Optimizing Sensor Selection in Laparoscopic Simulators: Lessons Learned in a Robotic Platform

Kade MacWilliams[1], James Green[1], Ahmed Nasr[2], Georges Azzie[3], and Carlos Rossa[1]
[1]*Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada.*
[2]*Division of Pediatric Surgery, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada.*
[3]*Department of Surgery, Hospital for Sick Children, Toronto, ON, Canada.*

***Abstract*—Laparoscopic simulators provide a safe environment in which surgeons can practice and hone specific skills without risk to patients. However, providing effective performance feedback requires selecting the relevant metrics that most accurately reflect skill levels while remaining actionable for the trainee. This study investigates optimal sensor selection for laparoscopic simulators to enhance training assessment accuracy. Six common sensor types were tested across different combinations to evaluate their impact on recognizing surgical gestures, surgical tasks, and surgeon expertise levels using convolutional neural networks and multidimensional dynamic time-warping classifiers. The results show that linear velocity and gripper angle yield high classification accuracy across all metrics. For gesture recognition, velocity and gripper angle consistently appeared in the top-performing sensor combinations, demonstrating that these two parameters alone are highly indicative of a surgeon's intent and skill. Surprisingly, adding positional data does not improve accuracy, challenging the traditional emphasis on positional metrics in training systems. With the right sensor selection, surgical simulators can achieve accurate and actionable feedback while reducing complexity and cost without sacrificing performance, which can help make simulators more accessible and effective for training purposes.**

I. Introduction

Laparoscopy is a form of minimally invasive surgery where thin surgical instruments and a tube-shaped camera are passed through small incisions in the abdomen into the body cavity [1]. Compared to open surgery, laparoscopic procedures result in less trauma, a faster recovery time, and shorter hospital stays for patients [1], [2]. Unlike open surgery, the surgeon cannot directly see the operation site and must maneuver the tools based solely on the 2D video feed from the camera. This results in a steep learning curve as surgeons must translate 2D visual information into 3D spatial awareness and develop new skills in depth perception and hand-eye coordination [3], [4].

To alleviate the learning curve, various training methods have been developed, including intra-operative experience, cadaveric training (wet models), and simulators (dry models, which can employ physical and/or virtual technology). With intra-operative training, a resident surgeon first observes a surgical procedure before performing surgical gestures of increasing complexity. Cadaver-based training eliminates patient risk but is costly and lacks the realism of actual surgery. Both modalities have other major limitations, including limited patient access, variable case complexity, inconsistent training quality, and a lack of timely feedback.

The lack of standardized training led the Society of American Gastrointestinal and Endoscopic Surgeons to create a training curriculum with structured drills in a controlled environment using a training box. The box had openings through which surgical tools were inserted while a monitor displayed the video feed from a camera inside the box [5]. The trainer simulates motions commonly used in laparoscopy, such as bean drop, running string, block moving, suture foam, and a checkerboard drill. It was shown that 30-35 repetitions of each of these gestures resulted in improved surgical performance [6] and surgery residents improved their performance over a 6-month training period [7]. However, box simulators lack adaptability for varying patient sizes and are often tailored to adult patients with an average workspace volume of 34,225 cm^3. In pediatric patients, for example, the workspace is 20 times smaller than in adults, requiring a specialized training box such as that presented in [8] with a volume of 1,620 cm^3. Another major limitation of these trainers is the lack of realism and subjective performance assessment metrics.

Virtual reality (VR) trainers can restore a level of realism that box trainers lack. The surgeon is immersed in a VR environment to practice surgical tasks or the gestures found in traditional box simulators [9]. Resident surgeons trained using VR performed significantly shorter laparoscopic operations in the operating room, and supplementary training further decreased surgical time than practicing on a box trainer alone [6]. While box trainers require an expert to provide needed corrections to the trainee, VR trainers can collect and evaluate data automatically and provide immediate feedback [9], [10].

The majority of metrics evaluating surgical skills compare the trainee's tool path with that of an expert [11], [12]. For example, based on the path and speed of the centre of gravity of both hands, a neural network can determine whether the user is an expert or novice [11]. Another study using dynamic time warping identified surgical gestures in percutaneous nephrolithotomy based on tool path and compared them to that of an expert [13]. Markov models, recurrent neural networks (RNNs), and convolutional neural networks have also been

This research is supported by the Queen Elizabeth II Scholarship in Science and Technology, and the Children's Hospital Academic Medical Organization (CHAMO) innovation fund.

Cette recherche est appuyée par la bourse d'études supérieures en sciences et technologie de la reine Élizabeth II et par les fonds d'innovation CHAMO.

Email: kademacwilliams@cmail.carleton.ca (K. MacWilliams); jrgreen@sce.carleton.ca (J. Green); anasr@cheo.on.ca (A. Nasr); georges.azzie@sickkids.ca (G. Azzie), carlos.rossa@carleton.ca (C. Rossa).

TABLE I: Different instrumentation modalities in current laparoscopic simulators from least to most expensive

Simulator		Instrumentation Modality
Box Trainer	[17]	Camera
Box Trainer	[18]	Accelerometer, gyroscope, magnetometer
Box Trainer	[19]	Force sensors, IMU
LAPKaans	[20]	Roll, yaw, pitch, surge, speed, grasping force
Virtual Reality	[21]	Camera, Leap Motion, position, orientation
Lapro Apex	[22]	13 parameters - motion sensors and camera
BlueDRAGON	[23]	Position, force/torque, contact force sensors
DaVinci	[14]	76 kinematic variables

used in laparoscopy training [14], [15], [16].

While tool path is widely accepted as a useful metric for assessing surgical performance, there are many other kinematic features that can be measured. For example, the DaVinci surgical system measures 76 kinematic variables in real-time. Each simulator captures a different subset of kinematic variables; Table I summarizes different laparoscopy trainers and the sensing modalities they use. Whereas such extensive data provide detailed and useful information to a performance assessment algorithm, it does not necessarily offer direct, interpretable value to trainees seeking to improve their performance. Overloading them with complex metrics that may not align with direct improvements may create cognitive burden, making it difficult to interpret and apply feedback meaningfully. This then poses a risk of confusing trainees or diverting focus toward optimizing irrelevant metrics. Effective feedback should be concise to allow trainees to focus on improvements without being overwhelmed by details.

When designing a laparoscopic simulator and accompanying performance evaluation metrics, which sensors provide the most relevant and valuable information for accurately assessing a surgeon's performance and for providing feedback to a trainee? To answer this question, we use combinations of data acquired from different sensors during laparoscopic training in a classification algorithm. The algorithm then tries to identify the surgical gesture the trainee is performing, the surgical task they are attempting (a series of sequential gestures), and their level of expertise (expert, intermediate, or novice). The algorithm's classification accuracy is then compared across these different combinations of sensor data. In doing so, we make the assumption that the combination of sensor data that leads to the most accurate classification in these three categories is the most relevant data in the simulator to which trainees should pay particular attention. While dimension reduction has been used on surgical data [24], we take a different approach. Sensor data is clustered under different sensor categories. For example, when the position sensor is used, positional data from all measured degrees of freedom are used in the algorithm. By comparing their impact on the accuracy of classifying surgical gestures, tasks, and expertise levels, we identified the most important sensors for training effectiveness rather than just minimizing data complexity.

The results challenge the currently accepted paradigm that positional data is the most critical metric for assessing surgical proficiency, suggesting that other data may play a more significant role. The results also offer new insights to guide the design of cost-effective laparoscopic simulators (reducing unnecessary sensors) and algorithms that provide effective and actionable feedback to trainees. The analysis performed in this study uses kinematic data from a robotic laparoscopic surgical trainer, which is different from traditional laparoscopy. However, some useful insights obtained from kinematic data may generalize between the two forms of surgery.

II. Sensor Data Collection

Data from different sensing modalities are acquired as multidimensional temporal sequences. To provide meaningful feedback to a trainee, we will assume that these temporal sequences can be segmented into surgical gestures and surgical tasks. A surgical gesture is a primitive interaction between a surgeon and the instrument (e.g., positioning or orienting a needle, pulling suture) and a surgical task is a continuous sequence of predefined gestures (e.g., knot tying, which requires several needle positioning and suture pulling gestures). From the survey summarized in Table I, the most commonly used sensors in laparoscopic trainers may be clustered into 6 different groups:

- **Group 1** - Linear position: A linear position sensor that measures the 3D Cartesian position of each tool's tip;
- **Group 2** - Linear speed: A sensor that can measure the 3D linear speed of each tool's tip;
- **Group 3** - Angular position: A sensor that can measure the pitch, yaw, and roll of the tool shaft (gyroscope);
- **Group 4**: Angular speed: Sensors that can measure the 3D angular speed of the tool shaft, usually a gyroscope;
- **Group 5**: Accelerometers: Sensors that measure the linear acceleration of the tool tip, usually an inertial measurement unit (IMU);
- **Group 6**: Gripper angle: Detects the opening angle of each tool's gripper or if the gripper is opened or closed.

To determine the importance of each variable measured by these sensors on training performance, the temporal kinematic data from each of these sensors can be combined into all $2^6 = 64$ possible sensor combinations and passed through a classifier to identify a surgical task or gesture. We assume that the sensor combination that best predicts a predefined surgical gesture or task provides the most useful information to a trainee and specific areas they can focus on. For example, if the opening angle of the grasping tool (gripper angle) is a key predictor of a surgical gesture, task and/or expertise level, then the training surgeon must focus on controlling that movement throughout the procedure.

Let the sensor database be a multidimensional time series ${}^{w}X \in \mathbb{R}^{t\times n}$, where w is a sample and u is the total number of samples in the database, i.e., $1 \leq w \leq u$. Each of the n dimensions of X represents one dimension of a given measurement, all of which have length t:

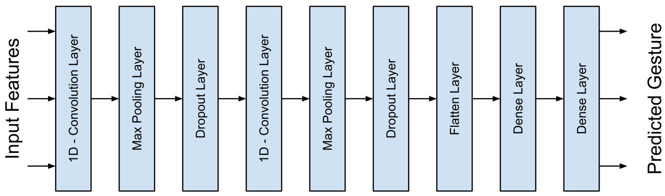

Fig. 1: Convolutional neural network structure for gesture and task recognition. The input layer is a convolution layer, followed by a max pooling, dropout, convolution, max pooling, dropout, flattening, and two dense layers.

$$^{w}X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t,1} & x_{t,2} & \dots & x_{t,n} \end{bmatrix}. \quad (1)$$

As per the notation above, each column of ^{w}X corresponds to a temporal sequence of the n_{th} measured dimension, and each row is a time stamp. We further assume that the each time series ^{w}X corresponds to a known class, corresponding to a given surgical gesture. Further, a surgical task ^{w}Y can be created by concatenating a sequence of p gestures, i.e. , $^{w}Y = [^{1}X \ ^{2}X \dots ^{p}X]^{T}$. With a database of known gestures, tasks and the expertise level of the performing surgeon, a classification algorithm can be implemented to identify the surgical task, gesture, and expertise level, using different dimensions of ^{w}X. When dimension $1 \leq c \leq n$ is not used in the classification algorithm, then column c of ^{w}X is removed.

III. Sensor Selection for Gesture, Task, and Expertise Level Classification

To identify surgical gestures and tasks using a specific combination of sensor data, several classification algorithms were tested. For gesture and task classification, a convolutional neural network (CNN) yielded the best results and is selected moving forward. While well suited for time series, it requires extensive training and a large dataset. Therefore, for expertise level recognition, dynamic time warping (DTW) was selected.

A. Sensor Selection for Gesture, and Task Classification

Classifying surgical gestures and tasks involves analyzing the measured data, comparing it with a database, and assigning a class to each gesture or task. A convolutional neural network (CNN) was selected, as it is well-suited to identifying patterns in temporal series data. The structure of the optimized CNN classifier is summarized in Fig. 1 and Table II. In the all convolution layers, the activation function is set to a rectified linear unit so that negative values resulting from the convolutional operation are set to 0, and the range of the resulting volume from the CNN is positive. This has been shown to increase speed of learning [25]. The CNN uses the "adam" optimizer and the loss function is the "categorical cross-entropy". For the convolutional layers, the stride value defaults to 1 and for the max pooling layers it defaults to the pool size of 2.

TABLE II: CNN model architecture with hyperparameters

Layer	Details
1D convolutional Layer	64 filters, 3 kernels, "relu" activation
Max pooling layer size	2
Dropout layer rate	0.5
1D convolutional layer	128 filters, 3 kernels, "relu" activation
Max pooling layer size	2
Dropout layer rate	0.5
Flatten layer	none
Dense layer	100 neurons, "relu" activation
Output dense layer	As many neurons as classes Softmax activation

Further optimization of the model architecture may improve the overall performance. However, the performance achieved with this model is in line with what other studies using the same dataset have achieved [16].

B. Sensor Selection for Expertise Recognition

Classifying the trainee's expertise level involves analyzing data in a task, comparing it with the database, and classifying the trainee as novice, intermediate, or expert. In a training context, the database will have a significant class imbalance, where non experts will be more represented than experts. Therefore, the classifier will be trained in a smaller dataset than when classifying gestures or tasks. CNN do not perform well on small training sets [26]. Multidimensional DTW is better suited to calculate the similarity between small sets of data [26]. A $k-$nearest neighbour (kNN) can then be used to assign the trainee to an expertise group. DTW involves stretching or compressing a signal to match the shape of another signal of different length. Given two samples $^{1}Y \in \mathbb{R}^{t_1 \times n}$ and $^{2}Y \in \mathbb{R}^{t_2 \times n}$ of lengths t_1 and t_2, DTW can be used to deform one of the samples until the minimum distance between them is found. To do so, n matrices $^{c}D \in \mathbb{R}^{t_1 \times t_2}$ can be created to compute the distance between each time stamp of column c of ^{1}Y and all time stamps of column c of ^{2}Y. Let element $d_{i,j}$ of D be the Euclidean distance between two entries in column c of two different samples, i.e.,

$$^{c}d_{i,j} = \sqrt{(^{1}y_{i,c})^2 + (^{2}y_{j,c})^2}, \quad (2)$$

where y_{ij} is the element in the i^{th} row and j^{th} column of ^{c}Y, with $1 \leq i \leq t_1$ and $1 \leq j \leq t_2$, the similarity between the respective dimensions of ^{1}Y and ^{2}Y can be quantified by finding the optimal alignment that minimizes the cumulative distance between them. The optimal aliment is known as the warping distance, or the continuous path in ^{c}D with the lowest distance ^{c}e between columns c:

$$^{c}e = d_{i,j} + \min \begin{cases} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} \end{cases}. \quad (3)$$

Considering all n columns of Y (each corresponding to a sensor measurement), the normalized warping distance is

$$e = \frac{1}{n} \sum_{c=1}^{n} {^{c}e}. \quad (4)$$

To determine each participant's expertise level, a kNN classifier identifies the database samples with the k smallest warping distances to the task being performed. The most common class in this subset is then assigned to the participant [26].

IV. Data Set and Data Pre-Processing

To implement the proposed algorithms, we use the publically available John Hopkins University Intuitive Surgical Inc. Gesture and Skill Assessment Working Set (JIGSAWS) [14]. It includes video recordings and 76 kinematic signals from 8 surgeons performing suturing, needle passing, and knot tying on the daVinci Surgical System. The three defined levels of surgical expertise are expert (an individual with over 100 hours of experience), intermediate (between 10 to 100 hours) and Novice (under 10 hours). Of the 8 surgeons who participated in the study, 2 self-identified as experts, 2 self-identified as intermediates, and 4 self-identified as novices.

The database is composed of 3 surgical tasks, namely suturing, knot tying, and needle passing. Each of these surgical tasks is broken down into a sequential combination of 15 possible annotated surgical gestures, as listed in Table III. In knot tying the surgeon picks up one end of a suture, while the other end is tied to a flexible tube. The surgeon ties a single loop knot. During suturing, the surgeon picks up a needle and passes it through the tissue three times. Finally, during needle passing the surgeon picks up a needle and passes it through small metal hoops from right to left[14]. Each surgeon performed each of these three tasks 5 times.

There is a total of 1703 gesture samples in the database and between 38 to 40 total samples of each surgical task, which resulted in a total of 118 total task samples (see Table III). The data was zero padded to ensure all samples used in the CNN have the same temporal length. The first derivative of each of the input dimensions was also taken and added to the sample. The gesture and task recognition CNN models were respectively given labeled samples from the gesture and task database in the form of ${}^{w}X$ and ${}^{w}Y$, which were then individually classified into 1 of 15 possible gestures or into 1 of 3 possible tasks. For gesture recognition, the gestures were divided into three subsets, each corresponding to a task, as seen in Table III. Then the CNN was trained and evaluated individually on each of these sets. This helps eliminate any kinematic differences caused by the task being performed, rather than by the gesture itself. For task recognition, all data is combined and used for training and validation.

To validate the CNN models, "Leave One User Out" cross validation was used and the average accuracy across the 8 trials (one for each user or surgeon) is reported. This helps ensure the models do not overfit and have not seen trials from the same surgeon in both training and testing. For expertise recognition, three data structures were created, one for each surgical task. The DTW-kNN algorithm was used to predict what level the surgeon was performing each task. This model was validated using a 80-20 train/test split. The needle passing and suturing data sets each contained 40 total samples. Due to two files either being empty or corrupted, the total number of samples for the knot tying task was 38. The number of neighbours in the k-NN classifier was set to $k = 3$.

TABLE III: List of gestures and tasks used in the evaluation, including what gestures compose a task, and the number of samples for each in the database [14].

Gesture	**Description**	Samples
G1	Reaching for needle with right hand	78
G2	Positioning needle	283
G3	Pushing needle through tissue	275
G4	Transferring needle from left to right	202
G5	Moving to centre with needle in grip	68
G6	Pulling suture with left hand	275
G7	Pulling suture with right hand	0
G8	Orienting needle	76
G9	Using right hand to help tighten suture	25
G10	Loosening more suture	5
G11	Dropping suture and moving to end point	100
G12	Reaching for needle with left hand	70
G13	Making C loop around right hand	75
G14	Reaching for suture with right hand	98
G15	Pulling suture with both hands	73
Task	**Gestures in Task**	
Suturing	G1-6, G8-11	40
Knot tying	G1, G11-15	38
Needle passing	G1-6, G8, G11	40

Since this paper aims to determine the importance of each sensing modality, each trial a specific subset of sensors was selected and thus only a subset of c dimensions of the input data was used to train and test the model. This resulted in 64 individual, identical models being trained and tested, for both surgical gesture recognition and surgical task recognition.

V. Results

Figure 2 shows the normalized cumulative accuracy of gesture recognition (left). All 15 gestures are classified using all 64 possible sensor combinations. The sensor combination is indicated by the 6 columns on the left: A shaded back square in a row indicates that the sensor was used, and unshared squares indicate the sensor was not used. The gesture classification results are separated per task, since each CNN was trained with gestures from the same task, (see Table III). The figure adds up the average classification accuracy of all gestures in a task, then normalizes the results to the highest classification accuracy. The right side of Figure 2 shows the cumulative accuracy of task recognition, also normalized to the highest classification accuracy. As can be seen, gripper angle and velocity are important sensing modalities in both surgical gesture and surgical task recognition. Interestingly, when all sensing modalities were included in the tested models, the performance was never the top ranked in terms of accuracy.

Table IV summarizes the best performing sensor combinations in order of accuracy, across each of the three classification exercises. The combination of velocity and gripper angle alone (sensor Group 2 and 6), are in the top accuracies, with each of these modalities being in every other top performing combination. Thus, by only measuring linear tool velocity and the tool gripper opening angle, a very good estimation of the surgeon's intention can be made. Table V compares the

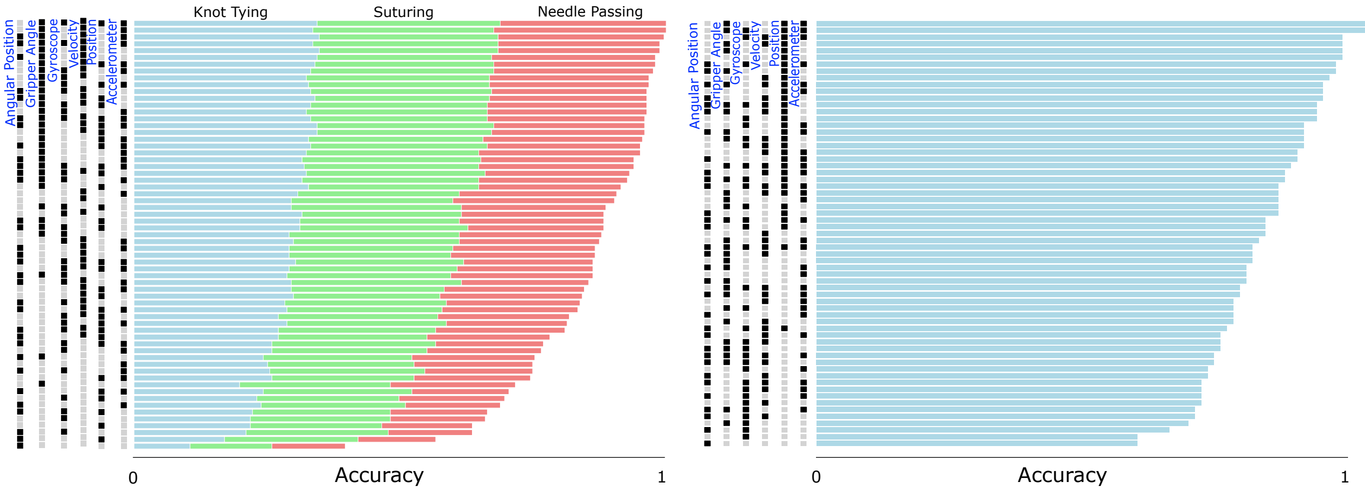

Fig. 2: Cumulative classification accuracy for gesture (left) and task recognition (right). On the left, each category corresponds the average classification accuracy of all gestures in a task (see Table III). The accuracy across all three classes ads up to 300% but is normalized to 1. A shaded black symbol indicates which sensing modality was used.

accuracy of the proposed methods with reduced dimensions against other gesture classification methods that used the same dataset. The proposed method outperforms these algorithms, reinforcing the notion that sensor selection plays a crucial when performing surgical gesture recognition and assessment.

Figure 3 shows the normalized cumulative accuracy of expertise level recognition. We see that the sensor combination of gripper angle, velocity and acceleration, is also the best performing, similar to gesture recognition, reaching an accuracy of 86%. Table IV further summarizes the sensor combinations that led to the most accurate expertise level recognition.

VI. Recommendations and Conclusion

Performing motion and surgical skill analysis within traditional laparoscopic box trainers involves selecting an appropriate sensing modalities to allow for accurate estimation of the trainee's intention. Deciding the suite of sensing modalities involves several considerations including resource cost, both monetary and computing, classification accuracy, and what actionable feedback they offer to the trainee.

As the results of gesture, task, and expertise recognition show, linear velocity and gripper angle are important parameters to be measured to provide accurate classification. These findings remained consistent across multiple configurations/model architectures tested; however, only the model described in Fig. 1 was included in the interest of space. These two sensing modalities appear in the top sensor array combination for all three classification exercises. Also, the top-performing combinations also include linear acceleration, which could be easily measured using an IMU. For gesture recognition, 8 of the top 9 performing combinations used velocity and gripper angle to classify the current gesture. Including linear position provided no gain in performance, as can be seen in Table IV. This leads to the recommendation that linear position could be determined from an alternative sensor than a kinematic position sensor, such as a camera.

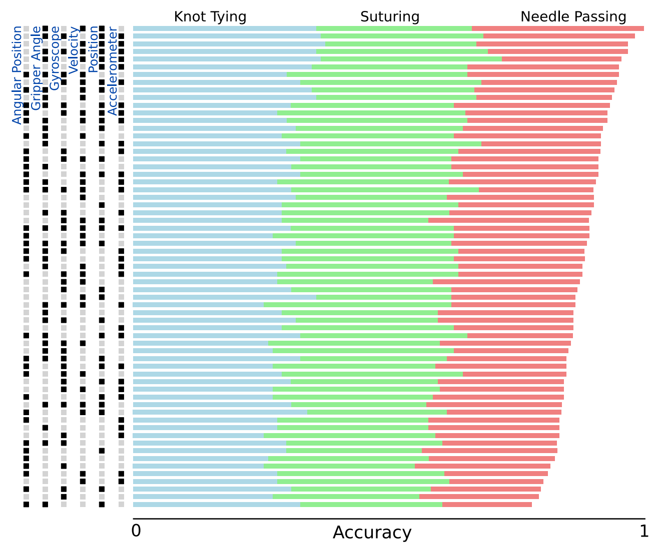

Fig. 3: Normalized expertise level classification accuracy. The 6 columns on the left indicate the sensors used in the classification algorithm.

This reduces the required hardware, as a camera is already available in most laparoscopic trainers.

The results from this paper suggest that when selecting sensors to equip a laparoscopic surgical training device, it is important to measure the gripper angle, velocity and acceleration of the tool tip. With only these three sensing modalities, our classification algorithm achieved the same or better accuracy than other algorithms that used the same data set [16], but with fewer data dimensions: The original dataset contained 76 dimensions of available data, yet our algorithm only requires four for a similar outcome. The achieved accuracy is within the is within the upper limit of range of human labeling performance of the JIGSAWS dataset [16].

The goal of this study is to inform the design of laparoscopic surgery simulators, in terms of which sensor modalities are

TABLE IV: Best average gesture, task, and expertise classification accuracy for different sensor combinations

	Sensors	Knot Tying	Suturing	Needle Passing
Gestures	Group 1, 2, 4, 5, 6	85%	85%	77%
	Group 2, 5, 6	83%	84%	80%
	Group 2, 3, 5, 6	86%	83%	77%
	Group 2-6	83%	86%	75%
	Group 2,6	86%	83%	75%
	All sensors	82%	82%	73%

Task		Expertise	
Sensors	**Accuracy**	**Sensors**	**Accuracy**
Group 1, 5, 6	87%	Group 1, 2, 4, 6	80%
Group 1, 4, 5, 6	86%	Group 1, 2, 6	78%
Group 1, 2, 6	82%	Group 1, 2, 5	77%
Group 1, 2, 4, 5	82%	Group 1-4	72%
Group 1, 3-5	81%	Group 1, 2, 5, 6	71%
All sensors	68%	All sensors	72%

TABLE V: Comparison of gesture recognition accuracy using a reduced dataset vs slave side data set using Gaussian Mixture Model - Hidden Markov Model classifier from [16].

Author	Method	Knot Tying	Suturing	Needle Pass.
Ahmidi et al. [16]	GMM-HMM	78.44%	80.83%	66.22%
This paper	CNN	86.00%	86.00%	80.00%

most informative when providing feedback to trainees. However, the data was obtained from the DaVinci robotic surgery system, where all movements are characterized by a full suite of sensor modalities, permitting the objective comparison of different sensors. One limitation of such an approach is that robotic surgery differs from traditional laparoscopy in terms of how the instruments are controlled, how the actions are implemented (e.g., some smoothing may be applied by the robotic control systems), and the constraints on which movements can be executed. While both types of surgery use a laparoscopy approach, these differences may limit our ability to generalize results from the current analysis to the design of traditional laparoscopy simulators. This study found gripper angle, tooltip linear velocity, and linear acceleration to be key sensing modalities in robotic surgery, aligning with previous studies on traditional laparoscopy research and revealing common kinematic trends in both approaches [27]. Future research will seek to confirm and validate these findings on data collected during traditional laparoscopic training exercises.

Analysis of surgical training data is a crucial step to give a training surgeon an interpretable performance metric, and actionable feedback when training. The results from this paper can guide the development of low cost laparoscopic training devices and algorithms that provide clear, concise, and actionable performance feedback.

REFERENCES

[1] A. Johnson, "Laparoscopic surgery," *The Lancet*, vol. 349, no. 9052, pp. 631–635, 1997.

[2] B. Jönsson and N. Zethraeus, "Costs and benefits of laparoscopic surgery—a review of the literature," *European Journal of Surgery*, vol. 166, no. S12, pp. 48–56, 2000.

[3] Y. Tanagho *et al.*, "2D versus 3D visualization: Impact on laparoscopic proficiency using the fundamentals of laparoscopic surgery skill set," *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 22, no. 9, pp. 865–870, 2012.

[4] S. Dreier, S. Mona, L. Konge, and F. Bjerrum, "Three-dimensional versus two-dimensional vision in laparoscopy: a systematic review," *Surgical Endoscopy*, vol. 30, no. 1, pp. 11–23, 2016.

[5] L. Villegas *et al.*, "Laparoscopic skills training," *Surgical Endoscopy And Interventional Techniques*, vol. 17, no. 12, pp. 1879–1888, 2003.

[6] J. Emken, E. Mcdougall, and R. Clayman, "Training and assessment of laparoscopic skills," *JSLS : Journal of the Society of Laparoendoscopic Surgeons*, vol. 8, pp. 195–9, 04 2004.

[7] M. Caban *et al.*, "Use of collapsible box trainer as a module for resident education," *JSLS*, vol. 17, no. 3, pp. 440–444, 2013.

[8] G. Azzie *et al.*, "Development and validation of a pediatric laparoscopic surgery simulator," *Journal of Pediatric Surgery*, vol. 46, no. 5, pp. 897–903, 2011.

[9] J. V. Taba *et al.*, "The development of laparoscopic skills using virtual reality simulations: A systematic review," *Plos one*, vol. 16, no. 6, p. e0252609, 2021.

[10] M. Varras *et al.*, "Role of laparoscopic simulators in the development and assessment of laparoscopic surgical skills in laparoscopic surgery and gynecology (review)," *World Academy of Sciences Journal*, vol. 2, no. 2, pp. 65–76, 2020.

[11] M. Uemura *et al.*, "Feasibility of an ai-based measure of the hand motions of expert and novice surgeons," *Computational and Mathematical Methods in Medicine*, p. 9873273, 2018.

[12] O. Trudeau *et al.*, "Construct validity and educational role for motion analysis in a laparoscopic trainer," *Surgical Endoscopy*, vol. 29, no. 9, pp. 2491–2495, 2015.

[13] A. Cotton, B. Sainsbury, and C. Rossa, "Sliding adaptive dynamic time warping for segment matching in surgical simulation data," in *Inter.Conf. on Smart Multimedia*, 03 2024.

[14] Y. Gao *et al.*, "JHU-ISI gesture and skill assessment working set (JIGSAWS): A surgical activity dataset for human motion modeling," *Modeling and Monitoring of Comp. Assisted Interventions*, 2014.

[15] B. van Amsterdam *et al.*, "Gesture recognition in robotic surgery: A review," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 6, pp. 2021–2035, 2021.

[16] N. Ahmidi *et al.*, "A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2025–2041, 2017.

[17] J. Wong *et al.*, "Construction and validation of a low-cost laparoscopic simulator for surgical education," *Journal of Surgical Education*, vol. 70, no. 4, pp. 443–450, 2013.

[18] C. Heiliger *et al.*, "Tracking and evaluating motion skills in laparoscopy with inertial sensors," *Surgical Endoscopy*, vol. 37, no. 7, pp. 5274–5284, 2023.

[19] B. Gavrilovic *et al.*, "Development of an open-source laparoscopic simulator capable of motion and force assessment: High tech at low cost," *Journal of Laparoendoscopic & Adv. Surgical Techniques*, 2018.

[20] L. H. Olivas-Alanis *et al.*, "LAPKaans: Tool-motion tracking and gripping force-sensing modular smart laparoscopic training system," *Sensors*, vol. 20, 2020.

[21] V. Lahanas *et al.*, "Virtual reality-based assessment of basic laparoscopic skills using the leap motion controller," *Surgical Endoscopy*, vol. 31, no. 12, pp. 5012–5023, 2017.

[22] Laparosimulators, "Apexpro laparoscopic simulator," accessed: 2024-10-23. [Online]. Available: https://laparosimulators.com/apexpro/

[23] J. Rosen *et al.*, "The bluedragon - a system for measuring the kinematics and dynamics of minimally invasive surgical tools in-vivo," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, vol. 2, 2002, pp. 1876–1881 vol.2.

[24] P. Hasani *et al.*, "Towards an efficient computational framework for surgical skill assessment: Suturing task by kinematic data," in *International Conf. on Robotics and Mechatronics*, 2021, pp. 171–176.

[25] R. K. Emanuel *et al.*, "The effect of activation functions on accuracy, convergence speed, and misclassification confidence in cnn text classification: a comprehensive exploration," *The Journal of Supercomputing*, vol. 80, no. 1, pp. 292–312, 2024.

[26] S. Albasri *et al.*, "A novel distance for automated surgical skill evaluation," in *2019 7th E-Health and Bioengineering Conference*, 2019.

[27] M. A. Farcas *et al.*, "Analysis of motion in laparoscopy: the deconstruction of an intra-corporeal suturing task," *Surgical Endoscopy*, vol. 31, no. 8, pp. 3130–3139, 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27928669/