surgeon and his or her senior. This process is expensive and time-consuming, and relies heavily upon the quality of communication between the junior and his or her senior. Assessing surgical practice is mandatory to ensure a smooth expertise transmission between senior and junior surgeons. This assessment requires a consistent understanding of surgical processes and has thus strongly supported the modeling of surgical processes. In recent years, many techniques have been proposed to compute LC for specific intervention types like laparoscopic colorectal surgery [40] or cardiothoracic and vascular surgery [2]. However, there is no consensus on the methods and variables that should be used to compute a LC [18]. Mimicking the aviation industry, a special interest has been given to create simulation environments to train surgeon and to use LC to evaluate their skills on simulators [16,3,12]. Alternatively, recording trajectories of surgical tools on surgical robots [24,23] was also considered to assess surgical skills and training [42,6].

For example, in [14], the authors propose a system based on the automatic analysis of laparoscopic training video to perform an automatic assessment of the trainee skills. This system allows to compute learning curves and to provide automatic feedback.

In this paper, we introduce an automatic method that aims at computing a learning curve from recordings of low-level activities performed by a surgeon during multiple interventions. We use the deviation of the practice at the low level as a proxy for *progress*; we hypothesize that:

1. At the start of the training of a new junior surgeon, his or her surgeries will be relatively different to each other and relatively different to the practice of senior surgeons.
2. As the training progresses, the surgeries should be more and more consistent and tend toward the general behavior of senior surgeons.

In this paper, we show that there is some evidence supporting this hypothesis and that even low-level descriptions of the surgeries can be used for this assessment.

The only information required by our method is the input data: a series of surgeries, where each surgery is represented by a low-level recording of the activities performed by the surgeon over the course of the surgeries (*e.g.*, *cut* the *skin* with a *scalpel*). Our method outputs a curve that describes the standardization process of the junior's surgical practice over time. Note that our method can easily be used in conjunction with any system able to recognize surgical activities [19]. We carry out experiments on data recorded in operating room and composed of 26 anterior cervical discectomy and fusion surgery recorded at the Neurosurgery Department of the Rennes University Hospital, France. The surgeries were performed by a junior surgeon over his first 3 years after residency and were all recorded by the same senior surgeon.

This paper is organized as follows. In "Surgical learning curves" section, we present related work on learning curves for surgery. In "Methods" section, we introduce our method for the automatic computation of learning curves from low-level surgical activities. In "Results" section, we conduct experiments that demonstrate the quality and performance of our approach. Finally, we conclude this work and describe future research in "Conclusion" section.

2 Surgical learning curves

The concept of learning curve is particularly interesting in surgery where the skill-set to master is important and the training generally last for years. It would indeed be useful to know how many interventions a surgeon have to perform before reaching an adequate and safe level of expertise. An important consideration when computing a LC is the variable that is studied to create the curve [18]. There are two main types of variables: (1) measuring the surgical process or (2) measuring patient outcomes [18]. Measures of surgical process include variables such as time to complete the procedure, the number of surgical actions, the success or completion rate of the procedure, etc. [5]. Measures of patient outcomes include the length of hospital stay, postoperative complications, mortality, etc. [17].

Operation time is one of the most used variables to measure the acquisition of skills. Multiple studies showed that senior surgeons are on average faster than junior surgeon [32]; this has led to using time as a proxy for seniority. However, it is obviously very undesirable for a junior to try to speed up his or her surgery for assessment purposes alone, without having reached the

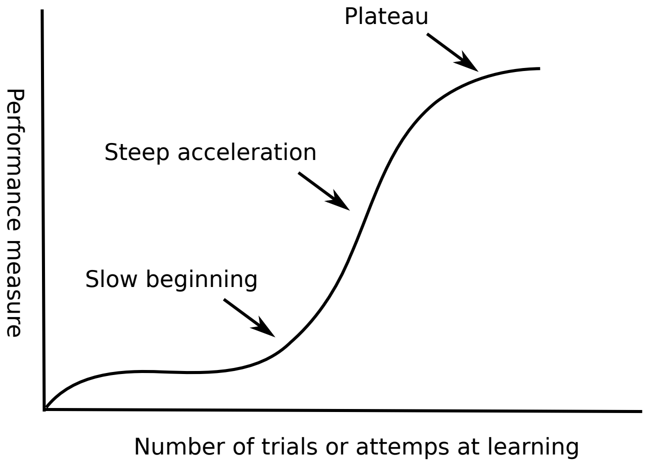


Fig. 1: Illustration of a learning curve [7].

dexterity and experience of senior surgeons. Operation time is also not well defined [5]; the most standard definition seems to be from the start at the incision to closure of the wound [1]. The problem of definition is also present with what variables to use for patient outcome [18].

Despite the growing number of initiatives in this field, there is no consensus on the variables to use and their definition [18]. Furthermore, all these measures are very general variables that only partially represent what the surgeon did during the surgery. One way to automatically assess skills is also to rely on existing evaluation skill methodologies like the Objective Structured Assessment of Technical Skill (OSATS) [41] test. This test is composed of scales to score the skills of surgical trainee who have to be assessed by an observer. Recent works try to automate this evaluation using video analysis [39]. Once automated, the automatic grading using OSATS could be use to grade the evolution of the performance of surgical trainee.

Somewhat surprisingly, only little attention has been put on the development of automatic methods to compute LC from sensors present in the operating rooms (ORs). Indeed, more and more ORs are getting equipped systems with sensing devices that can capture the surgeon's activities and environment. For example, in [38], the authors extracted descriptions of surgical processes and identified relationships between the course of a surgical process and the quality of its outcome. In total, 450 training sessions were manually recorded and compared to expert evaluation as per the quality of the sessions. Video processing has also being investigated to automatically evaluate surgical processes. For example, using cameras in pituitary surgery, both the phases of the surgery [21] and the low-level surgical tasks [19] can be detected and recorded automatically. In [19], using a dataset of 20 cataract surgeries, and identifying 25 possible pairs of activities, a frame-by-frame recognition rate of 64.5% was achieved with the proposed system. The surgical phases can also be predicted from low-level activities [11]. The task performed by the surgeon can also be automatically inferred by combining RFID chips on instruments (for identification) with accelerometers [29].

3 Methods

3.1 Surgery as sequence of surgical activities

The data captured in the OR to represent surgery have a specific granularity level. A granularity level is defined as the level of abstraction at which the surgical procedure is described. MacKenzie et al. [25] were the first to propose a model of the surgical procedure that consists of different levels of granularity: the procedure, the step, the substep, the task, the subtask and the motion. Later, Lalys and Jannin [20] introduced a terminology consisting of phase defined as the major types of events occurring during surgery. Each phase is composed of several steps. A step is considered to be a sequence of activities used to achieve a surgical objective. The data used in this paper capture the activity of both hands for three different elements: *used instrument*, *performed action* and *targeted anatomical structure* [28]. Note that the recordings of the performed activities only partially represent the process of surgery. As surgery is a complex task, it involves a difficult decision-making process influenced by multiple factors. Thus, acquiring data that *represent* surgery is very challenging. In this work, we decided to focus surgical activities as they make possible the assessment of *procedural knowledge*. Procedural knowledge only partially covers the skill-set required to master surgery, as it also includes conceptual knowledge, cognitive skills, interpersonal skills, etc. However, monitoring technical or procedural skill acquisition is important, as these are shown to correlate with a reduction in patient complications [8]. One can note that we currently consider all actions to be of equivalent importance. While it would be interesting to take into account the importance, quality or precision of actions, these characteristics are still very difficult to assess for a single action in the context of an entire surgery.

3.2 Proposed method for computing learning curve

The goal of our method is to evaluate how surgical practice evolves through time. Let $\mathbb{S} = \{S_1, \cdots, S_N\}$ be the set of N surgeries (*i.e.*, sequences of activities) performed by the same surgeon and ordered by increasing operating date, and $S = \langle s_1, \cdots, s_l \rangle$ be one sequence of this set. We propose to use the evolution through time of the heterogeneity inside the set $\mathbb{S}$ in order to assess the evolution of surgical practice. The assumption is that junior surgeons, while learning how to operate, do not have an homogeneous practice. Indeed, when learning how to perform a task, the first attempts are generally different from each other. However, with time and practice, the tasks as they are repeated tend to be performed in a similar way. This tendency has been observed previously while comparing the surgical behaviors of junior and senior surgeons [10]. To evaluate this phenomena, we proposed to study the evolution through time of the heterogeneity (*i.e.*, the dispersion) inside a set of surgeries $\mathbb{S}$. The heterogeneity H of a set of surgeries is defined as:

$$H(\mathbb{S}) = \frac{1}{|\mathbb{S}|^2 - |\mathbb{S}|} \sum_{S_i \in \mathbb{S}} \sum_{S_j \in \mathbb{S}}^{S_j \neq S_i} \mathrm{sim}(S_i, S_j) \tag{1}$$

It corresponds to the average similarity between all pairs of surgery present in the set. A low value will indicates that, on average, the surgeries are similar, while a high value will indicate that they are different.

In order to compute the heterogeneity (Eq. 1), a similarity measure sim between surgeries (*i.e.*, sequences of low-level surgical activities) has to be defined. Following previous work on comparing surgical processes [9], we used DTW (dynamic time warping) [37] to evaluate the similarity between surgeries. The DTW similarity measure makes it possible to find the optimal alignment of two sequences (and thus register them) and provide an alignment score that we used as an assessment of the similarity between the sequences. The similarity function used between two surgical activities weighs each of the three components (action, anatomical structure and instrument) equally by 1/3 [9].

To compute a learning curve using this measure, we create a set of sets of surgeries, $\mathcal{S} = \{\mathbb{S}_1, \cdots, \mathbb{S}_M\}$, by using the date of the interventions as a partial ordering in $\mathbb{S}$ (*i.e.*, S_0 is the oldest recording and S_N the most recent one). The first set, $\mathbb{S}_0$, is composed of the two oldest surgeries: $\{S_0, S_1\}$ (*i.e.*, at least two elements are needed to compute $H(\mathbb{S})$). Then, new sets are created by adding one by one the recordings according to the intervention dates (*i.e.*, $\{S_0, S_1\}$, $\{S_0, S_1, S_2\}$, $\{S_0, S_1, S_2, S_3\}$, etc.). The last set, $\mathbb{S}_M$, contains all N recordings.

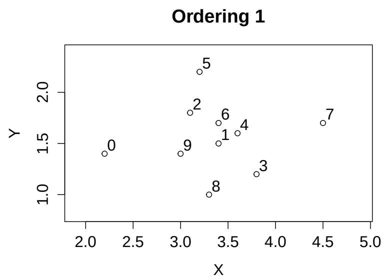
The heterogeneity (Eq. 1) is then computed for each set of $\mathcal{S}$ in order to create the points of the curve. Regression can then be used to compute the learning curve (*e.g.*, least-squares regression, logarithmic or negative exponential curves etc.) [33]. The squared residual of the regression can then be used as a proxy for the correctness of the learning curve. A low value indicates that the learning progression is very smooth and progressive as it means that the polygon is a good approximation.

3.3 Illustrative example of learning curve computation

In this section, we illustrate on a simple example how the proposed method works. For simplicity, we run this example on ten simple data points (x, y), each point representing one surgery. We investigate how the ordering of these ten data points influence our method in building the learning curve. Figure 2a, c, and e illustrates three different orderings of the ten data points: Ordering 1, Ordering 2 and Ordering 3. The number associated with each point is used to sort the data points, like the dates are used to sort the surgeries. Zero (0) indicates the first point, and nine (9) indicates the last point. In Figure 2a, the data points were randomly placed. In Figure 2c, the data points were sorted inside out, while in Figure 2a, the data points were sorted outside in. To apply our method and to compute learning curves, we used the heterogeneity (*i.e.*, Eq. 1) with the euclidean distance as a dissimilarity metric. In order to compute the learning curve, the heterogeneity was computed for each set of sets following the three orderings. Figures 2b, 2d, and 2f show the learning curves for the three orderings. As one can see, when the ordering is random (Figure 2b) the learning curve does not have a distinctive trend. The heterogeneity values increase and decrease randomly. Figures 2d, and 2f are presenting the two extreme cases, in Figures 2d the heterogeneity values are continuously decreasing while in 2f they are continuously increasing. Intuitively, the shape of Figure 2f is the trend we are targeting to evaluate the acquisition of skills as its trend correspond to what it is expected from a learning curve (*i.e.*, as presented in Figure 1). This simple example shows that depending on the ordering of the elements, the learning curve that is produced can have different shapes. Thus, these curves can be used to assess the fact that the set of surgeries tend to be more homogeneous (*i.e.*, if the average distance between them is decreasing).

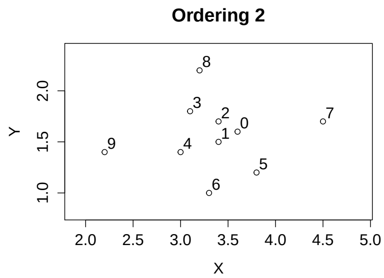
4 Results

The framework was evaluated using clinical data composed of 26 cervical disc herniation surgeries recorded at the Neurosurgery Department of the Rennes University Hospital, France. This procedure is very standard-

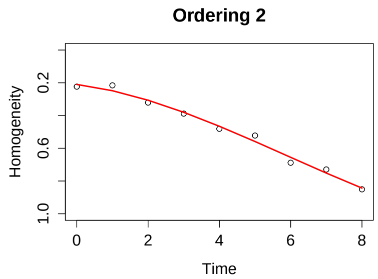


(a) Ordering 1.

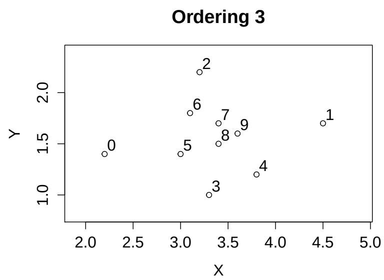
(b) Learning curve for Ordering 1.



(c) Ordering 2.



(d) Learning curve for Ordering 2.



(e) Ordering 3

(f) Learning curve for Ordering 3

Fig. 2: Illustrative example with 3 orderings (a,c,e) of 10 data points and their corresponding learning curves (b,d,f). a Ordering 1. b Learning curve for Ordering 1. c Ordering 2. d Learning curve for Ordering 2. e Ordering 3. f Learning curve for Ordering 3.

ized: The same techniques, instruments and synthetic implants were used for the 26 recordings. The surgeries involved 15 male and 11 female patients, with a median age of 52 years. These cervical disc surgeries are divided into five main steps: (1) approach of the disc, (2) discectomy. (3) hemostasis, (4) arthrodesis and (5) closure. Depending on the patient, multiple hemostasis phases are required. The herniated disc is approached via a right anterior cervical route. The surgeries were performed by a junior neurosurgeon over its first 3 years after residency. The recordings were performed by the same senior surgeon using the ICCAS surgical workflow editor [30]. A total of 693 days passed between the first and the last recordings. The number of days between two recordings was on average of 28 days with a maximum of 119 days.

For this surgery, the list of actions is: *cut, swab, sew, coagulate, install, dissect, irrigate, drill, remove* and *hold*. The list of anatomical structures is: *muscle, vertebra, skin, fascia, disc* and *ligament*. And the list of surgical instruments is: *scalpel, needle-holders, curettes, hooks, rongeurs, high-speed-drill, arthrodesis, dissectors, drainage, scissors, suction tube, forceps, saline solution, retractors* and *cottonoids*. Our dataset contains 87 different activities; note that all triples are not present (Some triples of action, instrument, anatomical

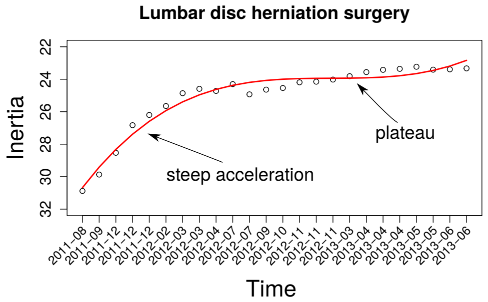


Fig. 3: Learning curve for 26 cervical disc herniation surgeries.

structure are irrelevant.). The decomposition in phases was decided by the observer during the recording of the surgical activities.

The Figure 3 presents the learning curve for the 26 cervical disc herniation surgeries. The year and the month of the operating dates are provided for each surgery. The specific day in the month was omitted to preserve anonymity. This curve was computed using the whole surgeries as input without using the phase information. Figure 4 presents the learning curves for 17 surgeries subdivided into surgical phases. These surgeries were selected because they had the exact same number of phases (*i.e.*, only one hemostasis phase). For each phase, the learning curve was computed by only using the surgical activities that were performed during this specific phase. The average number of activities is 44 for the approach phase, 62 for the disectomy phase, 16 for the hemostasis phase, 7 for the arthrodesis phase and finally 10 for the closure phase.

5 Discussion

The learning curve for the entire produce (Figure 3) shows a smooth progression before reaching a plateau. It exhibits the classical LC pattern: an initial period of difficulty followed by an improvement and stabilization of performance. Note that only 25 sets were evaluated from a total of 26 surgeries. Indeed, a set has to contain at least two elements to compute the heterogeneity (Eq. 1). For visualization purpose, the y-axis has been inverted as a reduction in the heterogeneity is interpreted as a skill progression. The heterogeneity started from 30.88 and ended at 23.33 which means a reduction of 24%. The sum of the residuals of the regression is equal to 8.05. This value is difficult to interpret from a single curve, but could be used to compare multiple surgeons to each others.

The Figure 4 presents learning curves according to the different surgical phases for 17 surgeries. Decomposing a learning curve according to phases allows to perform a more precise analysis of skills progression. In the following, we analyzed the learning curves phase by phase.

The LC of the *Approach* phase (Figure 4 (a)) has a disrupted trend. Indeed, if we consider all the points (continuous curve), the regression does not show a clear increase in the skills and is almost flat. However, if we focus on the first half (except the first recording) and the second half by performing two distinct regressions (dashed curves), we obtain two learning curves with a very slight trend of skill progression. The approach phase is very standard and requires less technical skill than the other phases of the surgery. This can explain the lack of clear skill progression on the learning curve. Furthermore, the approach phase is also similar in other types of surgery that might have been performed by the trainee (and not used in our study). The flat behavior of the curve indicates that the surgeon performed all the approach phase in a similar way.

The LC of the *Disectomy* phase (Figure 4 (b)) has a similar behavior than the entire procedure (Figure 3). The *Disectomy* phase is the most important and difficult phase of this type of surgery. It is also the longest phase and it is known to be the most characteristic phase of the surgeon technique. The *Disectomy* learning curve reveals a smooth progression of the junior surgeon technique over time. This progression is similar to the progression witnessed by the senior surgeon that performed the recordings. The *Hemostasis* phase (Figure 4 (c)) has also a similar trend.

The LC of the *Arthrodesis* phase (Figure 4 (d)) has a more hectic behavior if we consider all the interventions (continuous curve). If we consider only the first half (and removing the first two), the learning curve (dashed curve) exhibits a reduction in the heterogeneity. However, the addition of the following surgeries increased the heterogeneity of the set. This result shows that adding more surgeries to a set does not always increase its heterogeneity. The specific behavior of the *Arthrodesis* learning curve could be explained by the inner variability of this phase. Indeed, this phase relies heavily on the scrub nurse skills (not recorded in this work) who has the responsibility to propose the implant. Furthermore, this phase contains a very limited number surgical activities (7 on average). Thus, small variations can have an important influence while comparing the sequences. It is thus difficult to have a clear explanation for the obtained trends.

Finally, the LC of the *Closure* phase (Figure 4 (e)) has a similar behavior than the LC of the entire procedure expect for the two first interventions. We computed the regression with all the interventions (contin-

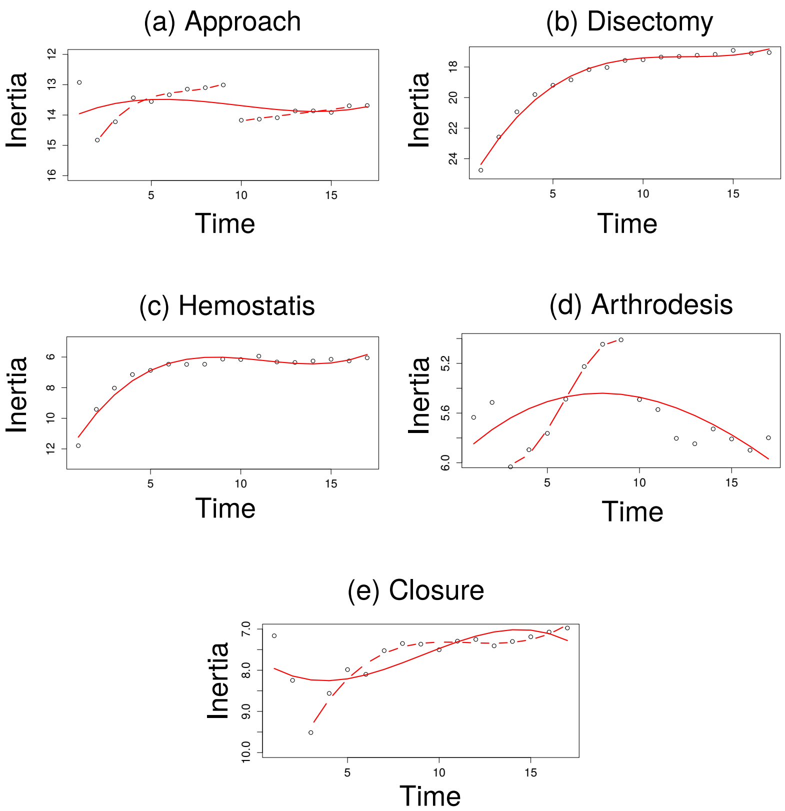


Fig. 4: Learning curves for 17 lumbar disc herniation surgeries by phases : Approach, Disectomy, Hemostasis, Arthrodesis and Closure. Continuous curves show the results of regressions using all the interventions and dashed curves only consider a subset of interventions.

uous curve) and all but the two first (dashed curve). This result can be explained by specificities in patient anatomy for these two interventions.

All these results show that our method is able to correctly assess the evaluation of surgical skills. Note that the source code implementing the method is distributed as an open-source software and is available for download[1]. A web application is also available in the same page to illustrate how the method works.

One can note that getting more experienced can also mean moving away from homogeneity when required. This is why we selected the procedures that share common features (*e.g.* patient age, difficulty of the cases, etc.). The goal of our system is to evaluate the acquisition of the core skills that a young surgeon should master. Thus, we focused this study on evaluating the acquisition of skills as the reduction in the heterogeneity in a set of performed surgeries. However, differences in patient anatomy, but also emergencies, complications and other nonstandard occurrences can make deviation from the standard surgical behavior the right thing to do. This is not currently handled by the proposed method, which currently relies on internal evaluation. It means that we only consider the surgeries performed by the evaluated surgeon to compute the learning curve.

An alternative would be to compare the behavior of the trainee with a database of recorded and annotated surgeries and use the relative distance to "expert behaviors" as a proxy to evaluate the acquisition of skills. Comparing sequences of surgical activities has already been investigated in the previous work [9,10], but never with the goal of computing learning curves.

As future work, we are planning to acquire additional datasets with more formal evaluation of the junior surgeons (*e.g.* OSATS [26] results) in order to assess the correlation between automatically computed learning curves and these formal evaluations.

6 Conclusion

In this paper, we presented a method to automatically compute learning curves from recordings of low-level surgical activities. We used the evolution of the heterogeneity as a criteria to evaluate the skill progression of surgical practice. Experiments were performed on 26

[1] http://germain-forestier.info/src/ipcai2018/

anterior cervical discectomy and fusion surgery. They revealed the ability of the method to accurately represent the surgical skill evolution. We also showed that the learning curves can be computed by phases allowing a finer evaluation of the skill progression.

In future work, we are planing to take into account the time gap between recordings. We also want to go further into the analysis in order to identify more precisely subsequences of activities that influence skill progression. Finally, we are planning to study the correlation of our results with classical techniques of skill assessment like OSATS [26].

Conflict of Interest: The authors declare that they have no conflict of interest.
Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.
Informed consent: Informed consent was obtained from all individual participants included in the study.

Acknowledgements Dr François Petitjean is the recipient of an Australian Research Council Discovery Early Career Award (Project Number DE170100037) funded by the Australian Government.

References

1. T. Akiyoshi, H. Kuroyanagi, M. Ueno, M. Oya, Y. Fujimoto, T. Konishi, and T. Yamaguchi. Learning curve for standardized laparoscopic surgery for colorectal cancer under supervision: a single-center experience. *Surgical endoscopy*, 25(5):1409–1414, 2011.
2. K. S. Arora, N. Khan, H. Abboudi, M. S. Khan, P. Dasgupta, and K. Ahmed. Learning curves for cardiothoracic and vascular surgical procedures-a systematic review. *Postgraduate medicine*, (0):1–13, 2014.
3. J. Barrie, D. G. Jayne, J. Wright, C. J. C. Murray, F. J. Collinson, and S. H. Pavitt. Attaining surgical competency and its implications in surgical clinical trial design: a systematic review of the learning curve in laparoscopic and robot-assisted laparoscopic colorectal cancer surgery. *Annals of surgical oncology*, 21(3):829–840, 2014.
4. J. D. Birkmeyer, J. F. Finks, A. O'Reilly, M. Oerline, A. M. Carlin, A. R. Nunn, J. Dimick, M. Banerjee, and N. J. Birkmeyer. Surgical skill and complication rates after bariatric surgery. *New England Journal of Medicine*, 369(15):1434–1442, 2013.
5. D. H. Choi, W. K. Jeong, S.-W. Lim, T. S. Chung, J.-I. Park, S.-B. Lim, H. S. Choi, B.-H. Nam, H. J. Chang, and S.-Y. Jeong. Learning curves for laparoscopic sigmoidectomy used to manage curable sigmoid colon cancer: single-institute, three-surgeon experience. *Surgical endoscopy*, 23(3):622–628, 2009.
6. 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*, 2015.
7. R. A. Dewey. *Psychology: an introduction.* Russ Dewey, 2007.
8. B. J. Dlouhy and R. C. Rao. Surgical skill and complication rates after bariatric surgery. *The New England Journal of Medicine*, 370(3):285–285, 2014.
9. G. Forestier, F. Lalys, L. Riffaud, B. Trelhu, and P. Jannin. Classification of surgical processes using dynamic time warping. *Journal of Biomedical Informatics*, 45(2):255–264, 2012.
10. G. Forestier, F. Lalys, R. Riffaud, L. Collins, J. Meixensberger, S. N. Wassef, T. Neumuth, B. Goulet, and P. Jannin. Multi-site study of surgical practice in neurosurgery based on surgical process models. *Journal of Biomedical Informatics*, 46(5):822–829, 2013.
11. G. Forestier, L. Riffaud, and P. Jannin. Automatic phase prediction from low-level surgical activities. *International journal of computer assisted radiology and surgery*, pages 1–9, 2015.
12. M. Hanzly, A. Frederick, T. Creighton, K. Atwood, D. Mehedint, E. C. Kauffman, H. L. Kim, and T. Schwaab. Learning curves for robot-assisted and laparoscopic partial nephrectomy. *Journal of Endourology*, 2014.
13. A. Hopper, M. Jamison, and W. Lewis. Learning curves in surgical practice. *Postgraduate Medical Journal*, 83(986):777–779, 2007.
14. G. Islam, K. Kahol, B. Li, M. Smith, and V. L. Patel. Affordable, web-based surgical skill training and evaluation tool. *Journal of biomedical informatics*, 59:102–114, 2016.
15. C. Jackson and K. Gibbin. 'per ardua...'training tomorrow's surgeons using inter alia lessons from aviation. *Journal of the Royal Society of Medicine*, 99(11):554–558, 2006.
16. R. M. Jiménez-Rodríguez, J. M. Díaz-Pavón, F. d. l. P. de Juan, E. Prendes-Sillero, H. C. Dussort, and J. Padillo. Learning curve for robotic-assisted laparoscopic rectal cancer surgery. *International Journal of Colorectal Disease*, 28(6):815–821, 2013.
17. J.-C. Kang, S.-W. Jao, M.-H. Chung, C.-C. Feng, and Y.-J. Chang. The learning curve for hand-assisted laparoscopic colectomy: a single surgeon's experience. *Surgical endoscopy*, 21(2):234–237, 2007.
18. N. Khan, H. Abboudi, M. S. Khan, P. Dasgupta, and K. Ahmed. Measuring the surgical learning curve: methods, variables and competency. *BJU International*, 113(3):504–508, 2014.
19. F. Lalys, D. Bouget, L. Riffaud, and P. Jannin. Automatic knowledge-based recognition of low-level tasks in ophthalmological procedures. *International journal of computer assisted radiology and surgery*, 8(1):39–49, 2013.
20. F. Lalys and P. Jannin. Surgical process modelling: a review. *International Journal of Computer Assisted Radiology and Surgery*, 8(5):1–17, 2013.
21. F. Lalys, L. Riffaud, X. Morandi, and P. Jannin. Automatic phases recognition in pituitary surgeries by microscope images classification. In *Information Processing in Computer-Assisted Interventions*, pages 34–44. Springer, 2010.
22. P.-J. Le Reste, P.-L. Henaux, L. Riffaud, C. Haegelen, and X. Morandi. Influence of cumulative surgical experience on the outcome of poor-grade patients with ruptured intracranial aneurysm. *Acta neurochirurgica*, 157(1):1–7, 2015.
23. H. C. Lin, I. Shafran, T. E. Murphy, A. M. Okamura, D. D. Yuh, and G. D. Hager. Automatic detection and segmentation of robot-assisted surgical motions. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pages 802–810. Springer, 2005.
24. 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.
25. L. MacKenzie, J. Ibbotson, C. Cao, and A. Lomax. Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. *Minimally Invasive Therapy & Allied Technologies*, 10(3):121–127, 2001.

26. J. Martin, G. Regehr, R. Reznick, H. MacRae, J. Murnaghan, C. Hutchison, and M. Brown. Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery*, 84(2):273–278, 1997.
27. J. E. Mazur and R. Hastie. Learning as accumulation: a reexamination of the learning curve. *Psychological Bulletin*, 85(6):1256, 1978.
28. N. Mehta, R. Haluck, M. Frecker, and A. Snyder. Sequence and task analysis of instrument use in common laparoscopic procedures. *Surgical endoscopy*, 16(2):280–285, 2002.
29. C. Meißner, J. Meixensberger, A. Pretschner, and T. Neumuth. Sensor-based surgical activity recognition in unconstrained environments. *Minimally Invasive Therapy & Allied Technologies*, 2014.
30. T. Neumuth, N. Durstewitz, M. Fischer, G. Strauß, A. Dietz, J. Meixensberger, P. Jannin, K. Cleary, H. U. Lemke, and O. Burgert. Structured recording of intraoperative surgical workflows. In *Medical imaging*, pages 61450A–61450A. International Society for Optics and Photonics, 2006.
31. S.-H. PARK1a, I. H. Suh, J.-h. Chien, J. Paik, F. E. Ritter, D. Oleynikov, and K.-C. Siu. Modeling surgical skill learning with cognitive simulation. *Medicine Meets Virtual Reality 18: NextMed*, 163:428, 2011.
32. 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, 2012.
33. C. R. Ramsay, A. Grant, S. Wallace, P. Garthwaite, A. Monk, and I. Russell. *Statistical assessment of the learning curves of health technologies.* Core Research, 2001.
34. F. E. Ritter and L. J. Schooler. The learning curve. *International encyclopedia of the social and behavioral sciences*, 13:8602–8605, 2001.
35. J. Rodriguez-Paz, M. Kennedy, E. Salas, A. Wu, J. Sexton, E. Hunt, and P. Pronovost. Beyond “see one, do one, teach one”: toward a different training paradigm. *Quality and Safety in Health Care*, 18(1):63–68, 2009.
36. S. O. Rogers, A. A. Gawande, M. Kwaan, A. L. Puopolo, C. Yoon, T. A. Brennan, and D. M. Studdert. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery*, 140(1):25–33, 2006.
37. H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):43–49, 1978.
38. S. Schumann, U. Bühligen, and T. Neumuth. Outcome quality assessment by surgical process compliance measures in laparoscopic surgery. *Artificial intelligence in medicine*, 63(2):85–90, 2015.
39. Y. Sharma, T. Plötz, N. Hammerld, S. Mellor, R. McNaney, P. Olivier, S. Deshmukh, A. McCaskie, and I. Essa. Automated surgical osats prediction from videos. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 461–464. IEEE, 2014.
40. P. P. Tekkis, A. J. Senagore, C. P. Delaney, and V. W. Fazio. Evaluation of the learning curve in laparoscopic colorectal surgery: comparison of right-sided and left-sided resections. *Annals of surgery*, 242(1):83, 2005.
41. P. Van Hove, G. Tuijthof, E. Verdaasdonk, L. Stassen, and J. Dankelman. Objective assessment of technical surgical skills. *British Journal of Surgery*, 97(7):972–987, 2010.
42. B. Varadarajan, C. Reiley, H. Lin, S. Khudanpur, and G. Hager. Data-derived models for segmentation with application to surgical assessment and training. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 426–434. Springer, 2009.
43. T. P. Wright. Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences (Institute of the Aeronautical Sciences)*, 3(4), 2012.
44. L. E. Yelle. The learning curve: Historical review and comprehensive survey. *Decision Sciences*, 10(2):302–328, 1979.