INVESTIGATING RIGHT HAND AND LEFT-HAND SYMMETRY/ASYMMETRY WITH AN INTEGRATED FORCE SENSOR IN LAPAROSCOPIC BOX TRAINING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ANKARA UNIVERSITY

by

Sebahat Selin ŞAHİN

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN
BIOMEDICAL ENGINEERING

ANKARA
2025

All rights reserved

ABSTRACT

Master's Thesis

INVESTIGATING RIGHT HAND AND LEFT-HAND SYMMETRY/ASYMMETRY WITH AN INTEGRATED FORCE SENSOR IN LAPAROSCOPIC BOX TRAINING

Sebahat Selin ŞAHİN

Ankara University
Graduate School of Natural and Applied Sciences
Department of Biomedical Engineering

Supervisor: Assist. Prof. Hasan Onur KELEŞ

Our objectives were to integrate a pressure-sensing unit into box-trainer simulators and laparoscopic tools to validate tool pressure features as objective indicators of laparoscopic surgical skills, and to combine EEG-derived power and phase-locking value (PLV) features with tool pressure data to evaluate the classification accuracy of machine learning models. We recruited 23 participants, including 13 without laparoscopy experience and 10 surgeons with varying experience (5–100 procedures). A pressure sensor was installed on both grasper handles to monitor thumb force. Alongside tool pressure data, NASA-TLX, completion time, error rate, reaction time, ECG, and mobile EEG (Mentalab, Germany) were collected during the task. Right–left asymmetry was calculated as $(R - L) / (R + L)$, where R and L represented average amplitudes over time. EEG data underwent power and PLV analysis across delta (0–4 Hz), theta (4–8 Hz), and alpha (8–12 Hz). HRV was also calculated. Three machine learning models, Random Forest Classifier (RFC), Gaussian Process (GP), and AdaBoost (ABC), were tested. Right–left pressure asymmetry was an important metric for skill assessment. Using only the asymmetry index during the task, RFC reached 78% accuracy. The highest performance occurred when combining theta-band power with asymmetry features during the task, with RFC and ABC both reaching 86% accuracy, F1 = 0.83, and AUC = 0.92 (RFC) and 0.85 (ABC). Theta-band findings support its relevance for surgical skill assessment. Combining pressure features with EEG features enhances objective evaluation of surgical skills, offering potential for real-time feedback in laparoscopic training.

September, 66 pages

Key Words: Laparoscopy, Surgical Training, Force Sensor, EEG, HRV, Mental workload

ÖZET

Yüksek Lisans Tezi

LAPAROSKOPİK KUTU EĞİTİMİNDE ENTEGRE EDİLMİŞ KUVVET SENSÖRÜ İLE SAĞ EL VE SOL EL SİMETRİSİ/ASİMETRİSİNİ İNCELEME

Sebahat Selin ŞAHİN

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Biyomedikal Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Hasan Onur KELEŞ

Amacımız, laparoskopik cerrahi becerilerin objektif göstergeleri olarak alet basıncı özniteliklerini doğrulamak için mevcut box-trainer simülatörlerine ve laparoskopik aletlere basınca duyarlı bir ünite entegre etmek ve EEG'den elde edilen güç ile faz kilitleme değeri (PLV) özniteliklerini basınç verileriyle birleştirerek makine öğrenmesi modellerinin sınıflandırma doğruluğunu değerlendirmektir. Çalışmaya 23 katılımcı dâhil edilmiştir; bunların 13'ü laparoskopi deneyimi olmayan gönüllüler, 10'u ise farklı düzeylerde deneyime sahip (5–100 prosedür) cerrahlardan oluşmaktadır. Başparmaklarla uygulanan kuvveti ölçmek için her iki grasper koluna özel basınç sensörleri yerleştirilmiştir. Basınç verilerine ek olarak NASA-TLX, tamamlanma süresi, hata oranı, reaksiyon süresi, EKG ve mobil EEG (Mentalab, Almanya) kayıtları görev sırasında toplanmıştır. Sağ–sol asimetrisi (R − L) / (R + L) formülüyle hesaplanmış; burada R ve L, zamana göre ortalama genlikleri temsil etmiştir. EEG verileri frekans güç analizi ve delta (0–4 Hz), teta (4–8 Hz) ve alfa (8–12 Hz) bantlarında PLV hesaplamasına tabi tutulmuştur. Ayrıca HRV de hesaplanmıştır. Cerrahi becerilerin sınıflandırılması için üç makine öğrenmesi modeli — Random Forest Classifier (RFC), Gaussian Process (GP) ve AdaBoost (ABC) — test edilmiştir. Sağ–sol basınç asimetrisi beceri değerlendirmesinde önemli bir öznitelik olarak öne çıkmış, yalnızca asimetri indeksiyle görev sırasında RFC %78 doğruluğa ulaşmıştır. En yüksek performans, teta bant gücü ile asimetri özniteliklerinin birlikte kullanıldığı durumda elde edilmiş; RFC ve ABC %86 doğruluk, F1 = 0.83, AUC = 0.92 (RFC) ve 0.85 (ABC) sağlamıştır. Teta bant bulguları cerrahi beceri değerlendirmesindeki önemini desteklemiştir. Basınç ve EEG özniteliklerinin birleştirilmesi, cerrahi becerilerin objektif değerlendirilmesini geliştirmekte ve laparoskopik eğitimde gerçek zamanlı geri bildirim için potansiyel sunmaktadır.

Eylül, 66 sayfa

Anahtar Kelimeler: Laparoskopi, Cerrahi Eğitim, Kuvvet Sensörü, EEG, HRV, Mental İş Yükü

FOREWORD AND ACKNOWLEDGEMENTS

Completing this thesis has been a journey of learning, effort, and self-discovery. It has challenged me, shaped my perspective, and allowed me to grow both academically and personally. Along the way, I have been fortunate to receive guidance and support from many people, to whom I am deeply grateful.

I would like to sincerely thank my supervisor, Assistant Professor Hasan Onur KELEŞ, for his trust, patience, and guidance throughout this process. His insights and constructive feedback have been invaluable in shaping the quality and direction of this work.

I also wish to thank Assistant Professor Mehmet YÜKSEKKAYA for his contributions and support during this study. My thanks also go to Çağrı ZENGİN for his collaboration and valuable input during the thesis process.

I am especially grateful to Res. Asst. Sena HAMARAT, Res. Asst Mustafa TÜFEKÇİ, and Res. Asst Abdullah EYİDOĞAN. Their encouragement, kindness, and friendship have been a constant source of motivation, and their presence has made this demanding process far more meaningful and enjoyable. Having the opportunity to share this journey with such supportive and understanding people has been one of the most rewarding parts of my graduate experience.

I would also like to extend my appreciation to all my lab colleagues for their help and for creating an environment where collaboration and mutual support have always been present. Working alongside such dedicated individuals has made this experience not only productive but also personally fulfilling.

I would also like to acknowledge the Ankara University Biomedical Engineering Department for providing the facilities and resources necessary for this study. Additionally, I am grateful to NÖROM (Neuroscience Research Laboratory), where all

experiments for this thesis were conducted, for offering the proper setting to carry out this work.

Finally, my deepest gratitude goes to my family. To my mother, Seval Şahin, for her endless love, patience, and sacrifices, and to my brother, Ali Furkan Şahin, for his unwavering support and belief in me. Their presence has been my greatest strength throughout this journey.

This thesis reflects not only my own efforts but also the contributions and support of everyone who has accompanied me along the way. I am sincerely thankful to each of you for helping me reach this milestone.

Sebahat Selin ŞAHİN
Ankara, September 2025

TABLE OF CONTENTS

THESIS APPROVAL
ETHIC ..... **i**
ABSTRACT ..... **ii**
ÖZET ..... **iii**
FOREWORD AND ACKNOWLEDGEMENTS ..... **iv**
ABBREVIATIONS ..... **vii**
LIST OF FIGURES ..... **ix**
LIST OF TABLES ..... **xi**
1. INTRODUCTION **1**
2. LITERATURE REVIEW **3**
2.1 Laparoscopic Surgery **3**
2.2 Surgical Skill Assessment **4**
2.3 Force-Based Skill Assessment **6**
2.4 Neurophysiological-Based Skill Assessment **9**
2.5 Subjective and Behavioral-Based Skill Assessment **12**
2.6 Other Skill Assessment **14**
2.7 Machine Learning in Surgical Skill Assessment **16**
3. MATERIALS AND METHOD **21**
3.1 Participants **21**
3.2 Experimental Design **21**
3.2.1 Sensor design **24**
3.3 Data Collection **25**
3.4 Data Analysis **27**
3.4.1 Behavioral and performance data **27**
3.4.2 Sensor measures **27**
3.4.3 EEG measures **29**
3.4.4 Statistical and machine learning analyses **32**
4. RESULTS **35**
4.1 Behavioral and Performance Results **35**
4.2 Force Sensor Data Results **37**
4.3 Inter-Variable Relationship Assessment **39**
4.4 Classification Results **41**
4.5 Confusion Matrix Analysis **45**
5. DISCUSSION **51**
5.1 Limitations **52**
6.CONCLUSION **54**
REFERENCES **55**
CURRICULUM VITAE **66**

SYMBOLS AND ABBREVIATIONS

ABC	AdaBoost Classifier
AD	Active Duration
AD_L	Active Left Duration Left
AD_L	Left Active Duration Left
AD_R	Active Right Duration
BR	Blink Rate
CL	Cognitive Load
CNNs	Convolutional Neural Networks
ECG	Electrocardiogram
EDA	Electrodermal Activity
EMG	Electromyography
FLS	Fundamentals of Laparoscopic Surgery
fNIRS	Functional Near-Infrared Spectroscopy
FTI	Force Time Integral
GOALS	Global Operative Assessment of Laparoscopic Skills
GPC	Gaussian Process Classifier
GSR	Galvanic Skin Response
HbO	Oxyhemoglobin
HbR	Deoxyhemoglobin
HR	Heart Rate
HRV	Heart Rate Variability
ICA	Independent Component Analysis
IMUs	Inertial Measurement Units
IoT	Internet of Things
ISPC	Inter-Site Phase Clustering
L	Left
M1	Primary Motor Cortex
ML	Machine Learning
MNI	Montreal Neurological Institute
NASA-TLX	NASA Task Load Index

OSATS	Objective Structured Assessment of Technical Skills
PAC	Phase–Amplitude Coupling
PFC	Prefrontal Cortex
PLV	Phase Locking Value
PSD	Power Spectral Density
R	Right
RFC	Random Forest Classifier
RLA	Right–Left Asymmetry
RLO	Right–Left Overlap
sEMG	Surface Electromyography
SMA	Supplementary Motor Area
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
tDCS	Transcranial Direct Current Stimulation
XGB	Extreme Gradient Boosting
α	Alpha
β	Beta
δ	Delta
θ	Theta

LIST OF FIGURES

Figure 3.1 Peg transfer task 23

Figure 3.2 Laparoscopic tool (grasper) schematic 24

Figure 3.3 Circuit for pressure measurement using flexible force sensor 25

Figure 3.4 Custom-designed electronic circuit board used for pressure sensor data acquisition 25

Figure 3.5 Electrode placement map based on the international 10–20 system 26

Figure 3.6 Experimental setup of the multimodal data acquisition system. The surgeon performs laparoscopic box trainer tasks while EEG signals are recorded via a mobile EEG device and pressure sensors capture hand force data from the laparoscopic tools. Representative EEG waveforms and force signal outputs are shown on the right 27

Figure 4.1 Comparison between surgeons (Surg) and non-experienced participants (Non-Exp) across four measures: HRV, number of errors, task completion time, and total NASA-TLX scores Metrics include heart rate variability (HRV), number of errors, task completion time, and total NASA-TLX scores 36

Figure 4.2 Force signals recorded from the right and left grasper handles during the peg transfer task. The lower panel presents the normalized and thresholded force signals 37

Figure 4.3 Sensor Comparison of sensor-derived force metrics between surgeons (Surg) and non-experienced participants (Non-Exp). A trend toward lower asymmetry was observed in the Surg group ($p = 0.06$), suggesting more balanced bimanual coordination. No significant differences were found for R–L overlap duration or active force duration. Error bars indicate standard deviation 39

Figure 4.4 Simple Linear regression analyses examining the relationships between NASA-TLX subscale scores and two predictors: task completion time (top two rows) and tool-use asymmetry index (bottom two rows) 40

Figure 4.5 Classification accuracies across four EEG frequency bands (delta, theta, alpha, and beta) using three different classifiers: Random Forest Classifier (RFC), Gaussian Process Classifier (GPC), and AdaBoost Classifier (ABC). Three feature sets were compared: asymmetry-only (red), EEG power spectral density (PSD) combined with asymmetry (light gray), and EEG phase locking value (PLV) combined with asymmetry (dark gray) 42

Figure 4.6 Confusion matrices of the classification results using EEG power features and asymmetry indices in different frequency bands. The matrices show the performance of the RFC, GPC, and ABC models in distinguishing surgeon and control participants during the task 45

Figure 4.7 Confusion matrices of the classification results using PLV-based functional connectivity features with asymmetry indices in the α, β, θ, and δ bands. The figure shows that ABC classifiers generally produce more balanced predictions, GPC shows class bias, especially in recognizing surgeon class, and RFC provides higher sensitivity for surgeons in the θ and δ bands..... 50

LIST OF TABLES

Table 3.1 Demographics of participants 21

Table 3.2 Experimental design 21

Table 4.1 Classification performances using EEG power features and asymmetry indices for the δ, θ, α, and β frequency bands under task and rest conditions. The table shows the accuracy, F1 score, precision, sensitivity, and AUC values for three classifiers (RFC, GPC, ABC), highlighting the best-performing models for each band 44

Table 4.2 Performance metrics for multimodal feature combinations including EEG-based PLV, asymmetry, and sensor-derived force measurements. Results are given in different frequency bands for RFC, GPC, and ABC classifiers, demonstrating the additional contribution of sensor data when integrated with neurophysiological measurements 47

1. INTRODUCTION

Laparoscopy is widely used in various surgical procedures today, offering significant advantages such as faster patient recovery, shorter hospital stays, and reduced postoperative pain. However, performing laparoscopic surgery presents unique challenges and requires surgeons to develop advanced perceptual and cognitive skills (Zakeri et al., 2020). Limited visual access and reduced tactile feedback necessitate surgeons to adopt new techniques characterized by significant learning curves and require advanced training methods beyond traditional master-apprentice models. Surgical skill level plays a crucial role in reducing medical errors and complications (Keleş et al., 2021).

Objective assessment of surgical skills is crucial in laparoscopic training, as accurate feedback is essential for improving performance. Recent technological advancements have facilitated more precise and continuous assessments of surgical performance (Soangra et al., 2022;). Psychophysiological measurements overcome the limitations of traditional skill assessment methods and provide objective and continuous assessments, providing an alternative to methods based on large-scale equipment, behavioral metrics, or subjective questionnaires (Omurtag et al., 2019).

Sensor-based technologies, such as electrocardiogram (ECG) including Heart Rate (HR) and Heart Rate Variability (HRV) (Farah et al., 2024), motion analysis via video cameras (Zia et al., 2015), eye tracking (Tien et al., 2014), electromyography (EMG) (Soangra et al., 2022), functional near-infrared spectroscopy (fNIRS) (Keles et al., 2021), and force sensors measuring tool-tissue interactions (Cundy et al., 2015), have been utilized to objectively assess surgical skill levels. These measures have revealed differences in psychomotor performance among surgeons of varying expertise, offering complementary insights into trainee development.

Surgeon-tool interactions, specifically involving the grasper instrument (laparoscopic tool), are critical to surgical performance. Measurement methods for tool forces during laparoscopic procedures have demonstrated that novice surgeons typically exert higher

and more variable forces compared to experts (Cundy et al., 2015). This study specifically investigates pressure asymmetry, defined as the imbalance of forces applied by surgeons across different surgical instruments, a concept particularly relevant to laparoscopic procedures requiring precise tool control.

Despite extensive research utilizing neuroimaging modalities independently to assess surgical skills, there remains a significant gap regarding the integration of tool pressure data with neuroimaging techniques. To the best of our knowledge, no multimodal studies have specifically combined tool pressure data and EEG to classify surgical expertise using machine learning techniques.

This study investigates the relationship between tool-pressure features, particularly right-left pressure asymmetry, and physiological and behavioral markers of surgical expertise during laparoscopic training. Our goal is to integrate a pressure-sensing unit into existing box trainer simulators and tools to achieve two main objectives: (1) validate right-left pressure asymmetry as an objective measure of laparoscopic skill, and (2) improve classification accuracy by combining EEG power and phase-locking value (PLV) features with tool pressure. By integrating neurophysiological and sensor-based metrics, our approach provides a novel, automated objective assessment method that can significantly enhance laparoscopic training programs.

2. LITERATURE REVIEW

2.1 Laparoscopic Surgery

Laparoscopic surgery, a type of minimally invasive procedure, involves performing operations through small incisions using a camera and specialized instruments. Compared to traditional open surgery, it is generally associated with less internal trauma, reduced blood loss, smaller scars, a lower risk of infection, shorter hospital stays, faster recovery, and less postoperative pain (Cleveland Clinic, 2024).

Surgical complications occur at a rate of 8% to 12% worldwide and cost the healthcare system approximately $9 billion annually in the United States alone. To prevent surgery-related complications, surgeons must acquire the necessary surgical skills before performing operations. (Aghazadeh et al.,2025). Surgical performance directly impacts patient outcomes (Keles,2022). Surgical skill is not solely dependent on innate ability; it can be improved with appropriate training and effective techniques (Rosser et al., 1997). However, laparoscopic skills are more challenging than those required in open surgery. Minimally invasive surgery presents several challenges that do not present in open surgery, including reduced tactile feedback, the fulcrum effect, and the complexity of working in a three-dimensional space using two-dimensional imaging (Adams et al., 2012).

Insufficient training contributes to increased surgical error rates. Surgical residency programs aim to educate professionals with the necessary technical, personal, and cognitive skills demanded by society (Aggarwal et al., 2004). However, traditionally, surgical skills have been taught through a master-apprentice model implemented in the operating room under the supervision of senior surgeons (Keles, 2022). Medical students have limited opportunities to develop their surgical skills in a safe and controlled environment before performing on actual patients (Porte et al., 2007). Achieving proficiency in these specialized skills goes beyond what can be gained through hands-on experience in the operating room alone. Given the constraints of the

long hours surgical residency work week, the implementation of innovative and structured training methods has become essential. (Adams et al.,2012)

Furthermore, laparoscopic surgery also requires highly specialized visuospatial reasoning and advanced manual dexterity (Adams et al.,2012). Very few training programs use objective tests of dexterity and surgical skills in their usual surgical training and evaluation methods. One reason for this is that the relationship between surgical skill and manual dexterity has been controversial for many years (Rosser et al., 1998). However, more recent research suggests that manual dexterity and visuospatial abilities play an important role in the development of laparoscopic skills. However, it has been noted that these variables alone are insufficient to explain technical performance; cognitive and psychological factors also play a decisive role (Beattie et al., 2022).

2.2 Surgical Skill Assessment

Objective and reliable assessment of surgical skills is a critical component of modern surgical training, enabling standardized evaluation, personalized feedback, and the development of proficiency benchmarks. Over the past two decades, a wide range of sensor-based technologies have been integrated into surgical education and research, each offering distinct advantages and limitations depending on the type of data collected. These methods can be categorized as tool motion tracking, body motion monitoring, cognitive load assessment, and object–tissue interaction analysis.

Object–tissue interaction has been studied using force sensors embedded in surgical tools, offering direct quantification of applied grip force and tool–tissue dynamics. Despite their low-profile and flexible design, the accuracy of such sensors varies depending on device type and applied force ranges.

Overall, these sensor modalities represent complementary approaches to surgical skill assessment. Tool and body motion sensors primarily quantify motor performance,

physiological sensors provide insights into cognitive and affective states, and force sensors characterize the interaction with the surgical environment. The integration of multimodal sensing is increasingly recognized as a promising strategy to capture the complex interplay between technical execution and cognitive load during surgical training.

Cognitive load measurement has gained attention as surgical performance is strongly influenced by mental workload. Smart glasses with integrated eye tracking allow unobtrusive monitoring of gaze behavior, while electroencephalography (EEG) provides high temporal resolution of neural activity. Additional physiological signals, such as electromyography (EMG), electrodermal activity (EDA), and heart rate variability (HRV), have also been employed to infer stress and workload, though these methods are often susceptible to noise, movement artifacts, and user discomfort.

Tool motion tracking has been performed using mechanical sensors, electromagnetic systems, inertial measurement units (IMUs), and optical markers.

Cognitive load (CL) plays a crucial role in understanding human learning and performance, referring to the perceived mental demands placed on an individual during a task (Wu et al., 2025). It refers to the amount of mental effort required to perform a task and plays a central role in understanding surgical performance and skill acquisition. When cognitive demands exceed the available mental resources, task efficiency decreases, reaction times slow down, and error rates increase (Almukhtar et al., 2025). Managing cognitive workload effectively is therefore critical to achieving optimal surgical performance, especially during complex laparoscopic tasks.

Bimanual coordination is referred as the skill of using both hands in a coordinated manner. It is essential for complex tasks and relies on temporal and spatial synchronization across limbs (Yeganeh Doost et al., 2017). Research shows that practicing asymmetric or complex two-handed tasks significantly improves coordination quality and reduces movement variability, thus enhancing precision (Roman-Liu et al., 2020). Furthermore, effective bimanual coordination is supported by

the central nervous system's predictive and adaptive mechanisms that help stabilize movements, which in turn reduces cognitive effort (Ridderikhoff, 2008).

Motor control is defined as the central nervous system's ability to plan, organize, and execute voluntary movements by integrating sensory feedback and motor commands to produce smooth and coordinated actions (Shumway-Cook & Woollacott, 2017). It plays a critical role in complex tasks because efficient motor control enables precise adjustments of movement patterns in response to changing task demands and environmental constraints (Latash, 2012). Research shows that repetitive practice of goal-directed motor tasks enhances neural efficiency and improves movement accuracy and stability, leading to better task performance (Dayan & Cohen, 2011). Furthermore, effective motor control is supported by the brain's predictive and adaptive mechanisms, which optimize motor planning and facilitate automatized, efficient execution during skilled performance (Krakauer & Mazzoni, 2011).

These three concepts are closely interrelated in the acquisition and execution of surgical skills. Studies show that improved motor control and well-developed bimanual coordination decrease unnecessary cognitive demands, enabling experts to perform tasks more smoothly and efficiently (Nemani et al., 2018). In laparoscopic training, practice with asymmetric or complex two-handed tasks has been found to enhance interhemispheric connectivity and coordination precision, leading to greater task automatization and reduced variability (Huang et al., 2025). Moreover, repeated practice strengthens adaptive motor mechanisms and stabilizes visuomotor integration, which supports the efficient allocation of cognitive resources during demanding procedures (Roman-Liu et al., 2020).

2.3 Force-Based Skill Assessment

Force-based skill assessment has emerged as a powerful and objective method for assessing surgical expertise because it captures not only movement patterns but also the magnitude, variability, and symmetry of the force applied during instrument-tissue interaction. Unlike traditional subjective rating scales, metrics derived from force and

movement allow for the detailed quantification of motor control, tactile awareness, and bimanual coordination in surgical training. Motor control refers to the central nervous system's ability to regulate and coordinate muscle activity, while bimanual coordination describes the simultaneous and complementary use of both hands, enabling the performance of complex tasks.

In this context, early studies demonstrated the feasibility of sensor-based measurement for psychomotor assessment. For example**,** they employed a haptic device to evaluate reaching and pointing tasks, showing that surgeons exhibited shorter trial times, smoother trajectories, and greater target accuracy compared to non-surgeons (Ahmmad et al.,2014). Building on these findings, another study subsequently introduced force variability as an objective marker of skill, surgeons showed fewer errors, better force stability, and more consistent control compared to non-surgeons (Ahmmad et al.,2015). Similarly, Sugiyama et al. (2018) measured tool–tissue interaction forces in neurosurgery using force-sensing bipolar forceps and found that experienced surgeons applied lower peak forces with less variability compared to novices, highlighting fine force control as a key indicator of surgical expertise.

Furthermore, a central dimension of force assessment is the evaluation of hand asymmetry and bimanual coordination. In a study, tractive force patterns during laparoscopic knot tying were analyzed and found that experts produced reproducible and balanced force trajectories across hands, whereas novices demonstrated unstable and asymmetric force application, often compromising knot security (Takayasu et al., 2018).

In addition, force measurement has also played an important role in endovascular skill assessment. Extending this approach, another study reported that novices frequently exceeded the safe loading threshold using catheter tip force sensors, while experts applied controlled and consistent forces, establishing construct validity for force-derived metrics in vascular navigation (Rafii-Tari et al.,2017).

More recently, newer platforms have shifted toward simulator-based force integration. For example, a pediatric thoracoscopic simulator equipped with force and motion sensors was developed, and this system more effectively distinguished expert from novice performance than traditional box simulators. This finding highlights the importance of anatomically realistic and force-sensitive environments in pediatric surgical training (Takazawa et al., 2016). Moreover, quantitative kinematic analyses have been shown to effectively identify subtle deficiencies in fine motor control, providing objective insights into the smoothness and efficiency of hand movements (Harada et al., 2015). In addition, instrument contact vibration analyses in robotic surgery revealed a strong negative correlation with technical proficiency, indicating that higher skill levels are associated with lower instrument vibrations (Gomez et al., 2016).

Differences in grip strength between experienced and novice surgeons have been studied in laparoscopic procedures performed on pigs using forceps equipped with pressure sensors. Findings revealed that experts applied more consistent and moderate forces and were more fluid and controlled in instrument use, while novices exhibited higher peak forces, greater variability, and abrupt force changes. This has been reported as a finding that may increase the risk of tissue damage (Araki et al., 2017). Furthermore, Markov modeling of force/torque data has shown that experts perform more consistent and efficient transitions, while novices display erratic and inefficient force patterns (Rosen et al., 2001).

A comparative study using force-sensored instruments found that experts applied lower and more stable grip forces, while novices produced excessive and irregular pressures (de Mathelin et al., 2019). Significant differences in intraoperative force application have also been demonstrated between experienced and novice surgeons (Hanna et al., 2008). More recently, novice surgeons have been shown to exhibit less synchronized bimanual control and more variable grip forces, whereas experts maintain smoother force patterns and better inter-hand coordination. Overall, these findings highlight that unstable and inconsistent force control is a key indicator of novice performance (Soleymani et al., 2024)

Beyond the surgical context, motor control research has also emphasized the importance of balance between the two hands. Right-handed individuals have generally been found to exhibit more pronounced asymmetry in the dominant hand, while left-handed individuals exhibit more balanced bimanual abilities (Ozcan et al., 2004). Additionally, it has been reported that men exhibit stronger dominant hand asymmetry, while women exhibit more symmetrical hand performance (Park, 2022). Beyond these inherent differences, effective motor performance relies heavily on the coordination between grip force and load force during object manipulation. Disruptions in this grip–load force coupling lead to unstable object control, increased movement variability, and ultimately reduced efficiency (Jaric & Uygur, 2013). Task complexity further intensifies these challenges, as bidirectional manipulations require greater adjustments in force and have been associated with higher force variability and less precise movement control (de Freitas et al., 2008).

2.4 Neurophysiological-Based Skill Assessment

In neurophysiological based skill assessment, few studies have investigated the relationship between surgical performance and the experience levels of assistants and expert surgeons. Techniques such as fNIRS and EEG are frequently employed to better understand the cognitive processes essential for surgical training and practice.

Functional near-infrared spectroscopy (fNIRS) is a functional imaging technique that measures changes in blood flow within the cerebral cortex to indirectly assess brain activity (Keleş, 2022). It offers a complementary view of cognitive processes by monitoring cortical oxygenation via HbO/HbR changes, with sensitivity in the prefrontal cortex (PFC). Novice trainees typically display diffuse and elevated prefrontal responses during surgical tasks, whereas experts show more focal and efficient activation patterns consistent with the automation of cognitive strategies (Leff, 2009). Subsequent studies have confirmed that surgical expertise can be reliably distinguished using neurophysiological measures. Nemani et al. (2018) demonstrated that novices exhibit higher prefrontal HbO levels, which are associated with greater error rates and longer completion times, whereas experts show lower PFC activation

and greater motor-area engagement, reflecting more efficient resource allocation. Crewther et al. (2016) further reported that repeated laparoscopic training leads to a progressive reduction in prefrontal activation, indicating a decrease in cognitive workload as surgical skills improve. Modi et al. (2018) conducted a study with surgical residents at different experience levels and reported that time pressure led to reduced prefrontal activation and decreased task performance.

In further applications, fNIRS has been combined with machine learning to classify expertise and predict subjective workload, with high-density optical imaging demonstrating that prefrontal oxygenation patterns estimate performance quality and mental effort with high fidelity (Keles et al., 2021).

Electroencephalography (EEG) remains one of the fundamental tools for studying mental workload and skill acquisition in surgery, capturing the rapid cortical dynamics associated with attention, working memory, and motor planning on the millisecond scale (Gevins et al., 1998). Increases in working memory demand are typically accompanied by increases in frontal theta activity and decreases in alpha rhythms; in laparoscopic and robotic contexts, novices exhibit higher frontal theta, while experts show more stable neural activation profiles (Antonenko et al., 2010; Hubert et al., 2013).

Previous EEG studies investigating surgical expertise have highlighted distinct neural signatures underlying cognitive workload and attentional control. For instance, Frontal theta activity (4–7 Hz) has been shown to systematically increase in novices compared to experts, reflecting higher mental effort and reliance on executive resources during laparoscopic training (Balkhoyor et al., 2020). In contrast, experts demonstrate lower frontal theta power, indicating more efficient and automatic cognitive processing strategies developed through practice. Occipital alpha rhythms have also been reported to be modulated more effectively in expert surgeons, enabling enhanced attentional control and better suppression of irrelevant visual information, whereas novices exhibit weaker modulation, reflecting less efficient attentional allocation (Reyes et al., 2021). In one study with surgeons with different levels of laparoscopic expertise and novices

performed standardized laparoscopic training tasks. Performance metrics, subjective workload ratings, and prefrontal cortical activity were assessed. Results showed that experienced surgeons completed tasks more rapidly, reported lower cognitive load, and exhibited minimal alterations in left prefrontal activation compared to novices (H. O. Keles & Omurtag, 2021)

Connectivity-oriented analyses extend these observations by showing that novices recruit diffuse and inefficient networks, whereas experts display more synchronized activity across frontal and motor cortices that supports bimanual coordination, with metrics such as phase-locking value and inter-site phase clustering capturing expertise-related differences in functional integration (Keles & Omurtag, 2021; Omurtag et al., 2021; Jian et al., 2017; Bertrand et al., 2016). It has been shown that as skills mature, EEG connectivity becomes more structured and efficient in parallel with reductions in perceived workload, underscoring network-level efficiency as a neural signature of proficiency (Omurtag et al., 2025).

Neural communication relies on the precise synchronization of oscillatory phases between pre-synaptic and post-synaptic neurons, occurring both within localized cortical areas and across distributed brain networks. The Phase Locking Value (PLV) quantifies the strength of neural phase coupling by measuring the consistency of EEG signal synchronization within a specific frequency band (Aydore et al., 2013). This method assumes that the relative phase difference between two connected brain regions remains relatively stable over time. PLV values range from 0 to 1, where 0 indicates complete phase independence, while 1 represents perfect synchronization (Schmidt et al., 2014).

Multimodal EEG–fNIRS systems enhance accuracy and ecological validity because EEG captures fast electrophysiological oscillations while fNIRS measures slower hemodynamics, together providing both temporal dynamics and spatial distribution of cognitive activity during complex tasks. In a study, simultaneous fNIRS and EEG recordings were obtained from participants to examine how physiological responses reflect cognitive effort during training tasks. By analyzing heart rate (HR), heart rate

variability (HRV), and blink rate (BR), the authors identified clear links between autonomic indicators and task-related workload (Zakeri et al.,2020). In another study, the relationship between neural oscillations, multimodal brain signals, and localized hemodynamic activity was studied. Their findings revealed delayed hemodynamic responses in posterior cortical regions, particularly within the alpha band (8–16 Hz), alongside notable beta-band (16–32 Hz) modulations that were interpreted because of alpha–beta coupling in EEG signals (Keles et al. 2016).

Recent studies have moved beyond single-modal classifiers in predicting expertise and workload by integrating combined EEG–fNIRS features with machine learning, leveraging complementary signatures such as EEG phase-locking indices and fNIRS HbO/HbR dynamics to build robust, generalizable models and to dissociate cognitive from motor workload components (Keles, 2022; Omurtag et al., 2024). From a neuroergonomic perspective, mental workload, engagement, and performance co-vary with cortical activity, particularly in prefrontal and parietal regions; optimal engagement occurs in the intermediate workload band, and outcomes are negatively affected when demand is too low or too high (Dehais et al., 2020).

2.5 Subjective and Behavioral-Based Skill Assessment

Among the most frequently used methods for evaluating surgical skill and mental workload are subjective assessments, which capture perceived task demands through structured self-reports (Wilson et al., 2011; Dias et al., 2018). Widely adopted examples include the Global Operative Assessment of Laparoscopic Skills (GOALS), designed to evaluate laparoscopic performance (Vassiliou et al., 2005), the Objective Structured Assessment of Technical Skills (OSATS), developed for structured observation of technical proficiency in surgery (Martin et al., 1997), and the NASA Task Load Index (NASA-TLX), a multidimensional workload rating scale measuring mental, physical, temporal, and performance-related demands (Hart & Staveland, 1988; Wilson et al., 2011). Among these, the NASA-TLX has emerged as the most extensively applied tool in both surgical and cognitive workload research (Dias et al., 2018). Several studies have applied these measures within laparoscopic skill assessment. For instance, trainees

performing standardized simulator tasks reported higher NASA-TLX scores, particularly for mental and physical demands, while also demonstrating lower efficiency compared to more experienced surgeons (Süer et al., 2025).

In addition to subjective metrics, behavioral measures provide objective indicators of performance. These often include primary task outcomes—such as completion time, accuracy, error rate, and instrument handling—and secondary task performance, like reaction time to an auditory cue. Higher cognitive load is typically associated with longer task durations, lower accuracy, and a greater probability of errors (Yurko et al., 2010). A dual-task paradigm involving visual detection during laparoscopic suturing effectively differentiates expertise levels by quantifying cognitive load. In this context, experts demonstrate superior performance in both primary and secondary tasks, supporting the utility of dual-task metrics in mental workload assessment (Zheng et al., 2010). Similarly, single-incision laparoscopic approaches impose significantly higher workloads than conventional multiport surgery, leading to decreases in secondary task performance and increases in primary task errors (Scerbo et al., 2017).

The relationship between workload and surgical performance has been consistently highlighted in the literature. Higher NASA-TLX scores are strongly associated with increased error rates and reduced instrument efficiency, supporting the role of workload monitoring in training optimization (Yurko et al., 2010).

Beyond laparoscopic surgery, research from related domains provides additional evidence. For instance, combining NASA-TLX with performance metrics revealed that novice surgeons reported higher workload and performed less efficiently than experts, confirming the sensitivity of subjective measures to skill level (Carswell et al., 2005). Findings from other contexts, including human-computer interaction, driving, and physically demanding tasks, similarly show that higher subjective workload is consistently associated with slower reactions, reduced situational awareness, and increased strain when cognitive and physical demands overlap (Longo, 2017; Paxion et al., 2014; DiDomenico & Nussbaum, 2008).

Overall, these findings demonstrate that subjective and behavioral measures are valuable tools for assessing surgical skill and workload. While questionnaires such as NASA-TLX provide structured insights into perceived demands, behavioral performance metrics offer objective indicators of efficiency and error propensity. Although limited by their reliance on self-report and task context, these measures remain essential components of workload assessment. Importantly, they establish a baseline for more advanced physiological and neurophysiological methods and, when combined, provide a multidimensional perspective that enhances the reliability and validity of surgical skill evaluation.

2.6 Other Skill Assessment

In addition to traditional behavioral, subjective, and neurophysiological methods, several complementary approaches have been explored to assess cognitive workload and surgical performance objectively. These methods frequently integrate multimodal physiological, ocular, and ergonomic indicators, providing a more holistic view of how surgeons and other professionals manage physical and cognitive demands under high-pressure conditions.

In a study, the neural and psychophysiological responses of participants exposed to stress-inducing and cognitively demanding tasks were examined. The study revealed that increased mental workload was associated with elevated frontal theta and beta EEG activity, reduced HRV, and poorer task performance. These findings suggest that both neural and autonomic signals reliably capture stress-related workload changes, allowing them to guide workload management strategies in operational environments (Mandrick et al., 2016).

Recent studies show that using several physiological measures together helps to assess workload and surgical expertise. During laparoscopic training, increases in task demands have been associated with stronger changes in galvanic skin response (GSR) and higher stress levels in novices, whereas experts tend to remain more stable (Guzmán-García et al., 2022). Evidence also indicates that the combination of

electroencephalography (EEG), heart rate variability (HRV), and GSR provides a more comprehensive understanding of cognitive workload (Diarra et al., 2025). In this context, EEG reflects brain activity, HRV indicates autonomic nervous system regulation, and GSR measures stress responses. Considering these signals together enables the identification of expertise levels and the detection of performance problems under high workload.

Recent reviews highlight that physiological measures can provide valuable insights into workload assessment across different domains. Increased task difficulty has been consistently linked to higher heart rate (HR), lower heart rate variability (HRV), elevated electroencephalography (EEG) activity, and stronger pupillary and electrodermal responses (Pagnotta et al., 2022). The integration of multiple measures, including EEG, HRV, pupillometry, and electrodermal activity, has been suggested to increase the sensitivity and reliability of workload monitoring in applied domains such as healthcare, aviation, and ergonomics (Charles & Nixon, 2019; Tao et al., 2019). Such multimodal approaches may also improve measurement accuracy and temporal resolution, thereby supporting more effective adaptive training strategies and task allocation (Tao et al., 2019)

Experimental evidence reinforces this integrative approach by demonstrating that workload changes produce consistent patterns across HRV, EEG, and pupil dilation. The simultaneous use of multiple modalities enables more precise monitoring of cognitive states in complex human-machine interactions such as surgery (Marinescu et al., 2018). Combining surface EMG (sEMG) and eye-tracking data in laparoscopic surgeons revealed that greater pupil dilation and longer fixations were associated with increased muscle activation, especially when task complexity was high. The link between cognitive and physical demands was more pronounced in novices compared to experts (Zhang et al., 2017).

Competitive training scenarios have also been shown to increase perceived workload (NASA-TLX) and muscle activation, and particularly increase stress in novices (Yu et al., 2015). Complementary ergonomic approaches support these findings. Novice

participants reported higher NASA-TLX scores and exhibited lower tool efficiency than experts. Auditory distractors in the operating room have been reported to significantly increase pupil dilation, error rates, and task times (Gao et al., 2019).

2.7 Machine Learning in Surgical Skill Assessment

The growing complexity of laparoscopic minimally invasive surgical procedures has increased the demand for objective, automated, and reliable approaches to assess surgical skills. Traditional evaluation methods, such as expert observation, structured rating scales, and task completion times, are limited by subjectivity, inter-observer variability, and lack of scalability, which reduce their effectiveness in modern surgical education and certification systems (Lam et al., 2022; Pedrett et al., 2023). Machine learning (ML) has emerged as a powerful solution by enabling the analysis of large, multimodal datasets and extracting patterns that are difficult to identify using traditional assessment techniques (Zia et al., 2018). Through the integration of kinematic, force, video, and neurophysiological data, ML-based frameworks provide quantitative, reproducible, and automated performance assessments that can distinguish expertise levels, predict learning curves, and support personalized surgical training.

At a systems level, a study provides a comprehensive review of IoT-based skill assessment, reporting that force sensors are central to quantifying tool–tissue interaction, often used alongside motion, eye-tracking, and physiological data. They noted that multimodal data fusion—especially when processed with machine learning algorithms such as SVM (Support Vector Machine) or neural networks—consistently improved the classification of skill levels (Castillo-Segura et al., 2021). Building on these developments, predictive modelling approaches have been introduced to estimate individual surgical learning curves using metrics such as task completion time, error rates, and movement efficiency. These models were able to accurately forecast performance trajectories and differentiate between fast and slow learners, highlighting the value of ML for tailoring training programs based on individual progress (Gao et al., 2020).

Kinematic features derived from laparoscopic instruments have become one of the most widely used inputs in ML-based expertise classification. Fard et al. (2016) extracted six kinematic features, including completion time, path length, velocity, smoothness, curvature, and depth perception, from robotic-assisted suturing tasks. Using a Support Vector Machine (SVM) classifier and a leave-one-super-trial-out validation strategy, they achieved 85.3% accuracy in distinguishing novices, intermediates, and experts. Results showed that expert surgeons followed shorter and smoother instrument trajectories, while novices demonstrated longer, less controlled paths, indicating that kinematic motion signatures can objectively capture differences in expertise. In a large-scale study with 50 participants performing virtual neurosurgical simulations, movement efficiency and force control were identified as the most predictive features for skill classification. The machine learning model classified four expertise levels with 90% accuracy. Experts demonstrated more controlled movements and efficient bleeding control, whereas novices showed less stable performance and higher error rates, highlighting that integrating simulation platforms with machine learning techniques can effectively support both objective performance assessment and personalized surgical training (Winkler-Schwartz et al., 2019).

Systematic reviews have provided a comprehensive overview of ML applications in surgical skill assessment. One review of 66 studies reported that kinematic data were used in 40 studies, video-based features in 19 studies, and multi-modal approaches in 7 studies. Among these, multi-modal frameworks consistently achieved the highest accuracies, often exceeding 94% when combining complementary data streams such as video and motion data (Lam et al., 2022). Another review of 50 studies highlighted the rapid adoption of deep learning techniques, particularly convolutional neural networks (CNNs), which were used in 68% of the studies. CNN-based models trained on surgical videos, instrument motion, and robotic datasets achieved classification accuracies ranging between 85% and 97%, outperforming traditional ML methods in complex task environments (Pedrett et al., 2023).

Neurophysiological data have increasingly been integrated into machine learning (ML)-based models to enhance classification performance. High-density functional near-

infrared spectroscopy (fNIRS) recordings have been used to evaluate bimanual motor skills, revealing significant differences in cortical activation patterns between expertise levels. By combining activation data from the prefrontal cortex (PFC), primary motor cortex (M1), and supplementary motor area (SMA), models achieved nearly 100% accuracy in distinguishing experts from novices during laparoscopic tasks (Nemani et al., 2018).

Similarly, high-density fNIRS recordings have been shown to predict both surgical skill levels and subjective mental workload measured using the NASA Task Load Index (NASA-TLX), achieving classification accuracies close to 90% (Keles et al., 2021). Extending these findings, the combination of electroencephalography (EEG) and fNIRS has provided additional insights into error-related brain states during complex suturing tasks. Using EEG microstate clustering combined with fNIRS data, a simple machine learning approach effectively distinguished skill-related brain states. Experts demonstrated focused neural activation and more efficient cortical engagement, while novices exhibited widespread prefrontal activation, reflecting higher cognitive demands during error detection and correction. The study also achieved an accuracy of about 87% in distinguishing experts from novices during laparoscopic suturing tasks (Walia et al., 2022).

The integration of eye-tracking data with neurophysiological signals has further improved performance classification. In robot-assisted surgical procedures, combining EEG metrics with gaze features captured differences in visual attention strategies and cognitive workload. A study focusing on three subtasks — blunt dissection, retraction, and burn dissection — reported classification accuracies of up to 93% when EEG and eye-tracking data were analysed together, outperforming models based on single-modality inputs (Shafiei et al., 2023). Expanding on this approach, a multi-class ML model was developed to simultaneously classify three surgical subtasks and three levels of expertise, achieving an overall accuracy of 88.5% using extreme gradient boosting (XGB) algorithms (Shafiei et al., 2024).

These findings demonstrate that combining cognitive, visual, and behavioral features substantially improves the accuracy of ML-based assessment frameworks, providing richer insights into expertise levels, workload, and attentional strategies compared to single-modality approaches.

Deep learning approaches have further advanced surgical skill assessment by enabling automated feature extraction from complex multimodal datasets. In robotic-assisted procedures, deep learning-based models have been used to track instrument motion and quantify performance, achieving high accuracy in differentiating expertise levels (Lee et al., 2020). Convolutional neural networks (CNNs) applied to EEG topographic features during laparoscopic suturing tasks have also demonstrated superior accuracy compared to traditional linear models, highlighting the value of spatiotemporal brain representations for distinguishing experts from novices (Manabe et al., 2023). Similarly, EEG-based neural network models trained on virtual neurosurgical simulations achieved near-perfect discrimination between skilled and less-skilled participants, with spectral features such as low-alpha power, beta activity, and theta/beta ratios identified as key predictors of expertise (Natheir, 2021).

In parallel, the introduction of large, multimodal datasets has enabled benchmarking of ML models in surgical education. Recent initiatives that integrate neuroimaging signals with standardized performance metrics provide valuable resources for expertise classification, learning curve modelling, and the development of adaptive training strategies (Kamat et al., 2023). Multi-modal frameworks that combine EEG, fNIRS, video-based kinematics, and eye-tracking are increasingly used to provide a more comprehensive assessment of motor performance and cognitive load. Classification performance improves when EEG and fNIRS are combined, achieving higher accuracy (EEG: 71%, fNIRS: 64%) compared to using each modality alone, as they offer complementary insights into brain activity (Sánchez-Reolid et al., 2024).

Advances in optical neuroimaging and non-invasive neuromodulation techniques, such as transcranial direct current stimulation (tDCS), have also contributed to this progress. Combining high-density fNIRS with ML enables objective monitoring of cortical

activation and mental workload during surgical tasks, while tDCS has shown potential to accelerate motor skill acquisition and improve performance under stress (Goble et al., 2023).

3. MATERIALS AND METHOD

3.1 Participants

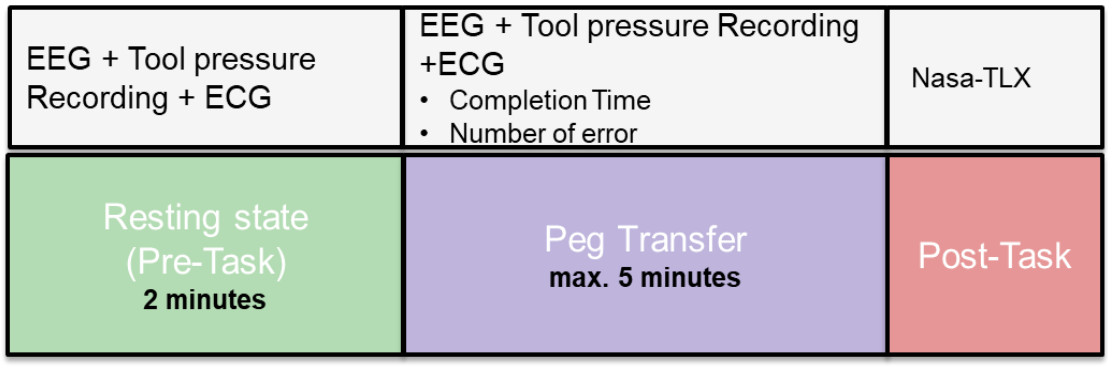
The study involved two groups of participants: 10 surgeons with diverse levels in laparoscopy practice, and 13 students without laparoscopy experience. Participant demographics are summarized in Table 3.1. All participants were informed with written consent before the experimental session. The research protocol was approved by the Ankara University Human Research Ethics Committee (AUTFKAEK No: 2024/333), and the study was conducted at the Neuroscience and Neurotechnology Center of Excellence (NÖROM) in accordance with relevant ethical guidelines.

Table 3.1 Demographics of participants

	Female	**Male**	**Age**	**Lap. Experience (Operation)**
Surgeons	4	6	29 ± 3	33 ± 37
Non-experienced	4	9	26±6	0

3.2 Experimental Design

Table 3.2 Experimental design



The experiment was carried out at the Neuroscience and Neurotechnology Center of Excellence (NÖROM) using a custom-built laparoscopy simulation box, a single monitor, and in-house designed pressure sensors. The total duration of the experiment was approximately 30 minutes, which included both task performance and the setup of the system and devices. At the beginning, participants were given time to complete a practice trial to familiarize themselves with the surgical instruments and to receive an overview of the experimental procedures.

For this study, a laparoscopic training box was specifically designed. A camera was mounted inside the simulator, with a training table positioned beneath it and centered within the camera's field of view. The live video feed was presented on a monitor placed directly in front of the participant, aligned with their line of sight. Additionally, custom-made pressure sensors were attached to both the right and left laparoscopic instruments to capture tool interactions during the task. One of the FLS (Fundamentals of Laparoscopic Surgery) tasks, the peg transfer task, was replaced inside the box. This task is a fundamental exercise that forms the foundation of laparoscopic surgical skills and aims to improve surgeons' ability to manipulate objects using their hands. It is designed to enhance surgeons' capability to perform precise and coordinated movements with laparoscopic instruments. For achieving the task, participants were required to transfer three rings from one rod to another to complete the task. The peg transfer task was arranged using four rods labeled A, B, C, and D, positioned at the lower left, upper left, upper right, and lower right corners, respectively (Figure 3.1). Participants started with three rings on rod A. Using the left grasper, they picked up a ring, transferred it mid-air to the right grasper, and placed it on rod B. They then picked up each ring from rod B with the left grasper and moved it to rod C using the right grasper. Finally, rings from rod C were transferred one by one with the left grasper and placed on rod D. If a ring was dropped onto the platform, participants could return it to the tray and continue the task. Dropping a ring onto the platform during transfer was recorded as an error. Completion time and the number of errors were recorded for each participant. Participants began the task after a two-minute rest period. Completion times and errors made during the task were recorded.

In addition to the experimental setup, handmade pressure sensors were integrated into both the right and left laparoscopic tools. These sensors precisely measured the pressure applied by the participants' thumbs during task execution. A series of metrics were designed to capture distinctions in skill level and training. Active time segment was defined as those periods when pressure exceeded a specified threshold, indicating engagement. To measure bilateral coordination, the Right-Left Overlap metric was introduced. It is defined as the duration in which pressures from both the right and left sides are simultaneously active. Additionally, the Right-Left Asymmetry metric was developed, which is calculated as **|R-L| / (R+L).** Here, R and L represent the time-averaged amplitudes of the right and left tool pressures, allowing us to quantify any imbalance between sides.

These metrics provide robust insights, and variations in threshold settings were found to have minimal impact on their reliability. Furthermore, the calculated measures remained stable across a wide range of threshold values, ensuring consistent results regardless of threshold adjustments (Modi et al., 2017).

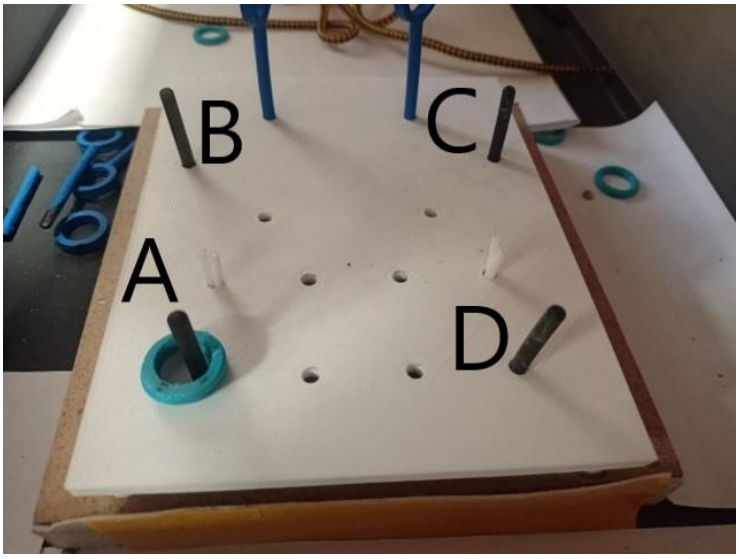


Figure 3.1 Peg transfer task

3.2.1 Sensor design

Accurate measurement of instrument forces in laparoscopic surgery is a key element in objectively assessing technical performance and preventing inadvertent tissue damage. Finger pressure measurements are essential to determine whether the force applied to laparoscopic graspers is excessive or insufficient. Laparoscopic tools (graspers) have handles where the surgeon applies the necessary pressure to grasp or cut tissue. The applied pressure can be measured at two main points: handle grips or directly at the instrument tips (Van den Dobbelsteen et al., 2012). For this project, the handle grip was chosen as the sensor insertion point.

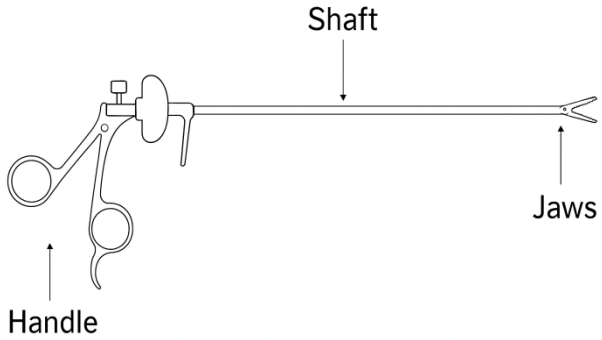


Figure 3.2 Laparoskopic tool (grasper) schematic

In the handles, one part is stable while the other is movable. The movable part is controlled by the thumbs (Rajakumar & Varadhan, 2022). Thumbs are placed on the handle grips and press against the inner side of the grip. In this project, to sense the pressure, pressure sensors are attached to the inner side of the grips. The grips' shape and hole size should not be changed to maintain the same conditions, so very thin sensors are used. These sensors are flexible force sensors (Drewniak et al., 2007). They are usually resistive type, and their resistance decreases according to the applied force. The resistance change is detected by an amplifier circuit as shown in Figure 3.3 (Almassri et al., 2014). A constant voltage is applied to the flexible pressure sensors, and the current changes according to the change in resistance. This current is converted to voltage by a current-to-voltage converter amplifier. This analog voltage value is then

converted to digital by a microcontroller with an analog-to-digital converter module. Thus, the final data represents the applied force in the form of digital values.

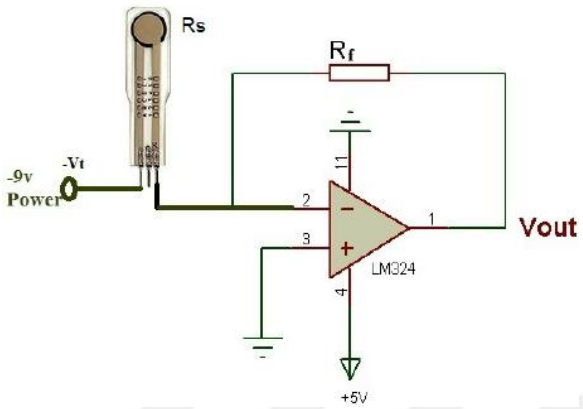


Figure 3.3 Circuit for pressure measurement using flexible force sensor

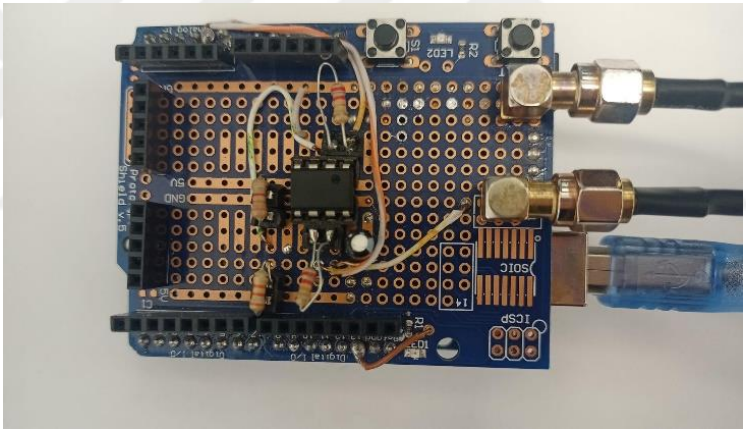


Figure 3.4 Custom-designed electronic circuit board used for pressure sensor data acquisition

3.3 Data Collection

In this study, data were collected from both participant groups, surgeons and inexperienced, in Neuroscience and Neurotechnology Center of Excellence (NÖROM), Virtual Reality laboratory. For the purpose of this research, sensor data from the tool device was collected simultaneously with electroencephalography (EEG) and electrocardiography (ECG) recordings throughout the experimental session. To measure brain and cardiac activity, the Mentalab Portable EEG device, a wireless eight-channel

system with a sampling rate of 250 Hz, was used. One channel was assigned for recording ECG data, while the remaining seven channels were utilized to gather EEG data related to brain activities. Electrode placement followed the international 10–20 system, with seven channels which positioned at Fz, F3, F4, C3, C4, P3, and P4. This electrode arrangement facilitated the examination of neural activities across various cortical regions, including the frontal lobe which is associated with decision-making and attention, the central lobe which is involved in motor activities, and the parietal lobe which is responsible for sensory processing. Additionally, heart rate variability (HRV) was analysed from the ECG data. As behavioral data, task completion times and reaction times of the participants were recorded. As part of the setup, both the left and right tool handles were fitted with in-house pressure sensors to monitor the thumb force applied by participants while performing the task.

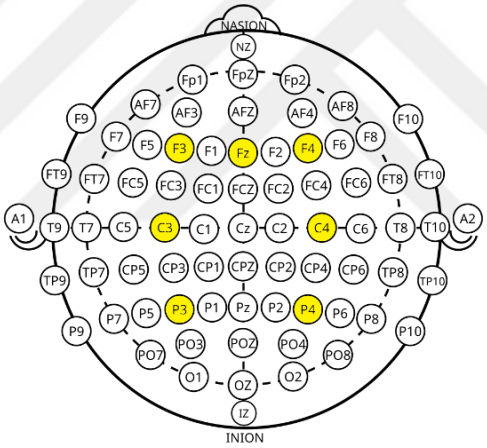


Figure 3.5 Electrode placement map based on the international 10–20 system

After the experiment, participants were asked to fill out the NASA Task Load Index (NASA-TLX) questionnaire. This multidimensional tool determines mental workload across multiple subscales each rated on a 20-point visual analog scale, providing insight into task demands and participant–task interaction.

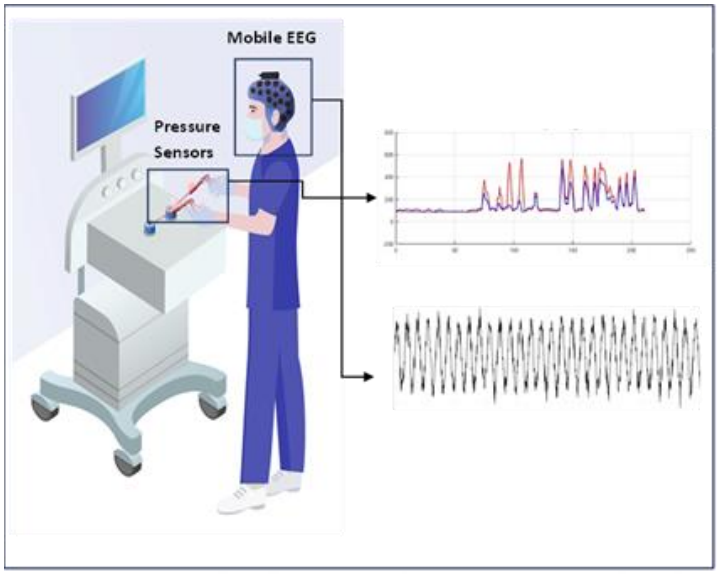


Figure 3.6 Experimental setup of the multimodal data acquisition system. The surgeon performs laparoscopic box trainer tasks while EEG signals are recorded via a mobile EEG device and pressure sensors capture hand force data from the laparoscopic tools. Representative EEG waveforms and force signal outputs are shown on the right.

3.4 Data Analysis

3.4.1 Behavioral and performance data

Task completion times for the peg transfer exercise were recorded. Participants assessed their workload using the NASA-TLX, which evaluates six subscales: mental demand, physical demand, temporal demand, effort, frustration, and performance. Each subscale was rated on a 20-point visual analog scale, and an overall mental workload score was calculated by summing the subscale ratings. Linear regression analyses were conducted to explore the relationships between participants' task completion times, reaction times, NASA-TLX scores, HRV metrics, and other relevant variables.

3.4.2 Sensor measures

Several quantitative metrics were extracted from the force sensor signals to assess motor performance, focusing specifically on tool engagement patterns and inter-hand coordination during laparoscopic tasks. Force data were acquired simultaneously from

the right and left grasper handles, sampled at 20 Hz, and expressed in arbitrary pressure units.

A pressure signal from each handle was classified as active when it exceeded a predefined threshold (θ). This threshold was determined prior to analysis and applied consistently within each trial to distinguish periods of active tool manipulation from rest. Based on this definition, the following metrics were calculated:

Active Duration (AD): The total time, expressed in seconds, during which each handle remained active. Separate values were computed for the right (AD_R) and left (AD_L) handles. This measure captures the temporal extent of tool engagement for each hand individually and can reveal tendencies such as prolonged reliance on one hand over the other.

Right–Left Overlap (RLO): The cumulative time during which both hands were simultaneously active, i.e., both pressures exceeded the threshold. This metric reflects the degree of bimanual engagement and the extent to which both hands contributed concurrently to the task. Higher RLO values indicate more synchronous hand usage, which is often associated with improved bimanual coordination in skilled performance.

Mean Active Force (R, L): The average pressure amplitude recorded during active periods for the right (R) and left (L) handles, respectively. By focusing only on active periods, this metric avoids diluting force values with inactive segments and provides a more accurate representation of effective force application.

Right–Left Asymmetry (RLA): A normalized difference between the mean active pressures of the two hands, defined as:

$$RLA = \frac{\mathrm{R} - \mathrm{L}}{R + L}$$

where R and L correspond to the mean active pressures for the right and left handles, respectively. Positive RLA values indicate right-hand dominance, while negative values indicate left-hand dominance. An RLA value near zero reflects balanced force application between the two hands.

All metrics were calculated on a per-trial basis, allowing for within-subject and between-subject comparisons. The resulting data were stored in structured tables including participant identifiers, AD_R, AD_L, RLO, R, L, and RLA values. These tables were subsequently aggregated across trials and exported in spreadsheet format for statistical analysis. This approach ensured that both temporal (duration-based) and amplitude (force-based) aspects of motor performance were systematically quantified, enabling comprehensive evaluation of tool handling proficiency.

3.4.3 EEG measures

3.4.3.1 Preprocessing

To remove non-neural artifacts and preserve brain activity signals for future analysis, the raw EEG data underwent a series of pre-processing steps. Segments containing high-frequency, high-amplitude fluctuations—likely caused by gross body movements—were identified and deleted using a sliding window approach and removed. Also, channels exhibiting kurtosis values greater than 5 were considered invalid and excluded from the dataset. The remaining signals were then band-pass filtered between 0.16 Hz and 40 Hz to eliminate slow drifts and high-frequency noise.

3.4.3.2 Power spectral analysis

Power spectral density (PSD) analysis was performed to extract frequency-specific information from EEG signals. The spectrogram was obtained using short-time Fourier transform with 1-s windows, 50% overlap, and a frequency resolution of 1 Hz. Power values were integrated over the standard frequency bands delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), and beta sub bands ranging from 12–32 Hz. A cutoff point was

applied at 32 Hz to avoid artifacts caused by electromyographic activity. For each participant and trial, band-specific power values were averaged across task epochs to represent neural oscillatory activity during laparoscopic performance. These power characteristics are generally associated with mental workload, with enhancement in the power of theta wave and lowers in alpha power being associated with higher cognitive demands and increased attentional focus.

3.4.3.3 Functional connectivity analysis

Phase Locking Value (PLV) is a extensively used method for measuring phase synchrony between two EEG signals, reflecting the consistency of phase differences over time. In cognitive neuroscience, it is frequently used to assess the connectivity between functions and different parts of the brain and provides information about network-level coordination during a task (Lachaux et al., 1999; Vinck et al., 2011).

Functional linkage between cortical regions was quantified using the Inter-Site Phase Clustering (ISPC), also known as the Phase Locking Value (PLV). ISPC was calculated for each not sorted electrode pair 1 Hz-wide frequency bands centered at integer frequencies between 1 Hz and 40 Hz. This measure reflects the strength of phase-based synchronization between electrode sites, providing an estimate of functional coupling between underlying neuronal populations

The cortical distance for each ISPC value was obtained using the Montreal Neurological Institute (MNI) coordinates of the neuronal populations corresponding to each electrode. Narrow-band frequency analysis was utilized to preserve detailed spectral information and obtain more accurate phase estimates. This approach aimed to hinder the probable loss of information that might occur when combining broad frequency bands.

To tone down the impact of volume conduction, a Laplacian fitting was applied, in which the instantaneous signal from each electrode was re-referenced to the average of its proximate neighbors. Substitutional methods unresponsive to volume conduction,

such as the weighted phase lag index, were not used because both modeling and experimental evidence indicate that zero-delay synchronization can occur even when conduction delays are present and that this can have a critical role in visuomotor performance (Roelfsema et al., 1997).

ISPC was chosen over other connectivity measures because of its high within-subject test-retest reliability and specificity for phase relationships. In contrast, coherence—a commonly used alternative—incorporates band power to weight individual phase angles. This choice allowed the analyses to emphasize synchrony based solely on phase matching, independent of power fluctuations.

3.4.3.4 Feature extraction

Various features representing different aspects of brain activity were extracted from EEG data. The main features used in this study are Power Spectral Density (PSD) and Phase Locking Value (PLV).

PSD indicates how can the EEG signals power distributed across its frequency components and is widely used to quantify brainwave activity in specific frequency bands. The calculations of the spectrogram was completed with the help of the Short-Time Fourier Transform (STFT) using 1-second windows, 50% overlap, and a frequency resolution of 1 Hz. As a result of calculations, power yield to eight different frequency bands, each with a width of 4 Hz, covering the range from 0 to 32 Hz. These bands were labeled as delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), continued as five beta sub-bands differ between low beta (12–16 Hz) to high beta (28–32 Hz). For convenience, they were also represented as f1, f2, …, f8. The power of each EEG frequency band was calculated for every epoch through integration of the signal's power across the respective frequency range. The cutoff frequency of 32 Hz was imposed because in EEG analysis stronger frequencies are not generally viewed as informative related to cortical activity (Goncharova et al., 2003).

As a measure of phase synchrony, PLV was derived from chosen EEG electrode pairs to estimate interactions between distinct neuronal populations and quantify inter-area synchrony (Vinck et al., 2011). The types of synchronies were paired with electrodes and represented as follow:

- Intra-hemispheric electrode pairs: F3–P3, F4–P4, Fc1–Cp1, Fc2–Cp2, Fz–Poz
- Symmetric inter-hemispheric pairs: F7–F8, F3–F4, Fc1–Fc2, C3–C4, T3–T4, Cp1–Cp2, P3–P4, O1–O2
- Asymmetric inter-hemispheric pairs: F3–P4, F4–P3, Fc1–Cp2, Fc2–Cp1

Phase-locking value (PLV) was computed for four frequency bands of interest: delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), and beta sub bands ranging from 12–32 Hz.

3.4.4 Statistical and machine learning analyses

In order to assess the statistical significance of the difference between two groups of non-paired results, we used the non-parametric Mann-Whitney test. We did not utilize null hypotheses whose rejection would have required corrections for multiple comparisons or false discovery. When conducting a regression analysis comparing 2 numerical variables, linear fit with analysis of variance was used. The descriptive results comparing two groups, such as asymmetry with NASA TLX subscores; completion time with NASA TLX subscores.

EEG and sensor data are used to train various machine learning algorithms for distinguishing between surgeon and control groups across different frequency bands. We considered that having too few folds may not have a sufficient number of partitions to reveal the true accuracy, while too many folds may have an insufficient number of observations per partition for proper training. We found no clear differences between these cases; however, 5-fold was selected as a potential compromise. To ensure

balanced evaluation, Stratified 5-fold cross-validation is applied, dividing the data into five parts where each part serves as a validation set once, allowing the model to train five times. Before proceeding with the full set of computations, we considered how the accuracy of a 3-fold, 5-fold, and 7-fold cross-validation would differ by performing a limited set of classifications to discriminate surgeons from normal volunteers.

We considered that having too few folds may not have a sufficient number of partitions to reveal the true accuracy, while too many folds may have an insufficient number of observations per partition for proper training. We did not use leave-one-out cross-validation as this scheme has only one possible partition and does not allow one to generate distribution and may generate biased performance estimates each feature was standardized by centering and scaling with the mean and standard deviation of the corresponding column of the feature matrix. The linear kernel was selected for Support Vector Machine and its scale was computed by MATLAB using a heuristic procedure. We used a fixed random number of seed for the reproducibility of the results.

Due to lack of data, Feature selection is conducted using the SelectKBest method with different criterias such as chi square and mutual information. These methods are a test that measures the degree of dependence of features on the target variable and are used to select features, especially in categorical data and classification problems. This test evaluates the relationship of each feature with the class label; small p-values indicate that the feature has a strong relationship with the target, while large p-values indicate independence. In this way, features that do not have a strong relationship with the target variable can be eliminated and a better model can be created. This step ranks the best features and selects the top 5 for the dataset. By selecting only the most relevant features, the model achieves improved accuracy and reduced processing time.

GridSearchCV is utilized to identify optimal hyperparameters for each model. It systematically tries different combinations of certain hyperparameters to improve the performance of the model and finds the combination that gives the best result. It evaluates the performance metrics such as accuracy and F1 score of the model for each hyperparameter combination using the classical cross-validation method. This process

helps to choose the most appropriate values of the hyperparameters to improve the overall accuracy of the model. GridSearchCV saves time and provides more efficient results, especially by automating the hyperparameter settings instead of manually selecting them. For instance, 'C', 'kernel', 'gamma', and 'degree' parameters are fine-tuned for the SVM model, while 'n_estimators' and 'criterion' are optimized for the Random Forest model. The cross-validation accuracy results guide the selection of the best hyperparameters, ensuring performance optimization as each model's accuracy is reported as the mean of cross-validation outcomes, ultimately aiming for the highest success rate with these configurations.

4. RESULTS

We collected data from 23 voluntary participants (12 females, 11 males). Each participant first completed a resting period, followed by the peg transfer task, a laparoscopic training task which is widely used to assess bimanual coordination. EEG and grip force data were continuously and simultaneously recorded throughout the task. During the task, the completion time and the error were recorded. After task completion, participants were required to complete the NASA-TLX questionnaire for evaluation of their comprehension of cognitive workload.

Our analyses proceeded in three main steps. First, we examined participants' behavioral performance metrics, including task completion time and error rates. Second, we explored the relationship between behavioral performance, sensor-based force metrics, and self-reported workload subscales from the NASA-TLX. Finally, Shapiro evaluated and compared multiple machine learning classification models using combinations of behavioral and sensor-derived features to predict participants' skill level groupings.

4.1 Behavioral and Performance Results

First, behavioral performance and physiological measurements were compared between participants with surgeons (Surg) and non-experienced (Non-Exp) groups. In Figure 4.1, the number of errors experienced in the peg transfer task was substantially less in the surgeon group compared to the non-exp group ($p < 0.05$), indicating that surgical expertise was strongly associated with higher accuracy.

In addition to this significant difference in accuracy, other measures also showed favorable but non-significant trends for the Surg group. Task completion time was generally shorter among surgeons, indicating more efficient performance on the peg transfer task. Furthermore, heart rate variability (HRV) values were slightly higher in the Surg group, suggesting that surgeons may have tended to adapt their physiological

regulation more effectively during the task. However, this difference was not statistically significant.

Total NASA-TLX scores, reflecting the overall subjective workload, did not make great difference between the two groups. It suggested that despite their superior performance, surgeons perceived the task as comparably demanding as non-experienced group.

Overall, these results demonstrate that surgical training and experience are most clearly reflected in improved task accuracy, while efficiency and physiological adaptability show consistent but non-significant advantages.

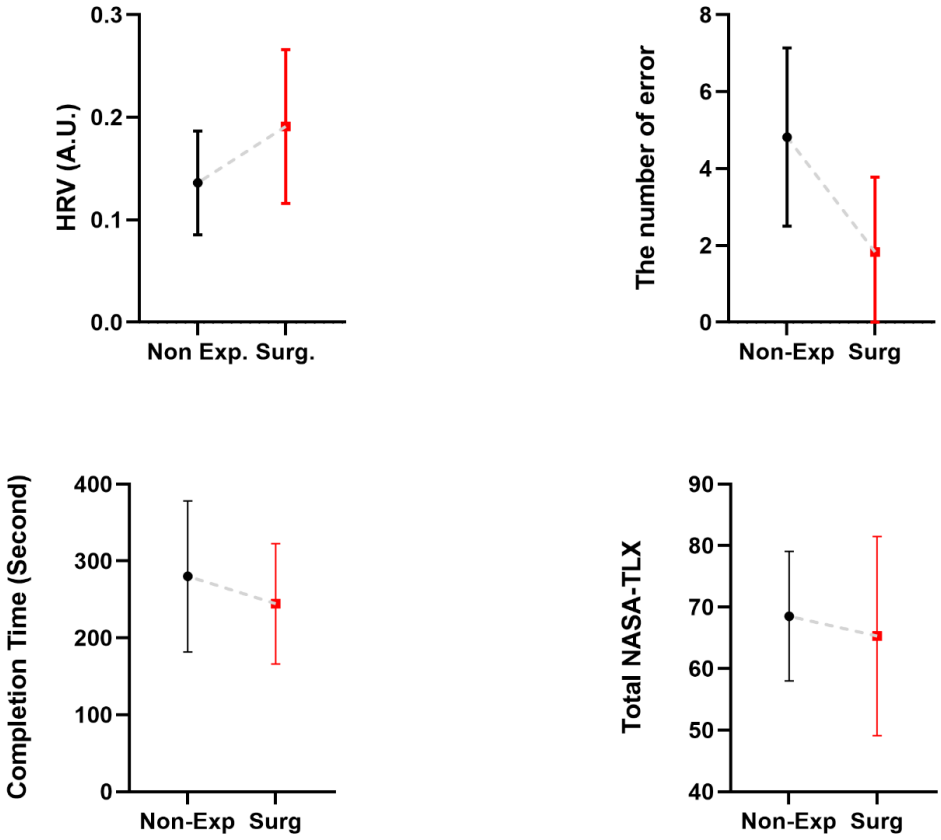


Figure 4.1 Comparison between surgeons (Surg) and non-experienced (Non-Exp) participants across four measures: HRV, number of errors, task completion time, and total NASA-TLX scores Metrics include heart rate variability (HRV), number of errors, task completion time, and total NASA-TLX scores.

4.2 Force Sensor Data Results

The raw force signals recorded from both grasper handles during the peg transfer task are illustrated in Figure 4.2. Each trace represents the pressure exerted by the right and left thumbs over time. A threshold of approximately 90 units was applied to distinguish between active and inactive periods of tool engagement. Segments exceeding this threshold were defined as active force application intervals, whereas lower values indicated resting phases. This allowed the identification of discrete episodes of tool use, including onset and offset points for each hand. The figure also demonstrates that both groups exhibited alternating patterns of right- and left-hand activation, with periods of simultaneous force application.

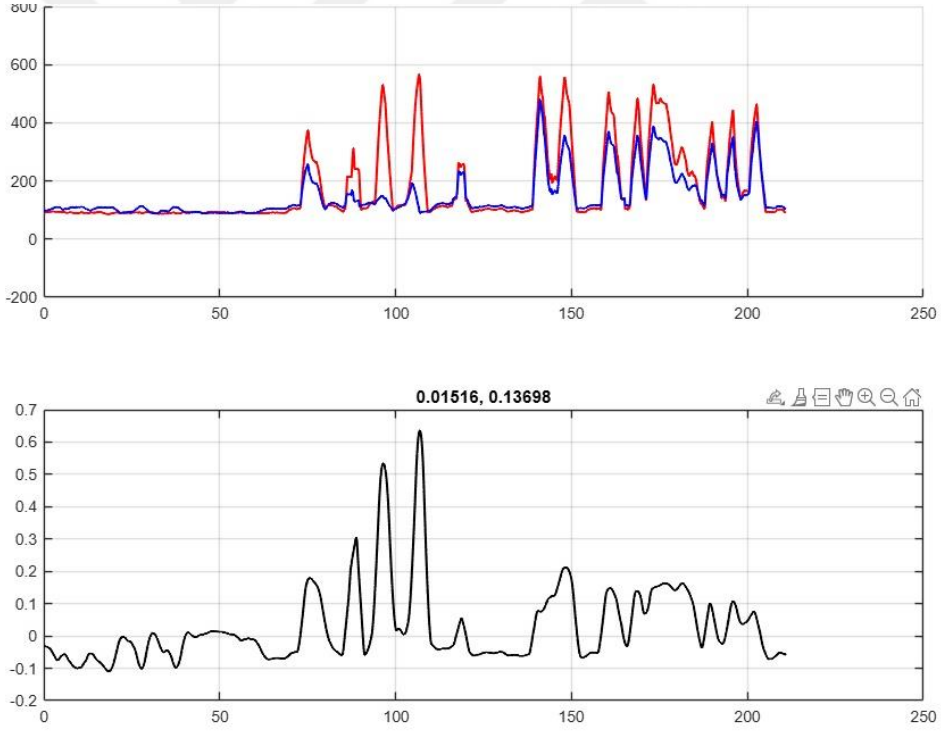


Figure 4.2 Force signals recorded from the right and left grasper handles during the peg transfer task. The lower panel presents the normalized and thresholded force signals.

Time-domain features derived from force signals recorded during the peg transfer task were comprehensively evaluated. These features included maximum slope, peak and

average force amplitudes, force variability (standard deviation), force time integral (FTI), number of peaks, time to peak, and smoothness.

There is not any remarkable differences was identified in both groups related to these parameters. This suggests that both surgeons and inexperienced participants exhibited similar temporal and amplitude-based motor characteristics throughout the task.

Then, coordination-related force metrics were examined: the asymmetry index, right–left (R–L) overlap, and active force duration (Fig. 4.3). The asymmetry index, which represents the normalized imbalance between left- and right-hand pressures, showed a trend toward significance ($p = 0.06$). The non-experienced group exhibited higher asymmetry values (0.087 ± 0.052) compared to the Surgeon group (0.030 ± 0.031). This difference is also clearly evident in the bar graphs; non-experienced participants consistently had a higher mean value with a wider margin of error. This suggests greater variability in their bimanual coordination. In contrast, surgeons appear to have more balanced and stable bilateral force control, with lower asymmetry values and narrower error ranges.

The right-left overlap represents the total time (in ms) required to use both hands simultaneously. This time was higher in the surgeon group (183.69 ± 55.94 ms), while non-experienced participants recorded 168.68 ± 49.94 ms. The corresponding plot illustrated this difference, with surgeons showing a slight upward shift and broader variability, suggesting more frequent but also less uniform bilateral engagement. Although this discrepancy was not statistically notable, the trend indicates that surgeons tend to sustain more coordinated bimanual activity during the task.

Finally, no standout difference was obtained in between both groups related to active force duration, which represents the total number of force instances exceeding a predetermined threshold. Both groups displayed similar values (non-experienced: 132.14 ± 47.22; Surgeon: 134.75 ± 62.18). The wide error bars in both groups highlight considerable inter-individual variability, implying that overall force exertion strategies were not substantially different between surgeons and non-experienced participants.

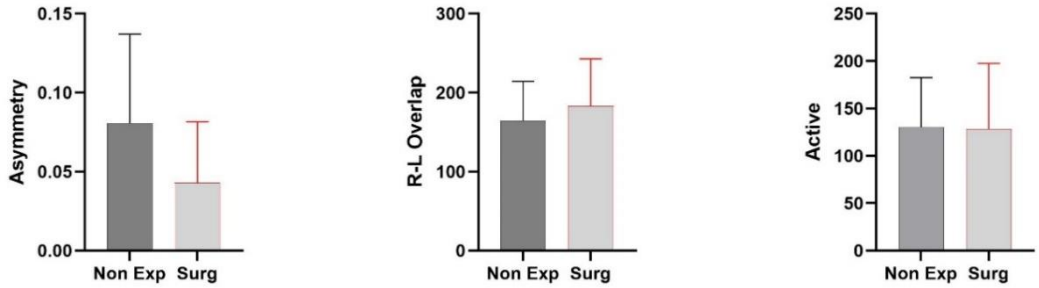


Figure 4.3 Sensor Comparison of sensor-derived force metrics between surgeons (Surg) and non-experienced participants (Non-Exp). A trend toward lower asymmetry was observed in the Surg group ($p = 0.06$), suggesting more balanced bimanual coordination. No significant differences were found for R–L overlap duration or active force duration. Error bars indicate standard deviation.

4.3 Inter-Variable Relationship Assessment

To analyzed task performance and perceived workload, regression analyses were conducted between NASA-TLX subscale scores and two predictors: task completion time and asymmetry index (Fig. 4.4). Each workload dimension was assessed separately to determine how subjective scores were affected by performance time and bimanual coordination.

For completion time, mental demand exhibited a slight positive slope, indicating that participants requiring longer durations tended to report higher cognitive effort, although this relationship was not statistically significant. Physical demand also showed a similar positive relationship, suggesting that prolonged engagement was related to increased muscular strain, though not at a statistically robust level. Temporal demand demonstrated a weak negative association, with faster performers reporting relatively higher temporal stress compared to slower performers. Also, this effect did not reach significance. Performance evaluations were not affected by task duration. In contrast, effort ($p = 0.01$) and frustration ($p = 0.01$) both showed significant positive correlations with task duration, indicating that longer completion time was associated with increased subjective effort and emotional burden.

For asymmetry index, mental demand showed a weak negative slope, which was not statistically significant. Physical demand demonstrated a significant positive association

($p < 0.05$), with greater imbalance between hands corresponding to higher physical workload. Temporal demand showed a modest positive slope that was not statistically significant. Performance ratings did not vary with asymmetry, while effort showed only a weak positive relationship ($p > 0.1$). Frustration did not demonstrate any meaningful association with asymmetry.

In summary, task completion time was significantly associated with effort and frustration, while the asymmetry index was only significantly associated with physical demand. The other workload dimensions did not show statistically significant relationships with either predictor.

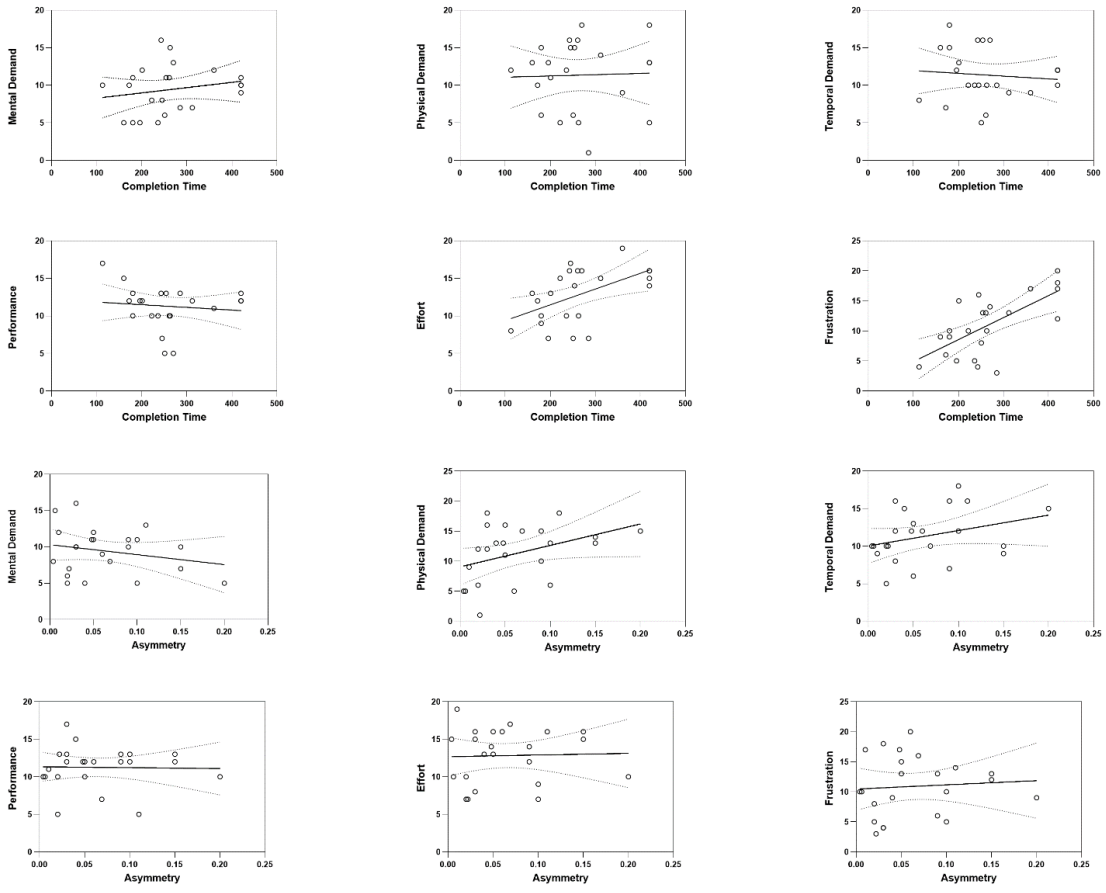


Figure 4.4 Simple Linear regression analyses examining the relationships between NASA-TLX subscale scores and two predictors: task completion time (top two rows) and tool-use asymmetry index (bottom two rows).

4.4 Classification Results

Classification performance was systematically evaluated using Power Spectral Density (PSD) and Phase Locking Value (PLV) features obtained across four key frequency bands (delta, theta, alpha, and beta) of the EEG in both task and rest conditions. To ensure robustness, three different supervised learning algorithms were applied: Random Forest Classifier (RFC), Gaussian Process Classifier (GPC), and AdaBoost Classifier (ABC). The predictive capacity of three different feature sets was examined: (1) asymmetry index only, (2) PSD features combined with asymmetry, and (3) PLV features combined with asymmetry (Fig. 4.5).

When only the asymmetry index was used as the input feature, classification accuracies varied significantly between models. The RFC achieved an accuracy of 78%, while ABC similarly achieved the highest value at 79%. In contrast, the GPC showed a substantially lower performance with 60% accuracy. These findings suggest that asymmetry-based motor coordination features by themselves carry moderate discriminative power for differentiating between expertise levels, with ensemble-based methods such as RFC and ABC yielding the most reliable outcomes.

The addition of EEG spectral power features to the asymmetry index further improved classification accuracy. In the PSD + asymmetry condition, accuracy values increased across most classifiers, particularly in the alpha and theta bands, where RFC and ABC consistently outperformed GPC. This pattern suggests that spectral dynamics, when integrated with motor coordination measures, enhance the ability to capture differences between novice and expert participants.

Similarly, the inclusion of PLV features with asymmetry provided an additional discriminative signal. While improvements were less pronounced than those observed with PSD features, notable gains were observed in the beta band, where RFC achieved the highest classification rates among all tested combinations. This finding suggests that functional connectivity, as represented by PLV, provides complementary information to motor asymmetry for expertise classification.

Overall, the findings suggest that classification based solely on asymmetry has moderate but significant discrimination, while integrating EEG-based PSD and PLV features with asymmetry slightly improves classification performance, particularly in the theta band. Limited improvement was determined with RFC in the alpha band, while contributions were weaker in the beta and delta bands. The most consistent and high improvements were achieved by the RFC and ABC classifiers.

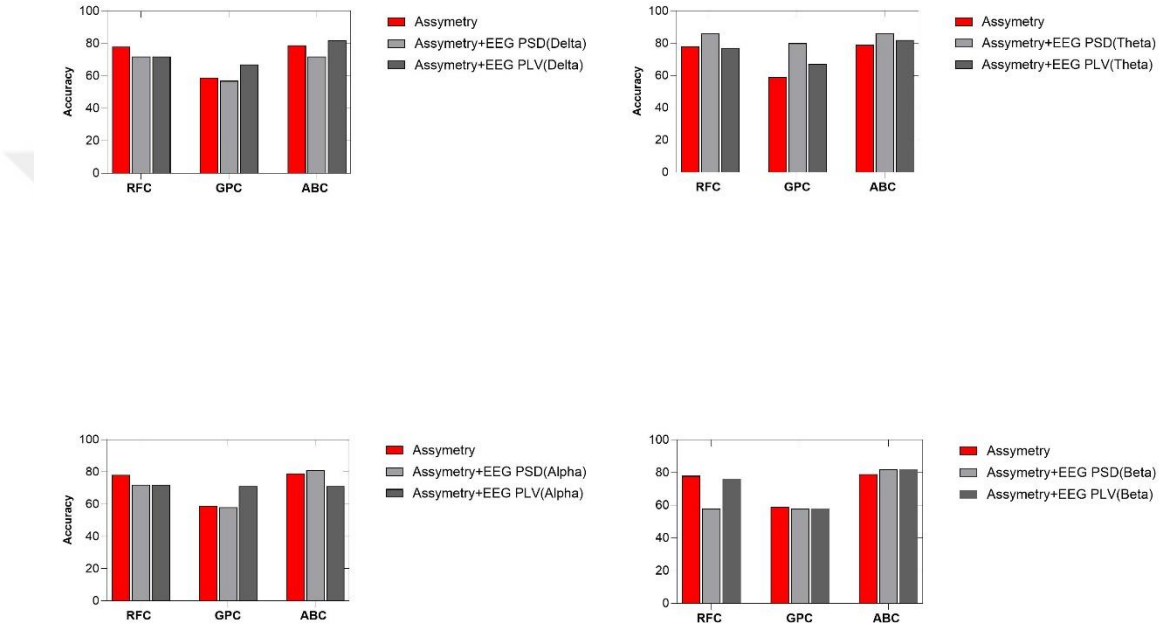


Figure 4.5 Classification accuracies across four EEG frequency(bands delta, theta, alpha, and beta) using three different classifiers: Random Forest Classifier (RFC), Gaussian Process Classifier (GPC), and AdaBoost Classifier (ABC). Three feature sets were compared: asymmetry-only (red), EEG power spectral density (PSD) combined with asymmetry (light gray), and EEG phase locking value (PLV) combined with asymmetry (dark gray).

Table 4.1 shows the classification performance of three machine learning models—Random Forest Classifier (RFC), Gaussian Process Classifier (GPC), and AdaBoost Classifier (ABC)—trained using a combination of EEG power band features (δ – delta, θ – theta, α – alpha, β – beta) and motor asymmetry metrics under both task and resting-state conditions. Model evaluation was carried out using accuracy, F1 score, precision, recall, and AUC score to provide an overview of performance.

During task condition, the strongest classification was observed in the theta band combined with asymmetry features. Both RFC and ABC achieved the highest accuracy

(0.86), with F1 scores of 0.83 and AUC values of 0.92 (RFC) and 0.85 (ABC), respectively. These outcomes are coherent with the established role of theta activity in attentional control and processes and suggest that theta oscillations are strongly engaged during complex motor execution. In contrast, GPC performed slightly lower (accuracy = 0.80, AUC = 0.79), but was found to be competitive range.

For the delta band under task condition, RFC and ABC reached moderate performance levels (accuracy = 0.72 for each), while GPC dropped to 0.57 accuracy, reflecting weaker generalization. The alpha band yielded mixed results: ABC achieved a relatively high accuracy of 0.81 with strong precision (0.93), whereas RFC lagged at 0.72 accuracy and GPC at 0.58 accuracy. A clear distinction was observed in the beta band. While ABC maintained robust accuracy (0.82, AUC = 0.82), both RFC and GPC were considerably weaker (accuracy = 0.58 each), with GPC in particular showing very poor discriminative power (AUC = 0.45).

In the resting-state condition, overall performance was lower compared to the task-based features. However, ABC consistently produced the best results across bands, with accuracies ranging from 0.77–0.82 and high F1 and AUC values (e.g., δ band: accuracy = 0.82, AUC = 0.87; θ band: accuracy = 0.82, AUC = 0.82). In contrast, RFC and GPC, exhibited lower and more unstable outcomes, particularly in the alpha band where RFC accuracy dropping to 0.52 (AUC = 0.67). These findings indicate that EEG signals recorded during task engagement contain richer and more discriminative information than those obtained during resting state.

Overall, ABC consistently demonstrated high and stable performance, particularly in terms of F1 and AUC scores, highlighting its robustness in modeling non-linear feature interactions. RFC also produced competitive results in specific bands (e.g., θ task condition), but its performance fluctuated across conditions. In contrast, GPC generally exhibited lower accuracy and poor generalization, particularly in the beta-band under task conditions where its AUC score dropping to 0.45.

In summary, models trained with task-related EEG and asymmetry features outperformed those using resting-state data, emphasizing the value of active engagement signals for distinguishing surgeons from non-experienced participants. Theta band features obtained during the task emerged as the strongest predictor, while ABC provided the most reliable classification across all frequency bands.

Table 4.1 Classification performances using EEG power features and asymmetry indices for the δ, θ, α, and β frequency bands under task and rest conditions. The table shows the accuracy, F1 score, precision, sensitivity, and AUC values for three classifiers (RFC, GPC, ABC), highlighting the best-performing models for each band.

Band	Classifier	Accuracy	F1 Score	Precision	Recall	AUC Score
δ Band Task + Assy.	RFC	0.72	0.73	0.7	0.8	0.73
	GPC	0.57	0.61	0.6	0.7	0.6
	ABC	0.72	0.7	0.8	0.7	0.8
θ Band Task + Assy.	RFC	0.86	0.83	0.93	0.8	0.92
	GPC	0.8	0.79	0.83	0.8	0.9
	ABC	0.86	0.83	0.93	0.8	0.85
α Band Task + Assy.	RFC	0.72	0.6	0.7	0.6	0.88
	GPC	0.58	0.51	0.6	0.5	0.77
	ABC	0.81	0.76	0.93	0.7	0.85
β Band Task + Assy.	RFC	0.58	0.57	0.6	0.6	0.62
	GPC	0.58	0.41	0.48	0.5	0.45
	ABC	0.82	0.77	0.9	0.8	0.82
δ Band Rest + Assy.	RFC	0.71	0.59	0.67	0.6	0.73
	GPC	0.57	0.59	0.53	0.7	0.48
	ABC	0.82	0.8	0.9	0.8	0.87
θ Band Rest + Assy.	RFC	0.62	0.6	0.7	0.6	0.72
	GPC	0.63	0.61	0.67	0.7	0.73
	ABC	0.82	0.77	0.9	0.8	0.82
α Band Rest + Assy.	RFC	0.52	0.49	0.53	0.5	0.67
	GPC	0.61	0.57	0.68	0.6	0.72
	ABC	0.82	0.77	0.9	0.7	0.82
β Band Rest + Assy.	RFC	0.68	0.7	0.77	0.7	0.82
	GPC	0.63	0.61	0.77	0.7	0.7
	ABC	0.77	0.73	0.9	0.7	0.85

4.5 Confusion Matrix Analysis

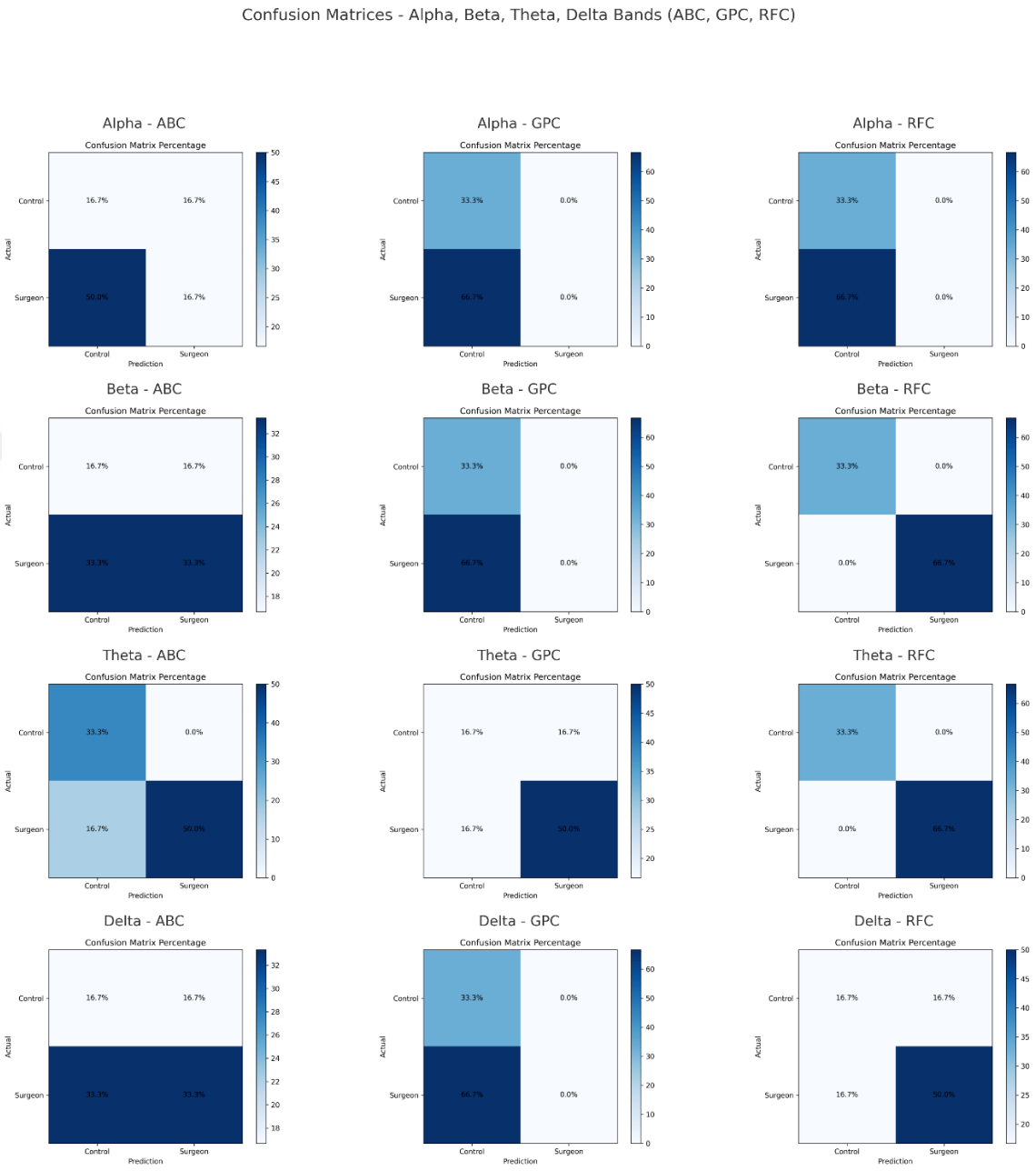


Figure 4.6 Confusion matrices of the classification results using EEG power features and asymmetry indices in different frequency bands. The matrices show the performance of the RFC, GPC, and ABC models in distinguishing surgeon and control participants during the task.

To examine the classification behavior of the models, Figure 4.6 presents the confusion matrices across EEG frequency bands (δ – delta, θ – theta, α – alpha, β – beta) and classifiers (ABC, GPC, RFC). Each matrix provides insight into the dissemination of

correct and incorrect classifications by showing the predicted and actual group memberships (Control vs. Surgeon).

The AdaBoost Classifier (ABC, left column) generally produced the most balanced classification results, effectively discriminating between surgeons and control participants. For example, in the Alpha-ABC model, both groups were identified with equal accuracy (33.3% for each), demonstrating ABC's ability to capture non-linear patterns in the data.

In contrast, the Gaussian Process Classifier (GPC, middle column) exhibited a marked tendency to predict all samples as belonging to a single class, mostly the control group. This resulted in the surgeon group not being recognized at all in several conditions, as observed in the Alpha-GPC and Delta-GPC models. Such behavior indicates overfitting and poor generalization capacity for this classifier.

The Random Forest Classifier (RFC, right column) showed consistently a high recognition rate for the surgeon group, reaching 66.7% in the Theta-RFC condition. However, this success was accompanied by reduced accuracy for the control group, which dropped to 33.3%, suggesting a mild bias toward the surgeon class.

Overall, the confusion matrix analysis highlights the superiority of ABC, followed closely by RFC, in achieving reliable classification performance. On the other hand, GPC failed to generalize effectively across frequency bands. Furthermore, results once again highlight that features extracted during task-related EEG activity—particularly in the theta band—yielded more accurate classification than resting-state features, reinforcing the importance of cognitive and motor involvement in distinguishing levels of surgical expertise.

Table 4.2 Performance metrics for multimodal feature combinations including EEG-based PLV, asymmetry, and sensor-derived force measurements. Results are given in different frequency bands for RFC, GPC, and ABC classifiers, demonstrating the additional contribution of sensor data when integrated with neurophysiological measurements.

Band	Classifier	Accuracy	F1 Score	Precision	Recall	AUC Score
δ Band Task + Assy.	RFC	0.72	0.67	0.9	0.6	0.82
	GPC	0.67	0.6	0.8	0.6	0.68
	ABC	0.82	0.77	0.9	0.7	0.82
θ Band Task + Assy.	RFC	0.77	0.75	0.83	0.8	0.73
	GPC	0.67	0.69	0.67	0.8	0.72
	ABC	0.82	0.77	0.9	0.7	0.82
α Band Task + Assy.	RFC	0.72	0.67	0.8	0.6	0.79
	GPC	0.71	0.71	0.6	0.9	0.85
	ABC	0.71	0.72	0.7	0.8	0.74
β Band Task + Assy.	RFC	0.76	0.75	0.8	0.8	0.83
	GPC	0.58	0.59	0.53	0.7	0.67
	ABC	0.72	0.68	0.8	0.7	0.88
δ Band Rest + Assy.	RFC	0.81	0.81	0.8	0.9	0.93
	GPC	0.86	0.88	0.8	1.0	0.9
	ABC	0.82	0.77	0.9	0.8	0.88
θ Band Rest + Assy.	RFC	0.82	0.79	0.83	0.8	0.85
	GPC	0.67	0.7	0.67	0.9	0.6
	ABC	0.77	0.75	0.77	0.8	0.92
α Band Rest + Assy.	RFC	0.72	0.74	0.7	0.8	0.8
	GPC	0.72	0.73	0.77	0.9	0.87
	ABC	0.67	0.67	0.77	0.7	0.81
β Band Rest + Assy.	RFC	0.72	0.63	0.9	0.5	0.82
	GPC	0.68	0.63	0.68	0.8	0.73
	ABC	0.82	0.77	0.9	0.7	0.82

Table 4.2 provides a detailed comparison of classification performances obtained when PLV-based functional connectivity features were combined with measures of motor asymmetry metrics. Unlike the power spectral analyses reported earlier, this table emphasizes the role of phase synchrony across four frequency bands (δ, θ, α, β) and evaluates these bands separately under task and rest conditions. Classification results obtained by integrating PLV-based connectivity features with motor asymmetry indices demonstrated that resting-state conditions generally provided higher predictive power compared to task-related conditions. Among all frequency bands, the delta rest condition produced the most robust performance. In this condition, RFC achieved 81% accuracy and an AUC of 0.93, while GPC reached even higher accuracy (86%) and an AUC of 0.90. Similarly, theta-band PLV features under rest conditions also yielded

excellent results. As shown in Table 4.2, ABC achieved an AUC of 0.92 demonstrated a balanced performance with 77% accuracy and an F1 score of 0.75. These findings highlight the critical role of theta synchrony in expertise-related neural differentiation.

In contrast, task-based PLV features produced relatively lower classification results, though the beta band still exhibited notable discriminative potential. Specifically, beta features in task conditions enabled RFC to reach an accuracy of 0.76 with an AUC of 0.83, while ABC produced an AUC of 0.88. This suggests that phase-based connectivity metrics retain some predictive value even under active motor execution. However, the overall results suggest that resting-state connectivity patterns are more informative for distinguishing surgeons from non-experienced participants than task-based PLV activity.

Among classifiers, ensemble-based approaches such as RFC and ABC consistently delivered higher and more stable performance than GPC across nearly all conditions. The GPC frequently showed reduced accuracy, poor generalization, and a tendency to collapse predictions into a single class, particularly under task conditions where surgeon participants were often misclassified. These findings reinforce the utility of ensemble classifiers in capturing the nonlinear dynamics of EEG connectivity features and highlight the superior discriminatory power of resting delta and theta PLV activity when combined with measures of motor asymmetry.

Figure 4.7 presents the confusion matrices for all frequency band-classifier combinations obtained by using PLV-based functional connectivity metrics and motor asymmetry features together. Results demonstrate that classifier performance is strongly influenced by both the type of algorithm employed and the frequency band considered.

For the ABC classifiers, performance varied considerably depending on the band. In the Alpha-ABC model, surgeons were correctly identified in 16.7% of cases, while controls also achieved 16.7% recall; however, 50.0% of surgeons were misclassified as controls, indicating limited discriminative power in this band. The Beta-ABC model achieved more balanced performance, with surgeons correctly classified in 33.3% of cases and

controls recognized at 16.7%, reflecting a moderate but more symmetric prediction behavior. In the Theta-ABC model, surgeon detection improved substantially, reaching 50.0% recall, while controls were correctly classified in 33.3% of cases. Finally, the Delta-ABC model showed a similar trend, with surgeons detected at 50.0% recall and controls at 16.7%, suggesting better sensitivity to surgeons but moderate recognition of controls.

In contrast, the GPC classifiers systematically failed to identify surgeons across all frequency bands. In the Alpha-GPC model, controls were correctly classified in 33.3% of cases, while 0.0% of surgeons were recognized. The same pattern persisted for Beta-GPC, Theta-GPC, and Delta-GPC, where all surgeon cases were misclassified as controls, resulting in 0% surgeon recall. This indicates a strong prediction bias toward the control class and highlights GPC's poor suitability for this classification task.

The RFC classifiers demonstrated the most robust and consistent performance, particularly in the Beta and Theta bands. The Beta-RFC model achieved good separation, with controls correctly classified in 33.3% of cases and surgeons in 66.7%, reaching 100% recall for both groups. Similarly, the Theta-RFC model exhibited good sensitivity to surgeons, correctly identifying 66.7% of surgeon cases while achieving 16.7% recall for controls. The Delta-RFC model yielded strong results as well, with surgeons recognized in 50.0% of cases and controls in 16.7%, maintaining a stable, moderately balanced prediction profile. However, the Alpha-RFC model performed poorly, with 0.0% surgeon recall and 33.3% control recall, similar to the GPC classifiers.

Overall, these results show that the combination of PLV-based functional connectivity and asymmetry metrics carries meaningful discriminative value, particularly in the Beta and Theta frequency bands. Among the classifiers, RFC consistently outperforms GPC and frequently surpasses ABC, achieving surgeon recall rates up to 66.7% while maintaining non-negligible control recognition. Conversely, GPC consistently collapses toward the control class, making it unsuitable for EEG-based expertise classification in this context.

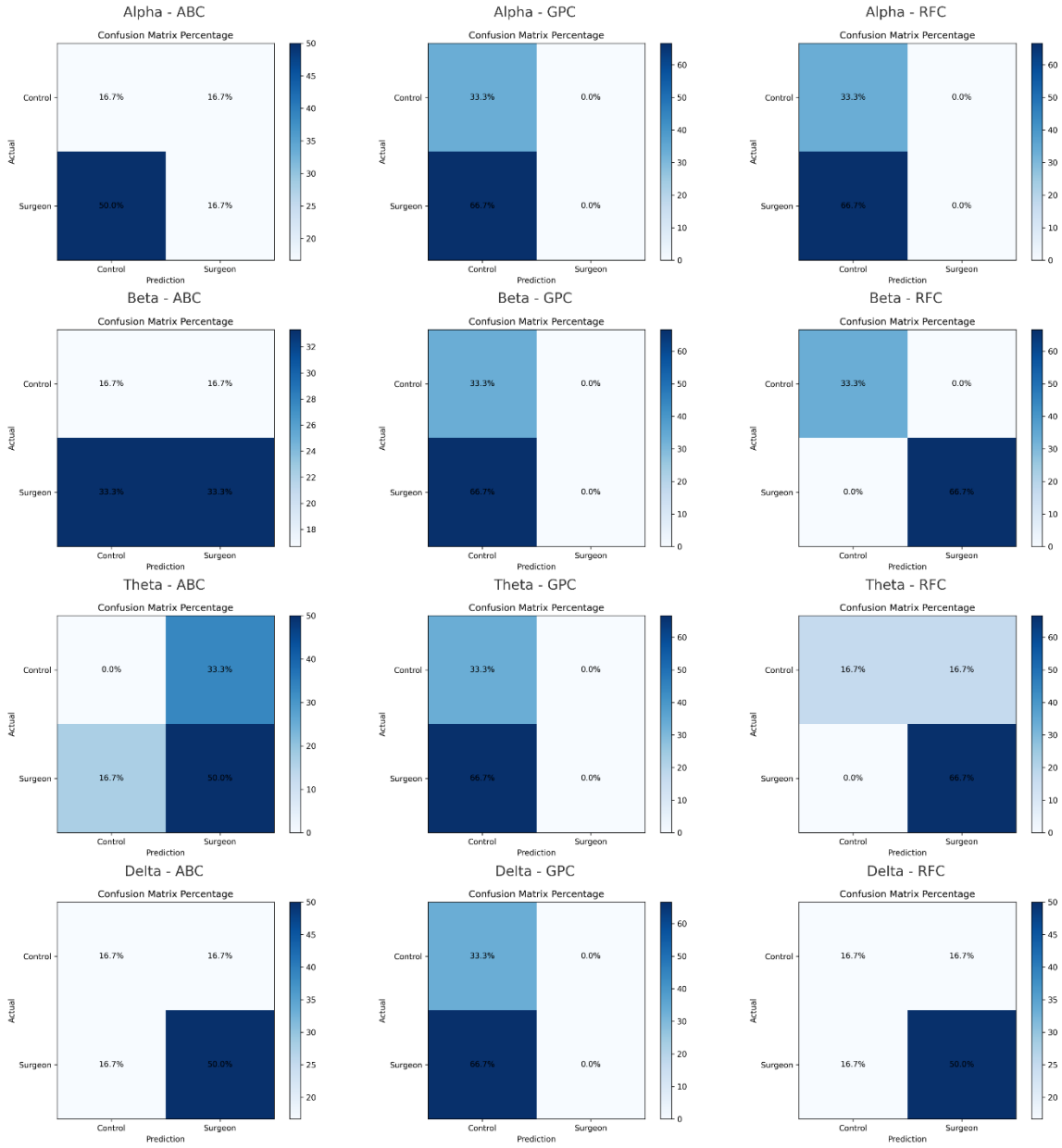


Figure 4.7 Confusion matrices of the classification results using PLV-based functional connectivity features with asymmetry indices in the α, β, θ, and δ bands. The figure shows that ABC classifiers generally produce more balanced predictions, GPC shows class bias, especially in recognizing surgeon class, and RFC provides higher sensitivity for surgeons in the θ and δ bands.

5. DISCUSSION

This study demonstrates that surgical expertise can be objectively distinguished using behavioral, physiological, and neurophysiological measures. As expected, surgeons outperformed novices, particularly in terms of error rate, with the expert group having a significantly lower error rate. While other metrics, such as completion time, HRV indices, and subjective workload, showed trends in favor of surgeons, these differences did not crucial role. This may be due to individual differences and limited sample size. However, the observed trends are consistent with previous studies emphasizing the motor efficiency and cognitive stability associated with surgical expertise (Araki et al., 2017; Zheng et al., 2025).

One of the most novel contributions of this study is the analysis of force symmetry and asymmetry. It was found that surgeons applied more even pressure between both hands, while novices overloaded their dominant hand, creating a significant asymmetry. Metrics such as Right–Left Overlap (RLO) and Right–Left Asymmetry (RLA) emerged as strong indicators of skill level, reflecting both the coordination and distribution of motor effort. These findings suggest that surgical training should emphasize bimanual coordination and that force-based parameters can be used as reliable, objective indicators of technical proficiency.

Between the groups, no difference was found significantly in subjective workload assessments measured with the NASA-TLX in terms of total scores. However, subscale analyses revealed notable differences. Strong correlations were observed between task duration and effort, frustration, and temporal demand scores in novices. This suggests that novices are more sensitive to time pressure and cognitive load. Surgeons, on the other hand, made more consistent workload assessments regardless of task duration, suggesting more efficient use of mental resources and the deployment of automated strategies. This observation is consistent with the concept of "cognitive offloading," which suggests that task demands in experts are managed with less subjective strain through automatic skill execution.

Neurophysiological measurements also supported these behavioral and subjective findings. EEG analyses revealed that theta band activity was particularly decisive in distinguishing groups; this band has been associated with cognitive control, error monitoring, and motor planning processes. While EEG alone offered valuable discriminatory power, its use in conjunction with motor asymmetry features significantly increased classification accuracy. This multimodal integration highlights the complementary nature of cognitive and motor domains, providing a more comprehensive framework for assessing expertise.

Machine learning analyses confirmed the additional value of multimodal data integration. Models trained with EEG features alone or with force metrics alone achieved moderate accuracy, but classification performance improved significantly when the features were combined. In particular, the integration of theta band features with RLA metrics produced the highest discriminatory performance, demonstrating the potential of hybrid models for surgical skill assessment. The findings are consistent with recent studies advocating for multimodal classification approaches to increase the robustness of expertise assessment (Lotte et al., 2018; Wutzl et al., 2024).

Overall, this study demonstrates that experienced surgeons differ from novices not only in terms of error rate and task efficiency, but also in their ability to maintain motor balance and cognitive stability under task demands. Integrating EEG theta-band features with force symmetry/asymmetry metrics improved classification performance and provided a reliable multimodal biomarker framework for distinguishing expertise. These findings directly contribute to the improvement of objective training protocols, real-time monitoring models, and personalized feedback strategies, potentially improving the safety and efficiency of surgical training.

5.1 Limitations

Despite the promising findings, several limitations should be acknowledged. The relatively small sample size reduces the generalizability of the results and may have contributed to the absence of statistical significance in some measures. Including more

participants with varying levels of surgical expertise would improve the robustness and applicability of future research.

The study was also restricted to a single laparoscopic simulation task. While this provided controlled conditions, it may not fully reflect the variability and complexity of real surgical settings. Expanding the range of tasks and contexts would allow for a more comprehensive and realistic assessment of surgical performance.

Another constraint arises from the cross-sectional design, which prevents observation of how motor coordination, cognitive workload, and neurophysiological markers evolve over time. Longitudinal approaches are needed to capture changes across different stages of surgical training and to better understand learning trajectories.

In addition, the analysis focused on a specific combination of EEG, force-based metrics, and subjective assessments. Although this multimodal framework showed promising classification performance, incorporating further physiological measures—such as fNIRS or eye-tracking—could enhance predictive accuracy and yield deeper insights into the interplay between cognitive workload and motor control.

6. CONCLUSION

This thesis investigated the integration of EEG spectral asymmetry features with force-based asymmetry indices to assess surgical expertise in a laparoscopic simulation task. Findings demonstrated that the combined use of neurophysiological and behavioral measures provides a powerful framework for distinguishing surgeons from inexperienced participants. Task-related EEG asymmetry, particularly in the theta and alpha bands, emerged as a sensitive biomarker of expertise, reflecting the cognitive control and attentional mechanisms engaged during surgical performance. These neural indicators, when combined with motor asymmetry metrics derived from instrument pressure data, yielded higher classification accuracy compared to single approaches.

Among the tested classifiers, the AdaBoost Classifier (ABC) performed higher and more consistently than Random Forest and Gaussian Process models. This result highlights the effectiveness of ensemble learning methods in capturing the multidimensional nature of surgical skills. The integration of EEG-based asymmetry with motor coordination measures demonstrates not only the feasibility but also the additional value of multimodal approaches for objective surgical assessment.

This thesis demonstrates that spectral asymmetry derived from EEG signals can serve as a robust neuroergonomic marker of surgical expertise. By integrating cognitive (EEG) and motor (force asymmetry) dimensions, the study provides an objective and multimodal framework for distinguishing expert surgeons from novices. The findings highlight the added value of multimodal features in improving classification performance and emphasize their potential use in training evaluation, certification processes, and clinical ergonomics.

REFERENCES

Adams, B. J., Margaron, F., & Kaplan, B. J. (2012). Comparing video games and laparoscopic simulators in the development of laparoscopic skills in surgical residents. Journal of surgical education, 69(6), 714-717.

Aghazadeh, F., & Zheng, B. (2025). A new approach to laparoscopic skill assessment: Motion smoothness and bimanual coordination. Laparoscopic, Endoscopic and Robotic Surgery.

Ahmmad, S. N. Z., Ming, E. S. L., Fai, Y. C., & Narayanan, A. L. T. (2014). Experimental study of surgeon's psychomotor skill using sensor-based measurement. Procedia Computer Science, 42, 130-137.

Ahmmad, S. N. Z., San, C. Z., Ming, E. S. L., & Fai, Y. C. (2015). Force variability as an objective measure of surgical skill. Jurnal Teknologi (Sciences & Engineering), 74(6).

Alan Gevins, Michael E. Smith, Harrison Leong, Linda McEvoy, Susan Whitfield, Robert Du, & Georgia Rush. (1998). Monitoring Working Memory Load during Computer-Based Tasks with EEG Pattern Recognition Methods.

Almassri, Ahmed M., Mostafa B. Abuitbel, W. Z. WanHasan, S. A. Ahmad, and Ahmad H. Sabry. "Real-time control for robotic hand application based on pressure sensor measurement." In 2014 IEEE International Symposium on Robotics and Manufacturing Automation (ROMA), pp. 80-85. IEEE, 2014.

Almukhtar, A., Caddick, V., Naik, R., Goble, M., Mylonas, G., Darzi, A., ... & Leff, D. R. (2025). Objective assessment of cognitive workload in surgery: a systematic review. Annals of surgery, 281(6), 942-951.

Anton, N. E., Howley, L. D., Pimentel, M., Davis, C. K., Brown, C., & Stefanidis, D. (2016). Effectiveness of a mental skills curriculum to reduce novices' stress. Journal of surgical research, 206(1), 199-205.

Antonenko, P., Paas, F., Grabner, R., & Van Gog, T. (2010). Using electroencephalography to measure cognitive load. Educational psychology review, 22(4), 425-438.

Araki, A., Makiyama, K., Yamanaka, H., Ueno, D., Osaka, K., Nagasaka, M., ... & Yao, M. (2017). Comparison of the performance of experienced and novice surgeons: measurement of gripping force during laparoscopic surgery performed on pigs using forceps with pressure sensors. Surgical endoscopy, 31(4), 1999-2005.

Ayaz, H., & Dehais, F. (Eds.). (2018). Neuroergonomics: The brain at work and in everyday life. Academic Press. Paperback ISBN: 9780128119266

Aydore, S., Pantazis, D., & Leahy, R. M. (2013). A note on the phase locking value and its properties. Neuroimage, 74, 231-244.

Balkhoyor, A. M., Awais, M., Biyani, S., Schaefer, A., Craddock, M., Jones, O., ... & Mushtaq, F. (2020). Frontal theta brain activity varies as a function of surgical experience and task error. BMJ Surgery, Interventions, & Health Technologies, 2(1), e000040.

Beattie, K.L., Hill, A., Horswill, M.S. et al. Aptitude and attitude: predictors of performance during and after basic laparoscopic skills training. Surg Endosc 36, 3467–3479 (2022).

Bertrand, J. A., McIntosh, A. R., Postuma, R. B., Kovacevic, N., Latreille, V., Panisset, M., ... & Gagnon, J. F. (2016). Brain connectivity alterations are associated with the development of dementia in Parkinson's disease. Brain connectivity, 6(3), 216-224.

Bez, M., Maimon, N. B., Ddobot, D., Molcho, L., Intrator, N., Kakiashvilli, E., & Bickel, A. Continuous monitoring of mental load during virtual simulator training for laparoscopic surgery, reflects laparoscopic dexterity. A comparative study using a novel wireless device.

Carswell, C. M., Clarke, D., & Seales, W. B. (2005). Assessing mental workload during laparoscopic surgery. Surgical innovation, 12(1), 80-90.

Castillo-Segura, P., Fernandez-Panadero, C., Alario-Hoyos, C., Munoz-Merino, P. J., & Kloos, C. D. (2021). Objective and automated assessment of surgical technical skills with IoT systems: A systematic literature review. Artificial Intelligence in Medicine, 112, 102007.

Chacko, R. V., Kim, B., Jung, S. W., Daitch, A. L., Roland, J. L., Metcalf, N. V., ... & Leuthardt, E. C. (2018). Distinct phase-amplitude couplings distinguish cognitive processes in human attention. NeuroImage, 175, 111-121.

Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. Applied ergonomics, 74, 221-232.

Cleveland Clinic. (2024). Laparoscopy. Retrieved August 25, 2025, from https://my.clevelandclinic.org/health/procedures/4819-laparoscopy

Crewther, B. T., Shetty, K., Jarchi, D., Selvadurai, S., Cook, C. J., Leff, D. R., Darzi, A., & Yang, G. Z. (2016). Skill acquisition and stress adaptations following laparoscopic surgery training and detraining in novice surgeons. Surgical Endoscopy, 30(7), 2961–2968.

Cundy, T. P., Thangaraj, E., Rafii-Tari, H., Payne, C. J., Azzie, G., Sodergren, M. H., ... & Darzi, A. (2015). Force-Sensing Enhanced Simulation Environment (ForSense) for laparoscopic surgery training and assessment. Surgery, 157(4), 723-731.

Dayan, E., & Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. Neuron, 72(3), 443–454.

de Freitas, P. B., Krishnan, V., & Jaric, S. (2008). Force coordination in object manipulation. Journal of Human Kinetics, 20(1), 37-50. DOI 10.2478/v10078-008-0016-8

de Mathelin, M., Nageotte, F., Zanne, P., & Dresp-Langley, B. (2019). Sensors for expert grip force profiling: Towards benchmarking manual control of a robotic device for surgical tool movements. Sensors, 19(20), 4575.

Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A neuroergonomics approach to mental workload, engagement and human performance. Frontiers in neuroscience, 14, 268.

Diarra, M., Theurel, J., & Paty, B. (2025). Systematic review of neurophysiological assessment techniques and metrics for mental workload evaluation in real-world settings. Frontiers in Neuroergonomics, 6, 1584736.

Dias, R. D., Ngo-Howard, M. C., Boskovski, M. T., Zenati, M. A., & Yule, S. J. (2018). Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *Journal of British Surgery*, *105*(5), 491-501.

DiDomenico, A., & Nussbaum, M. A. (2008). Interactive effects of physical and mental workload on subjective workload assessment. International journal of industrial ergonomics, 38(11-12), 977-983.

Drewniak, Elizabeth I., Joseph J. Crisco, David B. Spenciner, and Braden C. Fleming. "Accuracy of circular contact area measurements with thin-film pressure sensors." Journal of biomechanics 40, no. 11 (2007): 2569-2572.

Dvorak, D., & Fenton, A. A. (2014). Toward a proper estimation of phase–amplitude coupling in neural oscillations. Journal of Neuroscience methods, 225, 42-56.

Ershad, M., Rege, R., & Fey, A. M. (2018). Meaningful assessment of robotic surgical style using the wisdom of crowds. International journal of computer assisted radiology and surgery, 13(7), 1037-1048.

Farah, E., Desir, A., Marques, C., Hegde, S. R., Abreu, A., Polanco, P. M., ... & Sankaranarayanan, G. (2024). Heart rate variability: an objective measure of mental stress in surgical simulation. Global Surgical Education-Journal of the Association for Surgical Education, 3(1), 25.

Fard, M. J., Ameri, S., Chinnam, R. B., Pandya, A. K., Klein, M. D., & Ellis, R. D. (2016). Machine learning approach for skill evaluation in robotic-assisted surgery. arXiv preprint arXiv:1611.05136.

Gao, J., Liu, S., Feng, Q., Zhang, X., Jiang, M., Wang, L., ... & Zhang, Q. (2019). Subjective and objective quantification of the effect of distraction on physician's workload and performance during simulated laparoscopic surgery. Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 25, 3127.

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

Gao, Y., Yan, P., Kruger, U., Cavuoto, L., Schwaitzberg, S., De, S., & Intes, X. (2020). Functional brain imaging reliably predicts bimanual motor skill performance in a standardized surgical task. IEEE Transactions on Biomedical Engineering, 68(7), 2058-2066.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science, 4(1–2), 113–131.

Goble, M., Caddick, V., Patel, R., Modi, H., Darzi, A., Orihuela-Espina, F., & Leff, D. R. (2023). Optical neuroimaging and neurostimulation in surgical training and assessment: A state-of-the-art review. Frontiers in Neuroergonomics, 4, 1142182.

Gomez, E. D., Aggarwal, R., McMahan, W., Bark, K., & Kuchenbecker, K. J. (2016). Objective assessment of robotic surgical skill using instrument contact vibrations. Surgical endoscopy, 30(4), 1419-1431.

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2003). EMG contamination of EEG: spectral and topographical characteristics. Clinical neurophysiology, 114(9), 1580-1593.

Guzmán-García, C., Sánchez-González, P., Margallo, J. A. S., Snoriguzzi, N., Rabazo, J. C., Margallo, F. M. S., ... & Oropesa, I. (2022). Correlating personal resourcefulness and psychomotor skills: an analysis of stress, visual attention and technical metrics. Sensors, 22(3), 837.

Hanna, G. B., Drew, T., Arnold, G., Fakhry, M., & Cuschieri, A. (2008). Development of force measurement system for clinical use in minimal access surgery. Surgical endoscopy, 22(2), 467-471.

Harada, K., Morita, A., Minakawa, Y., Baek, Y. M., Sora, S., Sugita, N., ... & Mitsuishi, M. (2015). Assessing microneurosurgical skill with medico-engineering technology. World neurosurgery, 84(4), 964-971.

Harris, D. J., Vine, S. J., Wilson, M. R., McGrath, J. S., LeBel, M. E., & Buckingham, G. (2017). The effect of observing novice and expert performance on acquisition of surgical skills on a robotic platform. PLoS One, 12(11), e0188233.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society, 904–908.

Huang, Y. et al. (2025). Bimanual visuomotor training enhances interhemispheric connectivity and coordination precision. Scientific Reports.

Hubert, N., Gilles, M., Desbrosses, K., Meyer, J., Felblinger, J., & Hubert, J. (2013). Ergonomic assessment of the surgeon's physical workload during standard and

robotic assisted laparoscopic procedures. International Journal of Medical Robotics and Computer Assisted Surgery, 9(2), 142–147.

Jaric, S., & Uygur, M. (2013). Assessment of hand function through the coordination of contact forces in manipulation tasks. Journal of Human Kinetics, 36, 5.

Jian, W., Chen, M., & McFarland, D. J. (2017). Use of phase-locking value in sensorimotor rhythm-based brain–computer interface: zero-phase coupling and effects of spatial filters. Medical & biological engineering & computing, 55(11), 1915-1926.

Kamat, A., Eastmond, C., Gao, Y., Nemani, A., Yanik, E., Cavuoto, L., ... & Intes, X. (2023). Assessment of Surgical Tasks Using Neuroimaging Dataset (ASTaUND). Scientific Data, 10(1), 699.

Keleş, H. O. (2022). Cerrahların Zihinsel İş Yükü Değerlendirmesinde İşlevsel Beyin Görüntüleme Uygulamaları.

Keles, H. O., & Omurtag, A. (2021). Video game experience affects performance, cognitive load, and brain activity in laparoscopic surgery training. Research Square.

Keles, H. O., Barbour, R. L., & Omurtag, A. (2016). Hemodynamic correlates of spontaneous neural activity measured by human whole-head resting state EEG + fNIRS. NeuroImage, 138, 76–87.

Keles, H. O., Cengiz, C., Demiral, I., Ozmen, M. M., & Omurtag, A. (2021). High density optical neuroimaging predicts surgeons's subjective experience and skill levels. PLoS ONE, 16(2 February).

Keles, H., Cengiz, C., Demiral, I., Ozmen, M., & Omurtag, A. (2021). 155 A Machine Learning- Based Approach for Predicting Surgeons' Subjective Experience and Skill Levels: Neuroimaging Study. British Journal of Surgery, 108.

Krakauer, J. W., & Mazzoni, P. (2011). Human sensorimotor learning: adaptation, skill, and beyond. Current Opinion in Neurobiology, 21(4), 636–644.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring Phase Synchrony in Brain Signals. In Hum Brain Mapping (Vol. 8).

Lam, K., Chen, J., Wang, Z., Iqbal, F. M., Darzi, A., Lo, B., ... & Kinross, J. M. (2022). Machine learning for technical skill assessment in surgery: a systematic review. NPJ digital medicine, 5(1), 24.

Latash, M. L. (2012). Fundamentals of motor control. Academic Press.

Lee, D., Yu, H. W., Kwon, H., Kong, H. J., Lee, K. E., & Kim, H. C. (2020). Evaluation of surgical skills during robotic surgery by deep learning-based multiple surgical instrument tracking in training and actual operations. Journal of clinical medicine, 9(6), 1964.

Leff, D. R. (2009). An assessment of cortical function in expert and novice surgeons (Doctoral dissertation, Imperial College London (University of London)).

Longo, L. (2017, September). Subjective usability, mental workload assessments and their impact on objective human performance. In IFIP Conference on Human-Computer Interaction (pp. 202-223). Cham: Springer International Publishing.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. Journal of neural engineering, 15(3), 031005.

Manabe, T., Rahul, F. N. U., Fu, Y., Intes, X., Schwaitzberg, S. D., De, S., ... & Dutta, A. (2023). Distinguishing laparoscopic surgery experts from novices using EEG topographic features. Brain Sciences, 13(12), 1706.

Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. Biological psychology, 121, 62-73.

Marinescu, A. C., Sharples, S., Ritchie, A. C., Sanchez Lopez, T., McDowell, M., & Morvan, H. P. (2018). Physiological parameter response to variation of mental workload. Human factors, 60(1), 31-56.

Martin, J. A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C., & Brown, M. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *British journal of surgery*, *84*(2), 273-278.

Mentis, H. M., Chellali, A., Manser, K., Cao, C. G. L., & Schwaitzberg, S. D. (2016). A systematic review of the effect of distraction on surgeon performance: directions for operating room policy and surgical training. Surgical Endoscopy, 30(5), 1713–1724.

Miller, G. A. (1956). The Psychological Review The Magical Number Seven, Plus Or Minus Two: Some Limits On Our Capacity For Processing Information 1.

Miyake, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. International journal of psychophysiology, 40(3), 233-238.

Modi, H. N., Singh, H., Yang, G. Z., Darzi, A., & Leff, D. R. (2017). A decade of imaging surgeons' brain function (part I): terminology, techniques, and clinical translation. Surgery, 162(5), 1121-1130.

Natheir, S. (2021). Utilizing Machine Learning and Electroencephalography to Assess Expertise in Virtual Neurosurgical Performance (Master's thesis, McGill University (Canada)).

Nemani, A., Yücel, M. A., Kruger, U., Gee, D. W., Cooper, C., Schwaitzberg, S. D., ... & Intes, X. (2018). Assessing bimanual motor skills with optical neuroimaging. Science advances, 4(10), eaat3807.

Omurtag, A., Roy, R. N., Dehais, F., Chatty, L., & Garbey, M. (2019). Tracking team mental workload by multimodal measurements in the operating room.

Omurtag, A., Sunderland, C., Mansfield, N. J., & Zakeri, Z. (2024). Neural Mechanisms of Early Motor Learning in Laparoscopic Surgery: EEG Connectivity, BDNF, and Cognitive Load.

Omurtag, A., Sunderland, C., Mansfield, N. J., & Zakeri, Z. (2025). EEG connectivity and BDNF correlates of fast motor learning in laparoscopic surgery. Scientific reports, 15(1), 7399.

Ozcan, A., Tulum, Z., PInar, L., & Başkurt, F. (2004). Comparison of pressure pain threshold, grip strength, dexterity and touch pressure of dominant and non-dominant hands within and between right-and left-handed subjects. Journal of Korean medical science, 19(6), 874-878.

Pagnotta, M., Jacobs, D. M., de Frutos, P. L., Rodríguez, R., Ibáñez-Gijón, J., & Travieso, D. (2022). Task difficulty and physiological measures of mental workload in air traffic control: a scoping review. Ergonomics, 65(8), 1095-1118.

Park, C. H. (2022). Comparison of grip force, coordination, and dexterity between dominant and non-dominant hand according to gender. PNF and Movement, 20(1), 73-81. https://doi.org/10.21598/JKPNFA.2022.20.1.73

Paxion, J., Galy, E., & Berthelon, C. (2014). Mental workload and driving. Frontiers in psychology, 5, 1344.

Pedrett, R., Mascagni, P., Beldi, G., Padoy, N., & Lavanchy, J. L. (2023). Technical skill assessment in minimally invasive surgery using artificial intelligence: a systematic review. Surgical endoscopy, 37(10), 7412-7424

Peterson, L. R., & Peterson, M. J. (1959). Short-Term Retention Of Individual Verbal Items 1. In Journal of Experimental Psychology (Vol. 58, Issue 3).

Rafii-Tari, H., Payne, C. J., Bicknell, C., Kwok, K. W., Cheshire, N. J., Riga, C., & Yang, G. Z. (2017). Objective assessment of endovascular navigation skills with force sensing. Annals of biomedical engineering, 45(5), 1315-1327.

Rafii-Tari, H., Payne, C. J., Liu, J., Riga, C., Bicknell, C., & Yang, G. Z. (2015, May). Towards automated surgical skill evaluation of endovascular catheterization tasks based on force and motion signatures. In 2015 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1789-1794). IEEE.

Rajakumar, Banuvathy, and S. K. M. Varadhan. "Evidence to support the mechanical advantage hypothesis of grasping at low force levels." Scientific Reports 12, no. 1 (2022): 20834.

Reyes, S., Quiñones, M., Kreither, J., López-Calderón, J., & Aylwin, M. D. L. L. (2021). Modulation of Visual Alpha Oscillations in Expert and Novice Surgeons Performing Sutures.

Ridderikhoff, A. (2008). Attentional loads associated with interlimb interactions. Experimental Brain Research

Roelfsema, P. R., Engel, A. K., König, P., & Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. Nature, 385(6612), 157-161.

Roman-Liu, D. et al. (2020). Effectiveness of bimanual coordination tasks on coordination quality, musculoskeletal load, and cognitive functions. PLoS One

Rosen, J., Hannaford, B., Richards, C. G., & Sinanan, M. N. (2001). Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE transactions on Biomedical Engineering, 48(5), 579-591.

Rosser Jr, J. C., Rosser, L. E., & Savalgi, R. S. (1998). Objective evaluation of a laparoscopic surgical skill program for residents and senior surgeons. Archives of surgery, 133(6), 657-661.

Sánchez-Reolid, D., García-Pérez, E., Borja, A. L., Fernández-Caballero, A., & Sánchez-Reolid, R. (2024). EEG and fNIRS Signal-Based Emotion Identification by Means of Machine Learning Algorithms During Visual Stimuli Exposure. Electronics, 13(23), 4797.

Sankaranarayanan, G., Odlozil, C. A., Wells, K. O., Leeds, S. G., Chauhan, S., Fleshman, J. W., ... & De, S. (2020). Training with cognitive load improves performance under similar conditions in a real surgical task. The American Journal of Surgery, 220(3), 620-629.

Scerbo, M. W., Britt, R. C., & Stefanidis, D. (2017). Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task. The American Journal of Surgery, 213(2), 244-248.

Schmidt, B. T., Ghuman, A. S., & Huppert, T. J. (2014). Whole brain functional connectivity using phase locking measures of resting state magnetoencephalography. Frontiers in neuroscience, 8, 141.

Shafiei, S. B., Shadpour, S., Mohler, J. L., Kauffman, E. C., Holden, M., & Gutierrez, C. (2024). Classification of subtask types and skill levels in robot-assisted surgery using EEG, eye-tracking, and machine learning. Surgical Endoscopy, 38(9), 5137-5147

Shafiei, S. B., Shadpour, S., Mohler, J. L., Sasangohar, F., Gutierrez, C., Seilanian Toussi, M., & Shafqat, A. (2023). Surgical skill level classification model development using EEG and eye-gaze data and machine learning algorithms. Journal of robotic surgery, 17(6), 2963-2971.

Shahbazi, M., Poursartip, B., Siroen, K., Schlachta, C. M., & Patel, R. V. (2018, July). Robotics-assisted surgical skills evaluation based on electrocortical activity. In

2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 3673-3676). IEEE.

Shrivastava, S. Physiological correlates of cognitive load in laparoscopic surgery.

Shumway-Cook, A., & Woollacott, M. (2017). Motor Control: Translating Research into Clinical Practice (5th ed.). Lippincott Williams & Wilkins.

Soangra, R., Sivakumar, R., Anirudh, E. R., Reddy Y, S. V., & John, E. B. (2022). Evaluation of surgical skill using machine learning with optimal wearable sensor locations. PLoS One, 17(6), e0267936.

Soleymani, A., Tavakoli, M., Aghazadeh, F., Ou, Y., Rouhani, H., Zheng, B., & Li, X. (2024). Hands collaboration evaluation for surgical skills assessment: An information theoretical approach. IEEE Transactions on Medical Robotics and Bionics.

Spaun, G. O., Zheng, B., Martinec, D. V., Arnold, B. N., & Swanström, L. L. (2010). A comparison of early learning curves for complex bimanual coordination with open, laparoscopic, and flexible endoscopic instrumentation. Surgical endoscopy, 24(9), 2145-2155.

Süer, M. S., Demir, S., Çanakçı, M. H., Değirmencioğlu, G., & Akıncı, M. (2025). Training Laparoscopic Surgeons: Assessing Workload and Skill Using Nasa-Tlx. Journal of Surgical Education, 82(9), 103588.

Sugiyama, T., Lama, S., Gan, L. S., Maddahi, Y., Zareinia, K., & Sutherland, G. R. (2018). Forces of tool-tissue interaction to assess surgical skill level. JAMA surgery, 153(3), 234-242.

Takayasu, K., Yoshida, K., Kinoshita, H., Yoshimoto, S., Oshiro, O., & Matsuda, T. (2018). Analysis of the tractive force pattern on a knot by force measurement during laparoscopic knot tying. The American Journal of Surgery, 216(2), 314-318.

Takazawa, S., Ishimaru, T., Harada, K., Deie, K., Fujishiro, J., Sugita, N., ... & Iwanaka, T. (2016). Pediatric thoracoscopic surgical simulation using a rapid-prototyped chest model and motion sensors can better identify skilled surgeons than a conventional box trainer. Journal of Laparoendoscopic & Advanced Surgical Techniques, 26(9), 740-747.

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A systematic review of physiological measures of mental workload. International journal of environmental research and public health, 16(15), 2716.

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G. Z., & Darzi, A. (2014). Eye tracking for skills assessment and training: a systematic review. journal of surgical research, 191(1), 169-178.

Van den Dobbelsteen, John J., Ruben A. Lee, Maarten van Noorden, and Jenny Dankelman. "Indirect measurement of pinch and pull forces at the shaft of laparoscopic graspers." Medical & biological engineering & computing 50 (2012): 215-221.

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. NeuroImage, 145, 166-179.

Vassiliou, M. C., Feldman, L. S., Andrew, C. G., Bergman, S., Leffondré, K., Stanbridge, D., & Fried, G. M. (2005). A global assessment tool for evaluation of intraoperative laparoscopic skills. *The American journal of surgery*, *190*(1), 107-113.

Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., & Pennartz, C. M. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage, 55(4), 1548-1565.

Walia, P., Fu, Y., Norfleet, J., Schwaitzberg, S. D., Intes, X., De, S., ... & Dutta, A. (2022). Error-related brain state analysis using electroencephalography in conjunction with functional near-infrared spectroscopy during a complex surgical motor task. Brain Informatics, 9(1), 29.

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

Wilson, M. R., McGrath, J. S., Vine, S. J., Brewer, J., Defriend, D., & Masters, R. S. (2011). Perceptual impairment and psychomotor control in virtual laparoscopic surgery. Surgical endoscopy, 25(7), 2268-2274. https://doi.org/10.1007/s00464-010-1546-4

Wilson, M. R., Poolton, J. M., Malhotra, N., Ngo, K., Bright, E., & Masters, R. S. (2011). Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). *World journal of surgery*, *35*(9), 1961-1969.

Winkler-Schwartz, A., Yilmaz, R., Mirchi, N., Bissonnette, V., Ledwos, N., Siyar, S., ... & Del Maestro, R. (2019). Machine learning identification of surgical and operative factors associated with surgical expertise in virtual reality simulation. JAMA network open, 2(8), e198363-e198363.

Wu, Y., Zhu, Y., & Zheng, B. (2025). Enhancing surgical training through cognitive load assessment. Laparoscopic, Endoscopic and Robotic Surgery.

Wutzl, B., Leibnitz, K., & Murata, M. (2024). An Analysis of the Correlation between the Asymmetry of Different EEG-Sensor Locations in Diverse Frequency Bands and Short-Term Subjective Well-Being Changes. Brain Sciences, 14(3), 267.

Yeganeh Doost, M. Y., Afsarlar, M., & Swinnen, S. P. (2017). Two Processes in Early Bimanual Motor Skill Learning. Frontiers in Human Neuroscience.

Yu, D., Abdelrahman, A. M., Buckarma, E. H., Lowndes, B. R., Gas, B. L., Finnesgard, E. J., ... & Hallbeck, S. (2015, September). Mental and physical workloads in a competitive laparoscopic skills training environment: a pilot study. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 59, No. 1, pp. 508-512). Sage CA: Los Angeles, CA: SAGE Publications.

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. Simulation in healthcare, 5(5), 267-271.

Zakeri, Z., Mansfield, N., Sunderland, C., & Omurtag, A. (2020). Physiological correlates of cognitive load in laparoscopic surgery. Scientific reports, 10(1), 12927.

Zander, T. O., Shetty, K., Lorenz, R., Leff, D. R., Krol, L. R., Darzi, A. W., ... & Yang, G. Z. (2017). Automated task load detection with electroencephalography: towards passive brain–computer interfacing in robotic surgery. Journal of Medical Robotics Research, 2(01), 1750003.

Zhang, J. Y., Liu, S. L., Feng, Q. M., Gao, J. Q., & Zhang, Q. (2017). Correlative evaluation of mental and physical workload of laparoscopic surgeons based on surface electromyography and eye-tracking signals. Scientific reports, 7(1), 11095.

Zheng, B., Cassera, M. A., Martinec, D. V., Spaun, G. O., & Swanström, L. L. (2010). Measuring mental workload during the performance of advanced laparoscopic tasks. Surgical endoscopy, 24(1), 45-50.

Zhou, T., Cha, J. S., Gonzalez, G. T., Wachs, J. P., Sundaram, C., & Yu, D. (2018, March). Joint surgeon attributes estimation in robot-assisted surgery. In Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 285-286).

Zia, A., Sharma, Y., Bettadapura, V., Sarin, E. L., & Essa, I. (2018). Video and accelerometer-based motion analysis for automated surgical skills assessment. International journal of computer assisted radiology and surgery, 13(3), 443-455.

Zia, A., Sharma, Y., Bettadapura, V., Sarin, E. L., Clements, M. A., & Essa, I. (2015, October). Automated assessment of surgical skills using frequency analysis. In International conference on medical image computing and computer-assisted intervention (pp. 430-438). Cham: Springer International Publishing.