Journal of Medical Robotics Research, Vol. 0, No. 0 (2013) 1–11

Determining the Significant Kinematic Features for Characterizing greenStress during Surgical Tasks Using Spatial Attention

Yi Zheng[a], Grey Leonard[b], Herbert Zeh[b], Ann Majewicz Fey[a,b]

[a] *Department of Mechanical Engineering, the University of Texas at Austin, Address, Austin, TX, USA*
E-mail: yi.zheng@austin.utexas.edu

[b] *Department of Surgery, the University of Texas Southwestern Medical Center, Address, Dallas, TX, USA*

It has been shown that intraoperative stress can have a negative effect on surgeon surgical skills during laparoscopic procedures. Stressful conditions can lead to significantly higher velocity, acceleration, and jerk of the surgical instrument tips, resulting in faster but less smooth movements. However, it is still not clear which of these kinematic features (velocity, acceleration, or jerk) is the best marker for identifying the normal and stressed conditions. Therefore, in order to find the most significant kinematic feature that is affected by intraoperative stress, we implemented a spatial attention-based Long-Short-Term-Memory (LSTM) classifier. In a prior IRB approved experiment, we collected data from medical students performing an extended peg transfer task who were randomized into a control group and a group performing the task under external psychological stresses. In our prior work, we obtained "representative" normal or stressed movements from this dataset using kinematic data as the input. In this study, a spatial attention mechanism is used to describe the contribution of each kinematic feature to the classification of normal/stressed movements. We tested our classifier under Leave-One-User-Out (LOUO) cross-validation, and the classifier reached an overall accuracy of 77.11% for classifying "representative" normal and stressed movements using kinematic features as the input. More importantly, we also studied the spatial attention extracted from the proposed classifier. Velocity and acceleration on both sides had significantly higher attention for classifying a normal movement ($p <= 0.0001$); Velocity ($p <= 0.015$) and jerk ($p <= 0.001$) on non-dominant hand side had significant higher attention for classifying a stressed movement, and it is worthy noting that the attention of jerk on non-dominant hand side had the largest increment when moving from describing normal movements to stressed movements ($p = 0.0000$). In general, we found that the jerk on non-dominant hand side can be used for characterizing the stressed movements more effectively.

Keywords: Laparoscopic surgery; Surgical stress; Deep learning; Motion analysis.

1. Introduction

Excessive intraoperative stress can have a negative effect on surgeon technical skills and therefore compromise patient safety [1–4]. Laparoscopic surgery, in particular, represents a very complex motor control learning task [5], and it has been shown that external stressors can adversely affect motor performance [6]. Detecting the presence of operative stress and its potential detrimental effect on motor performance is an important problem for the surgical training community. Conventional methods for measuring human stress have included physiological sensing techniques such as measuring cortisol levels, heart rate, heart rate variability, and skin conductance levels [7–11]. In practice, physiological sensing techniques can be invasive, time consuming, and may require surgeons to wear sensors on their bodies that could interfere with the technical performance.

Alternatively, kinematic data promises to be a less invasive measurement technique than physiological sensing techniques as this data can be measured directly from robotic encoders in the case of robotic surgery, or through the use of computer vision [12] or other simple sensors [13]. Kinematic data has also been shown to be a powerful tool in other types of surgical skill evaluation [14–16]. For example, Wang et al. implemented a convolutional neural network and used kinematic data as input for real-time surgical skill assessment [17]. In our recent studies, we have validated the feasibility of using kinematic features of the laparoscopic instrument tips (velocity, acceleration and jerk) to distinguish between stressed and non-stressed (normal) conditions during laparoscopic training tasks using statistical analysis. The results indicated that the subjects had significantly higher velocity, acceleration, and jerk in both non-dominant and dominant hand sides when they were under stressed conditions [13, 18]. However, it is not clear which kinematic features can best characterize stressed conditions. In other words, our goal in this study is to find the kinematic feature which is most affected by external stressors as this data stream could hold the most promise for real-time stress detection and mitigation measures.

Deep learning algorithms, such as the attention mechanism with Recurrent Neural Networks (RNN) and, in particular, Long-Short-Term-Memory (LSTM) models [19] could help identify the best metrics for stress identification. LSTM can overcome the limitations in traditional RNNs, for example, traditional RNNs have a problem of vanishing gradients and thus are not able to capture long-term dependencies [20]. The attention mechanism in LSTM can select more critical information from numerous input features [21]. Recently, the attention mechanism has been widely used in variety of sequence modeling projects, such as machine translation [22, 23] and sentiment classification [24]. Qin et al. introduced a dual-stage attention-based Long-Short-Term-Memory (LSTM) model for time-series forecasting [25]. According to this study, the first stage was an input attention mechanism, or spatial attention mechanism, to adaptively extract relevant input features at each time step. Li et al. implemented a novel RNN-based spatial attention model for human manipulation skill assessment from video input. The attention in videos helped them focus on critically important video regions for better skill assessment [26]. In the field of robotic-assisted surgery, Qin et al. implemented a dual-stage attention-based LSTM model for predicting surgical movements and surgical states [27]. As inspired by these studies using attention mechanism on input features, we chose to implement a spatial attention-based LSTM classifier to extract the most important kinematic features for characterizing either a normal or a stressed movement.

With the recent development of robotics-assisted surgical platforms, the kinematic data can be streamed directly from encoders on robot joints without any additional sensors. More importantly, the actuated surgeon side end-effectors could be used for advanced control techniques to provide the surgeon with stress coping strategies in the form of force feedback applied by the surgeon side end-effectors to the surgeon's hands while the surgeon side end-effectors are controlled by surgeons to teleoperate the patient side end-effectors [28], for example, slowing down or pausing [29]. Once we are able to find the kinematic feature which can describe the stressed movements most significantly, the slowing down haptic strategies for coping with external stress can be designed based on this significant kinematic feature.

2. Background and Previous Work

We have raised a question before the study: "What characterizes a stressful movement and how do we detect it?" In order to answer this question, we conducted an experiment in which subjects were provided with commonly experienced intraoperative stressors while performing surgical training tasks in a randomized fashion [13, 18]. Then we studied the negative effect of stressors as well as implementing a deep learning algorithm to extract and detect the stressed movements. The details of this experiment will be summarized in Section 2.1.

2.1. *Identifying Stress*

2.1.1. *Experimental Design*

In this experiment, 30 medical students (29 were right-handed and 1 was left-handed) at the University of Texas Southwestern Medical Center were recruited. The study was IRB approved and informed consent was obtained (approved by UTD IRB office (UTD # 14-57) and UTSW IRB offices (STU #032015-053)). Each subject completed a 10-minute tutorial on the FLS peg transfer task to be familiarized with the instruments and the requirements of the experiment. Then the subjects were randomly divided into a control (n = 15) group or stressed (n = 15) group. greenThe random number sequence for control/stressed group assignment was generated using the random number generator in R programming language. The subject recruiter and the person who analyzed the data were separate. Therefore, it prevented the individuals analyzing the data from knowing which group a subject was assigned to in advance.

During the experiment, each subject was required to finish a 6-minute peg transfer task on the FLS trainer which was placed in the abdominal section of a medical manikin. A pair of electromagnetic (EM) trackers were mounted to the handles of the laparoscopic instruments to capture the time-series data of movements. redWe used the $trakSTAR^{TM}$ electromagnetic 6 DoF tracking system from Ascension Technology Corporation. The data collected by the EM trackers included x_h, y_h, z_h positional coordinates in space and quaternions q_0, q_1, q_2, q_3 at a frequency of 256 Hz. The instrument tip positions were calculated using the handle positions (x_h, y_h, z_h), a rigid body transformation obtained by handle rotations (q_0, q_1, q_2, q_3) and the instrument geometry.

The stressors in this study included the vital signs of the medical manikin and the moderator's feedback during the task. In the control group, each subject proceeded while hearing normal vital signs and with no feedback from the moderator. In the stressed group, each subject performed the task under a period of progressively deteriorating vital signs, with an distinct increase in intensity beginning at the 3-minute mark (the middle point of the 6-minute task). The moderator provided feedback to the stressed group and the feedback culminated in 30 seconds of cardiac arrest and the expiration of the medical manikin.

Besides the kinematic data from EM trackers, other data was collected and evaluated through video review, such as the number of pegs transferred and the number of errors made. Additionally, a blinded independent reviewer with training in OSATS scoring graded each subject using a modified OSATS (mOSATS) rubric [30]. red-mOSATS is a subsection of OSATS including respect for tissue (RFT), time and motion (TM), instrument handling (IH) and the total score (TOT). The subjects also completed a State-Trait-Anxiety-Inventory (STAI) to measure subjective stress after the experiment [31].

2.1.2. *Previous Results*

We removed the data of one subject (in the control group, right-handed) due to redthe loss of connections between sensors and computer during experiment. We down-sampled the kinematic data to 5Hz redto remove noise and smooth the data and organized the data of both instrument tips based on each subject's handedness, so the overall dataset of 29 subjects resulted in approximately 52,200 samples of six features x_{ND}, y_{ND}, z_{ND}, x_D, y_D, z_D (the subscript D indicates data from the dominant hand and ND is non-dominant hand) [32]. redAfter down-sampling the data, the kinematic metrics velocity (V), acceleration (A) and jerk (J) of the instrument tips were also calculated: red

$$V_t = \frac{\sqrt{(P_{t+1} - P_t)^T (P_{t+1} - P_t)}}{T_{t+1} - T_t} \tag{1}$$

redP_t is the 3-D position at time t, and T_t is the time stamp at time t.

redThe Acceleration (A) and the jerk (J) were time series data calculated in the similar way: red

$$A_t = \frac{V_{t+1} - V_t}{T_{t+1} - T_t}, J_t = \frac{A_{t+1} - A_t}{T_{t+1} - T_t} \tag{2}$$

redThe stressed group had significantly higher velocity, acceleration, and jerk than the control group for both hands. redIn stressed group, the second 3-minute half of the experiment had significantly higher velocity, acceleration, jerk, path length and lower economy of volume than the first 3-minute half for both hands. Other standard metrics were also analyzed, for example, the stressed group had smaller numbers of peg transferred and larger numbers of errors made, indicating worse performance under stressful conditions red [13]. Lower mOSATS scores and higher scores for the change from baseline (trait) to during the scenario (state) in STAI were found to be significant in the stressed group (Fig 1). These significance differences between control and stressed groups in our studies indicated that the kinematic data can be related to increased stress levels.

We also extracted the movements that were more significantly affected by the stress using an temporal attention-based LSTM classifier in another study [33]. We first implemented a trial-wise classifier with the attention mechanism which took the time-series instrument tips positional data (x_{ND}, y_{ND}, z_{ND}, x_D, y_D, z_D) of each trial as the input, and returned $y = 0$:control (normal) or $y = 1$: *stressed* as the output. The classifier returned the temporal attention for each trial, which was a vector containing the importance of each time step within a trial that contributed to classification of control or stressed trial. After obtaining the temporal attention vector of each trial, we used a sliding window to organize the temporal attention sequence and the input sequence into frames. We calculated the sum of each attention frame and considered any frame with an attention greater than the third quartile to be "important": a frame with an attention sum greater than the third quartile in a control (normal) trial was considered to be a "representative" normal movement; A frame with an attention sum greater than the third quartile in a stressed trial was considered to be a "representative" stressed movement.

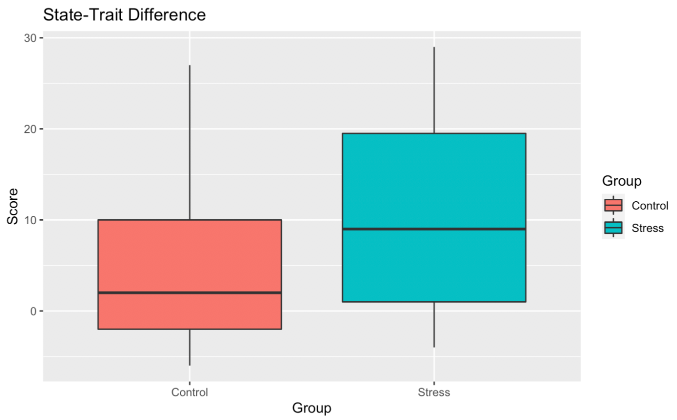


Fig. 1: The change from trait to state. Higher scores in the stressed group.

Finally, a subset of the original dataset containing the "representative" normal and stressed movements could be extracted based on the temporal attention.

2.2. *Goals of This Study*

With the first question mentioned in the first paragraph of Section 2.1 answered by our studies in Section 2.1.2, we decided to move forward to finding the answers of the second question:"Which type of haptic cues on telerobotic platform could improve the stressful movements significantly?"

Even through the results in Section 2.1.2 indicate that stress leads to significantly higher velocity, acceleration and jerk, it is still not clear that which kinematic feature has more contribution to identifying the stressed movements and therefore, should be improved using haptic cues.

3. Methods

~~Based on the results in Section 2.1.2, we aim to find the kinematic feature that was mostly affected by the stress. In the studies described in Section 2.1.2, we first found~~

~~that the velocity, acceleration and jerk of the instrument movement were higher when the subject experienced intra-operative stress, but it is not clear which feature has the best potential to characterize the stress. Then, we implemented a temporal attention-based LSTM classifier and used position data of the instrument tips as the input to distinguish between control (normal) and stressed trials. We also obtained the "representative" control (normal) and "representative" stressed surgical movements from each trial based on the attention mechanism since the attention could tell us the importance of each time step to the final classification of control/stressed.~~

In this study, the kinematic features (velocity, acceleration and jerk) of the obtained "representative" movements were used as the input of our newly proposed spatial attention-based LSTM classifier.

The classifier returns: first, whether a movement is a normal or a stressed movement; second, the spatial attention vector that describes the importance of each input feature (velocity, acceleration and jerk) that contributes to the classification of a normal/stressed movement. Instead of capturing the importance of each time step, namely temporal attention as described in previous sections, spatial-attention calculates the importance of each input feature at each time step for classification [25].

3.1. *Model Architecture*

The architecture of proposed spatial attention-based LSTM classifier is illustrated in Fig 2. The input sequence $\{x_1, x_2, ..., x_T\}$ was the kinematic features of each "representative" movement. As mentioned above, each x contained six kinematic features extracted from both instrument tips, velocity, acceleration and jerk, respectively $(V_{ND}, A_{ND}, J_{ND}, V_D, A_D, J_D)$. For each input:

$$x_j = [V_{NDj}, A_{NDj}, J_{NDj}, V_{Dj}, A_{Dj}, J_{Dj}]^T, j = 1...T \quad (3)$$

The subscript D is dominant hand side and ND is non-dominant hand side. The ground truth label $y = 0 or 1$ was assigned to be either a "representative" control (normal) movement or a "representative" stressed movement.

We measured the importance of each input feature by computing a $tanh$ function of input $\mathbf{x}$ with $units = 6$:

$$e_{ij} = tanh(x_j) = tanh(x_{1j}, x_{2j}, ..., x_{3j}), i = 1, ..., 6 \quad (4)$$

e_{ij} was called "energy" which calculated the contribution of each feature at each time step j to the final classification of control (normal) or stressed movement.

Then, the spatial attention weights β_{ij} at each time step j were obtained by passing e_{ij} to a $Softmax$ function to ensure all spatial attention weight as each time step sum to 1:

$$\beta_{ij} = \frac{exp(e_{ij})}{\sum_{i=1}^{n} exp(e_{ij})} \quad (5)$$

The spatial attention weight β_{ij} indicates how much attention the final output label y should pay to the i^{th} input feature at time step j.

Next, we calculated the context vector $(c_1, c_2, ..., c_T)$ as a weighted linear combination of all input features at each time step x_i:

$$c_i = \sum_{i=1}^{6} \beta_{ij} x_{ij} \quad (6)$$

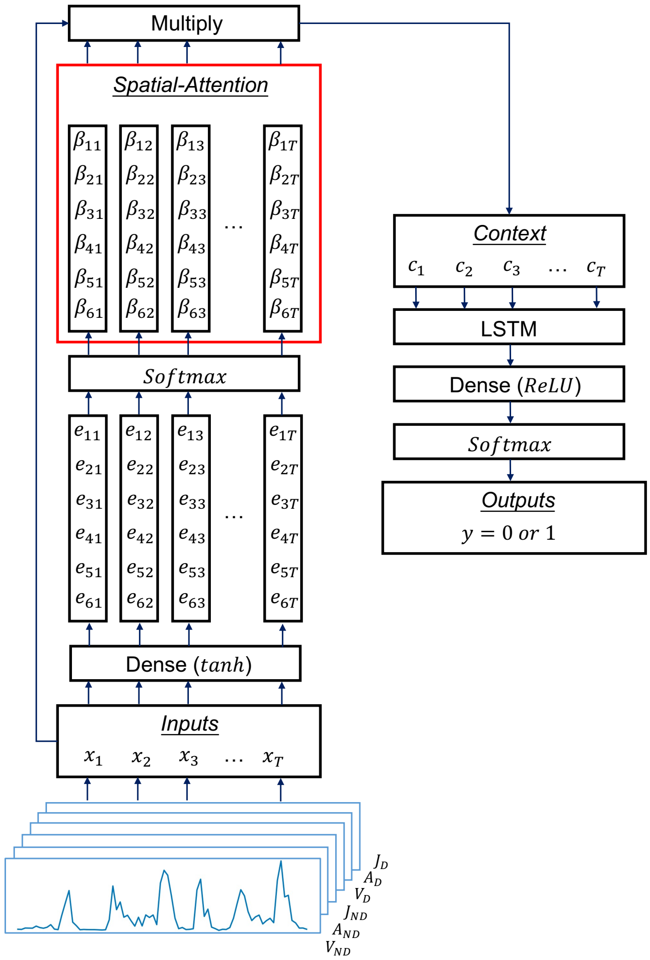


Fig. 2: Model architecture of proposed spatial attention-based LSTM classifier for "representative" movements classification and the extraction of input feature importance. The input had 6 features including time-series velocity, acceleration and jerk of both instrument tips.

Finally, we passed the context vector to an LSTM ($units = 100$). The final output of LSTM was sent to two fully-connected layers with activation functions of

$ReLU, units = 20$ and $Softmax, units = 2$ to output the prediction $\hat{y}$. redThe hyperparameters in the model were selected through grid search. During grid search, we shuffled and split the data into training set and testing set using 70/30 split. Then, we chose the set of hyperparameters which showed the best accuracy in grid search for classifier design.

greenDifferent from grid search, we adopted Leave-One-User-Out cross-validation to evaluate the performance of our proposed model which will be described in Section 3.2.

We obtained two outputs from the proposed classifier. First, reda classification result of the input movement deciding whether the input movement was normal or stressed; Second, the spatial attention vector that could tell the importance of each input feature to classify the movements as normal or stressed movements.

3.2. *Cross-Validation*

~~It is standard to test the deep learning model by leaving aside a portion of the data as the testing dataset and use the remaining portion to train the model.~~

To evaluate the performance of our proposed classifier, we adopted Leave-One-User-Out (LOUO) cross-validation. The LOUO used the i^{th} subject as testing dataset and the rest for training, and iterated throughout all the 29 subjects. The mean values of all 29 iterations' performance metrics were reported and will be shown in the following sections. LOUO was designed to test if the classifiers were generalized enough for unseen data.

3.3. *Model Performance Metrics*

To evaluate the performance of our proposed classifier, four commonly used metrics were used in our study - Accuracy, Precision, Recall, and F1-score. Accuracy is the ratio of correct predictions $(T_p + T_n)$ to the total predictions $(T_p + F_p + T_n + F_n)$; Precision is the ratio of correct positive predictions (T_p) to the total positive results $(T_p + F_p)$ predicted by the classifier; Recall is the ratio of correct positive predictions (T_p) to the total actual results $(T_p + F_n)$. F1-score is a measure of a classifier's accuracy which takes the harmonic mean of the precision and recall.

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + T_n + F_n}, \tag{7}$$

$$Precision = \frac{T_p}{T_p + F_p}, \tag{8}$$

$$Recall = \frac{T_p}{T_p + F_n}, \tag{9}$$

$$F1 - score = \frac{2(Recall * Precision)}{Recall + Precision}. \tag{10}$$

4. Results and Discussion

To investigate which kinematic features can potentially characterize either a normal movement or a stressed movement, we used the kinematic features of the "representative" movements as the input of our proposed spatial attention-based LSTM model.

The dataset we used in this study was from our previous experiment which has been discussed in Section 2.1. In order to validate our approach, first, the performance of our proposed spatial attention-based LSTM classifier was evaluated. Second, the spatial attention of all six kinematic features were obtained from the proposed classifier. Data analysis was carried out to determine the significant differences among the spatial attention of all six kinematic features.

4.1. *Classifier Performance*

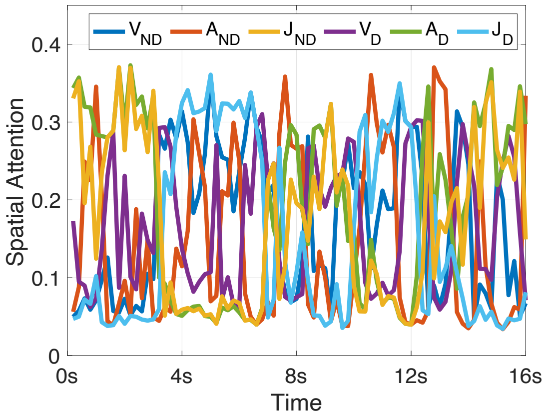
In this study, we aimed to classify between the "representative" normal/stressed movements using the kinematic features (velocity, acceleration and jerk). The input of our classifier was the kinematic features of each "representative" normal or stressed movement. And the output returned if the input was a normal movement $(y = 0)$ or a stressed movement $(y = 1)$ as well as the spatial attention vector describing the contribution of each kinematic feature.

Based on the previous study, the input movement length was 16 secondsred [33]. Under LOUO cross-validation, we used the movements of the i^th subject as the testing dataset and the remaining for training the model. And the same process iterated throughout all 29 subjects red(11 movements from each subject). Accuracy was obtained through averaging throughout LOUO cross-validation red(Mean: 77.11%, Standard Deviation: 17.32%). redSince we were using LOUO and each User's movement could only be either normal or stressed, it was not appropriate to calculate Precision, Recall and F1-score individually. Instead, we added the confusion matrix of all LOUO iterations and used the summed confusion matrix for calculations (Precision: 77.26%, Recall: 77.23%, F1-score: 77.24%).

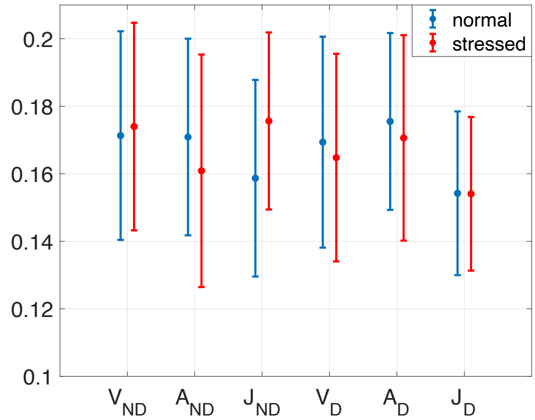
4.2. *Spatial Attention of Kinematic Features*

The classifier also returned the spatial attention vector β of each input feature that contributed to the classification. In other word, the spatial attention tells us which kinematic features had the most potential to characterize either a normal or a stressed movement.

As described in Section 3.1, the classifier returns a vector of attention at each time step j for a given input movement $([\beta_{1j}, \beta_{2j}, \beta_{3j}, \beta_{4j}, \beta_{5j}, \beta_{6j}]^T)$. In order to compare the attention among different kinematic features, we then took the average of the spatial attention across all time steps for each input movement. As a result, the averaged spatial attention of each kinematic feature across all time steps was



(a) An example showing the spatial attention of all kinematic features in a randomly selected stressed movement. In statistical analysis, the spatial attention was averaged across all time step, therefore resulting in a vector of six numbers for each movement.



(b) The means and standard deviations of averaged spatial attention across all normal and stressed movements were plotted.

Fig. 3: Comparing the spatial attention of different kinematic features in normal and stressed movements.

used in statistical analysis to determine significant differences in the six kinematic features in normal and stressed movements. The normality test to identify a normal distribution in the averaged spatial attention was rejected and thus, the Kruskal Wallis test was used to identify the significance.

The spatial attention of each kinematic feature to describe a normal movement is shown in the blue lines in Fig. 3b. The results of statistical analysis to determine the differences among the six features are summarized in Table. 1. As shown in Fig. 3b and Table. 1, the velocity and acceleration for both non-dominant (V_{ND}, A_{ND}) and dominant (V_D, A_D) hand sides had significantly higher attention than the jerk (J_{ND} and J_D) in normal movements. It means that the velocity and acceleration have more potential to describe a normal movement.

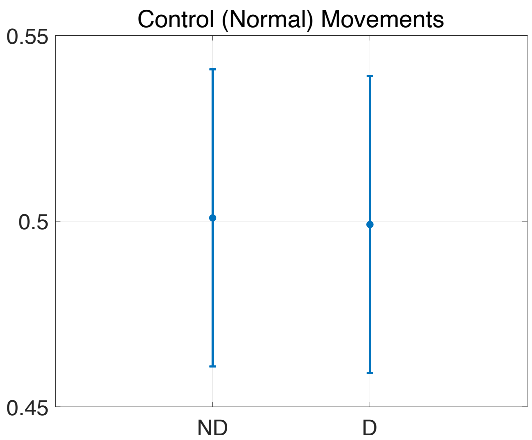
However, in stressed movements, as shown in the red lines in Fig. 3b and Table. 2, velocity and jerk on non-dominant hand side (V_{ND}, J_{ND}) had significantly higher attention than acceleration on non-dominant hand side (A_{ND}), velocity and jerk on the dominant hand side (V_D and J_D). Besides, according to Table. 2, acceleration on the dominant hand side (A_D) also showed significantly higher attention than jerk on the dominant hand side (J_D). The results indicate that velocity and jerk on non-dominant hand side and acceleration on the dominant hand side have a better potential to describe a stressed movement.

redWhen comparing the kinematic feature attentions between normal movements and stressed movements in Fig. 3b, we noticed that the attention values of V_{ND}, V_D and J_D did not show a clear difference between normal and stressed movements. However, the attention value of J_{ND} had an clear increment from describing normal movements to describing stressed movements. It means that J_{ND} received a higher attention from the classifier when describing a stressed movement. Similarly, the attention values of A_{ND} and A_D had an clear increment from describing stressed movements to describing the normal movements. It means that A_{ND} and A_D received a higher attention when describing a normal movement. Then, we used a Wilcoxon rank sum test to compare each kinematic feature between normal and control movements in Table. 3. Therefore, we can say that J_{ND} was mostly affected by the stress, and it can be used to characterize the stress more effectively; A_{ND} and A_D can be used to characterize the normal movements.

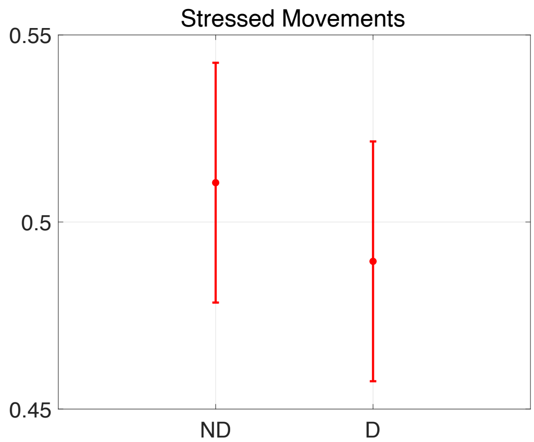
4.3. *Spatial Attention of non-dominant and Dominant Hand Sides*

We also examined the importance of hand sides to characterize the normal or stressed movement. Instead of analyzing the spatial attention of each kinematic feature separately, we took the sum of the spatial attention of kinematic features of both non-dominant hand side and dominant hand side.

As shown in Fig. 4a and the last row of Table. 1, no significant difference between non-dominant hand side and dominant hand side can be found. It means the movement on both sides has equal importance for describing a normal condition. This finding is easy to be explained since in normal movements, the subjects were performing under normal conditions and the movements on both sides were



(a) Comparing the sum of attention between non-dominant hand side and dominant hand side in normal movements.



(b) Comparing the sum of attention between non-dominant hand side and dominant hand side in normal movements.

Fig. 4: Comparing the sum of spatial attention between non-dominant hand side and dominant hand side for characterizing a normal and stressed movement. In stressed movements, the non-dominant hand size received more attention than dominant hand side.

not affected by the intraoperative stressors, so there is no difference between the two hands.

However, in stressed movements, where the subjects' performance was negatively affected by the stressors, the importance of both sides to characterize the stressed movements has changed. As shown in Fig. 4b and the last row of Table. 2, the non-dominant hand side showed significantly higher attention than the dominant hand side which means the kinematic features on the non-dominant hand side have more potential to characterize the stressed movements. The reason behind this finding is that the movement on the non-dominant hand side is less skilled and less dexterous. redInterestingly, recent work from our lab has also shown that when two hands are moving simultaneously, the non-dominant hand actually suffers in performance relative to if it was moving alone [34]. We think these results could indicate that because the non-dominant hand is arguably the weaker of the two hands, studying its movements is useful as it is more prone to performance degradations in challenging conditions. Therefore, the movement on the non-dominant hand side is more likely to be negatively affected by the intraoperative stressors and it is reflected as a higher attention on non-dominant hand side during classification of stressed movements.

Table 1: Normal movements: statistical analysis summary of the spatial attention of six kinematic features.

	Significance	$p - value$
Kinematic Features	$V_{ND} > J_{ND}, J_D$	$p < 0.0001$
	$A_{ND} > J_{ND}, J_D$	$p = 0.0001$
	$V_D > J_{ND}, J_D$	$p = 0.0001$
	$A_D > J_{ND}, J_D$	$p < 0.0001$
ND vs. D	N/A	$p = 0.8198$

Table 2: Stressed movements: statistical analysis summary of the spatial attention of six kinematic features.

	Significance	$p - value$
Kinematic Features	$V_{ND} > A_{ND}, V_D, J_D$	$p < 0.015$
	$J_{ND} > A_{ND}, V_D, J_D$	$p < 0.001$
	$A_D > J_D$	$p = 0.0002$
ND vs. D	$ND > D$	$p < 0.0001$

red

Table 3: Comparisons of the spatial attention of kinematic features between normal and stressed movements.

Kinematic Features	Significance	$p-value$
V_{ND}	N/A	$p=0.9821$
A_{ND}	Normal > Stressed	$p=0.023$
J_{ND}	Normal < Stressed	$p=0.0000$
V_D	N/A	$p=0.0724$
A_D	Normal > Stressed	$p=0.0002$
J_D	N/A	$p=0.8388$

5. Conclusion

In this study, we implemented a spatial attention-based LSTM model and used kinematic features (velocity, acceleration and jerk) as input for the classification of "representative" normal and stressed movements which were obtained from our previous studies [13, 33].

Our proposed classifier was able to distinguish between "representative" normal and stressed movement with an accuracy of 77.11% under LOUO cross-validation, and it showed that our classifier was generalized to unseen data. More importantly, the classifier also returned the spatial attention vector which was able to tell us the contribution of each kinematic feature to the final classification labels.

We also conducted statistical analysis to study the obtained spatial attention of six kinematic features. In normal movements, velocity and acceleration on both non-dominant and dominant hand sides had significantly higher attention than jerk. It means that velocity and acceleration contributed more to the classification of a normal movement, and therefore, can be used for characterizing a normal movement.

In stressed movements, velocity and jerk on the non-dominant hand side had significantly higher attention than acceleration on non-dominant hand side, velocity and jerk on dominant hand side. Although it is not significant, the jerk also had higher attention than velocity on non-dominant hand side.

redWhen comparing the kinematic feature attentions between normal and stressed movements in Fig. 3b and Table. 3, we noticed that the attention of the jerk on non-dominant hand side had the significant change when moving from normal movement to stressed movement. It means that jerk on the non-dominant hand side was the most significant kinematic feature to be affected by stress, therefore, had the best potential for characterizing a stressed movement. Similarly, in normal movements, the acceleration on both hand sides also had significantly higher spatial attentions than stressed movements, which means the accelerations had the best potential for characterizing a normal movement.

We also conducted analysis based on non-dominant and dominant hand sides. In normal movements, the spatial attention sums on both sides did not show any significant differences. However, in stressed movements, non-dominant hand side had significantly higher spatial attention than dominant hand side which means the kinematic features on non-dominant hand side had better potential to describe a stressed movement and the performance of non-dominant hand is more likely to be negatively affected by intraoperative stress.

redOne limitation of this study is the lack of expertise levels. We only had medical students recruited and only one trial (control or stressed) for each subject. A better generalization of this deep learning approach can be made if subjects could include a wider range of expertise levels, for example, attending, fellow and resident surgeons in a large number, therefore, reduce the probability of overfitting the model.

In general, in this paper, we answered the question raised in Section 2.2: "Which type of haptic cues on telerobotic platform could improve the stressful movements significantly?". redBased on the results, the jerk on non-dominant hand and the accelerations on both hand sides are most likely to be affected by stress. And according to our previous study, the stress led to significant greater values of jerk meaning less smooth movements under stressed conditions. These findings can be integrated to create haptic cues based on jerk, especially on non-dominant hand side on telerobotic platforms to help surgeons cope with intraoperative stress and therefore, mitigate the negative effect of stress. In future work, we will need to determine how to develop an effective haptic feedback cue that can mitigate changes in movement jerk. This is not a trivial problem as jerk-based measurements are prone to noise and it is not clear how to provide jerk-based haptic feedback in a stable way.

Acknowledgments

The authors would like to thank the UTSW Training Resident Doctors As Innovators in Science (TARDIS) program, supporting co-Author Grey Leonard. We also thank the UTSW Simulation Center director and staff for their invaluable assistance with data collection. This work was supported in part by NSF grant number #1846726 and NIH grant number #1R01EB030125.

Disclosure

The authors declared that they have no conflict of interest.

References

[1] S. Arora, N. Sevdalis, R. Aggarwal, P. Sirimanna, A. Darzi and R. Kneebone, Stress impairs psychomotor performance in novice laparoscopic surgeons, Surgical Endoscopy **24**(10) (2010) 2588–2593.

[2] S. Arora, N. Sevdalis, D. Nestel, T. Tierney, M. Woloshynowych and R. Kneebone, Managing intraoperative stress: what do surgeons want from a crisis training program?, American Journal of Surgery **197** (apr 2009) 537–543.

[3] K. Moorthy, Y. Munz, A. Dosis, S. Bann and A. Darzi, The effect of stress-inducing conditions on the performance of a laparoscopic task, Surgical Endoscopy and Other Interventional Techniques **17** (sep 2003) 1481–1484.

[4] K. H. Goodell, C. G. Cao and S. D. Schwaitzberg, Effects of cognitive distraction on performance of laparoscopic surgical tasks, Journal of Laparoendoscopic and Advanced Surgical Techniques **16** (apr 2006) 94–98.

[5] E. N. Spruit, G. P. Band, J. F. Hamming and K. R. Ridderinkhof, Optimal training design for procedural motor skills: a review and application to laparoscopic surgery, Psychological research **78**(6) (2014) 878–891.

[6] E. D. Ryan, Effects of stress on motor performance and learning, Research Quarterly. American Association for Health, Physical Education and Recreation **33**(1) (1962) 111–119.

[7] B. J. van Holland, M. H. Frings-Dresen and J. K. Sluiter, Measuring short-term and long-term physiological stress effects by cortisol reactivity in saliva and hair, International archives of occupational and environmental health **85**(8) (2012) 849–852.

[8] B. Böhm, N. Rötting, W. Schwenk, S. Grebe and U. Mansmann, A prospective randomized trial on heart rate variability of the surgical team during laparoscopic and conventional sigmoid resection, Archives of Surgery **136**(3) (2001) 305–310.

[9] E. Czyzewska, K. Kiczka, A. Czarnecki and P. Pokinko, The surgeon's mental load during decision making at various stages of operations, European Journal of Applied Physiology and Occupational Physiology **51**(3) (1983) 441–446.

[10] A. P. Tendulkar, G. P. Victorino, T. J. Chong, M. K. Bullard, T. H. Liu and A. H. Harken, Quantification of surgical resident stress "on call", Journal of the American College of Surgeons **201** (oct 2005) 560–564.

[11] W. Boucsein, Electrodermal activity (Springer Science & Business Media, 2012).

[12] F. Chadebecq, F. Vasconcelos, E. Mazomenos and D. Stoyanov, Computer vision in the surgical operating room, Visceral Medicine **36**(6) (2020) 456–462.

[13] Y. Zheng, G. Leonard, J. Tellez, H. Zeh and A. M. Fey, Identifying kinematic markers associated with intraoperative stress during surgical training tasks, 2021 International Symposium on Medical Robotics (ISMR), IEEE (2021), pp. 1–7.

[14] M. J. Fard, S. Ameri, R. Darin Ellis, R. B. Chinnam, A. K. Pandya and M. D. Klein, Automated robot-assisted surgical skill evaluation: Predictive analytics approach, The International Journal of Medical Robotics and Computer Assisted Surgery **14** (feb 2018) p. e1850.

[15] S. S. Vedula, A. Malpani, N. Ahmidi, S. Khudanpur, G. Hager and C. C. G. Chen, Task-Level vs. Segment-Level Quantitative Metrics for Surgical Skill Assessment, Journal of Surgical Education **73** (may 2016) 482–489.

[16] M. Ershad, R. Rege and M. Fey, A., Meaningful assessment of robotic surgical style using the wisdom of crowds, International Journal of Computer Assisted Radiology and Surgery **13** (2018) 1037–1048.

[17] Z. Wang and A. M. Fey, Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery, International journal of computer assisted radiology and surgery **13**(12) (2018) 1959–1970.

[18] G. Leonard, J. Cao, S. Scielzo, Y. Zheng, J. Tellez, H. J. Zeh and A. M. Fey, The Effect of Stress and Conscientiousness on Simulated Surgical Performance in Unbalanced Groups: A Bayesian Hierarchical Model, Journal of the American College of Surgeons **231** (oct 2020) p. S258.

[19] S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, Neural Computation **9** (nov 1997) 1735–1780.

[20] Y. Bengio, P. Simard and P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE transactions on neural networks **5**(2) (1994) 157–166.

[21] Y. Ding, Y. Zhu, J. Feng, P. Zhang and Z. Cheng, Interpretable spatio-temporal attention LSTM model for flood forecasting, Neurocomputing **403** (aug 2020) 348–359.

[22] D. Bahdanau, K. H. Cho and Y. Bengio, Neural machine translation by jointly learning to align and translate, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, (International Conference on Learning Representations, ICLR, sep 2015).

[23] B. Zhang, D. Xiong and J. Su, Neural Machine Translation with Deep Attention, IEEE Transactions on Pattern Analysis and Machine Intelligence **42** (jan 2020) 154–163.

[24] D. Ma, S. Li, X. Zhang and H. Wang, Interactive attention networks for aspect-level sentiment classification, Proceedings of the 26th International Joint Conference on Artificial Intelligence, (2017), pp. 4068–4074.

[25] Y. Qin, D. Song, H. Cheng, W. Cheng, G. Jiang and G. W. Cottrell, A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction, IJCAI International Joint Conference on Artificial Intelligence **0** (apr 2017) 2627–2633.

[26] Z. Li, Y. Huang, M. Cai and Y. Sato, Manipulation-skill assessment from videos with spatial attention network, 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), IEEE (2019), pp. 4385–4395.

[27] Y. Qin, S. Feyzabadi, M. Allan, J. W. Burdick and M. Azizian, davincinet: Joint prediction of motion and surgical state in robot-assisted surgery, 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE pp. 2921–2928.

[28] N. Enayati, A. M. Okamura, A. Mariani, E. Pellegrini, M. M. Coad, G. Ferrigno and E. De Momi, Robotic Assistance-as-Needed for Enhanced Visuomo-

tor Learning in Surgical Robotics Training: An Experimental Study (Institute of Electrical and Electronics Engineers Inc., sep 2018), pp. 6631–6636.

[29] N. E. Anton, P. N. Montero, L. D. Howley, C. Brown and D. Stefanidis, What stress coping strategies are surgeons relying upon during surgery?, American Journal of Surgery, **210**, (Elsevier Inc., nov 2015), pp. 846–851.

[30] J. A. Martin, G. Regehr, R. Reznich, H. Macrae, J. Murnaghan, C. Hutchison and M. Brown, Objective structured assessment of technical skill (OSATS) for surgical residents, British Journal of Surgery **84** (feb 1997) 273–278.

[31] C. Sielberger, R. Gorsuch, P. Vagg and G. Jacobs, Manual for the state-trait anxiety inventory (form y) (1983).

[32] T. N. Judkins, D. Oleynikov and N. Stergiou, Objective evaluation of expert and novice performance during robotic surgical training tasks, Surgical Endoscopy and Other Interventional Techniques **23** (mar 2009) 590–597.

[33] Y. Zheng, G. Leonard, H. Zeh and A. M. Fey, Framewise detection of surgeon stress levels during laparoscopic training using kinematic data, International Journal of Computer Assisted Radiology and Surgery (2022) 1–10.

[34] J. R. Boehm, N. P. Fey and A. M. Fey, Effects of scaling and sequence on performance of dynamic bimanual path following tasks, Journal of Medical Robotics Research **5**(03n04) (2020) p. 2042001.

Yi Zheng received the B.Eng. degree in Mechanical Engineering from Xi'an Jiaotong University, Shaanxi, China in 2015, and the M.S. degree in Mechanical Engineering from Boston University, Boston, MA in 2017. He is currently pursuing his Ph.D. in Mechanical Engineering at the University of Texas at Austin, Austin, TX.

Grey Leonard received a BS in Biology from the United States Air Force Academy in 2008, an MD from University of South Florida Morsani College of Medicine in 2015 and an MS in Applied Statistics Data Analytics from Southern Methodist University in 2020. He is currently a 4th year resident in general surgery at UT Southwestern Medical Center. He recently completed a 2-year research fellowship in the Training Resident Doctors as Innovators in Science Program, also at UT Southwestern.

Herbert Zeh received his M.D. from the University of Pittsburgh School of Medicine in 1994. Dr. Zeh completed his surgical residency and fellowship training in gastrointestinal surgery and surgical oncology at The Johns Hopkins Hospital. Dr. Zeh returned to Pittsburgh in 2002 and remained on faculty until 2018, where he served as the Chief of the Division of Gastrointestinal Surgical Oncology at the University of Pittsburgh Medical Center (UPMC) Hillman Cancer Center and Co- Director of the UPMC Pancreatic Cancer Center. During his tenure at UPMC, Dr. Zeh and his colleagues accumulated research and published extensively on one of the world's largest experiences with robotic pancreatic surgery. Dr. Zeh was recruited to Dallas in 2018 as Professor and Chair of the Department of Surgery at UT Southwestern Medical Center. He holds the Hall and Mary Lucile Shannon Distinguished Chair in Surgery. Dr. Zeh has conducted numerous investigator-initiated clinical trials examining novel treatments for patients with pancreatic cancer and directs a translational research lab that examines damage-associated molecular pattern molecules in the disease. An innovator and leader in the field of pancreatic diseases and pancreatic cancer, he has been a practicing surgeon for more than 20 years.

Ann Majewicz-Fey received the B.S. degrees in Mechanical Engineering and Electrical Engineering from the University of St. Thomas, St. Paul, MN, in 2008, the M.S. degree in Mechanical Engineering from Johns Hopkins University, Baltimore, MD, in 2010, and the Ph.D. degree from Stanford University, Stanford, CA, in 2014, all in mechanical engineering. She is currently an Associate Professor in the Department of Mechanical Engineering at the University of Texas at Austin, Austin, TX, where she holds a joint appointment in the Department of Surgery at UT Southwestern Medical Center, Dallas, TX. She directs the Human-Enabled Robotic Technology

(HeRo) Laboratory where she is responsible for research projects in the areas of robot-assisted surgery, teleoperation, haptics, and human-centric modeling. Dr. Majewicz Fey received the 2015 National Science Foundation CISE Research Initiation Initiative (CRII) award and the 2019 National Science Foundation CAREER Award.