

# Improving surgical motor skill assessment and acquisition via neuromodulation, neuroimaging, and machine learning

Yuanyuan Gao



## **Context – Modern surgical technique**



- + Minimal incision
- + Reduced pain
- + Shorter recovery time.
- Complex bimanual motor skills and hand-eye coordination
- Variable adverse event rate<sup>1</sup>
- Most errors during the learning phase





## International accredited program



Society of American Gastrointestinal and Endoscopic Surgeons (SAGES)

- Fundamentals of Laparoscopic Surgery (FLS) is a pre-requisite for Board certification to every general and Ob/Gyn surgeon<sup>2</sup>.
- Two components cognitive (high stakes exam) + psychomotor (trainer box)
- FLS trainer box is effective in teaching technical motor skills<sup>2</sup>.
- FLS score = f(completion time, performance error);
- FLS score: 0-300 (higher score -> more skilled)



**Peg Transfer** 



Pattern Cutting



Ligating Loop





Intracorporeal & Extracorporeal Suture

2. H. Peters, G. M. Fried, L. L. Swanstrom, N. J. Soper, L. F. Sillin, B. Schirmer, and K. Hoffman, "Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery," *Surgery*, vol. 135, no. 1, pp. 21–27, Jan. 2004.



## **Motivations**



#### **Collaborator:**



Professor and Chairman -Department of Surgery Jacob School of Medicine and Biomedical Science University at Buffalo Former chair of the FLS committee

## Challenges in learning curve prediction:

- Skill acquisition procedure is analyzed post-hoc.
  - Learning curve factors could not be predicted before the completion of the training.
  - Impede training protocol customization.

#### Challenges in skill assessment:

- FLS score is manually calculated by proctor (www.flsprogram.org).
  - Time consuming- two to three weeks;
  - Labor intensive trained proctor needed.

#### ---> Challenges in surgical training:

- The training protocol relies on repetition.
  - Time consuming;
  - Proficiency not guaranteed.



## Tools





5

# S. A. 1: to predict learning curve features in the early stage of training.

## **Previous work:**

- Graph or table displaying
- T-test, ANOVA to grouped data split by practice number
- Curve fitting
- Cumulative summation (CUSUM).

# **Drawback:**

Post-hoc analysis

# **Goal:** predict learning curve features from beginning **Data:**

Study	Platform	Motor task	No. of learning curves
Nemani et al. 2017	FLS	Pattern	4
		cutting	
Nemani et al. 2017	VBLaST	Pattern	6
		cutting	
Linsk et al. 2017	VBLaST	Pattern	2
		cutting	
Fu et al. 2019	FLS*	Suturing	3
Total			15





6

# S. A. 1: Results

# Hypothesis #1:

• The initial performance of a trainee can predict the number of trials required to achieve proficiency and the final proficiency level.

# **Algorithm:**

- Kernel Partial Least Squares (KPLS)
  - Small sample size
  - High dimensional variables
- Cross-validation: leave-one-out



Prediction	KPLS	Log-linear model
Initial performance   Number of trials to reach proficiency	$R^2 = 0.72$	$R^2 = -109.55$
Initial performance  Final performance level	$R^2 = 0.89$	$R^2 = -3.36$

**Gao, Y.,** Kruger, U., Intes, X., Schwaitzberg, S. and De, S., 2020. A machine learning approach to predict surgical learning curves. *Surgery*, 167(2), pp.321-327.



# Hypothesis #2:

• A single factor can describe the learning curve factors.

# **Algorithm:**

- Kernel Principle Component Analysis (KPCA)
  - Small sample size
  - High dimensional variables
- Cross-validation: leave-one-out





- K-means grouping results
  - Using the learning curve factors



• Using the extracted LI value



• The extracted single factor is enough to classify the learners



Specific aim 1

".. is to predict learning curve features in the early stage of training"

We established that:

- Learning curve factors can be predicted from the initial performance;
- Single factor can represent the learning curve factors.

Impact:

- We are the first to suggest predict learning curve from the early learning stage;
- It is a vital step towards surgical training remediation;
- Understanding of different learning abilities: the importance of training remediation

Gao, Y., Kruger, U., Intes, X., Schwaitzberg, S. and De, S., 2020. A machine learning approach to predict surgical learning curves. *Surgery*, 167(2), pp.321-327.



## **Motivations**





Professor and Chairman -Department of Surgery Jacob School of Medicine and Biomedical Science University at Buffalo Former chair of the FLS committee

## Challenges in learning curve prediction:

- Skill acquisition procedure is analyzed post-hoc.
  - Learning curve factors could not be predicted before the completion of the training.
  - Impede training protocol customization.

#### Challenges in skill assessment:

- FLS score is manually calculated by proctor (www.flsprogram.org).
  - Time consuming- two to three weeks;
  - Labor intensive trained proctor needed.

#### ---- Challenges in surgical training:

- The training protocol relies on repetition.
  - Time consuming;
  - Proficiency not guaranteed.

Specific aim 2 : to predict FLS scores via neuroimaging.



#### **Neural basis of motor learning**

Neuroplasticity



Two steps:

- Rapid reinforcement of preestablished organic pathways;
- Later formation of new pathways





Cortex areas

# Background

- Functional Near infrared spectroscopy (fNIRS)
  - Non-invasive imaging technique
  - Delivers infrared light on the surface of the scalp via source probes.
  - Infrared light scatters through turbid tissue and the backscattered light is detected



- Since attenuated light is related to functional chromophores (such as oxy-HbO2 and deoxy-HbO2), the relative concentration of these chromophores can be determined and finally be correlated with brain activity.
- Why should we use NIRS to measure brain activity?
  - High temporal resolution (~ 100Hz)
  - High depth penetration (~1.5 cm)
  - Allows for complex tasks to be performed





#### **Previous work- fNIRS in motor classification**

- fNIRS data could classify motor tasks.
  - Simple vs. complex motor tasks (Holper and Wolf 2011).
  - Left vs. right hand motion (Fazli et al. 2012, Naseer and Hong 2013).
  - Arm lifting vs. knee extension (Shin and Jeong 2014).
- fNIRS data could classify surgical levels.
  - PFC correlation to surgical skill level (Leff et al. 2008; Ohuchida et al. 2009; James et al. 2011)
  - Surgical skill levels could be classified by fNIRS (Nemani et al. 2018 & 2019).

Question: Can fNIRS data predict FLS score?



## **Experiment setup and data collection**

# **Population:**

- 13 novice medical students;
   FLS task:
  - Pattern cutting task
- Instrumentation:
  - TechEn CW6; continuous wave; 690&830 nm



Pattern Cutting



HOMER2 software suite used for all infrared data processing







# **Ease of use and implementation**

- Restricted ourselves to PFC
- Data types  $\rightarrow$  moments of temporal HbO<sub>2</sub> and Hb traces





## **Prediction models**

- 1. <u>Machine Learning</u>: Random Forest (RF), Kernel Partial Least Squares (KPLS), Support Vector Regression (SVR)
- 2. <u>Deep Learning</u>: Brain-NET

#### **Simple Architecture**

- 1D convolutional kernels along the
- 1D convolutional kernels along the Pro location direction;
- Flatten layer;
- Two fully connected layers.



training

30 rounds of ten-fold cross validation using randomly shuffled samples for each round





### What is the effect of sample size?

- →Sample size: 13 medical students; 1000 raw data; 700 processed data (due to motion artifacts)
- $\rightarrow$  <u>Metric</u>: R<sup>2</sup> value
- $\rightarrow$  ML>DL for small sample sizes (<600)
- $\rightarrow$  DL>ML when #samples >600





#### $\rightarrow$ Full dataset: 702 samples

- $\rightarrow$  <u>Methodology</u>: Remove-one-component scheme
- $\rightarrow$  Feature: Removing the HbR slope produces the lowest R<sup>2</sup>;
- → Region: Removing the left middle PFC region produces the lowest R<sup>2</sup>
- $\rightarrow$  Overall: Best R<sup>2</sup> obtain with 7 features of the 2 regions



<b>Component removed</b>	<b>R</b> <sup>2</sup>
None	0.73
<b>Rightmost PFC</b>	0.66
Second right PFC	0.63
Medial right PFC	0.68
Leftmost PFC	0.70
Second left PFC	0.61
Medial left PFC	0.66

Component removed	HbO	HbR
None	0.	73
Mean	0.72	0.68
Slop	0.67	0.57
Variance	0.68	0.72
Skewness	0.71	0.67
Kurtosis	0.72	0.66
Max	0.67	0.63
Min	0.67	0.66



## $\rightarrow$ Of particular interests

## →Cross-validation: ten-fold and Leave-One-User-Out

 $\rightarrow$  An additional experiment: 6 new participants

			Subject No.	Total	True score	Predicted	Absolute
Madal	10 fold	ΙΟΠΟ		trials		score	error
	10-1010	1000	. 1	10	163	147	16
Brain-NET	$0.67 \pm 0.04$	0.61	2	10	180	162	18
KPLS	$0.64 \pm 0.01$	0.34	3	5	121	133	12
SVP	$0.58 \pm 0.01$	0.26	4	9	185	174	11
SVIC	0.30 <u>1</u> 0.01	0.20	5	5	238	240	2
RF	$0.53 \pm 0.01$	0.31	6	10	207	216	9



## **Prediction accuracy**





## **Board certification accuracy?**

- AUC value of the ROC curve (Hajian-Tilaki et al. 2013)
- Conventional for medical test accuracy: above 0.90 indicates good classifier.



#### Conclusion

Brain-NET demonstrate good prediction accuracy of FLS score.





The prefortal cortex grants us executive function, or the ability to make complex and socially mindful decisions. The prefrontal cortex must filter out unmitportant sensations, focus our attention on the task at hand, and contextualize our decisions with memories of the past and predictions about how our behavior will influence the future. Behaviors and personality traits unque to humans often depend on the preformal cortex.



#### Limitations:

1. Feature extraction is require



- 2. Limited sample size
  - Recruitment of specific cohort
  - Trial discard due to motion artifacts (300 in 1000 discarded)





• What is motion artifacts?

fNIRS data processing flow



• Efficient removal of motion artifacts could save trials and increase the accuracy of the information.



- Motion artifact removal methods in fNIRS
  - Spline (Scholkmann et al, 2010) Wavelet (Molavi and Dumont 2012)
  - Kalman (Izzetoglu et al. 2010)
     PCA (Zhang et al. 2005)
- Cbsi (Cui et al. 2010)

Drawbacks

- Depend on assumptions to describe motion artifacts
- Subjective selection of associated tuning of parameters
- Goal: A model that does not rely on assumptions or require subjective fine-tuning
- Propose: Deep learning (denoising autoencoder structure (DAE))
  - Have been shown superior in denoising of medical images
- Motivation:





#### **In-silico training**





26

First loss: mean squared error (MSE) loss

$$L_{mse} = \frac{1}{n} \sum_{i} (y_i - \hat{y}_i)^2$$

Second loss: the total variation of the predicted signal

$$L_{var} = \frac{1}{n} \sum_{i} (\hat{y}_i - \hat{\mu})^2$$

- Third loss: the number of motion artifacts
  - the standard deviation (std) exceeds the standard deviation threshold

$$L_{std} = \frac{1}{N_{std}} \sum_{i=1}^{T} \sum_{j \in \{j \mid \Delta_i dc_j > m\}} \Delta_i dc_j$$

the amplitude change (amp) exceeds the amplitude threshold

$$L_{amp} = \frac{1}{N_{amp}} \sum_{i=1}^{T} \sum_{j \in \{j \mid \Delta_i dc_j > C_{amp}\}} \Delta_i dc_j$$

Loss function: weighted sum of the above losses

$$Loss = L_{mse} + \theta_1 \times L_{var} + \theta_2 \times L_{std} + \theta_3 \times L_{amp}$$



#### **Denoise performance on simulation data**



Mean (std)	MSE ( $\mu Mol^2$ )	CNR
No correction	0.49 (0.69)	5.30 (4.69)
Spline	0.47 (0.90)*	5.47 (4.55)*
Wavelet05	0.44 (0.67)*	6.17 (6.22)*
Wavelet35	0.45 (0.67)*	6.04 (6.00)*
Kalman	0.48 (0.66)*	5.37 (5.02)
PCA99	0.81 (2.14)*	5.09 (4.98)*
PCA50	0.55 (0.93)*	5.48 (5.21)*
Cbsi	5.03 (10.31)*	6.49 (4.94)*
DAE	0.35 (0.53)*	5.51 (4.43)*

\*: significantly different from 'No correction' by t-test (p < 0.05)



#### **Denoise performance on simulation data**





#### **Denoise performance on experimental data**



Model	Computation time (s)
Spline	0.44
Wavelet05	81.22
Wavelet35	85.92
Kalman	1.48
PCA99	0.27
PCA50	0.18
Cbsi	0.03
DAE	0.19



#### **Denoise performance on experimental data**



Model	AUC0-2 ( <i>µMol</i> * <i>s</i> )	AUCratio
No correction	0.13 (0.21)	26.60 (98.54)
Spline	0.09 (0.13)*	64.75 (271.01)*
Wavelet05	0.03 (0.03)*	33.49 (86.91)
Wavelet35	0.05 (0.05)*	23.85 (43.85)
Kalman	0.17 (0.22)	22.74 (48.73)
PCA99	0.11 (0.15)*	85.24 (482.75)
PCA50	0.17 (0.25)*	54.69 (277.42)
Cbsi	0.07 (0.10)*	57.88 (205.15)
DAE	<b>0.02</b> (0.03)*	<b>98.26</b> (686.87)

\*: significantly different from 'No correction' by t-test (p < 0.05)

Brigadoi, Sabrina, et al. "Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data." *Neuroimage* 85 (2014): 181-191.



#### **Denoise performance on experimental data**





### **Brain-NET**

Provide a fast, accurate, and robust method to assess FLS score.

## DAE

Provide an assumption-free and effective method to remove fNIRS motion artifact.

## Significance:

- A step toward real time skill assessment;
- Impact the training





## **Motivations**



#### **Collaborator:**



Professor and Chairman -Department of Surgery Jacob School of Medicine and Biomedical Science University at Buffalo Former chair of the FLS committee

#### → Challenges in surgical training:

- The training protocol relies on repetition.
  - Time consuming;
  - Proficiency not guaranteed.

Specific aim 3 : to investigate whether surgical skill acquisition, retention, and transfer can be enhanced via neuromodulation.



#### **Neural basis of motor learning**

Neuroplasticity



Two steps:

- Rapid reinforcement of preestablished organic pathways;
- Later formation of new pathways





#### Cortex areas

Noninvasive brain stimulation





#### **Experiment design**

Hypothesis: tES (tDCS and tRNS) enhances the long-term acquisition, retention, and transfer of learning of the complex surgical skills.





'tDCS group', n = 5; 'tRNS group', n = 5; 'Sham group', n = 7;

37

#### tES setup



Gao, Y., et al. 2020. The effects of transcranial electrical stimulation on human motor functions: a comprehensive review of functional neuroimaging studies. Frontiers in Neuroscience, accepted.



38

#### fNIRS setup



13

#### **Commercial Instrument: Nirx**

• LED 760&850nm



A Nemani, et al., "Assessing bimanual motor skills with optical neuroimaging," Science Advances 4: eat3807 (2018).



#### **Behavioral Results – Time & FLS score**



> No significant difference prior to training or at the end of the training



#### **Behavioral Results - Error**





> No change in Sham

- ➤ tRNS↓ error in the beginning than Sham
- ➤ tDCS↓ error during the training than Sham



& IMAGING IN MEDICINE

**Standard deviation:** the consistency of the behavioral output.



> **tDCS**  $\downarrow$  the std of **FLS score and time** than **tRNS** and **Sham** 

ightarrow tDCS and tRNS  $\downarrow$  the std of error than Sham



#### **Brain Activation Results**



- Sham <sup>↑</sup> M1 activation (Floyer-Lea et al. 2005; Ma et al. 2010)
- **tDCS** ↑ M1 activation than Sham
- ➤ tRNS ↓ total activation than Sham



#### **Behavioral Results – retention**



- **tDCS** and **tRNS** \$\frac{1}{2}\$ error than Sham
- **tDCS †** time than Sham
- FLS score is not different



**Brain Activation Results – Retention** 



- **tDCS** ↑ brain activation than Sham
- ► tRNS ↓ brain activation than Sham
- brain activation is not concentrated in M1 region



**Behavioral Results – Follow Up & Transfer** 



> There is **no significant difference** between the three groups



#### **Brain Activation Results – Transfer**

mPFC

mPFC

rPFC

IIM1

lmM1

rmM1

rPFC

IIM1



> There is **no significant difference** between the three groups



tDCS tRNS Sham

ĪĮ

SMA

SMA

rlM1

rIM1

rmM1

lmM1

#### **Summary**

Specific aim 3 : to investigate whether surgical skill performance, acquisition, retention, and transfer can be enhanced via neuromodulation.

- Conclusion:
  - tDCS facilitated surgical motor learning by lowering the performance error.
  - The M1 cortex excitation was enhanced with tDCS.
- Impact:
  - Offers a tool to enhance surgical motor skills, especially decrease the surgical errors



Specific aim 1: to predict learning curve features in the early stage of training.

- Learning curve features can be predicted from the initial performance;
- Single factor can represent the learning curve features;

Specific aim 2 : to predict FLS scores via neuroimaging.

- Neuroimaging can predict FLS score;
- DAE model is superior in fNIRS motion artifact removal

Specific aim 3 : to investigate whether surgical skill performance, acquisition, retention, and transfer can be enhanced via neuromodulation.

 tDCS enhanced the lengthy learning procedure by reducing the performance error via strengthening brain activation in the M1 region.



#### **Overview and Future Work**





#### **Publications**

#### <u>Journal papers</u>

#### Published/Accepted:

Gao, Y., Kruger, U., Intes, X., Schwaitzberg, S. and De, S., 2020. A machine learning approach to predict surgical learning curves. *Surgery*, 167(2), pp.321-327. (S.A. #1)

Gao, Y., Cavuoto, L., Schwaitzberg, S., Norfleet, J., Intes, X. and De, S., 2020. The effects of transcranial electrical stimulation on human motor functions: a comprehensive review of functional neuroimaging studies. *Frontiers in Neuroscience*, accepted. (review for S.A. #3)

#### Under review:

Gao, Y., Yan, P., Kruger, U., Cavuoto, L., Schwaitzberg, S., De, S. and Intes, X. 2020. Functional brain imaging reliably predicts bimanual motor skill performance in a standardized surgical task. *IEEE TBME*, under review. (S.A. #2)
Gao, Y., Chao, H., Cavuoto, L., Yan, P., Kruger, U., Norfleet, J., Makled B., Schwaitzberg, S., De, S. and Intes, X., 2020. Deep learning-based motion artifact removal in functional near-infrared spectroscopy (fNIRS). *NeuroImage*, under review. (S.A. #2)



#### Journal papers (continued)

#### In Preparation:

Gao, Y., Cavuoto, L., Dutta, A., Kruger, U., Yan, P., Norfleet, J., Makled B., Silverstri, J., Schwaitzberg, S., Intes, X. and De, S., 2020. tDCS-induced primary motor cortex activation diminishes errors in fine bimanual motor skill learning. *Nature biomedical engineering*, in polish. (S.A. #3)

A draft for the transfer learning is also under preparation (S.A. #3);

A Nemani, **Gao**, **Y.**, M Yucel, D Gee, C Cooper, S Schwaitzberg, X Intes and S De, "Functional brain connectivity distinguishes surgical skill dexterity in both physical and virtual simulation environments," Neuroimage, in polish.



#### **Conference papers**

Yuanyuan Gao, Pingkun Yan, Uwe Kruger, Suvranu De and Xavier Intes, "Neuroimaging biomarkers for surgical skill level prediction", SPIE.bios, San Francisco, CA, February 2019.

**Yuanyuan Gao**, Pingkun Yan, Uwe Kruger, Suvranu De and Xavier Intes, "fNIRS as a quantitative tool to asses and predict surgical skills", OSA Biophotonics Congress: Optics in the Life Sciences, Tucson, AZ, April 2019.

Yuanyuan Gao, Lora Cavuoto, L., Pingkun Yan, Uwe Kruger, Steven Schwaitzberg, Suvranu De, and Xavier Intes, "A deep learning approach to remove motion artifacts in fNIRS data analysis". *In Optics and the Brain, Optical Society of America, 2020.* 

**Yuanyuan Gao**, Lora Cavuoto, L., Pingkun Yan, Uwe Kruger, Jessica Silvestri, Steven Schwaitzberg, Xavier Intes and Suvranu De, "Monitoring the effect of transcranial Electric current Stimulation (tES) during a bimanual motor task via functional Near-InfraRed Spectroscopy (fNIRS)". *In Optics and the Brain, Optical Society of America, 2020*.

Yuanyuan Gao, Lora Cavuoto, L., Pingkun Yan, Uwe Kruger, Jessica Silvestri, Steven Schwaitzberg, Jack E. Norfleet, Basiel A. Makled, Xavier Intes and Suvranu De, "Transcranial direct current stimulation speeds up surgical bimanual motor learning and increases functional activation", *In MHSRS, Young Investigator breakout session award paper, August 2020*.



# **Acknowledgements**

- Committee members
  - Suvranu De, Xavier Intes, Pingkun Yan, Lucy Zhang, Emily Liu
- CeMSIM and Intes lab members
- NIH grant funding
- Tech support: TechEn (Dr. A. "Buzz" DiMartino), NIRx, NVIDIA corp.
- Collaborators:



Steven D. Schwaitzberg MD University at Buffalo



Lora Cavuoto PhD University at Buffalo



Anirban Dutta PhD University at Buffalo





Alvaro Pascual-Leone PhD Harvard Medical School



Emiliano Santarnecchi, PhD. Harvard Medical School



Rensselaer

CNBS

University

at Buffalo

**Uwe Kruger** RPI

M. Yucel and D. Boas, PhD **Boston University** 







BU



# Thank you!



# Backups



## **Pilot study**

- Hypothesis: tES (tDCS and tRNS) enhances the surgical bimanual task performance in a short term.
- Experimental design:





#### **Experiment #1**

- Hypothesis: tRNS enhances the surgical bimanual task performance in a short term and changes the brain activation.
- Experimental design:



Motor task	Pattern cutting
Task repetition	4
N of trainees	12

Nemani, Arun, et al. "Assessing bimanual motor skills with optical neuroimaging." Science advances 4.10 (2018): eaat3807.



# **Pilot study**



The FLS score increased significantly under tRNS condition.



#### **Experiment #1**



- The FLS score increased under tRNS condition;
- Time decreased for both conditions;
- Error decreased under tRNS condition.



#### **Experiment #1**







Preliminary results



