

**Objective assessment of bimanual laparoscopic surgical skills via  
functional near infrared spectroscopy (fNIRS)**

by

Arun Nemani

A Thesis Submitted to the Graduate  
Faculty of Rensselaer Polytechnic Institute  
in Partial Fulfillment of the  
Requirements for the Degree of  
DOCTOR OF PHILOSOPHY  
Major Subject: Biomedical Engineering

Approved by the  
Examining Committee:

---

Suvaranu De, Thesis Co-Adviser

---

Xavier Intes, Thesis Co-Adviser

---

Ge Wang, Member

---

David Boas, Member

---

Ganesh Sankaranarayanan, Member

Rensselaer Polytechnic Institute  
Troy, New York  
December, 2017  
(For Graduation December 2017)

© Copyright 2017  
by  
Arun Nemani  
All Rights Reserved

# CONTENTS

LIST OF TABLES.....	vi
LIST OF FIGURES .....	vii
ACKNOWLEDGMENT .....	xv
ABSTRACT .....	xvii
1. Introduction.....	1
1.1 Methods of surgical technical skill assessment.....	1
1.1.1 Direct observation methods: Global rating scales and checklist based assessment.....	2
1.1.2 Bimanual dexterity based assessment .....	5
1.1.3 Physical simulation assessment .....	6
1.1.4 Virtual reality (VR) simulation assessment .....	8
1.1.5 Robot assisted surgical skill assessment .....	9
1.1.6 Alternative surgical skill assessment .....	9
1.1.7 Limitations of current assessment methods .....	10
1.2 Non-invasive brain imaging methods for motor skill assessment .....	12
1.2.1 Electroencephalography (EEG) .....	12
1.2.2 Magnetoencephalography (MEG).....	13
1.2.3 Functional magnetic resonance imaging (fMRI) .....	13
1.2.4 Positron emission tomography (PET) .....	14
1.2.5 Limitation of non-invasive brain imaging methods.....	14
1.3 Functional near-infrared spectroscopy (fNIRS).....	16
1.3.1 fNIRS modeling .....	16
1.3.2 fNIRS hardware .....	18
1.3.3 fNIRS in bimanual surgical motor skills.....	20
1.4 Neurophysiology of human motor skill learning and retention .....	21
1.4.1 Motor control physiology.....	21
1.4.2 Motor skill learning.....	25
1.4.3 Motor skill retention.....	29
1.4.4 Gaps in bimanual motor skill neurophysiology knowledge.....	30
1.5 Specific aims .....	31

1.6	Research plan and thesis outline .....	32
2.	Assessing bimanual motor skills with optical neuroimaging .....	33
2.1	Overview .....	33
2.2	Methods.....	35
2.2.1	Hardware and equipment .....	35
2.2.2	Participants and experimental design.....	36
2.2.3	NIRS post processing.....	39
2.2.4	Task performance metrics, statistical, and classification methods .....	39
2.3	Results.....	40
2.3.1	Surgical training task performance assessment.....	40
2.3.2	Optical neuroimaging assessment of established surgical skill levels .....	43
2.3.3	Optical neuroimaging assessment of surgical skill level during training .....	48
2.3.4	Classification of subjects with varying surgical expertise levels.....	52
2.4	Discussion .....	54
2.5	Summary .....	61
3.	Objective assessment of surgical skill transfer using non-invasive brain imaging .....	62
3.1	Overview .....	62
3.2	Methods.....	64
3.2.1	Subject recruitment .....	64
3.2.2	Simulation hardware .....	64
3.2.3	Study design, fNIRS processing, and statistical methods .....	65
3.2.4	Learning curve and task retention study design.....	65
3.2.5	Transfer task study design.....	66
3.2.6	Task performance metrics .....	67
3.3	Results.....	67
3.3.1	FLS and VBLaST simulator performance learning curves.....	67
3.3.2	Differentiation and classification of motor skill transfer based on traditional task performance.....	70

3.3.3	Neuroimaging-based metrics for differentiation and classification of motor skill transfer.....	72
3.4	Discussion .....	75
3.5	Summary .....	78
4.	Brain connectivity analysis for surgical skill assessment in physical and virtual surgical simulators .....	79
4.1	Overview .....	79
4.2	Methods.....	80
4.2.1	Subject recruitment .....	80
4.2.2	Hardware, study design, and fNIRS processing.....	81
4.2.3	Wavelet coherence and wavelet phase coherence.....	81
4.3	Results .....	85
4.3.1	Wavelet coherence between surgical experts and novices.....	85
4.3.2	Wavelet coherence between surgically trained and untrained subjects.....	86
4.4	Discussion .....	88
4.5	Summary .....	89
5.	Summary and future work .....	90
5.1	Thesis summary .....	90
5.2	Future work .....	91
5.2.1	Expansion of cortical coverage .....	91
5.2.2	High density probe measurements .....	92
5.2.3	Prediction of motor skill levels .....	92
5.2.4	Brain imaging for non-technical surgical skills .....	93
	REFERENCES .....	94

## LIST OF TABLES

Table 1.1:	Summary of surgical technical skill assessment methods and major limitations. ....	11
Table 1.2:	Advantages and disadvantages of currently used functional brain imaging modalities [60], [61]. ....	15
Table 1.3:	Summary of fNIRS studies to assess surgical motor skill. ....	20
Table 2.1:	Subject demographics and descriptive data. ....	37
Table 2.2:	Expert vs Novice classification results for fNIRS (with and without short separation regression) and FLS metrics. ....	56
Table 4.1:	Study subject demographics and training procedures completed. ....	81
Table 4.2:	Frequency bandwidth intervals with their associated physiology [156], [182]–[184], [195]. ....	84

## LIST OF FIGURES

Figure 1.1:	Sample OSATS form for trainees performing laparoscopic suturing and intracorporeal knot-tying tasks [8].....	3
Figure 1.2:	Sample GRITS rating form to assess trainee laparoscopic technical skills [3]. .....	5
Figure 1.3:	FLS training and assessment tasks that are required for board certification in general surgery. (a) The FLS box trainer trains and assesses laparoscopic training tasks such as (b) peg transfer, (c) pattern cutting, (d) ligation loop, (e) extracorporeal knot tying, and (f) intracorporeal knot trying [31].....	7
Figure 1.4:	Schematic outlining the temporal and spatial resolution estimates for common non-invasive brain imaging methods [64]. .....	15
Figure 1.5:	Molar coefficients of HbO <sub>2</sub> and HbR within the optical window (redrawn from [67]). .....	17
Figure 1.6:	Different modalities of fNIRS imaging system that include (a) continuous-wave, (b) frequency domain, and (c) time domain systems [70]. .....	19
Figure 1.7:	Somatotopic representation of body parts from medial to lateral direction in the precentral gyrus[84].....	22
Figure 1.8:	Corticospinal tract diagram illustrating specific pathway of motor neurons from the cerebrum to muscle motor units [83]. .....	24
Figure 1.9:	Schematic illustrating the motor skill learning pathway involving spatial and motor sequences [82].....	27
Figure 1.10:	Schematic illustrating neural substrates and their activations during slow motor learning [113]. .....	29
Figure 2.1:	Infrared probe geometry positioning. Schematic of probe placement projected on cortical locations specific to the PFC, M1, and SMA. Optodes are placed for maximum coverage over the PFC, M1, and SMA. Red dots indicate infrared sources, blue dots indicate long separation detectors, and textured blue dots indicate short separation detectors. The PFC has three sources (1-3), the M1 has four sources (4-7), and the SMA has one source (8). Each of the sources are connected to their corresponding long and short separation detectors [142].....	36

Figure 2.2:	Subjects performing FLS pattern cutting task with fNIRS measurements. (a) An example medical subject, part of the FLS training group, and an example novice surgeon (b) performing a FLS pattern cutting task on the official FLS box trainer while undergoing fNIRS measurements in real time.....	38
Figure 2.3:	Schematic outlining cohort and study design. FLS training group (gray) and VBLaST training group (purple) underwent training regiments whereas the untrained control group (orange), surgical novices (green) and surgical experts (red) underwent no training. <i>m</i> is the number of trials per each session block and <i>n</i> is the number of pattern cutting trials per each session block. ....	38
Figure 2.4:	Schematic depicting the FLS box simulator where trainees perform the bimanual dexterity task. A continuous wave spectrometer is used to measure functional brain activation via raw fNIRS signals in real-time. ....	41
Figure 2.5:	FLS performance scores for Novice surgeons (green) and Expert surgeons (maroon) where Expert surgeons significantly outperformed Novice surgeons. Two sample t-tests were used for statistical differentiation ( <sup>n.s.</sup> not significant, * <i>p</i> <0.05).....	41
Figure 2.6:	FLS performance scores for all training subjects (black) with respect to days trained compared to untrained Control subjects (orange) ( <sup>n.s.</sup> not significant, * <i>p</i> <0.05). ....	42
Figure 2.7:	CUSUM scores for each subject with respect to trials. The $H_0$ threshold indicates that the probability of any given subject is mislabeled as a “Skilled trainee” is less than 0.05, and is subsequently labeled as a “Skilled trainee” subject. Results indicate that three subjects, FLS-2, FLS-3, and FLS-5 are labeled as “Skilled trainees”. The remaining subjects that do not cross the $H_0$ line are labeled “Unskilled trainees”. ....	43
Figure 2.8:	Group average hemodynamic response functions with respect to cortical regions. Group average of all trials during the post-test for expert surgeons (maroon) and novice surgeons (green). Stimulus onset begins at zero seconds (dashed black line) indicating that the trial has started. Negative time indicates the baseline measurements used for calibration before each trial. ....	44
Figure 2.9:	Differentiation and classification of motor skill between Novice and Expert surgeons. (a) Brain region labels are shown for prefrontal cortex (PFC), primary motor cortex (M1) and supplementary motor area (SMA) regions. Average functional activation for all subjects	

	in the Novice and Expert surgeon groups are shown as spatial maps while subjects perform the FLS task. ....	44
Figure 2.10:	Average changes in hemoglobin concentration during the FLS task duration with respect to specific brain regions for Novice (green) and Expert (maroon) surgeons. Two sample t-tests were used for statistical tests ( <sup>n.s.</sup> not significant, * p<0.05). ....	45
Figure 2.11:	LDA classification results between Experts and Novices for FLS scores and all combinations of fNIRS metrics. ....	46
Figure 2.12:	Leave-one-out cross-validation results show the ratio of samples that are below misclassification error rates of 0.05 for FLS scores and all other combinations of fNIRS metrics. ....	47
Figure 2.13:	Probability density functions (PDFs) for projected LDA classification models. PDFs derived from kernel density estimation of normalized FLS performance for Novices, Experts, Skilled trainees, Unskilled trainees, and Control during the post-test. fNIRS metrics used for classification are functional activation in the PFC, LMM1, and SMA. The type I error is defined as 0.05 for all cases. (a) Using only FLS task performance as the only metric, results show that the probability for a Novice surgeon being misclassified as an Expert surgeon is 53% (MCE <sub>2</sub> ) and the probability that a Novice surgeon is misclassified as an Expert surgeon is 61% (MCE <sub>1</sub> ). (b) fNIRS based classification results show that the probability for a Novice surgeon being misclassified as an Expert surgeon is 4.4% (MCE <sub>2</sub> ) and the probability that a Novice surgeon is misclassified as an Expert surgeon is 4.3% (MCE <sub>1</sub> ). Similarly, Control subjects are classified against Unskilled and Skilled subjects. ....	48
Figure 2.14:	Differentiation and classification of motor skill between Control, Skilled, and Unskilled trainees. (a) Spatial maps of average functional activation for all subjects in each respective group during the FLS training task on the post-test day. ....	49
Figure 2.15:	Average changes in hemoglobin concentration during stimulus duration with respect to specific brain regions for untrained Control subjects (orange) and all FLS training students (black). Two sample t-tests were used for statistical differentiation ( <sup>n.s.</sup> not significant, * p<0.05). Type I error is defined as 0.05 for all cases. ....	50
Figure 2.16:	Inter and intra-group misclassification errors for each subject population (Control, Skilled and Unskilled trainees) with respect to training days. MCE <sub>12</sub> and MCE <sub>21</sub> values significantly decrease	

below 5% when classifying pre-test Skilled and Unskilled trainees on the final training day. Furthermore, misclassification errors are also low when classifying Skilled and Unskilled trainees on the final training day, along with Skilled trainees and untrained Control subjects. .... 51

Figure 2.17: Misclassification errors are reported for each combination of training groups (Control, Skilled, and Unskilled trainees) with respect to pre-test, post-test, and final training days. MCEs are substantially low when classifying Skilled trainees and Control subjects along with inter-Skilled trainee group classification. Unskilled trainees, however, showed high misclassification errors even when compared to Unskilled trainees and Control subjects during the post-test. As a measure of skill retention, classification models were also applied for all subject groups from the final training day to the post-test..... 52

Figure 2.18: (a) The probability that an untrained Control subject is misclassified as an Unskilled trainee is 46% ( $MCE_1$ ) and the probability that an Unskilled trainee is misclassified as a Control is 50% ( $MCE_2$ ). (b) Conversely, the probability that a Control is misclassified as a Skilled trainee is 16% ( $MCE_1$ ). Whereas the probability that a Skilled trainee is misclassified as a Control is 9.5% ( $MCE_2$ )..... 52

Figure 2.19: Cross-validation results for classification across all subjects with varying degree of motor skills. Each box represents one trial per expertise group during the post-test, where the shaded regions indicate the MCE if that given trial is removed from the classification model. Cross-validation results with their respective ratio of samples that are below misclassification error rates of 5% for Expert surgeons vs Skilled trainees (28/35 samples), Expert surgeons vs Unskilled trainees (29/ 38), Expert vs Novice surgeons (43/43), Expert surgeons vs untrained Control subjects (34/38), Skilled trainees vs Unskilled trainees (15/21), Skilled trainees vs Novice surgeons (24/26), Skilled trainees vs untrained Control subjects (18/21), Unskilled trainees vs Novice surgeons (16/29 samples), Unskilled trainees vs untrained Control subjects (11/24), and finally Novice surgeons vs untrained Control subjects (9/29)..... 53

Figure 2.20: Cross-sectional diagram of infrared light propagation through cortical tissue. The short separation detectors are placed 8mm away from each source to ensure the backscattered light is solely due to superficial tissue, such as scalp, skull, dura, arachnoid, and pia matter. The large separation detectors are placed 3-4cm away from each source to ensure sufficient light penetration depth into the cortex is achieved. .... 54

Figure 2.21:	Differentiation and classification of motor skill between Novice and Expert surgeons without short separation regression. (a) Average changes in hemoglobin concentration during the FLS task duration with respect to specific brain regions for Novice (green) and Expert (maroon) surgeons. Two sample t-tests were used for statistical tests ( <sup>n.s.</sup> not significant, <sup>*</sup> p<0.05).....	55
Figure 2.22:	LDA classification results for FLS scores and all combinations of fNIRS metrics using fNIRS metrics with superficial tissue signals included.....	56
Figure 2.23:	Leave-one-out cross-validation results show the ratio of samples that are below misclassification error rates of 0.05 for FLS scores and all other combinations of fNIRS metrics when measurements include superficial tissue signals. ....	56
Figure 2.24:	Quadratic support vector machine (SVM) classification of Expert and Novice surgeons. (a) Unsupervised, quadratic polynomial support vector machines were used for bivariate classification of Expert and Novice surgeons. Quadratic support vectors outlining the decision boundaries are shown. (b) Receiving operating characteristic (ROC) curve showing the sensitivity vs specificity of SVM based classification. (c) Weights from the corresponding LDA classification of Expert vs Novice surgeons were used to combine the cortical activations as a single metric. (d) The corresponding ROC curves for the weighted fNIRS metric based classification vs traditional FLS scores. (e-f) Confusion matrices showing the specific results of true positive and true negative classes (green) along with the false positive and false negative classes (maroon). Results indicate that SVM based on weighted fNIRS classification show no cases of false negatives. ....	58
Figure 2.25:	Quadratic support vector machine (SVM) classification of Skilled vs Unskilled trainees. (a) Unsupervised, quadratic polynomial support vector machines were used for bivariate classification of Skilled and Unskilled trainees. Quadratic support vectors outlining the decision boundaries are shown. (b) Receiving operating characteristic (ROC) curve showing the sensitivity vs specificity of SVM based classification. (c) Weights from the corresponding LDA classification of Skilled vs Unskilled trainees were used to combine the cortical activations as a single metric. (d) The corresponding ROC curves for the weighted fNIRS metric based classification vs traditional FLS scores. (e-f) Confusion matrices showing the specific results of true positive and true negative classes (green) along with the false positive and false negative classes (red). ....	59

Figure 3.1:	FLS and VBLaST simulators. The physical FLS pattern cutting (PC) box trainer (left) and the VBLaST PC simulator (right) used in this study [142].	65
Figure 3.2:	Schematic illustrating the learning curve study design. Two training groups, VBLaST (blue) and FLS (magenta), undergo a training period whereas the control group (green) only perform the baseline test (Day 1), retention, and transfer task tests.	66
Figure 3.3:	Pattern cutting transfer task <i>ex-vivo</i> sample. (a) <i>Ex-vivo</i> peritoneum sample prior to transfer task completion for FLS trained subject 3. (b) Completed pattern cutting transfer task for FLS trained subject 3 with the pattern cutting task replicated and the marked peritoneal tissue resected [142].	67
Figure 3.4:	FLS pattern cutting performance are shown with respect to training day. FLS training students (magenta) are compared to untrained control students (green). FLS pattern cutting task retention scores are shown for trained FLS students (magenta), untrained control subjects (green), and VBLaST trained subjects (blue). Mann-Whitney U tests were used to statistically differentiate the control and FLS training groups ( <sup>n.s.</sup> not significant, * $p<0.05$ ).	68
Figure 3.5:	VBLaST pattern cutting performance are shown with respect to training day. VBLaST training students (blue) are compared to untrained control students (green). VBLaST pattern cutting task retention scores are shown for trained VBLaST students (blue), untrained control subjects (green), and FLS trained subjects (magenta). Mann-Whitney U tests were used to statistically differentiate the control and FLS training groups ( <sup>n.s.</sup> not significant, * $p<0.05$ ).	69
Figure 3.6:	CUSUM scores for trained FLS and VBLaST groups. CUSUM scores for each subject with respect to number of trials. The threshold score to be considered a senior in the pattern cutting task is 63[16]. (a) CUSUM scores indicate that three subjects (FLS2, FLS3, FLS5) achieved the level of senior during the FLS training period. (b) CUSUM scores indicate that four subjects (VBLaST1, VBLaST4 – 6) achieved the level of senior during the VBLaST training period.	70
Figure 3.7:	Transfer task completion times for the trained FLS, untrained control, and trained VBLaST subjects (* $p<0.05$ ).	71

Figure 3.8:	(a) LDA classification of trained FLS and control subjects during the transfer task based on completion times and (b) corresponding crossvalidation results.....	71
Figure 3.9:	(a) LDA classification of trained VBLaST and control subjects during the transfer task based on completion times and (b) corresponding crossvalidation results.....	72
Figure 3.10:	Changes in cortical activation during the transfer task with respect to cortical regions. Average changes in hemoglobin ( $\Delta\text{HbO}_2$ ) concentration as a measure of functional activation with respect to different cortical regions for FLS trained subjects (magenta), untrained control subjects (cyan), and VBLaST trained subjects (black) while all subjects perform the <i>ex-vivo</i> transfer task. ....	73
Figure 3.11:	The cumulative set of $\text{MCE}_1$ and $\text{MCE}_2$ for all combinations of fNIRS metrics and the transfer task completion time to classify FLS trained and control subjects. ....	74
Figure 3.12:	The cumulative set of $\text{MCE}_1$ and $\text{MCE}_2$ for all combinations of fNIRS and transfer task metrics to classify VBLaST trained and control subjects. ....	74
Figure 3.13:	Leave-one-out crossvalidation results indicate the misclassification errors for each sample treated as an independent sample for the LDA model using all combinations of fNIRs and transfer task metrics, where the percent of samples that have MCE below 0.05 for each possible metric combination are shown. ....	75
Figure 4.1:	Example wavelet coherence between two different fNIRS time series data. (a) Two time series data from the left lateral PFC (LPFC) and left medial M1 (LMM1) channels for a surgical expert during one FLS task trial (b) Wavelet coherence magnitude between the two time series data based on time and frequency domains. Wavelet coherence magnitude values are shown via the color bar. Only values within the cone of influence range, indicated by a dashed white line, are included for wavelet coherence power magnitude and phase coherence calculations. (c) Time-averaged wavelet coherence magnitudes and (d) wavelet phase coherence magnitudes between the two example time series shown in (a).....	84
Figure 4.2:	Wavelet coherence and wavelet phase coherence magnitude changes between experts and novices on physical or virtual simulators. (a-b) Wavelet coherence magnitudes and wavelet phase coherence magnitudes for FLS experts (red) vs novices (green) within the neurovascular coupling activity frequency range. (c-d)	

	Wavelet coherence magnitudes and wavelet phase coherence magnitudes for VBLaST experts (purple) vs novices (blue) within the neurovascular coupling activity frequency range. ....	86
Figure 4.3:	Longitudinal wavelet phase coherence in neurogenic activity frequency range with increasing surgical skill training. WPCO magnitudes within the neurovascular coupling activity frequency range (V) for between the CPFC and SMA channels for untrained control subjects, trained FLS (a) and trained VBLaST (b) subjects as training and motor skill proficiency increases with training days.....	87
Figure 4.4:	Functional connectivity schematics for training and untrained subjects post training on virtual or physical simulators. Schematic showing the functional connectivity differences, as shown by significant changes in WPCO in the neurogenic activity frequency ranges, between trained and untrained subjects.....	87

## ACKNOWLEDGMENT

With my graduate experience nearly ending, I am filled with nostalgic memories of the past 7 years at RPI. Many of these experiences have made a lifelong impact on my beliefs, relationships, and character. I would like to thank all my colleagues, friends, and family that have supported me through this journey. I would not have been able to achieve this dream without all of you.

First and foremost, I would like to acknowledge my co-advisors Drs. Suvaranu De and Xavier Intes for their tremendous support and trust. Both have been excellent mentors throughout my PhD career and constantly pushing me towards a standard of scientific rigor and excellence that I will take with me going forward. Suvaranu is someone who has incredible vision and has provided tremendous guidance on my research. We would have conversations where he encouraged me to think bigger and broader, and to make a meaning contribution with my work. Many of these conversations happened during our Boston road trips, where we would discuss research, political satire, science, and current events. In fact, my PhD topic spawned from one of these trips on one fateful late-night ride home from an NIH meeting in Boston.

I also thank Dr. Intes for his immense patience, guidance, and advice. Xavier was someone that I would turn to for technical discussions, career advice, and pretty much everything else. I admire his leadership and management skills and I truly learned a tremendous amount from simply observing him. He would often walk into the office and say “Papers? Papers?!?” and immediately walk out. This kind of humor, his drive, his passion for science, would motivate us daily to keep working hard. Both Suvaranu and Xavier have been mentors that have literally shaped my drive, work ethic, and passion for research, and for this I am extremely grateful. I would like to thank my committee members Drs. Ge Wang, David Boas, and Ganesh Sanakaranarayanan. Their inputs throughout my thesis project have been very valuable and I am very grateful for them to take time out of their days and attend my meetings.

I would like to thank, the VPR office, the BME department, and RPI for providing the academic, financial, and administrative support that allowed me to pursue my doctorate. Funding grants such as the NIH / NIBIB 1R01EB014305, NHBLI 1R01HL119248, and NCI 1R01CA197491 allowed us researchers to pursue advances in

science, and I am very grateful for such programs. I would like to specifically thank Dr. Uwe Kruger. His efforts and guidance in extracting the most out of my dataset truly added a new dimension to this work (pun intended). I am extremely grateful for his input and wisdom regarding my work. Lastly, I would also like to thank Dr. Deanna Thompson for believing in me when I needed it most during my transition to the PhD program.

I would like to thank all my colleagues during my time at RPI. First, I would like to thank our clinical collaborators Drs. Steven Schwaitzberg, Clairice Cooper, and Denise Gee. Each of them provided so much support to help me complete my studies and am very grateful for their guidance and help. I would also like to thank Dr. Boas and the Martinos Imaging center team for their immense support for helping me complete my studies. Dr. Meryem Yucel, for example, spent many hours training and educating me with infrared systems to make sure I would succeed when I started my studies. She was also instrumental in helping me with data processing and I am very grateful she is a true colleague in every sense of the word.

I would also like to thank my lab mates and friends during my time at RPI. I have bonded with so many people and created so many memories that there are simply too many to name. From intellectual discussions in both CeMSIM and Intes lab offices to NYC road trips, I am grateful to all my friends at RPI that made my graduate experience so fun. Many of these lab mates and friends I consider family, and will cherish for life.

My parents and my brother, often hundreds of miles away, always felt so close during my time at RPI. They would always support me during the tough times and am extremely grateful for their unwavering love and support. Finally, I would like to thank my wife, Akhila. We met when I was a third year PhD student and she has been by my side in every aspect during my graduate school career. Her support and unconditional love provided me with inspiration and reassurance when I needed it most. I am grateful and honored to consider her as a life partner and someone I can rely on for everything life has to offer.

## ABSTRACT

Surgical simulators are effective methods for training and assessing surgical technical skills, particularly those that are bimanual. These simulators are now ubiquitous in surgical training and assessment programs for residents. Simulators are used in programs such as the Fundamentals of Laparoscopic Surgery (FLS) and Fundamentals of Endoscopic Surgery (FES), which are pre-requisites for Board certification in general surgery. Although these surgical simulators have been validated for clinical use, they have significant limitations, such as subjectivity in assessment metrics, poor correlation of transfer from simulation to clinically relevant environments, poor correlation of task performance scores to learning motor skill levels, and ultimately inconsistent reliability of these assessment methods as an indicator of positive patient outcomes. These limitations present an opportunity for more objective and analytical approaches to assess surgical motor skills. To address these surgical skill assessment limitations, we present functional near-infrared spectroscopic (fNIRS), a non-invasive brain imaging method, to objectively differentiate and classify subjects with varying degrees of laparoscopic surgical motor skill levels based on measurements of functional activation changes.

In this work, we show that fNIRS based metrics can objectively differentiate and classify surgical motor skill levels with significantly more accuracy than established metrics. Using classification approaches such as multivariate linear discriminant analysis, we show evidence that fNIRS metrics reduce the misclassification error, defined as the probability that a trained subject is misclassified as an untrained subject and vice versa, from 53 – 61% to 4.2 – 4.4% compared to conventional metrics for surgical skill assessment. This evidence also translates to surgical skill transfer metrics, where such metrics assess surgical motor skill transfer from simulation to clinically relevant environments. Results indicate that fNIRS based metrics can successfully differentiate and classify surgical motor skill transfer levels by reducing the misclassification errors from 20 – 41 % to 2.2 – 9.1%, when compared to conventional surgical skill transfer assessment metrics. Furthermore, this work also shows evidence of high functional connectivity between the prefrontal cortex and primary motor cortex regions correlated to increases in surgical motor skill levels, addressing the gap in current literature in underlying neurophysiological responses to surgical motor skill learning.

This work is the first to show conclusive evidence that fNIRS based metrics can significantly improve subject classification for surgical motor skill assessment compared to metrics currently used in Board certification in general surgery. Our approach brings robustness, objectivity, and accuracy in not only assessing surgical motor skill levels but also validating the effectiveness of future surgical trainers in assessing and translating surgical motor skills to more clinically relevant environments. This non-invasive imaging approach for objective quantification for complex bimanual surgical motor skills will bring about a paradigm change in surgical certification and assessment, that may lead to significantly reduced negative patient outcomes. Ultimately, this approach can be generally applied for bimanual motor skill assessment and can be applied for other fields, such as brain computer interfaces (BCI), robotics, stroke and rehabilitation therapy.

# 1. Introduction

During the 1890s, William Halsted proposed the “see one, do one, teach one” approach for surgical training and assessment during residency. This approach comprised an apprenticeship model for residents to learn and retain surgical technical skills, and is largely still utilized in medical training programs today. With more recent advances in surgery, such as minimally invasive surgery (MIS) and laparoscopy based procedures, surgical technical skill requirements are increasingly more demanding due to the complexity of these procedures. While significant strides have been taken to improve surgical technical skill training and assessments, these methods are still subjective or underdeveloped. Furthermore, very little attention has been given to the underlying neurophysiological mechanisms of learning and retaining laparoscopic surgical motor skills. The purpose of this thesis is to propose non-invasive brain imaging, namely functional near-infrared spectroscopy (fNIRS), as a viable and clinically translatable method for accurate bimanual laparoscopic skill assessment.

This chapter starts with section 1.1, a review of the current assessments methods for surgical technical skill, especially for laparoscopic surgical procedures. Section 1.2 is a brief overview of the current non-invasive brain imaging methods utilized in research today, including their respective advantages and limitations. Section 1.3 is an overview of fNIRS theory and modeling, an overview of hardware types, and a literature review of fNIRS studies specific to bimanual surgical skill assessment. Section 1.4 provides an overview of motor skill learning, execution, and retention theories, specific to bimanual tasks. Section 1.5 defines each specific aim required to achieve research goals in this thesis. Finally, section 1.6 provides a thesis outline for the remainder chapters.

## 1.1 Methods of surgical technical skill assessment

Surgical skill assessment plays a vital role in surgical skill training and core competencies of surgical residents. This section is devoted to reviewing the current surgical skill assessment methodologies, that include direct observations methods, where an expert physician will directly observe a trainee and assess the core competencies of surgical technical skills, bimanual dexterity based analysis, simulation based assessment (physical and virtual). These methods, however, have significant limitations

that are addressed in this chapter and serve as the motivation for developing objective metrics based on functional connectivity, which is the focus of our work.

### **1.1.1 Direct observation methods: Global rating scales and checklist based assessment**

Traditional surgical assessment methods, such as direct observations by an experienced trainer to assess the skills of the trainee, are generally subjective and use global rating scales (GRS) to score competency. These observational methods allow experienced surgeons to use structured checklists for technical criteria and rate the surgical performance of the trainee under direct observation [1]–[4].

One such method is the Objective Structure Assessment of Technical Skills (OSATS) [4]–[6]. OSATS incorporates six different stations where trainees perform complex surgical training tasks on either live animals or benchtop models. Proctors, who are generally expert surgeons, will assess performance using two mechanisms. The first is a task-specific checklist consisting of maneuvers that have been deemed critical for that particular surgical task [5]. The second mechanism is a global rating form, that includes non-technical metrics, such as surgical behaviors, respect for tissue, economy of motion, and appropriate usage of surgical staff and assistants. Both of these mechanisms are generally combined onto a single form where experts use Likert scale based scores to assess each trainee according to each checklist component [7]–[9]. An example OSATS form is shown in Figure 1.1.

Date: \_\_\_\_\_ Study ID: \_\_\_\_\_ Evaluator Name: \_\_\_\_\_  
 Training level: (circle one) MS1 MS2 MS3 MS4 PGY1 PGY2 PGY3 PGY4

### Objective Structured Assessment of Technical Skills (OSATS) for Laparoscopic Suturing and Intracorporeal Knot Tying

**I. PROCEDURE-SPECIFIC COMPONENTS (PSC):**

Steps for Laparoscopic Suturing & Intracorporeal Knot Tying		YES	NO
<b>A. Needle Delivery/Load</b> Start Time: _____ End Time: _____			
1. Needle delivered atraumatically to the field (not caught in trocar)		1	0
2. Load needle perpendicular to needle driver		1	0
3. Choke needle 1/2 to 2/3 from needle tip		1	0
<b>B. Suturing</b> Start Time: _____ End Time: _____			
4. Place needle at 90 degree angle to tissue		1	0
5. Drive needle with wrist supination		1	0
6. Pull suture through to establish short free end (≤ 1 inch tail)		1	0
7. Suture placed accurately, on target		1	0
<b>C. Knot tying</b> Start Time: _____ End Time: _____			
9. First throw: a) Surgeon's knot, ie. 2 throws in same direction		1	0
b) Knot laid flat without air knots		1	0
c) Short free end maintained		1	0
10. 2nd throw: a) Square knot, ie. opposite direction from prior throw		1	0
b) Knot laid flat without air knots		1	0
11. Third throw: a) Square knot, ie. opposite direction from prior throw		1	0
b) Knot laid flat without air knots		1	0
12. Appropriate tissue re-approximation without strangulation		1	0
13. Good use of both hands to facilitate knot tying		1	0
<b>D. Suture cut and removal</b> Start Time: _____ End Time: _____			
14. Needle cut from suture under direct visualization		1	0
15. Needle safely removed under direct visualization		1	0
<b>E. General</b>			
16. Kept needle in view at all times when grasping needle		1	0
17. Non-dominant hand helps dominant hand in suturing		1	0

Completed Task within 5 minutes: YES \_\_\_\_\_ NO \_\_\_\_\_ Total Score: \_\_\_\_\_ / 20 points total  
 Initial Start Time: \_\_\_\_\_ Final End Time: \_\_\_\_\_ Total Time: \_\_\_\_\_ (time limit 5 min)

**II. GLOBAL COMPONENT: Global Rating Checklist (GRC)**

Tissue and Instrument Handling				
1 Rough movements; awkward handling of instruments and tissue (or model)	2	3 Careful handling of instruments and tissue (or model) overall, with occasional awkward movements	4	5 Consistently appropriate and careful handling of instruments and tissue (or model)
Depth Perception/ Accuracy				
1 Constantly misses target, slow to correct	2	3 Sometimes misses target, quick to correct	4	5 Accurately directs instrument to target
Dexterity/ Efficiency				
1 Uncertain, inefficient movements without progress	2	3 Efficient movements overall with some unnecessary moves	4	5 Fluid, efficient movements without wasted time or motion
Autonomy (Proficiency)				
1 Unable to complete entire task at this time	2	3 Able to perform task safely with some instruction	4	5 Able to perform task safely and independently

Total Score: \_\_\_\_\_ / 20 points total For this task, this candidate should: Pass \_\_\_\_\_ Fail \_\_\_\_\_

**Figure 1.1: Sample OSATS form for trainees performing laparoscopic suturing and intracorporeal knot-tying tasks [8].**

In the majority of surgical training programs, surgical skills are assessed using reports called In-Training Evaluation Reports (ITERS). Like OSATS, these reports utilize structure checklist based reports that evaluate surgical trainees on technical and non-technical surgery skills[10]. Expert physicians rate trainees on categories such as medical expertise, communication, collaboration skills, managerial skills, and basic technical skills[10]. Furthermore, these summative reports must be updated and periodically reassessed to ensure that training residents receive regular feedback through their surgical rotations.

With the advent of laparoscopic surgery, specialized training and practice is required due to the increased difficulties associated with two-dimensional visualization, unnatural hand-eye coordination and significantly reduced haptic interactions. However, OSATS and ITERs were not specifically designed to assess laparoscopic based procedures. As a consequence, Global Operative Assessment of Laparoscopic Skills (GOALS) was developed as an intraoperative assessment tool for such procedures [2]. GOALS consists of a 5-item global rating scale that specifically assess laparoscopic procedure competency and case difficulty using a 10-item checklist with visual analogue scales [2]. GOALS addresses the drawbacks of previous assessment methods by specifically designing rating scales for laparoscopic procedures and providing specific and focused feedback based on procedure type [2].

By utilizing commonalities in GOALS and OSATS based methodologies, another evaluation tool, called Global Rating Index for Technical Skills (GRITS), was developed to generalized surgical technical skills that focus on general markers of technical skill instead of procedure specific ratings. Using categories such as respect for tissue, instrument handling / knowledge, flow of operation, depth perception, time and motion, and bimanual dexterity, expert physician will rate trainees using a 5-point Likert scales in these categories [3]. GRITS expand upon ITERs by specifically focusing on technical skills and delivering an averaged Likert score for each training resident. The result is a more comprehensive checklist based assessment tool than ITERs or GOALS that is specifically tailored trainee assessment regarding minimally invasive procedures. Sample checklist based assessment forms are shown in Figure 1.2.

<b>Respect for Tissue</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Frequent unnecessary force on tissues or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage	Consistently handled tissue appropriately with minimal damage to tissues	
<b>Time and Motion</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Many unnecessary moves		Efficient time/motion but some unnecessary moves	Clear economy of movement. Maximum efficiency	
<b>Instrument Handling/Knowledge</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Tentative/awkward moves or inappropriate use		Competent use of instruments, occasionally awkward	Fluid moves with instruments. No awkwardness	
<b>Flow of Operation</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Frequently stopped, seemed unsure of next move		Some forward planning, reasonable progression	Obviously planned course, effortless flow	
<b>Knowledge of Specific Procedure</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Deficient knowledge. Required specific instruction at most steps		Knew all important steps of operation	Demonstrated familiarity with all steps of operation	
<b>Use of Assistants (if applicable)</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Consistently placed assistants poorly or failed to use		Appropriate use of assistants most of the time	Strategically used assistants to best advantage at all times	
<b>Communication Skills</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Frequent problems working with team or fails to communicate		Appropriate communication with team most of the time	Co-ordinates surgical team in a superior manner	
<b>Depth Perception (Laparoscopic Procedures Only)</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Constantly overshoots, swings wide, slow correction		Some overshooting but quick to correct	Accurately directs instruments in correct plane	
<b>Bimanual Dexterity (Laparoscopic Procedures Only)</b>				
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Uses only one hand, poor coordination between hands		Uses both hands but does not optimize their interaction	Expertly uses both hands to provide optimal exposure	

Figure 1.2: Sample GRITS rating form to assess trainee laparoscopic technical skills [3].

### 1.1.2 Bimanual dexterity based assessment

Dexterity analysis allows for more objective skill assessment than direct methods by utilizing hand-tracking techniques to discriminate surgical skill. For example, the Advanced Dundee Endoscopic Psychomotor Tester (ADEPT) utilized infrared optical motor sensor placed on the surgeon's arm to track extrapolated positional data and compare the performance between novices and experts [11], [12]. Due to possible optical occlusion for the sensors resulting in omitted positional data, this method is not widely accepted in operating rooms. The Imperial College Surgical Assessment Device (ICSAD) utilizes electromagnetic markers placed on each hand of the surgeon performing open and laparoscopic surgical simulated tasks. ICSAD uses metrics such as total procedure time,

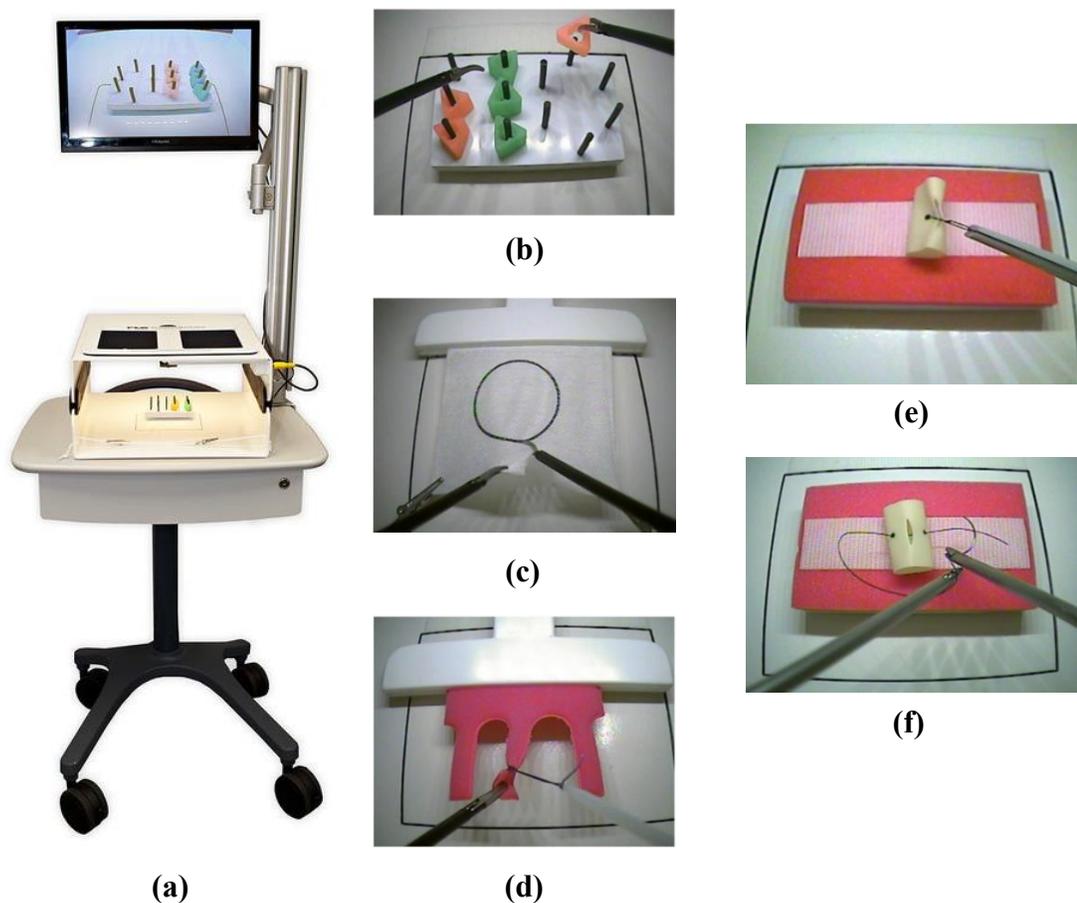
total travel path length, and speed of hand movements that have been used in several open surgical training tasks [12]. However, the ICSAD has not been clinically translated due to the infeasibility of attaching sensors to the surgeon's hands without impacting performance in the operating room [9]. Several approaches to modeling the kinematics of the surgical tools have also shown to differentiate motor skill between novices and experts as a means for surgical skill evaluation. These models often utilize Markov Models (MM) to discretize numerous surgical actions, such as grasping or cutting, into individual states. By analyzing tool kinematics for each state, these models have mapped tool behaviors for surgical experts and novices during basic surgical tasks with an accuracy of 87.5% [13]–[15]. However, these models require a significant amount of data set training to ensure an acceptable level accuracy and thus cannot be utilized in real time. While these methods have been published for use in surgical training environments, dexterity based approaches have not been directly implemented for real surgical procedures in the OR.

### **1.1.3 Physical simulation assessment**

To provide more objectivity and standardization for laparoscopic skills assessment, the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) was developed and validated as an effective simulator to teach and assess laparoscopic surgical skills [16]–[20]. The MISTELS system utilizes two 12-mm trocars where laparoscopic tools can be inserted into an enclosed box. Subjects can then manipulate and use these tools to practice surgical training tasks, such as pattern cutting, peg transfer, ligating loop, intra- and extracorporeal knot trying. These tasks are specifically designed to train important technical skills required for laparoscopic surgery. Beyond task training, the MISTELS system also offer objective scoring of each of the five training tasks, that are based on completion time and errors. These task assessment methods have been validated as effective in differentiation surgical skill between expert and trainee surgeons [16], [18], [20].

Upon rigorous validation and subsequent publications, a joint committee comprised of the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) and the American College of Surgeons (ACS) developed the MISTELS into a program called the Fundamentals of Laparoscopic Surgery (FLS). Consequently, the FLS program is now the

current standard is assessing proficiency in laparoscopic skills and is required for board certification since 2009[21]–[23] . FLS comprises of a didactic component, comprising of a web-based, multiple-choice examination, and a skill evaluation module, that includes hands-on skills testing for basic laparoscopic skills. An important component of the FLS program relies on a box trainer for training and measurements of technical skills during five basic laparoscopic surgical maneuvers. Surgical manual skills are evaluated via scoring the five skills tasks during timed trials with a maximum time limit. While the actual formulation for FLS scores are IP protected, studies have shown that FLS training is effective in teaching technical motor skills that are differentiable from untrained subjects and are retained for up to six months post-training [24]–[30].



**Figure 1.3: FLS training and assessment tasks that are required for board certification in general surgery. (a) The FLS box trainer trains and assesses laparoscopic training tasks such as (b) peg transfer, (c) pattern cutting, (d) ligation loop, (e) extracorporeal knot tying, and (f) intracorporeal knot trying [31].**

#### 1.1.4 Virtual reality (VR) simulation assessment

To address the general limitations of physical trainers, virtual reality based simulators have been developed and shown to provide a safe and effective training and assessment platform for laparoscopic surgical skills[24], [27]. As an analog to the MISTELS system, the minimally invasive surgical trainer-virtual reality (MIST-VR) system, is one of the first validated virtual simulator that trains and assesses laparoscopic skills *in silico*. By using metrics such as economy of motion, tool path length, instrument errors, task performance scores, and economy of time, studies have shown that MIST-VR is effective in training surgical technical skills specific to laparoscopic procedures [24], [29], [32], [33].

To specifically address the limitations of the FLS training simulator, we have developed the Virtual Basic Laparoscopic Skills Trainer (VBLaST) that is capable of simulating the five FLS task modules in real time[26], [34]–[37]. The benefits of the VBLaST system include automated and robust scoring, introduction of kinematic metrics that are correlated to task performance, dramatically increased objectivity in task performance assessment, and the elimination of high cost for administration or testing materials[26], [34]–[37]. As with any virtual reality based simulator, a thorough validation is required to demonstrate its effectiveness as a surgical training and performance assessment tool.

Commercial systems such as LapSim or Lap Mentor, are also high-fidelity VR systems that are capable of simulating, teaching, and assessing laparoscopic skills in real-time [38], [39]. These commercial systems train and assess laparoscopic motor skill using robust haptic feedback devices with moderate fidelity graphics. Significant research efforts have shown that commercial systems reduce operative time and increase trainee performance, and they have been shown to demonstrate positive correlation to operating room performance [24], [39]–[41]. Unfortunately, these commercial VR systems also present several drawbacks that include inaccurate force-feedback, high costs, limited realism that translates to the OR, and inconsistent results for skill transfer [38], [39]. Furthermore, these commercial systems report task performance scores based on simulated procedures, which may not be directly correlated to outcomes in clinical environments.

### **1.1.5 Robot assisted surgical skill assessment**

Robot assisted surgery (RAS) allows physicians to perform intricate and complex surgical procedure using sophisticated interactable robotic systems. Generally, these systems allow for increased precision, control, and flexibility than conventional methods in MIS. These trainers have seen an exponential rise in adoption within clinics geared towards numerous MIS based procedures. The top five VR simulators for RAS include the Surgical Education Platform (SEP), the dV Trainer, da Vinci Skills Simulator, the RobotiX, and the Robotic Surgical System [42]. While there is limited evidence that RAS trainers show specific increases in motor skill proficiency and conclusive surgical motor skill transfer, efforts to create a standardized training and assessment program called the Fundamentals of Robotic Surgery are already underway [42].

Many of these simulators utilize a scoring methodology to assess technical skill during training. Scoring metrics include, but not limited to, economy of motion, time of excessive force, instrument collisions, missed targets, economy of motion, mastery of workspace range, and task completion time [42]. However, these assessment methods are often proprietary and do not offer direct comparisons to outcomes in operative scenarios [42]. Beyond the impracticality of these systems due to high costs, many of these RAS simulators have inconsistencies in validation studies, no direct correlation of RAS training with positive operative outcomes and limited conclusive evidence of skill transfer from simulators to clinical environments [42].

### **1.1.6 Alternative surgical skill assessment**

Stress physiology is an important component of laparoscopic surgery where physicians must manage internal and external stimulus that evoke stress responses. In many cases, effective completion of surgical procedures with minimal negative patient outcomes while under operating room duress is a hallmark of surgical proficiency [43]. Recent reviews have indicated that stress physiology metrics, such as heart rate variability, average heart rate, self-reported stress metrics, and skin conductance changes, can differentiate surgical expertise between novices and experts [43]. These metrics are often used as the basis for crisis management training in teaching programs. While stress management is a crucial aspect of surgical training and carries through to proficiency,

accurate measurements of stress physiology is problematic. For example, many studies that utilize stress physiology metrics such as skin conductance, heart rate, or salivary production levels, assume that changes in these metrics are solely due to responses to stress stimuli. However, this may not be the case as subject variation and systemic physiology can dramatically impact these stress metrics, thus increasing the variability in reported data [43]. Furthermore, there is also a severe lack of critical evidence that suggests that changes in stress physiology are correlated to established surgical technical performance metrics [43].

Eye tracking has also been proposed for objective measures of surgical skill assessment. This methodology relies on the digital cameras or infrared optics to track pupil positioning in real time [44]. Beyond tracking the pupil centroids, other metrics such as pupil dilation, or fixation frequencies can also be tracked and have been utilized as measures of effort or concentration [44]–[46]. Note that eye tracking methodologies, while portable and very flexible for use in the operating room, have not shown rigorous validation for surgical attention or skill assessment.

### **1.1.7 Limitations of current assessment methods**

Several of the discussed tools for surgical technical skill assessment have been validated, where many are currently in use in medical institutions. However, each of the existing methods have major limitations that need to be addressed to have more objective and analytical methods of technical skill assessment that eventually correlate to positive outcomes in the OR [7], [47]–[51]. Table 1.1 summaries these methods along with their major limitations.

**Table 1.1: Summary of surgical technical skill assessment methods and major limitations.**

Assessment type	Assessment method	Assessment metrics	Major limitations
Global rating scales	OSATS	5-point Likert scale, general checklist based on proctor observations	<ul style="list-style-type: none"> <li>• Poor interrater reliability [7], [47]–[50]</li> <li>• Subjective metrics [7], [47]–[50]</li> <li>• High human resource cost [7], [47]–[50]</li> <li>• Not designed for laparoscopic procedures [2]</li> </ul>
	ITERS	Subjective reports, checklist based subsections	<ul style="list-style-type: none"> <li>• Distribution errors, recall bias, halo effects [2], [10]</li> <li>• Subjective metrics [7], [47]–[50]</li> <li>• High human resource cost [7], [47]–[50]</li> <li>• Not designed for laparoscopic procedures [2]</li> </ul>
	GOALS	Cumulative score based on 5-point Likert scale checklist	<ul style="list-style-type: none"> <li>• Subjective metrics [7], [47]–[50]</li> <li>• High human resource cost [7], [47]–[50]</li> </ul>
	GRITS	5-point Likert scale checklist	<ul style="list-style-type: none"> <li>• Halo effects [3]</li> <li>• Subjective metrics [7], [47]–[50]</li> <li>• Poor correlation to patient outcomes [7], [52]</li> </ul>
Dexterity analysis	ADEPT	Positional data metrics, hand velocity, hand acceleration	<ul style="list-style-type: none"> <li>• Impractical for OR usage due to optical occlusion [9]</li> <li>• Unimanual tracking only [9]</li> </ul>
	ICSAD	Hand positional data metrics, hand velocity, total task time, hand total path length	<ul style="list-style-type: none"> <li>• Impractical for OR usage due to motion artifacts and occluding wires [9]</li> </ul>
Physical simulators	MISTELS	Task performance scores, completion time, task errors	<ul style="list-style-type: none"> <li>• High cost for testing administration [18], [19], [34], [53]</li> <li>• Subjectivity in task assessment [18]–[20], [34], [53]</li> </ul>
	FLS	Proprietary task performance scores, task completion time, task errors	<ul style="list-style-type: none"> <li>• High cost for testing administration [18]–[20], [34], [53]</li> <li>• Subjectivity in task assessment [18]–[20], [34], [53]</li> <li>• Inconsistencies in FLS scoring methodology [18], [22], [34], [54]–[57]</li> <li>• No correlation of patient injury reduction due to FLS certification [18], [22], [34], [54]–[57]</li> </ul>
Virtual simulators	MIST-VR	Economy of time, task completion time, task performance scores,	<ul style="list-style-type: none"> <li>• Low fidelity virtual simulator [24], [27], [58]</li> <li>• No force sensory feedback [27]</li> <li>• Limited evidence of skill transfer to OR [9], [20], [24]</li> </ul>
	VBLaST	Kinematic metrics, force metrics, task performance scores, task completion time, task errors	<ul style="list-style-type: none"> <li>• Validation required for each new laparoscopic procedure</li> </ul>

## **1.2 Non-invasive brain imaging methods for motor skill assessment**

Now that current methods for surgical skill assessments have been identified, with associated limitations, one promising approach is to utilize brain imaging methods to quantify cortical changes as surgical motor skill increase. Several common brain imaging approaches are available with their inherent advantages and disadvantages and are addressed in this section. Direct methods of measuring functional activity include electroencephalography (EEG), and magnetoencephalography (MEG), while indirect methods include functional magnetic resonance imaging (fMRI), positron emitting tomography (PET), and functional near-infrared spectroscopy (fNIRS).

### **1.2.1 Electroencephalography (EEG)**

Perhaps the oldest neuroimaging technique, electroencephalography (EEG) records electrical current across cell membranes during a stimulus period. During a stimulus, neuronal depolarization with postsynaptic potentials (PSPs) occurs across the cellular membranes, which can be detected in an aggregate fashion in real time. EEG electrodes, which are generally composed on thin conductive discs, are commonly used in conjunction with the International 10-20 system of electrode placement for most functional connectivity studies [59]. Most EEG system utilize paired channels of up to 24 channels, although there are several systems that also utilize high electrode density approach that include 256 channels. Most cerebral signals measured via EEG fall within specific frequency ranges that are correlated to physiological responses. These frequency ranges include Delta, Theta, Alpha, and Beta waves. Delta waves (1-3 Hz) are normally apparent during sleep cycles, theta waves (4-7 Hz) are commonly observed in neonatal models or children, alpha waves (8-17 Hz) are observed during blinking or metal relaxation, and finally beta waves (12-30 Hz) are most notable during alert or anxious subjects [60], [61]. One of the prime benefits of EEG is the high temporal resolution during measurement acquisition. Signals can be acquired on the order of 100Hz and measure quick tasks or stimulus behavior, yet will have poor spatial resolution that can lead to lack of specificity in cortical activation.

### **1.2.2 Magnetoencephalography (MEG)**

Magnetoencephalography (MEG) is a direct brain imaging technique that measures the magnetic field changes due to increased electrical activity within cortical activity as a stimulus is presented [62]. While these magnetic fields are fairly weak, MEG can measure functional activation in the cortical sulci and gyri. MEG also is not hindered by superficial tissue such as scalp or skull matter since magnetic fields will not significantly distort as they propagate through these turbid media. These specific advantages allow for finer spatial resolution than EEG, although these increases are marginal. Unfortunately, there are several drawbacks that disable MEG as a practical means for non-invasive brain imaging. MEG imaging can only be implemented in magnetically shielded rooms that do not distort the magnetic field from outside sources. Also, MEG measurements are not specific enough for short neuronal stimuli and are often difficult to decouple from other neurological sources of electrical activity [60], [61].

### **1.2.3 Functional magnetic resonance imaging (fMRI)**

One of the most commonly used brain imaging techniques for a variety of applications is the functional magnetic resonance imaging (fMRI) technique. fMRI takes advantage of the paramagnetic properties of deoxygenated properties of hemoglobin (HHb) to quantify functional activation changes in cortical tissue. This technique is based on the theory that neural correlates to stimuli result in increased cerebral blood flow (CBF). [60], [61]. Thus, the resulting increase in oxygenated hemoglobin is also accompanied by a decrease in deoxygenated hemoglobin. Due to four unpaired electrons for each molecule, deoxygenated hemoglobin presents a significantly large magnetic susceptibility effect that disrupts the magnetic field produced by the MRI [63]. This effect is not observed for oxygenated hemoglobin, since  $\text{HbO}_2$  is diamagnetic. As a result, HHb changes can be quantified using the Blood Oxygen Level Dependent (BOLD) contrast under a high magnetic field. BOLD signals are reported as contrast-enhancing agent when presented with neuronal activation. In numerous fMRI based studies, the BOLD signal is considered a metric that quantifies activation responses due to neuronal activity.

Practically speaking, fMRI offers very high spatial resolution compared to most other forms of non-invasive brain imaging. Generally, there is a tradeoff between spatial

resolution and temporal resolution, since acquisition and processing time for smaller voxels takes significantly more time [61]. Many studies utilize conventional 1-3 Tesla scanners, that offer voxel sizes of approximately  $2\text{mm}^3$  with temporal resolutions on the order of a few seconds [60], [61]. These advantages are the primary reasons that many functional brain imaging studies utilize fMRI as the primary method for functional activation and connectivity measures.

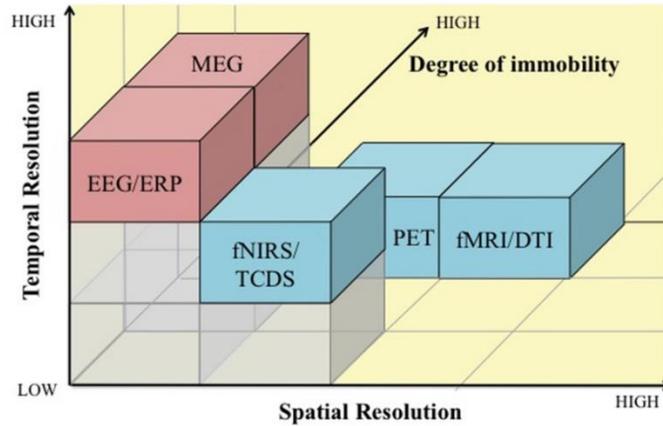
#### **1.2.4 Positron emission tomography (PET)**

Adopted well before MRI as a common practice for brain imaging methods, PET utilizes radiopharmaceutical agents that are tagged with a positron emitter. PET takes advantage of positron emitting radionuclides, that have very short half-lives, to map specific functional activation changes in the brain due to metabolic activity. These isotopes generally have very short half-lives and commonly consist of isotopes such as  $^{18}\text{F}$ ,  $^{15}\text{O}$ ,  $^{11}\text{C}$ , and  $^{13}\text{N}$  [61]. As functional activation increases during a stimulus onset, increased metabolic activity in cortical tissue increases the uptake of  $\text{O}_2$ , due to cellular glycolysis. These  $\text{O}_2$  molecules have radiotracers that are subsequently detected as tracer emitted positrons. This emitted positron will collide with nearby electrons to form a positronium, which eventually undergoes an annihilation process and emits two high energy photons with 511 KeV each [61]. A common technique of injecting  $^{15}\text{O}$  isotope labeled water, in the form of  $\text{H}_2^{15}\text{O}$ , directly into the bloodstream allows for PET to accurately measure regional changes in cerebral blood flow with spatial resolution on the order of 3-4mm [60], [61]. However, temporal resolution is significantly poor where each image scan and registration is on the order of 40 seconds [61].

#### **1.2.5 Limitation of non-invasive brain imaging methods**

Each of these common non-invasive brain imaging methods have inherent advantages and drawbacks that differentiate their applicability and features for various brain imaging studies. The most significant advantages and disadvantages of each method are listed in Table 1.2. A schematic detailing the spatial and temporal resolution trade-offs for brain imaging method is also shown in Figure 1.4. Note the inclusion of fNIRS for completeness

of comparisons. A more detailed review of fNIRS, however, is presented in the latter section 1.3.



**Figure 1.4:** Schematic outlining the temporal and spatial resolution estimates for common non-invasive brain imaging methods [64].

**Table 1.2:** Advantages and disadvantages of currently used functional brain imaging modalities [60], [61].

Modality	Advantages	Disadvantages
fMRI	<ul style="list-style-type: none"> <li>• Non-invasive</li> <li>• Sensitive to deep brain structures</li> <li>• High spatial resolution (~1mm)</li> </ul>	<ul style="list-style-type: none"> <li>• Low temporal resolution (0.1 Hz)</li> <li>• Confined to supine position</li> <li>• Sensitive to motion artifacts</li> <li>• Incompatible with metallic objects, particularly surgical tools</li> </ul>
EEG	<ul style="list-style-type: none"> <li>• High temporal resolution (10 – 250 Hz)</li> </ul>	<ul style="list-style-type: none"> <li>• Cumbersome gel application for adequate probe contact</li> <li>• Low spatial resolution (2-5 cm)</li> <li>• Low specificity to different cortical regions</li> <li>• Sensitive to motion artifacts</li> </ul>
MEG	<ul style="list-style-type: none"> <li>• Insensitive to superficial tissue (skull, scalp, dura, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>• Incompatible with metallic objects, particularly surgical tools</li> <li>• Highly sensitive to outside sources of magnetic fields</li> <li>• Low spatial resolution</li> <li>• Unspecific to transient cortical activation</li> </ul>
PET	<ul style="list-style-type: none"> <li>• Insensitive to motion artifacts</li> <li>• High spatial resolution (3-4mm)</li> </ul>	<ul style="list-style-type: none"> <li>• Significant radiation exposure</li> <li>• Short half-life of isotopes</li> <li>• Low temporal resolution (0.25 Hz)</li> <li>• Expensive</li> </ul>
fNIRS	<ul style="list-style-type: none"> <li>• Inexpensive, portable, and non-obtrusive to motor tasks</li> <li>• Insensitive to motion artifacts</li> <li>• High temporal resolution (100 Hz)</li> </ul>	<ul style="list-style-type: none"> <li>• Relatively low spatial resolution (~1 cm)</li> <li>• Spatial resolution dependent on probe design</li> <li>• Anatomical information cannot be directly measured</li> <li>• Sensitive to light absorbing media (hair)</li> </ul>

### 1.3 Functional near-infrared spectroscopy (fNIRS)

Functional near-infrared spectroscopy (fNIRS) is an indirect, non-invasive brain imaging method that measures cortical activation by detecting hemodynamic response changes due to a stimulus. Frans Jöbiss, widely considered the founding father of fNIRS, published the foundations for this concept by showing that near-infrared light at specific wavelengths can be attenuated and detected through turbid media [65]. Of course, several decades later, fNIRS instrumentation, post-processing, and applications have been significantly studied. However, our understating of bimanual skills and applications in surgical motor skills have been relatively untapped. This section is devoted to the fNIRS modeling, hardware and software methodologies, advantages and disadvantages of fNIRS along with a current literature review on fNIRS with specific applications to surgical skills.

#### 1.3.1 fNIRS modeling

To model near-infrared light propagation through turbid media, such as cortical tissue, we first start with the diffusion equation in highly turbid media as shown below [66]:

$$\nabla \cdot D(\mathbf{r})\nabla\Phi(\mathbf{r},t) - v\mu_a(\mathbf{r})\Phi(\mathbf{r},t) + vS(\mathbf{r},t) = \frac{\partial\Phi(\mathbf{r},t)}{\partial t} \quad (1-1)$$

, where  $\Phi(\mathbf{r},t)$  is the photon fluence rate with dimensions (photons / cm<sup>2</sup> \* s),  $v$  is the speed of light,  $S(\mathbf{r},t)$  is the isotropic source term that is the number of photons emitted at position  $\mathbf{r}$  and at time  $t$ ,  $\mu_a$  is the absorption coefficient,  $\mu_s$  is the scattering coefficient, and the photon diffusion coefficient  $D$ .

The diffusion coefficient  $D$  is defined [66]:

$$D = \frac{v}{3(\mu'_s + \mu_a)} \cong \frac{v}{3(\mu'_s)} \quad (1-2)$$

since  $\mu'_s \gg \mu_a$  is most tissue types [66]. The highly scattering media coefficient is defined below [66]:

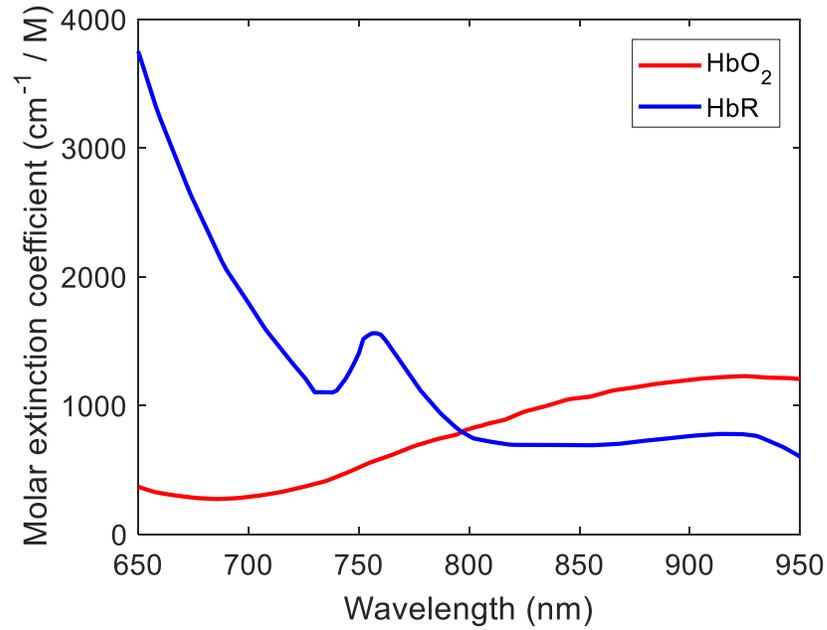
$$\mu'_s = (1 - g)\mu_s \quad (1-3)$$

where  $\mu_s$  is the inverse of the photon random walk pathlength  $l$ , and  $g = \langle \cos \theta \rangle$  where  $g$  is the ensemble averaged cosine of the scattering angle associated with single scattering events in the sample [66].

Note that absorption and scattering coefficient are wavelength dependent for each type of chromophore within the highly scattering media. For near-infrared wavelengths within the 650 - 950nm range, otherwise also known as the “optical window” [66], the primary chromophores are oxygenated and deoxygenated hemoglobin. With this knowledge, the wavelength dependent absorption coefficient can be defined below [66]:

$$\mu_a(\lambda) = \varepsilon_{HbO_2}(\lambda)[HbO_2] + \varepsilon_{HbR}(\lambda)[HbR] \quad (1-4)$$

where  $\lambda$  is the photon wavelength,  $\varepsilon$  is the extinction coefficient for HbO<sub>2</sub> and HbR that are wavelength dependent, and [HbO<sub>2</sub>] and [HbR] are the concentrations of oxygenated and deoxygenated hemoglobin, respectively. Molar coefficients for HbR and HbO<sub>2</sub> within the optical window are shown in Figure 1.5.



**Figure 1.5: Molar coefficients of HbO<sub>2</sub> and HbR within the optical window (redrawn from [67]).**

Now that we have defined our photon fluence equation, the change in optical density can be formulated using the modified Beer-Lambert law as shown below [66], [68], [69]:

$$\Delta OD(t, \lambda) = -\log\left(\frac{\Phi(t, \lambda)}{\Phi_0(t, \lambda)}\right) \Delta \mu_a(t, \lambda) L(\lambda) \quad (1-5)$$

where  $\Phi_0$  is the average photon fluence, L is the average path length of light through tissue. This equation can be further generalized with the following expression [70]:

$$\Delta OD(\Delta t, \lambda) = -\log_{10} \left( \frac{I(t_1, \lambda)}{I(t_0, \lambda)} \right) = \sum \varepsilon_i(\lambda) \Delta c_i DPF(\lambda) L \quad (1-6)$$

where  $I$  is the detected intensity of light with respect to the incident intensity of light  $I_0$ .  $\varepsilon$  is the extinction coefficient,  $c$  is the chromophore concentrations,  $L$  is the distance between the source and detector, and  $DPF$  is the differential path-length.

It is important to note that Equation (6) significantly underestimates the actual changes in  $HbO_2$  and  $HbR$  concentrations [66], [70] due to the assumption that the Beer-Lambert law is specific to homogenous tissue. Since cortical tissue is highly heterogeneous in nature, a correction can be made to the underestimating of hemoglobin concentration by incorporating the partial differential pathlengths [66], [70].

Finally, the change in optical density data can be converted to the hemoglobin concentration changes using the following expression [66], [68]–[70]:

$$\begin{bmatrix} \Delta[HHb] \\ \Delta[HbO_2] \end{bmatrix} = L^{-1} \begin{bmatrix} \varepsilon_{\lambda 1}^{HHb} & \varepsilon_{\lambda 1}^{HbO_2} \\ \varepsilon_{\lambda 2}^{HHb} & \varepsilon_{\lambda 2}^{HbO_2} \end{bmatrix}^{-1} \begin{bmatrix} \Delta OD(\Delta t, \lambda_1) / DPF(\lambda_1) \\ \Delta OD(\Delta t, \lambda_2) / DPF(\lambda_2) \end{bmatrix} \quad (1-7)$$

where the final oxy and deoxy hemoglobin concentration changes with respect to time can be calculated in the cortex due to a stimulus response. This formulation is used to create the hemodynamic response function (HRF) that is used for reporting of real-time hemoglobin concentrations changes in fNIRS studies.

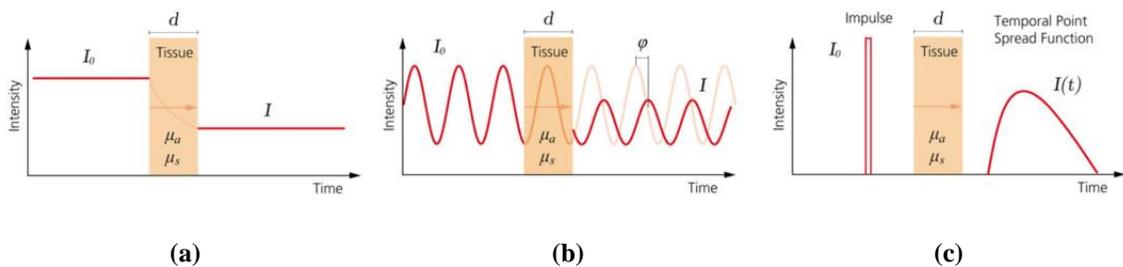
### 1.3.2 fNIRS hardware

fNIRS imaging modalities are split into three categories: continuous wave, frequency domain, and time domain systems. Continuous wave (CW) systems are only dependent on light intensity measurements, where near-infrared light at defined wavelengths and intensity are delivered to turbid media and the backscattered light is subsequently measured [70]. The primary disadvantage of CW systems is the inability to determine absolute concentrations of chromophore concentrations, since absorption or scattering coefficients cannot be absolutely measured. With most CW system, a source and detector channel distance of  $>2.5\text{cm}$  is generally recommended to ensure a constant  $DPF$  in tissue types [71]. While CW systems are versatile and robust for human imaging studies, this

methodology also assumes that coupling and light scattering effects are constant over time, and that hemoglobin absorption changes are solely due to blood [71].

Frequency domain (FD) systems utilize a continuous laser source at specific frequencies while modulating this source to record the relative phase shift of the entering and exiting light [71]. Using this approach, it is possible to separate absorption and scattering coefficients for an assessment of not only chromophore concentration but also lifetime measurements. This leads to improved spatial resolution than CW systems [71]. Since FD based systems require modulated lasers along with phase measurements of light exiting from turbid media, this approach is generally more expensive than CW systems. The main advantages, however, are increased spatial resolution and specificity to absolute concentrations of hemoglobin concentrations [66], [71].

Time domain devices utilize short pulsed laser sources (generally picosecond resolution) to detect photon intensity and ultimately calculate the delay between pulse emission and reception [71]. Time domain systems often contain the most information rich datasets that can also subsequently be used to create CW or FD datasets. Furthermore, the spatial resolution of time domain systems is significantly higher than CW or FD systems, but is accompanied with drawbacks including high instrumentation costs, bulky systems, and lower temporal resolutions. [70], [71]. Figure 1.6 shows the infrared light delivery and measurement approaches for CW, TD, and FD fNIRS systems. This work utilizes a 32-channel continuous-wave near-infrared spectrometer developed by TechEn Inc. (MA, USA) due to the versatility geared towards human subject research.



**Figure 1.6: Different modalities of fNIRS imaging system that include (a) continuous-wave, (b) frequency domain, and (c) time domain systems [70].**

### 1.3.3 fNIRS in bimanual surgical motor skills

While relatively new, the concept of utilizing fNIRS for brain imaging applications in surgical skill assessment has been broached by several groups. The task paradigm, subject selection, and cortical region of interests are varied for these studies but have commonalities in specific findings. For example, in general, the prefrontal cortex activation is significantly higher in novices or trainees compared to experts in most fNIRS based studies. A summary of fNIRS specific studies on surgical motor skill is shown in Table 1.3.

**Table 1.3: Summary of fNIRS studies to assess surgical motor skill.**

Author	Task	Neurological findings correlated to surgical performance
Leff <i>et al</i> (2008) [72]	Open surgical knot tying	<ul style="list-style-type: none"> <li>• PFC activation: Novice have higher activation than experts, trainees have significantly different activation than experts</li> <li>• Significant changes in PFC activation with skills training</li> </ul>
Ohuchida <i>et al</i> (2009)[73]	Laparoscopic knot tying	<ul style="list-style-type: none"> <li>• Significant increases in PFC activation for trainees compared to experts</li> <li>• Significant increases in PFC activation for novices after 2-hour training</li> </ul>
James <i>et al</i> (2011)[74]	Endoscopic camera navigation task	<ul style="list-style-type: none"> <li>• High lateral PFC activation in novice and expert groups.</li> <li>• Significantly higher lateral PFC activation in specific channels for Experts</li> </ul>
Crewther <i>et al</i> (2016)[75]	Laparoscopic knot tying	<ul style="list-style-type: none"> <li>• Significantly higher PFC activation in novices compared to experts</li> <li>• Changes in PFC activation reverted to baseline after 8-hour period of training</li> </ul>
Andreu-Perez <i>et al</i> (2016)[76]	Laparoscopic knot tying	<ul style="list-style-type: none"> <li>• Significantly higher PFC and premotor activation in novices compared to experts</li> </ul>
Shewokis <i>et al</i> (2017)[77]	Laparoscopic cholecystectomy in VR environment	<ul style="list-style-type: none"> <li>• Significantly lower PFC activation with increased training</li> </ul>

While these studies have broached the concept of fNIRS usage for surgical motor skill differentiation, they present several limitations. These limitations are outlined below:

- Lack of primary motor cortex and supplementary motor area measurements for objective surgical skill differentiation [60]. The implications of these cortical regions are further discussed in section 1.4.
- Lack of fNIRS studies that measure functional activation changes during surgical skill transfer from simulation to clinically relevant environments
- Lack of studies that compared fNIRS metrics for surgical skill assessment with established assessment metrics [60], [78].
- Grossly overestimated hemodynamic responses due to inclusions of superficial tissue signals stemming from scalp, skull, dura, and pial matter [79]–[81].
- Lack of functional connectivity studies that correlate surgical skill learning with increased connectivity in specific cortical regions.

This work aims to comprehensively address these limitations in current fNIRS studies for surgical skill assessment.

## **1.4 Neurophysiology of human motor skill learning and retention**

Now that fNIRS has been identified as a dominant imaging technique with specific benefits for real-time measurements of cortical activity during a complex bimanual surgical task, we need to identify the regions of interest for cerebral measurements. To determine this, a firm understanding of motor control physiology and the implications of various cortical regions on motor skill learning and execution is needed. This section is overview of motor skill physiology, relevant anatomy, and mechanisms for motor skill learning and retention.

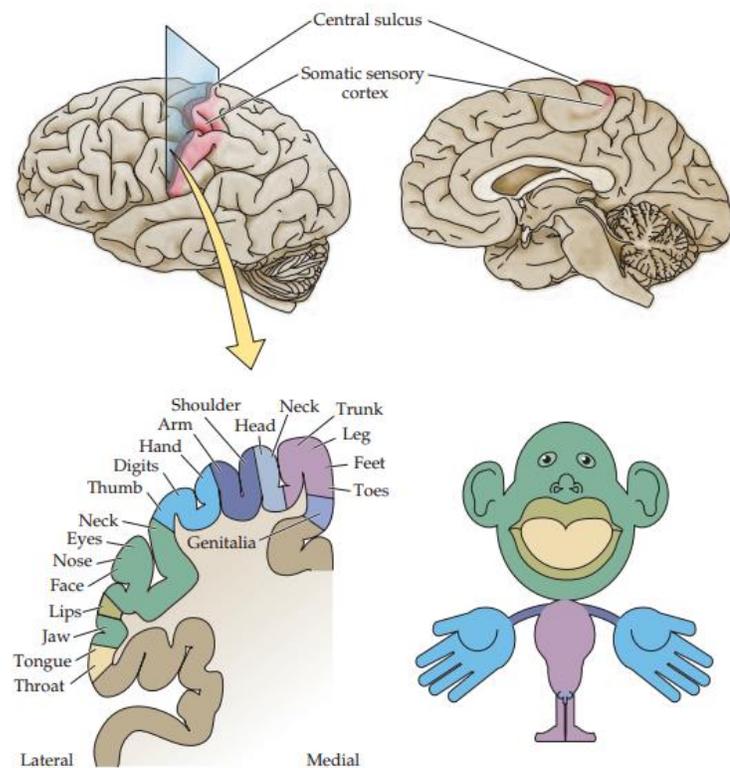
### **1.4.1 Motor control physiology**

Human motor control involves several cortical regions that function dependently with each other for voluntary motor function. The fundamental mechanism is where neuronal axons from the cortical regions and brain stem will descend through specific motor pathways to control specific muscle motor units. Five cortical regions, the prefrontal

cortex (PFC), primary motor cortex (M1), supplementary motor area (SMA), premotor area, thalamus, and the basal ganglia will be discussed in this section regarding their specific implications for motor control [82].

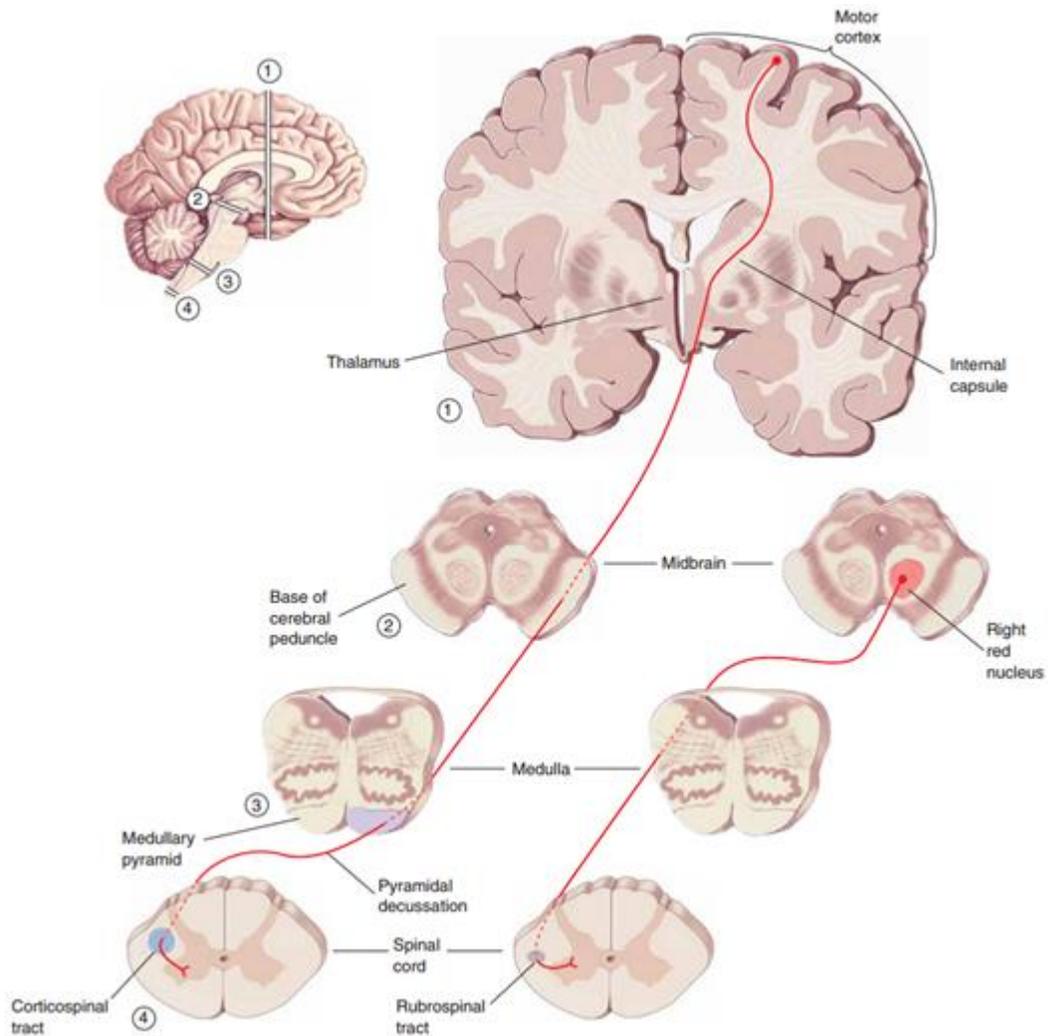
The first region, located in the frontal lobe, is the prefrontal cortex. This region is located anterior to the premotor cortex and includes a number of substructures such as the granular frontal area 9, frontopolar area, Brodmann’s prefrontal area 11 [83], [84]. The PFC is directly involved with other cortical lobes such as the temporal or parietal due to its role in high order cognitive functions, such as intellectual, judgmental, and predicative functions [60].

Located anterior to the central sulcus is the precentral gyrus, which is also called the primary motor cortex (M1). The precentral gyrus can also be represented using a motor homunculus. This visual depiction shows that the precentral gyrus, which comprises of the M1, is specific to degree of precision instead of the size of each body part. Figure 1.7 shows the somatotopic representation, also known as the motor homunculus, for one hemisphere of the central sulcus.



**Figure 1.7: Somatotopic representation of body parts from medial to lateral direction in the precentral gyrus[84].**

The primary function of the M1 is voluntary control of motor movements. The primary mechanism of motor control is where neuronal axons travel down to synapses directly attached to motor neurons in muscle motor units. The pathway for the axons is called the corticospinal tract where axons descend from the cerebral hemisphere by passing through the corona radiata and into the crus cerebri of the midbrain. These groups of axons, called corticospinal fibers, pass through the pons onto the ventral surface of the medulla oblongata. These corticospinal fibers decussate in the caudal medulla before terminating in several ventral horns of the spinal cord. Upon fiber termination, monosynaptic contacts are made with motor neurons associated with each limb. Figure 1.9 shows the corticospinal tracts of axonal fiber pathways for motor skill execution. Note that a significant amount of fibers will decussate upon leaving the medulla, thus indicating that motor function is associated with contralateral activation in the cerebrum.



**Figure 1.8: Corticospinal tract diagram illustrating specific pathway of motor neurons from the cerebrum to muscle motor units [83].**

Located on the medial surface of the premotor cortex, the supplementary motor area (SMA) is another region that is heavily involved in the voluntary control of motor movements. Neurons in this region are correlated to the selection of specific movements based on the task at hand. The SMA is mainly responsible for functions such as postural stability, bimanual coordination, and the initiation of motor movements [83], [84]. However, the specific pathways of how the SMA is involved for these various functions is still largely unknown.

The thalamus is almost completely surrounded by the cerebral hemisphere and primarily acts as a relay of information for neural fibers from the basal ganglia and the cerebellum during motor control. Comprised of the ventral anterior and ventral lateral

nuclei, the thalamus is an important region as it directly influences normal movement control with specific neuronal pathways to the M1, PFC, and SMA [60], [83], [84]. More specifically, the ventral lateral nucleus also has connections to the M1 via regions in the Thalamus called the pars oralis (VLo), pars medialis (VLm) and the pars caudalis (VLc) [83].

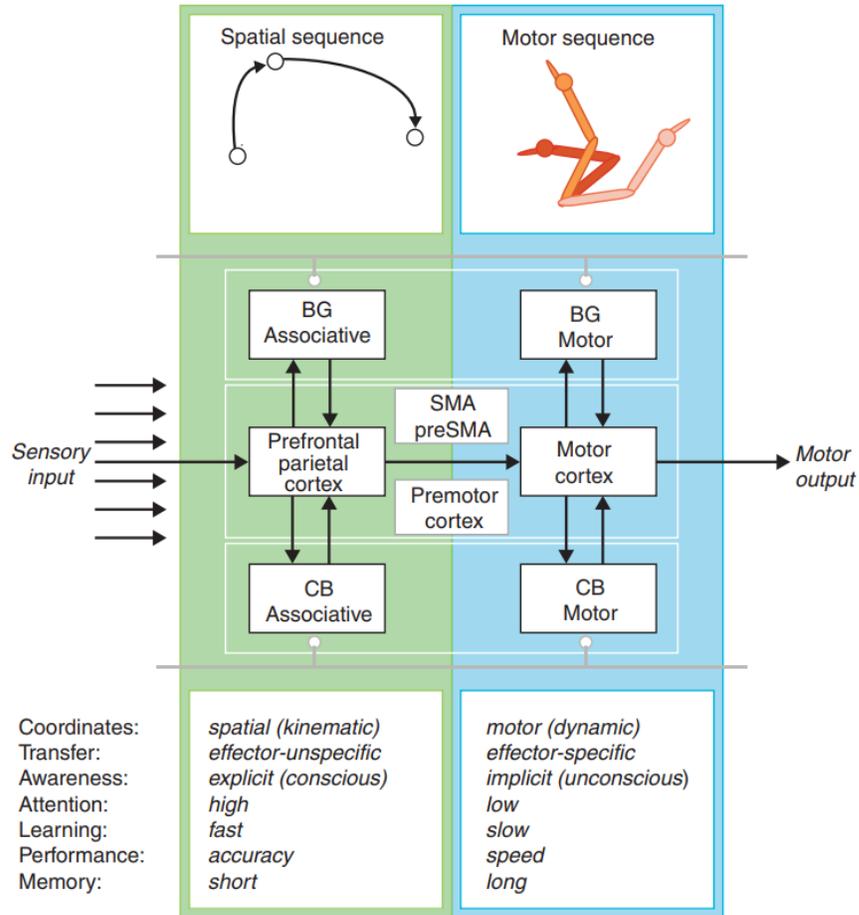
The basal ganglia lies deep within the cerebral hemisphere and is mainly comprised of the caudate nucleus, putamen, and the globus pallidus. These three regions combined are generally called the corpus striatum [83], [84]. Functionally, the basal ganglia facilitates specific movements along with strict suppression of unwanted movements. For example, when the M1 region initiates neuronal discharge through the corticospinal tract, these fibers will pass through the neostriatum. The neostriatum has a direct pathway, that results in overall excitatory effects of motor neurons stemming from the M1, along with an indirect pathway, which leads through the subthalamic nuclei to suppress unwanted movements that may hinder specific fine movements initiated by the M1 [83].

#### **1.4.2 Motor skill learning**

Motor skill learning follows a complex neurophysiological pathway that has only recently been understood with the advent of brain imaging techniques. Since a complex motor skill is often composed of multiple sequences of movements [85], to truly understand the mechanism of motor skill learning, the fundamentals of how motor sequences are acquired must be understood. However, how exactly are such motor sequences acquired for long term memory? Precisely designed to address this question, the 2x5 task has been used in many studies that allows the subject (human or primate) to learn to press buttons in the cored order via trial and error [86]. Results have shown that the presupplementary motor is activated during the learning of new sequences, but not for the performance of these learned sequences [82], [87]. Furthermore, deficits in new motor sequence learning have also been shown with the functional blockage of the SMA [82], [88]. Other areas of the motor cortex such as the anterior cingulate cortex ventral to the SMA and the primary motor cortex (M1) not only contribute to motor learning but undergo functional and structural changes [89]–[92].

Functional imaging studies involving the 2x5 task and fMRI have also shown corroborative evidence that SMA and the PFC are associated with the early stages of learning. A study by Sakai *et al.* showed that dorsolateral PFC and the SMA were activated during the early stages of motor skill learning whereas the parietal areas (intraparietal sulcus and the precuneus) were activated at later stages [93]. Another study showed results that the awareness of performance, also called explicit learning, induces prefrontal cortex and preSMA activation but not the sensorimotor cortex [94]. Several behavioral studies also show that different parts of the brain account for different aspects of motor skill acquisition. For example, the accuracy of motor performance is acquired before speed [86]. Accuracy has also been shown to be effector unspecific where speed is effector specific [82], [95], [96]. Once the learning period has been established, the motor skill is maintained for a long duration in the form of speed [86].

In addition to the preSMA, dorsolateral prefrontal cortex, and the primary motor cortex, there are other structures within the brain that are crucial for skill acquisition and motor learning. Functional studies have shown that the basal ganglia (BG) and the cerebellum (CB) are significantly involved in motor sequence learning [97], [98]. For example, the group activity of striatal neurons changes in the BG with long-term motor learning [99]. Furthermore, depletions in dopamine and blockages of the posterior striatum lead to performance deficiencies of motor skills, whereas the anterior striatum of the BG leads to deficiencies in learning new motor skills [99]–[101]. Like the BG, the cerebellum (CB) is also involved in the learning of new motor skills. Studies involving specific lesions in the cerebellum indicate that it impairs motor sequence learning but not conditional visuomotor learning or spatial working memory [82], [102]. Studies have shown that the blockade of the dorsal part of the dentate nucleus, which is part of M1, does not affect learning new sequences but disrupts skill performance [103]. Other studies however have shown that the long term motor skill memories are stored in the CB [104], indicating there is inconsistency within the literature on the role of CB for motor skill learning. Figure 1.9 below is a model developed by Hikosaka *et al.* illustrating the pathway for motor skill learning regarding spatial and motor sequences.

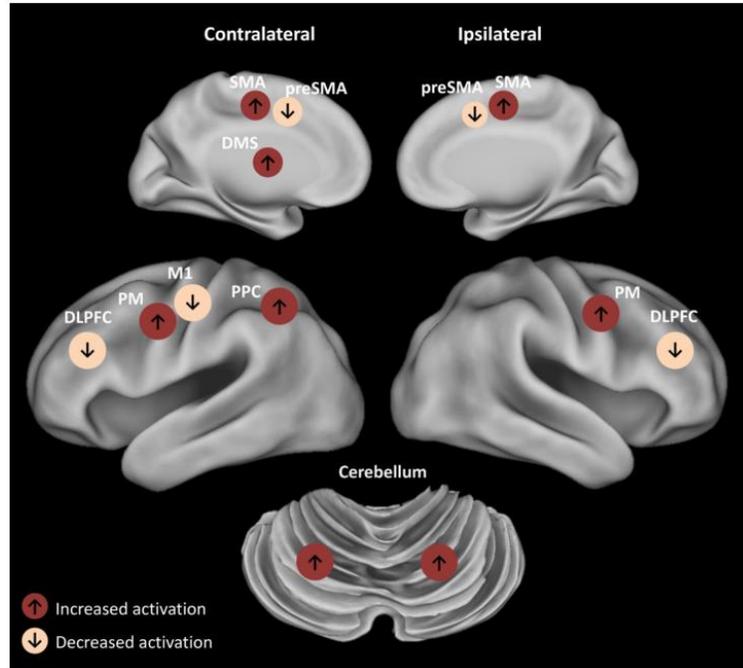


**Figure 1.9: Schematic illustrating the motor skill learning pathway involving spatial and motor sequences [82].**

With the advent of modern medical imaging, our understanding of complex motor learning pathways, consolidation, and retrieval have been vastly increased. For example, the neural substrates of learning stages were studied using positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). Specific learning paradigms, such as fast learning of sequential motor tasks, were shown to modulate regional brain activity in the dorsolateral prefrontal cortex, primary motor cortex (M1) and the preSMA [105], [106]. Furthermore, these studies showed decreased activation as learning progresses in the premotor cortex, SMA, parietal regions, striatum and the cerebellum. Conversely, all of these areas showed increased activity during motor task learning via PET studies [94], [105], [107]. Thus, learning is associated with differential modulation of blood oxygenation level-dependent (BOLD) activity.

Another key region for fast learning, as stated before, is the M1. Numerous functional studies have been conducted to implicate the role of M1 during the initial stages of learning. For example, Costa *et al*, used behavioral mice models to study long term potentiation and depression using a rotarod task. They were able to show the substantial recruitment of neurons in M1 during the initial motor skill learning of the rotarod task [108]. Similarly, PET and fMRI studies have shown the same consistency in humans where the learning a motor task modulates long-term potentiation plasticity in the M1 region [109], [110].

Slow motor skill learning involves quantitatively smaller changes than those during fast learning, and often develop at a much slower pace [111], [112]. These differ significantly from fast motor skills since they often reach a plateau of performance during learning. For example, this can be seen when an artist plays musical pieces on a violin, or a surgeon performs many procedures. Complex tasks such as these often take several instances of practice and even years to master. To corroborate this statement, studies have used fMRI to show functional brain activation shifts from anterior to posterior regions of the brain when progressing from early to late stages of motor skill [105]. Furthermore, more functional imaging studies have shown that the progression from fast to slow motor skill learning is also associated with a shift in activation from the associative to sensorimotor striatum [113], [114]. Slow learning has also been shown to elicit large functional reorganization in the brain. For example, a motor task was used for subjects over a 4-week period. fMRI results indicated that decreased integration in the premotor-associative striatum cerebellar network but high connectivity within the M1-sensorimotor striatum network. [105], [114]. Large plastic changes in the M1 function due to slow motor learning have been well studied in animal models. For example, reorganization of movement based motor learning have been documented in squirrel monkeys due to continued slow learning of tasks with substantial practice periods [115], [116]. However, similar studies are minimal for human based models. Specifically, the extent to which changes in motor maps due to slow motor skill training in humans is still not understood [112]. A schematic illustrating the functional activation responses due to slow motor learning in specific cortical regions is shown in Figure 1.10 below.



**Figure 1.10:** Schematic illustrating neural substrates and their activations during slow motor learning [112].

### 1.4.3 Motor skill retention

Motor memory retention, or consolidation, is defined as a set of processes that are post-acquisition and time dependent when a motor memory becomes more stages with time [117]–[119]. These processes of consolidation may be behaviorally evident as an improvement in performance between practice sessions for the given motor task [120], [121]. Studies have shown that the importance of memory consolidation is directly correlated to perturbations in the consolidation processes, and subsequently the retention of motor skill [122]–[124]. For example, low frequency repetitive transcranial magnetic stimulation (rTMS) over the primary motor cortex M1 immediately following practice of a reaction time task blocked any offline improvements in the skill [124]. This indicates the implication of the M1 cortex as an intermediary for the long-term retention of motor skill. Disruptive effects of rTMS on learning were only present when stimulation is applied immediately after practice and delayed by hours [124]. Ultimately, this indicates that there is active post-practice consolidation process that is responsible for offline learning, which results in the attainment of long term motor skill mastery. Compounding the consolidation process directly after practice, evidence suggests that motor memory also undergoes

consolidation over the sleep period [125], [126]. During sleep, consolidation occurs during repeated cycles of non-REM sleep followed by REM sleep [127].

Memory retrieval is a key process in the overall retention of skill mastery. However, few studies have assessed the retrieval process for motor skill acquisition. Specific to motor skill, evidence exists to suggest that practice conditions that promote cognitive processes are required for retrieval of skills, often to enhance retention performance [117]. Another study for motor skill transfer and retrieval was conducted by Lin *et al* [128]. They investigated functional brain activation with fMRI during the retention of three finger sequences that were practiced in blocked or random orders. During acquisition the random-order practice was associated with longer response times. However, during motor skill retention, the random-order finger sequences had shorter response times, which is essentially higher performance. Since higher cognitive demands during random practice invoke premotor and prefrontal network activation, these networks became more efficient in the cognitive processing of information, as shown by reduced BOLD signals despite superior performance via MRI [117], [128].

#### **1.4.4 Gaps in bimanual motor skill neurophysiology knowledge**

While tremendous research efforts have shown significant insights into the underlying neural mechanisms for motor skill control and learning, several gaps still exist in current literature for a comprehensive understanding of motor skill learning and execution mechanisms, particularly in complex bimanual motor skills.

Majority of studies that address neurophysiological mechanisms for motor skill levels utilize motor movements that depend on event timing, such as finger tapping, or visual cues for finger flexion tasks [129]. Evoked neural correlates from these unimanual based tasks are often the basis of reporting fundamental functional changes with respect to motor skill learning and do not address the continuous movements often present in bimanual task [129]. Another limitation is that majority of motor skill learning studies utilize tasks that isolate specific movement frequencies. These movements are very transient in nature, focus on hand dominance, and focus on specific event marked movements. Complex surgical tasks, however, are often comprised of continuous hand movements, asynchronous motor behavior, and are largely unstudied in motor skill learning studies.

Another major limitation is that majority of motor skill neurophysiology studies do not specifically distinguish between motor skill learning and performance [129]. For example, many studies have focused on using EEG or MRI based approaches on the initial stages of motor learning and do not specifically resolve the transition states from early to late stages of motor skill learning [129]. This can be addressed with studies that specifically measure neural correlates for motor skill between experts, novices, and subjects that transition from early to late stages of motor learning. Such studies would comprehensively assess neural correlates for bimanual task production as motor skill learning stages transition and ultimately provide a deeper understanding of motor skill performance [129].

## 1.5 Specific aims

Based on the discussions above, fNIRS was chosen as the noninvasive imaging modality for assessing bimanual fine motor skills that underpin surgical expertise. However, the existing literature in this field point to the following major gaps:

- Lack of brain imaging studies that differentiate and classify surgical motor skill levels with direct comparisons to established skill assessment metrics
- Lack of brain imaging studies that differentiate and classify surgical motor skill transfer from simulation to clinically relevant environments
- Lack of brain imaging studies that quantify functional connectivity changes with varying degrees of surgical motor skills

***Aim 1* is to establish fNIRS-based metrics to objectively differentiate and classify surgical motor skill proficiency.** The *milestone* for Aim 1 is when fNIRS is more accurate and specific in surgical skill assessment than established measures, which are based on performance time and errors.

***Aim 2* is to establish metrics to objectively differentiate and classify surgical motor skill transfer between trained and untrained subjects based on fNIRS.** The *milestone* for Aim 2 is when fNIRS is more accurate and specific in surgical skill transfer assessment than established measures, which are based on performance time alone.

***Aim 3 is to identify cortical regions within the PFC, M1, and SMA that have significantly high functional connectivity correlations to surgical motor skill proficiency, using established measures of functional connectivity.*** The *milestone* for Aim3 is met when functional connectivity metrics between specific combinations of PFC, M1, and SMA cortical regions is higher than 0.5 and are significantly higher for subjects with increased surgical motor skill proficiency regardless of simulation platform.

## **1.6 Research plan and thesis outline**

To accomplish Aim 1, subjects with varying degrees of surgical motor skills were recruited to perform complex, bimanual surgical training tasks while undergoing fNIRS imaging. Multivariate analysis was used to classify subjects based on surgical motor skill proficiency and directly compared to established assessment metrics.

For Aim 2, surgical simulator trained and untrained subjects from Aim 1 were asked to perform a surgical transfer task on cadaveric tissue to represent surgical skill transfer. The resulting multivariate fNIRS metrics was used to classify successful and unsuccessful surgical skill transfer subjects and directly compared to established transfer assessment metrics.

For Aim 3, functional connectivity metrics will be derived from cortical activations in in the PFC, M1, and SMA for all subjects from Aim 1 and 2. These resulting metrics will be compared between inter and intra subject to establish correlation of functional connectivity to surgical motor skill.

The remainder of this thesis is organized as follows. Chapter 2 focuses on examining Aim 1 by proposing fNIRS as a mean for accurate differentiation and classification of surgical skill levels. Chapter 3 focuses on addressing Aim 2 to establish fNIRS metrics as accurate classification of trained and untrained subjects for surgical motor skill transfer. Chapter 4 focuses on addressing Aim 3 to examine specific cortical regions that exhibit high functional connectivity, and are also correlated to increased levels of surgical motor skills. Finally, thesis summary along with potential areas for future research are presented in Chapter 5.

## 2. Assessing bimanual motor skills with optical neuroimaging

### 2.1 Overview

Motor skills that involve bimanual motor coordination are essential in performing numerous tasks ranging from simple daily activities to complex motor actions performed by highly skilled individuals. Hence, metrics to assess motor task performance are critical in numerous fields including neuropathology and neurological recovery, surgical training and certification, and athletic performance [49], [130]–[135]. In the vast majority of fields, however, current metrics are human-administered, subjective, and require significant personnel resources and time. Thus, there is critical need for more automated, analytical, and objective evaluation methods [7], [47]–[50]. From a neuroscience perspective, bimanual task assessment provides insights into motor skill expertise, motor dysfunctions, interconnectivity between brain regions, and higher cognitive and executive functions, such as motor perception, motor action, and task multitasking [135], [136]. Therefore, incorporating the underlying neurological responses in bimanual motor skill assessment is a logical step towards providing robust, objective metrics, which ultimately may lead to greatly improving our understanding of motor skill processes and facilitating bimanual-based task certification.

Among all non-invasive functional brain imaging techniques, functional near infrared spectroscopy (fNIRS) offers the unique ability to monitor and quantify fast functional brain activations over numerous cortical areas without constraining and interfering with bimanual task execution. Hence, fNIRS is a promising neuroimaging modality to study cortical brain activations but to date, only a very limited number of studies have been reported in regards to assessing fine surgical motor skills [60].

---

Portions of this chapter previously appeared as:

A. Nemani, W. Ahn, C. Cooper, S. Schwaitzberg, and S. De, “Convergent validation and transfer of learning studies of a virtual reality-based pattern cutting simulator,” *Surg. Endosc.*, to be published. doi: 10.1007/s00464-017-5802-8.

A. Nemani, X. Intes, and S. De, “Monte Carlo based simulation of sensitivity curvature for evaluating optimal probe geometry,” in *Biomedical Optics 2014*, pp. 3-36.

A. Nemani *et al.*, “Assessing bimanual motor skills with optical neuroimaging,” submitted for publication.

These exploratory studies have reported differentiation in functional cortical activations between groups with varying surgical motor skills [60], [72], [73], [137], [138]. However, they suffer from recognized limitations [60] such as such as the lack of signal specificity between scalp and cortical hemodynamics [80], [81], the lack of multivariate statistical approaches that leverage changes in functional brain activity across multiple brain regions, and benchmarking against established metrics. Hence, they have not impacted current practice of professional bimanual skill proficiency assessment. Here, we present a fNIRS-based optical neuroimaging methodology that overcomes all these shortcomings at once. For the first time in the field, we measure concurrently functional activations in the prefrontal cortex (PFC), the primary motor cortex (M1), and the supplementary motor area (SMA) to map the distributed brain functions associated with motor task strategy, motor task planning, and fine motor control in complex bimanual tasks[82], [87], [92]–[94], [139], [140]. Moreover, we increase the specificity of optical measurements to cortical tissue hemodynamics by regressing signals from scalp tissues<sup>31,35,36</sup>. Furthermore, we leverage changes in intraregional activation and interregional coupling of cerebral regions via multivariate statistical approaches to classify subjects according to motor skill levels. Finally, we compare our fNIRS based approach with currently employed metrics in surgical certification by assessing bimanual motor tasks that are a part of surgical training accreditation.

The performance of the reported optical neuroimaging methodology enables the objective assessment of complex bimanual motor skills as seen in laparoscopic surgery. Indeed, imaging distributed task-based functional responses demonstrated significant cortical activation differences between subjects with varying surgical expertise. By leveraging connected cerebral regions correlated to fine motor skills, we report increased specificity in discriminating surgical motor skills via fNIRS based metrics. For the first time, we show that our approach is significantly more accurate than currently established metrics employed for certification in general surgery, as reported via estimated misclassification errors. These results demonstrate that the combination of advanced fNIRS imaging with multivariate statistical approaches offers a practical and quantitative method to assess complex bimanual tasks. Topically, the reported optical neuroimaging

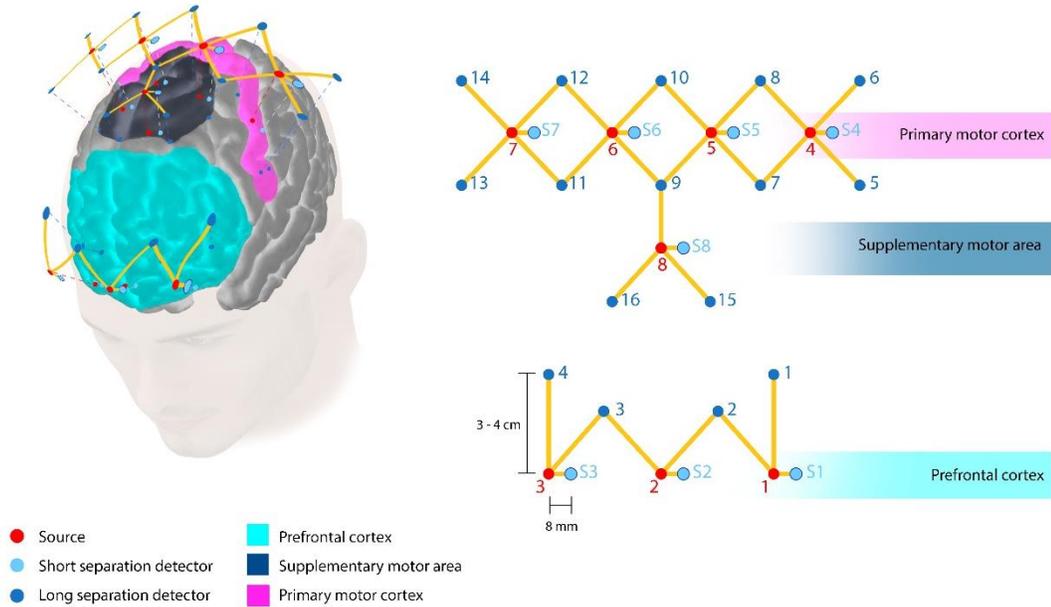
methodology is well suited to provide quantitative and standardized metrics for bimanual skill-based professional certifications.

## **2.2 Methods**

The study was approved by the Institutional Review Board of Massachusetts General Hospital, University of Buffalo, and Rensselaer Polytechnic Institute.

### **2.2.1 Hardware and equipment**

We utilize a validated continuous-wave, 32-channel near- infrared spectrometer for this study, which delivered infrared light at 690nm and 830nm (CW6 system, TechEn Inc., MA, USA). The system employed eight long distance and eight short distance illumination fibers coupled to 16 detectors. The long-distance channels comprised all the measurements within a 30 – 40mm distance between the source and the detector, and the short distance channels comprised all the measurements within a ~8mm distance between the source and the detector. The short channels are limited to probing the superficial tissue layers, such as skin, bone, dura and pial surfaces, whereas the long channel probed both superficial layers and cortical surface. The probe design was assessed using Monte Carlo simulations and was characterized to have high sensitivity to functional changes in the PFC, M1, and SMA. A schematic of the geometric arrangement of probes are shown in Figure 2.2.



**Figure 2.1: Infrared probe geometry positioning.** Schematic of probe placement projected on cortical locations specific to the PFC, M1, and SMA. Optodes are placed for maximum coverage over the PFC, M1, and SMA. Red dots indicate infrared sources, blue dots indicate long separation detectors, and textured blue dots indicate short separation detectors. The PFC has three sources (1-3), the M1 has four sources (4-7), and the SMA has one source (8). Each of the sources are connected to their corresponding long and short separation detectors [54].

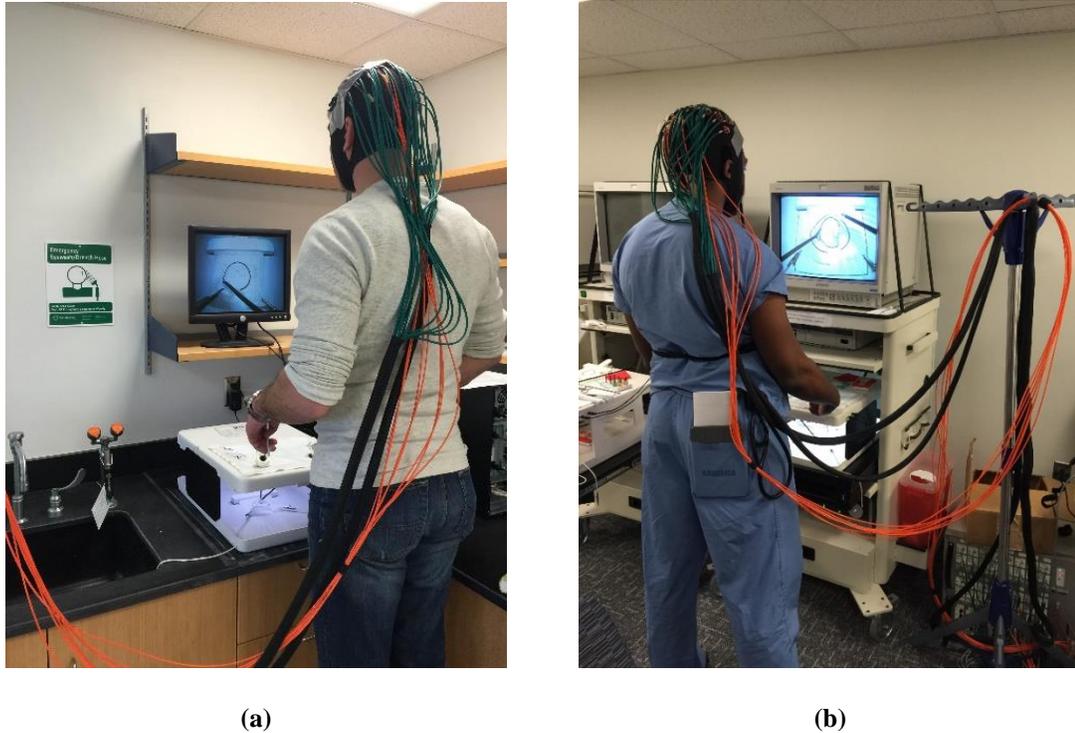
### 2.2.2 Participants and experimental design

17 surgeons and 13 medical students participated in this study. The minimum number of samples required for this study was determined a priori using power analysis according to the two-sample t-test comparing the means between two groups. Based on an initial pilot study, a conservative effect size ( $d=1.4$ ) was chosen for the prefrontal and motor cortices. Furthermore, with 95% confidence interval, and a minimum power of 0.80, it was determined that a minimum of 8 samples were required per group, which was calculated by a statistical software G\*Power[141]. The sample population was distributed within Novices ( $n=9$ , 1st – 3rd year residents with mean age  $31 \pm 2$ ) and Experts ( $n=8$ , 4th and 5th year residents and attending surgeons with mean age  $35 \pm 6$ ) surgeons. Example subjects performing the FLS pattern cutting task are shown in Figure 2.2 and subject demographics are shown in Table 2.1.

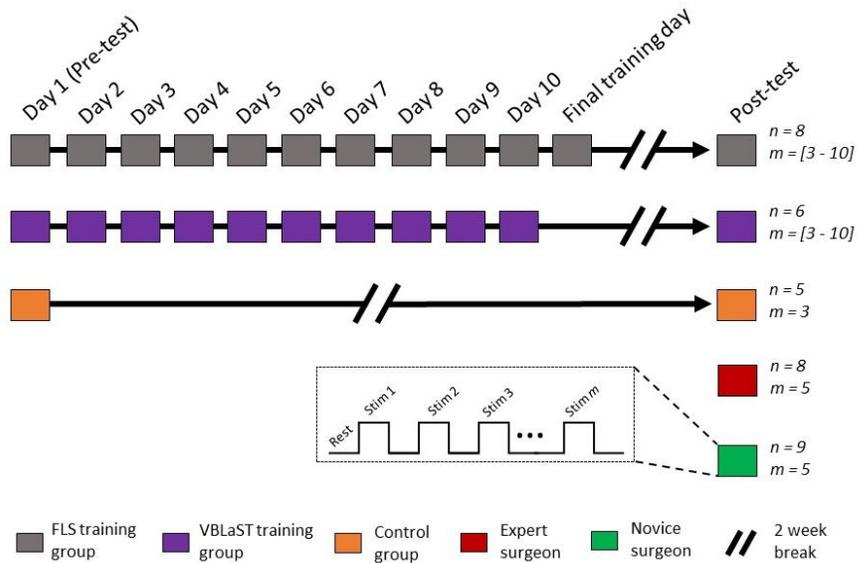
**Table 2.1: Subject demographics and descriptive data.**

Cohort	# of subjects	Mean age	Training / Certification	Average # of laparoscopic procedures	Average # of completed FLS pattern cutting trials during study	Color group
Expert surgeon	8	35	PGY 4-5 or attending surgeons	700	5	■
Novice surgeon	9	31	PGY 1 - 3	60	5	■
Skilled trainee	3	27	Medical school year 1-4	0	>85	■
Unskilled trainee	4	24	Medical school year 1-4	0	>85	■
Control	5	26	Medical school year 1-4	0	6	■

To avoid any issues regarding hemisphere specific activation, only right-handed participants were selected. All participants were instructed on how to perform the task with standardized verbal instructions indicating the goal of the task and rules for task completion. The optical probes were positioned on the participant with great care to avoid any hair between the source/detector and scalp, as well as, robust coupling with the skin. The cap holding the fibers on the participant as well as the fibers did not hinder the participant’s movement during bimanual tasks. The participants were asked to perform the FLS pattern cutting task using a FLS certified simulator, where the goal is to use laparoscopic tools to cut a marked piece of gauze as quickly and as accurately as possible. The experiment for each participant consisted of a block design of rest and stimulus period (cutting task). The surgical cutting task was performed until completion or stopped after five minutes. Then a rest period of one minute was observed. The cycle of cutting task and rest periods was repeated five times per participant. The experimental protocol design is shown in Figure 2.3. The following measurements were recorded simultaneously for each participant during each trial: total task time, light intensity (raw NIRS data), and performance scores for the pattern cutting task based on the FLS metrics.



**Figure 2.2: Subjects performing FLS pattern cutting task with fNIRS measurements. (a) An example medical subject, part of the FLS training group, and an example novice surgeon (b) performing a FLS pattern cutting task on the official FLS box trainer while undergoing fNIRS measurements in real time.**



**Figure 2.3: Schematic outlining cohort and study design. FLS training group (gray) and VBLaST training group (purple) underwent training regiments whereas the untrained control group (orange), surgical novices (green) and surgical experts (red) underwent no training.  $m$  is the number of trials per each session block and  $m$  is the number of pattern cutting trials per each session block.**

### **2.2.3 NIRS post processing**

Data processing was completed using the open source software HOMER2[142], which is implemented in Matlab (Mathworks, Natick, MA). First, channels with signal quality outside of the range of 80dB to 140dB were excluded. The remaining raw optical signals (intensity at 690nm and 830nm) were converted into optical density using the modified Beer-Lambert law with a partial path-length factor of 6.4 (690nm) and 5.8 (830nm) [68], [69], [143]. Motion artifacts and systemic physiology interference were corrected using recursive principal component analysis and low-pass filters [142], [144], [145]. The filtered optical density data is used to derive the delta concentrations of oxy and deoxy-hemoglobin.

The short distance channels are regressed from the long-distance channels to remove any interference originating from superficial layers. This is achieved by using a consecutive sequence of Gaussian basis functions via ordinary least squares to regress scalp and dura activation data collected from the short separation fibers to create the hemodynamic response function (HRF) [80], [81], [146]. Finally, the corresponding source and detector pairs for each source were averaged over each subject's task completion time. The result is a scalar value for the change in oxy-hemoglobin according to different brain regions for all participants.

### **2.2.4 Task performance metrics, statistical, and classification methods**

The FLS scores were determined using the standardized FLS scoring metric formulation for the pattern cutting task based on time and error. This formulation is IP protected and was obtained with consent under a non-disclosure agreement from the FLS Committee, and hence its details cannot be reported in this paper. Descriptive and inferential statistics were performed using SPSS (IBM Inc., NY, USA). Two sample t-tests were used to determine statistically significant differences in functional activation between two groups. All box plots display median values (red bar) along with standard deviations. A confidence level of 95% was selected as the minimum required to reject the null hypothesis.

Linear discriminant analysis (LDA) was used to classify the populations based on their FLS scores and functional brain activation metrics. Prior to the analysis of LDA, all

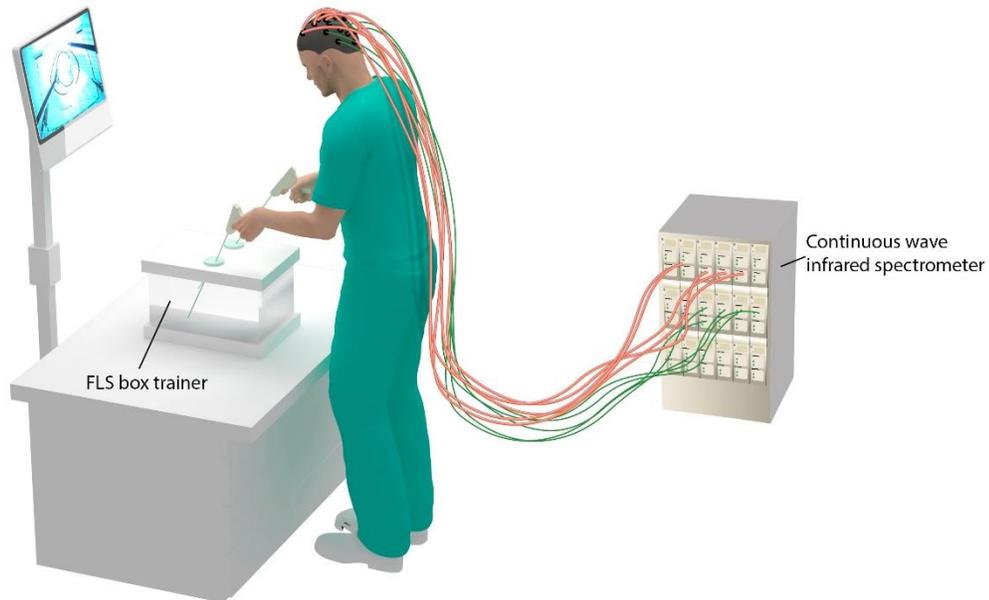
recorded metrics were first normalized, *i.e.* the sample mean and variance is 0 and 1. LDA determines the optimal vector,  $v$ , such that the projected metrics of two classes (ex. Novice and Expert surgeons) in the  $v$  direction has the highest separation between the classes with the lowest variance for each class [147]. The resulting LDA scores are objectively compared for each class and the degree of separation objectively quantified as misclassification errors.

## **2.3 Results**

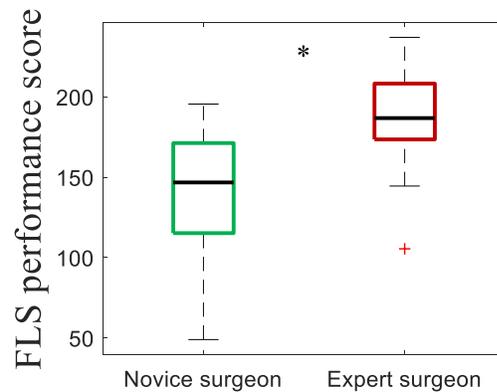
### **2.3.1 Surgical training task performance assessment**

To demonstrate the potential of neuroimaging as an objective tool to assess bimanual task expertise, we selected a challenging bimanual pattern cutting task, which is part of the fundamentals of laparoscopic surgery (FLS) program. Demonstrating proficiency in the FLS is now required for certification in general surgery by the American Board of Surgery. For our study, we recruited a population with varying laparoscopic surgical expertise as defined via the FLS program and conventional professional nomenclature. The subjects were either classified into established skill levels, such as Novice surgeons (1<sup>st</sup> – 3<sup>rd</sup> year surgical residents), Expert surgeons (4<sup>th</sup> – 5<sup>th</sup> year residents and attending surgeons) or into trained medical students that are labeled as Skilled or Unskilled trainees (see Table 2.1). The Control group constituted of medical students that underwent no training at all. Note that all groups were independent, *i.e.*, each subject belonged to only one group. Each subject followed the official FLS pattern cutting task protocols. The experimental protocol followed by each cohort is provided in Figure 2.3. The FLS performance scores were recorded for all subjects and the cumulative sum control chart (CUSUM) computed for the population following a training protocol. It is important to note that this study is the first to acquire FLS performance scores simultaneously with the neuroimaging data. The FLS scoring methodology was obtained with consent under a non-disclosure agreement from the FLS Committee. Thus, this study is the first one to report on direct comparisons of neuroimaging metrics and FLS scores for validation. In all cases, the FLS performance score were acquired simultaneously with the neuroimaging data. Figure 2.4 shows a schematic of the surgical trainer along with the fNIRS setup that is used to measure real-time cortical activation. A physical depiction of the setup is also

provided in Figure 2.3. Figure 2.5 reports on the descriptive statistics of the FLS performance score for the Novice and Expert surgeons, where Experts significantly outperformed Novice surgeons ( $p < 0.05$ ).

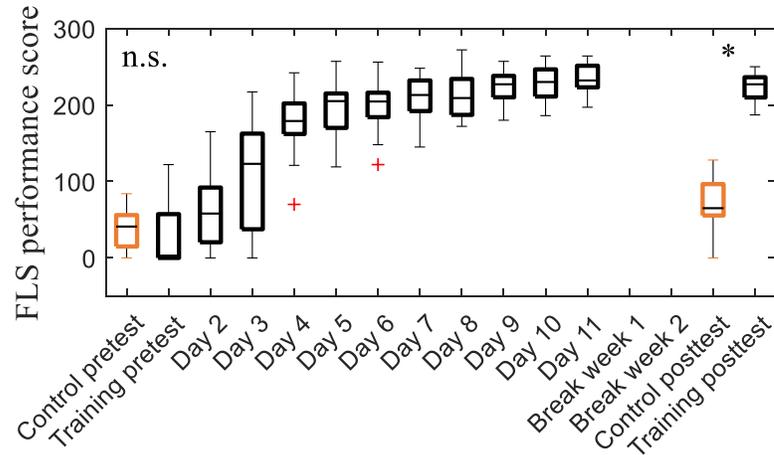


**Figure 2.4: Schematic depicting the FLS box simulator where trainees perform the bimanual dexterity task. A continuous wave spectrometer is used to measure functional brain activation via raw fNIRS signals in real-time.**



**Figure 2.5: FLS performance scores for Novice surgeons (green) and Expert surgeons (maroon) where Expert surgeons significantly outperformed Novice surgeons. Two sample t-tests were used for statistical differentiation ( $n.s.$  not significant,  $*p < 0.05$ ).**

Similarly, the descriptive statistics of FLS performance scores over the whole training period are provided for all FLS task training subjects and untrained Control subjects in Figure 2.6. Results indicate that there are no significant differences between the untrained Control subjects and training subjects on day 1 or the pretest ( $p>0.05$ ). However, the trained FLS students significantly outperformed the untrained Control students on the final post-test, which follows a two-week break period post training ( $p<0.05$ ).



**Figure 2.6: FLS performance scores for all training subjects (black) with respect to days trained compared to untrained Control subjects (orange) (n.s. not significant, \* $p<0.05$ ).**

To provide insight at the subject level, Figure 2.7 summarizes the CUSUM scores for each of the subjects with respect to trials performed. Trials that have a FLS performance score higher than 63 are considered a “success” and the respective CUSUM score is subtracted by 0.07 [16], [17]. Trials that have a performance score lower than 63 are considered a “failure” and the CUSUM score is added by 0.93 [16], [17]. Results indicate that three trained subjects (FLS 2, FLS 3, FLS 5) have passed the acceptable failure rate of 0.05 ( $H_0$ ) and thus are considered “Skilled” henceforth. The remaining trained four subjects (FLS 1, FLS 4, FLS 6, FLS 7) are considered “Unskilled” as they did not meet the FLS criteria for successful completion of the training program.

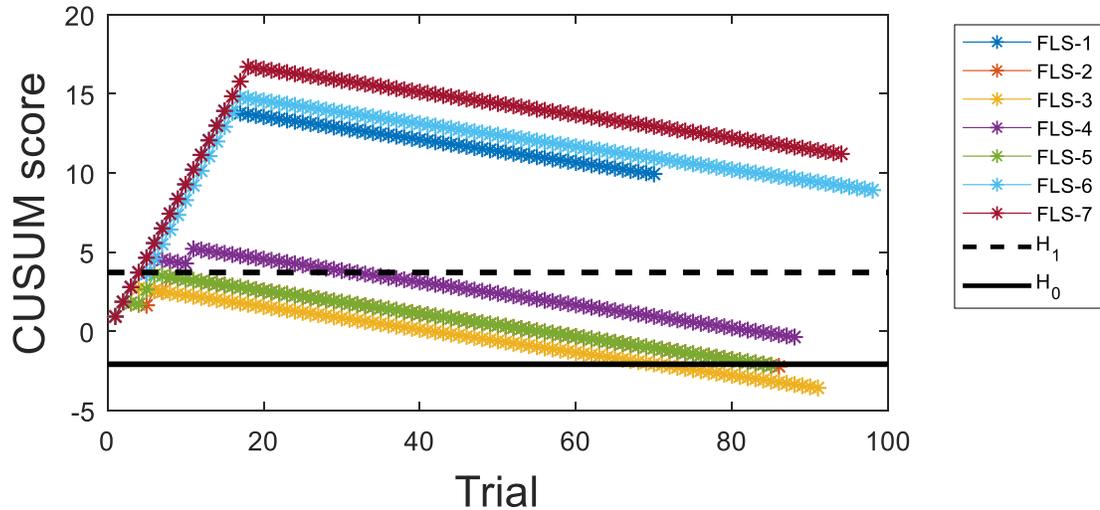
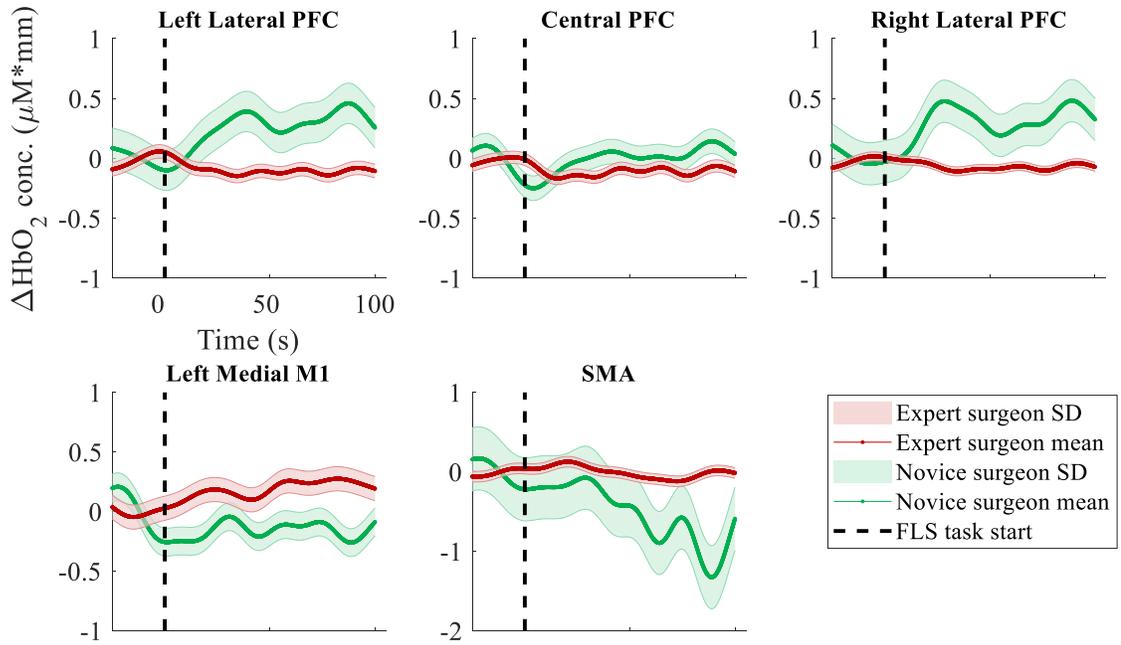


Figure 2.7: CUSUM scores for each subject with respect to trials. The  $H_0$  threshold indicates that the probability of any given subject is mislabeled as a “Skilled trainee” is less than 0.05, and is subsequently labeled as a “Skilled trainee” subject. Results indicate that three subjects, FLS-2, FLS-3, and FLS-5 are labeled as “Skilled trainees”. The remaining subjects that do not cross the  $H_0$  line are labeled “Unskilled trainees”.

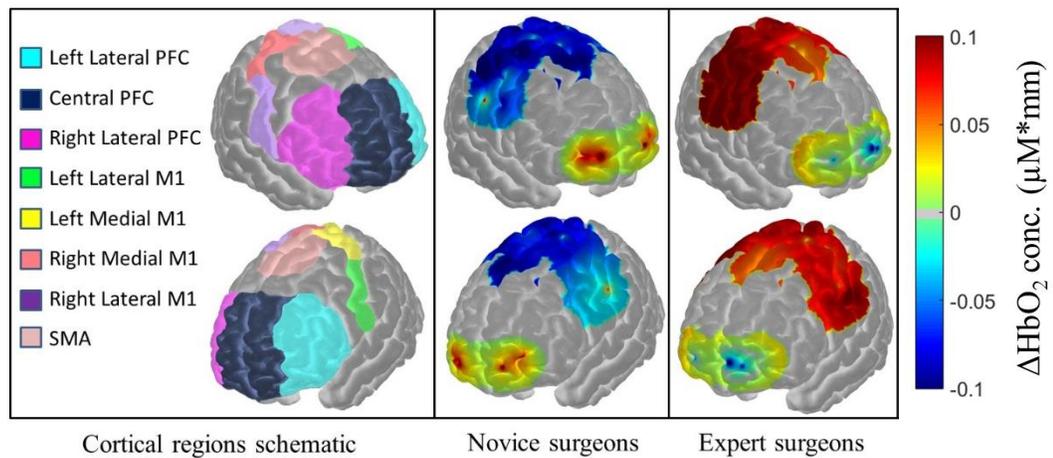
### 2.3.2 Optical neuroimaging assessment of established surgical skill levels

To ascertain that our neuroimaging methodology can discriminate between established skill levels, we quantified the real-time hemodynamic activation over the PFC, M1, and SMA cortical regions while Novice and Expert surgeons performed the standardized FLS bimanual pattern cutting (PC) task [16], [18], [20], where typical hemodynamic responses are shown in Figure 2.8.



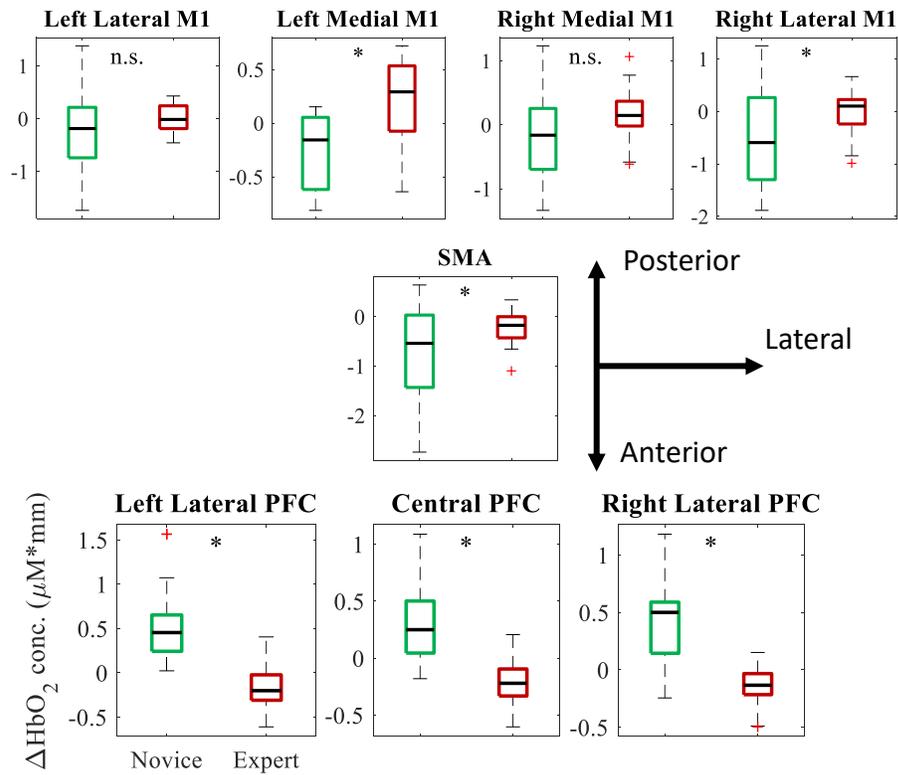
**Figure 2.8: Group average hemodynamic response functions with respect to cortical regions. Group average of all trials during the post-test for expert surgeons (maroon) and novice surgeons (green). Stimulus onset begins at zero seconds (dashed black line) indicating that the trial has started. Negative time indicates the baseline measurements used for calibration before each trial.**

Figure 2.9 depicts the spatial distribution of average changes in functional brain activation, as reported by  $\Delta[\text{HbO}_2]$  for all subjects in the surgical Novice and Expert groups.



**Figure 2.9: Differentiation and classification of motor skill between Novice and Expert surgeons. (a) Brain region labels are shown for prefrontal cortex (PFC), primary motor cortex (M1) and supplementary motor area (SMA) regions. Average functional activation for all subjects in the Novice and Expert surgeon groups are shown as spatial maps while subjects perform the FLS task.**

For the first time, significant differences were observed in all the PFC regions, the SMA, the left medial M1, and the right lateral M1 as depicted in Figure 2.10. More precisely, Novice surgeons have significantly higher functional activation in the PFC regions ( $p < 0.05$ ) and significantly lower functional activation in the left medial M1 and SMA regions when compared to Expert surgeons. Habituation, the phenomenon where a response to a stimulus is gradually reduced due to repetition[148], was not observed ( $p > 0.05$ ).

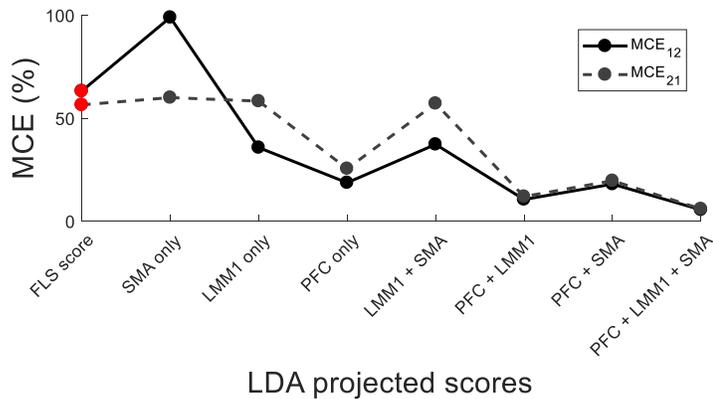


**Figure 2.10: Average changes in hemoglobin concentration during the FLS task duration with respect to specific brain regions for Novice (green) and Expert (maroon) surgeons. Two sample t-tests were used for statistical tests (n.s. not significant,  $p < 0.05$ ).**

While motor skill discrimination as reported via significant differences in the measurements from different cortical regions is typically central to neuroscience discovery studies, it does not provide insights into the utility of the data set to achieve robust classification based on quantitative metrics, such as accomplished during certification (*i.e.*, successfully pass a performance-based manual skills assessment). To quantify the performance accuracy of neuroimaging based classification of individuals in preset categories such as Novice surgeons (failed certification) and Expert surgeons (passed

certification), we post-computed misclassification errors (MCEs) associated with current accredited FLS performance scores and with our neuroimaging method.

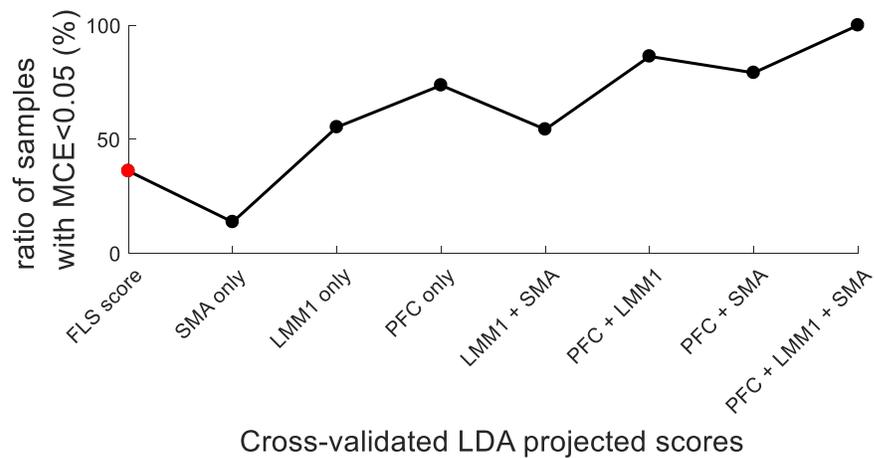
We employed a multivariate statistical method, namely linear discriminant analysis (LDA), to estimate the MCEs associated with the FLS and fNIRS based measurements. MCEs are defined as the probability that the first population is classified into the second population ( $MCE_{12}$ ) and the second population is classified into the first population ( $MCE_{21}$ ). Perfect classification is indicated by  $MCE = 0\%$  and complete misclassification is indicated by  $MCE = 100\%$ . Figure 2.11 reports on these two misclassification errors for FLS performance scores and all combinations of fNIRS metrics for the classification of surgical Experts and Novices. Results indicate that subject classification is relatively poor when considering FLS performance scores only ( $MCE_{12}=61\%$  and  $MCE_{21}=53\%$ ). On the other hand, neuroimaging based quantities provide lower errors (besides SMA only). Specifically, the combination of PFC, left medial M1 (LMM1) and the SMA leads to the overall lowest MCEs ( $MCE_{12}=4.4\%$  and  $MCE_{21}=4.2\%$ ).



**Figure 2.11: LDA classification results between Experts and Novices for FLS scores and all combinations of fNIRS metrics.**

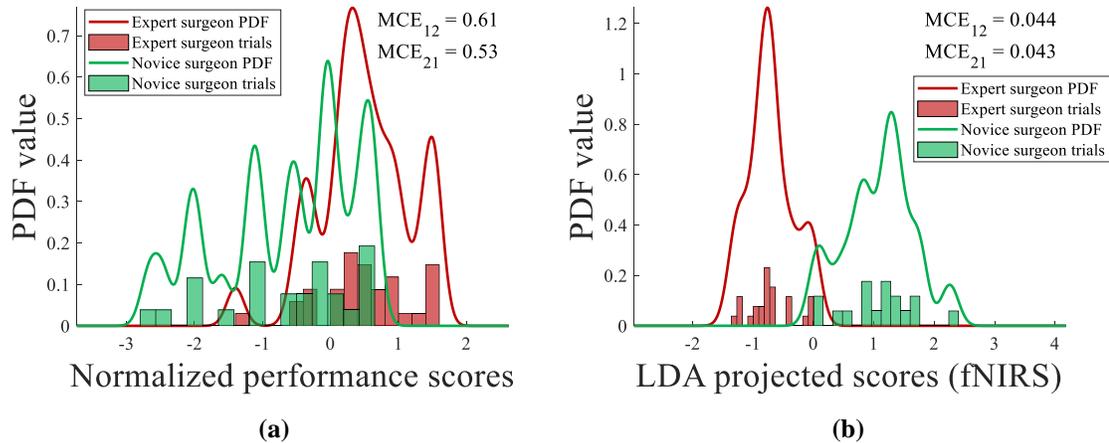
Additionally, we provide the leave-one-out cross-validation results for the LDA classification models used for this data set, as seen in Figure 2.12. This approach assesses the robustness of the LDA classification model, where each sample is systematically not used to build the LDA model and is treated independently. Results show that the combination of PFC + LMM1 + SMA leads to the most robust and best performing data

sets to build the classification model, as demonstrated by the fact that 100% of the samples in the leave-one-out cross-validation have MCEs < 5%.



**Figure 2.12: Leave-one-out cross-validation results show the ratio of samples that are below misclassification error rates of 0.05 for FLS scores and all other combinations of fNIRS metrics.**

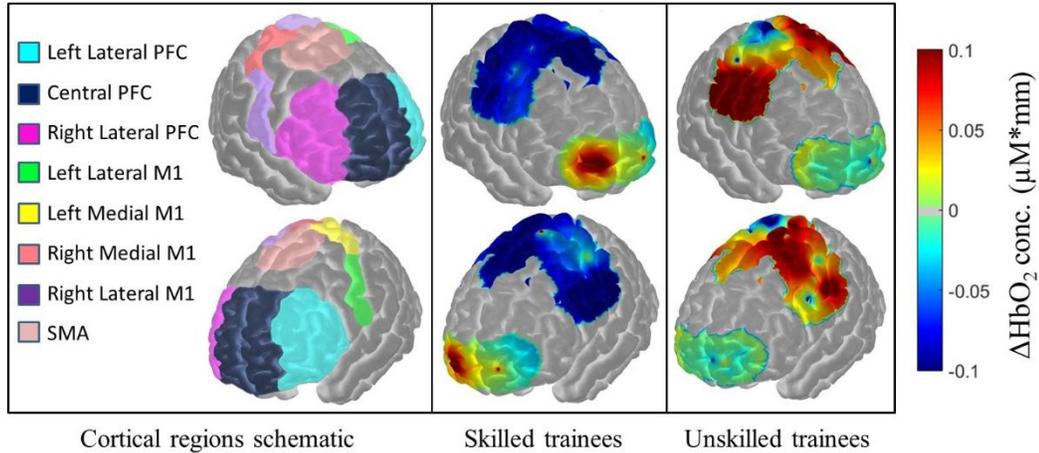
The specific distributions of the classification results are shown in Figure 2.13. Furthermore, weights for each cortical region and their respective contribution to the total LDA model were also determined to show the correlation between different cortical regions on motor skill proficiency. The weights for left lateral PFC (0.58), medial PFC (0.23), right lateral PFC (0.29), left medial M1 (-0.70), SMA (0.14) contribute to the entire discriminant function with the norm of all the weights equal to 1.0. Three regions (left lateral PFC, right lateral PFC, and left medial M1) account for 96.18% of discriminant function indicating the preponderance of these regions for a robust and accurate subject classification.



**Figure 2.13: Probability density functions (PDFs) for projected LDA classification models.** PDFs derived from kernel density estimation of normalized FLS performance for Novices, Experts, Skilled trainees, Unskilled trainees, and Control during the post-test. fNIRS metrics used for classification are functional activation in the PFC, LMM1, and SMA. The type I error is defined as 0.05 for all cases. (a) Using only FLS task performance as the only metric, results show that the probability for a Novice surgeon being misclassified as an Expert surgeon is 53% ( $MCE_2$ ) and the probability that a Novice surgeon is misclassified as an Expert surgeon is 61% ( $MCE_1$ ). (b) fNIRS based classification results show that the probability for a Novice surgeon being misclassified as an Expert surgeon is 4.4% ( $MCE_2$ ) and the probability that a Novice surgeon is misclassified as an Expert surgeon is 4.3% ( $MCE_1$ ). Similarly, Control subjects are classified against Unskilled and Skilled subjects.

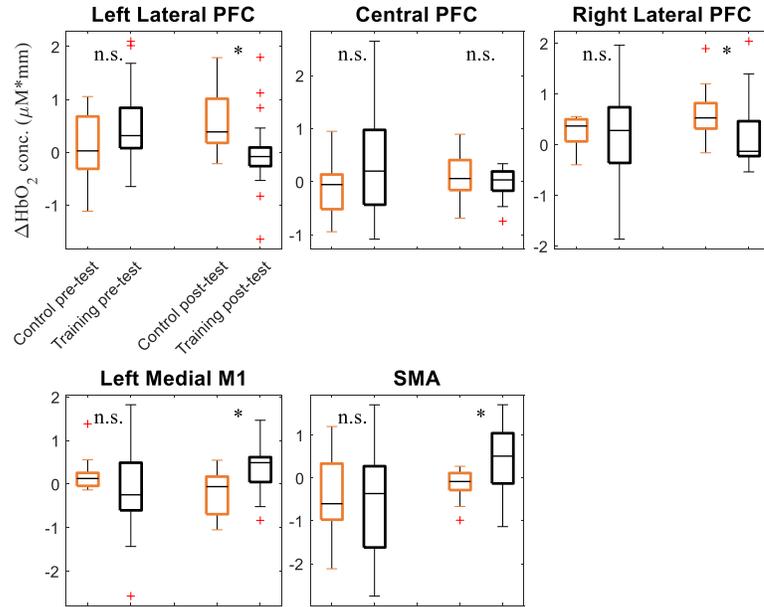
### 2.3.3 Optical neuroimaging assessment of surgical skill level during training

Beyond determining skill levels of individuals compared to established groups, one key challenge in bimanual skill assessment and in laparoscopic surgery is the evaluation of bimanual motor skill acquisition during training. We applied our neuroimaging methodology to the FLS pattern cutting task over an eleven-day training period for inexperienced medical students. Based on the established FLS metrics currently employed in the field, the enrolled medical student population was divided into Skilled and Unskilled trainees at the completion of the training program as previously shown in Figure 2.7. Additionally, five medical students with no prior experience in laparoscopic surgery were recruited as the Control group that underwent no training. Figure 2.14 shows a visual spatial map conveying the average cortical activation of all Skilled trainees or Unskilled trainees while performing the post-test (i.e. simulated certification exam). Like Expert vs Novice surgeons as shown in Figure 2.9, Skilled trainees exhibit increased cortical activation in the left medial M1 and SMA and decreased PFC activation when compared to Unskilled trainees upon training completion and after a two-week break.



**Figure 2.14: Differentiation and classification of motor skill between Control, Skilled, and Unskilled trainees. (a) Spatial maps of average functional activation for all subjects in each respective group during the FLS training task on the post-test day.**

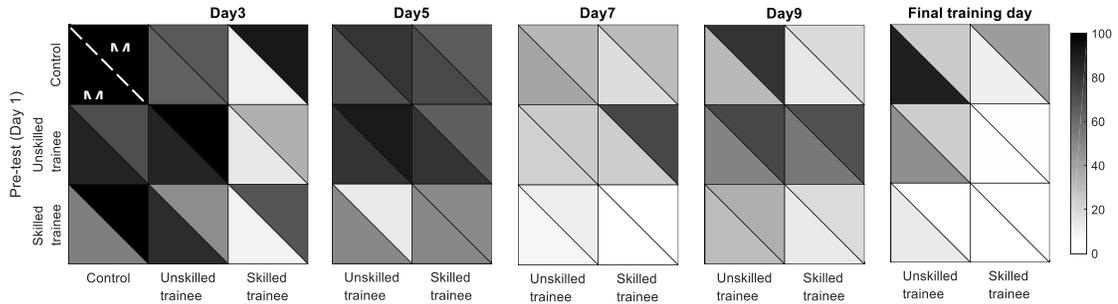
To provide a more global view of the training outcome, we present the descriptive statistics of functional activation between untrained Control students and all trained FLS students for pre-test (Day1) and post-test (final day after two-week break period) with respect to different brain regions in Figure 2.15. Results indicate that there are no significant differences between the Control and all training students (Skilled and Unskilled trainees) at the onset of the training program ( $p > 0.05$ ). However, at the completion of the training and after a two weeks break period, both Skilled and Unskilled trainees exhibit a significantly lower functional activation in the left lateral and right lateral PFC compared to the untrained Control students ( $p < 0.05$ ). Furthermore, trained FLS students have significantly higher left medial M1 and SMA activation than untrained Control students during the post-test ( $p < 0.05$ ). These results reinforce the findings of the previous section regarding functional activation differences between Expert and Novice surgeons. To further stress the fact that our neuroimaging modality enables to provide a more granular view of training outcomes, we computed the MCEs for the three populations involved in this surgical training study (Control, Skilled and Unskilled trainees).



**Figure 2.15: Average changes in hemoglobin concentration during stimulus duration with respect to specific brain regions for untrained Control subjects (orange) and all FLS training students (black). Two sample t-tests were used for statistical differentiation (n.s.: not significant, \* $p < 0.05$ ). Type I error is defined as 0.05 for all cases.**

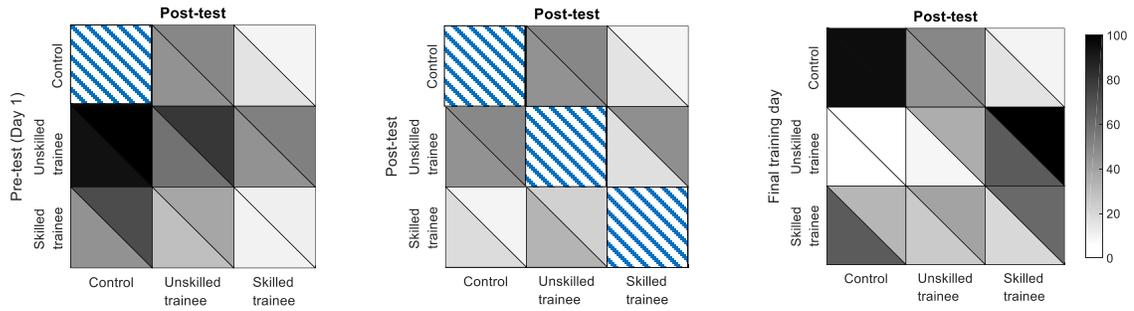
Figure 2.16 and Figure 2.17 report on the MCEs for each potential combination of medical student populations at different stages or end points of training. These MCEs were computed using the combined PFC, LMM1 and SMA brain functional optical measurements. The longitudinal MCEs of pretest populations versus odd days of training indicate that at the onset of the training, the populations could not be distinguished as reported by large inter-group misclassification errors, as shown in Figure 2.16.

However, after day 7, the Skilled trainee population demonstrated a significantly different neuroimaging distributed response compared to the first day of training, as demonstrated by very low intra-group misclassification errors. Conversely, the Unskilled trainee population did not exhibit such marked trends. Even during the final training day (day 11), poor intra-group misclassification errors were observed for the Unskilled trainee population ( $MCE_{12} = 24\%$  and  $MCE_{21} = 47\%$ ). In contrast, Skilled trainees on the final training day were completely classified from Skilled trainees on the pre-test, with  $MCE_{12} = 0\%$  and  $MCE_{21} = 0\%$ .

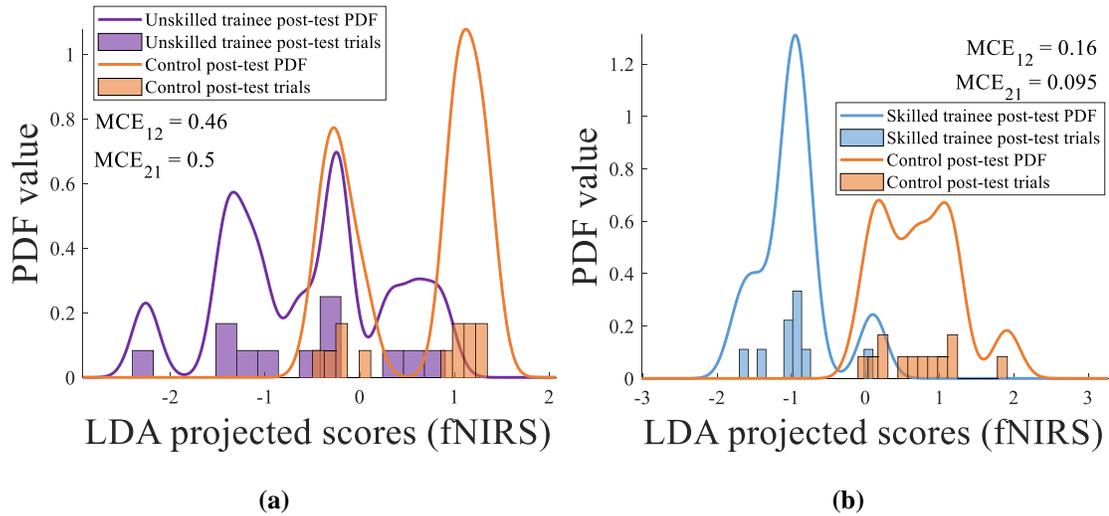


**Figure 2.16: Inter and intra-group misclassification errors for each subject population (Control, Skilled and Unskilled trainees) with respect to training days.  $MCE_{12}$  and  $MCE_{21}$  values significantly decrease below 5% when classifying pre-test Skilled and Unskilled trainees on the final training day. Furthermore, misclassification errors are also low when classifying Skilled and Unskilled trainees on the final training day, along with Skilled trainees and untrained Control subjects.**

Similar results were observed when looking at the same intra-group misclassifications between the pre-test and post-test conditions, as shown in Figure 2.17. Classification continues to remain poor for Unskilled trainees when comparing this population from the pre-test and the post-test, with  $MCE_{12} = 58\%$  and  $MCE_{21} = 80\%$ . Yet, Skilled trainees during the pre-test are successfully classified from Skilled trainees during the post-test, with  $MCE_{12} = 10\%$  and  $MCE_{21} = 11\%$ . While the Unskilled and Skilled trainee inter-groups were successfully classified at the end of the training session compared to the pre-test, the two populations did exhibit some intra-group overlap in their associated probability density function during the post-test. Of importance, both trainee populations did not exhibit marked differences between the final training day and post-test measurements as indicated by relatively high MCEs. Classification of Skilled trainees and Control subjects during the post-test also yielded in very low misclassification errors, whereas classification of Unskilled trainees and Control subjects still yielded in high misclassification errors, as shown in further detail in Figure 2.18. These cross-validated classification methods show that cortical activation has significantly changed for Skilled trainees during the post-test when compared to Skilled trainees on the pre-test or untrained Control subjects whereas Unskilled trainees do not exhibit such a marked trend.



**Figure 2.17:** Misclassification errors are reported for each combination of training groups (Control, Skilled, and Unskilled trainees) with respect to pre-test, post-test, and final training days. MCEs are substantially low when classifying Skilled trainees and Control subjects along with inter-Skilled trainee group classification. Unskilled trainees, however, showed high misclassification errors even when compared to Unskilled trainees and Control subjects during the post-test. As a measure of skill retention, classification models were also applied for all subject groups from the final training day to the post-test.



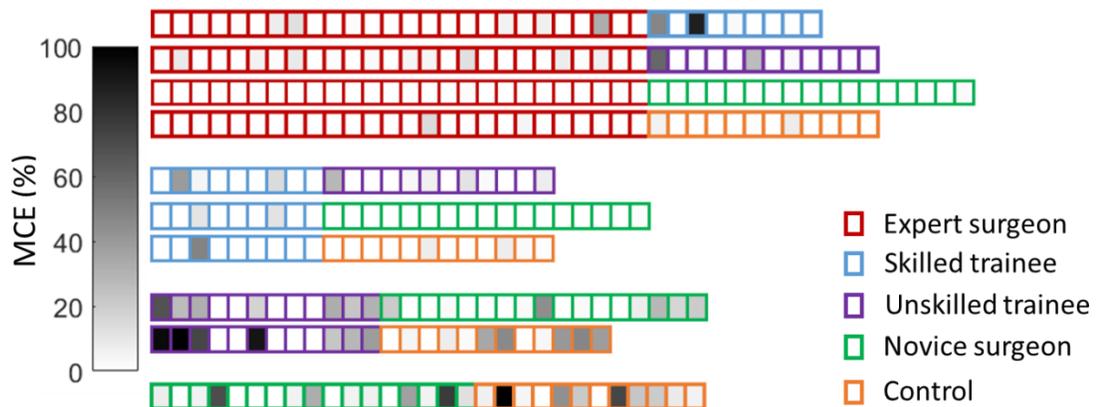
**Figure 2.18:** (a) The probability that an untrained Control subject is misclassified as an Unskilled trainee is 46% ( $MCE_{12}$ ) and the probability that an Unskilled trainee is misclassified as a Control is 50% ( $MCE_{21}$ ). (b) Conversely, the probability that a Control is misclassified as a Skilled trainee is 16% ( $MCE_{12}$ ). Whereas the probability that a Skilled trainee is misclassified as a Control is 9.5% ( $MCE_{21}$ ).

### 2.3.4 Classification of subjects with varying surgical expertise levels

For our neuroimaging based approach for motor skill differentiation to be formative, it is important to validate the classification-test models across all subject populations, especially since the studies associated with assessment of established skill levels and FLS training were performed independently in two different institutions. The subject population represents the full spectrum of laparoscopic surgical expertise, from Novices

to certified attending surgeons, including Skilled and Unskilled medical student trainees. Regarding the number of procedures and associated level of expertise (at the completion of the training protocol), it is expected that the distribution in terms of surgical skills levels, from more proficient to less proficient is distributed as follow at the group level: Expert surgeons, Skilled trainees, Unskilled trainees, Novice surgeons, and Control.

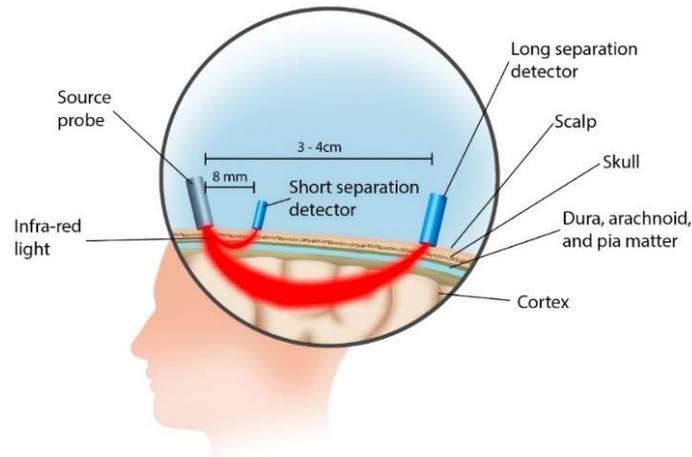
Figure 2.19 shows the cross-validated classification model results comparing all subject population groups with varying expertise levels. Each box corresponds to a single trial for each expertise group, as shown via different colored borders. Shaded regions within each box indicate the MCE if that trial is removed from the classification model. For example, the first trial in cross-validation results for classifying Expert surgeons from Skilled trainees shows a MCE of 0%, as indicated by a white shade. However, the 29<sup>th</sup> sample in the classification model, or the third trial in the Skilled trainee group, shows a MCE of 89% when removed from the classification model. The latter is an indication that the LDA classification model fails to reliably classify Experts and Skilled trainees if the third sample in the Skilled trainee group is removed.



**Figure 2.19: Cross-validation results for classification across all subjects with varying degree of motor skills. Each box represents one trial per expertise group during the post-test, where the shaded regions indicate the MCE if that given trial is removed from the classification model. Cross-validation results with their respective ratio of samples that are below misclassification error rates of 5% for Expert surgeons vs Skilled trainees (28/35 samples), Expert surgeons vs Unskilled trainees (29/ 38), Expert vs Novice surgeons (43/43), Expert surgeons vs untrained Control subjects (34/38), Skilled trainees vs Unskilled trainees (15/21), Skilled trainees vs Novice surgeons (24/26), Skilled trainees vs untrained Control subjects (18/21), Unskilled trainees vs Novice surgeons (16/29 samples), Unskilled trainees vs untrained Control subjects (11/24), and finally Novice surgeons vs untrained Control subjects (9/29).**

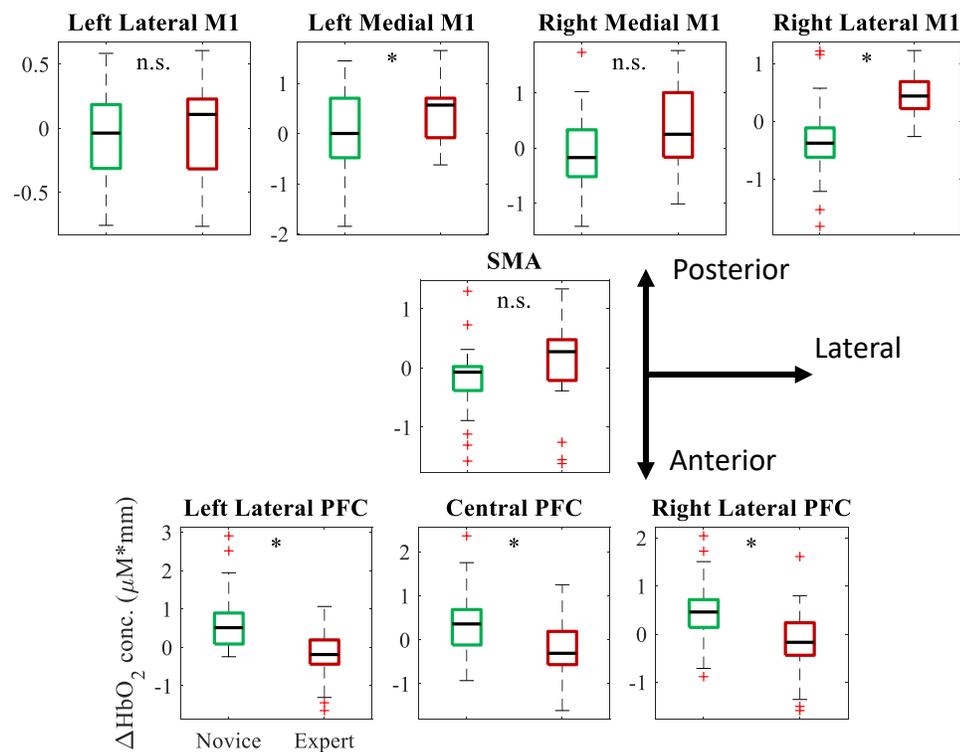
## 2.4 Discussion

While there have been extensive efforts in the surgical community to confirm training effectiveness and validation of the FLS program [18], [20]–[23], the surgical skill scoring component has received little attention and has garnered criticisms, such as subjectivity in scoring, inconsistencies in FLS score interpretations, and no correlation of patient injury reduction due to FLS certification [18], [22], [34], [54]–[57]. Despite the lack of rigorous evaluation of the FLS scoring methodology, the program has become the de-facto evaluation method for accreditation of skills required for general surgery [55]. Given the high-stakes nature of surgical assessment in the FLS program and its implications on training future surgeons, there is a current gap in the rigorous validation of FLS scores as a robust and objective methodology [55]. In this regard, previous studies have broached the concept of non-invasive brain imaging as a means for objectively assessing surgical skills [60], [72]–[74]. However, they suffer from methodological limitations that are now well-recognized by the fNIRS community, namely the contamination of superficial tissue, such as scalp, dura, or pia matter, in the recorded measurements[80], [81] as shown in Figure 2.20.

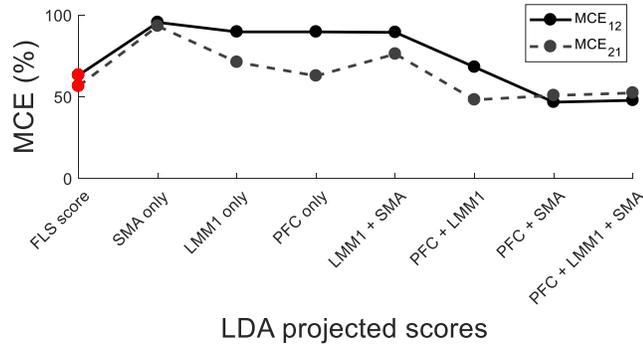


**Figure 2.20: Cross-sectional diagram of infrared light propagation through cortical tissue. The short separation detectors are placed 8mm away from each source to ensure the backscattered light is solely due to superficial tissue, such as scalp, skull, dura, arachnoid, and pia matter. The large separation detectors are placed 3-4cm away from each source to ensure sufficient light penetration depth into the cortex is achieved.**

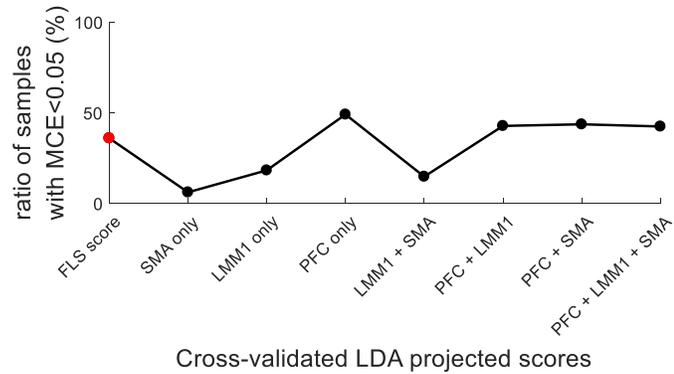
To highlight this point, results from the Expert and Novice surgeon cohort in this study were reprocessed without the regression of superficial tissue data and are provided in Figure 2.21, Figure 2.22, and Figure 2.23. These results clearly demonstrate that previously reported fNIRS based metrics with the inclusion of superficial tissue responses can statistically differentiate surgical novices and experts [60], [73], [74], [137], [138], yet fail to classify subjects based on motor skill proficiency and perform as poorly as current surgical skill assessment metrics. In contrast, regressing shallow tissue hemodynamics from the optical measurements significantly reduces the false omission rate, where a surgical novice is mistakenly classified as an expert, to 0% whereas previous approaches still maintain false omission rates of 13-18%, as shown in Table 2.2.



**Figure 2.21: Differentiation and classification of motor skill between Novice and Expert surgeons without short separation regression. (a) Average changes in hemoglobin concentration during the FLS task duration with respect to specific brain regions for Novice (green) and Expert (maroon) surgeons. Two sample t-tests were used for statistical tests (n.s. not significant,  $p < 0.05$ ).**



**Figure 2.22: LDA classification results for FLS scores and all combinations of fNIRS metrics using fNIRS metrics with superficial tissue signals included.**

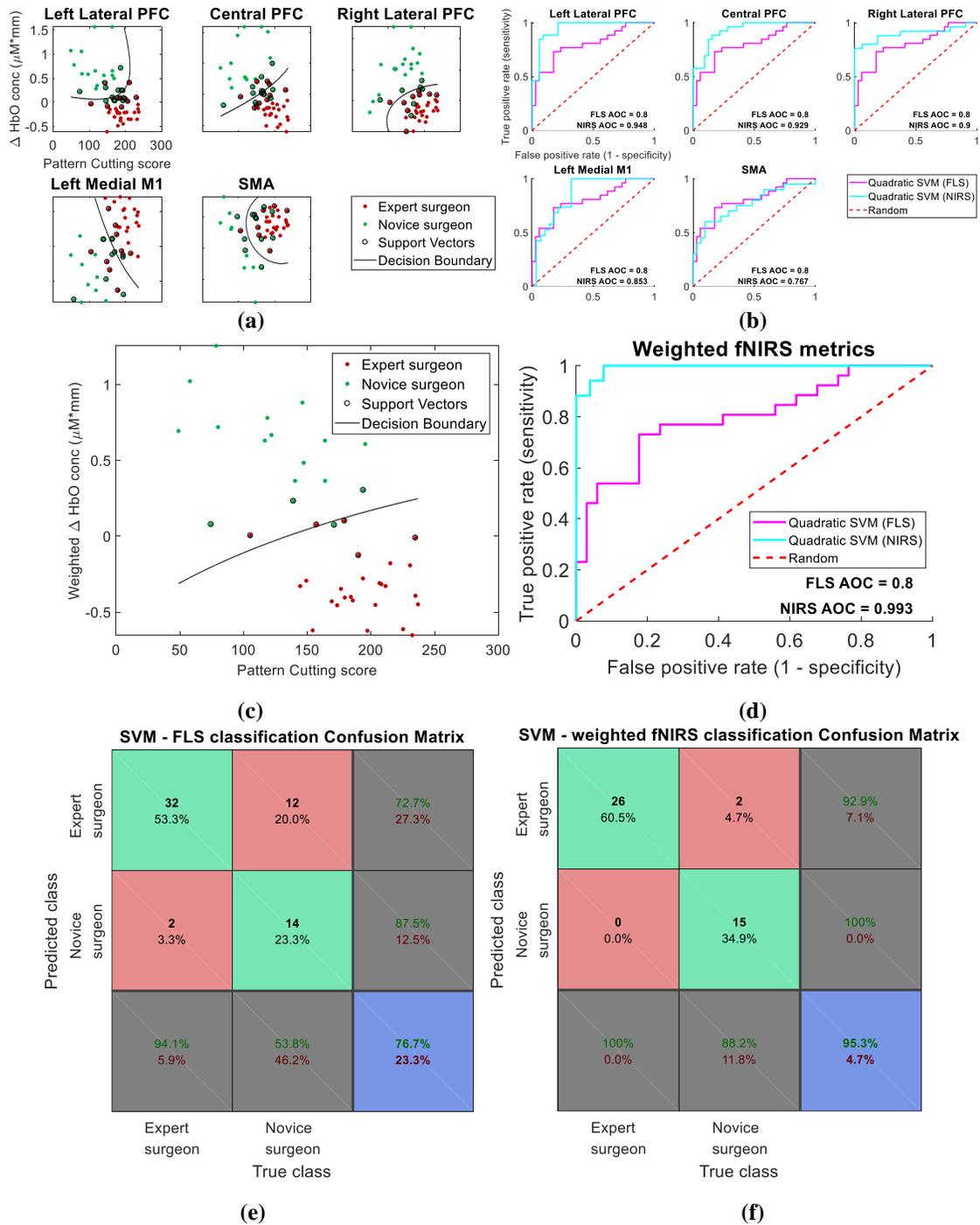


**Figure 2.23: Leave-one-out cross-validation results show the ratio of samples that are below misclassification error rates of 0.05 for FLS scores and all other combinations of fNIRS metrics when measurements include superficial tissue signals.**

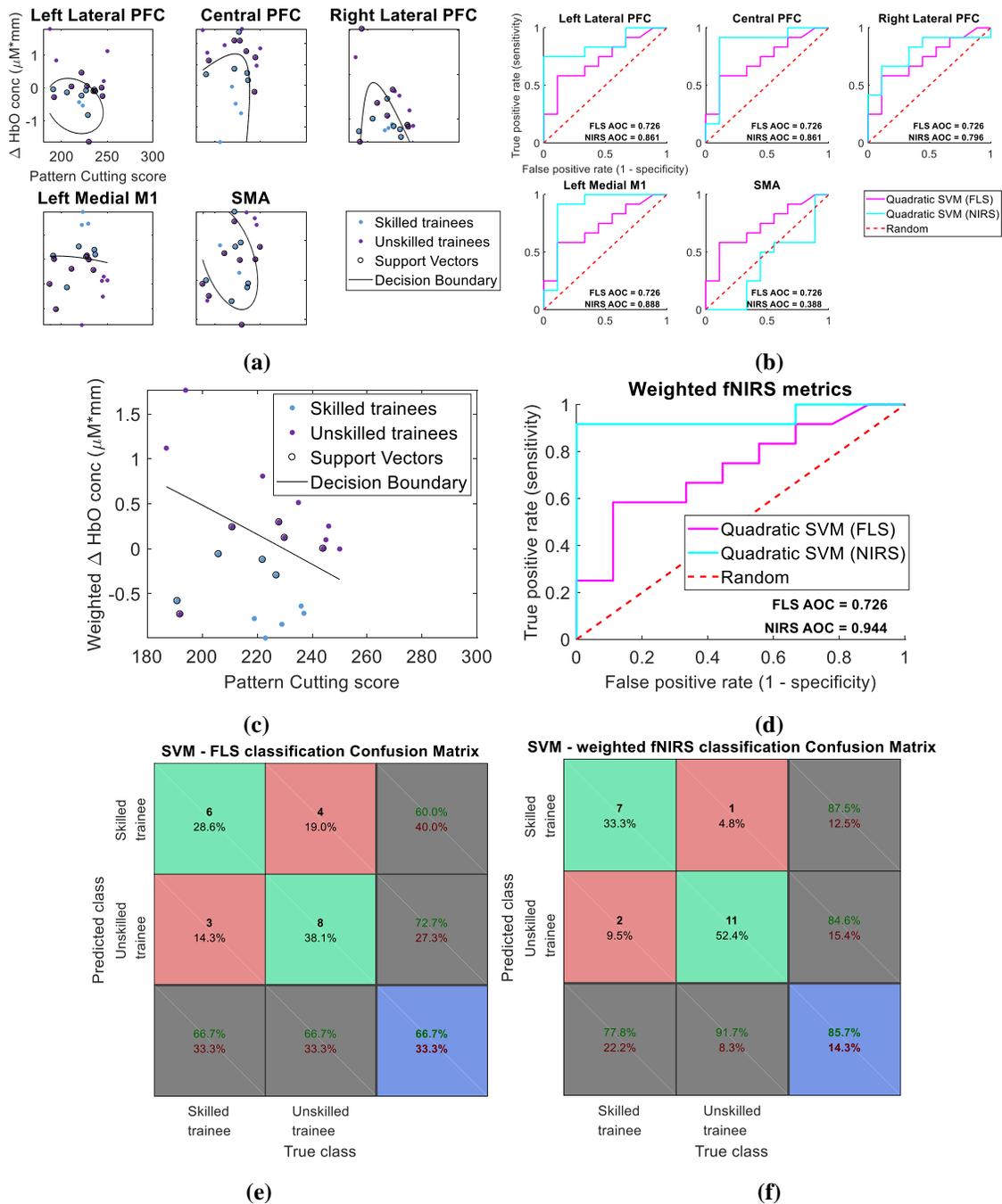
**Table 2.2: Expert vs Novice classification results for fNIRS (with and without short separation regression) and FLS metrics.**

Weighted Quadratic SVM classification	Sensitivity	Specificity	Positive predictive value (PPV)	False omission rate (FOR)	ROC – AUC
FLS scores only	0.94	0.54	0.73	0.13	0.8
Without short separation regression	0.88	0.76	0.82	0.18	0.91
With short separation regression	1.00	0.88	0.93	0.00	0.99

Beyond improving the robustness of optical measurements sensitivity to cortical activations, this work is also the first to measure functional activation in a multivariate fashion to determine critical cortical regions that are correlated to surgical motor skill differentiation and classification. More specifically, this is the first report of measuring functional activations in the PFC, M1, and SMA cortical regions that are putatively associated with motor task strategy, motor task planning, and fine motor control [73], [74], [132], [137], [149]–[154]. Our results demonstrate that the inclusion of these cortical regions significantly improves the utility of fNIRS in assessing bimanual skills and can offer improved objective metrics over conventional FLS-based metrics currently used for certification in general surgery. Of importance, while using single regional readouts lead to enhanced population differentiation, the combination of the three above mentioned cortical regions provide excellent classification performances (for completeness, we also provide bivariate classification results using support vector machines in Figure 2.24 and Figure 2.25. Indeed, when combining measurements from these three brain regions, optical neuroimaging enables a remarkably robust classification of subjects based on their proven surgical skills levels, including novice, intermediate and expert skill levels. More precisely, our methodology allows for: (1) highly accurate classification of subjects with well-defined bimanual skills levels with better performance than currently employed metrics, (2) longitudinally assessing the acquisition of surgical skills during the FLS training program, and (3) performing robust classifications of populations recruited from multiple institutions with varying skill levels.



**Figure 2.24: Quadratic support vector machine (SVM) classification of Expert and Novice surgeons.** (a) Unsupervised, quadratic polynomial support vector machines were used for bivariate classification of Expert and Novice surgeons. Quadratic support vectors outlining the decision boundaries are shown. (b) Receiving operating characteristic (ROC) curve showing the sensitivity vs specificity of SVM based classification. (c) Weights from the corresponding LDA classification of Expert vs Novice surgeons were used to combine the cortical activations as a single metric. (d) The corresponding ROC curves for the weighted fNIRS metric based classification vs traditional FLS scores. (e-f) Confusion matrices showing the specific results of true positive and true negative classes (green) along with the false positive and false negative classes (maroon). Results indicate that SVM based on weighted fNIRS classification show no cases of false negatives.



**Figure 2.25: Quadratic support vector machine (SVM) classification of Skilled vs Unskilled trainees.** (a) Unsupervised, quadratic polynomial support vector machines were used for bivariate classification of Skilled and Unskilled trainees. Quadratic support vectors outlining the decision boundaries are shown. (b) Receiving operating characteristic (ROC) curve showing the sensitivity vs specificity of SVM based classification. (c) Weights from the corresponding LDA classification of Skilled vs Unskilled trainees were used to combine the cortical activations as a single metric. (d) The corresponding ROC curves for the weighted fNIRS metric based classification vs traditional FLS scores. (e-f) Confusion matrices showing the specific results of true positive and true negative classes (green) along with the false positive and false negative classes (red).

On a practical side, it is important to note that even if our methods leverage the most recent technical developments in the field of fNIRS, the instrumental and algorithmic platforms employed herein are readily available for wide-dissemination and use in surgical training facilities. Moreover, as more neuroscience-driven investigations focus on mapping distributed brain function, the positioning of the optodes (source or detector) on the subject scalp become increasingly challenging with extended spatial coverage. One key consideration is to ensure effective coupling is minimally affected by natural movements and not compromised by the subject's hair. Hence, positioning of the optodes can be a lengthy process that is not suitable for professional environments that are time-constrained either by cost or throughput considerations. In this regard, our study identifies that the PFC, SMA and left medial M1 regions are sufficient for accurately assessing bimanual skill-based task execution. Thus, probe placement can be completed in a short amount of time without any impact on task execution, both critical factors for an acceptance of our surgical skill assessment methodology by the surgical community.

Beyond bimanual skills assessment and objective classification of individuals based on their skill levels, the work herein provides a sound foundation to further investigate the neurophysiology underlying bimanual skill acquisition and retention. Herein, we deliberately focused on reading the brain outputs as a mean to provide objective and quantitative measures of bimanual task execution without delving into the mechanistic understanding of the underlying physiology and functional connectivity. However, current neurophysiological knowledge supports the overall findings of our studies, namely increases in left medial M1 and SMA activation, and significant decreases in PFC activation across all groups with increasing motor task performance [60], [73], [74], [132], [137], [149]–[154]. It is also important to note that previous studies utilize motor tasks that are deliberately designed to decrease variability in studying cortical activation changes, such as finger tapping or simple visual or virtual based unimanual tasks. Conversely, the FLS task at hand is a complex bimanual task that involves visuospatial coordination, varying degrees of synchronicity between hands, motion frequency and range, and exerted forces on the surgical tools for task completion. Consequently, it is not feasible to ensure that each session replicates the same conditions and hence, the same

cortical responses. Moreover, the cortical activations and interactions associated with the task planning and execution are dynamic by nature from expected explicit control in the early stages of learning to more implicit or automatic control in the later stages of motor learning. Thus, mapping the cortical networks and their dynamical changes associated with task execution and skill acquisition should be the next step.

Indeed, there is currently great interest in investigating dynamic functional connectivity (DFC) in neuroscience. Typically, DFC studies are conducted using fMRI, which is not appropriate for protocols requiring supine positions and/or non-elicited task execution. Recent studies have demonstrated that fNIRS is well positioned in such scenarios [155]–[157]. We foresee that implementing such approaches in the context of bimanual skill assessment can lead to refined skill level assessment metrics as well as potentially provide predictive models of skill acquisition. For instance, composite cognitive metrics, possibly obtained by weighting regional cortical measurements using the LDA weights for best classification between Skilled and Unskilled trainees, could be central to developing tailored surgical training program for optimal skill acquisition and retention assessment, as shown in Figure 2.25. Furthermore, these methodologies can be easily applied to other fields including rehabilitation, brain computer interfaces (BCI), robotics, stroke and rehabilitation therapy [158]–[160]. In summary, we believe this non-invasive imaging approach for objective quantification for complex bimanual motor skills will bring about a paradigm change in broad applications such as surgical certification and assessment, aviation training, and motor skill rehabilitation and therapy.

## **2.5 Summary**

In this chapter, we propose fNIRS as a non-invasive imaging method to quantify bimanual motor skills and show evidence for two major conclusions. First, we show that functional activation decreases in the PFC and increases in the left medial M1 and SMA when motor skill levels increase. Secondly, fNIRS measurements of combined cortical activation in the PFC, M1, and SMA are sufficient to successfully classify participants based on varying degrees of motor skills levels with significantly lower misclassification errors when compared to established FLS metrics.

### 3. Objective assessment of surgical skill transfer using non-invasive brain imaging

#### 3.1 Overview

With mounting concerns about patient safety and the need to have objective measures of surgical technical competence, simulation as a means of surgical training and certification is rapidly gaining ground [161]. The Fundamentals of Laparoscopic Surgery (FLS), which employs a box-trainer, and the Fundamentals of Endoscopic Surgery (FES) with a virtual reality-based simulator, have been recently adopted by the American Board of Surgery as pre-requisites for certification in general surgery [16], [17], [20], [23], [162]–[166]. However, prior to acceptance, each simulator, real or virtual, must undergo extensive validation to demonstrate effectiveness. One critical measure of simulator effectiveness is the evidence of successful transfer of technical skills from the simulation environment to the clinical environment [161], [167], [168]. While metrics for technical skills assessment may vary, the current standard in assessing successful transfer of skills from the simulation environment to a clinical setting is direct observations by an expert clinician [1] using a checklist such as the Objective Structured Assessment of Technical Skills (OSATS) or Global Operative Assessment of Laparoscopic Surgery (GOALS) [1]–[3]. Alternative metrics such as task completion time have also been reported for assessing technical skill transfer [169]. Despite the current widespread usage of these generalized rating or completion time based assessments, there are significant drawbacks to these methods that include personnel resource costs, poor interrater reliability between proctors, and poor correlation of learned technical skills from the simulator to outcomes in the operating room [7], [52], [169]. These limitations necessitate a need for more objective and analytical methods to assess surgical skill transfer [49], [51].

---

Portions of this chapter previously appeared as:

A. Nemani, W. Ahn, C. Cooper, S. Schwaitzberg, and S. De, “Convergent validation and transfer of learning studies of a virtual reality-based pattern cutting simulator,” *Surg. Endosc.*, to be published. doi: 10.1007/s00464-017-5802-8.

A. Nemani, C. Cooper, X. Intes, S. De, and S. D. Schwaitzberg, “Noninvasive brain imaging demonstrates that surgical skills transfer from training simulators to *ex-vivo* models,” *J. Am. Coll. Surg.*, vol. 225, no. 4, pp. 22, Oct. 2017.

A promising technique that is objective in determining surgical motor skills is non-invasive brain imaging. Among all the non-invasive brain imaging methods currently available, functional near-infrared spectroscopy (fNIRS) offers the unique features to be portable, non-invasive, non-obtrusive to performing the surgical task, fast and relatively inexpensive [137], [170]. Investigators have used fNIRS to study brain activation responses between surgical experts and novices during the performance of surgical training tasks by measuring the fluctuations of hemodynamics signals, namely changes in concentration of oxy- (HbO) and deoxy- (HbR) hemoglobin [60], [72]–[74], [76], [77], [171]. However, these studies are limited in scope as they are subject to signal contamination from superficial tissue, and show no evidence of surgical skill transfer to more clinically relevant environments.

The purpose of this study is to determine if fNIRS can accurately assess motor skill transfer from simulation to *ex-vivo* environments for trained and untrained subjects as they perform an established surgical training task. We hypothesize that fNIRS based metrics can classify different levels of surgical motor skill transfer with more accuracy than established methods. To test this hypothesis, subjects trained on a physical or virtual surgical simulator where subjects practiced a surgical training task and subsequently performed a surgical transfer task post-training. The physical simulator utilized in this study is the official FLS box trainer, whereas the virtual trainer is the Virtual Basic Laparoscopic Skills Trainer (VBLaST), which is a validated virtual simulator that simulates the five FLS task modules in real time [26], [34]–[37], [54]. In order to measure cortical activation changes during the transfer task, we measure functional activation specifically in the prefrontal cortex (PFC), primary motor cortex (M1) and the supplementary motor area (SMA), as these cortical regions are directly involved in fine motor skill learning, planning, and execution [60], [82], [134], [136], [140], [160]. Multivariate statistical approaches were then used to objectively differentiate and classify subjects that exhibit successful motor skill transfer.

## 3.2 Methods

The study was approved by the Institutional Review Board of University at Buffalo and Rensselaer Polytechnic Institute.

### 3.2.1 Subject recruitment

Prior to subject recruitment, we performed an *a priori* analysis according to the Mann-Whitney U test to determine the minimum number of subjects required for the FLS training group, VBLaST training group, and the control group. Using pilot data, we estimated conservative effect sizes for the FLS and VBLaST groups and show that  $d = 5.67$  and  $d = 2.57$ , respectively. Based on these effect sizes, a 95% confidence interval, and a minimum power of 0.80, we determined that a minimum of four subjects are required for the FLS training group, three subjects are required for the VBLaST training group, and four subjects are required for the control. Consequently, we recruited seven subjects for the FLS training group, six subjects for the VBLaST training group, and five subjects for the control group. To eliminate any bias due to handedness, all the recruited subjects had no prior skills in laparoscopic surgery and were right handed. Subjects were monetarily compensated for their participation. The statistical software G\*Power was used to determine the effect sizes and the minimum number of subjects required for this study[141].

### 3.2.2 Simulation hardware

Two different simulators were used over the course of this learning curve study. The FLS group trained on a standard SAGES certified FLS box trainer with the official supplementary materials to administer the pattern cutting task. The VBLaST group trained on the VBLaST system, specifically on the pattern cutting module. The VBLaST system consists of two major components: hardware interface, and the simulation software suite. The hardware interface utilizes two PHANTOM Omni haptic devices (Geomagic, Morrisville, North Carolina), connected to appropriate surgical tool interfaces, that provide positional tracking and real-time force feedback in the virtual environment. The simulation software uses custom developed algorithms and software to simulate tool to

cloth interactions in the virtual environment. Figure 3.1 displays both the FLS box trainer and the VBLaST simulator.



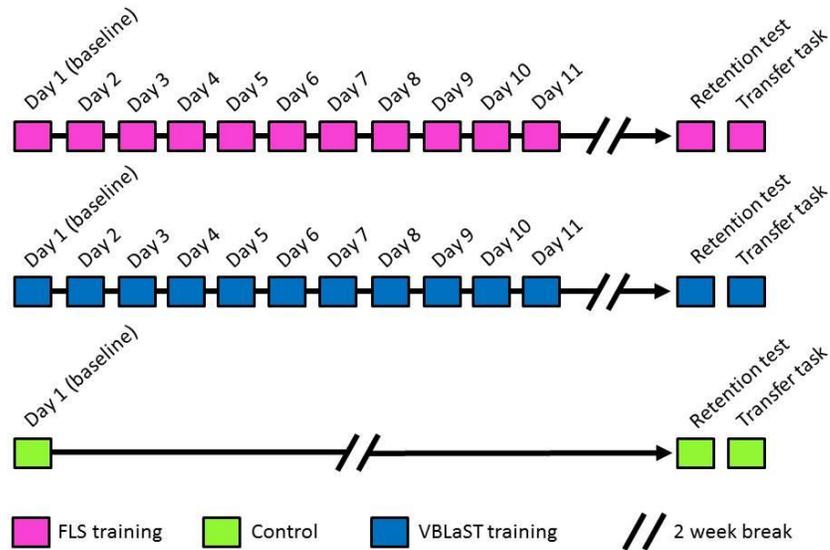
**Figure 3.1: FLS and VBLaST simulators. The physical FLS pattern cutting (PC) box trainer (left) and the VBLaST PC simulator (right) used in this study [54].**

### **3.2.3 Study design, fNIRS processing, and statistical methods**

Information regarding the study design, fNIRS hardware and processing, and statistical methods for this chapter have been previously discussed in section 2.2.

### **3.2.4 Learning curve and task retention study design**

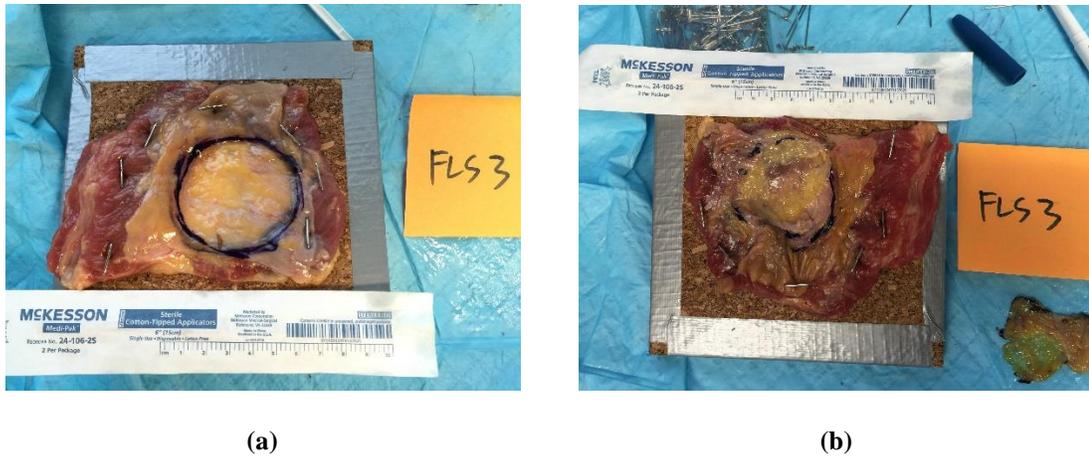
Recruited subjects were split into three groups: FLS training group, VBLaST training group, and control group with no training. All the subjects were given standardized instructions on how to successfully complete the pattern cutting task for the FLS and VBLaST simulators. The untrained control group performed three FLS trials and three VBLaST trials on the first day. The control group then performed three FLS trials and three VBLaST trials as part of the final task retention day without undergoing any laparoscopic skills training. The FLS and VBLaST training groups were instructed to complete up to 10 trials per day for twelve consecutive days on each group's respective simulator. Following twelve days of training, each group was instructed to wait two weeks without undergoing any laparoscopic training before performing three FLS and three VBLaST trials each as part of the final task retention day. A schematic outlining the study design is shown in Figure 3.2.



**Figure 3.2: Schematic illustrating the learning curve study design. Two training groups, VBLaST (blue) and FLS (magenta), undergo a training period whereas the control group (green) only perform the baseline test (Day 1), retention, and transfer task tests.**

### 3.2.5 Transfer task study design

Following the task retention trials, each subject was asked to perform a FLS pattern cutting task on *ex-vivo* cadaveric peritoneal tissue to simulate motor skill transfer from the simulation environment to *ex-vivo* tissue models. The transfer task consisted of replicating the FLS pattern cut task on marked excised cadaveric abdominal tissue samples. The official FLS pattern cutting gauze pads were used as a stencil to draw circles on *ex-vivo* samples to ensure that all the diameters for marked samples remain the same for each sample. Using a standardized set of instructions, the subjects were told to resect the marked peritoneal tissue as accurately and as quickly as possible without damaging the underlying fascia or muscle tissue. Each tissue sample was photographed before and after the completion of the transfer task. Figure 2 shows sample images of before and after the transfer task completion for an example subject.



**Figure 3.3: Pattern cutting transfer task *ex-vivo* sample. (a) *Ex-vivo* peritoneum sample prior to transfer task completion for FLS trained subject 3. (b) Completed pattern cutting transfer task for FLS trained subject 3 with the pattern cutting task replicated and the marked peritoneal tissue resected [54].**

### 3.2.6 Task performance metrics

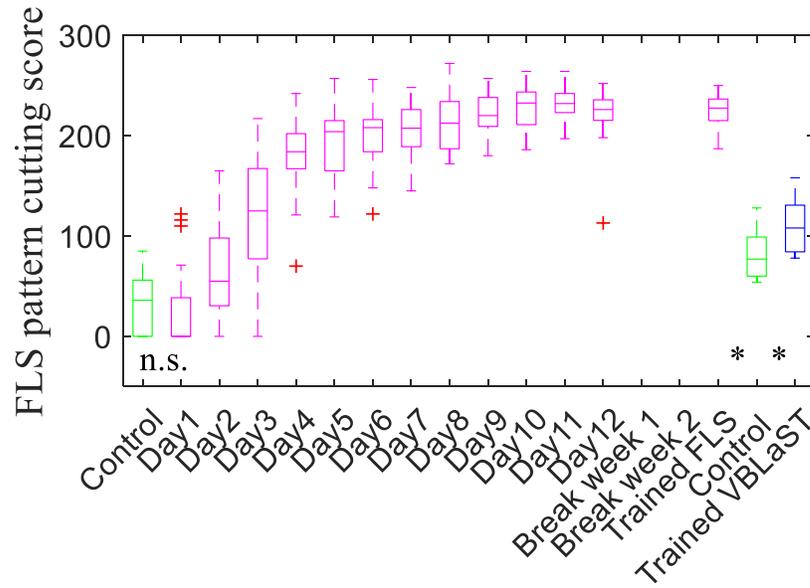
The proprietary FLS scoring metrics for the pattern cutting task was used to manually score each trial for each subject [16]. Each FLS pattern cutting trial completion time was subjectively recorded with an accuracy of  $\pm 1$  second. FLS scoring metrics were obtained from the FLS committee under a non-disclosure agreement, and hence its details cannot be reproduced in this paper. The VBLaST task performance metric reproduces the same undisclosed FLS scoring formulation in the automated VR environment[37]. The FLS and VBLaST pattern cutting performance scores were used as outcomes measure for the learning curve and task retention tests. Since video recording was not allowed according to institute policies at the gross anatomy lab, the performance metric for the *ex-vivo* based transfer task was completion time. Completion time consisted of the total time (minutes) required to completely resect the circle-marked peritoneal tissue from the tissue sample. Each transfer task trial's completion time was subjectively recorded with an accuracy of  $\pm 1$  second.

## 3.3 Results

### 3.3.1 FLS and VBLaST simulator performance learning curves

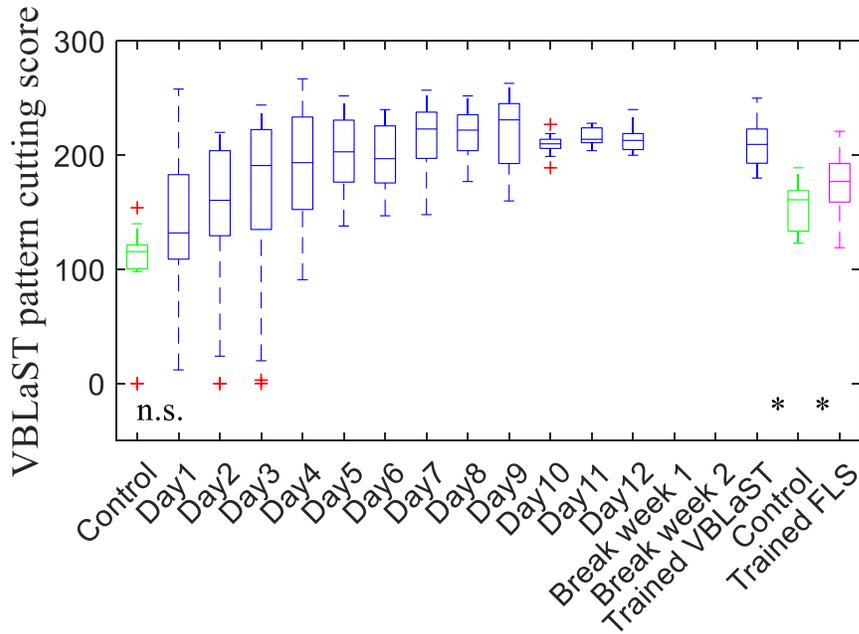
Figure 3.4 shows the FLS pattern cutting performance scores, with respect to training days, for the FLS training and control groups. Results show that there are no significant

differences between the FLS training group and the control group for the first day of training. FLS pattern cutting retention task scores show that both the FLS trained ( $223.5 \pm 18$ ) and VBLaST trained ( $109.6 \pm 26.8$ ) groups significantly outperformed the untrained control group ( $81.5 \pm 25$ ,  $p < 0.05$ ).



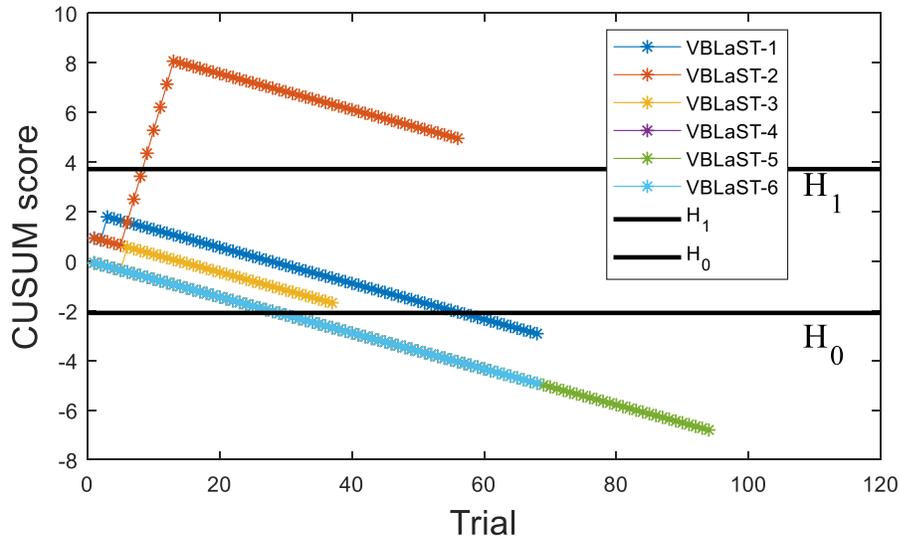
**Figure 3.4:** FLS pattern cutting performance are shown with respect to training day. FLS training students (magenta) are compared to untrained control students (green). FLS pattern cutting task retention scores are shown for trained FLS students (magenta), untrained control subjects (green), and VBLaST trained subjects (blue). Mann-Whitney U tests were used to statistically differentiate the control and FLS training groups (n.s. not significant, \* $p < 0.05$ ).

Figure 3.5 shows the VBLaST pattern cutting performance scores, with respect to training days, for the VBLaST training and control groups. Results indicate that there is no significant differences between the VBLaST training group and the control group for the first day of training. However, VBLaST pattern cutting retention task scores indicate that both the VBLaST trained ( $209.4 \pm 21$ ) and the FLS trained ( $175.2 \pm 26.3$ ) significantly outperformed the untrained control group ( $155 \pm 21.2$ ).



**Figure 3.5: VBLaST pattern cutting performance are shown with respect to training day. VBLaST training students (blue) are compared to untrained control students (green). VBLaST pattern cutting task retention scores are shown for trained VBLaST students (blue), untrained control subjects (green), and FLS trained subjects (magenta). Mann-Whitney U tests were used to statistically differentiate the control and FLS training groups (<sup>n.s.</sup> not significant, \* $p < 0.05$ ).**

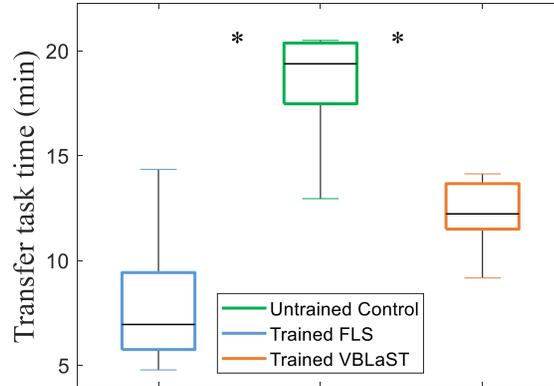
Figure 2.7, as shown previously, details the CUSUM learning curve results for subjects trained in the FLS simulator. Three subjects, FLS2, FLS3, and FLS5, passed the acceptable failure rate of 5% ( $H_0$ ) over the course of the twelve day training period. Specifically, FLS2, FLS3, and FLS5 subjects passed the acceptable failure rate at trials 71, 85, and 85, respectively. Figure 3.6 shows the CUSUM learning curve results for subjects training in the VBLaST simulator where four subjects, VBLaST1, VBLaST4, VBLaST5, and VBLaST6 all passed the acceptable failure rate of 5% ( $H_0$ ) over the course of the training period. Specifically, the four subjects VBLaST1, VBLaST4, VBLaST5, and VBLaST6 passed the acceptable failure rate at trials 57, 29, 29, and 29, respectively.



**Figure 3.6: CUSUM scores for trained FLS and VBLaST groups. CUSUM scores for each subject with respect to number of trials. The threshold score to be considered a senior in the pattern cutting task is 63[16]. (a) CUSUM scores indicate that three subjects (FLS2, FLS3, FLS5) achieved the level of senior during the FLS training period. (b) CUSUM scores indicate that four subjects (VBLaST1, VBLaST4 – 6) achieved the level of senior during the VBLaST training period.**

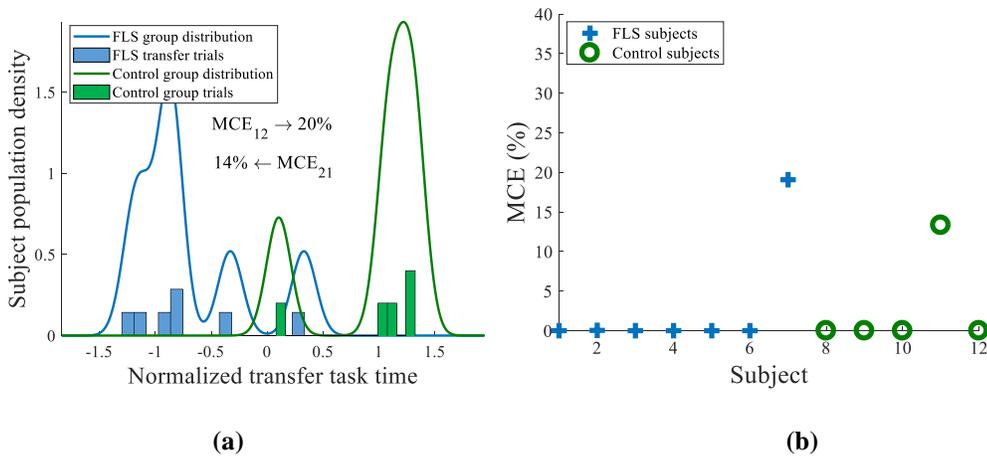
### 3.3.2 Differentiation and classification of motor skill transfer based on traditional task performance

To investigate whether trained subjects significantly outperform untrained subjects in the *ex-vivo* environment, first we report transfer task completion times for trained FLS, trained VBLaST, and untrained control subjects. As shown in Figure 3.7, results indicate that both the trained FLS ( $7.9 \pm 3.3$  min) and trained VBLAST ( $12.2 \pm 1.8$  min) groups completed the transfer task significantly faster than the untrained control group ( $18.3 \pm 3.1$  min,  $p < 0.05$ ). While results show that transfer task time can statistically differentiate trained and untrained subjects during a transfer task, they do not address the accuracy of differentiation.



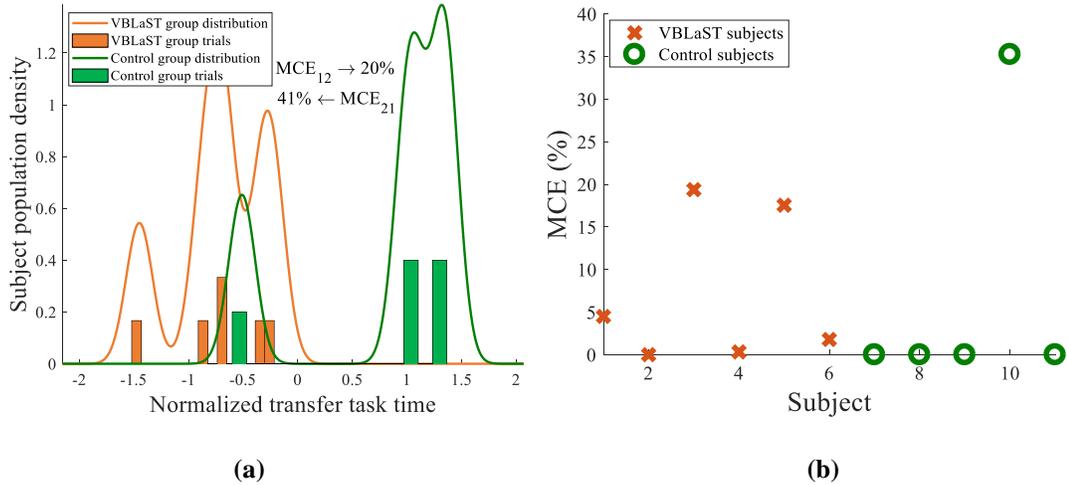
**Figure 3.7: Transfer task completion times for the trained FLS, untrained control, and trained VBLaST subjects (\* $p < 0.05$ ).**

In this context, LDA based classification was used to classify trained and untrained subjects based on completion time. Figure 3.8 shows that classification based on transfer task completion time of trained FLS and untrained control subjects is poor, as shown by high MCEs ( $MCE_1 = 20\%$ ,  $MCE_2 = 14\%$ ). These results indicate that a trained FLS student has a 20% probability of being misclassified as a control subject and an untrained control subject has a 14% probability of misclassified as a FLS trained subjects. Cross-validation results, as seen in Figure 3.8, show that 10/12 or 83% of the samples have MCEs less than 5%, indicating that the classification model is valid for potentially future datasets.



**Figure 3.8: (a) LDA classification of trained FLS and control subjects during the transfer task based on completion times and (b) corresponding crossvalidation results.**

The same classification approach was applied for the virtual simulator trained (VBLaST) subjects vs untrained control subjects as shown in Figure 3.9. Once again, subject classification based on transfer task completion time is poor, indicated by high MCEs ( $MCE_1 = 20\%$ ,  $MCE_2 = 41\%$ ). Furthermore, cross validation results show that 8/11 or 72% of the samples have MCEs less than 5%, as shown in Figure 3.9.

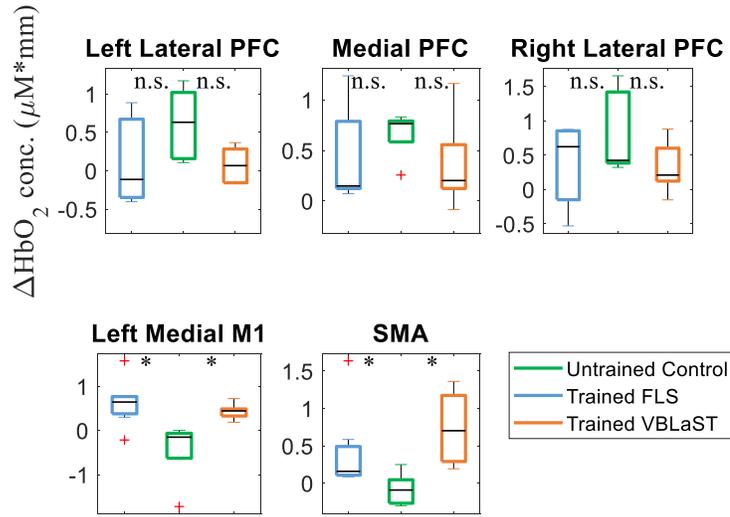


**Figure 3.9:** (a) LDA classification of trained VBLaST and control subjects during the transfer task based on completion times and (b) corresponding crossvalidation results.

### 3.3.3 Neuroimaging-based metrics for differentiation and classification of motor skill transfer

Due to high misclassification errors encountered in assessing transfer task performance based on task time, we propose subject classification based on fNIRS metrics. Prior to classification, we determine if fNIRS is sensitive to changes in cortical activation during the transfer task for trained and untrained subjects, specifically in the prefrontal cortex (PFC), left medial M1, and the SMA. Results indicate that all simulator trained subjects show no significant differences in all PFC cortical regions when compared to control subjects ( $p > 0.05$ ). However, both FLS and VBLaST simulator trained subjects have significantly higher functional activation in the left medial M1 ( $0.64 \pm 0.54$  and  $0.44 \pm 0.18 \Delta HbO_2$  conc.  $\mu M \cdot mm$ , respectively) compared to untrained control subjects ( $-0.44 \pm 0.72 \Delta HbO_2$  conc.  $\mu M \cdot mm$ ,  $p < 0.05$ ). Furthermore, both FLS and VBLaST trained subjects also showed significant increases in functional activation in the SMA ( $0.42 \pm 0.56$  and  $0.74 \pm 0.47 \Delta HbO_2$  conc.  $\mu M \cdot mm$ , respectively) when compared to untrained control

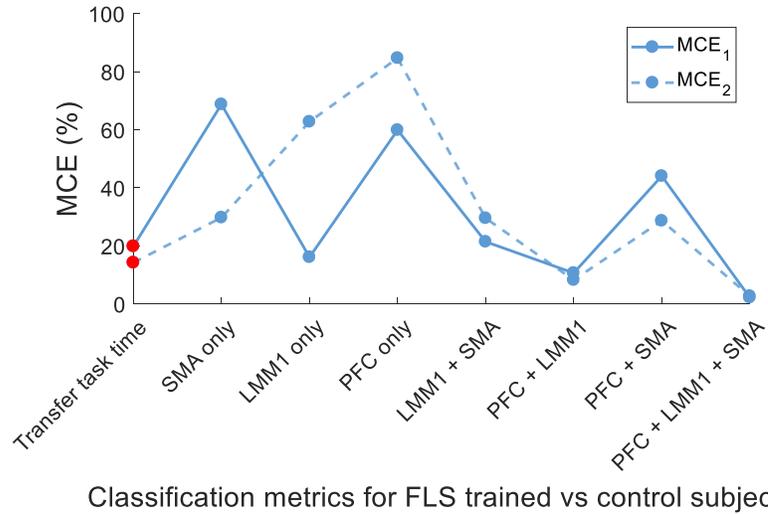
subjects ( $-0.08 \pm 0.22 \Delta\text{HbO}_2$  conc.  $\mu\text{M}^*\text{mm}$ ,  $p < 0.05$ ). These results are summarized in Figure 3.10.



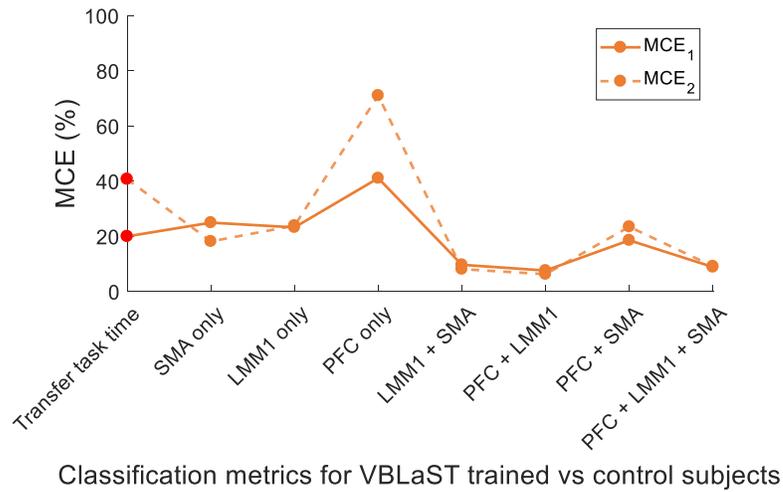
**Figure 3.10: Changes in cortical activation during the transfer task with respect to cortical regions. Average changes in hemoglobin ( $\Delta\text{HbO}_2$ ) concentration as a measure of functional activation with respect to different cortical regions for FLS trained subjects (magenta), untrained control subjects (cyan), and VBLaST trained subjects (black) while all subjects perform the *ex-vivo* transfer task.**

In order to compare the accuracy of subject classification based on transfer task completion time or fNIRS based metrics, several combinations of metrics are used for the classification models. These combinations include transfer task performance time only, PFC only, left medial M1 only, SMA only, left medial M1 + SMA, PFC + left medial M1, PFC + SMA, and PFC + left medial M1 + SMA. Figure 3.11 shows the relative MCEs for various combinations of performance and fNIRS metrics to classify FLS trained subjects from untrained control subjects. The fNIRS metrics combination of PFC + LMM1 + SMA used for the FLS classification model yields very low misclassification errors ( $\text{MCE}_1 = 2.2\%$ ,  $\text{MCE}_2 = 2.7\%$ ). Similarly, the fNIRS metrics combination of PFC + LMM1 + SMA used for the VBLaST classification model yields very low misclassification errors ( $\text{MCE}_1 = 8.9\%$ ,  $\text{MCE}_2 = 9.1\%$ ), as shown in Figure 3.12. Figure 3.13 shows the cross-validation results of various classification models to classify trained FLS or VBLaST subjects with untrained control subjects. FLS trained vs control subjects classification based on transfer performance scores and PFC + LMM1 + SMA combinations yield results where 83% of the samples have misclassification errors less than 0.05. In a similar fashion, VBLaST

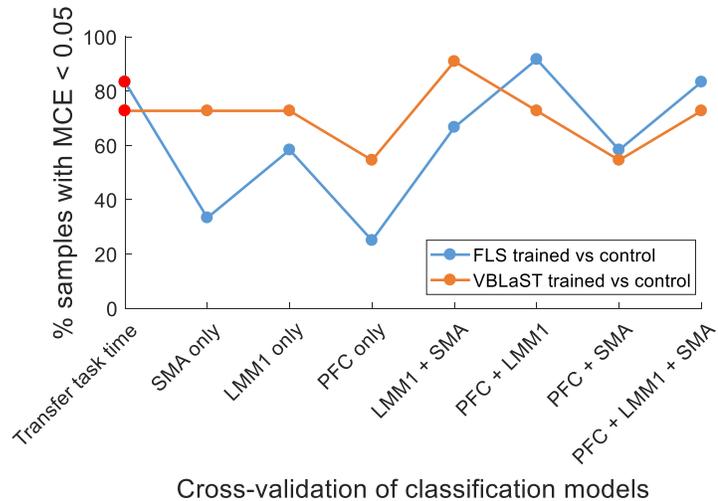
trained vs control subjects classification models show that the transfer task performance score and PFC + LMM1 + SMA metric combinations yield in 72% of the samples with misclassification errors less than 0.05.



**Figure 3.11:** The cumulative set of MCE<sub>1</sub> and MCE<sub>2</sub> for all combinations of fNIRS metrics and the transfer task completion time to classify FLS trained and control subjects.



**Figure 3.12:** The cumulative set of MCE<sub>1</sub> and MCE<sub>2</sub> for all combinations of fNIRS and transfer task metrics to classify VBLaST trained and control subjects.



**Figure 3.13: Leave-one-out crossvalidation results indicate the misclassification errors for each sample treated as an independent sample for the LDA model using all combinations of fNIRs and transfer task metrics, where the percent of samples that have MCE below 0.05 for each possible metric combination are shown.**

### 3.4 Discussion

Accurate and objective assessment of surgical skills transfer from simulation environments to clinical settings is vital in determining the effectiveness of surgical training. Current standards utilizing rating checklists or task completion time metrics are limited in reliability, when objectively determining motor skill transfer to clinical environments [7], [24], [52], [169], [172]. For the first time, we present evidence that a neuroimaging-based approach provides objective assessment of surgical skill transfer from simulation to clinically relevant environments. The results are independent of whether the simulated task was in a physical or a virtual simulator and have been cross-validated to be robust in classifying trained and untrained subjects.

Currently, only task performance scores are used to determine surgical motor skill performance on the FLS and VBLaST trainers. Studies have shown that other measure such as kinematic metrics can also be used as effective measure for assessing surgical skill [15], [173]. However, these metrics focus on the outcomes of task performance instead of assessing the underlying neurological responses to fine motor skills. OUR Neuroimaging based approach provides metrics that can be incorporated into surgical simulator can also provide objective measures of motor skill performance by directly measuring cortical activation during a given task.

Another limitation is the usage of CUSUM scores to objectively measure learning curve outcomes for longitudinal studies. The CUSUM method utilizes a threshold that assigns a binary value of “success” or “failure” trials depending on whether the threshold condition is met. However, many learning curve rates are often nonlinear, and this nonlinearity is not captured in the CUSUM method. Moreover, CUSUM scores utilize arbitrary threshold values that may not directly translate from one simulation environment to another. Traditionally, transfer tasks have been performed on live patients or animal models to show transfer of laparoscopic motor skills from the simulation environment to clinical environments[40], [162], [174], [175]. Due to the complexity and variability of *in-vivo* clinical environments, it is often difficult to standardize the transfer task for each subject. Several of these limitations are addressed by our neuroimaging approach by provide objective assessments of surgical motor skill transfer that are independent from task performance metrics.

It is important to note that the defacto metric used in numerous validation studies to show the surgical skill transfer is performance time [169]. While the results in our study also corroborate the notion that decreases in task performance time are features of expert surgical skills, utilizing this metric alone leads to inconsistencies in literature [24], [169], [176], [177]. This point is further supported by our classification models where task performance time metrics present 20% - 41% misclassification errors indicating a lack of robustness. A univariate metric, such as task performance time, alone may not be an indication of the quality of the task performance [169], [176]. Unfortunately, measures for task quality are not standardized for simulation paradigms and are often subjective in nature, prompting a need for alternative methods such as our neuroimaging based approach [7], [52], [169], [176].

Using fNIRS as a means to measure functional brain activation in real-time, we have shown that FLS and VBLaST trained subjects show significant increases in activation in the left medial M1 and SMA, however no significant differences in the PFC. These regions have been deliberately chosen due to their influence on motor task planning, execution, and fine motor control for complex motor tasks and their critical role in motor skill learning [82], [87], [92]–[94], [139], [140]. Specifically, the PFC is associated with motor strategy and the early stages of motor skill learning. The M1 and SMA are associated with

execution and fine motor control and show increased activation during the later stages of motor skill learning as an indication of procurement of fine motor skills. Our results are consistent with literature findings that indicate that subjects with fine motor skills in complex motor tasks exhibit higher M1 and SMA activation, particularly for bimanual motor tasks [136], [158], [160]. Furthermore, since all the subjects are right handed, majority of the fine motor manipulations employed during the pattern cutting task is via the right hand. Since right handed motor tasks evoke contralateral activation in the left hemisphere of the cortex, we expect increased activity in the left medial M1 [82], [87], [92]–[94], [139], [140]. Although we do not report any significant cortical activation differences between the untrained and trained subjects in the PFC during the transfer task, this is an expected result since all the subjects are expected to recruit the PFC to develop a motor strategy for this unfamiliar task.

Using well-established neurophysiological principles, our work integrates most recent advances in neuroimaging and assessment of surgical competence during transfer of skills from a simulation environment. This work also expands on several technical improvements over previous work involving fNIRS with applications in surgical skill differentiation. Since fNIRS signals are heavily contaminated by superficial tissue, techniques such as dense spatial sampling via diffuse optical tomography and short separation channel regression can be used to further isolate cortical brain activation signals from superficial tissue [80], [81], [178]. Such approaches provide more robust estimations of the underlying hemodynamic responses associated with surgical tasks, which were not reported in previous fNIRS surgical studies. Furthermore, we utilize a multivariate approach to statistically differentiate and classify subjects based on surgical motor skill levels to accurately determine motor skill transfer. Since no single metric itself, such as task completion time, can demonstrate surgical skill proficiency between trained and untrained subjects [169], [176], our fNIRS metrics based multivariate approach on classifying trained and untrained subjects brings robustness in surgical skill transfer assessment.

### **3.5 Summary**

In this chapter, fNIRS is proposed as a non-invasive real-time imaging method to successfully differentiate and classify surgical motor skills that transfer from simulation to *ex-vivo* environments. First, we show that two brain regions, the left medial M1 and the SMA, have significantly increased cortical activation for subjects that have trained on physical and virtual surgical simulators when compared to untrained subjects. We also show that fNIRS based metrics have significantly lower misclassification errors than simple metrics, such as task completion time, when used to classify trained and untrained subjects in assessing surgical motor skill transfer. fNIRS based approaches to objectively quantify motor skill transfer may be a paradigm change for the surgical community in determining the effectiveness of surgical trainers in training technical skills that ultimately transfer to the operating room.

## **4. Brain connectivity analysis for surgical skill assessment in physical and virtual surgical simulators**

### **4.1 Overview**

Surgical training has traditionally followed an apprenticeship based model where technical skills are taught in the operating room [161], [179]. However, this approach is often costly, time consuming, and presents significant negative patient outcomes due to the inexperience of the trainee [161], [179]. Furthermore, with the advent of minimally invasive surgery and laparoscopic procedures, programs such as the Fundamentals of Laparoscopic Surgery (FLS) and the Fundamentals of Endoscopic Surgery (FES) have been adopted by the American Board of Surgery as accredited means for assessing technical surgical skills [16], [17], [20], [23], [162]–[166]. Surgical skill assessment in these simulator-based training methods often utilize rating scales, rudimentary performance metrics, or direct observation methods to rate and assess surgical task performance [1]–[3], [23], [55], [56], [180]. While the usage of these metrics are standard of practice in surgical skill training and assessment, they have been cited to have poor interrater reliability and poor correlation of simulator based performance metrics to clinical outcomes in the operating room [7], [52], [169].

Compounding the lack of robust surgical skill assessment metrics, there is a lack of studies that comprehensively address the underlying neurophysiological responses to varying surgical motor skill levels. Recent studies have shown the potential of non-invasive brain imaging to quantify cortical activation differences for subjects with varying degrees of surgical motor skills [54], [60], [72], [74], [137], [138]. While these studies have shown significant differences in functional activation in the prefrontal cortex (PFC), primary motor cortex (M1) and the supplementary motor area (SMA) due to their direct involvement in motor skill learning [54], [60], [72]–[74], [101], [134]–[138], [140], [160], the underlying functional connectivity between cortical regions that are correlated to surgical motor skills has not been studied.

---

Portions of this chapter have been submitted to:

A. Nemani, X. Intes, and S. De, “Functional brain connectivity distinguished surgical skill learning with surgical simulators,” submitted for publication.

Techniques to quantify brain functional connectivity, such as wavelet coherence (WCO) and wavelet phase coherence (WPCO), have been utilized in multiple fNIRS studies [155], [171], [181]–[186]. Since WCO and WPCO analyses can objectively quantify functional connectivity and strong temporal correlations by determining significantly high common power and phase-locked behavior between two specific cortical channels [155], [187], this approach can strongly address the neurophysiological knowledge gap of surgical motor skill learning and assessment.

The purpose of this study is to determine whether there is significant functional connectivity between the PFC, the M1, and the SMA cortical regions that are correlated to surgical motor skill proficiency. Indeed, we hypothesize that are significant functional connectivity changes in these cortical regions that depend solely on surgical motor skill levels and are independent of task performance metrics or surgical skill assessment platforms, namely physical or virtual surgical simulators. To test this hypothesis, subjects with varying degrees of surgical motor expertise performed a complex surgical training task on physical and virtual simulators while undergoing fNIRS imaging in real time. To quantify functional connectivity between cortical regions, WCO and WPCO were calculated during each trial for each subject as they perform the surgical training task.

## **4.2 Methods**

### **4.2.1 Subject recruitment**

36 subjects were recruited in this IRB approved study conducted at Massachusetts General Hospital and University at Buffalo. The subjects were split into two cohorts. The first cohort included novice and expert surgeons and the second cohort included training medical students. The second cohort was further split into three distinct groups: FLS training group, VBLaST training group, and control group. An *a priori* power analysis, based on two sample t-tests, was completed to determine the minimum number of samples required for both cohorts in this study. Using pilot study data and the power estimation software G\*Power [141], we estimated a conservative effect size,  $d = 1.4$ , for the novice and expert surgeon cohorts. Furthermore, we estimated conservative effect sizes for the FLS and VBLaST training groups,  $d = 5.67$  and  $d = 2.57$ , respectively. With a 95% confidence interval and a minimum power of 0.80, a minimum of 8 subjects each for the

expert and novice surgeon cohort group, four subjects for the FLS training group, three subjects for the VBLaST training group, and four subjects for the control group were estimated for this study. Subject demographics are summarized in Table 4.1 and further details on study design is shown in section 2.2.

**Table 4.1: Study subject demographics and training procedures completed.**

<b>Cohort</b>	<b># of subjects</b>	<b>Mean age</b>	<b>Training / Certification</b>	<b>Average # of laparoscopic procedures</b>	<b># of completed FLS pattern cutting trials</b>	<b># of completed VBLaST pattern cutting trials</b>
Expert surgeon	8	35	Postgraduate year 4-5 or attending surgeons	700	5	5
Novice surgeon	9	31	Postgraduate year 1 – 3	60	5	5
FLS training group	9	25	Medical school year 1-4	0	>100	0
VBLaST training group	8	24	Medical school year 1-4	0	0	>85
Control	5	26	Medical school year 1-4	0	3	3

#### **4.2.2 Hardware, study design, and fNIRS processing**

Information regarding the hardware, study design, and fNIRS processing methodologies for this chapter have been previously discussed in section 2.2. The notable exception is digital filters were not applied to each fNIRS time series data to preserve the entire frequency bandwidth of each channel.

#### **4.2.3 Wavelet coherence and wavelet phase coherence**

To objectively quantify functional connectivity between time series from different cortical regions, we utilize wavelet coherence and wavelet phase coherence metrics. First,

the wavelet transforms for each time series are computed using the Morlet wavelet as defined below [181], [183]:

$$W(s, u) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} \psi\left(\frac{t-u}{s}\right) g(t) dt \quad (4-1)$$

$$\psi(t) = \frac{1}{\sqrt[4]{\pi}} \cdot e^{-i\omega_0 t} \cdot e^{-t^2/2} \quad (4-2)$$

where  $W(s, u)$  is the generated complex wavelet transform,  $g(t)$  is the input time series,  $\Psi(t)$  is the Morlet wavelet scaled by a  $s$  and translated in time by  $u$ ,  $\omega_0$  is the basic frequency, and  $i = \sqrt{-1}$ . Since wavelet scaling is sensitive to oscillations in different frequencies and wavelet translation shown time series spectra evolution over time, it is important to note that the translation from scale to frequency depends on the wavelet type choice [181]. Consequently, the Morlet wavelet, with a basic frequency of  $\omega_0 = 2\pi$ , was chosen as it provides the best time-frequency localization [181], [188]–[191]. Using the wavelet transform values, the complex oscillatory time series can be computed with the following expression [181], [183]:

$$w_k(t_n) = W_k(f, t_n) \cdot e^{i\phi_k(f, t_n)} \quad (4-3)$$

where  $k = 1, 2$ , and  $n$  is each time step in the entire complex time series. Using the transformed oscillatory time series, the wavelet coherence (WCO) as a function of frequency is defined below [181], [183]:

$$WCO(f) = \frac{\left[ \frac{1}{N} \sum_{n=1}^N w_1(t_n) w_2^*(t_n) \right] \left[ \frac{1}{N} \sum_{m=1}^N w_1^*(t_m) w_2(t_m) \right]}{P_1(f) P_2(f)} \quad (4-4)$$

where  $w_1$  and  $w_2$  are complex oscillatory Morlet wavelet transforms of the first and second time series,  $N$  is the total number of time steps of each time series,  $*$  is the complex conjugate, and  $P_1(f)$  and  $P_2(f)$  are the wavelet power at frequency  $f$ . The time averaged wavelet phase coherence (WPCO) is also defined below [181], [183]:

$$WPCO(f) = \sqrt{\langle \cos \Delta\phi(f) \rangle^2 + \langle \sin \Delta\phi(f) \rangle^2} \quad (4-5)$$

$$\langle \cos \Delta\phi(f) \rangle = \frac{1}{N} \sum_{n=1}^N \cos \Delta\phi(f, t_n) \quad (4-6)$$

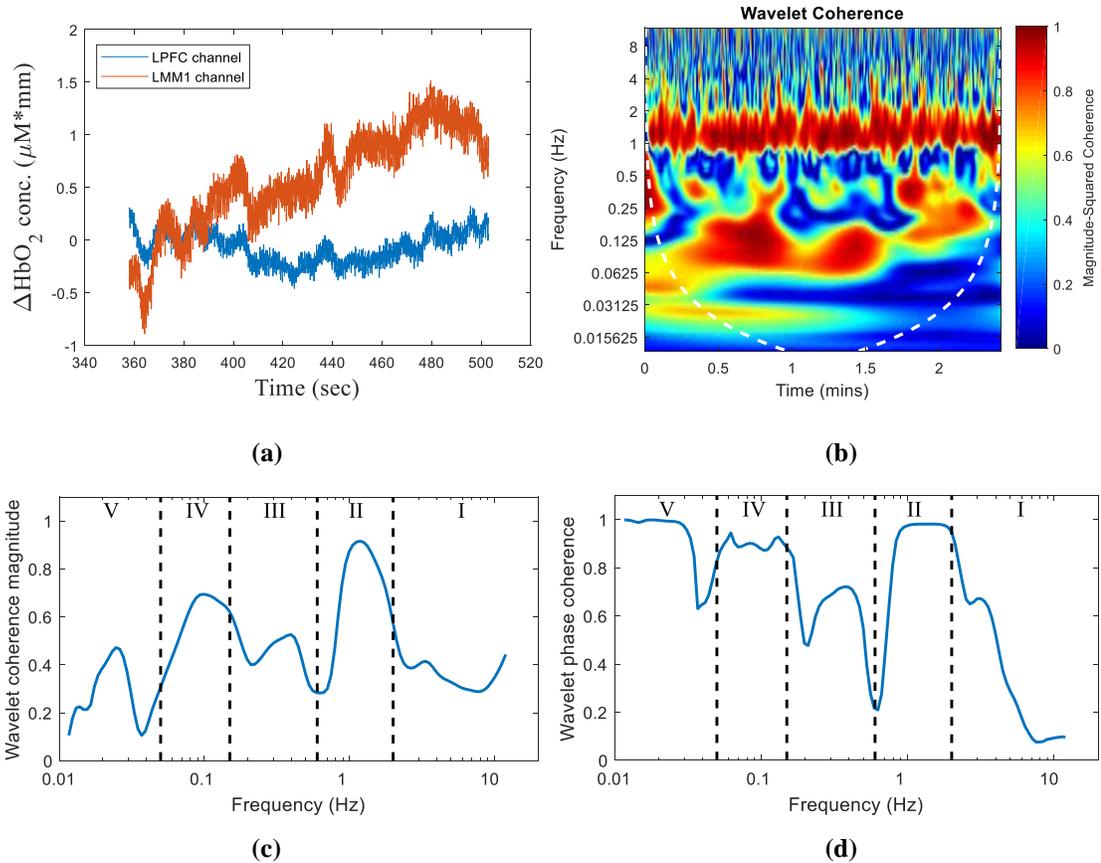
$$\langle \sin \Delta\phi(f) \rangle = \frac{1}{N} \sum_{n=1}^N \sin \Delta\phi(f, t_n) \quad (4-7)$$

where  $\Delta\phi(f, t_n)$  is the instantaneous phase difference between two complex oscillatory time series. The coefficients  $\cos \Delta\phi(f, t_n)$  and  $\sin \Delta\phi(f, t_n)$  are then time averaged across the entire time series. The significance of these metrics is that they can objectively quantify correlations of two independent time series with specificity to frequency and time step changes [181]. A value of 0 for both WCO and WPCO indicates that two time-series are completely unrelated in phase changes and coherence magnitudes. A value of 1 for both WCO and WPCO indicates a significant linear relationship between the two time series and that oscillatory phase changes are significantly correlated [181], [183], [190], [192]. As shown in Table 4.2 below, the entire frequency bandwidth of the resulting WCO and WPCO vectors is split into five different intervals that are correlated to different physiological activities.

Furthermore, a WCO and WPCO analysis is shown for two example fNIRS time series in Figure 4.1. Figure 4.1a shows two example channels, the left lateral PFC and the left medial M1, for one subject while performing the FLS pattern cutting task. Figure 4.1b shows the corresponding wavelet coherence magnitude plot for each frequency and time step between the two example channels. Figure 4.1c-d are the time averaged wavelet coherence and wavelet phase coherence magnitudes. Furthermore, the frequency intervals are depicted to show the specific coherence magnitudes ranges for each associated physiology. For example, the average WCO within the neurovascular coupling range is 0.44 and the WPCO magnitude is 0.94. Note that only WCO and WPCO values within cone of influence, depicted as a shaded white line, are used for analysis due to edge effects that may bias the analysis. Regarding statistical tests, two tailed Mann-Whitney U tests were utilized within a 95% confidence interval to determine significant differences of WCO and WPCO metrics between populations.

**Table 4.2: Frequency bandwidth intervals with their associated physiology [155], [181]–[183], [193].**

Frequency interval	Frequency range (Hz)	Associated physiology
I	0.6 – 2	Cardiac activity
II	0.15 – 0.6	Respiratory activity
III	0.05 – 0.15	Myogenic smooth muscle activity
IV	0.0095 – 0.02	Neurovascular coupling and autonomic control in the cortex
V	0.005 – 0.0095	Nitric oxide related endothelial metabolic activity

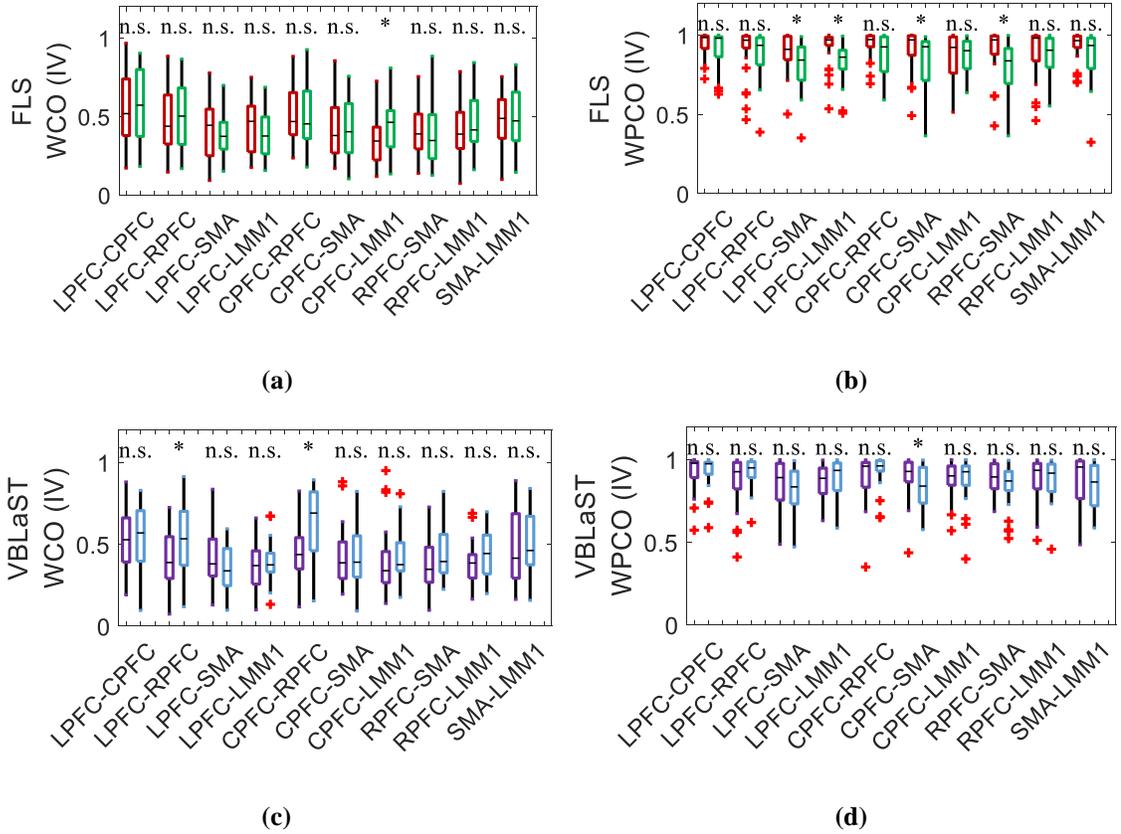


**Figure 4.1: Example wavelet coherence between two different fNIRS time series data. (a) Two time series data from the left lateral PFC (LPFC) and left medial M1 (LMM1) channels for a surgical expert during one FLS task trial (b) Wavelet coherence magnitude between the two time series data based on time and frequency domains. Wavelet coherence magnitude values are shown via the color bar. Only values within the cone of influence range, indicated by a dashed white line, are included for wavelet coherence power magnitude and phase coherence calculations. (c) Time-averaged wavelet coherence magnitudes and (d) wavelet phase coherence magnitudes between the two example time series shown in (a).**

## 4.3 Results

### 4.3.1 Wavelet coherence between surgical experts and novices

To investigate significant functional connectivity differences in cortical regions associated with varying degrees of surgical motor skills, we report, in Figure 4.2, the WCO and WPCO metrics between each combination of the following cortical channels: left lateral prefrontal cortex (LPFC), central prefrontal cortex (CPFC), right lateral prefrontal cortex (RPFC), left medial primary motor cortex (LMM1), and the supplementary motor area (SMA). First, we compare functional connectivity differences between the expert and novice surgeon cohorts during the pattern cutting task for both physical (FLS) and virtual (VBLaST) simulators. Results indicate that WPCO values, between the CPFC and SMA channels in the neuro-coupling activity frequency range (IV), are significantly higher for experts ( $0.913 \pm 0.125$ ) compared to novices ( $0.827 \pm 0.183$ ,  $p = 0.007$ ). While there are other cortical channels that have significant functional connectivity differences between experts and novices, only the CPFC – SMA channels are consistent across the physical FLS and virtual VBLaST simulators.

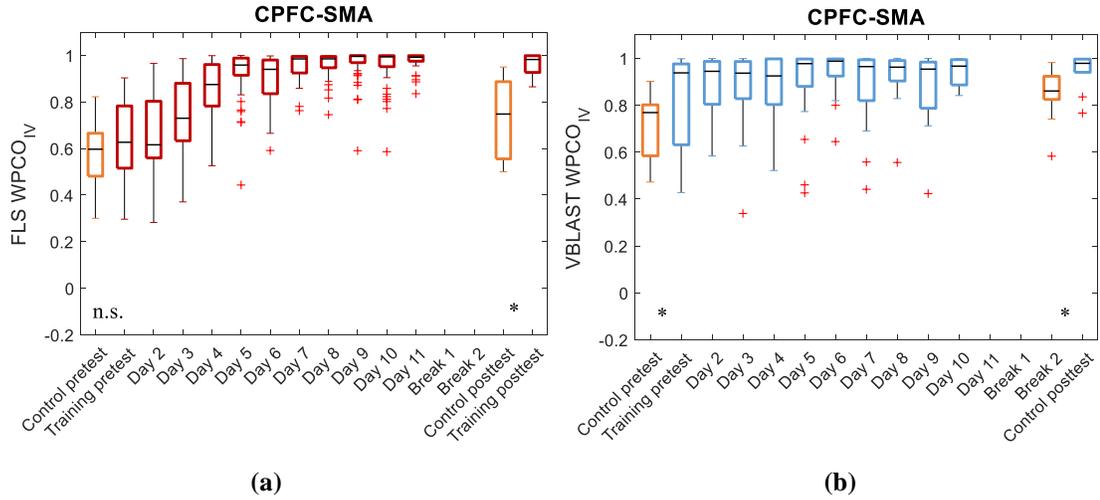


**Figure 4.2: Wavelet coherence and wavelet phase coherence magnitude changes between experts and novices on physical or virtual simulators. (a-b) Wavelet coherence magnitudes and wavelet phase coherence magnitudes for FLS experts (red) vs novices (green) within the neurovascular coupling activity frequency range. (c-d) Wavelet coherence magnitudes and wavelet phase coherence magnitudes for VBLaST experts (purple) vs novices (blue) within the neurovascular coupling activity frequency range.**

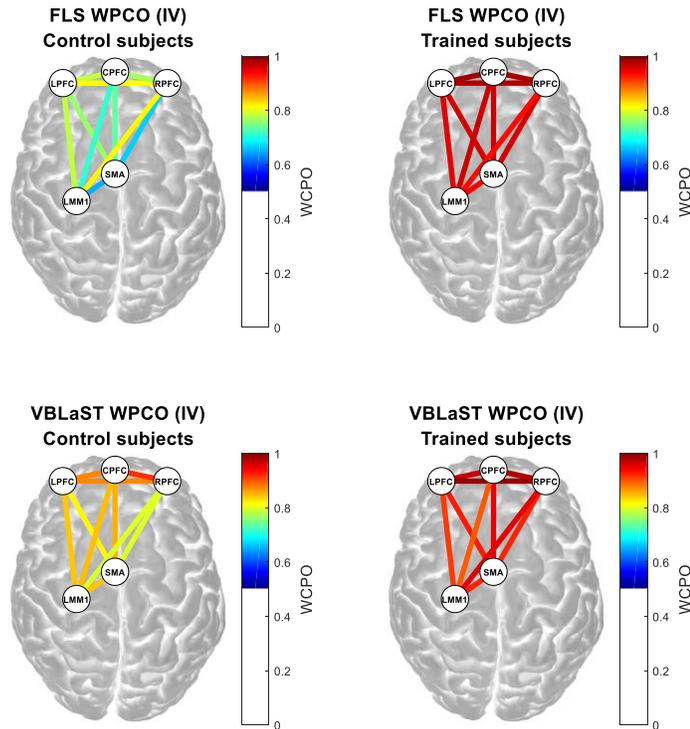
### 4.3.2 Wavelet coherence between surgically trained and untrained subjects

To determine functional connectivity changes with increasing surgical motor skill proficiency due to increased practice, we calculate WCO and WPCO metrics for both the FLS and VBLaST student training groups. Next, these WCO and WPCO metrics are compared in the CPFC and SMA channels with untrained control subjects. Figure 4.3 shows the longitudinal functional connectivity results of the training cohort group. Figure 4.3a shows significant WPCO differences in the neuro-coupling activity frequency range (IV) between untrained control students ( $0.735 \pm 0.177$ ) vs physical simulator trained students ( $0.960 \pm 0.045$ ,  $p < 0.001$ ). Similarly, Figure 4.3b also shows significant WPCO differences in the neuro-coupling activity frequency range between untrained control students ( $0.853 \pm 0.109$ ) vs virtual simulator trained students ( $0.944 \pm 0.079$ ,  $p = 0.0166$ ).

Figure 4.3 shows a visual schematic of the functional activations in the neurovascular coupling frequency range for subjects within the FLS and VBLaST simulator frameworks.



**Figure 4.3: Longitudinal wavelet phase coherence in neurogenic activity frequency range with increasing surgical skill training. WPCO magnitudes within the neurovascular coupling activity frequency range (V) for between the CPFC and SMA channels for untrained control subjects, trained FLS (a) and trained VBLaST (b) subjects as training and motor skill proficiency increases with training days.**



**Figure 4.4: Functional connectivity schematics for training and untrained subjects post training on virtual or physical simulators. Schematic showing the functional connectivity differences, as shown by significant changes in WPCO in the neurogenic activity frequency ranges, between trained and untrained subjects.**

## 4.4 Discussion

While surgical simulators are significantly gaining ground for use in surgical skill training and assessments[161], the underlying neurological mechanisms or functional connectivity between correlated cortical regions are largely unstudied. This study compares the functional connectivity of cortical regions associated with fine motor skills for subjects with varying degrees of surgical motor skill. More specifically, WCO and WPCO in the neurovascular coupling frequency range (0.0095 – 0.02 Hz) between the CPFC and SMA channels yielded significant differences, where experts and trained subjects show sustained increased in functional connectivity compared to novices or untrained subjects.

Functional connectivity within the neuro-coupling frequency ranges between the CPFC and SMA also have underlying physiological mechanisms that drive increased connectivity. Cerebral oscillations within the neuro-coupling frequency range are affected by regional cerebral blood supply. Thus when localized brain activation occurs, cerebral blood flow (CBF) significantly increases [155]. This concept is called functional hyperemia and changes constantly during the surgical task trial stimulus duration to maintain CBF [192], [194]. Furthermore, significant decreases in functional connectivity, as measured by WCO and WPCO, may also be attributed to decreased blood supply due to long-term imbalance of hyperemia [155], particularly for novices or subjects with insufficient motor skill compared to experts. Significant uniform decreases of WCO and WPCO in the prefrontal cortex regions for surgical novices or untrained subjects may indicate overall functional connectivity levels for this cohort group. Cerebral oscillations within the neuro-coupling frequency range may also stem from local activity that is responsible for contractions in vascular smooth muscles [193]. During fine motor tasks, particularly during complex bimanual surgical tasks, the primary motor cortex should maintain increases in excitation to the correlated motor neuron end effector, to ensure steady and sustained bimanual force production [155], [195]. Fine motor skills, particularly employed in this study, required bimanual coordination that significantly employs the primary motor cortex, leading in increased active motor units and stimulus coherence within neuro-coupling range. Ultimately, our results show evidence of this

concept where there are significant increases in WCO and WPCO for experts or surgically trained subjects compared to novices and untrained subjects.

From a practicality standpoint, our neuroimaging approach utilizes the most recent advances in neuroimaging with a focus in functional imaging via fNIRS, such as increased specificity to cortical tissue due to short separation regression [80], [81]. Such approaches provide accurate estimations of the hemodynamics of cortical tissue during complex bimanual surgical tasks for both virtual and physical simulators, which have not been reported previously. There has also been criticism for virtual simulators that state that they lack the robustness in performance assessment compared to physical simulators [161]. Our approach to quantify functional connectivity solely based on neuroimaging based metrics show promise that specific markers of surgical motor skill proficiency, such as WCO and WPCO metrics, can be utilized for future surgical assessment and training. More importantly, our approach is not dependent on task performance metrics or simulation paradigms, which have large inconsistencies in literature due to the lack of standardization and robust metrics in the virtual surgical simulation field [24], [161], [176].

## **4.5 Summary**

In this chapter, functional connectivity metrics, namely WCO and WPCO, were used to quantify significant correlations between connectivity in specific cortical regions to increases in surgical motor skill proficiency. My measuring multiple channels corresponding to cortical regions in the PFC, M1, and SMA, fNIRS imaging shows evidence that increases in functional connectivity between the CPFC and the SMA are positively correlated to increased surgical motor skill. Furthermore, this increase in functional connectivity is significantly higher in surgical experts compared to novices and similarly in simulator trained subjects compared to untrained subjects. Note that functional connectivity increases between the CPFC and SMA are independent of simulation platform or task performance metrics.

## 5. Summary and future work

### 5.1 Thesis summary

Accurate assessment of surgical skills is becoming ever more paramount as new surgical techniques and procedures continue to rely on MIS techniques that are inherently difficult to master and embody complex bimanual tasks. Current surgical motor skill assessment methods, however, have severe limitations such as high subjectivity, poor correlations to patient outcomes, high human resource costs, and inconsistencies in assessment methodologies. Compounding this lack of robustness in skill assessment, the underlying neurophysiological responses to increased surgical skill learning have not been studied in detail. This presents a clear need for more objective and accurate assessment methods for surgical motor skill levels.

In this work, we propose fNIRS to objectively quantify surgical motor skills in real time and compare our approach with established assessment metrics. We show that fNIRS metrics can not only differentiate varying degrees of surgical motor skill proficiency, but also classify subjects according to motor skill levels with significantly higher accuracy than established metrics. More specifically, classification using multivariate metrics including the PFC, M1, and SMA cortical regions drastically reduce misclassification errors from 53 – 61% to 4.2 – 4.4% when compared to established metrics.

In this work, we also propose fNIRS as a non-invasive real-time imaging method to successfully differentiate and classify surgical motor skills that transfer from simulation to *ex-vivo* environments. We present evidence that multivariate metrics from the PFC, M1, and SMA regions while subjects perform a surgical transfer task can significantly reduce misclassification errors from 14 – 41% to 2.2 – 9.1% when compared to conventional metrics. Our approach is significantly more accurate in classifying trained and untrained subjects with respect to surgical motor skill transfer than established metrics. Furthermore, the novelty of our approach is due to independence of task performance metrics or simulation environment.

For the first time, we also evidence that functional connectivity changes are correlated to surgical motor skill proficiency and these connectivity changes depend on specific neurovascular coupling mechanisms in cortical regions and are independent of simulation

environments. Our study shows that surgical experts and surgically trained subjects exhibit significant functional activation correlations between the CPFC and SMA within the neurovascular coupling frequency range. These results further our understanding of functional neural correlates of fine motor skills associated with surgical performance and can be used for future training and assessment paradigms.

This work addresses the limitations of current subjective and underdeveloped surgical skill assessment metrics by proposing fNIRS measures as robust and accurate means of significantly improving surgical skill assessment accuracy. While further studies are required to establish repeatability, non-invasive brain imaging approaches, using near-infrared spectroscopy, present a paradigm change in surgical skill assessment that may drastically improve training efficacy for resident programs that incorporate surgical simulators and ultimately reduce negative patient outcomes.

## **5.2 Future work**

While this work has established the usage of fNIRS metrics as robust means of surgical skill assessment, future research may be undertaken to further improve the specificity, robustness, and application of fNIRS measures for motor skill proficiency assessment.

### **5.2.1 Expansion of cortical coverage**

The most obvious improvement is the expansion of probe measurements to other cortical regions that are correlated to motor skill task performance. The occipital lobe, for example, is crucial in discriminating colors, shapes, and movement that directly impact motor task performance via visuospatial feedback mechanisms for motor output [196], [197]. For example, studies involving unimanual motor task learning have shown evidence of BOLD signal decreases in functional activation within the occipital lobe [197]. Generally, studies addressing visuospatial task learning and performance have utilized the fMRI approach in a univariate manner to report occipital lobe activations. Functional activation measurements within this region can be incorporated in future studies to potentially increase the robustness of classification of surgical motor skill levels. Note that current studies using fNIRS for surgical skill assessment, including this work, only address

the motor skill performance aspect of surgical technical skills and do not address neural correlates of feedback mechanisms for visuospatial tasks. By expanding cortical measurements to the occipital lobe, it may be possible to further increase the robustness of subject classification based on surgical motor skill levels.

### **5.2.2 High density probe measurements**

The biggest limitation of fNIRS approaches to brain imaging is the limited spatial resolution. fNIRS imaging studies, including this work, typically utilize sparse arrangements of source-detector channels to ensure sufficient depth penetration, while maintaining moderate spatial resolution [198]. However, fNIRS spatial resolution is still significantly lower than fMRI approaches and recent proposal of high density probe measurements have shown promising results to shorten this gap in spatial resolutions between fMRI and fNIRS [178], [199]. High density probe measurements utilize significantly larger amounts of sources and detectors and can range up to 1200 separate channels that span across the cerebrum during functional studies. This approach presents multiple advantages that include significantly larger field of views, spatial resolution on the order of 13mm, and high specificity to activation on the gyrus scale. While these advances are promising, several limitations to this approach as also evident, such as increased cross-talk between channel measurements posing a significant signal encoding challenge. Furthermore, the inclusion of over 90 source and detectors probes to achieve over 1200 measurement channels further complicates accurate probe placement and ease of use for clinical applications.

### **5.2.3 Prediction of motor skill levels**

Another potential area of work that has not been addressed in literature is the usage of cortical measurements to predict motor skill learning curves for tailored learning programs. In this work, learning curves have been quantified using the CUSUM method to establish whether subjects have achieved a minimum proficiency level for surgical skill expertise. However, this approach does not account for the rate of motor skill learning or provide any information for accurate modeling of learning curve rates and motor skill plateaus for each subject. By utilizing regression methods such as partial least squares, or

non-linear regression it may be possible to utilize fNIRS metrics to predict plateaus in motor skill learning. This approach can also be expanded with the usage of multivariate convolutional neural networks (CNNs) to accurately predict motor skill learning curves using training datasets comprised of varying surgical motor skill learning curves with a wide variety of plateau ranges and learning rates. Ultimately, this approach may provide specific training programs for subjects that may significantly reduce required training times, and reduce surgical training program costs.

#### **5.2.4 Brain imaging for non-technical surgical skills**

Of course, bimanual motor skills are only a piece of the whole concept of surgical expertise. Non-technical skills such as cognitive decision making, communication, stress management, teamwork, and confidence all play a significant role in differentiating surgical expertise [200]. Non-technical skills are inherently difficult to objectively assess yet studies have cited that almost 43% of errors made in surgery are due to communication, cognitive, and diagnostic errors [200], [201]. While the significance of learning non-technical skills is clearly apparent, current studies focus on rating based checklists or subjective metrics to assess non-technical skills. This presents a significant opportunity to utilize non-invasive brain imaging approaches to quantify non-technical surgical skills.

Cortical regions such as the prefrontal cortex, bilateral anterior insula, and the anterior cingulate cortex are heavily implicated in the decision-making process during a given surgical procedure [202]. This work has already shown the utility of imaging the PFC specifically for surgical motor skill learning, and can be expanded for future studies involving efficiencies in decision making and communication. Furthermore, coupled imaging modalities, such as fMRI-fNIRS systems, would allow simultaneous measurements of the PFC, and deeper structures such as the hypothalamus, and basal ganglia to study their role in improvements in non-technical surgical skills over time.

## REFERENCES

- [1] R. Aggarwal *et al.*, “An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room,” *Ann. Surg.*, vol. 245, no. 6, pp. 992–999, Jun. 2007.
- [2] M. C. Vassiliou *et al.*, “A global assessment tool for evaluation of intraoperative laparoscopic skills,” *Am. J. Surg.*, vol. 190, no. 1, pp. 107–113, Jul. 2005.
- [3] J. D. Doyle, E. M. Webber, and R. S. Sidhu, “A universal global rating scale for the evaluation of technical skills in the operating room,” *Am. J. Surg.*, vol. 193, no. 5, pp. 551–555, May 2007.
- [4] J. A. Martin *et al.*, “Objective structured assessment of technical skill (OSATS) for surgical residents,” *Br. J. Surg.*, vol. 84, no. 2, pp. 273–278, Feb. 1997.
- [5] R. K. Reznick and H. MacRae, “Teaching surgical skills - changes in the wind,” *N. Engl. J. Med.*, vol. 355, no. 25, pp. 2664–2669, Dec. 2006.
- [6] R. Reznick, G. Regehr, H. MacRae, J. Martin, and W. McCulloch, “Testing technical skill via an innovative ‘bench station’ examination,” *Am. J. Surg.*, vol. 173, no. 3, pp. 226–230, Mar. 1997.
- [7] K. Moorthy and Y. Munz, “Objective assessment of technical skills in surgery,” *Br. Med. J.*, vol. 327, no. 7422, pp. 1032–1037, Nov. 2003.
- [8] O. H. Chang, L. P. King, A. M. Modest, and H.-C. Hur, “Developing an objective structured assessment of technical skills for laparoscopic suturing and intracorporeal knot tying,” *J. Surg. Educ.*, vol. 73, no. 2, pp. 258–263, Mar. 2016.
- [9] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, “Review of methods for objective surgical skill evaluation,” *Surg. Endosc.*, vol. 25, no. 2, pp. 356–366, Feb. 2011.
- [10] R. S. Sidhu, E. D. Grober, L. J. Musselman, and R. K. Reznick, “Assessing competency in surgery: Where to begin?,” *Surgery*, vol. 135, no. 1, pp. 6–20, Jan. 2004.
- [11] N. K. Francis, G. B. Hanna, and A. Cuschieri, “The performance of master surgeons on the advanced dundee endoscopic psychomotor tester: Contrast validity study,” *Arch. Surg.*, vol. 137, no. 7, pp. 841–844, Jul. 2002.
- [12] V. Datta, S. Mackay, M. Mandalia, and A. Darzi, “The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model,” *J. Am. Coll. Surg.*, vol. 193, no. 5, pp. 479–485, Nov. 2001.
- [13] G. Megali, S. Sinigaglia, O. Tonet, and P. Dario, “Modelling and evaluation of

- surgical performance using hidden markov models,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1911–1919, Oct. 2006.
- [14] J. Rosen, J. D. Brown, L. Chang, M. Barreca, M. Sinanan, and B. Hannaford, “The BlueDRAGON - a system for measuring the kinematics and dynamics of minimally invasive surgical tools in-vivo,” in *Proceedings 2002 IEEE International Conference on Robotics and Automation*, 2002, vol. 2, pp. 1876–1881.
- [15] J. Rosen, J. D. Brown, L. Chang, M. N. Sinanan, and B. Hannaford, “Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 399–413, Mar. 2006.
- [16] S. A. Fraser, D. R. Klassen, L. S. Feldman, G. A. Ghitulescu, D. Stanbridge, and G. M. Fried, “Evaluating laparoscopic skills: Setting the pass/fail score for the MISTELS system,” *Surg. Endosc.*, vol. 17, no. 6, pp. 964–967, Jun. 2003.
- [17] S. A. Fraser, L. S. Feldman, D. Stanbridge, and G. M. Fried, “Characterizing the learning curve for a basic laparoscopic drill,” *Surg. Endosc.*, vol. 19, no. 12, pp. 1572–1578, Dec. 2005.
- [18] M. C. Vassiliou *et al.*, “The MISTELS program to measure technical skill in laparoscopic surgery : Evidence for reliability,” *Surg. Endosc.*, vol. 20, no. 5, pp. 744–747, May 2006.
- [19] L. S. Feldman, S. E. Hagarty, G. Ghitulescu, D. Stanbridge, and G. M. Fried, “Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents,” *J. Am. Coll. Surg.*, vol. 198, no. 1, pp. 105–110, Jan. 2004.
- [20] G. M. Fried *et al.*, “Proving the value of simulation in laparoscopic surgery,” *Ann. Surg.*, vol. 240, no. 3, pp. 518–525, Sep. 2004.
- [21] N. J. Soper and G. M. Fried, “The fundamentals of laparoscopic surgery: Its time has come,” *Bull. Am. Coll. Surg.*, vol. 93, no. 9, pp. 30–32, Sep. 2008.
- [22] J. H. Peters *et al.*, “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.
- [23] G. M. Fried, “FLS assessment of competency using simulated laparoscopic tasks,” *J. Gastrointest. Surg.*, vol. 12, no. 2, pp. 210–212, Feb. 2008.
- [24] N. E. Seymour *et al.*, “Virtual reality training improves operating room performance: Results of a randomized, double-blinded study,” *Ann. Surg.*, vol. 236, no. 4, pp. 458–463, Oct. 2002.
- [25] J. R. Korndorffer, R. Seirra, J. B. Dunne, and D. Scott, “Simulator training for

- laparoscopic suturing using performance goals translates to the OR,” *J. Am. Coll. Surg.*, vol. 199, no. 3, pp. 73–74, Sep. 2004.
- [26] V. S. Arikatla *et al.*, “Face and construct validation of a virtual peg transfer simulator,” *Surg. Endosc.*, vol. 27, no. 5, pp. 1721–1729, May 2013.
- [27] A. G. Gallagher *et al.*, “Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training,” *Ann. Surg.*, vol. 241, no. 2, pp. 364–372, Feb. 2005.
- [28] D. Stefanidis, J. R. Korndorffer, R. Sierra, C. Touchard, J. B. Dunne, and D. J. Scott, “Skill retention following proficiency-based laparoscopic simulator training,” *Surgery*, vol. 138, no. 2, pp. 165–170, Aug. 2005.
- [29] R. Aggarwal, T. Grantcharov, K. Moorthy, J. Hance, and A. Darzi, “A competency based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill,” *Am. J. Surg.*, vol. 191, no. 1, pp. 128–133, Jan. 2006.
- [30] G. B. Healy, “The college should be instrumental in adapting simulators to education,” *Bull. Am. Coll. Surg.*, vol. 87, no. 11, pp. 10–11, Nov. 2002.
- [31] Limbs & Things, “Specialist Skills | FLS.” [Online]. Available: <https://www.limbsandthings.com/us/our-products/details/f-l-s-fundamentals-of-laparoscopic-surgery-trainer-system>. Accessed on: Nov. 9, 2017.
- [32] A. J. Debes, R. Aggarwal, I. Balasundaram, and M. B. Jacobsen, “A tale of two trainers: Virtual reality versus a video trainer for acquisition of basic laparoscopic skills,” *Am. J. Surg.*, vol. 199, no. 6, pp. 840–845, Jun. 2010.
- [33] M. S. Wilson, A. Middlebrook, C. Sutton, R. Stone, and R. F. McCloy, “MIST VR: A virtual reality trainer for laparoscopic surgery assesses performance,” *Ann. R. Coll. Surg. Engl.*, vol. 79, no. 6, pp. 403–404, Nov. 1997.
- [34] L. Zhang *et al.*, “Characterizing the learning curve of the VBLaST-PT(©) (Virtual basic laparoscopic skill trainer),” *Surg. Endosc.*, vol. 27, no. 10, pp. 3603–3615, Oct. 2013.
- [35] A. Maciel, Y. Liu, W. Ahn, T. P. Singh, W. Dunnican, and S. De, “Development of the VBLaST: A virtual basic laparoscopic skill trainer,” *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 4, no. 2, pp. 131–138, Jun. 2008.
- [36] G. Sankaranarayanan *et al.*, “Preliminary face and construct validation study of a virtual basic laparoscopic skill trainer,” *J. Laparoendosc. Adv. Surg. Tech.*, vol. 20, no. 2, pp. 153–157, Mar. 2010.
- [37] A. Chellali *et al.*, “Preliminary evaluation of the pattern cutting and the ligating loop virtual laparoscopic trainers,” *Surg. Endosc.*, vol. 29, no. 4, pp. 815–821, Apr. 2015.

- [38] I. Badash, K. Burt, C. A. Solorzano, and J. N. Carey, "Innovations in surgery simulation: A review of past, current and future techniques," *Ann. Transl. Med.*, vol. 4, no. 23, p. 453, Dec. 2016.
- [39] M. Alaker, G. R. Wynn, and T. Arulampalam, "Virtual reality training in laparoscopic surgery: A systematic review & meta-analysis," *Int. J. Surg.*, vol. 29, pp. 85–94, May 2016.
- [40] A. Hyltander, E. Liljegren, P. H. Rhodin, and H. Lönroth, "The transfer of basic skills learned in a laparoscopic simulator to the operating room," *Surg. Endosc.*, vol. 16, no. 9, pp. 1324–1328, Sep. 2002.
- [41] P. S. Kundhal and T. P. Grantcharov, "Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room," *Surg. Endosc.*, vol. 23, no. 3, pp. 645–649, Mar. 2009.
- [42] A. Moglia, V. Ferrari, L. Morelli, M. Ferrari, F. Mosca, and A. Cuschieri, "A systematic review of virtual reality simulators for robot-assisted surgery," *Eur. Urol.*, vol. 69, no. 6, pp. 1065–1080, Jun. 2016.
- [43] S. Arora, N. Sevdalis, D. Nestel, M. Woloshynowych, A. Darzi, and R. Kneebone, "The impact of stress on surgical performance: A systematic review of the literature," *Surgery*, vol. 147, no. 3, pp. 318–330, Mar. 2010.
- [44] T. Tien, P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G.-Z. Yang, and A. Darzi, "Eye tracking for skills assessment and training: A systematic review," *J. Surg. Res.*, vol. 191, no. 1, pp. 169–178, Sep. 2014.
- [45] R. S. A. Khan, G. Tien, M. S. Atkins, B. Zheng, O. N. M. Panton, and A. T. Meneghetti, "Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation?," *Surg. Endosc.*, vol. 26, no. 12, pp. 3536–3540, Dec. 2012.
- [46] J. L. Marquard, P. L. Henneman, Z. He, J. Jo, D. L. Fisher, and E. A. Henneman, "Nurses' behaviors and visual scanning patterns may reduce patient identification errors," *J. Exp. Psychol. Appl.*, vol. 17, no. 3, pp. 247–256, Sep. 2011.
- [47] A. Darzi, S. Smith, and N. Taffinder, "Assessing operative skill: Needs to become more objective," *Br. Med. J.*, vol. 318, no. 7188, pp. 887–888, Apr. 1999.
- [48] K. R. Wanzel, S. J. Hamstra, D. J. Anastakis, E. D. Matsumoto, and M. D. Cusimano, "Effect of visual-spatial ability on learning of spatially-complex surgical skills," *Lancet*, vol. 359, no. 9302, pp. 230–231, Jan. 2002.
- [49] C. Bosecker, L. Dipietro, B. Volpe, and H. I. Krebs, "Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke," *Neurorehabil. Neural Repair*, vol. 24, no. 1, pp. 62–69, Jan. 2010.

- [50] J. Shah and A. Darzi, "Surgical skills assessment: An ongoing debate," *BJU Int.*, vol. 88, no. 7, pp. 655–660, Dec. 2001.
- [51] C. E. Buckley, D. O. Kavanagh, O. Traynor, and P. C. Neary, "Is the skillset obtained in surgical simulation transferable to the operating theatre?," *Am. J. Surg.*, vol. 207, no. 1, pp. 146–157, Jan. 2014.
- [52] N. J. Hogle *et al.*, "Validation of laparoscopic surgical skills training outside the operating room: A long road," *Surg. Endosc.*, vol. 23, no. 7, pp. 1476–1482, Jul. 2009.
- [53] L. S. Feldman, V. Sherman, and G. M. Fried, "Using simulators to assess laparoscopic competence: Ready for widespread use?," *Surgery*, vol. 135, no. 1, pp. 28–42, 2004.
- [54] A. Nemani, W. Ahn, C. Cooper, S. Schwaitzberg, and S. De, "Convergent validation and transfer of learning studies of a virtual reality-based pattern cutting simulator," *Surg. Endosc.*, to be published. doi: 10.1007/s00464-017-5802-8.
- [55] B. Zendejas, R. K. Ruparel, and D. A. Cook, "Validity evidence for the fundamentals of laparoscopic surgery (FLS) program as an assessment tool: A systematic review," *Surg. Endosc.*, vol. 30, no. 2, pp. 512–20, Feb. 2016.
- [56] B. Zheng, H.-C. Hur, S. Johnson, and L. L. Swanström, "Validity of using fundamentals of laparoscopic surgery (FLS) program to assess laparoscopic competence for gynecologists," *Surg. Endosc.*, vol. 24, no. 1, pp. 152–160, Jan. 2010.
- [57] S. D. Schwaitzberg *et al.*, "Threefold increased bile duct injury rate is associated with less surgeon experience in an insurance claims database," *Surg. Endosc.*, vol. 28, no. 11, pp. 3068–3073, Nov. 2014.
- [58] A. Chaudhry, C. Sutton, J. Wood, R. Stone, and R. McCloy, "Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: Quality of human-computer interface," *Ann. R. Coll. Surg. Engl.*, vol. 81, no. 4, pp. 281–286, Jul. 1999.
- [59] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10-20 system electrode placement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 4, pp. 376–382, Apr. 1987.
- [60] H. N. Modi *et al.*, "A decade of imaging surgeons' brain function (part I): Terminology, techniques, and clinical translation," *Surgery*, vol. 28, no. 0, pp. 2189–2198, Aug. 2017.
- [61] B. Crosson *et al.*, "Functional imaging and related techniques: An introduction for rehabilitation researchers," *J. Rehabil. Res. Dev.*, vol. 47, no. 2, pp. 7–34, 2010.

- [62] D. Cohen, "Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents," *Science*, vol. 161, no. 3843, pp. 784–786, Aug. 1968.
- [63] S. Ogawa *et al.*, "Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: A comparison of signal characteristics with a biophysical model," *Biophys. J.*, vol. 64, no. 3, pp. 803–812, Mar. 1993.
- [64] R. K. Mehta and R. Parasuraman, "Neuroergonomics: A review of applications to physical and cognitive work," *Front. Hum. Neurosci.*, vol. 7, no. 889, pp. 1–10, Dec. 2013.
- [65] F. F. Jöbsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science*, vol. 198, no. 4323, pp. 1264–1267, Dec. 1977.
- [66] D. A. Boas, A. M. Dale, and M. A. Franceschini, "Diffuse optical imaging of brain activation: Approaches to optimizing image sensitivity, resolution, and accuracy," *Neuroimage*, vol. 23 Suppl 1, pp. 275–288, Jan. 2004.
- [67] S. Prahl, "Optical Absorption of Hemoglobin," 1999. [Online]. Available: <http://omlc.org/spectra/hemoglobin/index.html>. Accessed on Oct. 22, 2017.
- [68] M. Cope and D. T. Delpy, "System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination," *Med. Biol. Eng. Comput.*, vol. 26, no. 3, pp. 289–294, May 1988.
- [69] D. T. Delpy, M. Cope, P. van der Zee, S. Arridge, S. Wray, and J. Wyatt, "Estimation of optical pathlength through tissue from direct time of flight measurement," *Phys. Med. Biol.*, vol. 33, no. 12, pp. 1433–1442, Dec. 1988.
- [70] F. Scholkmann *et al.*, "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *Neuroimage*, vol. 85, pp. 6–27, Jan. 2014.
- [71] A. Pellicer and M. del C. Bravo, "Near-infrared spectroscopy: A methodology-focused review," *Semin. Fetal Neonatal Med.*, vol. 16, no. 1, pp. 42–49, Feb. 2011.
- [72] D. R. Leff *et al.*, "Changes in prefrontal cortical behaviour depend upon familiarity on a bimanual co-ordination task: An fNIRS study," *Neuroimage*, vol. 39, no. 2, pp. 805–813, Jan. 2008.
- [73] K. Ohuchida *et al.*, "The frontal cortex is activated during learning of endoscopic procedures," *Surg. Endosc.*, vol. 23, no. 10, pp. 2296–2301, Oct. 2009.
- [74] D. R. C. James *et al.*, "The ergonomics of natural orifice transluminal endoscopic surgery (NOTES) navigation in terms of performance, stress, and cognitive behavior," *Surgery*, vol. 149, no. 4, pp. 525–533, Apr. 2011.

- [75] B. T. Crewther *et al.*, “Skill acquisition and stress adaptations following laparoscopic surgery training and detraining in novice surgeons,” *Surg. Endosc.*, vol. 30, no. 7, pp. 2961–2968, Jul. 2016.
- [76] J. Andreu-Perez, D. R. Leff, K. Shetty, A. Darzi, and G.-Z. Yang, “Disparity in frontal lobe connectivity on a complex bimanual motor task aids in classification of operator skill level,” *Brain Connect.*, vol. 6, no. 5, pp. 375–388, Jun. 2016.
- [77] P. A. Shewokis, F. U. Shariff, Y. Liu, H. Ayaz, A. Castellanos, and D. S. Lind, “Acquisition, retention and transfer of simulated laparoscopic tasks using fNIR and a contextual interference paradigm,” *Am. J. Surg.*, vol. 213, no. 2, pp. 336–345, Feb. 2017.
- [78] H. N. Modi, H. Singh, G.-Z. Yang, A. Darzi, and D. R. Leff, “A decade of imaging surgeons’ brain function (part II): A systematic review of applications for technical and nontechnical skills assessment,” *Surgery*, vol. 162, no. 5, pp. 1130–1139, Nov. 2017.
- [79] S. Tak and J. C. Ye, “Statistical analysis of fNIRS data: A comprehensive review,” *Neuroimage*, vol. 85, pp. 72–91, Jan. 2014.
- [80] L. Gagnon, M. A. Yücel, D. A. Boas, and R. J. Cooper, “Further improvement in reducing superficial contamination in NIRS using double short separation measurements,” *Neuroimage*, vol. 85, pp. 127–135, Jan. 2014.
- [81] L. Gagnon, R. J. Cooper, M. A. Yücel, K. L. Perdue, D. N. Greve, and D. A. Boas, “Short separation channel location impacts the performance of short channel regression in NIRS,” *Neuroimage*, vol. 59, no. 3, pp. 2518–2528, Mar. 2012.
- [82] O. Hikosaka, K. Nakamura, K. Sakai, and H. Nakahara, “Central mechanisms of motor skill learning,” *Curr. Opin. Neurobiol.*, vol. 12, no. 2, pp. 217–222, Apr. 2002.
- [83] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain*, 4th ed. Philadelphia, PA, USA: Wolters Kluwer, 2016.
- [84] D. Purves *et al.*, *Neuroscience*, 4th ed. Sunderland, MA, USA: Sinauer Associates, 2008.
- [85] D. B. Willingham, “A neuropsychological theory of motor skill learning,” *Psychol. Rev.*, vol. 105, no. 3, pp. 558–584, Jul. 1998.
- [86] O. Hikosaka, M. K. Rand, S. Miyachi, and K. Miyashita, “Learning of sequential movements in the monkey: Process of learning and retention of memory,” *J. Neurophysiol.*, vol. 74, no. 4, pp. 1652–1661, Oct. 1995.
- [87] K. Nakamura, K. Sakai, and O. Hikosaka, “Neuronal activity in medial frontal cortex during learning of sequential procedures,” *J. Neurophysiol.*, vol. 80, no. 5,

pp. 2671–2687, Nov. 1998.

- [88] K. Nakamura, K. Sakai, and O. Hikosaka, “Effects of local inactivation of monkey medial frontal cortex in learning of sequential procedures,” *J. Neurophysiol.*, vol. 82, no. 2, pp. 1063–1068, Aug. 1999.
- [89] E. Procyk, Y. L. Tanaka, and J. P. Joseph, “Anterior cingulate activity during routine and non-routine sequential behaviors in macaques,” *Nat. Neurosci.*, vol. 3, no. 5, pp. 502–508, May 2000.
- [90] M. S. Rioult-Pedotti, D. Friedman, and J. P. Donoghue, “Learning-induced LTP in neocortex,” *Science*, vol. 290, no. 5491, pp. 533–536, Oct. 2000.
- [91] C. S. Li, C. Padoa-Schioppa, and E. Bizzi, “Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field,” *Neuron*, vol. 30, no. 2, pp. 593–607, May 2001.
- [92] A. Y. Klintsova and W. T. Greenough, “Synaptic plasticity in cortical systems,” *Curr. Opin. Neurobiol.*, vol. 9, no. 2, pp. 203–208, Apr. 1999.
- [93] K. Sakai, O. Hikosaka, S. Miyauchi, R. Takino, Y. Sasaki, and B. Pütz, “Transition of brain activation from frontal to parietal areas in visuomotor sequence learning,” *J. Neurosci.*, vol. 18, no. 5, pp. 1827–1840, Mar. 1998.
- [94] M. Honda, “Dynamic cortical involvement in implicit and explicit motor sequence learning: A PET study,” *Brain*, vol. 121, no. 11, pp. 2159–2173, Nov. 1998.
- [95] M. K. Rand, O. Hikosaka, S. Miyachi, X. Lu, and K. Miyashita, “Characteristics of a long-term procedural skill in the monkey,” *Exp. Brain Res.*, vol. 118, no. 3, pp. 293–297, Feb. 1998.
- [96] M. K. Rand *et al.*, “Characteristics of sequential movements during early learning period in monkeys,” *Exp. Brain Res.*, vol. 131, no. 3, pp. 293–304, Apr. 2000.
- [97] J. Doyon *et al.*, “Role of the striatum, cerebellum, and frontal lobes in the learning of a visuomotor sequence,” *Brain Cogn.*, vol. 34, no. 2, pp. 218–245, Jul. 1997.
- [98] M. Jueptner, C. D. Frith, D. J. Brooks, R. S. Frackowiak, and R. E. Passingham, “Anatomy of motor learning: Subcortical structures and learning by trial and error,” *J. Neurophysiol.*, vol. 77, no. 3, pp. 1325–1337, Mar. 1997.
- [99] M. S. Jog, Y. Kubota, C. I. Connolly, V. Hillegaart, and A. M. Graybiel, “Building neural representations of habits,” *Science*, vol. 286, no. 5445, pp. 1745–1749, Nov. 1999.
- [100] N. Matsumoto, T. Hanakawa, S. Maki, A. M. Graybiel, and M. Kimura, “Role of nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner,” *J. Neurophysiol.*, vol. 82, no. 2, pp. 978–998, Aug. 1999.

- [101] S. Miyachi, O. Hikosaka, K. Miyashita, Z. Kárádi, and M. K. Rand, “Differential roles of monkey striatum in learning of sequential hand movement,” *Exp. Brain Res.*, vol. 115, no. 1, pp. 1–5, Jun. 1997.
- [102] P. D. Nixon and R. E. Passingham, “The cerebellum and cognition: Cerebellar lesions impair sequence learning but not conditional visuomotor learning in monkeys,” *Neuropsychologia*, vol. 38, no. 7, pp. 1054–1072, Jan. 2000.
- [103] X. Lu, O. Hikosaka, and S. Miyachi, “Role of monkey cerebellar nuclei in skill for sequential movement,” *J. Neurophysiol.*, vol. 79, no. 5, pp. 2245–2254, May 1998.
- [104] H. Imamizu *et al.*, “Human cerebellar activity reflecting an acquired internal model of a new tool,” *Nature*, vol. 403, no. 6766, pp. 192–195, Jan. 2000.
- [105] A. Floyer-Lea and P. M. Matthews, “Distinguishable brain activation networks for short- and long-term motor skill learning,” *J. Neurophysiol.*, vol. 94, no. 1, pp. 512–518, Jul. 2005.
- [106] K. Sakai, O. Hikosaka, S. Miyauchi, Y. Sasaki, N. Fujimaki, and B. Pütz, “Presupplementary motor area activation during sequence learning reflects visuo-motor association,” *J. Neurosci.*, vol. 19, no. 10, pp. 1–6, May 1999.
- [107] S. T. Grafton, E. Hazeltine, and R. B. Ivry, “Motor sequence learning with the nondominant left hand: A PET functional imaging study,” *Exp. Brain Res.*, vol. 146, no. 3, pp. 369–378, Oct. 2002.
- [108] R. M. Costa, D. Cohen, and M. A. L. Nicolelis, “Differential corticostriatal plasticity during fast and slow motor skill learning in mice,” *Curr. Biol.*, vol. 14, no. 13, pp. 1124–1134, Jul. 2004.
- [109] U. Ziemann, T. V Ilić, T. V Ilić, C. Pauli, F. Meintzschel, and D. Ruge, “Learning modifies subsequent induction of long-term potentiation-like and long-term depression-like plasticity in human motor cortex,” *J. Neurosci.*, vol. 24, no. 7, pp. 1666–1672, Feb. 2004.
- [110] K. Stefan *et al.*, “Temporary occlusion of associative motor cortical plasticity by prior dynamic motor training,” *Cereb. Cortex*, vol. 16, no. 3, pp. 376–385, Mar. 2006.
- [111] J. Doyon and H. Benali, “Reorganization and plasticity in the adult brain during learning of motor skills,” *Curr. Opin. Neurobiol.*, vol. 15, no. 2, pp. 161–167, Apr. 2005.
- [112] E. Dayan and L. G. G. Cohen, “Neuroplasticity subserving motor skill learning,” *Neuron*, vol. 72, no. 3, pp. 443–454, Nov. 2011.
- [113] D. Coynel *et al.*, “Dynamics of motor-related functional integration during motor sequence learning,” *Neuroimage*, vol. 49, no. 1, pp. 759–766, Jan. 2010.

- [114] S. Lehericy *et al.*, “Distinct basal ganglia territories are engaged in early and advanced motor sequence learning,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 35, pp. 12566–12571, Aug. 2005.
- [115] R. J. Nudo, G. W. Milliken, W. M. Jenkins, and M. M. Merzenich, “Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys,” *J. Neurosci.*, vol. 16, no. 2, pp. 785–807, Jan. 1996.
- [116] J. A. Kleim, E. Lussnig, E. R. Schwarz, T. A. Comery, and W. T. Greenough, “Synaptogenesis and FOS expression in the motor cortex of the adult rat after motor skill learning,” *J. Neurosci.*, vol. 16, no. 14, pp. 4529–4535, Jul. 1996.
- [117] S. S. Kantak and C. J. Winstein, “Learning-performance distinction and memory processes for motor skills: A focused review and perspective,” *Behav. Brain Res.*, vol. 228, no. 1, pp. 219–231, Mar. 2012.
- [118] J. W. Krakauer and R. Shadmehr, “Consolidation of motor memory,” *Trends Neurosci.*, vol. 29, no. 1, pp. 58–64, Jan. 2006.
- [119] E. M. Robertson, A. Pascual-Leone, and R. C. Miall, “Current concepts in procedural consolidation,” *Nat. Rev. Neurosci.*, vol. 5, no. 7, pp. 576–582, Jul. 2004.
- [120] E. M. Robertson and D. A. Cohen, “Understanding consolidation through the architecture of memories,” *Neurosci.*, vol. 12, no. 3, pp. 261–271, Jun. 2006.
- [121] E. M. Robertson, “Skill learning: Putting procedural consolidation in context,” *Curr. Biol.*, vol. 14, no. 24, pp. 1061–1063, Dec. 2004.
- [122] P. Baraduc, N. Lang, J. C. Rothwell, and D. M. Wolpert, “Consolidation of dynamic motor learning is not disrupted by rTMS of primary motor cortex,” *Curr. Biol.*, vol. 14, no. 3, pp. 252–256, Feb. 2004.
- [123] W. Muellbacher *et al.*, “Early consolidation in human primary motor cortex,” *Nature*, vol. 415, no. 6872, pp. 640–644, Mar. 2002.
- [124] E. M. Robertson, D. Z. Press, and A. Pascual-Leone, “Off-line learning and the primary motor cortex,” *J. Neurosci.*, vol. 25, no. 27, pp. 6372–6378, Jul. 2005.
- [125] M. P. Walker and R. Stickgold, “Sleep-dependent learning and memory consolidation,” *Neuron*, vol. 44, no. 1, pp. 121–133, Sep. 2004.
- [126] M. P. Walker, T. Brakefield, J. Seidman, A. Morgan, J. A. Hobson, and R. Stickgold, “Sleep and the time course of motor skill learning,” *Learn. Mem.*, vol. 10, no. 4, pp. 275–284, Jan. 2003.
- [127] C. F. Siengsukon and L. A. Boyd, “Does sleep promote motor learning? Implications for physical rehabilitation,” *Phys. Ther.*, vol. 89, no. 4, pp. 370–383,

Apr. 2009.

- [128] C.-H. J. Lin, B. J. Knowlton, M.-C. Chiang, M. Iacoboni, P. Udompholkul, and A. D. Wu, “Brain-behavior correlates of optimizing learning through interleaved practice,” *Neuroimage*, vol. 56, no. 3, pp. 1758–1772, Jun. 2011.
- [129] L. M. Rueda-Delgado, E. Solesio-Jofre, D. J. Serrien, D. Mantini, A. Daffertshofer, and S. P. Swinnen, “Understanding bimanual coordination across small time scales from an electrophysiological perspective,” *Neurosci. Biobehav. Rev.*, vol. 47, pp. 614–635, Nov. 2014.
- [130] S. B. Issenberg *et al.*, “Simulation technology for health care professional skills training and assessment,” *JAMA*, vol. 282, no. 9, p. 861, Sep. 1999.
- [131] G. Wulf, C. Shea, and R. Lewthwaite, “Motor skill learning and performance: A review of influential factors,” *Med. Educ.*, vol. 44, no. 1, pp. 75–84, Jan. 2010.
- [132] S. Shimada, “Modulation of motor area activity by the outcome for a player during observation of a baseball game,” *PLoS One*, vol. 4, no. 11, pp. 1–6, Nov. 2009.
- [133] R. Aggarwal, K. Moorthy, and A. Darzi, “Laparoscopic skills training and assessment,” *Br. J. Surg.*, vol. 91, no. 12, pp. 1549–1558, Dec. 2004.
- [134] S. P. Swinnen, “Intermanual coordination: From behavioural principles to neural-network interactions,” *Nat. Rev. Neurosci.*, vol. 3, no. 5, pp. 348–359, May 2002.
- [135] C. Maes, J. Gooijers, J.-J. Orban de Xivry, S. P. Swinnen, and M. P. Boisgontier, “Two hands, one brain, and aging,” *Neurosci. Biobehav. Rev.*, vol. 75, pp. 234–256, Apr. 2017.
- [136] S. P. Swinnen and J. Gooijers, “Bimanual Coordination,” in *Brain Mapping*, 1st ed., San Diego, CA USA: Elsevier, 2015, pp. 475–482.
- [137] D. R. Leff *et al.*, “Assessment of the cerebral cortex during motor task behaviours in adults: A systematic review of functional near infrared spectroscopy (fNIRS) studies,” *Neuroimage*, vol. 54, no. 4, pp. 2922–2936, Feb. 2011.
- [138] J. Andreu-Perez, D. R. Leff, K. Shetty, A. Darzi, and G.-Z. Yang, “Disparity in frontal lobe connectivity on a complex bimanual motor task aids in classification of operator skill level,” *Brain Connect.*, vol. 6, no. 5, pp. 375–388, Jun. 2016.
- [139] D. M. Wolpert, Z. Ghahramani, and J. R. Flanagan, “Perspectives and problems in motor learning,” *Trends Cogn. Sci.*, vol. 5, no. 11, pp. 487–494, Nov. 2001.
- [140] D. M. Wolpert, J. Diedrichsen, and J. R. Flanagan, “Principles of sensorimotor learning,” *Nat. Rev. Neurosci.*, vol. 12, no. 12, pp. 739–751, Oct. 2011.
- [141] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, “G\*Power 3: A flexible

- statistical power analysis program for the social, behavioral, and biomedical sciences,” *Behav. Res. Methods*, vol. 39, no. 2, pp. 175–191, May 2007.
- [142] T. J. Huppert, S. G. Diamond, M. A. Franceschini, and D. A. Boas, “Homer: A review of time-series analysis methods for near-infrared spectroscopy of the brain,” *Appl. Opt.*, vol. 48, no. 10, pp. 280–298, Mar. 2009.
- [143] A. Duncan *et al.*, “Measurement of cranial optical path length as a function of age using phase resolved near infrared spectroscopy,” *Pediatr. Res.*, vol. 39, no. 5, pp. 889–894, May 1996.
- [144] M. A. Franceschini, D. K. Joseph, T. J. Huppert, S. G. Diamond, and D. A. Boas, “Diffuse optical imaging of the whole head,” *J. Biomed. Opt.*, vol. 11, no. 5, pp. 1–22, Jan. 2006.
- [145] Y. Zhang, D. H. Brooks, M. A. Franceschini, and D. A. Boas, “Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging,” *J. Biomed. Opt.*, vol. 10, no. 11014, pp. 1–11, Jan. 2005.
- [146] J. C. Ye, S. Tak, K. E. Jang, J. Jung, and J. Jang, “NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy,” *Neuroimage*, vol. 44, no. 2, pp. 428–447, Jan. 2009.
- [147] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics, 5th ed.* Needham Heights, MA, USA: Allyn & Bacon, 2007.
- [148] P. M. Groves and R. F. Thompson, “Habituation: A dual-process theory,” *Psychol. Rev.*, vol. 77, no. 5, pp. 419–450, Sep. 1970.
- [149] E. Watanabe, Y. Yamashita, A. Maki, Y. Ito, and H. Koizumi, “Non-invasive functional mapping with multi-channel near infra-red spectroscopic topography in humans,” *Neurosci. Lett.*, vol. 205, no. 1, pp. 41–44, Feb. 1996.
- [150] M. M. Plichta, M. J. Herrmann, A.-C. Ehlis, C. G. Baehne, M. M. Richter, and A. J. Fallgatter, “Event-related visual versus blocked motor task: Detection of specific cortical activation patterns with functional near-infrared spectroscopy,” *Neuropsychobiology*, vol. 53, no. 2, pp. 77–82, Jan. 2006.
- [151] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, “Brain–computer interfaces for communication and rehabilitation,” *Nat. Rev. Neurol.*, vol. 12, no. 9, pp. 513–525, Aug. 2016.
- [152] M. Mihara *et al.*, “Near-infrared spectroscopy-mediated neurofeedback enhances efficacy of motor imagery-based training in poststroke victims: A pilot study,” *Stroke*, vol. 44, no. 4, pp. 1091–1098, Apr. 2013.
- [153] K. Yarrow, P. Brown, and J. W. Krakauer, “Inside the brain of an elite athlete: The neural processes that support high achievement in sports,” *Nat. Rev. Neurosci.*, vol.

10, no. 8, pp. 585–596, Aug. 2009.

- [154] T. Durduran *et al.*, “Diffuse optical measurement of blood flow, blood oxygenation, and metabolism in a human brain during sensorimotor cortex activation,” *Opt. Lett.*, vol. 29, no. 15, pp. 1766–1768, Aug. 2004.
- [155] L. Xu, B. Wang, G. Xu, W. Wang, Z. Liu, and Z. Li, “Functional connectivity analysis using fNIRS in healthy subjects during prolonged simulated driving,” *Neurosci. Lett.*, vol. 640, pp. 21–28, Feb. 2017.
- [156] F. S. Racz, P. Mukli, Z. Nagy, and A. Eke, “Increased prefrontal cortex connectivity during cognitive challenge assessed by fNIRS imaging,” *Biomed. Opt. Express*, vol. 8, no. 8, pp. 3842–3855, Aug. 2017.
- [157] N. Arizono, Y. Ohmura, S. Yano, and T. Kondo, “Functional connectivity analysis of NIRS data under rubber hand illusion to find a biomarker of sense of ownership,” *Neural Plast.*, vol. 2016, pp. 1–9, Jun. 2016.
- [158] D. J. Serrien, R. B. Ivry, and S. P. Swinnen, “Dynamics of hemispheric specialization and integration in the context of motor control,” *Nat. Rev. Neurosci.*, vol. 7, no. 2, pp. 160–166, Feb. 2006.
- [159] J. R. Flanagan, M. C. Bowman, and R. S. Johansson, “Control strategies in object manipulation tasks,” *Curr. Opin. Neurobiol.*, vol. 16, no. 6, pp. 650–659, Dec. 2006.
- [160] S. P. Swinnen and N. Wenderoth, “Two hands, one brain: Cognitive neuroscience of bimanual skill,” *Trends Cogn. Sci.*, vol. 8, no. 1, pp. 18–25, Jan. 2004.
- [161] S. R. Dawe, J. A. Windsor, J. A. J. L. Broeders, P. C. Cregan, P. J. Hewett, and G. J. Maddern, “A systematic review of surgical skills transfer after simulation-based training,” *Ann. Surg.*, vol. 259, no. 2, pp. 236–248, Feb. 2014.
- [162] A. L. McCluney *et al.*, “FLS simulator performance predicts intraoperative laparoscopic skill,” *Surg. Endosc.*, vol. 21, no. 11, pp. 1991–1995, Oct. 2007.
- [163] D. J. Scott, E. M. Ritter, S. T. Tesfay, E. A. Pimentel, A. Nagji, and G. M. Fried, “Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training,” *Surg. Endosc.*, vol. 22, no. 8, pp. 1887–1893, Aug. 2008.
- [164] R. M. Satava, “Emerging trends that herald the future of surgical simulation,” *Surg. Clin. North Am.*, vol. 90, no. 3, pp. 623–633, Jun. 2010.
- [165] B. K. Poulouse *et al.*, “Fundamentals of endoscopic surgery cognitive examination: Development and validity evidence,” *Surg. Endosc.*, vol. 28, no. 2, pp. 631–638, Feb. 2014.

- [166] M. C. Vassiliou *et al.*, “Fundamentals of endoscopic surgery: Creation and validation of the hands-on test,” *Surg. Endosc.*, vol. 28, no. 3, pp. 704–711, Mar. 2014.
- [167] E. M. McDougall, “Validation of surgical simulators,” *J. Endourol.*, vol. 21, no. 3, pp. 244–247, Mar. 2007.
- [168] D. J. Scott, C. M. Pugh, E. M. Ritter, L. M. Jacobs, C. A. Pellegrini, and A. K. Sachdeva, “New directions in simulation-based surgical education and training: Validation and transfer of surgical skills, use of nonsurgeons as faculty, use of simulation to screen and select surgery residents, and long-term follow-up of learners,” *Surgery*, vol. 149, no. 6, pp. 735–744, Jun. 2011.
- [169] S. R. Dawe *et al.*, “Systematic review of skills transfer after surgical simulation-based training,” *Br. J. Surg.*, vol. 101, no. 9, pp. 1063–1076, Aug. 2014.
- [170] X. Dai, T. Zhang, H. Yang, J. Tang, P. R. Carney, and H. Jiang, “Fast non-invasive functional diffuse optical tomography for brain imaging,” *J. Biophotonics.*, to be published. doi: 10.1002/jbio.201600267.
- [171] A. Nemani, U. Kruger, X. Intes, and S. De, “Increased sensitivity in discriminating surgical motor skills using prefrontal cortex activation over established metrics,” presented at the Conf. Optics in the Life Sciences Congress, San Diego, CA, USA, Apr. 2-5, 2017.
- [172] G. Ahlberg *et al.*, “Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies,” *Am. J. Surg.*, vol. 193, no. 6, pp. 797–804, Jun. 2007.
- [173] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager, “Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions,” *Comput. Aided Surg.*, vol. 11, no. 5, pp. 220–230, Jan. 2006.
- [174] J. R. Korndorffer, J. B. Dunne, R. Sierra, D. Stefanidis, C. L. Touchard, and D. J. Scott, “Simulator training for laparoscopic suturing using performance goals translates to the operating room,” *J. Am. Coll. Surg.*, vol. 201, no. 1, pp. 23–29, Jul. 2005.
- [175] G. Sroka, L. S. Feldman, M. C. Vassiliou, P. A. Kaneva, R. Fayez, and G. M. Fried, “Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room—A randomized controlled trial,” *Am. J. Surg.*, vol. 199, no. 1, pp. 115–120, Jan. 2010.
- [176] G. Ahlberg, R. Hultcrantz, E. Jaramillo, A. Lindblom, and D. Arvidsson, “Virtual reality colonoscopy simulation: A compulsory practice for the future colonoscopist?,” *Endoscopy*, vol. 37, no. 12, pp. 1198–1204, Dec. 2005.
- [177] T. P. Grantcharov, V. B. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P.

- Funch-Jensen, “Randomized clinical trial of virtual reality simulation for laparoscopic skills training,” *Br. J. Surg.*, vol. 91, no. 2, pp. 146–150, Feb. 2004.
- [178] A. T. Eggebrecht *et al.*, “Mapping distributed brain function and networks with diffuse optical tomography,” *Nat. Photonics*, vol. 8, no. 6, pp. 448–454, Jun. 2014.
- [179] B. Zendejas, R. Brydges, S. J. Hamstra, and D. A. Cook, “State of the evidence on simulation-based training for laparoscopic surgery,” *Ann. Surg.*, vol. 257, no. 4, pp. 586–593, Apr. 2013.
- [180] M. C. Vassiliou, B. J. Dunkin, J. M. Marks, and G. M. Fried, “FLS and FES: Comprehensive models of training and assessment,” *Surg. Clin. North Am.*, vol. 90, no. 3, pp. 535–558, Jun. 2010.
- [181] Q. Tan *et al.*, “Frequency-specific functional connectivity revealed by wavelet-based coherence analysis in elderly subjects with cerebral infarction using NIRS method,” *Med. Phys.*, vol. 42, no. 9, pp. 5391–5403, Aug. 2015.
- [182] X. Cui, D. M. Bryant, and A. L. Reiss, “NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation,” *Neuroimage*, vol. 59, no. 3, pp. 2430–2437, Feb. 2012.
- [183] R. Cui *et al.*, “Wavelet coherence analysis of spontaneous oscillations in cerebral tissue oxyhemoglobin concentrations and arterial blood pressure in elderly subjects,” *Microvasc. Res.*, vol. 93, pp. 14–20, May 2014.
- [184] L. Holper, F. Scholkmann, and M. Wolf, “Between-brain connectivity during imitation measured by fNIRS,” *Neuroimage*, vol. 63, no. 1, pp. 212–222, Oct. 2012.
- [185] A. Bandrivskyy, A. Bernjak, P. McClintock, and A. Stefanovska, “Wavelet phase coherence analysis: Application to skin temperature and blood flow,” *Cardiovasc. Eng.*, vol. 4, no. 1, pp. 89–93, Mar. 2004.
- [186] F. Tian, T. Tarumi, H. Liu, R. Zhang, and L. Chalak, “Wavelet coherence analysis of dynamic cerebral autoregulation in neonatal hypoxic–ischemic encephalopathy,” *NeuroImage Clin.*, vol. 11, pp. 124–132, Jan. 2016.
- [187] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, “Functional connectivity in the motor cortex of resting human brain using echo-planar MRI,” *Magn. Reson. Med.*, vol. 34, no. 4, pp. 537–541, Oct. 1995.
- [188] T. Söderström, A. Stefanovska, M. Veber, and H. Svensson, “Involvement of sympathetic nerve activity in skin blood flow oscillations in humans,” *Am. J. Physiol. Heart Circ. Physiol.*, vol. 284, no. 5, pp. 1638–1646, May 2003.
- [189] L. W. Sheppard, A. Stefanovska, and P. V. E. McClintock, “Testing for time-localized coherence in bivariate data,” *Phys. Rev. E*, vol. 85, no. 4, pp. 1–16, Apr. 2012.

- [190] A. Bernjak, A. Stefanovska, P. McClintock, J. Owen-Lynch, and P. Clarkson, “Coherence between fluctuations in blood flow and oxygen saturation,” *Fluct. Noise Lett.*, vol. 11, no. 1, pp. 1–11, Mar. 2012.
- [191] A. Stefanovska, M. Bracic, and H. D. Kvernmo, “Wavelet analysis of oscillations in the peripheral blood circulation measured by laser doppler technique,” *IEEE Trans. Biomed. Eng.*, vol. 46, no. 10, pp. 1230–1239, Oct. 1999.
- [192] A. B. Rowley *et al.*, “Synchronization between arterial blood pressure and cerebral oxyhaemoglobin concentration investigated by wavelet cross-correlation,” *Physiol. Meas.*, vol. 28, no. 2, pp. 161–173, Feb. 2007.
- [193] Y. Shiogai, A. Stefanovska, and P. V. E. McClintock, “Nonlinear dynamics of cardiovascular ageing,” *Phys. Rep.*, vol. 488, no. 2–3, pp. 51–110, Mar. 2010.
- [194] C. Iadecola and M. Nedergaard, “Glial regulation of the cerebral microvasculature,” *Nat. Neurosci.*, vol. 10, no. 11, pp. 1369–1376, Nov. 2007.
- [195] H. van Duinen, R. Renken, N. Maurits, and I. Zijdwind, “Effects of motor fatigue on human brain activity, an fMRI study,” *Neuroimage*, vol. 35, no. 4, pp. 1438–1449, May 2007.
- [196] E. B. Johnson *et al.*, “The impact of occipital lobe cortical thickness on cognitive task performance: An investigation in huntington’s disease,” *Neuropsychologia*, vol. 79, pp. 138–146, Dec. 2015.
- [197] M. W. Parsons, D. L. Harrington, and S. M. Rao, “Distinct neural systems underlie learning visuomotor and spatial representations of motor skills,” *Hum. Brain Mapp.*, vol. 24, no. 3, pp. 229–247, Mar. 2005.
- [198] H. Obrig and A. Villringer, “Beyond the visible—imaging the human brain with light,” *J. Cereb. Blood Flow Metab.*, vol. 23, no. 1, pp. 1–18, Jan. 2003.
- [199] B. R. White and J. P. Culver, “Quantitative evaluation of high-density diffuse optical tomography: in vivo resolution and mapping performance,” *J. Biomed. Opt.*, vol. 15, no. 2, pp. 1–14, Mar. 2010.
- [200] S. Yule, R. Flin, S. Paterson-Brown, and N. Maran, “Non-technical skills for surgeons in the operating room: A review of the literature,” *Surgery*, vol. 139, no. 2, pp. 140–149, Feb. 2006.
- [201] A. A. Gawande, M. J. Zinner, D. M. Studdert, and T. A. Brennan, “Analysis of errors reported by surgeons at three teaching hospitals,” *Surgery*, vol. 133, no. 6, pp. 614–621, Jun. 2003.
- [202] A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen, “The neural basis of economic decision-making in the ultimatum game,” *Science*, vol. 300, no. 5626, pp. 1755–1758, Jun. 2003.

