

Chapman University

Chapman University Digital Commons

Physical Therapy Faculty Articles and Research

Physical Therapy

6-3-2022

Evaluation of Surgical Skill Using Machine Learning with Optimal Wearable Sensor Locations

Rahul Soangra

R. Sivakumar

E. R. Anirudh

Sai Viswanth Reddy Y.

Emmanuel B. John

Follow this and additional works at: https://digitalcommons.chapman.edu/pt_articles



Part of the [Surgery Commons](#)

Evaluation of Surgical Skill Using Machine Learning with Optimal Wearable Sensor Locations

Comments

This article was originally published in *PLoS ONE*, volume 17, issue 6, in 2022. <https://doi.org/10.1371/journal.pone.0267936>

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

Copyright

The authors

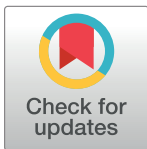
RESEARCH ARTICLE

Evaluation of surgical skill using machine learning with optimal wearable sensor locations

Rahul Soangra^{1,2*}, R. Sivakumar³, E. R. Anirudh³, Sai Viswanth Reddy Y.³, Emmanuel B. John¹

1 Department of Physical Therapy, Crean College of Health and Behavioral Sciences, Chapman University, Irvine, California, United States of America, **2** Department of Electrical and Computer Science Engineering, Fowler School of Engineering, Chapman University, Orange, California, United States of America, **3** Department of Sensor and Biomedical Technology, School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

* soangra@chapman.edu



Abstract

Evaluation of surgical skills during minimally invasive surgeries is needed when recruiting new surgeons. Although surgeons' differentiation by skill level is highly complex, performance in specific clinical tasks such as pegboard transfer and knot tying could be determined using wearable EMG and accelerometer sensors. A wireless wearable platform has made it feasible to collect movement and muscle activation signals for quick skill evaluation during surgical tasks. However, it is challenging since the placement of multiple wireless wearable sensors may interfere with their performance in the assessment. This study utilizes machine learning techniques to identify optimal muscles and features critical for accurate skill evaluation. This study enrolled a total of twenty-six surgeons of different skill levels: novice ($n = 11$), intermediaries ($n = 12$), and experts ($n = 3$). Twelve wireless wearable sensors consisting of surface EMGs and accelerometers were placed bilaterally on bicep brachii, tricep brachii, anterior deltoid, flexor carpi ulnaris (FCU), extensor carpi ulnaris (ECU), and thenar eminence (TE) muscles to assess muscle activations and movement variability profiles. We found features related to movement complexity such as approximate entropy, sample entropy, and multiscale entropy played a critical role in skill level identification. We found that skill level was classified with highest accuracy by i) ECU for Random Forest Classifier (RFC), ii) deltoid for Support Vector Machines (SVM) and iii) biceps for Naive Bayes Classifier with classification accuracies 61%, 57% and 47%. We found RFC classifier performed best with highest classification accuracy when muscles are combined i) ECU and deltoid (58%), ii) ECU and biceps (53%), and iii) ECU, biceps and deltoid (52%). Our findings suggest that quick surgical skill evaluation is possible using wearables sensors, and features from ECU, deltoid, and biceps muscles contribute an important role in surgical skill evaluation.

OPEN ACCESS

Citation: Soangra R, Sivakumar R, Anirudh ER, Reddy Y. SV, John EB (2022) Evaluation of surgical skill using machine learning with optimal wearable sensor locations. PLoS ONE 17(6): e0267936. <https://doi.org/10.1371/journal.pone.0267936>

Editor: Anwar P.P. Abdul Majeed, Universiti Malaysia Pahang, MALAYSIA

Received: November 17, 2021

Accepted: April 20, 2022

Published: June 3, 2022

Copyright: © 2022 Soangra et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information](#) files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Surgical skills and techniques are central components of a surgeon's skill set and directly correlate with patient benefits [1]. It is critical to assess surgical skills among trainees in surgical specialties to identify their competence and confidence to practice independently [2]. Surgical skills are influenced by decreasing work hours and increasing sub-specializations focusing on minimally invasive surgeries (MIS). These changes impact residency training since there are concerns for current surgery residents' adequate operative experience and technical training [3, 4]. To determine the level of surgical skills and clinical competency, a multitude of surgical skill assessment tools, consisting of experts' evaluation of surgical skill videos, which is often costly and time-consuming [5, 6] are used.

Furthermore, evaluation by experts often lacks objective data and may sometimes be biased by the reviewer [7, 8]. Thus, there is a lack of objective high-fidelity tools to evaluate surgical skills in hospital environments. Previously researchers have estimated surgical skills using i) kinematic data and convolutional neural network [9], ii) kinematic data as putative markers and deep neural networks [10], iii) virtual reality spinal task and machine learning algorithms (support vector machines, k-nearest neighbors, least discriminant analysis, naïve bayes and decision tree) [11], iv) image processing and deep neural network during robotic surgery [12–14], v) kinematic data from da Vinci robot and global rating score and machine learning (kNN, logistic regression, SVM) [15]. Recently deep learning-based haptic guidance systems have been used for surgical skill development [16]. Moreover, it is unknown which features from wearable sensors such as EMG and accelerometers could contribute more to skill identification. It is also unknown which location on the upper extremity can be helpful for accurate skill classification with minimal hindrance. Instead of evaluating surgical time, one needs to assess objective trajectorial movement and muscle activity features for correct skill classification and surgical ergonomics. In surgical skill evaluations, some studies evaluated the muscular workload during surgical procedures using surface electromyography (sEMG) (17–20) for muscle activity. Other researchers have utilized wearable inertial sensors consisting of accelerometers to determine efficiency between experts and trainees performing the same surgical task [17]. To some extent, both movement accelerations and muscle activities have shown the potential to distinguish subtle details on successfully identifying the skill level of surgeons. EMG signals are a reliable measure for physiological stress detection in the laboratory [18], and earlier studies have reported the upper trapezius muscle as an essential stress indicator [19, 20].

Nonlinear movement variability features can quantify neuromuscular connections (feedback) and subtle movement changes [21–23]. Researchers have reported that entropy measures such as approximate entropy, sample entropy, and multiscale entropy can estimate specific feedback mechanisms and spontaneous properties of interconnected neurons and are characterized as regularity [24–26]. Both linear and nonlinear features can be extracted from EMG and accelerometer sensors, respectively. Feature extraction is crucial for accumulating information relevant to surgical skill evaluation. The selection of feature vectors from EMG and accelerometer signals during surgical tasks needs careful consideration since many muscles and features may contain redundant information [27, 28]. This study extracted muscle work-related EMG features and linear and nonlinear movement variability features from accelerometer signals. Appropriate features will result in higher classification accuracy with maximum skill separability and robustness [27, 28]. With recent developments in machine learning, we will explore critical objective biomarkers for identifying surgical skillsets with high accuracy; however, there are multiple sites where wearable sensors such as EMG and inertial sensors can be affixed. This study will consider the performance analysis for feature

extraction and selection algorithms from two fundamental perspectives: 1) Which features varied significantly among three surgical tasks and best differentiate surgical skillset? 2) Which sensor position offers the skill classification. This study will attempt to establish the groundwork for identifying i) important sites of sensor placement and ii) important linear and non-linear features which can distinguish skill levels amongst residents, medical students, and expert surgeons. Our results will help develop wearable devices integrated with machine learning technologies to evaluate surgeons' performance during minimally invasive surgical tasks.

Materials and methods

Twenty-six participants categorized with experience from the Department of Urology at the University of California, Irvine, (UCI) participated in this study. All participants signed the written consent form approved by UCI (HS#: 2018–4407). The three groups were categorized based on their surgical experience; i) novice (undergraduate or medical students without prior surgical experience) (N = 10), ii) intermediate (urology residents postgraduate year 1–5) (N = 11), and expert surgeons (urology physicians with more than five years experience) (N = 5). Participants performed basic surgical tasks; i) pegboard transfer, ii) knot tying, and iii) robotic suturing. At least three trials were conducted for each surgical task. The participants were asked to perform these tasks at their normal pace. To identify optimal locations for surgical workload assessment, electromyography (EMG) sensors with in-built accelerometers were used to estimate muscular activation level and timing. A total of twelve EMG surface electrodes (DELSYS[®] Trigno[™] Wireless, Boston, MA) were attached bilaterally on bicep brachii, tricep brachii, anterior deltoid, flexor carpi ulnaris (FCU), extensor carpi ulnaris (ECU), and thenar eminence (TE) (Fig 1). Resting data was collected during sitting and standing postures. These muscle groups were selected since previous studies on surgical ergonomics have highlighted their importance and activations [29]. Deltoid was selected since the shoulder is the most common site for musculoskeletal symptoms reported by laproscopic surgeons [30]. Muscle activities were normalized using Maximum Voluntary Contraction (MVC) of each muscle group.

Surgical skills tasks

All participants performed four surgical tasks. i) *Open knot tying*: Two-handed surgeon's knot with other three-square knot throws was tied over two rubber bands mounted on a knot tying practice board (Ethicon Inc., Somerville, NJ) (Fig 2). ii) *Laparoscopic peg transfer*: The task aims to transfer the objects (rubber pegs) from one side of the board to the other and back using two laparoscopic graspers. Six pegs are placed on the left side of the board; each peg is picked with the left grasper, transferred midair to the right grasper, and then placed over a post on the right side of the board. Once all pegs have been transferred, the process is reversed (Fig 3). Laparoscopic peg transfer was completed on an SZABO-BERCI-SACKIER Laparoscopic Trainer (Karl Storz, Tuttlingen, Germany). iii) *Robotic suturing*: A needle is passed through a rubber Penrose drain using da Vinci[®] Si surgical system robotic needle drivers (Intuitive, Sunnyvale, CA). The instrument knot is then reinforced with three additional square knots (Fig 4). EMG data were recorded at 2000 Hz, and accelerometer data were sampled at 100Hz. For EMG processing, raw signals were detrended to remove bias. EMG signals were then band-pass filtered from 20–500 Hz followed by a notch filter of 60 Hz to attenuate 60Hz electrical noise. Filtered signals were rectified using Root Mean Square (RMS) utilizing a window size of 6 data points and filtered using a fourth-order low-pass filter with a cutoff frequency of 5 Hz.

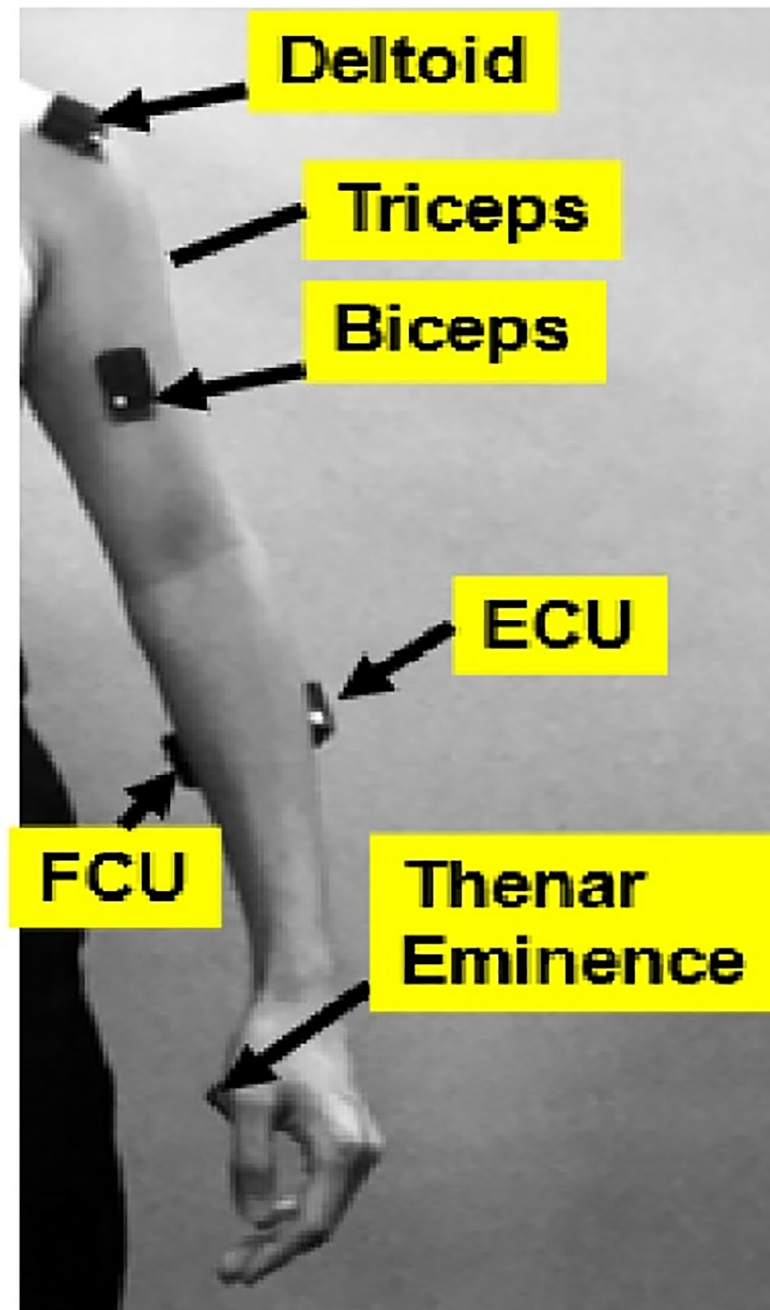


Fig 1. Muscle sites for Delsys Trigno EMG sensor attachment. The muscle sites included deltoid, triceps, biceps, Extensor Carpi Ulnaris (ECU), Flexor Carpi Ulnaris (FCU), Thenar Eminence (TE) for the left and right side.

<https://doi.org/10.1371/journal.pone.0267936.g001>

EMG preprocessing and feature extraction

i) Muscular workload-related features: Cumulative muscular workload represents the total amount of work for each muscle group, normalized for MVC, as a function of time. The average muscular workload was computed as the total cumulative work done per second. **ii) Time-domain feature:** The time-domain features are quick and easy to compute [31–33]. The EMG

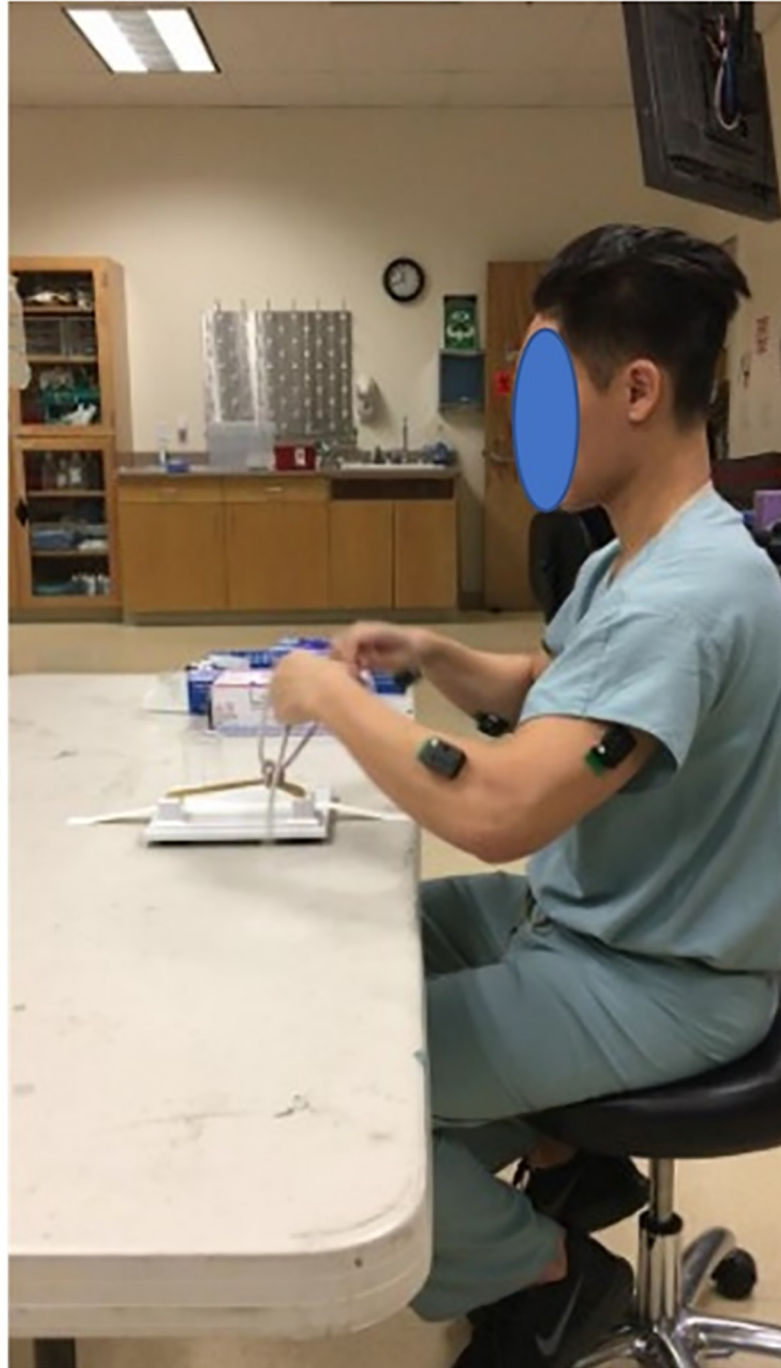


Fig 2. Knot tying task performed by participants.

<https://doi.org/10.1371/journal.pone.0267936.g002>

signals were rectified using windowed RMS and then low-pass filtered. The muscle activation time was computed using a threshold from baseline data and five standard deviations, as shown in Table 1 below. Total time of muscular activity was computed when muscle onset and muscle cessation points were determined, **iii) linear variability features:** variability features such as signal range and RMS were evaluated for quantifying muscle firing variability during a

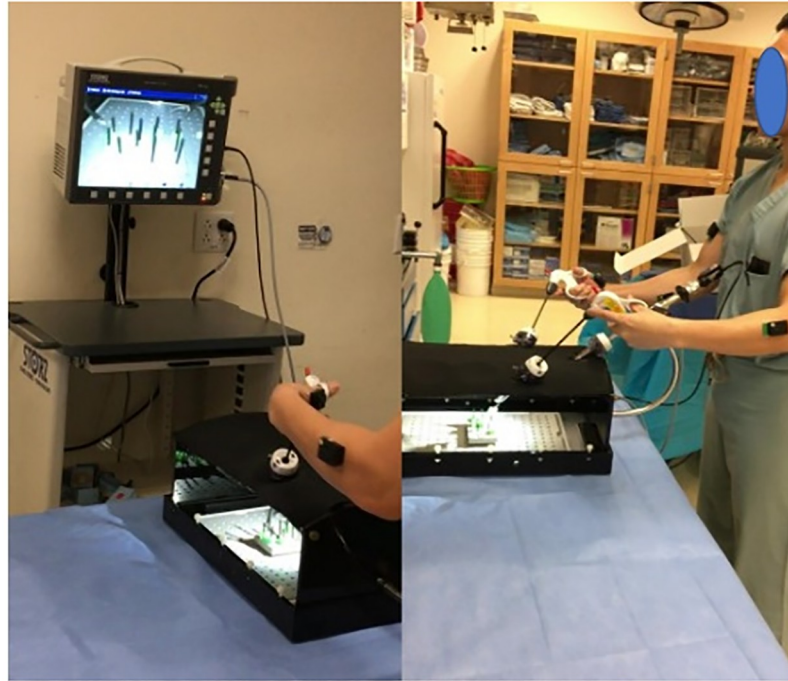


Fig 3. Pegboard transfer task performed by participants.

<https://doi.org/10.1371/journal.pone.0267936.g003>

task. iv) **frequency-domain feature**: dominant muscle firing frequency was evaluated for each muscle.

Accelerometer feature extraction

Both linear and nonlinear features were extracted from accelerometer data, as shown in [Table 2](#) below.



Fig 4. (a) position of six muscles sites where surface EMG sensors were affixed (b) Pegboard transfer task (c) Robotic suturing task using Da Vinci (d) Knot tying task (e) Ureteroscopy task.

<https://doi.org/10.1371/journal.pone.0267936.g004>

Table 1. Temporal and frequency domain features extracted from EMG signals.

| Features | Feature Evaluation Method |
|--|---|
| Cumulative Muscular Workflow (CMW) was computed through the integration of normalized EMG data for time throughout the task | $\text{Cumulative muscular workload (CMW)} = \int_0^t \frac{\text{Rectified EMG}}{\text{MVC}}$ |
| Average Muscular Work done per second | $\text{Average Work per Second (AWS)} = \frac{\text{CMW}}{\text{Performance Time}}$ |
| Total Time: Total time was evaluated utilizing a threshold; start time was recorded when EMG activity was more than the threshold. | $\text{Threshold} = \text{Mean} + 5 \times \text{SD}$ |
| RMS: Root means square measures linear variability through quantifying fluctuations in a signal, where n is the length of the EMG signal. | $\text{EMG}_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \text{EMG}_i^2}$ |
| Signal Range: Signal range was computed as the difference between the maximum and minimum EMG activity | $\text{Range} = \text{EMG}_{\text{max}} - \text{EMG}_{\text{min}}$ |
| Signal Frequency: The dominant frequency carries more energy to other frequencies on the EMG spectrum. | <ul style="list-style-type: none"> • EMG signal is detrended • zero-padded to the nearest higher power of 2 • Fast Fourier transform (FFT) was computed • Modulus of signal FFT (complex number) is squared to get raw power spectral density (PSD) • The frequency with the highest amplitude in the PSD is the dominant frequency of the EMG signal. |

<https://doi.org/10.1371/journal.pone.0267936.t001>

Feature Selection (FS)

Feature selection algorithms are necessary since they can remove redundant or unnecessary information from features that do not improve classification accuracy and help lower computational costs [41]. The best subset of features is associated with sensors placed at different muscles. Feature selection was utilized to reduce the number of input features (dimension reduction) or remove the redundant features. This dimension reduction allows lower computing speed and even reduces space complexity. We explored filter and wrapper feature selection methods such as Correlation-based Feature Selection (CFS) and Recursive Feature Elimination (RFE). CFS shows how well two features are correlated with a value ranging from -1 to 1, whereas RFE provides a ranking of features by giving weights to a particular part.

Filter feature selection methods are usually the first choice as it is not computationally expensive compared with wrapper and embedded feature methods. Filter methods are advantageous since they prevent the model from overfitting solely depending on the dataset's statistical distribution.

Correlation-based Filter Feature Selection (CFS)

CFS works well on supervised than unsupervised problems. It groups a subset of highly correlated features with the target attribute (i.e., novice, intermediate, expert). CFS measures the relationship between two variables with statistical data but is limited to identifying only the linear relationship between the features. It works on the principle that we could drop the features with lower correlation coefficients if the predictor variables are correlated. We evaluated pairwise correlation among feature sets. The value of the correlation coefficient varies between +1 and -1. A value of ± 1 indicates a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the variables is

Table 2. Linear and nonlinear features extracted from accelerometer data.

| Features | Feature Evaluation Method |
|--|--|
| <p>Resultant Acceleration Timeseries: A time series was extracted from each sensor using the resultant of 3-dimensional acceleration as given by equation 1.</p> | $R_{Acc} = \sqrt{Acc_x^2 + Acc_y^2 + Acc_z^2} \quad (1)$ <p>Acc_x, Acc_y, and Acc_z are acceleration in X, Y, and Z-directions.</p> |
| <p>Correlation Dimension: The correlation dimension was introduced by Grassberger and Procaccia [34]. The correlation dimension was computed from resultant acceleration and is a fractal dimension unrevealing system’s complexity [35, 36] and quantifies inherent dynamical behaviors with a single number.</p> | <p>For an m-dimensional phase space, the correlation function $C(r)$ is given by equation 2 below.</p> $C(r) = \lim_{N \rightarrow \infty} \frac{2}{N(N-1)} \sum_{\substack{i,j \\ (1 \leq i < j \leq N)}} H(r - Y_i - Y_j) \quad (2)$ <p>Where H is the Heaviside step function, with $H(u) = 1$ for $u > 0$, and $H(u) = 0$ for $u \leq 0$, where $u = r - Y_i - Y_j$, r is the radius of a sphere centered on Y_i or Y_j, and N is the number of data points.</p> |
| <p>Multiscale Entropy (MSE): Entropy techniques effectively quantify the probability that neighboring points in the resultant acceleration time series will be within a predetermined range. Entropy is a measure of complexity in physiological systems denoting a highly adaptable network of neuromuscular connections attained by regular practice [37, 38]; increased MSE is indicative of a greater degree of complex movement dynamics [38, 39].</p> | <p>Utilizing the coarse-graining process to time series, new time series is constructed by averaging the data points within non-overlapping windows of increasing length, τ. Each time series element $y_j^{(\tau)}$, is given by equation 3.</p> $y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i \quad (3)$ <p>Where τ represents the scale factor and $1 \leq j \leq N/\tau$. The length of each coarse-grained time series is N/τ.</p> |
| <p>Rosenstein’s Lyapunov Exponent: Lyapunov exponents can detect the presence of chaos in a dynamical system by quantifying divergence in trajectories [40]. Resultant acceleration time series from sensors were utilized to assess nonlinear variability and chaotic properties.</p> | <p>i) A delayed reconstruction $Y_{i:N}$ with embedding dimension ‘m’ and lag ‘τ’.</p> <p>ii) For a point ‘i’, the algorithm finds the nearest point i^* that satisfies $\min\ Y_i - Y_{i^*}\$ such that $i - i^* > \text{‘Minimum Separation,’}$ where ‘Minimum Separation,’ the mean period, is the reciprocal of the mean frequency.</p> <p>iii) The Lyapunov Exponent for the entire expansion range is calculated as in equation 4,</p> $\lambda(i) = \frac{1}{(K_{max} - K_{min} + 1)dt} \sum_{K=K_{min}}^{K_{max}} \frac{1}{K} \ln \frac{\ Y_{i+K} - Y_{i+K}\ }{\ Y_i - Y_{i^*}\ } \quad (4)$ <p>K_{min} and K_{max} represent Expansion Range, ‘dt’ is the sampling time and divergence at ith point is given by equation 5.</p> $l_{div} = \ln \frac{\ Y_{i+K} - Y_{i+K}\ }{\ Y_i - Y_{i^*}\ } \quad (5)$ <p>iv) A single value for the Lyapunov exponent is then computed from the earlier step using polynomial fit as given in equation 6</p> $LyE = \text{polyfit}([K_{min} \ K_{max}], \lambda(i)) \quad (6)$ |

<https://doi.org/10.1371/journal.pone.0267936.t002>

weaker. Pearson correlation coefficients provide a relationship among features and use information about the mean and standard deviation from the data compared to nonparametric correlation, which only uses the ordinal information and scores of pairs. It is limited since it can only assess the linear relationship between the features. However, the correlation between the categorical attributes and the target attributes cannot be determined.

Recursive Feature Extraction (RFE)

RFE is an external estimator and assigns weights to features. The principle is to select features by recursively considering a smaller set of features. The model is trained on the initial set of features, and with each iteration, we can calculate the feature coefficient. Then the least essential features will be removed, and then again, the model will be trained. Since it takes a lot of iterations to complete the whole ranking process, the computational time is quite expensive. Feature selection methods are performed on the entire muscle dataset, including different independent features to evaluate the surgeon’s skill in that particular task. This dataset accumulates data generated by six muscle sensors attached to the surgeon. Feature selection methods have been implemented in the Jupyter notebook with the help of pandas python library for ML. Reducing the features from 26 to 10 has increased the accuracy by five and reduced computational time.

Model development

After feature selection, preprocessing techniques is used to transform the initial dataset to standardized values. Methods such as imputer and standardization are implemented to enhance the dataset's quality. We trained three supervised ML models to assess the dataset: Random Forest Classifier (RFC) with 100 estimators, Naive Bayes classifier (NB) with kernel set to Radial Basis Function (RBF), Support Vector Machine (SVM) with all the default parameters. The objective of these models is to classify a surgeon's skill based on the feature instances. For all the three supervised machine learning models, both train and test sizes were kept at the same ratio (70/30). We used 70% of the whole dataset for training and 30% for testing. According to the six muscle types, the entire dataset is divided into six sub-datasets. These six sub-datasets are then trained on three ML models to investigate which muscle is vital to identify the surgical skill. A few performance parameters were considered to evaluate the model to determine which model has done better on a particular sub-dataset. These parameters include precision, recall, f1 score (harmonic mean of precision and recall), true positive rate (TPR), false-positive rate(FPR). A **true positive (TP)** is an outcome where the model *correctly* predicts the *positive* class. Similarly, a **true negative (TN)** is an outcome where the model *predicts the negative class correctly*. A **false positive (FP)** is an outcome where the model *incorrectly* predicts the *positive* class. A **false negative(FN)** is an outcome where the model *incorrectly* predicts the *negative* class. The accuracy, recall, specificity, precision, and F1-score were evaluated using Eqs 7–11 below.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

$$Precision = \frac{TP}{TP + FN} \quad (10)$$

$$F1\ Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (11)$$

Supervised classifiers

Supervised learning algorithms are helpful in model correlation and the dependencies among their input features to predict the output values for new data based on the relationships learned from the training data sets. This study used supervised classification algorithms to classify three distinct skill groups—Novice, Intermediate, Expert.

Feature selection process

Preprocessing was conducted to remove impurities for higher model classification accuracy. Feature selection methods like CFS and RFE were used to identify features that could improve model accuracy significantly. The feature subset was divided into six subsets as per the muscle type. Each data subset was divided into training and testing sets (70/30), training data was

used to train the 3 model types (SVM, RFC, and Naïve Bayes). The model parameters such as precision, recall, and accuracy were evaluated (Fig 5).

This study investigated muscles that could potentially classify surgical skills with high accuracy. The muscles that achieved better statistical performance than others were chosen, and others were dropped in evaluating the surgical skill. The priority of muscles from all models was assigned equal importance and weightage, and finally, the muscles that showed the highest significance for all three ML models were selected.

Random Forest Classifiers (RFC)

Random Forest is a machine learning algorithm based on decision trees. Initially, all root nodes have different supervised classifiers, and each training data is predicted. The class selected as output for most by the classifiers is regarded as the final output by the RFC. RFC consists of multiple individual decision trees that operate as an ensemble. Each unique tree in the random forest divides a class of predictions, and the class with the most votes becomes the model's prediction.

In this study, RFC was trained using input linear and nonlinear EMG and accelerometer features to predict which target class is more suitable for the new instance. The hyperparameters for the RFC are set to default parameters levels (Table provided in S1 Table). The feature data set was divided into 6 sub-datasets based on muscle attributes. In this dataset, we have six different muscles groups. The performance test scores attained by this model for surgical skill classification across these datasets were evaluated.

Naïve Bayes

Naive Bayes is a rapid classification algorithm best suited for a massive chunk of data. It operates on the Bayes theorem of probability for the prediction of a new instance. The classification stage comprises two general phases: the learning and evaluation phases. The classifier trains its model on a given dataset in the learning phase, whereas the evaluation phase tests the classifier performance. This classifier is trained on our dataset and predominately predicts the probability of a class (surgical skill level) in the target attribute. The class with the highest chance will be considered for the particular instance. This model has been evaluated based on a few performance metrics such as precision, recall, and f1score.

Support Vector Machines

Support Vector Machines(SVM) are efficient with smaller datasets compared to other supervised learning classifiers. SVM is beneficial when nonlinear features are involved; in such cases, feature correlation with model output cannot be mapped. SVM distinguishes different classes by constructing a hyperplane in a multidimensional plane. It maximizes the margin between the classes so that there is probably less chance of a new instance overlapping two classes. SVM is considered in our approach because of the dataset being multi-dimensional (more features). Confusion metrics have been used to evaluate our model's performance. The results suggest that the model accurately recalls the true positives (correctly predicting the positive class). For further analysis, we have plotted the ROC curves to decide the suitability of this classifier with our dataset. It was inferred from the ROC curve that SVM classifiers outperformed other classifier models.

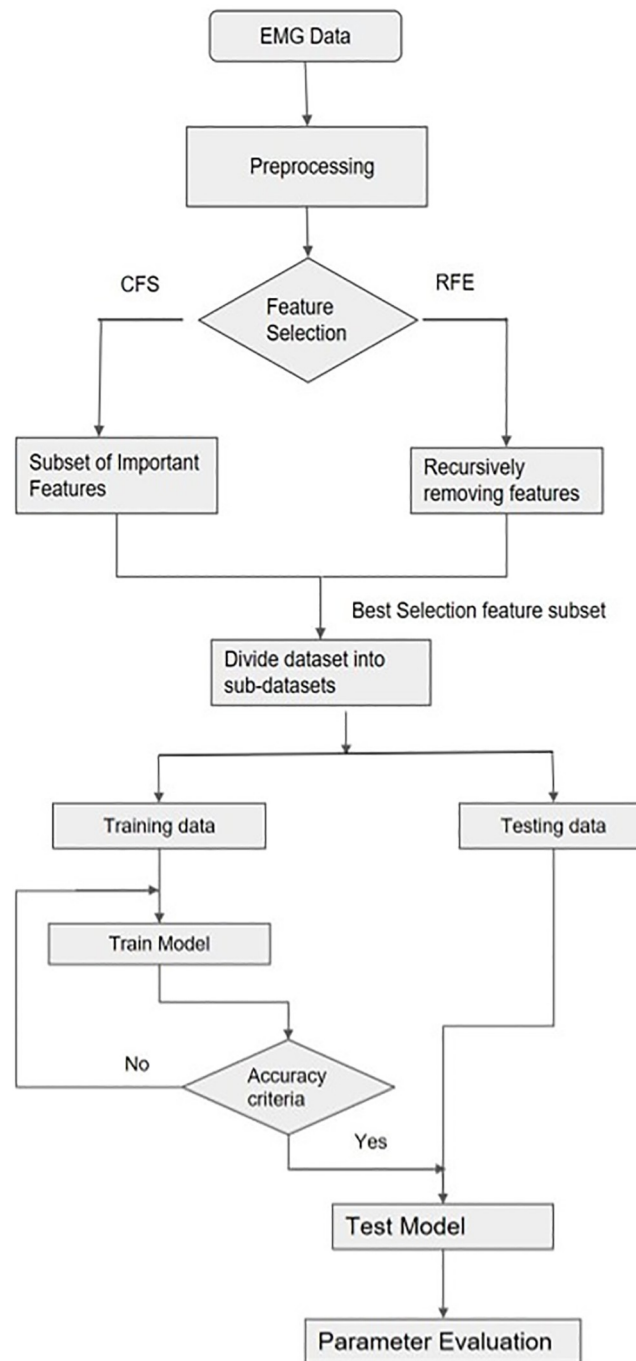


Fig 5. Flow chart of feature selection-based model development and identification of parameters of muscle importance for skill classification.

<https://doi.org/10.1371/journal.pone.0267936.g005>

Statistical analysis

An initial MANOVA examined 1) Total Time, 2) RMS, 3) Range, 4) Frequency, 5) CMW, 6) AWS as dependent variables, and 1) Skill Level (Expert, Intermediate, and Novice) and 2) Surgical Test (Knot Tying, Pegboard, Robotic suturing) as independent variables. An overall 3 X 4

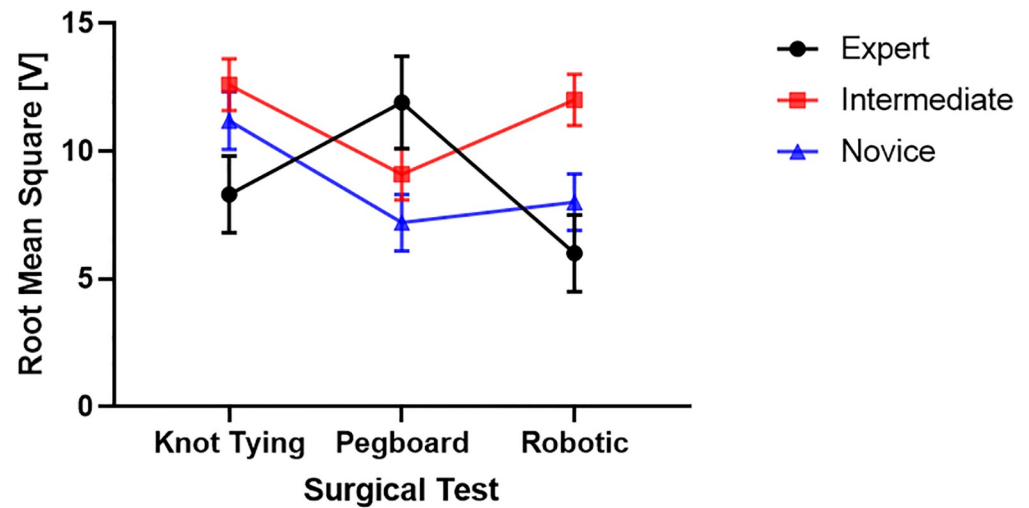


Fig 6. Linear variability as measured by RMS of EMG signals between groups of 3 skill levels.

<https://doi.org/10.1371/journal.pone.0267936.g006>

(Skill Level X Surgical Task) multivariate analysis of variance (MANOVA) with repeated measures design was applied to all data to investigate skill level using EMG and accelerometry signals. The main and interaction effects were analyzed using JMP[®] Pro (SAS Institute Inc., Cary, NC, 2020), with the significance level of $p = 0.05$.

Results

Six linear EMG features based on temporal and frequency domains (CMW, AWS, total time, RMS, Signal Range, and dominant frequency) (Table 1) and eleven nonlinear variability features were extracted from accelerometer data (Table 2). Feature selection attributed to important features from the dataset, which enhanced the predictive model's performance.

We found significant interaction effects among skill level and surgical test for total time ($F(4,24) = 218, p < 0.01$), RMS ($F(4,24) = 2.97, p < 0.05$), Range ($F(4,24) = 5.01, p < 0.01$), frequency ($F(4,24) = 337, p < 0.01$), CMW ($F(4,24) = 14.8, p < 0.01$), AWS ($F(4,24) = 2.49, p < 0.05$).

Fig 6 shows experts showed fewer fluctuations in muscle firing (RMS) in knot tying and robotic suturing tasks. In contrast, there was somewhat higher variability during pegboard tasks than intermediate skilled and novice surgeons.

Frequency profiles of experts, novice, and intermediate skilled surgeons (Fig 7) show EMG signal dominant frequencies during the performance of three surgical tasks.

We found that the cumulative muscular workload was higher for the novice group (Fig 8). Post hoc comparisons using the Tukey HSD test indicated that the CMW was significantly higher for the novice group than the intermediate group for pegboard ($M = 2494810, SD = 149982$ versus $M = 1320467, SD = 143597$) and robotic suturing ($M = 3304601, SD = 149980$ versus $M = 1528405, SD = 145634$) (Fig 7). We also found that the expert group's CMW ($M = 669617, SD = 203076$) was significantly lower than intermediate ($M = 1528405, SD = 145634$) and novice ($M = 3304601, SD = 149980$) groups.

Post hoc comparisons using the Tukey HSD test indicated that the AWS was significantly lower for the expert group than the intermediate group for the robotic suturing task ($M = 9400, SD = 2226$ versus $M = 18960, SD = 1596$) (Fig 9).

We also found significant interaction effects among skill level and the three tasks for MSE ($F(4,24) = 19.3, p < 0.01$). Post hoc comparisons using the Tukey HSD test indicated that the

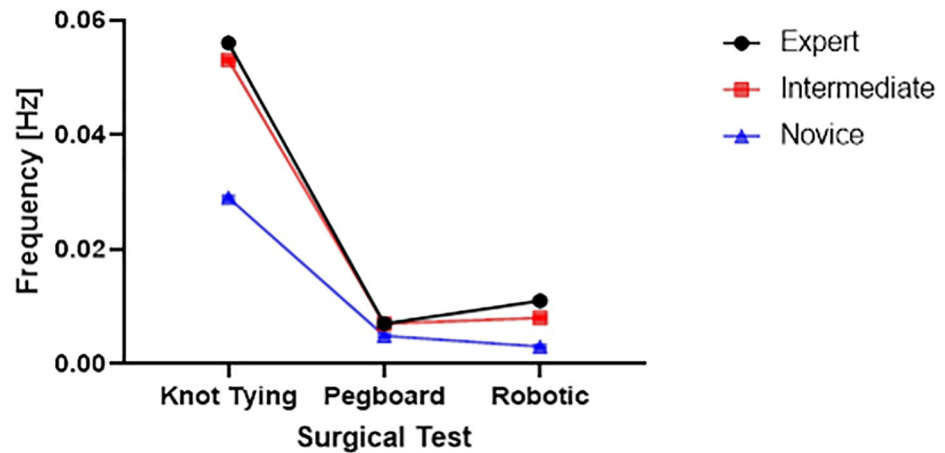


Fig 7. The dominant frequency of EMG signals during the performance of 3 surgical tasks (knot tying, pegboard transfer, and robotic suturing).

<https://doi.org/10.1371/journal.pone.0267936.g007>

MSE was significantly lower for the intermediate ($M = 2.09$, $SD = 0.09$) group compared to the novice ($M = 2.9$, $SD = 0.09$) group for the knot tying task (Fig 10).

Feature selection methods as i) Correlation-based filters (CFS) and ii) recursive feature elimination (RFE) identified important features with high accuracy. Both CFS (Fig 11) and RFE algorithms (Table 3) were tested by implementing machine learning algorithms such as Random Forest classifier, Naïve Bayes, and support vector machines. Inbuilt feature importance methods were utilized for forecasting for the most valuable variables and associated muscles on the model. The models were built using inbuilt feature importance methods, where only significant features are selected, and skill prediction results were evaluated. We found that nonlinear features of complexity such as approximate entropy, multiscale entropy, and sample entropy carried higher importance weightage for accurate classification of surgical

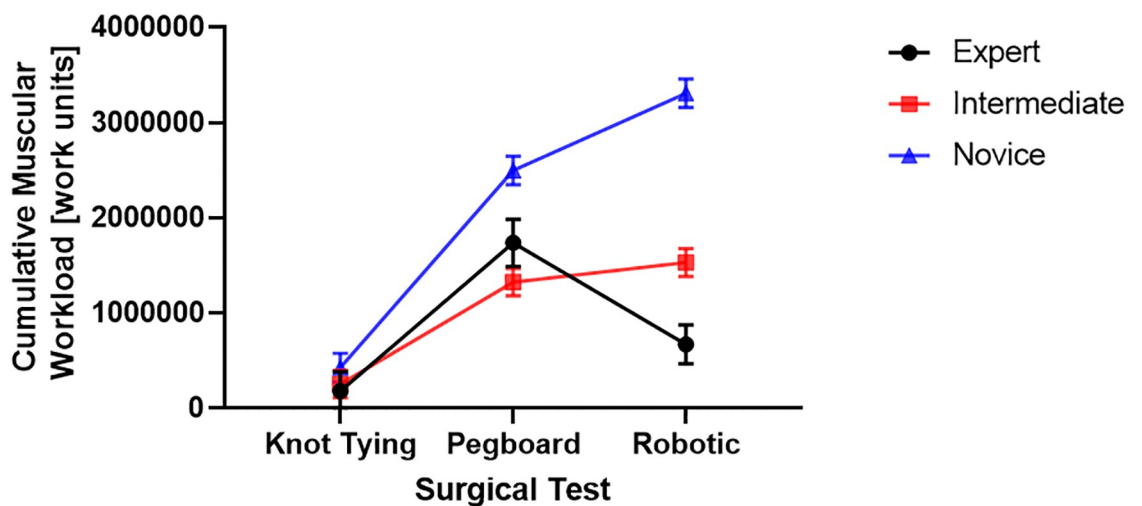


Fig 8. The cumulative muscular workload for completing surgical tasks (knot tying, pegboard transfer, and robotic suturing) by three groups of different skill levels (expert, intermediate, and novice).

<https://doi.org/10.1371/journal.pone.0267936.g008>

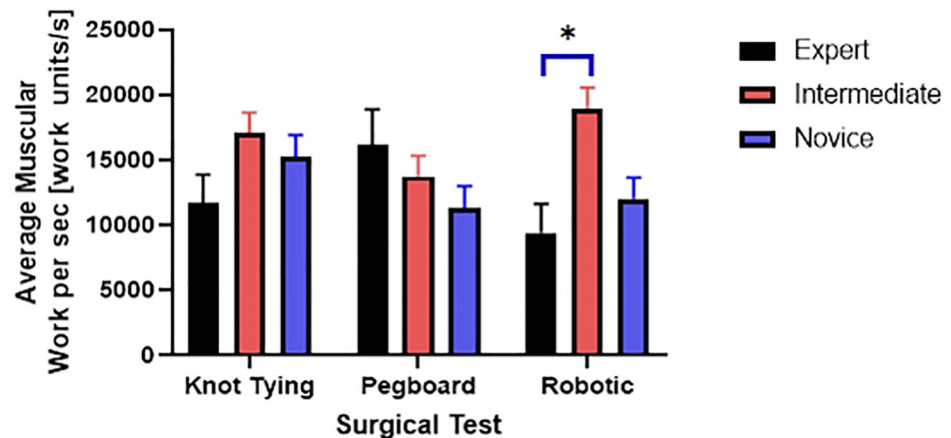


Fig 9. Average muscular work done per second and significant differences in the robotic suturing task.

<https://doi.org/10.1371/journal.pone.0267936.g009>

skill. We also found that nonlinear features such as complexity as measured by approximate entropy (ApEn), sample entropy (SampEn), and multiscale entropy (MSE) were significantly correlated (Fig 11). We also found long and short Lyapunov exponents were highly correlated.

Prediction accuracies for surgical skill were determined for all six muscles (biceps, triceps, deltoid, ECU, FCU, and TE) for all three machine learning algorithms using selected features (Table 4). Different ML algorithms utilized various muscle-derived features for higher surgical skill classification. We found i) random forest classifier showed higher accuracies with ECU-61%, Deltoid-55%, and TE-55%, ii) support vector machines showed higher accuracies with Deltoid-57%, biceps-45%, ECU-41%, iii) Naive Bayes showed higher accuracies with biceps-47%, and ECU-43% (Table 4). Since TE location hinders surgical tasks, thus was excluded for further analysis.

Since the three important muscles with the highest accuracies were found to be i) ECU, ii) deltoid and iii) biceps. Our investigation of muscle combinations revealed that the combination of three muscles for skill evaluation resulted in high accuracies for RFC (52%), SVM

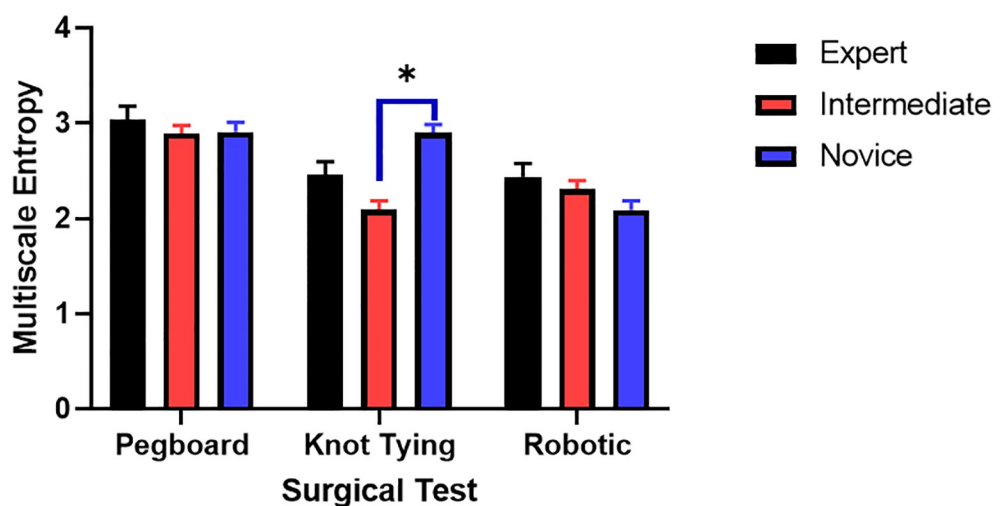


Fig 10. High complexity measured by multiscale entropy.

<https://doi.org/10.1371/journal.pone.0267936.g010>

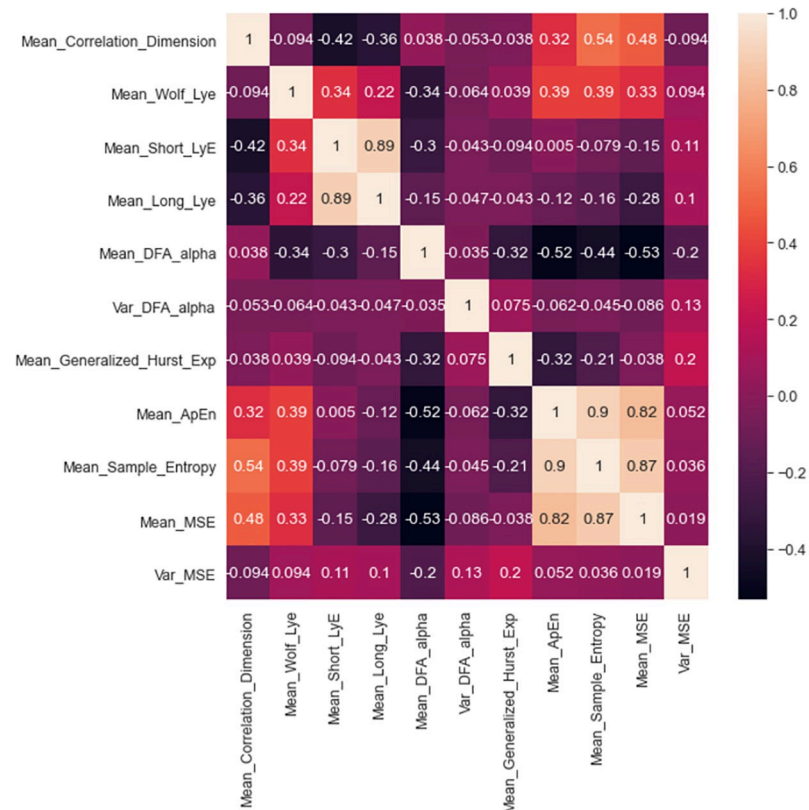


Fig 11. A correlation matrix shows correlation coefficients between features.

<https://doi.org/10.1371/journal.pone.0267936.g011>

(50%), and naïve Bayes (40%) algorithm (Table 5). However, combining two muscle locations for sensors such as ECU and deltoid resulted in higher classification accuracy for RFC (58%) and SVM (50%). A combination of ECU and biceps resulted in higher accuracy for the naïve Bayes algorithm (41%) (Table 5).

The receiver operating curves (ROC) evaluated models for their false positive and true positive rates for three surgical skill classifications (novice, intermediate, and expert). We found the combination of i) ECU and biceps (Fig 12), ii) ECU and deltoid (Fig 13), and iii) ECU, deltoid, and biceps (Fig 14) resulted in the best classification accuracy with minimum sensors on

Table 3. Feature importance as determined by recursive feature elimination method.

| Feature Priority | Weights |
|----------------------------|---------|
| Mean_Apen | 0.0453 |
| Mean_MSE | 0.0475 |
| Mean_Sample_Entropy | 0.0477 |
| Var_MSE | 0.0049 |
| Mean_Wolf_Lye | 0.0509 |
| Mean_Short_Lye | 0.0524 |
| Mean_Long_Lye | 0.0544 |
| Mean_Correlation_Dimension | 0.0596 |
| Mean_Generalized_Hurst_Exp | 0.0635 |

<https://doi.org/10.1371/journal.pone.0267936.t003>

Table 4. Performance metrics were evaluated by the three ML models (random forest classifier, SVM, and Naïve Bayes) with six muscle datasets. Highest accuracies are heighted in the table as bold.

| Muscles | Random Forest Classifier | | | | Support Vector Machine | | | | Naive Bayes | | | |
|---------|--------------------------|-----------|--------|----------|------------------------|-----------|--------|----------|-------------|-----------|--------|----------|
| | Accuracy | Precision | Recall | F1 score | Accuracy | Precision | Recall | F1 score | Accuracy | Precision | Recall | F1 score |
| Biceps | 0.40 | 0.27 | 0.33 | 0.29 | 0.45 | 0.46 | 0.39 | 0.37 | 0.47 | 0.46 | 0.45 | 0.45 |
| Deltoid | 0.55 | 0.40 | 0.43 | 0.41 | 0.57 | 0.39 | 0.45 | 0.41 | 0.28 | 0.3 | 0.25 | 0.26 |
| ECU | 0.61 | 0.73 | 0.53 | 0.51 | 0.41 | 0.30 | 0.35 | 0.30 | 0.43 | 0.40 | 0.41 | 0.39 |
| FCU | 0.45 | 0.31 | 0.37 | 0.32 | 0.35 | 0.25 | 0.30 | 0.24 | 0.35 | 0.25 | 0.26 | 0.25 |
| TE | 0.55 | 0.39 | 0.44 | 0.40 | 0.36 | 0.23 | 0.29 | 0.25 | 0.36 | 0.32 | 0.33 | 0.30 |
| Triceps | 0.51 | 0.36 | 0.41 | 0.37 | 0.53 | 0.35 | 0.42 | 0.38 | 0.31 | 0.31 | 0.28 | 0.28 |

<https://doi.org/10.1371/journal.pone.0267936.t004>

Table 5. This table shows the performance of a combination of muscles for surgical skill evaluation for each model separately. Accuracy, precision, recall, and F1 score are reported for all 3 classification models. Highest accuracies are highlighted in the table as bold.

| Algorithm | Deltoid, ECU, and Biceps | | | | ECU and deltoid | | | | ECU and biceps | | | |
|---------------|--------------------------|-----------|--------|----------|-----------------|-----------|--------|----------|----------------|-----------|--------|----------|
| | Accuracy | Precision | Recall | F1 score | Accuracy | Precision | Recall | F1 score | Accuracy | Precision | Recall | F1 score |
| Random Forest | 0.52 | 0.54 | 0.46 | 0.45 | 0.58 | 0.46 | 0.51 | 0.51 | 0.53 | 0.47 | 0.45 | 0.28 |
| SVM | 0.50 | 0.66 | 0.43 | 0.41 | 0.5 | 0.33 | 0.4 | 0.36 | 0.45 | 0.47 | 0.39 | 0.37 |
| Naive Bayes | 0.41 | 0.43 | 0.43 | 0.43 | 0.29 | 0.33 | 0.33 | 0.29 | 0.44 | 0.43 | 0.39 | 0.39 |

<https://doi.org/10.1371/journal.pone.0267936.t005>

upper extremities. Random forest resulted in the highest accuracy for skill classification among surgeons. We also found muscle ECU and deltoid could achieve an accuracy of 58% during classification.

Discussion

This work presents machine learning-based methods for selecting the best pairs of sensor features and sensor locations for accurate skill classification during three different surgical tasks. The primary goal of this study was to determine if i) EMG-derived variables such as total time, CMW, AWS, RMS, Range, Frequency, and ii) motion sensor-derived nonlinear variables such as Lyapunov exponent, MSE, correlation dimension can successfully distinguish surgical skill

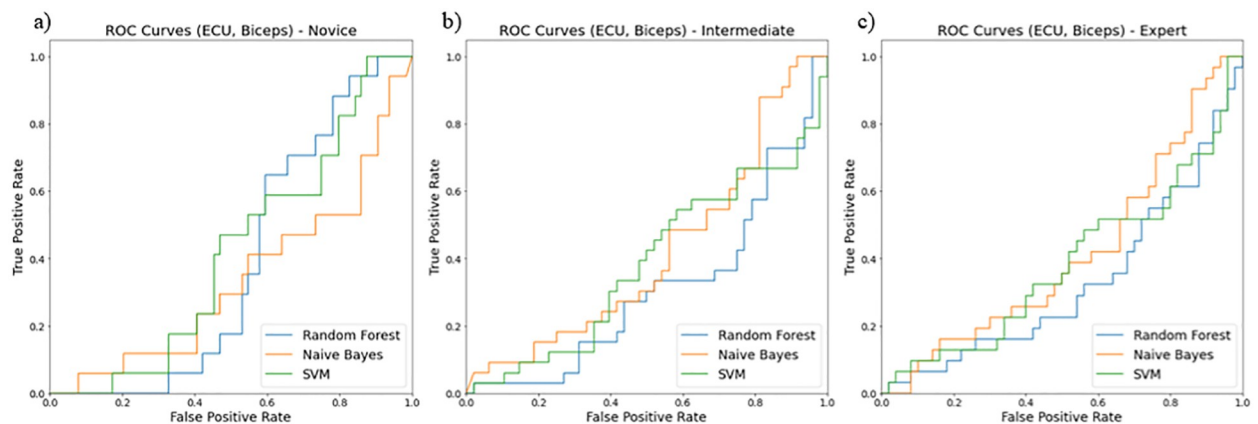


Fig 12. ROC curves showing performance of two muscles (ECU and Biceps) for skill level classification among a) novice, b) intermediate, and c) expert using three classifiers random forest (blue line), Naïve Bayes (orange line), SVM (green line).

<https://doi.org/10.1371/journal.pone.0267936.g012>

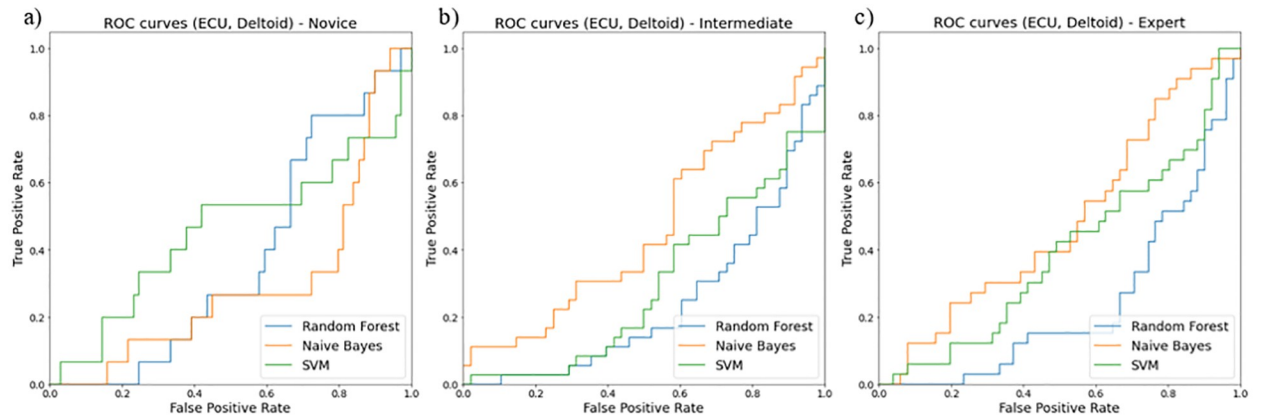


Fig 13. ROC curves showing performance of two muscles (ECU and Deltoid) for skill level classification among a) novice, b) intermediate, and c) expert using three classifiers random forest (blue line), Naïve Bayes (orange line), SVM (green line).

<https://doi.org/10.1371/journal.pone.0267936.g013>

levels. We utilized wearable sensors integrated with EMG and inertial sensors such as accelerometers to classify skill levels in 26 surgeons of three different skill levels (novice, intermediate, and expert). Our work suggests that EMG and accelerometer sensors and nonlinear variability measures can capture the predictability and irregularity of movement fluctuations and thus facilitate surgical skill assessment. Similar studies have adopted data fusion algorithms from wearable inertial and surface EMG sensors to evaluate upper extremity motor function [42].

Our secondary goal was to select the best features and sensor locations for higher classification accuracy. The quality of these selections was tested against features from both right and left limbs and both 3-directional accelerometer and EMG sensors. Nonlinear movement variability features extracted from accelerometer data such as approximate entropy, sample entropy, and multiscale entropy represent a high amount of information to classify skill groups (novice, intermediate, and expert). We also found the most optimal location for sensors to be i) ECU and deltoid, ii) ECU, deltoid, and biceps, and iii) ECU and biceps. These positions resulted in higher accuracy of skill classification among surgeons. As shown in Table 5, for the deltoid, ECU, biceps dataset, the accuracy for RFC is 52% which is 2% more than SVM and

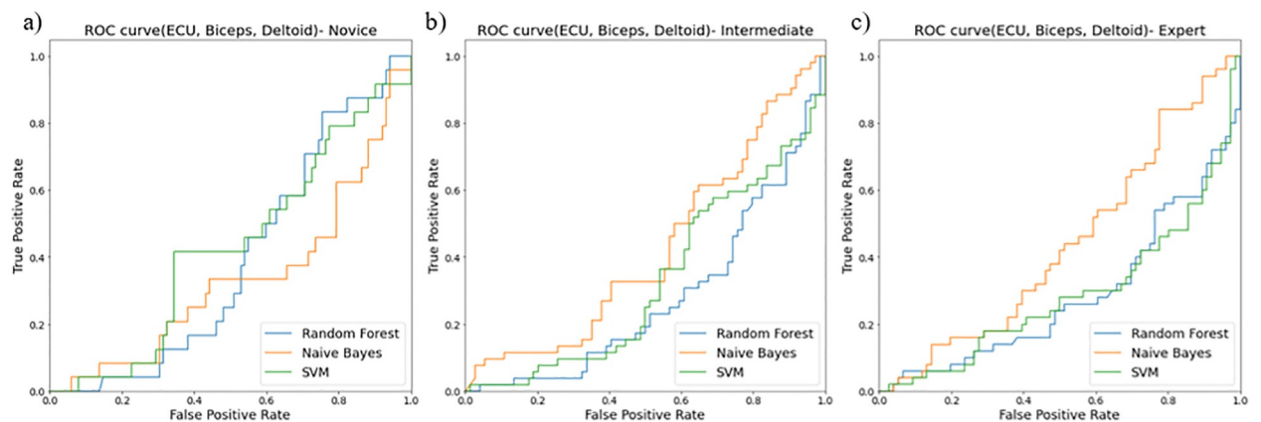


Fig 14. ROC curves showing performance of two muscles (ECU, Biceps and Deltoid) for skill level classification among a) novice, b) intermediate, and c) expert using three classifiers random forest (blue line), Naïve Bayes (orange line), SVM (green line).

<https://doi.org/10.1371/journal.pone.0267936.g014>

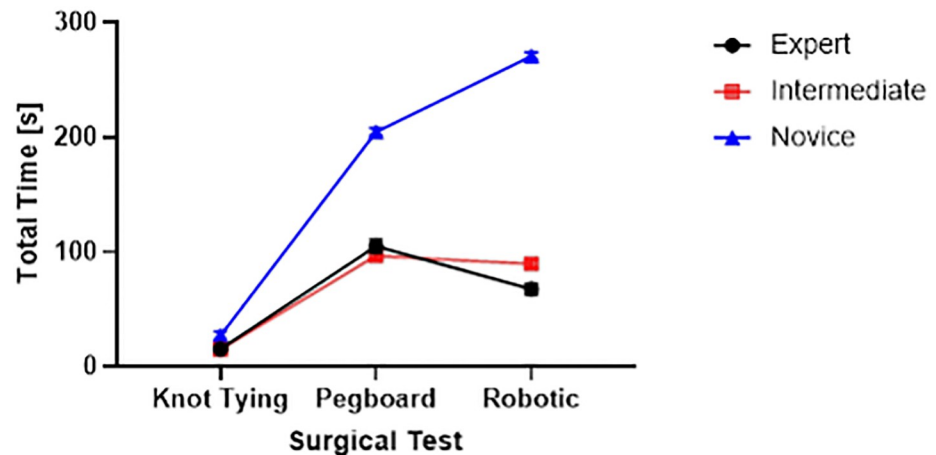


Fig 15. Total time taken for completing surgical tasks (knot tying, pegboard transfer and robotic suturing) by three groups of different skill level (expert, intermediate and novice).

<https://doi.org/10.1371/journal.pone.0267936.g015>

11% more than the Naïve Bayes model. For ECU and deltoid datasets, the accuracy for RFC is about 58% which is 8% more than SVM and 29% more than the Naïve Bayes model. Similarly, for ECU and biceps dataset, the accuracy for RFC is about 53% which is 8% more than SVM and 9% more than the Naïve Bayes model. We discarded the sensor at the TE position since it produces hindrance during surgical tasks and may not be feasible for skill evaluation during surgeries. Our study highlights the importance of nonlinear accelerometer variability measures for skill classification since all accelerometer features showed higher priority using RFE (Table 3).

Training highly skilled and competent surgeons are vital to ensure good quality of patient care and minimize treatment disparities. Surgeons are required to master specific skills during residency and surgical training. However, there is a lack of objective tools utilizing wearable sensors and signal processing which can help differentiate the surgical level of expertise among surgeons. Quantifying and documenting clinical competence and identifying peculiar muscle and movement-based features associated with skill level is a challenging area of research. The studies on clinical competence are limited due to the lack of wearable validated tools. Learning surgical skills involves continuous skills and improvement through feedback from supervising surgeons. Some medical schools have adopted tools like Objective structured assessment of technical skills (OSATS) [5], which are graded by specific criteria like respect for tissue, time and motion, instrument handling, flow in operation, and overall performance [43]. Some other research groups have previously assessed surgical skills utilizing techniques to analyze movement from videos [44–46] and wearable sensors [47–49].

Statistical differences in features among three skill groups

We found that the novice surgeon group took more time to complete the pegboard and robotic suturing task, but not the knot-tying task (Fig 15). Post hoc analysis revealed that the novice group took significantly more time than the expert and intermediate groups for pegboard and robotic suturing tasks. These results suggest that knot tying is equally tricky or straightforward for all groups, thus completing in somewhat similar total times. On the other hand, linear variability measures such as RMS revealed that the expert group had higher muscle firing fluctuations during pegboard transfers, indicating higher adaptability (Fig 6) and task experience.

Experts are well trained and may have learned and practiced more than one strategy to transfer cubes during pegboard transfer tasks. Such variability in muscle activations suggests a high degree of freedom and an extensive range of muscle activations among the expert group. We also found muscle firing frequency among surgeons decreased with fewer years of surgical experience (Fig 7). We observed that the novice group had significantly lower muscle firing frequency than the intermediate and expert groups for robotic suturing and knot tying tasks. This could be potentially due to less practice or experience of novice group. Previously researchers have reported that a lower firing rate is observed with impaired muscle function or reduced neuromuscular recruitment [50]. Thus, muscle firing frequency was found to be higher with a high level of surgical expertise. Muscular workload as measured by CMW was significantly higher for the novice group compared to the intermediate group in pegboard transfer. Thus, delineating more muscular work done by novice group compared to intermediate skill group for pegboard transfer task. Besides, CMW was also significantly higher than the expert group in the robotic suturing task (Fig 8). There may be a change in activation patterns after surgical training, with proximal becoming more relaxed and distal muscle groups becoming more active [51], thereby reducing muscular workload or musculoskeletal strain in a surgical expert group compared to novice and intermediate skill groups. When comparing the average rate of muscular work done (work units/s), we found that the intermediate group performed robotic suturing significantly faster than the expert group (Fig 9). This high muscular work rate in the intermediate group may lead to muscle fatigue, contributing to failed surgical procedures.

Correlation-based filter feature selection suggested that nonlinear features such as complexity were significantly correlated (Fig 11) and attained high importance when tested with the recursive feature elimination method (Table 3). High variability in MSE in experts demonstrates that experts can control motor strategies according to task requirements. Higher variability in the correlation dimension of time series is associated with more exchange of dynamical system information among expert groups (Fig 10). Interestingly, previous studies have reported that entropy-based features are more accurate for robotic suturing but not for knot tying [43]. However, the entropy analysis was limited to a single scale using approximate entropy [52]. In this study, we utilized MSE, which provides insights into the complexity of fluctuations over a range of time scales. MSE is advantageous since the time scale of relevance in the surgical task was unknown. MSE could differentiate knot tying skill levels among novice and intermediate groups (Fig 10).

Using features from CFS and RFE allowed identifying important sensor locations that could classify skill levels with higher accuracy. However, we discarded sensor data from the thenar eminence (TE) muscle since it is impractical to conduct surgeries with the sensor at this location due to hindrance. However, ECU, biceps, and deltoid showed the highest accuracy in classification compared to other sites (Tables 4 and 5). These sensor locations are practically feasible and expected to cause the slightest hindrance during surgical tasks-related movements. Our future studies with sensors at these locations will be helpful to build machine learning classification models for surgical skill assessment. An essential contribution of this study is to present new quantitative predictors from muscle activity and movement accelerations to different skill levels in surgeons. This is important over the last decade since increasing sub-specializations and expanded use of MIS techniques have impacted residency training and raised concerns about objectively evaluating residents' surgical experience or skill level. One of the significant limitations of this study is that it has primarily focused on the surgical task, ignoring mental workload and the support from assistants. Surgical assistants help surgeons, and their interaction is an additional passive task ignored in this study. Future studies are warranted to assess clinical skills better and address these deficits.

Surgical tasks are related to high stress, muscular workload, and variability in surgical task performance. Although objective assessment tools for surgical competence such as EMG and accelerometers are available, information is lacking on important sensor-based features and attachment sites of these wearable systems, which could differentiate surgical skills with the highest accuracy. This study recruited 26 surgeons with three different skill levels and utilized wearable sensors, nonlinear movement variability features, feature selection methods, and classifiers to identify sensor-sites that could distinguish skill levels with the highest accuracy. Advancements in signal processing for movement variability and muscle activation parameters can help develop further assessments. This research can potentially lead to wearable sensor-assisted high fidelity simulation technology, which can evaluate technical training issues with skill decay addressing specific training deficiencies in surgeons. These wearable sensor-based simulation training can assess progress in skill learning or deficits in confidence.

Supporting information

S1 Table. Default ML hyperparameters used for classification.
(DOCX)

Acknowledgments

We are grateful to Dr. Rayman Calph, Dr. Pengbo Jiang, Dr. Perry Xu, Dr. Daniel Haik, Dr. Andrew Brevik, Dr. Akhil Peta, Dr. Shlomi Tapiero, Dr. Jaime Landman and all UCI surgeons for their help in volunteering and facilitating data collection. We are also thankful to Michael Shiraishi and Christopher Hoang for data collection.

Author Contributions

Conceptualization: Rahul Soangra, Emmanuel B. John.

Data curation: R. Sivakumar, E. R. Anirudh, Sai Viswanth Reddy Y.

Formal analysis: Rahul Soangra, E. R. Anirudh, Sai Viswanth Reddy Y.

Funding acquisition: Emmanuel B. John.

Investigation: Rahul Soangra.

Methodology: Rahul Soangra, R. Sivakumar, E. R. Anirudh, Sai Viswanth Reddy Y., Emmanuel B. John.

Resources: E. R. Anirudh.

Software: Rahul Soangra.

Visualization: Rahul Soangra.

Writing – original draft: Rahul Soangra, Sai Viswanth Reddy Y.

Writing – review & editing: Rahul Soangra, Emmanuel B. John.

References

1. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013; 369(15):1434–42. <https://doi.org/10.1056/NEJMsa1300625> PMID: 24106936.
2. Fonseca AL, Reddy V, Longo WE, Gusberg RJ. Graduating general surgery resident operative confidence: perspective from a national survey. *Journal of Surgical Research*. 2014; 190(2):419–28. <https://doi.org/10.1016/j.jss.2014.05.014> PMID: 24908164

3. Bell RH Jr. Why Johnny cannot operate. *Surgery*. 2009; 146(4):533–42. Epub 2009/10/01. <https://doi.org/10.1016/j.surg.2009.06.044> PMID: 19789010.
4. Bell RH Jr., Biester TW, Tabuenca A, Rhodes RS, Cofer JB, Britt LD, et al. Operative experience of residents in US general surgery programs: a gap between expectation and experience. *Ann Surg*. 2009; 249(5):719–24. Epub 2009/04/24. <https://doi.org/10.1097/SLA.0b013e3181a38e59> PMID: 19387334.
5. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997; 84(2):273–8. <https://doi.org/10.1046/j.1365-2168.1997.02502.x> PMID: 9052454.
6. Niitsu H, Hirabayashi N, Yoshimitsu M, Mimura T, Taomoto J, Sugiyama Y, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today*. 2013; 43(3):271–5. <https://doi.org/10.1007/s00595-012-0313-7> PMID: 22941345.
7. Dill-Macky A, Hsu CH, Neumayer LA, Nfonsam VN, Turner AP. The Role of Implicit Bias in Surgical Resident Evaluations. *J Surg Educ*. 2021. Epub 20211229. <https://doi.org/10.1016/j.jsurg.2021.12.003> PMID: 34973900.
8. Menon D, Wong DJN, Harris S, Sahni A, Bedford JR, Cortes L, et al. Developing and validating subjective and objective risk-assessment measures for predicting mortality after major surgery: An international prospective cohort study. *PLOS Medicine*. 2020; 17(10). <https://doi.org/10.1371/journal.pmed.1003253> PMID: 33057333
9. Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, Muller P-A. Evaluating Surgical Skills from Kinematic Data Using Convolutional Neural Networks. *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018. Lecture Notes in Computer Science 2018*. p. 214–21.
10. Yanik E, Intes X, Kruger U, Yan P, Diller D, Van Voorst B, et al. Deep neural networks for the assessment of surgical skills: A systematic review. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*. 2021. <https://doi.org/10.1177/15485129211034586>
11. Bissonnette V, Mirchi N, Ledwos N, Alsidieri G, Winkler-Schwartz A, Del Maestro RF, et al. Artificial Intelligence Distinguishes Surgical Training Levels in a Virtual Reality Spinal Task. *J Bone Joint Surg Am*. 2019; 101(23):e127. <https://doi.org/10.2106/JBJS.18.01197> PMID: 31800431.
12. Lee D, Yu HW, Kwon H, Kong HJ, Lee KE, Kim HC. Evaluation of Surgical Skills during Robotic Surgery by Deep Learning-Based Multiple Surgical Instrument Tracking in Training and Actual Operations. *J Clin Med*. 2020; 9(6). Epub 20200623. <https://doi.org/10.3390/jcm9061964> PMID: 32585953.
13. Lavanchy JL, Zindel J, Kirtac K, Twick I, Hosgor E, Candinas D, et al. Automation of surgical skill assessment using a three-stage machine learning algorithm. *Scientific Reports*. 2021; 11(1). <https://doi.org/10.1038/s41598-021-84295-6> PMID: 33664317
14. Davids J, Makariou SG, Ashrafian H, Darzi A, Marcus HJ, Giannarou S. Automated Vision-Based Microsurgical Skill Analysis in Neurosurgery Using Deep Learning: Development and Preclinical Validation. *World Neurosurg*. 2021; 149:e669–e86. Epub 20210212. <https://doi.org/10.1016/j.wneu.2021.01.117> PMID: 33588081.
15. Fard MJ, Ameri S, Darin Ellis R, Chinnam RB, Pandya AK, Klein MD. Automated robot-assisted surgical skill evaluation: Predictive analytics approach. *Int J Med Robot*. 2018; 14(1). Epub 20170629. <https://doi.org/10.1002/rcs.1850> PMID: 28660725.
16. Fekri P, Dargahi J, Zadeh M. Deep Learning-Based Haptic Guidance for Surgical Skills Transfer. *Frontiers in Robotics and AI*. 2021; 7. <https://doi.org/10.3389/frobt.2020.586707> PMID: 33553246
17. Arbelaez-Garces G, Joseph D, Camargo M, Tran N, Morel L. Contribution to the objective assessment of technical skills for surgery students: An accelerometer based approach. *International Journal of Industrial Ergonomics*. 2018; 64:79–88. <https://doi.org/10.1016/j.ergon.2017.12.002>
18. S AA, P S, V S, S SK, A S, Akl TJ, et al. Electrodermal Activity Based Pre-surgery Stress Detection Using a Wrist Wearable. *IEEE J Biomed Health Inform*. 2020; 24(1):92–100. Epub 2019/01/23. <https://doi.org/10.1109/JBHI.2019.2893222> PMID: 30668508.
19. Wijsman J, Grundlehner B, Penders J, Hermens H. Trapezius muscle EMG as predictor of mental stress. *ACM Transactions on Embedded Computing Systems*. 2013; 12(4):1–20. <https://doi.org/10.1145/2485984.2485987>
20. Taib MF, Bahn S, Yun MH. The effect of psychosocial stress on muscle activity during computer work: Comparative study between desktop computer and mobile computing products. *Work*. 2016; 54(3):543–55. Epub 2016/07/04. <https://doi.org/10.3233/WOR-162334> PMID: 27372893.
21. Cavanaugh JT, Guskiewicz KM, Giuliani C, Marshall S, Mercer V, Stergiou N. Detecting altered postural control after cerebral concussion in athletes with normal postural stability. *Br J Sports Med*. 2005; 39(11):805–11. Epub 2005/10/26. <https://doi.org/10.1136/bjsm.2004.015909> PMID: 16244188.

22. Cavanaugh JT, Guskiewicz KM, Giuliani C, Marshall S, Mercer VS, Stergiou N. Recovery of postural control after cerebral concussion: new insights using approximate entropy. *J Athl Train.* 2006; 41(3):305–13. Epub 2006/10/18. PMID: [17043699](#).
23. Frames C, Soangra R, Lockhart T, Lach J, Ha D, Roberto K, et al. Dynamical Properties of Postural Control in Obese Community-Dwelling Older Adults. *Sensors.* 2018; 18(6). <https://doi.org/10.3390/s18061692> PMID: [29794998](#)
24. Khandoker AH, Palaniswami M, Begg RK. A comparative study on approximate entropy measure and poincare plot indexes of minimum foot clearance variability in the elderly during walking. *J Neuroeng Rehabil.* 2008; 5:4. Epub 2008/02/05. <https://doi.org/10.1186/1743-0003-5-4> PMID: [18241348](#).
25. Pincus S. Approximate entropy (ApEn) as a complexity measure. *Chaos.* 1995; 5(1):110–7. Epub 1995/03/01. <https://doi.org/10.1063/1.166092> PMID: [12780163](#).
26. Pincus SM. Approximate entropy as a measure of irregularity for psychiatric serial metrics. *Bipolar Disord.* 2006; 8(5 Pt 1):430–40. Epub 2006/10/18. <https://doi.org/10.1111/j.1399-5618.2006.00375.x> PMID: [17042881](#).
27. Boostani R, Moradi MH. Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiol Meas.* 2003; 24(2):309–19. Epub 2003/06/19. <https://doi.org/10.1088/0967-3334/24/2/307> PMID: [12812417](#).
28. Zardoshti-Kermani M, Wheeler BC, Badie K, Hashemi RM. EMG feature evaluation for movement control of upper extremity prostheses. *IEEE Transactions on Rehabilitation Engineering.* 1995; 3(4):324–33. <https://doi.org/10.1109/86.481972>
29. Lee GI, Lee MR, Clanton T, Sutton E, Park AE, Marohn MR. Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries. *Surg Endosc.* 2014; 28(2):456–65. Epub 2013/11/08. <https://doi.org/10.1007/s00464-013-3213-z> PMID: [24196542](#).
30. Park A, Lee G, Seagull FJ, Meenaghan N, Dexter D. Patients benefit while surgeons suffer: an impending epidemic. *J Am Coll Surg.* 2010; 210(3):306–13. Epub 2010/03/03. <https://doi.org/10.1016/j.jamcollsurg.2009.10.017> PMID: [20193893](#).
31. Hudgins B, Parker P, Scott RN. A new strategy for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering.* 1993; 40(1):82–94. <https://doi.org/10.1109/10.204774> PMID: [8468080](#)
32. Oskoei MA, Huosheng H. Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb. *IEEE Transactions on Biomedical Engineering.* 2008; 55(8):1956–65. <https://doi.org/10.1109/TBME.2008.919734> PMID: [18632358](#)
33. Tkach D, Huang H, Kuiken TA. Study of stability of time-domain features for electromyographic pattern recognition. *Journal of NeuroEngineering and Rehabilitation.* 2010; 7(1). <https://doi.org/10.1186/1743-0003-7-21> PMID: [20492713](#)
34. Grassberger P, Procaccia I. Characterization of Strange Attractors. *Physical Review Letters.* 1983; 50(5):346–9. <https://doi.org/10.1103/PhysRevLett.50.346>
35. Pecora LM, Carroll TL. Synchronization of chaotic systems. *Chaos: An Interdisciplinary Journal of Non-linear Science.* 2015; 25(9):097611. <https://doi.org/10.1063/1.4917383> PMID: [26428564](#)
36. Bader R. *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology*: Springer Berlin Heidelberg; 2013.
37. Lipsitz LA. Dynamics of Stability: The Physiologic Basis of Functional Health and Frailty. *The Journals of Gerontology: Series A.* 2002; 57(3):B115–B25. <https://doi.org/10.1093/gerona/57.3.b115> PMID: [11867648](#)
38. Lipsitz LA. Loss of 'Complexity' and Aging. *Jama.* 1992; 267(13):1806. <https://doi.org/10.1001/jama.1992.03480130122036> PMID: [1482430](#)
39. Kitano H. Systems Biology: A Brief Overview. *Science.* 2002; 295(5560):1662–4. <https://doi.org/10.1126/science.1069492> PMID: [11872829](#)
40. Rosenstein MT, Collins JJ, De Luca CJ. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena.* 1993; 65(1–2):117–34. [https://doi.org/10.1016/0167-2789\(93\)90009-p](https://doi.org/10.1016/0167-2789(93)90009-p)
41. Schenk J, Kaiser M, Rigoll G. Selecting Features in On-Line Handwritten Whiteboard Note Recognition: SFS or SFFS? 2009 10th International Conference on Document Analysis and Recognition 2009. p. 1251–4.
42. Li Y, Zhang X, Gong Y, Cheng Y, Gao X, Chen X. Motor Function Evaluation of Hemiplegic Upper-Extremities Using Data Fusion from Wearable Inertial and Surface EMG Sensors. *Sensors.* 2017; 17(3):582. <https://doi.org/10.3390/s17030582> PMID: [28335394](#)
43. Zia A, Sharma Y, Bettadapura V, Sarin EL, Essa I. Video and accelerometer-based motion analysis for automated surgical skills assessment. *International Journal of Computer Assisted Radiology and Surgery.* 2018; 13(3):443–55. <https://doi.org/10.1007/s11548-018-1704-z> PMID: [29380122](#)

44. Sharma Y, Bettadapura V, Plötz T, Hammerla NY, Mellor S, McNaney R, et al., editors. Video Based Assessment of OSATS Using Sequential Motion Textures 2014.
45. Zia A, Sharma Y, Bettadapura V, Sarin EL, Clements MA, Essa I. Automated Assessment of Surgical Skills Using Frequency Analysis. 2015; 9349:430–8. https://doi.org/10.1007/978-3-319-24553-9_53
46. Sharma Y, Plötz T, Hammerld N, Mellor S, McNaney R, Olivier P, et al., editors. Automated surgical OSATS prediction from videos. 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI); 2014 29 April–2 May 2014.
47. Trejos AL, Patel RV, Naish MD, Schlachta CM, editors. Design of a sensorized instrument for skills assessment and training in minimally invasive surgery. 2008 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics; 2008 19–22 Oct. 2008.
48. Nisky I, Che Y, Quek ZF, Weber M, Hsieh MH, Okamura AM, editors. Teleoperated versus open needle driving: Kinematic analysis of experienced surgeons and novice users. 2015 IEEE International Conference on Robotics and Automation (ICRA); 2015 26–30 May 2015.
49. Ershad M, Koesters Z, Rege R, Majewicz A. Meaningful Assessment of Surgical Expertise: Semantic Labeling with Data and Crowds. 2016; 9900:508–15. https://doi.org/10.1007/978-3-319-46720-7_59
50. Alway SE, Macgregor LJ, Hunter AM. High-threshold motor unit firing reflects force recovery following a bout of damaging eccentric exercise. Plos One. 2018; 13(4):e0195051. <https://doi.org/10.1371/journal.pone.0195051> PMID: 29630622
51. Nowakowski MM, Trybek P, Rubinkiewicz M, Cegielný T, Romaniszyn M, Pędziwiatr M, et al. Upper extremity surface electromyography signal changes after laparoscopic training. Videosurgery and Other Miniinvasive Techniques. 2018; 13(4):485–93. <https://doi.org/10.5114/wiitm.2018.78744> PMID: 30524619
52. Pincus SM. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences. 1991; 88(6):2297–301. <https://doi.org/10.1073/pnas.88.6.2297> PMID: 11607165