Automated and Objective Assessment of Surgical Training: Detection of Procedural Steps on Videotaped Performances

Siddarth Jain*§, Katherine A. Barsness†‡, Brenna Argall*†§

*Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA
†Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
‡Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
§Rehabilitation Institute of Chicago, Chicago, IL, USA
sidd@u.northwestern.edu, kbarsness@luriechildrens.org, brenna.argall@northwestern.edu

***Abstract*—With the rapid growth in competency-based performance requirements for medical resident education, there is a critical need for validated assessments of technical skills. Skill evaluation today is predominantly based on manual evaluations by expert surgeons—which is time-consuming, non-uniform, and cumbersome to work into the natural flow of training or testing events. In this paper, we propose an algorithm for automated recognition of surgical procedural steps using video analysis for the objective assessment of technical skills during surgical training. We employ a bouquet of computer vision techniques, such as template matching, for the automated detection of correct and incorrect surgical procedural steps during tracheoesophageal fistula repair. We consider specifically a simulated model of human infant anatomy used to train surgeons to perform the tracheoesophageal fistula repair procedure. Using a simulated model is key for gaining expertise in repairing this relatively rare but deadly abnormality. Our automated detection approach provides an appropriate, clinically-relevant scenario using a well-designed and validated simulator and produces a uniform, manageable and verifiable solution. The algorithm was validated on nine performances of thoracoscopic tracheoesophageal fistula ligation from surgeons with a broad range of surgical skills. The algorithm result imitates the groundtruth for the evaluations, and thus demonstrates the feasibility of the proposed work for efficient, practical and objective assessment of surgical skill during training.**

I. Introduction

Surgical errors result in thousands of injuries and deaths each year. A study by the Agency for Healthcare Research and Quality (AHRQ) reported over 32,000 surgery-related deaths, placing it among the major causes of death in the US [1]. These can be prevented by appropriate training and skill assessment methods. Existing validated methods of surgical performance assessment are highly subjective, and may suffer from high inter-observer variability in scoring. While evaluating surgical instrument motion and use can provide objective measures, there is no assessment as to whether the operation was performed with or without adverse events or errors during the procedure—which in practice are the final metrics used to assess the success of a surgical procedure.

In this work we propose an automated algorithm for the *objective assessment* of technical skills during training through the detection of *correct and incorrect surgical procedural steps* observed from video during the repair of a tracheoesophageal fistula (TEF). A congenital TEF is an aberrant connection between the esophagus (leading to the stomach) and the trachea (leading to the lungs), most often associated with a blind-ending upper esophageal pouch. TEF occurs in 1 out of 2,500-5,000 live births. TEF is a condition that results in infant death within days of birth if left untreated, and moreover is so rare that most pediatric surgeons perform the procedure only a handful of times over the course of their careers—making TEF repair a prime candidate for simulation training with rigorous assessment feedback. (Here *simulation* refers to a physical model using simulated tissue.) The surgical repair of esophageal atresia (abnormal closure) with a tracheoesophageal fistula (abnormal connection) necessitates ligation (joining) of the fistula flush with the tracheal wall, followed by a sutured anastomosis (connection) between the proximal and distal esophageal segments.

While overall survival for infants born with a TEF has increased in the last two decades, postoperative complications remain a concern for the short and long-term development of these infants—many of which can be directly or indirectly related to specific procedural missteps that occurred during the operation. For example, if the ligation of the fistula leaves a residual esophageal pouch on the trachea, the pooling secretions may lead to recurrent pneumonia and recurrent esophageal fistulae. In this example, the error may be due to a lack of knowledge (cognitive gap), lack of skill (technical gap), or an inability to communicate with an assisting surgeon (nontechnical gap). *The correct identification of these surgeon-specific educational gaps is critical to ensure appropriate, targeted educational interventions* that can ultimately improve the overall outcomes for these infants.

Unfortunately, few opportunities exist to increase a surgeon's cognitive, technical, and non-technical skills for TEF repair. In 2006, the average surgical trainee in pediatric surgery performed a mean of 4.4 TEF repairs [2]. Even more concerning, the average practicing pediatric surgeon performed a mean of only 1.1 repairs per year. Therefore, alternative educational strategies are critical to achieve expert performance for TEF repair. This has prompted the development of a simulated model (Figure 1) of TEF, which is used to educate surgical

978-1-4673-6795-0/15/$31.00 ©2015 IEEE

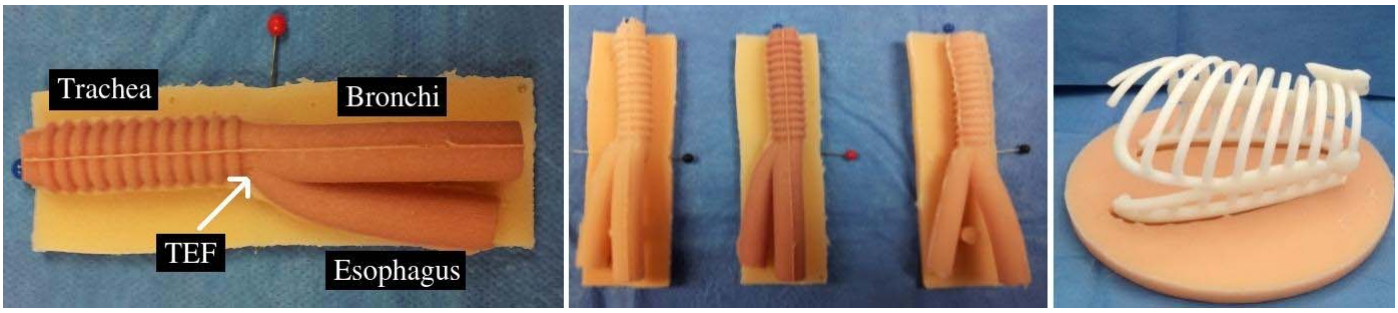


Fig. 1. Left to Right: The TEF simulated model, TEF models in physiologic tracheal shades (i.e. differing in color) and 3-D printed neonatal rib cage.

trainees on the TEF repair procedure [3]. This model is the first and the only validated simulator for the TEF repair procedure.

While the model provides trainees with more experience, the evaluation of their performance remains very labor intensive and, at best, subjective [4]. In addition, existing validated tools (e.g. OSATS [5]) to measure surgeon performance do not identify correct and incorrect procedural steps for specific operations. It was in this context that we developed an automated algorithm for the detection of correct and incorrect procedural steps of TEF repair during training on a size-relevant and anatomically-correct TEF repair simulation model. Our approach produces a uniform, manageable and verifiable solution and provides an appropriate clinically-relevant scenario for skill assessment and surgeon training.

II. Related Work

The apprenticeship model [6] for training surgeons involves direct trainer-trainee interaction and requires observation by an expert to provide feedback to the surgeon in training. This approach is beneficial but the evaluation is time-consuming, inherently subjective, and thus inappropriate for broad based surgical assessment. An extension of the traditional observational approach is the structured grading method that attempts to standardize skill evaluation using operation checklists. These methods involve expert observers and the grading is done using a global assessment chart. Examples of this assessment method are: OSCE [7]; OSATS [5]; and GRS [8]. These tools are popular in training curricula but are time-consuming to use. Furthermore, assignment of a global score to surgical skill for a performance does not provide useful information to the trainees about *where* in the operative procedure they need to improve. This suggests a need for more automated and analytical approaches to evaluate and train surgical skill.

The computational analysis of surgical tool motion offers the potential to assess skill more objectively and effectively than structured human grading. There exists works to support objective measures of surgical performance by evaluating surgical instrument motion and use. Malpani *et al*. [9] present a framework for assessing skill in suturing and knot-tying task using kinematic data. Most of the computational assessment methods on surgical skill evaluation have focused on dynamic or kinematic features—such as time to completion, forces, tool trajectories and velocities [10], [11], [12]. Some methods attach external force sensors to the surgical tools to track position and velocity in order to present a correlation to the surgeon's surgical dexterity level [13], [14]. While motion may differentiate between novice and expert learners, it does *not* provide any information on whether the operation was performed with or without adverse events or errors during the procedure. Additionally, motion metrics validation studies have occurred across a limited range of broadly generalized skills, such as intracorporeal suturing. Chmarra *et al*. [15] present a comparative review on both research and commercial tracking systems for minimally invasive surgery. They conclude that no single tracking system is advantageous for surgery and that the cost of commercial systems on the market is high.

Equally, if not more, important than surgical tool motion and dexterity is the surgeon's ability to make correct intraoperative decisions and the analysis of procedure outcomes. Lalys *et al*. [16] propose a framework for recognition of the phases of cataract surgery, using data provided by microscope videos. They recognize manually-defined visual cues to discriminate high-level tasks for situation recognition. Padoy *et al*. [17] also use low-level image features to process information from surgical tool positions, and Hidden Markov Models to recognize surgical steps during a laparoscopic surgery. Our approach similarly detects surgical steps, here for the particular instance of TEF repair, and additionally is developed specifically for application during surgical training with a simulated model—providing online feedback towards the aim of active learning by the surgeon trainee.

III. Method

This section presents our proposed algorithmic approach to identify the correct and incorrect procedural steps during TEF repair.

A. Simulation Model and Data Collection

The size-relevant, anatomically correct, TEF repair simulation model used in this work, consists of a 3-D printed neonatal rib cage, platinum-cured silicone TEF insert, stabilizing base and silicon skin were used to simulate TEF ligation (Figure 1).

A 4 mm 30 degree telescope (Karl Storz Endoscopy, Segundo, CA) is used for surgical field visualization (Operative Telescope, OT). This video stream is visible to the surgeon-in-training. A separate 4 mm 0 degree telescope (Karl Storz Endoscopy, Segundo, CA) is placed in the lumen of the simulated trachea (Tracheal Telescope, TT), for use by the automated algorithm. The TT video provides a view of both lumens and allows for the clear observation of all procedural steps (described next). This video stream is not visible to the

Fig. 2. The operative setup: Operative Telescope (OT) view used by the surgeon (left monitor) and Tracheal Telescope (TT) view used by the algorithm (right monitor).

surgeon, as this information would not be available during a live surgery. Figure 2 shows the operative setup.

During our evaluation, surgeons with varying levels of experience were provided with 3mm minimally invasive instruments and 4-0 braided suture, and instructed to complete thoracoscopic dissection and ligation of the TEF.

B. Surgical Training and Procedural Steps during TEF Repair

Key procedural steps during TEF repair consist of the following:

> **Complete Ligation (CL)** of the esophageal fistula, which is the desired outcome for the TEF repair.
>
> **Partial Ligation (PL)** of the esophageal fistula is an incorrect procedural step in which the lumen of the esophageal fistula remains partially open. PL is expected to result in postoperative recurrent fistulae, pneumonia(s) and pneumothoraces.
>
> **Bronchial Ligation (BL)** is an incorrect identification of the anatomy, with the right lung bronchus ligated (closed) instead of the fistula. Unrecognized/unrepaired BL leads to severe injury to the right nonventilated lung, with direct aspiration of oral secretions.
>
> **Tracheal Compression (TC)** that is brief and intermittent may be encountered during TEF repair. However when persistent (>8 seconds), tracheal compression adversely interferes with intraoperative oxygenation and ventilation of the normal left lung. Repeated TC also suggests inappropriate identification of the anatomy, which could also lead to devastating injury to the trachea.

In surgical training of TEF repair, the operative goal is complete ligation of the correct structure. Partial ligation means the learner needs additional practice in performing the procedure—a technical gap. Ligation of the bronchus (BL) is a critical error and signifies a cognitive educational gap—the trainee needs to start over with anatomy review before practicing again. (On board examinations, incorrect ligations (PL or BL) result in automatic failure.) In addition to ligation assessment, our approach allows for formative (on-going) feedback about compressions (TC), directly to the learner while practicing.

At the end of the training session, summative feedback provided by the algorithm consists of the total number of tracheal compressions, tracheal compressions longer than 8 seconds, complete or incomplete ligation of the esophagus, and whether there was bronchial ligation. In the case of complete (correct) ligation of esophagus, the trainee additionally may be assessed for where their skills lie on the learning curve, via an assessment informed by the continuous variables relating to tracheal compressions (fewer and less persistent is better) and the time to complete ligation (shorter is better) for the procedure. Converting these metrics into a validated skill score is an active area of our current and future work.

C. Algorithm to Detect Correct and Incorrect Procedural Steps

Our automated algorithm was developed using three videos recorded by an expert surgeon intentionally performing common mistakes and correct procedural steps during TEF repair.

The first step of the algorithm is to *segment the bronchial and esophogeal lumens*. The algorithm then proceeds to detect *tracheal compressions* (TC), followed by detection of the *procedural steps related to ligation* (PL, BL, CL).

To segment the bronchial and esophageal lumens, RGB image frames from the TT video data are converted to a high-contrast grayscale image by eliminating hue and saturation information while retaining the luminance, a binary image is obtained using thresholding and then the complement is applied (Figure 3). Regions of interest then are identified by connecting components and labeling. All the pixels in the 8-connected neighborhood of the current pixel are marked as

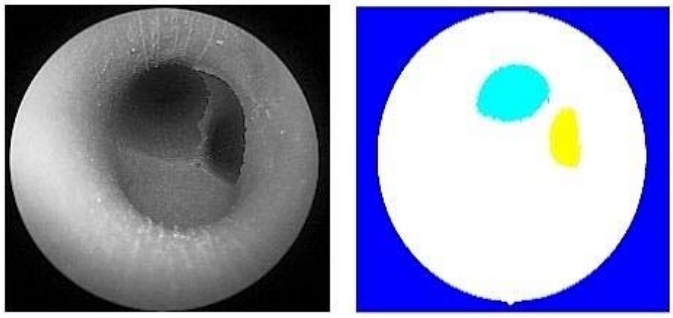


Fig. 3. Bronchial and esophageal lumen segmentation. *Left*: Grayscale image. *Right*: Segmented lumens marked in cyan and yellow color.

belonging to the current region. This procedure first scans row-wise until finding a boundary, and is continued recursively until all pixels belonging to the same region are counted. In the simulated model and surgical setup, the bronchial lumen resides on upper left and the esophageal fistula on the right. The lumens are segmented from the image and labeled (Figure 3, right).

To detect tracheal compressions (TC) is challenging, as they have no particular geometric pattern. We employ a template-matching approach based on the region around the airway in the anatomy for TC detection. The airway region is segmented as a Region of Interest (ROI), and used as a template pattern (Figure 4). In instances of TC the ROI region will be absent from the frame. (The input image and the ROI template are first reduced in size by image pyramids to improve the efficiency of the matching algorithm computations.) We determine the similarity measure of the ROI template image and the matching segment of the input image instance via cross-correlation, which is sum of pairwise multiplications of corresponding pixel values of the images.[1] In detail, Normalized Cross Correlation (NCC) [18] is employed, as the results are invariant to the global brightness changes and the correlation value is scaled to within [-1, 1]. Template matching is implemented in the frequency domain to find the ROI pattern in the input image instance. The pattern is localized by the maximum cross-correlation value, and the threshold on detection is fixed at 0.95. (Again, detection corresponds to a *negative* instance of TC.) A sampling of compression detections is provided in Figure 5).

To detect procedural steps related to ligation is the final step of the algorithm. In the first frame, the diameters (d) and centroids (c) of the segmented bronchial lumen $\langle d_B, c_B \rangle$

[1]An alternative approach based on machine learning was also explored to detect tracheal compressions, using an SVM classifier and Histogram of Gradient features (HoG). The classifier performed well but produced false positives in cases when both lumens were compressed.

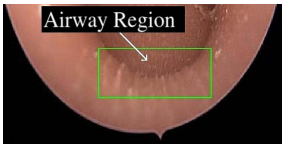


Fig. 4. Region of interest (airway) for tracheal compression detection.

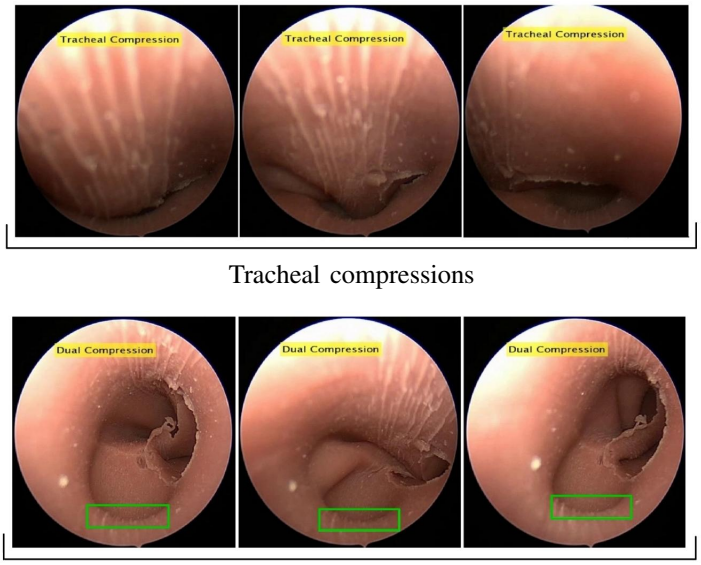


Fig. 5. A sampling of correct detections of tracheal compressions (TC) and lumen compressions (not dangerous, and so not detected by the algorithm). Note that for the lumen compressions, the ROI (green marker) remains visible.

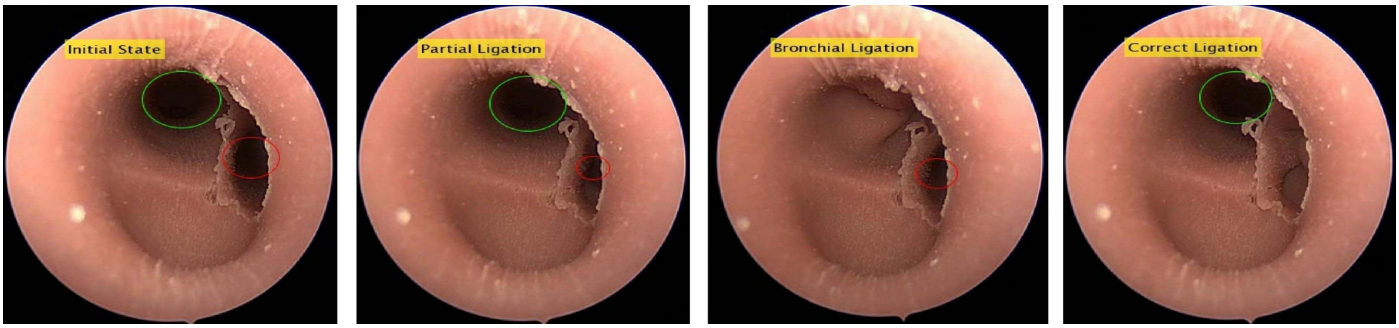
TABLE I. DETECTION CRITERION FOR PROCEDURAL STEPS

Procedural Step	# of Lumens	Detection Criterion
NL (no ligation)	2	$(d_B > \tau_B) \wedge (d_E > \tau_E)$
PL (incomplete ligation)	2	$(d_B > \tau_B) \wedge (d_E < \tau_E)$
BL (incorrect ligation)	1	$d_B \sim 0 \wedge (d_E > \tau_E)$
CL (correct ligation)	1	$(d_B > \tau_B) \wedge d_E \sim 0$

and esophageal lumen $\langle d_E, c_E \rangle$ are computed. The lumens are tracked, and the diameter and centroid variables are updated throughout the surgery procedure. Changes in these variables reflect the surgical state, as described in Table III-C. Parameters τ_B and τ_E correspond respectively to a 40% reduction in the diameters of the bronchial and espohageal lumen which, as advised by our expert surgeon, will result in a partial ligation (PL) or an imminent bronchial ligation (BL). Note that a closed lumen ($d \sim 0$) could correspond to a compression (temporary closure) or a ligation (permanent closure). The determination of ligation therefore is made until the end of the video—if indeed the closure is still present.

IV. EVALUATION AND RESULTS

Nine separate TT video performances of thoracoscopic fistula ligation were collected to provide a first evaluation of our algorithm. These nine videos used for evaluation were different from the three used to develop the algorithm, and neither did the expert surgeon who provided those three videos perform any of the nine procedures used for assessment. All three physiologic tracheal shades (Figure 1, middle) of the model were used. An expert surgeon provided groundtruth labels on each performance via manual review of videos.



No Ligation | Partial Ligation | Bronchial Ligation | Complete Ligation

Fig. 6. Detection of procedural steps related to ligation. Tracked variables (diameter, centroid) for the esophageal lumen are visualized as a red circle, and for the bronchial lumen as a green circle.

TABLE II. COMPARISON OF ALGORITHM PERFORMANCE (LEFT COLUMNS) VERSUS GROUNDTRUTH (RIGHT COLUMNS). DISAGREEMENT MARKED IN BOLD. TIME TO COMPLETE THE SURGERY REPORTED ONLY FOR CORRECTLY PERFORMED PROCEDURES (CL).

	Algorithm Prediction vs. Expert Review						
	# of TC		TC>8s		End-Result		Time (s)
V1 (Novice)	6	6	3	3	NL	NL	-
V2 (Novice)	**4**	**3**	0	0	PL	PL	-
V3 (Novice)	**13**	**9**	0	0	BL	BL	-
V4 (Intermediate)	0	0	0	0	PL	PL	-
V5 (Intermediate)	3	3	0	0	CL	CL	345
V6 (Intermediate)	3	3	0	0	CL	CL	198
V7 (Expert)	**5**	**3**	0	0	CL	CL	177
V8 (Expert)	3	3	0	0	CL	CL	172
V9 (Expert)	0	0	0	0	CL	CL	182

A. Algorithm Performance

The algorithm output consisted of the total number of tracheal compressions, those greater than 8 seconds, and the end-result of the surgical procedure. This output was compared to the groundtruth. The algorithm prediction of the end-result detection achieved 100% accuracy as compared to the manual review groundtruth for the nine evaluation performances (Table II). Our only source of disagreement with the groundtruth was in the detection of TC (mean absolute error 0.7); the algorithm is overly sensitive in this detection for transient compressions as compared to a human reviewer. However, extended (TC>8s) compressions were always in agreement with the groundtruth (mean absolute error 0).

Note also that the number of tracheal compressions decreases with expertise (Novice→ Intermediate→Expert), that the instances of correct ligations (CL) increase with expertise, and that the time to complete the surgical procedure decreases with expertise. Such trends speak to the feasibility of assigning a numerical score of surgical skill to a given performance. A validated mapping from these (and other) metrics to a score of surgical expertise will continue to be developed in our future work, as was discussed in Section III-B.

B. Trainee Performance Visualization

The algorithm output also produces a procedural steps diagram, consisting of correct and incorrect steps using time-stamped color coding. This represents a pictorial illustration of an individual surgical performance, as detected by the algorithm during the surgery. This diagram would allow the trainee to recognize and recover from incorrect procedures steps online, and for intraopertaive monitoring of surgical decision-making. A similar groundtruth diagram was created using the manual annotation provided by the expert. The correct and incorrect procedural steps diagrams from algorithm output imitated the groundtruth review in all of evaluations. Figure 7 provides representative results for three procedures.

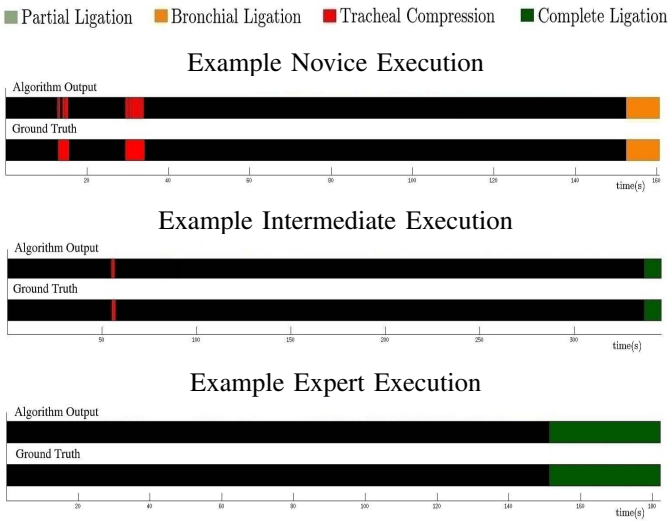


Fig. 7. Sample algorithm output of procedural steps detection, and comparison to groundtruth, for three representative performances (Novice, Intermediate and Expert).

V. DISCUSSION AND CONCLUSIONS

In a recent study evaluating adverse events in academic and community-based pediatric hospitals, surgical events accounted for nearly 33% of all adverse events identified [19]. In fact, children less than one year of age accounted for a staggering 75% of all procedure-related adverse events. These data illustrate that infants are a particularly vulnerable patient population for perioperative adverse events. Specific to TEF repair, early complications (anastomotic leak, anastomotic stricture and recurrent fistulae) are common, occurring in 16%-39% of all infants undergoing repair [20], [21], [22]. TEF repair thus remains a high priority for improved patient safety and overall outcomes.

Existing validated methods of surgical performance assessment are highly subjective, and may suffer from high inter-observer variability in scoring. All of the subjective methods of surgical assessment require real-time, or delayed videotape, review of the entire operative procedure in order to complete a full assessment. Thus, proper assessment of surgical performance becomes excessively time consuming. Our work makes use of a size-appropriate, anatomically correct TEF simulator, and provides automated procedural step detection that reliably detects three key errors of TEF repair: tracheal compression (impedes oxygenation/ventilation), bronchial ligation (instead of esophageal ligation) and incomplete esophageal ligation (leading to leaks, recurrent fistulae). The algorithm also detects tracheal compressions, which lead to airway obstruction and thus interfere with ventilation during the operation. This information is able to be provided in real-time to the surgeon-in-training, as they operate.

Experimental validation of the proposed algorithm has demonstrated the practical feasibility of online detection during training for the TEF procedure on the simulator. With the automated system, feedback can be immediate and not require expert review for training and evaluation. This allows for feedback that is *scalable* to many trainees, in addition to being *objective*. By immediately alerting the trainee to a near miss or actual adverse event, online correction and active learning becomes possible—opening the door to new and exciting methods for safer and smarter surgical training. Our future work will expand the detection capacity and validation of this algorithm—to the detection of more procedural steps (e.g. poor choice of ligation location), the assignment of a numerical value of surgical skill, and evaluation by more surgeons. (The evaluation as well on other simulators would be ideal, *if* there existed any other validated simulator models for TEF repair; however there do not.)

REFERENCES

[1] J. Xu, K. D. Kochanek, S. L. Murphy, and B. Tejada-Vera, "Deaths: final data for 2007." *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, vol. 58, no. 19, pp. 1–19, 2010.

[2] S. Sømme, M. Bronsert, A. Kempe, E. Morrato, and M. Ziegler, "Alignment of training curriculum and surgical practice: implications for competency, manpower, and practice modeling," *European Journal of Pediatric Surgery*, vol. 22, no. 1, p. 074, 2012.

[3] K. A. Barsness, D. M. Rooney, and L. M. Davis, "Collaboration in simulation: The development and initial validation of a novel thoracoscopic neonatal simulator," *Journal of pediatric surgery*, vol. 48, no. 6, pp. 1232–1238, 2013.

[4] K. A. Barsness, D. M. Rooney, L. M. Davis, and A. C. Chin, "Validation of measures from a thoracoscopic esophageal atresia/tracheoesophageal fistula repair simulator," *Journal of pediatric surgery*, vol. 49, no. 1, pp. 29–33, 2014.

[5] 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*, vol. 84, no. 2, pp. 273–278, 1997.

[6] B. N. Carter, "The fruition of halsted's concept of surgical training." *Surgery*, vol. 32, no. 3, p. 518, 1952.

[7] S. McAleer and R. Walker, "Objective structured clinical examination (OSCE)." *Occasional paper (Royal College of General Practitioners)*, no. 46, p. 39, 1990.

[8] J. D. Doyle, E. M. Webber, and R. S. Sidhu, "A universal global rating scale for the evaluation of technical skills in the operating room," *The American Journal of Surgery*, vol. 193, no. 5, pp. 551–555, 2007.

[9] 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 *Proceedings of the Information Processing in Computer-Assisted Interventions*, 2014.

[10] J. Rosen, B. Hannaford, C. G. Richards, and M. N. Sinanan, "Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 579–591, 2001.

[11] L. Vemer, D. Oleynikov, S. Holtmann, H. Haider, and L. Zhukov, "Measurements of the level of surgical expertise using flight path analysis from da Vinci robotic surgical system," *Medicine Meets Virtual Reality 11: NextMed: Health Horizon*, vol. 94, p. 373, 2003.

[12] 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*, vol. 11, no. 5, pp. 220–230, 2006.

[13] V. Datta, S. Mackay, M. Mandalia, and A. Darzi, "The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model," *Journal of the American College of Surgeons*, vol. 193, no. 5, pp. 479–485, 2001.

[14] V. Datta, A. Chang, S. Mackay, and A. Darzi, "The relationship between motion analysis and surgical technical assessments," *The American journal of surgery*, vol. 184, no. 1, pp. 70–73, 2002.

[15] M. Chmarra, C. Grimbergen, and J. Dankelman, "Systems for tracking minimally invasive surgical instruments," *Minimally Invasive Therapy & Allied Technologies*, vol. 16, no. 6, pp. 328–340, 2007.

[16] F. Lalys, L. Riffaud, D. Bouget, and P. Jannin, "A framework for the recognition of high-level surgical tasks from video images for cataract surgeries," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 966–976, 2012.

[17] N. Padoy, T. Blum, H. Feussner, M.-O. Berger, and N. Navab, "On-line recognition of surgical activity for monitoring in the operating room." in *Proceedings of the Innovative Applications of Artificial Intelligence*, 2008.

[18] K. Briechle and U. D. Hanebeck, "Template matching using fast normalized cross correlation," in *Proceedings of the SPIE Aerospace/Defense Sensing, Simulation, and Controls*, 2001.

[19] A. G. Matlow, G. R. Baker, V. Flintoft, D. Cochrane, M. Coffey, E. Cohen, C. M. Cronin, R. Damignani, R. Dubé, and R. Galbraith, "Adverse events among children in Canadian hospitals: the Canadian Paediatric adverse events study," *Canadian Medical Association Journal*, pp. cmaj–112 153, 2012.

[20] B. Allin, M. Knight, P. Johnson, and D. Burge, "Outcomes at one-year post anastomosis from a national cohort of infants with oesophageal atresia," *PloS one*, vol. 9, no. 8, p. e106149, 2014.

[21] C. Dingemann, C. Zoeller, and B. Ure, "Thoracoscopic repair of oesophageal atresia: results of a selective approach." *European Journal of Pediatric Surgery*, vol. 23, no. 1, pp. 14–18, 2013.

[22] P. O. Szavay, S. Zundel, G. Blumenstock, H. J. Kirschner, T. Luithle, M. Girisch, H. Luenig, and J. Fuchs, "Perioperative outcome of patients with esophageal atresia and tracheo-esophageal fistula undergoing open versus thoracoscopic surgery," *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 21, no. 5, pp. 439–443, 2011.