CONTROLLING COGNITIVE DEMANDS WITH SEMI-AUTONOMOUS SUCTION FRAMEWORK FOR ROBOTIC-ASSISTED SURGERY

by

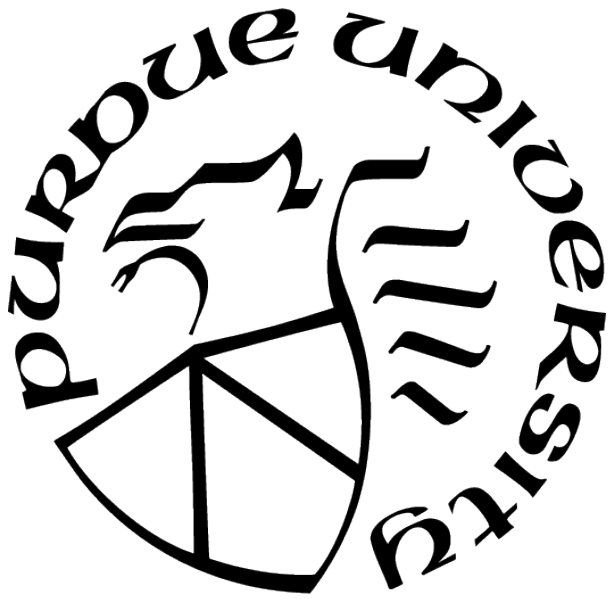
Juan Antonio Barragan

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Industrial Engineering



School of Industrial Engineering

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Juan P. Wachs

School of Industrial Engineering

Dr. Denny Yu

School of Industrial Engineering

Dr. Neera Jain

School of Mechanical engineering

Approved by:

Dr. Abhijit Deshmukh

I dedicate this thesis to my family, who have taught me to dream big and to never be stopped by today's limitations.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the amazing mentors and friends I had the fortune to meet. First and foremost, I want to express my gratitude to my advisor, Dr. Juan Pablo Wachs, for his mentorship and guidance during my master's program. I also want to thank my committee members, Dr. Denny Yu and Dr. Neera Jain, for their insightful comments to make my work the best possible.

To my dear friends and labmates from the ISAT laboratory, I want to let them know that I cannot imagine a more welcoming and collaborative place to work. I feel very fortunate for every dinner and tres leches cake that we had together. I especially want to thank Edgar, Glebys, and Naveen for being such outstanding mentors to me. Also, I want to thank Daniela Chanci for being such a fantastic friend and for our weekly and sometimes daily check-up calls.

I also want to thank my mom, dad, and brother for supporting and encouraging me to go above my limitations. This accomplishment would not have been possible without having you at every step of the way. To my roommate Catalina Vega, thanks for becoming such an incredible friend and my family here at Purdue. I will miss our late-night talks while having a cup of tea. To the rest of my family and friends, thanks. I cannot overstate how fortunate I have been to have in my life so many incredible people.

TABLE OF CONTENTS

LIST OF TABLES 8

LIST OF FIGURES 9

ABBREVIATIONS 11

ABSTRACT 12

1 INTRODUCTION 14

1.1 Problem Statement 16

1.1.1 Research Questions 17

1.2 Contributions 17

1.3 Summary 18

2 LITERATURE REVIEW 19

2.1 Surgical Robotics 19

2.2 Surgical Robotics Research Platforms 22

2.3 Computer Vision 24

2.3.1 Segmentation and Registration of Patient Anatomy and Instruments 24

2.3.2 Hemorrhage Detection and Tracking in Surgery 25

2.4 Autonomy in Surgical Robotics 26

2.5 Human-Robot Interaction in Surgery 29

2.6 Surgical Performance Metrics 30

2.7 Workload Assessment Techniques 31

2.8 Discussion 33

2.9 Summary 33

3 METHODOLOGY 35

3.1 Da Vinci Research Kit Description 35

3.2 Simulator Design 36

3.3 Autonomous Suction Robotic Assistant 37

3.3.1 Computer Vision Module . . . 38
3.3.1.1 Notation . . . 38
3.3.1.2 Manipulator Registration to the Camera Frame . . . 39
3.3.1.3 Semantic Segmentation of Endoscopic Images . . . 40
3.3.2 Robotic module . . . 44
3.3.3 Augmented Reality Visualization . . . 46
3.4 Cognitive Workload Sensing Module . . . 48
3.4.1 Physiological Sensors . . . 49
3.4.1.1 EEG Sensing . . . 49
3.4.1.2 Eye Tracker Based Sensing . . . 50
3.4.2 Feature Extraction . . . 50
3.4.2.1 EEG Frequency Features . . . 51
3.4.2.2 Eye Tracker Features . . . 52
3.4.2.3 Feature Synchronization . . . 53
3.4.3 Cognitive Workload Classification Models . . . 53
3.4.3.1 Recurrent Architecture for EEG Classification . . . 55
3.4.3.2 Feed Forward Model for Multi Sensor Classification . . . 57
3.4.4 Real-time Architecture and Deployment . . . 57
3.5 Summary . . . 58

4 EXPERIMENTS AND RESULTS . . . 59
4.1 Experiment 1: Evaluating Autonomous Suction Action on Users' Performance 59
4.1.1 User Study Protocol . . . 60
4.1.2 Evaluation Metrics and Statistical Analysis . . . 61
4.1.2.1 Objective Evaluation Metrics . . . 62
4.1.2.2 Subjective Evaluation Metrics . . . 62
4.1.3 Results . . . 63
4.1.4 Discussions . . . 64
4.2 Experiment 2: Evaluation of the Cognitive Detection Framework . . . 65
4.2.1 EEG Dataset Collection Protocols . . . 66

4.2.2 Evaluation of EEG models . . . 68
4.2.2.1 Session-to-Session Model Evaluation . . . 68
4.2.2.2 User-to-User Model Evaluation . . . 69
4.2.3 EEG Model Results . . . 69
4.2.3.1 Session-to-Session Results . . . 69
4.2.3.2 User-to-User Results . . . 70
4.2.3.3 Physiological Analysis of EEG Features . . . 71
4.2.4 Multi Sensors Dataset Collection Protocols . . . 73
4.2.5 Multi Sensors Model Evaluation . . . 74
4.2.5.1 Single User Model Evaluation . . . 74
4.2.5.2 Multi User Model Evaluation . . . 75
4.2.6 Multi Sensors Models Results . . . 75
4.2.7 Discussions . . . 75
4.2.7.1 EEG Models . . . 75
4.2.7.2 Multi Sensors Models . . . 77
4.3 Experiment 3: Evaluation of Cognitive-Triggered Autonomous Suction Action 78
4.3.1 User Study Protocol . . . 78
4.3.2 Evaluation Metrics . . . 79
4.3.3 Results . . . 79
4.3.4 Discussions . . . 82
4.4 Summary . . . 83

5 CONCLUSIONS AND FUTURE WORK . . . 84
5.1 Limitations and Future Work . . . 87
5.2 Research Questions . . . 89
5.3 Summary . . . 90

REFERENCES . . . 91

VITA . . . 109

PUBLICATION(S) . . . 110

LIST OF TABLES

3.1 Semantic segmentation network downsampling specifications. 42

4.1 Objective performance metrics results. Statistically significant results (α=0.05) were highlighted in bold. 63

4.2 Nasa-tlx results. Statistically significant results (α=0.05) were highlighted in bold. 63

4.3 (a) Analysis of accuracy's for multiple models and sequence lengths. (b) Results for the from the 2-factor ANOVA model analyzing the effect inter-subject variations and sequence length. 71

4.4 Single user model evaluation results. 76

4.5 Multi user model evaluation results 76

4.6 Objective performance metrics results. Statistically significant results ($\alpha = 0.05$) were highlighted in bold 80

4.7 Nasa-tlx results and measured cognitive workload. Statistically significant results (α=0.05) were highlighted in bold. 81

LIST OF FIGURES

2.1 Surgical cart of the da Vinci Xi robot. Image taken from [21] 20

2.2 Classification system for autonomous systems in surgical robotics. Image taken from [46] 27

3.1 DVRK central components. (a) Patient side components. (b) Console components. 36

3.2 Sequence of snapshots (from a to f) showing the simulated cavity filling with blood and the autonomous arm navigating through the workspace to provide suction. 37

3.3 Relevant coordinate frames of the proposed system 38

3.4 Proposed fully convolutional network with a VGG-16 backbone. The architecture uses the following color coding: (1) green blocks represent convolutional layers, (2) orange blocks represent maxpooling layers, (3) blue blocks represent upsampling layers and (4) purple layer represent softmax layers. 41

3.5 Example of semantic segmentation image and label. The following color code is used for the labels: (1) red pixels correspond to blood poolings, (2) purple pixels correspond to the simulated vessel, and (3) blue pixels correspond to the instrument arms. 43

3.6 Examples of augmented images in the dataset 44

3.7 Calibration markers used to calculate homography transformation. 46

3.8 Augmented reality result. (a) Transparent texture aligns with the autonomous manipulator. (b)-(c) Texture moves to indicate the robot intended motion. (d) After the texture stops, the manipulator starts its motion. 48

3.9 Augmented reality implementation diagram 49

3.10 EEG channel locations 50

3.11 Feature extraction pipeline for EEG signals 52

3.12 Diagram showing how the EEG and Eye tracker signals are synchronized for the cognitive load detection system. 54

3.13 LSTM cell inner mechanisms. 56

3.14 Cognitive load and autonomous module integration. 58

4.1 Sequence of steps to perform a running suture task. (a) Grasp the needle. (b) Position the needle close to the vessel. (c) Insert the needle. (d) retrieve the needle. 61

4.2 Box plots of NASA TLX workload components. Overhead asterisk indicates statistically significant differences between the two modalities. 64

4.3 Sequence of actions to perform a peg transfer. 67

4.4 Correlation between accuracy and sequence length 70

4.5 Transfer learning results 72
4.6 Spectral analysis of EEG features 73
4.7 Tasks used to construct a dataset of cognitive workload signals. (a) Low difficulty task: Needle pass exercise without bleeding events. (b) High difficulty task: the simulator slowly fills up with blood as the user performs the tasks. 74
4.8 Cognitive module average predictions box plot 82

ABBREVIATIONS

AI	Artificial Intelligence
AR	Augmented Reality
CNN	Convolutional Neural Network
CRTK	Collaborative Robotics Toolkit
CT	Computed tomography
DVRK	Da Vinci Research Kit
ECM	Endoscopic Camera Manipulator
EEG	Electroencephalogram
FCN	Fully convolutional Networks
HMM	Hidden Markov Model
HRSV	High resolution stereo viewer
LSTM	Long short-term memory
MIS	Minimally Invasive Surgery
MRI	Magnetic Resonance Imaging
MTM	Master Tool Manipulator
OR	Operating Room
OSATS	Objective Structured Assessment of Technical Skills
PSM	Patient Side Manipulator
RMIS	Robotic-assisted minimally invasive surgery
STFT	Short-time-fourier-transform
SVM	Support Vector Machine

ABSTRACT

Robotic-assisted minimally invasive surgery (RMIS) has been steadily increasing since its introduction in the early 2000s and now has become a medical standard of care in multiple surgical specialties. In RMIS, the leading surgeon teleoperates a surgical robot from a console distant from the patient. While on the patient side, there is at least one surgical assistant supporting the procedure by effective handling surgical instruments. One of the most important tasks done by the surgical assistant is blood suction and irrigation. This task is critical to maintaining a clear view of the surgical field and avoid contaminations and infections. When several tasks are competing for the surgical assistant's attention, taking care of blood suction implies leaving unattended other assistive tasks, such as the exchange of robotic instruments and handling of sutures. An alternative approach to handle bleeding events is having the leading surgeon teleoperate the suction tool. Likewise, this leads to less attention allocated to the patient and an increase in their cognitive load.

This thesis describes a semi-autonomous suction assistant to release the main surgeon of blood suction during a procedure and avoiding the associated cognitive demands of such task. At the heart of this system, there is a deep learning algorithm that segments and identifies the location of blood poolings from the endoscopic camera's images. With the segmented images, the system extracts navigational information to provide automatic suction, allowing the leading surgeon to focus exclusively on the main task. The system was integrated into a Da Vinci Research Kit robot (DVRK). Additionally, an augmented reality (AR) and a real-time cognitive workload assessment module were developed to improve human-robot work dynamics. The AR module displayed semi-transparent annotations indicating the robot's next target location. These annotations allowed the user to better coordinate his actions with the surgical robot. The cognitive workload assessment module allowed to classify the users' mental state into low and high cognitive workload. Using this information, the robotic assistant provided suction during periods of high mental demands.

To evaluate the proposed framework, a computational experiment and two user studies were conducted. The goal of the computational experiment was to assess the prediction performance of the proposed cognitive workload detection system under two different modalities:

single-user models and multi-user models. In the single-user modality, an average classification accuracy of 76% was achieved. This result demonstrates how to effectively use EEG and eye-tracker features to predict cognitive states in RMIS procedures. In the first user study, the goal was to evaluate the capability of the autonomous framework to improve the user's surgical performance. This was achieved by comparing the autonomous system against a condition of manual teleoperation of the suction tool. The study's main finding was a reduction in the completion time and the reported workload demands when using the autonomous system. The goal in the second study was to evaluate the integration of the autonomous system with the cognitive workload framework. In this setting, the robotic assistant would only act when the user's mental state was classified as high cognitive workload. To achieve the study's goal, the autonomous system was assessed against manual teleoperation of the suction tool. The main results for this study show a reduction in the completion time and improved human-robot collaboration fluency. Overall, the experiments' results shows how objective and real-time assessment of cognitive load can be used together with surgical autonomy to enhance RMIS surgical outcomes.

1. INTRODUCTION

Robotic-assisted minimally invasive surgery (RMIS) has become the medical standard for surgical procedures in multiple specialties such as urology, general surgery, and gynecology [1]. Currently, the most predominant robotic platform is the da Vinci surgical robot developed by Intuitive Surgical, Inc. To date, this company has installed more than 5900 da Vinci systems worldwide, allowing robotic surgeons to perform more than 8.5 million robotic-assisted procedures [2]. As technology improves and adoption rates increase, it is expected that the market of surgical robots will rise to a value of USD 11.8 billion by 2025 [3].

In RMIS, the operating or leading surgeon sits in a console separated from the patient and teleoperates the surgical robot at the patient side. On the patient side, there is at least one surgical assistant who helps with several support tasks, such as exchanging the robotic tools, handling of sutures and specimens, and providing suction and irrigation [4]. Among the assistant responsibilities, blood suction and irrigation tasks are critical to maintain a clear surgical field and avoid complications during a procedure often resulting from bleeding. Furthermore, bleeding has been associated with leading surgeon's high cognitive load and limited visibility of the surgical field. This can lead to increased surgical risks [5] and even mortality risks [6]. While blood suction does not stop bleeding, it can facilitate proper hemorrhage control by allowing the surgeon to localize the source of bleeding and treat it in a timely fashion. In the most severe hemorrhage control cases, the surgeon must undock the robot and switch to open surgery. However, conversion to open surgery has been associated with postoperative complications such as morbidity, infections, and increased risk of subsequent surgeries [7].

Having good visibility of the operating field can help prevent bleeding complications. A clear surgical field allows the surgeon to quickly identify and act upon the source of a hemorrhage [5] and stop bleeding. Increasing visibility during bleeding requires effective use of the suction/irrigation tools multiple times per procedure [5]. Standard medical practices require a surgical assistant to hold the suction tool at the side of the patient while the leading surgeon operates from a close-by console [8], [9]. The main drawback of such configuration

is that it limits the availability of the assisting surgeon to other essential support tasks, such as tissue retraction, instrument exchange, and suture [4].

In response, flexible suction systems, such as the ROSI, have been recently developed [10], [11]. These novel technologies allow the surgeon to control the suction tool directly, releasing the surgical assistant from suction/irrigation tasks. Paradoxically, having the leading surgeon in charge of blood suction leads to additional cognitive demands (since there are other tasks that the surgeon is responsible of) that can result in delays and, potentially, medical errors [12]. Furthermore, these additional cognitive demands can be particularly detrimental for surgeons earlier in the learning curve as they diverge their attention from the main procedure.

To alleviate these needs, we propose utilizing an autonomous robotic blood suction system. Such a framework would help to improve performance and reduce the cognitive demands of the medical staff. This system leverages on recent artificial intelligence (AI) advancements to segment automatically blood pooling regions in endoscopic camera views. Segmentation of endoscopic images is done with a novel fully convolutional network (FCN) [13]. Such a network assigns a label to every pixel; a task referred to as semantic segmentation. Later, the segmented images are used by a da Vinci robot to automatically extract navigational cues necessary for effective suction of blood accumulations.

To achieve our goal of autonomous blood suction, we first developed a surgical simulator that allows benchmarking the system's performance. We then built a platform to generate mock bleeding events. Such events were designed to increase the operator's cognitive workload and test the effectiveness of the system. Additionally, to improve the simulator's realism, ballistic gel was used to simulate the muscle tissue. Utilizing such simulator, a dataset of endoscopic images of users performing a running suturing exercise on the simulator was collected. This exercise was borrowed from the fundamentals of robotic surgery [14]. This dataset was used to train the semantic segmentation algorithms. Finally, a navigation and path planning algorithms were developed to move the surgical robot around the surgical workspace and provide suction.

Enhancing human-robot collaboration is a two-faced problem. On the one hand, it is required that the autonomous agent adapts its behavior according to the surgeon's mental state and needs at a precise moment. On the other hand, the robot must convey its inten-

tions to the surgeon [15]. In this regard, two additional components were added to boost the effectiveness of our autonomous framework. First, to improve the robot awareness of the surgeon's mental state, a cognitive workload detection algorithm based on EEG and eye-tracking sensors was designed and implemented. This framework allows to estimate the cognitive demands of the user in real-time and adapt the autonomous assistant accordingly. Second, to improve the surgeon's awareness of the robot's intentions, transparent AR annotations were developed that allowed the user to see the subsequent actions of the robot before they were executed. Thus, allowing the user to coordinate its actions with the robotic assistant.

To validate our approach, a user study with a first-generation Da Vinci Research Kit (DVRK) surgical robot that used the API version 1.7.1 was conducted. This study aimed to demonstrate that working with the robotic assistant can result in better performance and lower cognitive demands on the user. To achieve this, we evaluated the user performance in a surgical exercise under two modalities: manually controlled-suction tool (*manual* modality) and autonomy controlled-suction tool (*autonomous* modality). In the former, the user shifts between the teleoperation of the main arms and the suction tool as needed during the surgical training task. In the latter, the robotic assistant provides automatic blood suction directly. In both modalities, performance and workload metrics were collected to assess the effect of the AI system on task performance.

1.1 Problem Statement

This research focuses on the development of an algorithmic and mechanic framework to automate the blood suction surgical task. The autonomy algorithm is expected to enhance performance in the context of RMIS, therefore, indirectly reducing the cognitive demands of the user. Our framework was only tested on suction tasks due to recent studies indicating that a combination of irrigation and suction does not lead to additional surgical benefits compare to suction alone [16].

1.1.1 Research Questions

Research question 1 (RQ1): How to develop computer vision algorithms and control policies to automate the blood suction task in the context of RMIS?

This question deals with the technical requirements of developing an autonomous system for blood suction. We focus on the development of computer vision algorithms to extract critical cues such as the location of the blood pooling and the handling of robotic instruments.

Research question 2 (RQ2): How to predict cognitive load from multimodal physiological signals in real-time?

A critical step to improve human-robot collaboration in surgery is providing the autonomous system with information about the surgeon's cognitive demands. This information can be used to adapt the level of assistance of the autonomy, i.e., providing additional help on cognitive demanding situations. This question focuses on the algorithmic design and implementation of cognitive workload prediction models based on EEG and eye-tracking sensors.

Research question 3 (RQ3): How the autonomy-controlled suction action can improve the surgeon's task performance and cognitive demands?

Our baseline hypothesis is that automating the blood suction surgical subtask improves the surgeon's performance. In this regard, demonstrating that our autonomous system leads to better performance and lower cognitive demands than the suction tool's manual teleoperation is a critical step for this work.

1.2 Contributions

The main contributions of this work are:

1. Design and implementation of a semi-automatic blood suction system deployed in a DVRK robot.

2. Design and implementation of a real-time and unobtrusive cognitive workload prediction framework based on physiological sensing.
3. Implementation of transparent augmented reality annotations to improve human-robot collaboration.
4. Evaluation of the cognitive demands and surgical performance of the surgeon when working with the semi-autonomous robotic assistant.

1.3 Summary

This chapter discussed the motivation for developing an autonomous robotic blood suction system in the context of minimally invasive surgery and presented three research questions related to the design, development, and assessment of the proposed framework. The rest of this work is divided into the following sections. In chapter 2, we present a review of the relevant literature. Chapter 3 describes the methodology used to develop the autonomous system and the cognitive workload prediction algorithms. Chapter 4 discusses the results obtained from the user study and performance metrics from the system. Finally, in chapter 5, we present conclusions and future work.

2. LITERATURE REVIEW

This chapter reviews the main research topics in surgical robotics that are related to the presented research questions. In sections 2.1 and 2.2, we will describe current commercially available surgical robot platforms, and the most common research platforms. Next, in section 2.3, we delve into computer vision applications in surgery as this is a key enabling technology for the development of autonomous platforms. Section 2.4 and 2.5 cover the main topics of this thesis which are autonomy systems and human-robot integration in surgical robotics. Finally, sections 2.6 and 2.7, will go over objective performance metrics and workload assessment techniques.

2.1 Surgical Robotics

Minimally invasive surgery (MIS) is a surgical technique where the surgeon utilizes small ports (incisions) made in the patient's body to perform a surgical procedure. This technique has resulted in improved patient outcomes such as reduced recuperation times, lower risks of complications, and shorter hospital stays [17]. The main drawbacks of traditional MIS are (1) limited dexterity of laparoscopic instruments, (2) lack of tactile feedback, and (3) the loss of depth perception due to monocular endoscopic cameras and 2D displays. Robotic surgical platforms were introduced to the market in the early 2000s to improve the surgeon's dexterity and visualization. Since then, this technology has evolved into a growing industry encompassing surgical care in multiple specialties such as general surgery, gynecology, orthopedics, and radiotherapy [18]. The main drawbacks of current commercial platforms are their high costs and large footprints.

The field of surgical robotics originated with researchers and engineers that adapted industrial robots to clinical tasks. In their seminal work, Kwoh *et al.* proposed using a PUMA 200 robot to perform a guided brain biopsy with computed tomography images (CT). This work led to one of the first robotic surgical procedures in 1985. In 1992, THINK Surgical, Inc introduced the Robodoc, a system capable of autonomously assisting in orthopedic surgeries [19]. In 1994, Computer Motion (Goleta, California, USA) released the precursor of teleoperated surgical robots, the Automated Endoscopic System for Optimal Positioning

(AESOP). This system evolved into the ZEUS robot by adding remotely controlled manipulators for laparoscopic instruments. The direct competitor of ZEUS was the da Vinci system by Intuitive Surgical. Both companies received FDA approval for their products in the early 2000s. In 2003, these rival companies merged, resulting in the da Vinci as the only robot for MIS in the market for many years [18].

Today, the most prominent surgical robot is still the da Vinci robot. This system is composed of three central components: (1) a cable-driven surgical robot with three 7-DOF patient side manipulators and a 4-DOF endoscopic camera, (2) a master console with 3D immersive visualization of the surgical field and two teleoperation manipulators, and (3) a vision-tower with video processing hardware and generators of electrosurgical instruments. The latest version of the da Vinci is the Xi version, and it was released in 2014 (see figure 2.1). The main criticisms towards da Vinci's systems are the lack of tactile feedback, high acquisition price, and oversized design that requires specially designed ORs [20].

Figure 2.1. Surgical cart of the da Vinci Xi robot. Image taken from [21]

Currently, multiple other robotic companies have developed novel platforms that address the main limitations of the da Vinci system. The biggest competitor is the Senhance laparoscopic robot (TransEnterix, Morrisville, NC, USA). This robot has 4 robotic manipulators each one mounted on its own cart. Unlike the Da Vinci console, this system utilizes a combination of 3D glasses and 3D screens to provide the surgeon with a view of the surgical field. Additionally, it includes novel haptic feedback and eye-tracking technology. The console was designed to ease the transition from traditional laparoscopic surgery to robotic surgery. Two other robotic platforms are the BITRACK (Rob Surgical, Barcelona, Spain) and Revo-i robots (Meere Company, Seoul, South Korea). Both platforms are similar in size and functionalities to the Da Vinci and Senhance robot; however, they are expected to include haptic feedback technology and have a lower cost. Currently, the Revo-i and the BITRACK systems do not have FDA approval.

Other platforms under active development are attempting to further reduce the footprint of surgical robots. The main advantage of portable surgical robots is that they do not require specialized operating rooms, which would allow hospitals and small clinics to reduce costs [22] as current OR could be used to host such systems. The Versius surgical robot developed by Cambridge Medical Robotics (Cambridge, UK) has independent surgical arms mounted in a reduced cart size compared to the Sehance robot. This robot offers a non-immersive teleoperation platform and the unique feature of allowing the surgeon to stand while operating. The most compact designs have been achieved by mounting the robotic arms directly into the surgical table with systems such as the Mirosurge robot by DLR Institute of Robotics and Mechatronics (Germany).

Commercial surgical robots have also been developed for specialties such as orthopedics and radiotherapy. In orthopedic surgery, precise models of bones can be calculated with preoperative images such as CT scans. This allows robots to plan and complete a procedure autonomously. One of the most notable platforms for knee and hip replacement surgery (arthroplasty procedures) is the successor of the Robodoc platform, the TSolution-One by THINK Surgical Inc. (Fremont, CA) [19]. For spinal surgery, the Mazor X Stealth by Medtronic has been proposed for automatic screw placement. This robot requires a CT 3D model of the spine to calculate the optimal trajectories [23]. In radiotherapy, the CyberKnife

robot by Accuray (Sunnyvale, California, USA) is commercially used for radiation therapy of cancerous tissue. This computer-guided system uses preoperative images to calculated radiation angles that maximize exposure of the target and minimize the impact on healthy structures. The aforementioned are examples of high autonomy levels where the surgeon's role is to approve the pre-planned surgical plan instead of performing the procedure.

2.2 Surgical Robotics Research Platforms

The introduction of RMIS platforms became a driver of innovation in health sciences and engineering. For instance, the da Vinci surgical robot has led to more than 25,000 clinical and engineering peer-reviewed publications [24] alone. In the 2000s, there were significant barriers to the field of surgical robotics to multiple research institutions due to the high costs and commercial limitations that prevented accessing the robots' kinematic and video information. Consequently, only two options existed to conduct research in RMIS: customized surgical robots or industrial platforms adapted for surgery [24]. To address this situation, two open-source robotic platforms have been introduced: The Raven II robot [25] and the Da Vinci Research Kit (DVRK)[26]. These two platforms have been installed in more than 40 research institutions around the world and have become a key enabler for the formation of a growing research community in surgical robotics [24]. These common research platforms offer three main advantages: (1) the availability of open surgical systems that allow changes of the software stack, (2) open-source code repositories that facilitate reproducing and validating other researchers' work, and (3) access to low-cost surgical hardware.

The DVRK is a modular surgical platform based on first-generation da Vinci surgical systems provided free of cost by Intuitive Surgical Inc [26]. The controllers and software were developed in a collaboration between Johns Hopkins University (JHU) and the Worcester Polytechnic Institute (WPI). A controller based on Field programmable gate arrays (FPGAs) serves as a communication bridge between the actuators and sensors of the robot and an external computer used to control the robot. The main components of this system are the 7 DOF Patient side manipulators (PSM), a 4 DOF Endoscope camera manipulator (ECM), a 7DOF master tool manipulator (MTM), and a stereo-viewer visualization console. The

modular architecture of the DVRK allows for each component to be used independently or to be assembled, mimicking the clinical version of the da Vinci robot.

The Raven II robot is a 7 DOF cable-driven robotic platform developed in the year 2012 by the University of Washington and the University of California Santa Cruz [25]. This robot has two coarse 3 DOF positioning mechanisms and allows 4 DOF wrists-like instruments to be mounted. Additionally, the system was made compatible with DVRK's laparoscopic instruments. Its software is open-source and relies on ROS and Linux. Presently, the robot is being manufactured and distributed by Applied Dexterity Inc. Last, the Taurus Robot by SRI International (Menlo Park, USA) has been proposed to perform surgical tasks in austere environments such as the battlefield. This robot has two dexterous 7 DOF manipulators and a lightweight design that allows the robot to be mounted in a wheeled cart. Madapana *et al.* developed a dataset of surgical skills that was used to transfer surgical knowledge from the DVRK robot to the Taurus [27].

Multiple open-source simulation platforms have been developed for surgical robots for testing control algorithms and surgical training. Gondokaryono *et al.* developed a simulator of the DVRK robot in Gazebo based on a closed-kinematics model [28]. Additionally, Fontanelli *et al.* implemented a DVRK kinematic model in the V-rep simulation environment and 4 surgical scenes to showcase the potential applications of simulation in surgical robotics [29]. While these simulators provided full kinematic models for the DVRK, they lack the adaptation flexibility to different input controllers and robotic architectures. To alleviate this situation, Munawar *et al.* proposed a more general and flexible simulation platform called the Asynchronous Multi-Body Framework (AMBF) [30]. This framework allows working with multiple devices such as Razer Hydra and the DVRK master tool manipulators (MTM); and in addition, it supports the DVRK robot and the Raven II robot.

Last, there has been an effort in the robotics community to standardize the DVRK's API to be compatible with Raven II robots through the Collaborative Robotics Toolkit (CRTK) initiative. The goal of CRTK is to facilitate the reproduction of results and avoid unnecessary duplication of work by developing a standard interface and specifications for teleoperated surgical robots [24].

2.3 Computer Vision

The advent of minimally invasive techniques (MIS), e.g., laparoscopic and robotic-assisted procedures, has significantly increased the availability of surgical video data. In MIS procedures, the view of the surgical field is provided by either a monocular or stereo endoscopic camera. Easy access to datasets of surgical videos has resulted in the development of multiple computer-aided applications based on data-driven and machine-learning techniques. Examples of such applications are extracting postoperative information of the patient anatomy, evaluating the surgical performance of surgeons and residents, and providing real-time navigational information to improve surgical work.

Computer vision applications for MIS can be broadly separated into three categories [31]: (1) surgical phases understanding, (2) computer-aided detection, (3) navigation applications. The first category deals with the problem of automatically separating a surgical procedure into sub-phases and extracting contextual information about the surgical steps being performed at every phase. The granularity of the temporal divisions depends on the application. Hager *et. al* [32] proposed dividing the surgical maneuvers into atomic actions referred as to *surgemes*. The second category addresses the problem of automatically detecting relevant objects or events in the surgical field, such as the surgical instruments, anatomical structures, and anomalies. Examples of anomalies are lesions and hemorrhages. The last category focuses on extracting 3D navigational information that can be utilized to improve motion planning, or autonomous navigation of the arms, or to facilitate surgical instruments control within the patient's body. Overall, these applications attempt to harness the power of data-driven techniques to enhance medical staff performance and patient safety.

2.3.1 Segmentation and Registration of Patient Anatomy and Instruments

Segmentation and registration from surgical images are key techniques to improve performance and safety in the operating room. By registering the observed surgical field with preoperative scans and 3D models, patient anatomy's information can be overlayed on top of the endoscopic images to improve the surgeons' navigation and decision-making [33]. Combining registration with instrument segmentation can be used to create safe operating zones

that avoid vital structures and provide guidance during critical steps in the surgery. Additionally, instrument segmentation can be used to evaluate technical skills and generate labeled datasets for machine learning algorithms [31].

State-of-the-art models for instrument segmentation are based on convolutional neural networks (CNN's)[34]. These models allow segmenting relevant objects by assigning a label to each pixel in the surgical image. Examples of surgical regions of interest include anatomical structures such as organs, lesions, tissue, bone, and surgical instruments. [35] proposed a lesion detection system for gastrointestinal endoscopic images, which used Support Vector Machine (SVM) and CNNs. This system was designed to detect lesions such as ulcers, polyps, and bleeding automatically. [36] proposed segmenting connective tissue using a variant of the U-net architecture. It used RGB depth maps as inputs to the network and was part of an autonomous tissue retraction application. For instrument segmentation tasks, [37] introduced a fully convolutional variation of the residual network Resnet-18. This architecture utilized a combination of low stride convolutional layers with dilated convolutions to ensure dense feature maps.

Additional challenges for segmenting surgical images are the lack of large-scale surgical datasets and the occlusion generated due to bleeding and smoke from cutting devices. To resolve the lack of annotated data, [38] proposed the use surgical instruments on top of green screens to generate semi-synthetic labeled data quickly. Afterward, models trained with the synthetic datasets were shown to obtain comparable results as models trained with real data. To address challenges due to occlusion, [39] proposed a system to segment occluded organs by fusing preoperative 3D computed tomography scans (CT) and the stereo endoscopic vision of a surgical robot.

2.3.2 Hemorrhage Detection and Tracking in Surgery

Hemorrhages detection and tracking are needed for an autonomous blood suction and irrigation robotic system. Traditional blood detection algorithms are based on pixel-wise classifiers relying on color representations such as RGB and HSV. Fu *et al.* used RGB features from wireless capsule endoscopy (WCE), and trained an SVM model to automatically

segment bloodstains [40]. This same approach was employed by Okamoto *et al.* to detect bleeding sources during laparoscopic surgeries [41]. Although computationally efficient, these methods had limited performance when specular reflections or smoke from ablation was present in the images [42].

State-of-the-art models for blood detection are based on convolutional neural networks (CNN). These models provide robust blood tracking even with specular reflections, but they also require larger labeled training datasets. While these methods were initially developed for fluids detection, they were adapted to blood detection [42]. Schenck *et al.* demonstrated the effectiveness of fully convolutional networks (FCN) liquid detection on pouring demonstrations [43]. To build a large, annotated dataset quickly, they used a simulator where liquids were heated first, acquired using a thermal camera, and subsequently segmented and annotated automatically. For a liquid tracking application, Yamaguchi *et al.* demonstrated that optical flow using Lukas-Kanade method and stereo cameras could be used to create 3D models of flowing liquids [44]. Subsequently, these models were used by a robot for a liquid manipulation task. Instead of calculating optical flow with traditional methods, Ritcher *et al.* utilized pre-trained deep learning architecture to improve the tracking accuracy in the surgical setting [42]. The resulting optical flow was later filtered using Hidden Markov models (HMM) to temporally fuse the predictions and increase the tracking robustness.

We build on this previous work on FCN networks, as they provide the most robust results for blood detection. To alleviate the need for a large training dataset, we leverage pre-trained models with ImageNet [45]. These models were later retrained to the specific setting of blood detection in RMIS with a dataset of 180 images.

2.4 Autonomy in Surgical Robotics

Autonomous systems have been proposed as one of the many important stepstones in surgical robotics [46]. Because autonomous systems can take over the most repetitive and tiresome tasks of a surgical procedure, they would reduce the demands on the surgical staff, enhance the efficiency in the OR and improve the overall patient well-being [47], [48]. Furthermore, autonomous technologies can enable bringing surgical care to locations where

teleoperation is not possible due to delays and unreliable communication [49]. These hostile environments include war zones, disaster and recovery regions, and space explorations missions, among others [50].

Figure 2.2. Classification system for autonomous systems in surgical robotics. Image taken from [46]

To guide the progress of this field and address legal and ethical barriers Yang. *et. al* proposed a 6-level classification system for autonomous surgical platforms[46] (see figure 2.2). The levels of such system are: **(level 0)** No autonomy, **(level 1)** Robot assistance, **(level 2)** Task autonomy, **(level 3)** Conditional autonomy, **(level 4)** High autonomy, **(level 5)** Full autonomy. Level 1 encompasses any system that provides assistance to the surgeon but never takes full control, e.g., virtual fixtures [51], enhanced visualization [52], and haptic feedback [53]. Level 1 autonomy is often referred as *shared control* teleoperation [48]. Level 2 includes systems capable of performing autonomously specific subtasks or procedures. At this level, the goal is to develop systems that can take over the execution of non-critical and repetitive tasks such as tissue retraction and suturing [46]. Level 3 systems are capable of automatically generating strategies to complete a surgical goal; however, the surgeon is still in charge of selecting among those strategies [46] which one is the most effective. Level 4 systems can

perform complete surgical procedures under the supervision of a surgeon. Finally, level 5 systems are as proficient as surgeons and do not require human oversight. Currently, there are no commercial examples of levels 5 systems [54]. Our proposed autonomous blood suction system corresponds to a level 2 system.

Level 2 autonomous systems can account for changes in their environment to complete a specific goal even under uncertain conditions [54], such as anatomy variations from patient-to-patient. These systems differ from traditional automation-based systems that can only work under controlled and predictable conditions [54]. In this regard, a key capability of such a system is to estimate a model of the environment based on high-level information from raw sensory data, e.g., intraoperative images, ultrasound, CT scans [50]. After estimating an environment model, a mapping function is used to transform the estimated environment model into actions. This mapping function can be attained through either *explicit learning* or *implicit learning* [50]. In the former, the robot behavior is explicitly programmed. In the latter, the robot can learn the surgical skills by observing expert demonstrations or training in a virtual reality environment using reinforcement learning [48]. *Implicit learning* is a convenient task-independent method for teaching robotics; however, it is computationally expensive and currently does not generalize well to complex procedures with multiple steps [55].

Level 2 autonomous solutions have been proposed for Fundamental of Laparoscopic Surgery (FLS) tasks, such as peg transfer [56], pattern cutting, and surgical subtasks such as debridement, tissue retraction, and blunt dissection. Hwang *et al.* [57], developed a system that led to better performance when compared to surgical residents in a peg transfer task. This system used a Zivid RGBD camera sensor to obtain a 3D reconstruction of the board and the optimal grasping point for the pegs. This system's main innovation is high-precision control that uses a recurrent neural network to estimate the robot's dynamic state. Additionally, they proposed a calibration method to reduce inaccuracies due to cabling effects. Ettorre *et. al* proposed a system for autonomously grasping needles. To achieve this, they utilized a color-based needle tracking algorithm and a visual servoing system to approach the needle [58]. This work was expanded by Sundaresan *et al.* to automatically extract needles from tissue phantoms [59]. Their algorithm used a non-color-based and marker-less

segmentation algorithm to detect optimal grasping points and extraction trajectory. Surgical automation has also been applied to the task of tissue retraction. Nagy *et. al* proposed a soft tissue retraction system using a disparity map in combination with a fuzzy controller to obtain the optimal retraction angle [60]. This same vision system was used by Elek *et al.* to develop an automatic blunt dissection system [61].

2.5 Human-Robot Interaction in Surgery

Surgical performance and patient care can be improved by enhancing the human and the robot interaction in RMIS [18]. The challenges faced by surgeons during surgery can be coarsely organized into three categories: decision-making, navigation, and anatomy recognition capabilities. All of these can be improved by better integration between the robotic platform and the surgeon.

Decision-making in surgery can be improved by integrating different preoperative and intraoperative imaging sources at the operator's console [33]. Examples of such sources are ultrasound imaging, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) scans. A commercial example of a multi-input display is the TilePro^{TM} system developed for the Da Vinci surgical robot [62]. This system allows the surgeon to simultaneously visualize the endoscopic camera along with two other digital data sources. Particularly, ultrasound probes can be mounted on robotic manipulators to provide intraoperative information of the patient's anatomy. For instance, Mohareri *et al.* proposed a transrectal ultrasound system probe (TRUS) that can be teleoperated from the da Vinci robot console to obtain real-time information of cancerous tissue [63].

Navigation systems can be used to safely move the robotic instruments within the patient body to minimize damage to tissue and organs. The surgeon's navigation capabilities can be improved using advanced visualization techniques such as Augmented Reality (AR) and VR. AR can increase the surgeon's situational awareness by adding additional visual information about occluded regions in the anatomy [64], or identified lessions. Lee *et al.* proposed using AR in preoperative CT to highlight critical hidden nerves during robotic thyroid surgery [65]. To improve the surgical assistant visualization, Qian *et al.* proposed using a head-mounted

display to show 3D AR models of the laparoscopic tools and robotic tools in the surgical field.

Surgeons' performance can also be enhanced via collaboration with autonomous robotic assistants [48]. A commercial example of such a system is the VIKY UP robot (Endocontrol, La Tronche, France). This robot is a uterine positioning system that allows the surgeon to displace the uterus with voice commands. This system allows to reduce the risk of injury to the uterus during gynecology procedures [18]. In urology procedures, the AquaBeam System was used to autonomously resect tissue from the prostate utilizing a high-velocity water stream [66]. The surgeon working with this system creates a surgical plan by selecting the tissue to be removed, and then the robot performs the dissection autonomously under the surgeon's supervision.

An alternative scheme for human-robot collaboration is shared-control teleoperation. Under this scheme, the robotic end-effectors are controlled by the human, and an autonomous agent at times each. Kaplan *et al.* demonstrated how precision on a tumor detection task could be improved with a haptic feedback system and a shared-control strategy[67]. Padoy *et al.* developed a human-machine collaborative system that allowed to shift between a manual and an autonomous teleoperation mode to complete a surgical task[68]. This system used a Hidden Markov Model (HMM) to identify procedural sub-tasks that allowed a seamless transition between the two operation modes.

2.6 Surgical Performance Metrics

The long-established "See one, do one, teach one" has been the guiding principle for surgical training [69]. This training paradigm is based on the idea that becoming proficient in surgery requires first to learn by demonstration, then to learn by practicing, and finally to learn by teaching others[69]. This approach to teaching has been increasingly criticized for being inefficient and inadequate for the standardization of current surgical training [70], [71] [72]. One of the limiting factors towards surgical training is the lack of objective skill assessment metrics [73]. Traditionally, surgical skill has been quantified using patient outcomes and subjective structured assessments [72]. The former provides indirect metrics about sur-

gical skills but are highly variable between patients. The latter is based on surveys completed by surgical experts during proctoring sessions. One of the most common structured assessment surveys is the Objective Structured Assessment of Technical Skills (OSATS) [74]. The main shortcomings of this method are the rater-to-rater variability and the need to have an available expert for assessment.

A different and arguably more objective approach for skill assessment is to use motion analysis [72] metrics. This approach is objective and can provide real-time feedback to trainees. Motion Analysis techniques can be divided into (1) Descriptive statistics methods and (2) Language modeling methods [72]. Descriptive statistics refer to high-level features of the data of a procedure such as an instrument velocity, economy of motion, and completion time. In robotic surgery, Kumar *et al.* showed in a longitudinal study that motion statistics of the console manipulators, such as the total clutching time and the manipulators' total displacement, were strongly correlated with OSATS assessments [75].

In addition, motion objective metrics were successfully used to classify surgical expertise. Nguyen *et. al* proposed a set of automated performance metrics (APM) based on instrument motion and system events that were correlated with expertise and patient outcomes. In this thesis, we will leverage instrument motion objective metrics to compare the skill level of the users when doing a surgical exercise with a manually teleoperated suction action and an autonomy-teleoperated suction action.

2.7 Workload Assessment Techniques

Surgical robotic platforms have improved the surgeon's dexterity and ergonomic conditions. Nevertheless, surgeons are still susceptible to high cognitive demands due to the lack of tactile feedback, and overwhelming visual and audio signals in the OR [76]. Prolonged states of high cognitive load can considerably affect surgeons' performance and reaction to unexpected events [77], [78]. Long periods of high cognitive load can be reduced by efficiently predicting the current and future mental states of the surgeon and by providing real-time feedback to the surgeon about her cognitive state [79]. If a surgeon is made aware of his

cognitive load and its effect on performance, they can either request assistance or take a short break.

Traditionally, quantitative cognitive load assessments were conducted using subjective metrics such as self-reported evaluations (e.g., NASA TLX) and objective metrics such as the performance in secondary task [80]. However, these methods are not appropriate for measuring intraoperative cognitive load as they distract the surgeon and compromise the safety of the patient [12]. In contrast, the physiological analysis relies on sensor data to assess cognitive load in an objective, continuous, and unobtrusive manner. Among the physiological signals, electroencephalogram (EEG) stands out because it provides discriminative information of the operator's brain activity. Such activity correlates strongly with cognitive workload [81], [82]

Cognitive workload recognition based on EEG was traditionally achieved by classifying the temporal and spectral signal features into a set of mental states. For instance, the Filter Bank Common Spatial Pattern was popularly used to construct features based on subject-specific frequency bands [83]. Although traditional approaches yielded promising results, they are often user and task-dependent, i.e., trained models would only work for a specific task and user. In other words, cognitive assessment models based on EEG data are intra and inter-user dependent [84], [85].

Researchers have proposed using deep learning architectures as an alternative to the time-intensive feature engineering work required when using EEG classifiers [86]. EEG classification using deep learning architectures can be divided into two categories: temporal and frequency/spatial models. The temporal models utilize raw EEG data as an input and applies 1D convolutional operators across each channel to obtain low dimensional features maps. These models are advantageous because they allow building end-to-end workload classification systems, i.e., systems that skip intermediate preprocessing steps, making predictions directly from the EEG raw data [87], [88].

The frequency/spatial models utilize preprocessed data obtained via either spatial or frequency decomposition transformations. Such transformations result in features containing frequency information of the signal [89]. Common spectral representations for these models are the Short-Time-Fourier-Transform (STFT) and the spectral power band coefficients. The

primary benefits of these models are that they can be trained with smaller datasets compared to temporal models and are easier to understand [90]. For instance, *Wang et al.* proposed a deep learning approach to classify EEG signals encoded as spectrograms (Frequency and temporal) [91]. In that vein, *Jiao et al.* [92] improved that approach by combining the frequency information with the location of each electrode into a power scalp map.

For our cognitive workload detection module, we leverage previous work on frequency/spatial models as they are more data-efficient and easier to interpret than temporal models. Mainly, we used the spectral power band coefficients as they have been previously shown to be discriminative of mental workload demands [93].

2.8 Discussion

Collaborative autonomous robots in the surgical setting can improve performance and help reduce surgeons' cognitive demands during surgery [48]. Additionally, these lower mental demands could lead to improvements on patient care [12]. However, surgery is a very complex activity that requires physical and cognitive skills far above the machine's abilities in its current state of development. To overcome these barriers, novel sensing and visualization technologies are being developed to reach similar perception capabilities than those found in humans. Dexterous motor control and manipulation are being studied, together with computer vision algorithms to allow for safe autonomous navigation through the patient anatomy. Finally, improving the integration between the surgeon and the robot is critical to ensure technology acceptance and safety.

2.9 Summary

This chapter reviewed all relevant research areas related to semi-supervised telesurgical robots. First, commercial surgical robotics platforms were presented with an emphasis on future technologies. Such technologies have the potential to improve patient care and surgeons work performance. Second, surgical robot research platforms and their importance to technology development were presented. Third, an overview of current research trends in computer vision for surgical applications was discussed as they are a key enabling tech-

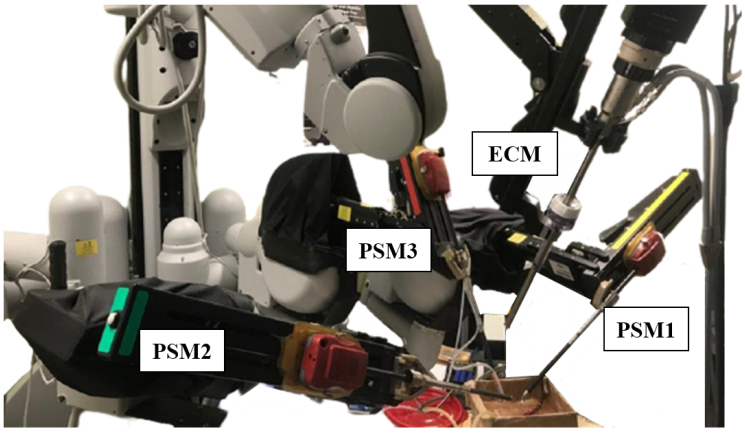
nology for autonomous systems. Then, autonomy in surgical robotics, the main topic of this thesis, was presented. This section focuses on task autonomy as our system provides autonomous assistance for blood suction surgical subtask. Then, human-interaction applications in RMIS were discussed. The last two sections describe objective skill assessment techniques for surgical skill and workload detection in RMIS.

3. METHODOLOGY

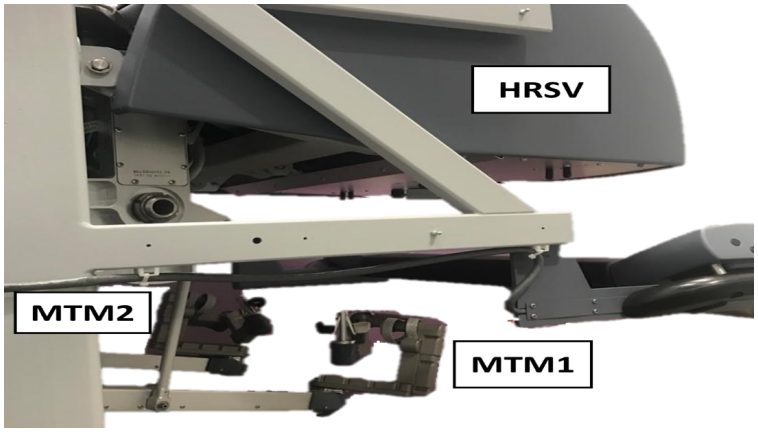
This chapter describes the techniques, algorithms, and apparatus used to develop and implement the autonomous robotic suction system. First, we describe the da Vinci Research Kit (DVRK), the open-source surgical robot used to deploy the proposed algorithms, and a developed surgical bleeding simulator used as a testbed for the system. Next, we present the core contributions of this dissertation: the autonomous robotic suction system and the cognitive workload prediction module. The section regarding the autonomous module is divided into the computer vision algorithms, the robot integration module, and the augmented reality visualizations. The section regarding the cognitive prediction module is divided into the description of the physiological sensors, feature extraction algorithms, classification models, and real-time synchronization and deployment.

3.1 Da Vinci Research Kit Description

The autonomous robotic system for blood suction was implemented on a DVRK robot (Refer to section 2.2) [26]. This robot is composed of four central components: (1) three Patient Side Manipulators (PSMs), (2) an Endoscopic Camera Manipulator (ECM), (3) two Master Tool Manipulators (MTMs), and (4) a High-Resolution Stereo Viewer (HRSV). The PSMs and the ECM are mounted in a surgical cart that optimally located the manipulators over the surgical workspace (See figure 3.1a). MTMS and the HRSV are located in a console from where the PSMs can be teleoperated (See figure 3.1b). The PSMs are 6 DOF wrist-like instrument arms that allow the surgeon to interact with the surgical environment. The ECM delivers a stream of stereo images of the robot workspace directly to the console at the user end. For this work, the PSM1 and PSM2 arms were teleoperated by the user, while the PSM3 moved autonomously following the proposed algorithm. To provide the robot with suction capabilities, we fabricated a custom suction tool with submersible motor pumps and silicone tubing attached to the PSM3. The DVRK only allows for the simultaneous teleoperation of two instrument arms. This means that if a user wanted to teleoperate the suction tool (PSM3), he would have to tap a pedal to swap the teleoperation between PSM2 and PSM3.



(a)



(b)

Figure 3.1. DVRK central components. (a) Patient side components. (b) Console components.

3.2 Simulator Design

To assess the capabilities of the implemented robotic autonomous system, a simulator was developed to perform a vessel ligation procedure [94]. The simulator was fabricated with colored ballistic gel, given its resemblance to human muscle tissue. The gel was molded to enable the formation of several cavities where the artificial blood could accumulate. In addition, small silicone tubes were incorporated in predetermined locations. One end of these tubes was connected to submersible water pumps to generate the bleeding events. The artificial blood used was created by combining water with food coloring. Finally, a human vessel was synthetically created using flexible rubber tube that was placed in the

middle of the simulator (See figure 3.7). During the evaluation experiments, the robot tool tip interacted with this vessel to perform a training exercise: running suturing exercise as described in the fundamentals of robotic surgery [14] (See subsection 4.1.1). Figure 3.2 shows a sequence of snapshots of the simulator cavities as they fill with blood and the autonomous robotic system performing the suction tasks.

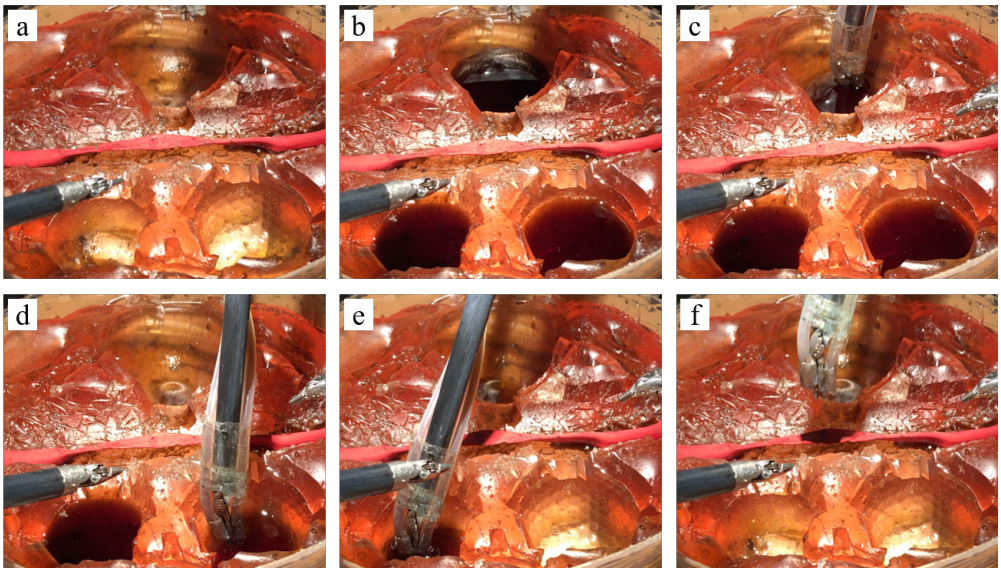


Figure 3.2. Sequence of snapshots (from a to f) showing the simulated cavity filling with blood and the autonomous arm navigating through the workspace to provide suction.

3.3 Autonomous Suction Robotic Assistant

This section describes the design and development of the autonomous suction assistant. This framework is composed of three main components. (1) The computer vision algorithms used to segment the blood accumulations automatically on the robot endoscopic images. (2) The robotic module used to transform the pixel locations of the segmented regions into spatial coordinates that the robotic manipulator can use. (3) The AR visualization modules which are used to indicate the future actions of the robot to the user.

3.3.1 Computer Vision Module

This section starts by describing the procedure to register the autonomous patient side manipulator (PSM3) to the camera frame. Then, the neural network architecture used to segment blood accumulations in the endoscopic images is described. The former procedure allows transforming locations specified in pixels into 3D coordinates for the robot to reach. The latter procedure was used to automatically segment the blood pooling in the simulator and obtain the centroid locations of each. Finally, the augmented reality visualizations used to enhance human and robot collaboration are described.

3.3.1.1 Notation

First, we define all the required coordinate frames and transformations to develop the autonomous robotic assistant. These coordinates frames are depicted in figure 3.3. As a convention for this work, all transformation between frames are represented with homogeneous transformation matrices. Vectors are assumed to be column vectors, and the 3×3 identity matrix is represented with I_3. Last, we denote the transformation from frame A to frame B as ${}^{B}T_{A}$ where A or B could be any of the following coordinate frames:

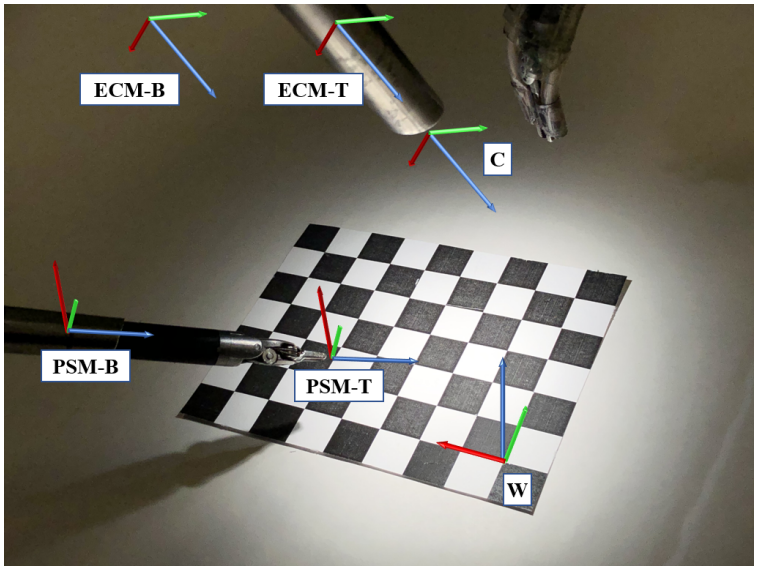


Figure 3.3. Relevant coordinate frames of the proposed system

1. C is the coordinate frame of the right camera of the endoscope located in the ECM.
2. ECM_T is the coordinate frame of the tool-tip of the ECM arm.
3. ECM_B is the coordinate frame of the base of the ECM.
4. $PSM3_T$ is the coordinate frame of the tool-tip of PSM3.
5. $PSM3_B$ is the coordinate frame of the base of PSM3.
6. W is an arbitrarily selected world coordinate system used during the registration process.

The transformation matrices ${}^{PSM3_B}T_{PSM3_T}$ and ${}^{ECM_B}T_{ECM_T}$ are provided by the python API of the DVRK robot. Additionally, the API dynamically updates these transformations as the robot moves.

3.3.1.2 Manipulator Registration to the Camera Frame

Manipulator registration to the camera frames requires finding the transformation between the PSM3 base frame to the camera frame. In other words, finding the transformation matrix ${}^{C}T_{PSM_B}$. This transformation is essential to plan trajectories for the manipulator and create the augmented reality visualizations. This transformation can be calculated by solving Equation 3.1, where ${}^{C}T_{ECM_T}$ and ${}^{ECM_B}T_{PSM_B}$ are unknowns.

$$ {}^{C}T_{PSM_B} = {}^{C}T_{ECM_T}\, {}^{ECM_T}T_{ECM_B}\, {}^{ECM_B}T_{PSM_B} \tag{3.1} $$

To calculate ${}^{C}T_{ECM_T}$, the intrinsic parameter matrices are computed first, for both cameras of the endoscope using a chessboard calibration pattern (chessboard in figure 3.3) and the OpenCV library. Then ${}^{C}T_{ECM_T}$ is calculated by constructing a system of equations of the form $AX = YB$, where A and B are a pair of measurable transformation matrices, and X and Y are fixed unknown transformations. In the literature, this equation is referred to as the hand-eye calibration problem. In our specific setting, this formulation takes the form of equation 3.2.

$$ {}^{W}T_{C}\, {}^{C}T_{ECM_T} = {}^{W}T_{ECM_B}\, {}^{ECM_B}T_{ECM_T} \tag{3.2} $$

Equation 3.2 was solved using the Kronecker Product formulation proposed by [95]. This formulation requires measuring ${}^{W}T_{C}$ and ${}^{W}T_{ECM_B}$ in different configurations of the endoscopic manipulator to estimate the unknown transformations. ${}^{W}T_{C}$ can be calculated with the intrinsic parameter matrices of the cameras and the calibration chessboard. The OpenCV function $solvePnP$ to identify the pose of the chessboard with respect to the camera frame is used. ${}^{W}T_{ECM_B}$ is given by the robot API.

To calculate ${}^{ECM_B}T_{PSM_B}$, the equation 3.3 is solved, where ${}^{W}T_{PSM3_B}$ is unknown. This transformation was calculated following the 3-point registration procedure described in [96]. Solving equation 3.3 completely specifies the transformation between the autonomous assistant base frame and the camera frame.

$$ {}^{ECM_B}T_{PSM_B} = {}^{ECM_B}T_{W}\,{}^{W}T_{PSM_B} \tag{3.3} $$

3.3.1.3 Semantic Segmentation of Endoscopic Images

To provide the autonomous manipulator with navigational information from the cameras, semantic segmentation architectures were used. Semantic segmentation means grouping pixels corresponding to the same object part, organ, or anatomic structure. This grouping can be done by assigning the same index to all the pixels that belong to the same group. These algorithms allow segmenting from the endoscopic images' relevant objects such as bleeding vessels, blood accumulations, and the robotic arms. Once the objects have been segmented, pixel locations are transformed into 3D coordinates in the robot frame using the calibration transformation matrices (see section 3.3.1.2). We collected a dataset of 180 images extracted from endoscopic videos while the robot was teleoperated in our surgical simulator to train the segmentation models. Two techniques were utilized to prevent overfitting of the models in our dataset. First, the backbone's weights of our FCN model were initialized with the weights of a VGG-16 pre-trained in the ImageNet dataset. Second, our dataset was augmented utilizing spatial and color-space data augmentations techniques.

Segmentation of the endoscopic images was done with fully convolutional networks (FCNs) [13]. These models are characterized by their lack of dense layers, meaning they are entirely

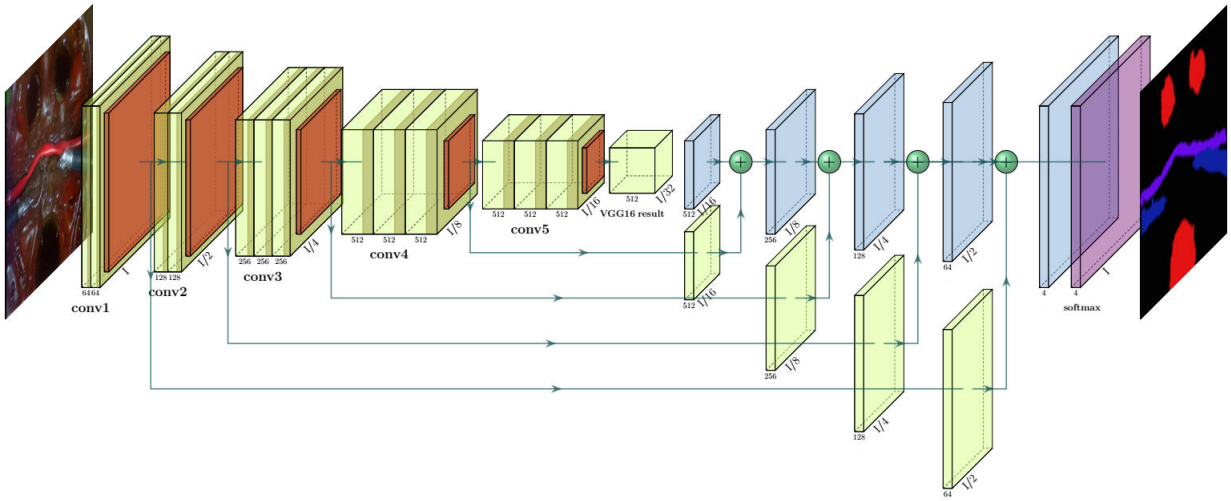


Figure 3.4. Proposed fully convolutional network with a VGG-16 backbone. The architecture uses the following color coding: (1) green blocks represent convolutional layers, (2) orange blocks represent maxpooling layers, (3) blue blocks represent upsampling layers and (4) purple layer represent softmax layers.

constructed from convolutional layers. Thus, any of the traditional classification architectures, e.g., InceptionNet, AlexNet, VGG, can be repurposed for semantic segmentation by removing the last dense layers. Another characteristic of FCNs is their use of transposed convolutions (also called fractionally-strided convolutional layers) to upsample the feature maps to the original resolution of the input image. As discussed in [13], upsampling operations are not able to accurately reconstruct fine spatial information. To address this, skip connections are used before the upsampling layers to merge activation maps from convolutional layers. Last, FCNs allow for variable shape inputs.

For this thesis, a fully convolutional variation of the VGG-16 architecture (see figure 3.4)[97] was utilized. This network is constructed with five downsampling blocks. Each block is composed of a combination of convolutional and max pooling layers. The former utilizes a kernel size of 3 and a zero-padding of 1 to preserve the input's resolution. The latter utilizes a kernel of size 2 and a stride of 2 to downsample the input by a factor of 2. As observed in table 3.1, the downsampling blocks reduce the image resolution by a factor of 32. Then, transposed convolutional layers were added to upsample the activation maps to the original resolutions. Skip connections were added between the downsampling and upsampling layers to maintain fine spatial details at the output.

Table 3.1. Semantic segmentation network downsampling specifications.

Block Descriptions		**Output Shape**	
Blocks	**Convolutional Layers**	**Input Resolution (H, W)**	**Input Resolution (384,480)**
1	2	(128, H/2, W/2)	(128, 192, 240)
2	2	(128, H/4, W/4)	(128, 96, 120)
3	3	(256, H/8, W/8)	(256, 48, 60)
4	3	(512, H/16, W/16)	(512, 24, 30)
5	3	(512, H/32, W/32)	(512, 12, 15)

A model with a dataset of 160 images of resolution (640x480) collected during a simulated surgical exercise was trained. These images were annotated using an online annotation tool

called SuperAnnotate [98]. Labels are matrices with the same width and height as the images, where every entry represents the class for a corresponding pixel in the image. This dataset had four different labels: blood, instrument, vessel, and background. A sample image and its corresponding label are shown in figure 3.5. The convolutional layers of the VGG-16 were initialized with pre-trained weights on the ImageNet dataset [45]. The model was trained using the RMSprop optimizer for 250 epochs and a learning rate of 1×10^{-4}. Last, a scheduler was used to decrease the learning rate by a factor of 0.5 every 50 epochs to improve training stability.

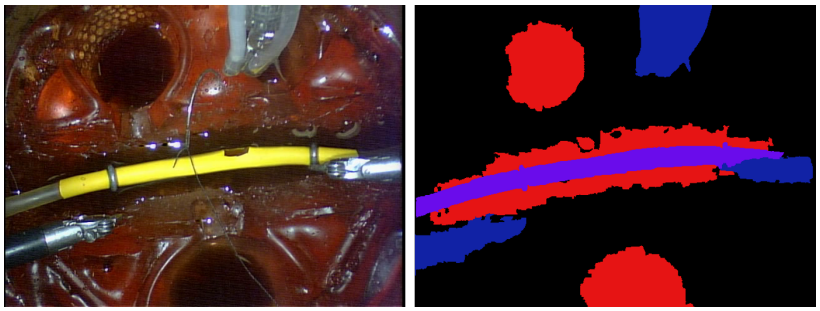


Figure 3.5. Example of semantic segmentation image and label. The following color code is used for the labels: (1) red pixels correspond to blood poolings, (2) purple pixels correspond to the simulated vessel, and (3) blue pixels correspond to the instrument arms.

Overfitting small training datasets is a well-known problem of deep learning architectures [99]. Deep models with many convolutional layers, such as the VGG-16, have a high learning capacity. Thus, when not enough training data is provided, the model can memorize the training dataset rather than learning a general representation of the data. As creating large-scale datasets is costly and time-consuming, data augmentation techniques have been proposed as an alternative. Data augmentation refers to a set of transformations that preserve the labels and create more variability in the data. For this thesis, we leverage four basic techniques: (1) random cropping, (2) random flipping, (3) color space transformations, and (4) noise injection techniques. Random cropping involved extracting a rectangular section of size (384x480) from the full image. This technique allows the model to better segment partially shown objects. Flipping involved mirroring the images along the vertical or horizontal axis. Color space transformations involved channel shuffling in RGB color space and

multiplying hue and saturation in HSV color space. Noise injection involved adding salt and pepper noise and additive white gaussian noise. All transformations were applied using the Imgaug python library [100]. The first row of figure 3.6 shows 4 images from our dataset. Rows 2 and 3 show two different versions of the original images obtained by applying in random order color space transformations and noise injection techniques.

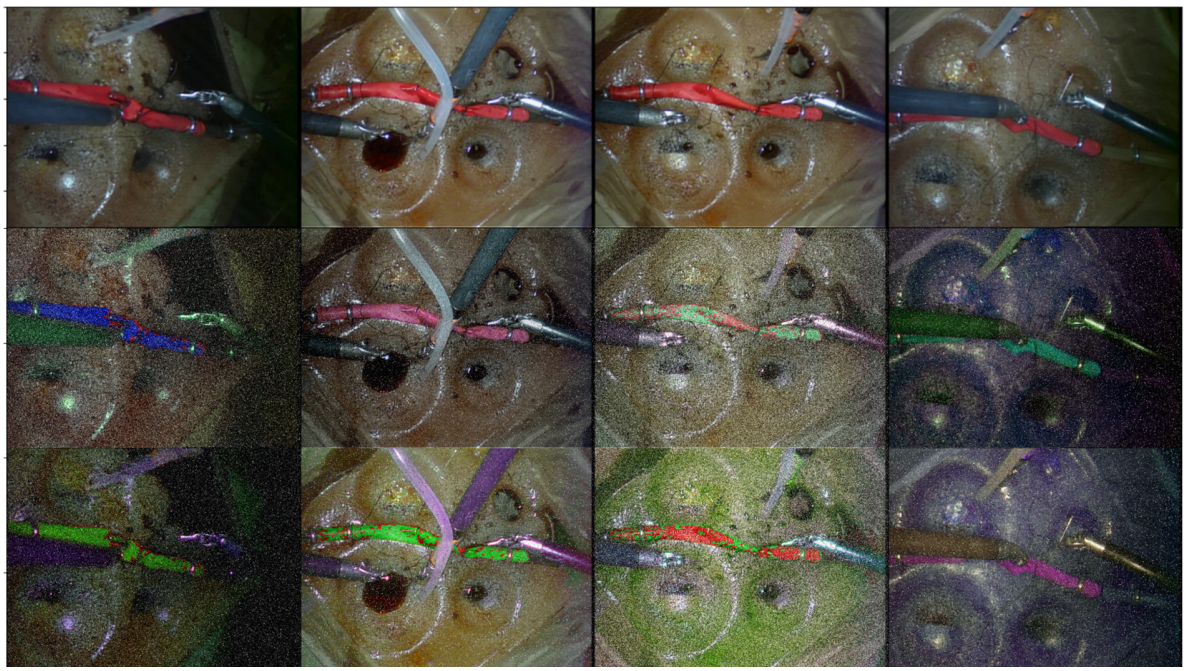


Figure 3.6. Examples of augmented images in the dataset

3.3.2 Robotic module

This section describes the integration between the computer vision algorithms and the robot to perform the autonomous suction of blood. After segmenting the endoscopic images with the proposed FCN model, the centroids and the area of the blood blobs are calculated. The pixel coordinates corresponding to the blob with the biggest area were selected as the next target of the robotic assistant. These target pixel coordinates were then transformed using a homography into spatial coordinates for the surgical robot. Last, a straight-line trajectory from the current position to the calculated target position was calculated and executed by the robot. The centroids of the accumulations were selected as the target

location to maximize the amount of blood that the robot could draw in a single retraction motion.

First, a calibration procedure was used to calculate the homography transformation, H, that mapped the pixel coordinates $[u, v]$ of the biggest blob into the x, y position of the end-effector (see equation 3.4).

$$\begin{bmatrix} x\prime \\ y\prime \\ w \end{bmatrix} = H_{3\times 3} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{3.4}$$

This matrix was estimated using the linear least-squares function in the OpenCV library [101]. The point correspondences, $(x_{\mathrm{i}}, y_{\mathrm{i}})$ and $(u_{\mathrm{i}}, v_{\mathrm{i}})$, for this function were obtained through a calibration procedure where the end-effector was manually moved to six different landmarks in the simulator (see figure 3.7). The calculation of the z coordinate was simplified by assuming a constant height. However, a more accurate solution would be to include force feedback to avoid tissue damage due to excessive force. The final step to obtaining the end-effector position vector is dividing the transformed $[x\prime, y\prime]$ by the w component as shown in equation 3.5 to obtain the in-homogeneous representation of the coordinates

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x\prime/w \\ y\prime/w \\ z \end{bmatrix} \tag{3.5}$$

The system computed the centroids and area of all blood pooling regions appearing in the segmentation maps on deployment. The centroids and the area were calculated using equation 3.6 and 3.7 respectively.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} M_{10}/M_{00} \\ M_{01}/M_{00} \end{bmatrix} \tag{3.6}$$

$$A = M_{00} \tag{3.7}$$

Figure 3.7. Calibration markers used to calculate homography transformation.

where M_{01}, M_{10}, and M_{00} are image moments. Defining $\hat{A}$ as the array containing the area of all the contours, the centroid $[u_{\mathrm{i}}, v_{\mathrm{i}}]$ was selected as the next target where $\mathrm{i} = argmax\{\hat{A}\}$ (the contour with the biggest area). Then, based on the previously obtained homography and the target position specified in pixels, the end-effector target position in world coordinates was estimated. Finally, the suction tool followed a straight-line trajectory from the predefined base position to the target position. After suction was completed, the robotic arm returned to the base position, following another straight-line trajectory.

3.3.3 Augmented Reality Visualization

This section describes the augmented reality (AR) module used in the GUI at the surgeon side. Such visualization was used to enhance the interaction between the surgeon and the autonomous suction manipulator. Effective human-robot collaboration requires the surgeon to anticipate the actions of the robotic assistant. In this regard, AR technology is an attractive alternative to provide visual cues about the future robot actions. The AR visualizations can be displayed in the surgical console, thus our module does not require any additional hardware. Furthermore, no markers were required in the surgical field as the API provides continuous kinematic information of the robot manipulators. The goal of this module was to provide transparent visuals that align with the initial robot pose and then updates its

location to display the immediate robot the trajectory (see figure 3.8). These visuals allowed the user to know what to expect when the robot starts moving.

Drawing the AR visuals requires establishing a correspondence between the world 3D coordinates and the image plane 2D pixels. Figure 3.9 shows all the relevant frames involved in this process. This mapping was achieved by representing the camera with the pin-hole model and calculating its projection matrix. After finding this matrix, the relation between 3D coordinates in the camera frame and the pixel coordinates was established through equation 3.8. Coordinates are written in their homogeneous representation

$$\overrightarrow{x} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}^{(c)} \tag{3.8}$$

This equation was further simplified to 3.9 where K represents the intrinsic camera parameter matrix. The matrix of dimensions 3×4 performing the mapping is referred to in the literature as the projection matrix.

$$\overrightarrow{x} = K \begin{bmatrix} I_3 & b \end{bmatrix} \overrightarrow{X}^{(C)} \tag{3.9}$$

The projection equation was modified to include the transformation matrix between the PSM-B and the camera frame, ${}^{C}T_{PSM_B}$ (see section 3.3.1.2). The final results is shown in equation 3.10.

$$\overrightarrow{x} = K \begin{bmatrix} I_3 & \overrightarrow{0} \end{bmatrix} {}^{C}T_{PSM_B} \overrightarrow{X}^{(PSM_B)} \tag{3.10}$$

Using the previous projection matrix, a transparent cuboid was drawn using OpenCV computer graphic functions to indicate the autonomous PSM's tool-tip location and orientation. The coordinates of the cuboid's faces were calculated in the PSM_B frame and then projected them with equation 3.10 into pixels. The cuboid axes correspond to the PSM_T coordinate frame. Thus it was obtained from the transformation matrix ${}^{PSM_B}T_{PSM_T}$ provided by the robot API. As shown by equation 3.11 the pose information of the tooltip was

encoded in a rotation matrix R and a translation vector t. Projecting $\overrightarrow{r_x}$, $\overrightarrow{r_y}$,$\overrightarrow{r_z}$, and $\overrightarrow{t}$ to the image plane provided all the necessary information to draw the AR visuals. The annotations were drawn in the initial position of the manipulator and then dynamically updated to match positions of the planned trajectory.

$$ {}^{PSM_B}T_{PSM_T} = \begin{bmatrix} R & \overrightarrow{t} \\ \overrightarrow{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \overrightarrow{r_x} & \overrightarrow{r_y} & \overrightarrow{r_z} & \overrightarrow{t} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.11} $$

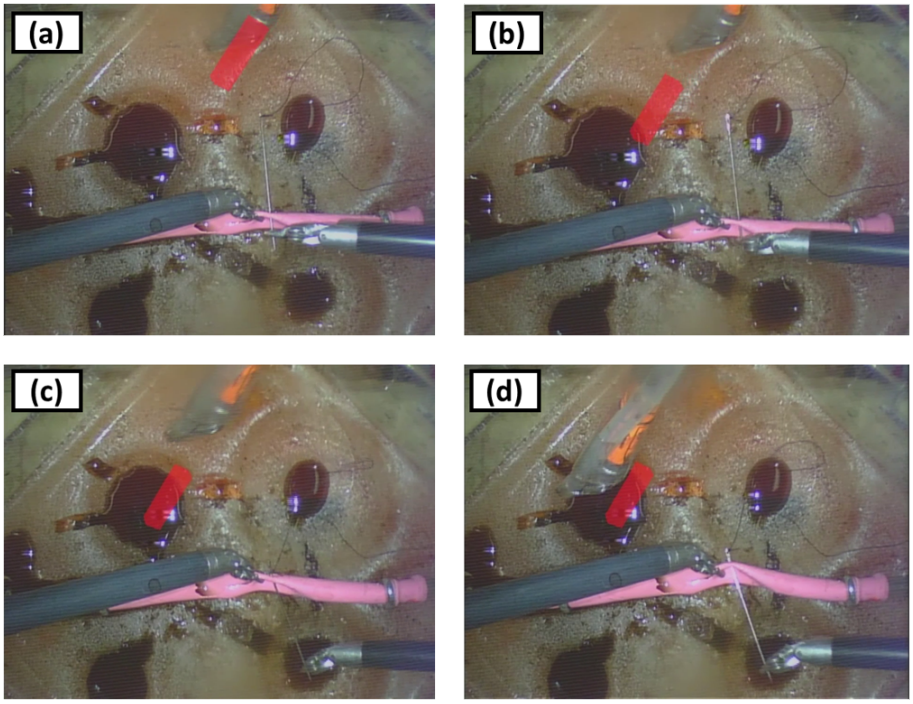


Figure 3.8. Augmented reality result. (a) Transparent texture aligns with the autonomous manipulator. (b)-(c) Texture moves to indicate the robot intended motion. (d) After the texture stops, the manipulator starts its motion.

3.4 Cognitive Workload Sensing Module

This section describes a complementary cognitive workload sensing module used to adapt the autonomy behavior based on the user's mental workload. The goal with this module was to construct a real-time binary classifier to recognize states of *low-cognitive* and *high-cognitive* load from EEG and eye tracker signals. First, we describe the hardware used to measure

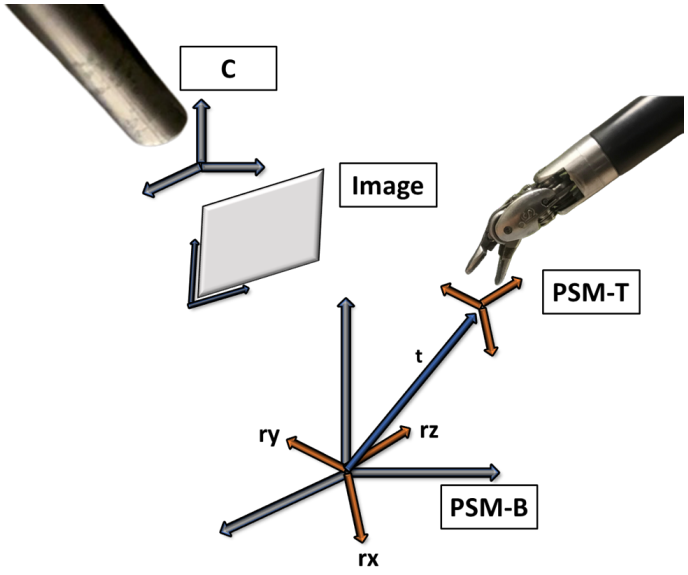


Figure 3.9. Augmented reality implementation diagram

EEG. Second, we detail the steps to extract frequency features from the raw signals. Third, we explain the learning algorithm used for the classification task. Two different learning algorithms are proposed for the cognitive estimation problem: a recurrent network used for workload classification from EEG features and a feed-forward network used for workload classification of combined features from both sensors. Finally, a framework for real-time prediction is presented.

3.4.1 Physiological Sensors

This subsection describes the two physiological sensors used to obtain unobtrusive and continuous physiological measurements. The two sensors are an EEG headset and eye-tracker glasses.

3.4.1.1 EEG Sensing

EEG recordings were made with a 32 channel G.Nautilus with active electrodes (Gel-based) from G.tec medical engineering GmbH, Austria. The electrodes' location in the users' head can be observed in figure 3.10. The data was recorded at 250Hz. Additionally, a band-

pass and notch filter were respectively applied between 0.5 and 30hz, and 58hz and 62hz with the proprietary g.NEEDaccess python client from G.tec [102]. All the configuration parameters for the device were chosen based on previous studies with the G.Nautilus [102], [103]. The channel AFZ is the device's ground, and the reference is the right earlobe. The preprocessing steps were minimized to allow easy translation to real-time scenarios.

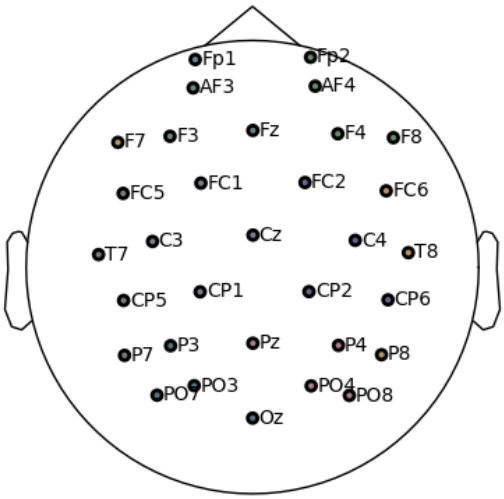


Figure 3.10. EEG channel locations

3.4.1.2 Eye Tracker Based Sensing

Eye movements were recorded with a Tobii Pro Glasses 2.0 (Tobii Technology AB, Danderyd, Sweden). This device has a pair of inner cameras that precisely track the eye movements and the user's pupil diameter. This sensor provided 2D and 3D gaze positions and the pupil diameter of both eyes at a sampling rate of 60Hz. No further preprocessing steps were done to the features provided by the sensor.

3.4.2 Feature Extraction

This subsection describes the low dimensional features extracted from the raw signals to train the cognitive workload classifiers.

3.4.2.1 EEG Frequency Features

Previous studies [92], [104], [105] have shown that EEG spectral features such as the band power coefficients are correlated to the cognitive workload. Therefore, this representation was used to train our recurrent model. Figure 3.11 illustrates the pipeline to transform EEG signals into a temporal sequence of band power feature vectors.

First, EEG signals were obtained from each trial into overlapping epochs. The epoch size was selected depending on the classification modality, i.e., 5 seconds for classification based on EEG and 15 seconds for multi-sensor classification. Each epoch was represented as a matrix $A^{(t)}$ with 32 rows, each one corresponding to a channel of the EEG headset, and n columns, corresponding to the number of data points in each epoch. Thus, the entry $a_{\mathrm{i,j}}$ represents the j^{th} raw sample of the i^{th} channel. Then, the power spectral density (PSD) was calculated for every row in matrix $A^{(t)}$ using Welch's method with a sliding window of 4 seconds. The resulting PSDs were concatenated into a new matrix $S^{(t)}$.

Finally, band power coefficients were calculated from the channels' PSDs. These are summary statistics that indicate the energy contribution of specific frequency bands. Following the brain theory of neural oscillations [93] coefficients were extracted from the following frequency ranges: the theta band (4-8Hz), the alpha band (8-12Hz), and the gamma band (12-30Hz). Additionally, the coefficients were normalized to the [0-1] range by dividing the band values by the total energy of the signal. This process resulted in the normalized feature vector $P^{(t)}$ of 96 elements (32 channels x 3 coefficients).

To use these features as an input for a recurrent model, a sliding window was used to group consecutive feature vectors into a matrix $G^{(k)}$ of dimensions $(96 \times L)$, where L is the size of the temporal sequence. The goal of this module was to classify the matrix $G^{(k)}$ as either *low-cognitive* or *high-cognitive*.For the workload classification based on eye-tracker and EEG, the average of each power coefficient across all channels was calculated. The reason for this operation was obtaining a similar number of features from the EEG and the eye tracker. After averaging, the final feature vector contained 3 coefficients for each epoch of raw data. These three features were concatenated at the end of the eye-tracker features to train the cognitive load classifier.

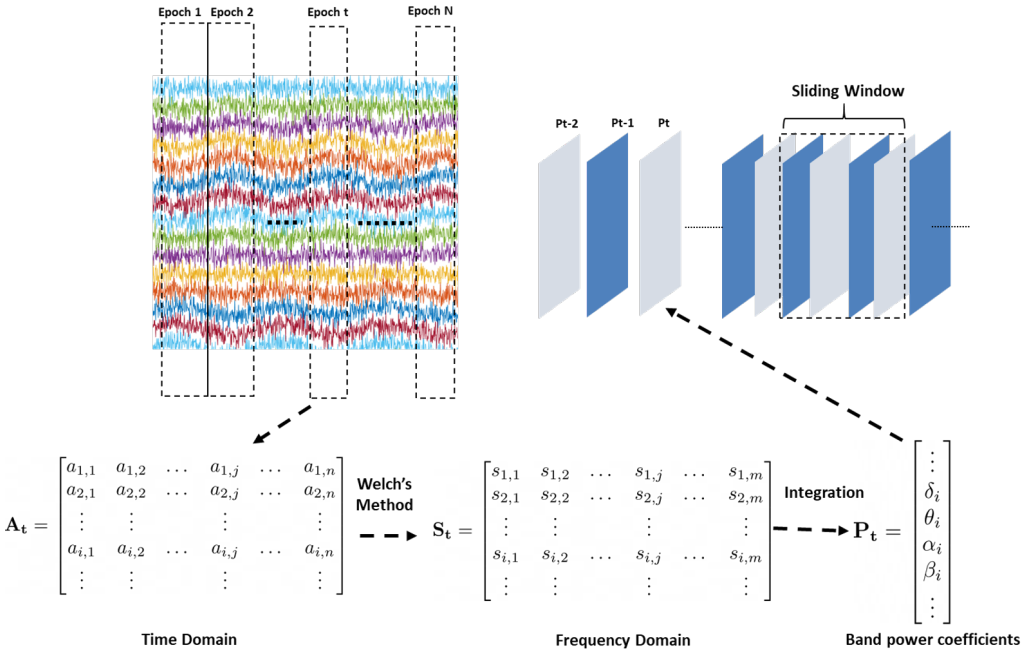


Figure 3.11. Feature extraction pipeline for EEG signals

3.4.2.2 Eye Tracker Features

The eye tracker provided with pupil diameter and 2D gaze position at a sampling rate of 60 Hz. First, the raw signals were divided into 15 seconds epochs to extract 5 features related to workload: (1) average pupil diameter $\hat{PD}$, (2) number of fixations (NF), (3) average fixation time($\hat{FT}$), (4) scan path length (SSP), and (5) nearest neighbor index (NNI). The following metrics were selected since they have been previously related to mental workload demands [106]. Fifteen seconds was the minimum epoch size to calculate the eye tracker features based on fixations. First, the average left eye pupil diameter $\hat{PD}$ was calculated for each window using the following equation

$$\hat{PD} = \frac{1}{N}\sum_{\mathrm{i}=0}^{N} PD_{\mathrm{i}} \tag{3.12}$$

To calculate the remaining features, the 2D gaze was transformed into a sequence of fixations points. A fixation was defined as the period where the gaze is relatively stationary [107]. Each fixation was associated with a corresponding duration in milliseconds. In this

experiment, only fixations of at least 85ms were considered. After obtaining the fixations of each window, the total number of fixations (NF) and the average fixation time ($\hat{FT}$) was calculated. Then, the scan path length was computed as the total Euclidean distance between consecutive fixations with the following equation

$$SSP = \sum_{\mathrm{i}=2}^{NF} d(f_{\mathrm{i}-1}, f_{\mathrm{i}}) \tag{3.13}$$

where d is the Euclidean distance operator and f_{i} is the $\mathrm{i}^t h$ fixation. Last, the nearest neighbor index was calculated as the ratio of the nearest neighbor distance of fixations $d(NN)$ and the average distance of a randomly distributed set of fixations $d(ran)$. The nearest neighbor distance, $d(NN)$, was calculated by applying equation 3.14 to the fixations set [106]. $d(ran)$ was calculated by applying 3.14 to a randomly generated set of fixations.

$$d(NN) = \sum_{\mathrm{i}=1}^{NF} \left[\frac{min(d_{\mathrm{ij}})}{N} \right], 1 < \mathrm{j} < NF \tag{3.14}$$

3.4.2.3 Feature Synchronization

Features from the EEG and eye tracker from 15 seconds epochs were extracted. For training purposes, a 50 percent overlap was used among consecutive epochs. For real-time deployment, an overlap of 14 seconds was used to obtain workload predictions at one Hz (see figure 3.12). Synchronization and recording of the signals were done with the Lab-StreamingLayer (LSL) [108]. The goal of this cognitive workload module was to classify the concatenated features from the EEG and the eye tracker as either *low-cognitive* or *high-cognitive*.

3.4.3 Cognitive Workload Classification Models

Two different learning models were proposed for the problem of cognitive workload prediction: a recurrent neural network used for the classification of EEG features and a feed-forward network for the combination of the eye tracker and EEG. The reason for these two models is that eye tracker features can only be calculated from an epoch at least 15 seconds,

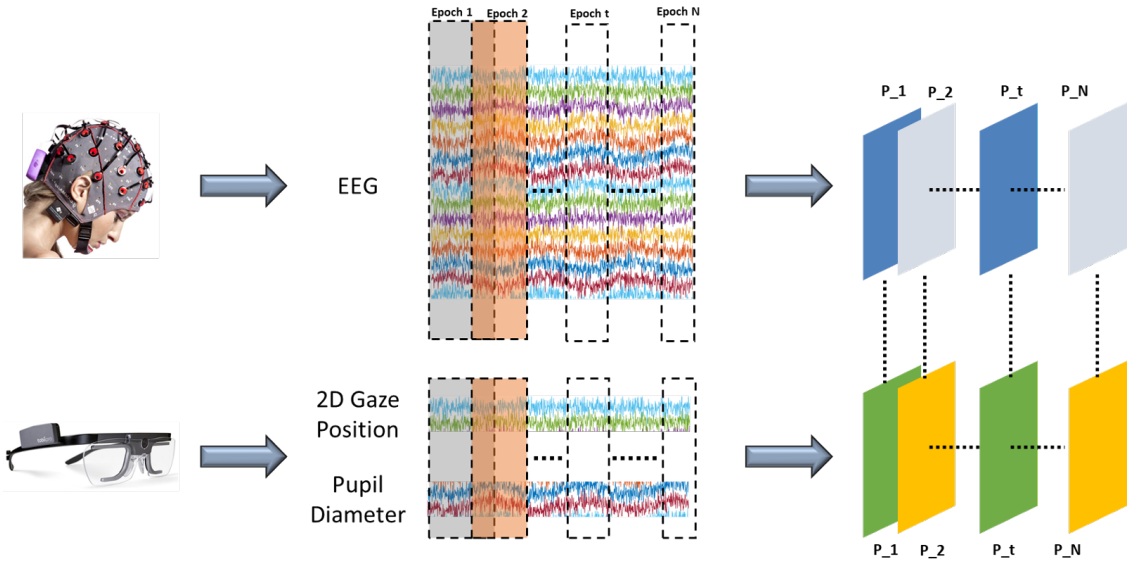


Figure 3.12. Diagram showing how the EEG and Eye tracker signals are synchronized for the cognitive load detection system.

while EEG spectral features can be calculated from much smaller window sizes . Therefore, using a recurrent model with eye tracker features would result in prediction delays.

3.4.3.1 Recurrent Architecture for EEG Classification

Classification of spectral feature vectors from the EEG was achieved with a compact architecture based on gated recurrent neural networks. This model was selected because it models the long-term dependencies of cognitive states. The architecture consisted of two stages and was implemented in the Keras framework [109]. In the first stage, an affine transformation was applied to the input at every time-step with equation 3.15. Thus transforming the feature vector $P^{(t)}$ into a smaller version $P'^{(t)}$. In the second stage, multiple time steps were combined with two bidirectional Long Short-Term Memory (LSTM) layers to predict the final workload.

$$P'^{(t)} = \sigma(\mathbf{b} + \mathbf{U}P^{(t)}) \tag{3.15}$$

LSTM networks were first introduced by Hochreiter et al. [110] to analyze long-term dependencies in sequences and as a solution to the problem of vanishing or exploding gradients. The basic LSTM cell (see figure 3.13) uses a combination of gated mechanisms to control the flow of information and gradients through time. The four gated mechanisms are (1) the forget gate $f^{(t)}$, (2) the output gate $q^{(t)}$, (3) the input gate $\mathrm{i}^{(t)}$, and (4) the external input gate $g^{(t)}$. The forward propagation equations of each gate are shown in 3.16

$$\begin{aligned} f^{(t)} &= \sigma(\mathbf{b_f} + \mathbf{U_f}P'^{(t)} + \mathbf{W_f}h^{(t-1)}) \\ \mathrm{i}^{(t)} &= \sigma(\mathbf{b_i} + \mathbf{U_i}P'^{(t)} + \mathbf{W_i}h^{(t-1)}) \\ g^{(t)} &= \sigma(\mathbf{b_g} + \mathbf{U_g}P'^{(t)} + \mathbf{W_g}h^{(t-1)}) \\ q^{(t)} &= \sigma(\mathbf{b_q} + \mathbf{U_q}P'^{(t)} + \mathbf{W_q}h^{(t-1)}) \end{aligned} \tag{3.16}$$

where U_f, U_q, U_{i}, U_g, W_f, W_q, W_{i} and W_g are learnable parameter matrices, and $P'^{(t)}$ and $h^{(t-1)}$ are respectively the input features and the hidden state vector.

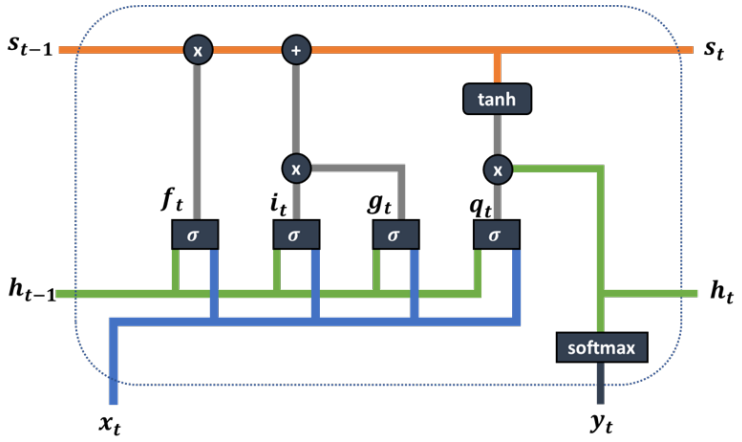


Figure 3.13. LSTM cell inner mechanisms.

Another key characteristic of the LSTM cell is a recurrent inner state $S^{(t)}$ that is directly controlled by the forget gate. Lastly, $h^{(t)}$ and $y^{(t)}$ are respectively the hidden state and the output vector. The former is used to control the gated mechanisms of succeeding cells in time. The latter provides a workload prediction for every time-step t. Equations for $s^{(t)}$, $h^{(t)}$ and $y^{(t)}$ are shown in 3.17.

$$\begin{aligned} S^{(t)} &= f^{(t)} \odot S^{(t-1)} + g^{(t)} \odot \mathrm{i}^{(t)} \\ h^{(t)} &= tanh(S^{(t)}) \odot q^{(t)} \\ y^{(t)} &= softmax(h^{(t)}) \end{aligned} \tag{3.17}$$

To train the described model, a binary cross-entropy loss function shown in equation 3.18 was used where L_{i} is the label for the i^{th} input sequence, N is the number of training samples, and $y_{\mathrm{i}}^{(T)}$ is the output of the last LSTM cell. To optimize the model architecture, a grid search cross-validation of the following hyperparameters was conducted: (1) number of LSTM layers, (2) connectivity of the recurrent layers, i.e., unidirectional or bidirectional,

and (3) dropout rates. The dropout rate adopted was from 0.5 to 0.45 with step size 1 and the number of LSTM layers from 1 to 3 with a step size of 1.

$$Loss = -\frac{1}{N}\sum_{\mathrm{i}=0}^{N} L_{\mathrm{i}} \cdot log\left(y_{\mathrm{i}}^{(T)}\right) + (1 - L_{\mathrm{i}}) \cdot log\left(1 - y_{\mathrm{i}}^{(T)}\right) \tag{3.18}$$

3.4.3.2 Feed Forward Model for Multi Sensor Classification

To classify the physiological features from multiple sensors, a feed-forward neural network was used. The network used only a combination of fully connected layers with Relu activation functions, dropout, and batch normalization layers. To train the described model, the binary cross-entropy function was used (see equation 3.18 and the Adam [111] optimization algorithm. Models were trained for 100 epochs using a batch size of 10 samples.

To optimize the model architecture, a grid search cross-validation of the following hyper-parameters was conducted: dropout rate and the number of hidden layers. The dropout rate from 0.5 to 0.45 was tested with step size one and the number of hidden layers from 4 to 8 with a step size of 1.

3.4.4 Real-time Architecture and Deployment

On deployment, the cognitive workload module receives the extracted features from the EEG and eye tracker in real-time. Then, the prediction module utilizes these features to generate a prediction score in the range $[0, 1]$ representing the likelihood that the user is in a state high cognitive load. This prediction was produced at a sampling rate of 1 Hz. If the score exceeded a previously predefined threshold, an activation trigger was sent to the autonomous system to assist the user. After finishing the suction of the blood, the autonomous system went back to a base position until the next cognitive trigger was received. The integration diagram of the system is presented in figure 3.14

The complete system allowed operators to perform the surgical task while the robotic blood suction assistant assisted during the periods of high cognitive load. This way the method presented allowed assistance to the main surgeon without taking complete autonomy of the task.

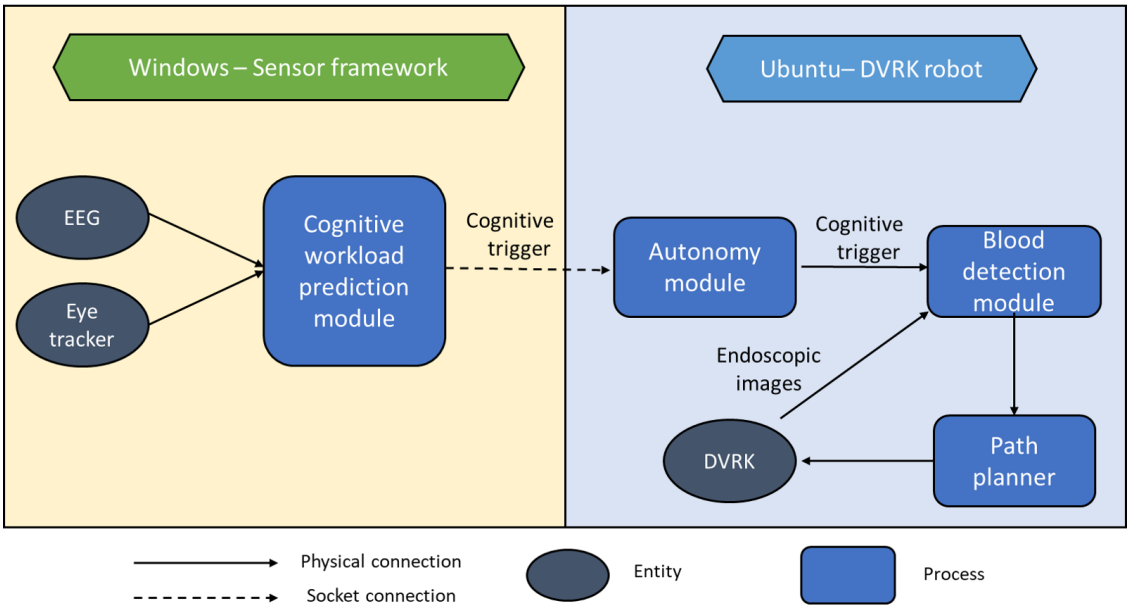


Figure 3.14. Cognitive load and autonomous module integration.

3.5 Summary

This chapter described all the components of the proposed autonomous suction assistant. First, we present the specifications of the DVRK robot used to deploy our system and a developed surgical simulator. Then, we delve into the specifications of the perception algorithms. This section included a description of the registration process of the autonomous manipulator to the endoscopic camera, the proposed semantic segmentation algorithm, and the augmented reality annotations module. Then, we describe how the integration between the perception module and the robot results in autonomous behavior. Finally, we present a complementary cognitive workload sensing module.

4. EXPERIMENTS AND RESULTS

The following chapter presents three experiments to evaluate the algorithms and systems proposed in the previous chapter. First, an experiment was designed to assess the performance and workload of the autonomous suction assistant in simulated surgical exercise. Second, experiment 2 focuses on evaluating the accuracy of our proposed cognitive workload sensing module. Last, experiment 3 evaluates the integration of the cognitive sensing module and autonomous algorithms.

4.1 Experiment 1: Evaluating Autonomous Suction Action on Users' Performance

The goal of experiment 1 was to evaluate the effectiveness of the proposed autonomous system in providing suction assistance. To achieve this, the users' performance and perceived workload were recorded under two different modalities: manually-teleoperated suction action(MS) and autonomous-suction action by the robotic assistant (AS). In MS, the users manually teleoperated the suction tool to facilitate their task. In AS, the autonomous robotic assistant controlled the suction tool. The underlying hypothesis (**H1**) is that the autonomous suction action (AS) will lead to better performance and lower mental demands than the manual teleoperationMS. This claim is supported by the fact that by having the autonomous system taking over the additional demands of bleeding events, the user can pay more attention to the primary task.

A within-subject experiment was designed to minimize the effect of individual differences. In other words, each user did the experimental task under both conditions. Finally, the order of the experimental conditions was randomized to account for possible learning effects. In other words, half of the users were randomly selected to start with AS modality and the remaining half to start with the MS modality. In this experiment, the autonomous system was configured to initiate a suction immediately after detecting any blood pooling event. Submersible motors were configured to pump blood into the simulator every 60 seconds to slowly fill up the simulator cavities as the user worked on the main task.

The rest of the subsections are divided as follows. First, the study protocol and the experimental task are described. Then, the objective and subjective metrics used in the study are explained. Finally, the results and discussion are presented. Part of the content of this section was published by Barragan *et al.* in [112] and reported here with modifications.

4.1.1 User Study Protocol

To achieve experiment's goal, a user study was conducted with eight graduate students. This study was approved by the Purdue University's Review Board under the protocol IRB-2021-22. For our study, users performed two surgical tasks sequentially: a running suture exercise and search and retrieve exercise. These tasks were selected from the Fundamentals of Robotic Surgery [14]. On the first task, users were required to do a 3-stitch running suture on the vessel of the simulator using a needle and a thread. On the second task, users had to search and retrieve 3 metallic pins that were scattered and attached to the ballistic gel (see figure 4.1). These tasks were selected as their difficulty increased with the number of blood accumulations events.

The experiment was divided into two sessions conducted on two subsequent. In the first session, users were informed about the experiment and signed a consent form. Then, users were familiarized with the teleoperation of the DVRK robot, i.e., moving the patient side manipulators and endoscopic camera, and using the clutch to swap controls between the main instruments and suction manipulator. The PSM1 and PSM2 (main patient side manipulators) were equipped with needle-drivers and the PSM3 with our custom-made suction tool (see figure 4.1). Afterward, users were required to train for 1 hour to become familiar with the selected surgical tasks. In this session, no simulated bleeding events were presented to the user.

In the second session, users were asked to complete the previously practiced tasks under the two experimental conditions; the *MS* and *AS* modalities. For this session, the bleeding functionality was activated while the users performed the tasks, i.e., the cavities of the simulator were slowly filled with blood. For the *MS* modality, users were instructed to use the suction tool in a way that would facilitate the completion of the main task. For the *AS*

modality, the suction tool was controlled by the autonomous algorithm. After completing each task, the user was asked to answer a NASA-TLX (National Aeronautics and Space Administration Task Load Index) questionnaire. During the experiment, kinematic data and the endoscopic video were recorded from the robot to assess the user's performance in each modality.

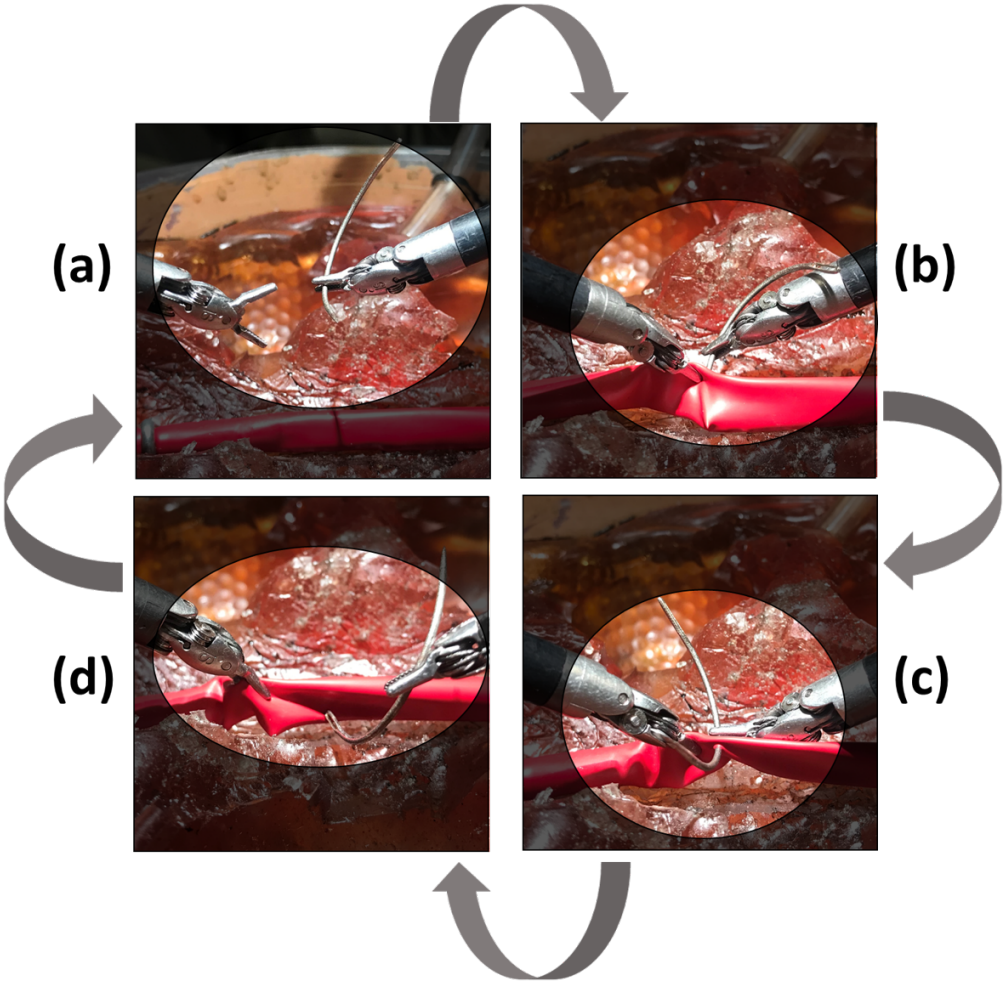


Figure 4.1. Sequence of steps to perform a running suture task. (a) Grasp the needle. (b) Position the needle close to the vessel. (c) Insert the needle. (d) retrieve the needle.

4.1.2 Evaluation Metrics and Statistical Analysis

To evaluate our system, a combination of objective performance and subjective workload metrics were used. The goal of the objective metrics was to assess whether the users'

surgical performance improved when working with the autonomous system compared to the direct teleoperation of the suction tool. Subjective metrics were used to evaluate the users' perceived workload on both conditions. Each subject completed the surgical tasks using both modalities, manual and autonomous. Thus, a paired t-test was used to evaluate the differences in performance and workload metrics. P-values below 0.05 were assumed to be statistically significant in this study.

4.1.2.1 Objective Evaluation Metrics

To evaluate the user's performance, we calculated objective metrics from the robot kinematic and video data. These metrics were grouped into four different categories: (1) time, (2) motion, (3) system events, and (4) blood metrics. On the time metrics, the completion and the total clutching time of each experimental condition were measured. On the motion metrics, the average velocity of the main robotic manipulators (PSM1 and PSM2) and the suction tool manipulator (PSM3) were calculated. System events accounted for the number of clutching and tool-changing events. Last, the average number of pixels corresponding to blood regions, which is a proxy for the volume of blood accumulated, was calculated. This calculation was accomplished by running our semantic segmentation algorithm on the recorded videos of each experimental condition. All the included metrics have been previously used to evaluate surgical expertise in clinical settings [113], [114].

4.1.2.2 Subjective Evaluation Metrics

To quantify workload demands, the subjects were asked to fill a NASA-TLX questionnaire [115] after completing each task. The NASA-TLX divides the workload demands into six components: (1) Mental Demand, (2) Physical Demand, (3) Temporal Demand, (4) Performance, (5) Effort, and (6) Frustration. A 0-10 score was assigned to each of them, where 0 is associated with low demand and 10 with high demand. A final workload score ranging from 0 to 60 was calculated by summing up the different components of the questionnaire.

Table 4.1. Objective performance metrics results. Statistically significant results (α=0.05) were highlighted in bold.

Type	***Name***	***Mean (std), N=8***		***T-test***	
		Autonomous	**Manual**	**T-statistic**	**p-value**
Time	completion time (s)	249.91 (87.68)	411.42 (210.64)	-3.133	**0.017**
	clutching time (s)	2.66 (3.54)	27.12 (31.18)	-2.328	0.053
Motion	psm1 velocity (cm/s)	1.1 (0.38)	0.79 (0.19)	3.250	**0.014**
	psm2 velocity (cm/s)	1.03 (0.43)	0.86 (0.39)	2.237	0.060
	psm3 velocity (cm/s)	1.21 (0.02)	0.27 (0.16)	18.283	0.000
Events	clutching events	0.88 (1.13)	11.5 (7.41)	-4.418	**0.003**
	tool changing events	0 (0)	2.75 (1.98)	-3.924	**0.006**
Blood	percentage of blood	0.11 (0.01)	0.18 (0.06)	-3.030	**0.019**

Table 4.2. Nasa-tlx results. Statistically significant results (α=0.05) were highlighted in bold.

Type	***Name***	***Mean (std), N=8***		***T-test***	
		Autonomous	**Manual**	**T-statistic**	**p-value**
NASA-TLX	Mental demand	3.62 (2.39)	4.81 (1.93)	-1.584	0.157
	Physical demand	2.69 (2.02)	3.88 (2.42)	-1.380	0.210
	Temporal demand	2.00 (1.28)	4.06 (2.53)	-2.855	**0.025**
	Performance	2.38 (1.36)	3.81 (1.25)	-2.959	**0.021**
	Effort	3.69 (2.17)	4.94 (1.70)	-1.909	0.098
	Frustration	2.50 (1.63)	4.31 (2.40)	-2.457	**0.044**
	Workload Score	16.88 (9.16)	25.81 (9.58)	-2.411	**0.047**

4.1.3 Results

Summary statistics, i.e., mean and standard, deviation were calculated for each performance metric and were shown in the Table 4.1. Statistically significant differences were found between the manual and autonomous modalities for all metrics except the clutching time and the velocity of the PSM2 arm. When the autonomous modality for the blood suction was used (249.91 $\pm$ 87.68s), users finished the task 162s faster than with the manual modality (411.425 $\pm$ 210.64s). The average speed of the main instrument arm (PSM1) was 0.3cm/s higher in the autonomous modality. In terms of the system events, users switched tools an average of 2.75 times while they used 11.5 clutching events when they were in the manual condition. In the autonomous modality, no tool changes were performed, and the

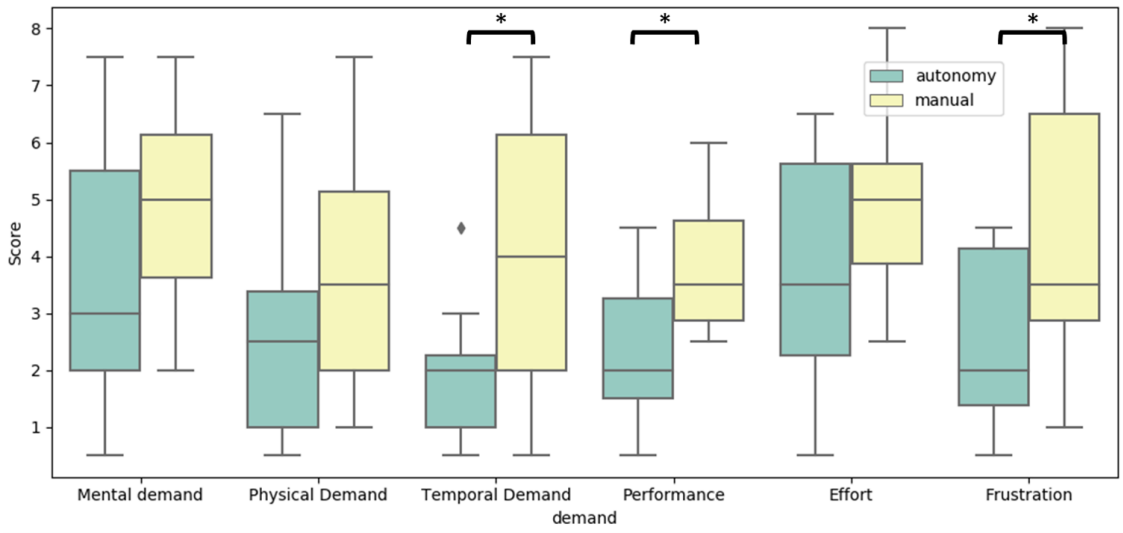


Figure 4.2. Box plots of NASA TLX workload components. Overhead asterisk indicates statistically significant differences between the two modalities.

number of clutching events was significantly reduced. Finally, the average percentage of blood detected in the video frames was reduced from 18% in the manual condition to 11% in the autonomous modality. This indicates that the autonomous modality controlled the bleeding more effectively.

Table 4.2 shows the self-assessed workload results. The mean of all the NASA-TLX components was found to be lower for the autonomous modality. However, it was only statistically significant for the performance (p=*0.021*), the temporal demand (p=*0.025*), and the frustration (p=*0.044*) indicators. Last, the mean workload score was significantly lower (p=0.047) for the autonomous modality, indicating the effectiveness of the robotic assistant. These results can be attributed to the way that the autonomous modality prevented cavities from flooding – that is only when the regions had a significant blood volume that would occlude the user's view.

4.1.4 Discussions

The results of the experiment indicate better performance and lower mental demands in *AS* (autonomous suction) than in the *MS* (manual suction). Users in the *AS*, completed the task 162 faster than in *MS*. This result indicates that switching between the main task and

the suction is a time-consuming effort that affects the overall efficiency of the procedure. Moreover, it indicates that the autonomous system effectively handles the bleeding events such that they do not affect the user's performance. Faster completion times on the *AS* modality can also be attributed to the higher average speed of the main instrument arms. In terms of system events, no tool changes were performed, and the number of clutching events was significantly reduced in the *AS*. These results imply that the autonomous assistant led to a more compact workspace at the console side, i.e., the total displacement of the console manipulators, which is a desirable quality in robotic-assisted procedures [75]. The size of the console workspace is increased due to the additional clutching of the console manipulators after manually teleoperating the suction tool. In the *AS*, these additional clutching events were not required as the autonomy was in charge of the suction. Last, the average percentage of the detected blood was significantly lower in the *AS*, implying that the autonomous suction system was more effective at providing suction than the user.

In terms of mental workload metrics, 3 components of the NASA-TLX (temporal demands, performance, and frustration) and the overall workload score were significantly lower in the *AS* modality. These results can be attributed to how the autonomous modality prevented cavities from flooding – that is only when the regions had a significant blood volume that would occlude the user's view. Conversely, in the manual modality, it was seen that users rushed to complete the suturing tasks before the blood level increased to inoperable levels. This explains why the users reported higher temporal demands in this condition. The experimental results ratify the underlying hypothesis (**H1**) that the autonomous suction action can lead to better performance and lower mental demands, demonstrating the benefits of automating surgical tasks (or subtasks) in the operating room.

4.2 Experiment 2: Evaluation of the Cognitive Detection Framework

The goal of experiment 2 is to evaluate the performance of our real-time cognitive workload detection framework. The baseline hypothesis (H2) is that cognitive workload can be predicted from a combination EEG and eye tracker metrics accurately and in real-time. The

intuition for this hypothesis comes from previous works showing correlations of EEG and eye tracker markers to different metal states in surgery [116], [117].

To achieve the goal of the experiment, two datasets of physiological signals were collected. The signals came from users while they were performing surgical exercises using the DVRK robot. The first dataset contained only EEG data that was collected over multiple sessions and multiple users. Participants recruited for this experiment performed a surgical exercise under two difficulty levels to elicit different mental workload states. The goals of this first dataset were two-fold. First, the session-to-session prediction accuracies of the proposed recurrent model were evaluated. This is the model's accuracy when tested on data from a new day. This evaluation was performed as physiological signals within the same subject can have high day-to-day variability [84]. The second objective of this dataset was evaluating the user-to-user predictions of our models. This is the accuracy when testing the models on unseen users. Last, a connection between the spectral features used in our models and cognitive load theory was provided [93] by plotting the features into 2D scalp plots.

The second dataset includes EEG and eye tracker signals from one session and multiple users. For this, the difficulty level of the tasks was modified by introducing simulated bleeding events that would elicit different mental states in the users. Models trained from this dataset were integrated with the autonomous blood suction robotic assistant to predict cognitive workload in real-time. Two different evaluations were performed on these models. (1) Single user model predictions where the training and testing data belonged to the same user and (2) user-to-user prediction accuracies.

4.2.1 EEG Dataset Collection Protocols

Participants were asked to perform the peg transfer task, a crucial part of the fundamentals of robotic surgery [14]. The task consisted of transferring an object from one side of a peg transfer board to the other for 5 minutes (Refer to the Figure 4.3). To perform the exercise correctly, the user had to pick up the object with one of the robotic tooltips, lift it carefully out of the pole, transfer it to the second robotic tooltip and place it in the peg on the other side of the board.

This peg-transfer task was designed to have two difficulty levels to elicit different states of cognitive load in the user. The easy task or less cognitive demanding task was the peg-transfer performed with the normal teleoperation mode of the robot. In this condition, the surgeon's hand and the robotic gripper would move in the same direction, e.g., when the surgeon's hand moves to the right, the robotic gripper moves to the right as well. The difficult task introduced a motion reversal effect in the teleoperation of the robot. In this condition, the robot's tooltip moved in the opposite direction from the user hand movement, e.g., when the surgeon's hand moves to the right, the robotic tooltip moves to the left (opposite direction). This inversion effect emulated the fulcrum motion effect seen in traditional laparoscopic surgery [118]. This task was inspired by the studies showing how mirroring the hand movements of a previously learned task resulted in significant differences in the EEG spectral content [119].

During each session, the user performed six trials of the peg transfer task, three times in the easy level and three times in the difficult level, each one for 5 minutes, accounting for the 30 minutes of EEG data per session. Finally, the difficulty level was used to partition the EEG trials into two categorical labels: low cognitive load and high cognitive load. In total, 8 subjects were recruited for this experiment, and each of them performed 4 sessions of data collection.

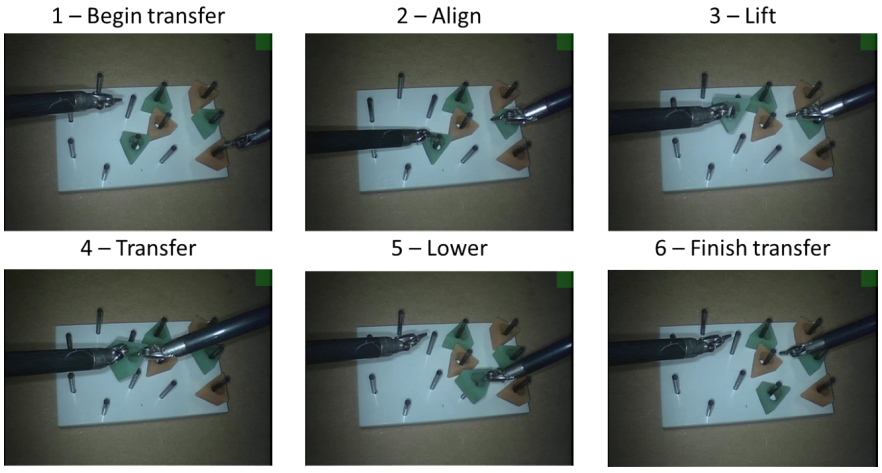


Figure 4.3. Sequence of actions to perform a peg transfer.

4.2.2 Evaluation of EEG models

Two evaluations were performed on the EEG models. First, the model's session-to-session prediction accuracy was evaluated. Since each session was collected on a different day, the models were trained with 3 sessions of data and tested on the same user's remaining session. Additionally, the effect of the recurrent models' sequence length (L) on the prediction accuracy was investigated.

Second, the accuracy was evaluated in a user-to-user transfer scenario. For this experiment, models were first trained on multiple users' data and tested on an unseen user. A small calibration set, and fine-tuning techniques were used to optimize the models to the unseen participant. The efficiency of our algorithm was analyzed by retraining the models with different proportions of the calibration set. Last, our model was compared against a convolutional network for EEG temporal features.

Multi-factor ANOVA models were fit on the testing accuracy of all the different trained models. These statistical tests were used to identify what hyperparameter configurations resulted in significant performance improvements. A significance level of 0.05 was used for all the statistical tests(α=0.05). Then, post-hoc Tukey's HSD (honestly significant difference) was used to analyze pairwise differences between factor levels. For the session-to-session evaluation, a two-way ANOVA model was used to analyze the effect of different sequences lengths (L) and the user's variability on the accuracy. For the user-to-user evaluation, an ANOVA model was used to determine the effect of the calibration's set size on the accuracy.

4.2.2.1 Session-to-Session Model Evaluation

The goal of the *session-to-session* evaluation was to quantify the generalization of the model to data from new unseen sessions. Since every session was collected on different days, this evaluation measures the generalization of the models across multiple days of data. A *leave-one-session-out* cross-validation strategy where the model was trained on 3 sessions and tested on the same user's remaining session was used. This process was repeated for each of the users in the dataset. Additionally, our proposed recurrent model was evaluated against a convolutional LSTM [120] and a plain feed-forward network.

4.2.2.2 User-to-User Model Evaluation

The goal of the *user-to-user* evaluation was to quantify the generalization of the model to data from new unseen users. For this evaluation, a *leave-one-session-out* cross-validation strategy was also used. In other words, only 1 session of data was used for testing on every cross-validation fold. However, in this case, the models were optimized using a two-stage training process. In the first round, the model was trained from a pool of multiple users' data. In the second round, the model was fine-tuned with the data calibration data from the unseen testing user. To build the training, calibration, and testing sets, the following procedure was used. One session was selected from the unseen user's data as the testing set, and the remaining 3 sessions were used as a calibration set. Data from the other participants was used as the training set for the first round of training. To optimize the model to the unseen user, the model was retrained using fine-tuning techniques and the calibration set during the second round of training. To evaluate the model performance with limited calibration data, the fine-tuning training step was repeated using different percentages of the calibration data ranging from 0% to 100%. 0% of data indicated no calibration data, and 100% represented the three sessions (90 minutes). This whole process was repeated for every session and user in the dataset. Last, the proposed model was evaluated against EEGNet [88], a convolutional EEG architecture that uses raw temporal features.

4.2.3 EEG Model Results

This section presents the results from the two evaluations performed to the EEG workload dataset: session-to-session prediction accuracies and user-to-user prediction accuracies. Last, scalp plots of the EEG spectral features are presented.

4.2.3.1 Session-to-Session Results

On figure 4.4, it can be seen a positive correlation between the sequence length and the session-to-session classification accuracy. Increasing the sequence length from 10 seconds to 100 seconds improved the model's prediction accuracy from 68% to 78%. Table 4.3b

shows the results of a two-factor ANOVA model on the sequence length and the different users. Different sequence lengths and users resulted in significant changes in the model's accuracy (P<0.000). Compared to the baseline models, the proposed model outperformed the baseline models at every sequence length (see 4.3b). The best classification accuracy of our model was 78% at a sequence length of 170 seconds. The second-best performing model was the convolutional LSTM with a prediction accuracy of 76%. These results indicate the effectiveness of the proposed models and the spectral EEG features to predict cognitive workload.

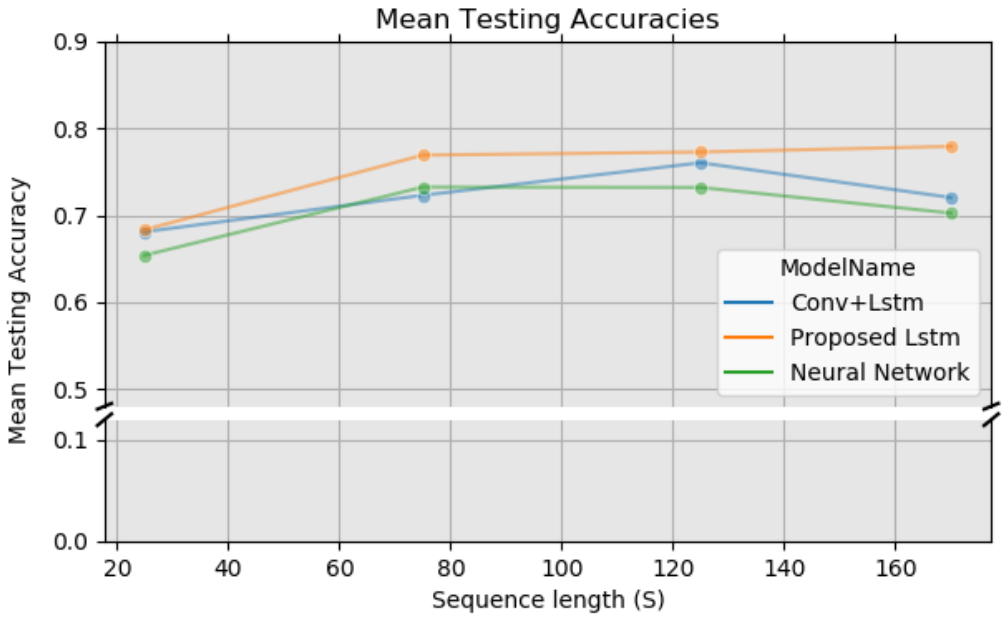


Figure 4.4. Correlation between accuracy and sequence length

4.2.3.2 User-to-User Results

Figure 4.5 shows the effect of applying the transfer learning training with different percentages of calibration data. The primary finding of this evaluation is that using transfer learning allowed the models to obtain a classification accuracy of 78.9% by using only 12.5% percent of the calibration data of the new unseen user. These results indicate that the cali-

Table 4.3. (a) Analysis of accuracy's for multiple models and sequence lengths. (b) Results for the from the 2-factor ANOVA model analyzing the effect inter-subject variations and sequence length.

	Mean Test Accuracy (%)			
Sequence Length	25	75	125	170
Models				
Conv+LSTM	68.09	72.30	76.04	72.04
Neural Network	65.39	73.24	73.20	70.25
Proposed B-LSTM	**68.33**	**76.92**	**77.28**	**77.92**

(a)

Source	**DF**	**F-Value**	**P-Value**
User	7.00	53.76	0.000
Sequence Length	7.00	25.84	0.000
User * Sequence Length	49.00	1.05	0.377

(b)

bration set's size can be significantly minimized by leveraging fine-tuning optimization. The ANOVA model indicated that changes in the accuracy due to the percentage of calibration ($p < 0.000$) were significantly different. Additionally, our proposed architecture outperforms the temporal baseline model. This result suggests that the spectral features are more likely to generalize to unseen users than the temporal raw features.

4.2.3.3 Physiological Analysis of EEG Features

To visually inspect the EEG features driving the classification of the models, the mean scalp topographical plots for each band power coefficient (Delta, Theta, Alpha, Beta) was calculated. This plot was built by calculating the mean spectral features of the low and high cognitive load conditions. Then, a baseline removal procedure was performed where the low cognitive features are subtracted from the high cognitive load features. This procedure allowed us to analyze the changes between the two mental states. After the baseline removal, the resulting feature vectors are averaged among all the participants and then projected to a 2D plane. Once in the 2D plane, data points are linearly interpolated to produce the scalp topography observed in figure 4.6.

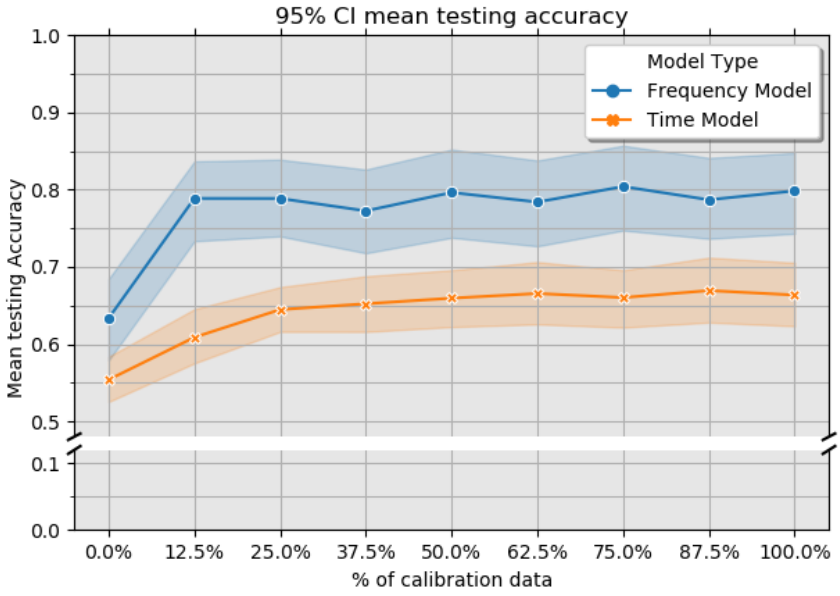


Figure 4.5. Transfer learning results

To evaluate statistical differences, a two-way ANOVA was calculated for each channel with a significance level of 0.01 ($\alpha = 0.01$). The two factors of the model were the user and workload condition. It is observed suppression of Delta activity on channels F7, FC5, FC6, T8. Increases in Theta activity were found on channels FP1, AF3, AF4, F7, F3, F4, F8, FC5, FC2, FC6, T7, C3, C4, T8, CP5. The alpha activity was also found to increase on channels F7, F8, FC5, FC6, T7, C4, T8, CP5, CP6. Finally, Beta activity increased on F8, FC6, T8, OZ while it is suppressed on FZ, F4, FC2, C3, CZ, CP5, CP2. It is remarked that most of the channels presenting significant differences due to cognitive load are in the frontal and temporal lobes of the brain.

Leveraging cognitive load theory [93] each of the significant trends previously described can be contextualized into specific cognitive functions that played a role during the experiment. First, Delta activity has been linked to the theory of Attention networks proposed by corbetta et al [121]. This theory states the existence of two different and competing attention mechanisms, one dedicated to the processing of external sensory information and another dedicated to internal concentration. Delta activity has been shown to be more active when internal concentration is required and suppressed on tasks where sensory information

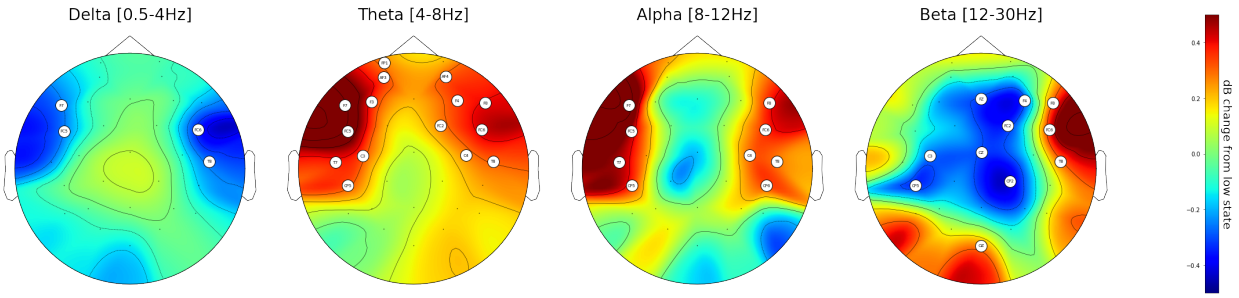


Figure 4.6. Spectral analysis of EEG features

is needed [122]–[124]. For our experiment, a suppression of Delta activity could imply that the users were highly dependent on the visual feedback during the inverted teleoperation condition.

Theta band activity has been associated with increasing use of working memory capacity [125]. Additionally, Puma et al. reported increased Theta and Alpha activity levels when increasing the number of subtasks in a multitask testing environment [126]. These studies align with the increasing levels of theta and alpha activity found in the fronto-temporal channels during the high workload conditions. Overall, these results indicate the effectiveness of our spectral band power features to discriminate between states of high and low cognitive workload.

4.2.4 Multi Sensors Dataset Collection Protocols

To build a dataset of physiological signals, a study with ten students was conducted at Purdue University. In this study, participants were required to teleoperate the DVRK robot to do a needle pass task [127] while wearing the physiological sensors. For this task, our surgical simulator was used. To elicit different workload demands, the task was designed to have two different levels of difficulty (see figure 4.7). In the low workload task, users performed the needle pass exercise with no bleeding events. In the high workload task, users completed the task as the cavity filled up with blood. Ground truth labels for our dataset

were assigned depending on the difficulty of the task. Users performed ten, 3 minutes trials, alternating between the difficult task and the easy task. This protocol accounted for 30 minutes of physiological data for each user.

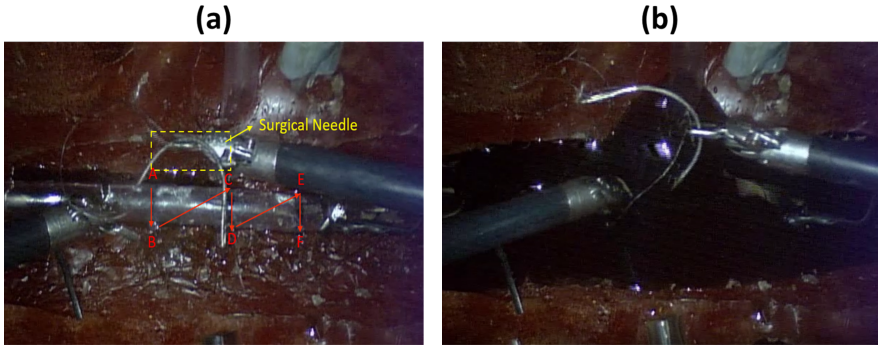


Figure 4.7. Tasks used to construct a dataset of cognitive workload signals. (a) Low difficulty task: Needle pass exercise without bleeding events. (b) High difficulty task: the simulator slowly fills up with blood as the user performs the tasks.

/

4.2.5 Multi Sensors Model Evaluation

To evaluate the performance of the multi sensor cognitive workload framework, two different methodologies were used. First, the performance of single-user models was assessed, i.e., the training and testing data comes from the samples of a single user. Second, multi-user models were evaluated, i.e., given K users, training data is obtained from K-1 users and testing data from the remaining one. In both evaluations, the number of correctly classified samples (accuracy) was used as a performance metric.

4.2.5.1 Single User Model Evaluation

As described in the data collection protocol (see subsection 4.2.4), ten trials were recorded from each user. To evaluate the single-user-model performance, a five-fold cross-validation procedure was utilized. First, the data from a user is partitioned into 5 folds, $\{\{X^{(k)}, Y^{(k)}\} \mid X^{(k)} \in \mathbf{R}^{l\times 10}, Y^{(k)} \in \{0,1\}\}$, where $k = 1,..,5$. Then, iterating over the k-folds, a model was trained

at each iteration using one of the folds as testing data and the remaining folds as the training data. This procedure was repeated for each one of the participants in the dataset.

4.2.5.2 Multi User Model Evaluation

To evaluate the multi-user model, a leave-one-user-out cross validation strategy was used. At each iteration, a model was trained using the data from 9 users. Then, the model was tested with the data of the remaining user. This evaluation allowed to quantify the generalization capabilities of the model to new users.

4.2.6 Multi Sensors Models Results

Results of single-user models can be observed in table 4.4. On average, a 75.90% $\pm$ 12.75 accuracy score was obtained for the cognitive load detection models. These results confirm that features from physiological signals provide discriminative information of the users' mental state. The highest individual accuracy was 91.36% and the lowest accuracy was 62.73%. This result could imply that the utilized set of features are not equally effective with all the users. Results from the multi-user models can be observed in table 4.5. The average accuracy in the multi-user evaluation (57.96% $\pm$ 1.23%) was 17.84% lower than the performance in the individual model evaluation. This drop in performance can be explained by the high subject-to-subject variability of physiological signals [84]. In this setting, the highest classification accuracy was 78.11% and the lowest classification 43.33%.

4.2.7 Discussions

In this section, discussion points from the two datasets of physiological signals are presented.

4.2.7.1 EEG Models

Real-time cognitive load assessment from physiological signals can improve performance and the learning of surgical skills in RAS. Nevertheless, there are still many challenges that

Table 4.4. Single user model evaluation results.

Single user model		
User id	Mean accuracy	Std
User 01	91.36%	4.73%
User 02	84.45%	10.16%
User 03	82.50%	8.90%
User 04	78.86%	6.15%
User 05	77.05%	15.23%
User 06	74.77%	25.31%
User 07	74.09%	13.72%
User 08	68.18%	9.20%
User 09	64.04%	20.62%
User 10	62.73%	13.53%
Average	75.80%	12.75%

Table 4.5. Multi user model evaluation results

Multi user model		
User id	Mean accuracy	Std
User 01	69.09%	1.13%
User 02	43.33%	1.94%
User 03	62.80%	0.75%
User 04	51.52%	0.43%
User 05	62.12%	0.92%
User 06	78.11%	1.88%
User 07	63.03%	2.15%
User 08	52.42%	0.84%
User 09	48.70%	0.57%
User 10	48.48%	1.68%
Average	57.96%	1.23%

need to be addressed, such as the high variability of the physiological signals. In this work, we proposed a recurrent neural network for predicting cognitive workload from EEG signals. Additionally, the proposed model's session-to-session and user-to-user prediction accuracies were evaluated.

In the session-to-session evaluation, the robustness of our models to the day-to-day variability of physiological signals was tested. In this scenario, the proposed models obtained a mean classification accuracy of 79.2% when using EEG segments of over 100 seconds. These experimental results show that the sequence length (i.e., the EEG windows size) and the

accuracy of the models' predictions are positively correlated. In other words, there exists a trade-off between the inference speed and the accuracy of the models. In this regard, the sequence length of the recurrent models can be fine-tuned according to the application. If accurate predictions are required, then the sequence length should be increased at the cost of increasing the latencies of the model. However, in real-time settings, the sequence length can be decreased to reduce prediction delays.

In a user-to-user scenario, a two-stage training procedure was used to improve the model's generalization to new unseen users. With this technique, a previously trained model on EEG data can significantly improve the classification by applying a recalibration procedure. Specifically, we showed that our transfer learning technique allowed our model to achieve a classification accuracy of 78.9% when using only 12.5% of the total calibration data of the new user.

4.2.7.2 Multi Sensors Models

For the multi-sensors dataset, a combination of EEG features and eye tracker features was used to classify the user's mental state as either *high cognitive* or *low cognitive* load. The best performing model achieved an accuracy of $75.90\% \pm 12.75$ in a single-user scenario, i.e., the training and testing data came from the same subject. This is evidence that combining eye tracker and EEG features provide discriminative information to detect different mental states in RAS. The main drawback of the single-user-models modality is that new users need to provide calibration before using the system. This requirement would limit the applicability of this system to time-constrained environments such as the OR.

The second evaluated modality was multi-user models. This modality is more desirable for deployment as new users can directly use the system without additional calibrations. However, prediction models have reduced accuracies due to the subject-to-subject variability of physiological signals. This limitation can be addressed by leveraging the transfer learning framework developed for the EEG dataset [84], [128]. On deployment, a calibration procedure can be applied to retrain available models to ensure optimal performance. Provided the

superior performance of single-user models, these models were used to drive the behavior of the autonomous suction system. Results from both datasets validate the **H2** hypothesis.

4.3 Experiment 3: Evaluation of Cognitive-Triggered Autonomous Suction Action

The goal of experiment 3 was to evaluate the system integration between the cognitive workload prediction module and the autonomous blood suction framework. In this scenario, the autonomous robotic assistant remained idle until the workload framework detected high mental demands on the user. Experiment 3 differs from experiment 1 in that the autonomous system would wait for cognitive load triggers rather than assisting immediately after detecting a blood accumulation. In this regard, the system only acts when the bleeding creates additional mental demands for the user. We hypothesize that this closed-loop feedback can improve surgical performance and lower mental demands on the user (H3 hypothesis). To assess the effectiveness of the integrated system, a user study with the same experimental conditions of experiment 1 was conducted. As a reminder the two experimental conditions were: manual teleoperation of the suction action(MS) and autonomous blood suction action by the robotic assistant (AS). Following the same guidelines as in experiment 1, a within-subject study was designed to account for individual differences. Also, the order of the experimental conditions was randomized to control for the learning effects (see 4.1). The rest of the subsections are divided as follows. First, the integration of the system is described. Then, the user study protocol and evaluation metrics are described. Finally, the experiment's results and conclusions are presented.

4.3.1 User Study Protocol

To evaluate the cognitive triggered autonomous system, a user study was conducted with 8 students. This study was approved by the Purdue University's Review Board under the protocol IRB-2021-22. For this study, the users were asked to come for two sessions occurring on subsequent days. In the first session, the users were familiarized with the teleoperation of the robot and the experimental task. For this experiment, the same surgical tasks as

in experiment 1 were used (see 4.1.1). On the second day, we asked the users to wear the EEG and eye tracker sensors before starting the experiments. After that, they performed the assigned tasks under the *AS* and *MS* modalities while their physiological signals were monitored in real-time.

4.3.2 Evaluation Metrics

To evaluate the system's performance, a combination of objective performance and workload metrics were used. Additionally, the users completed a NASA TLX questionnaire after each experimental trial to report subjective workload demands. The objective metrics included for this experiment are the same as in 4.1. As a reminder, this metrics are categorized into time, motion, system events, and blood metrics (see subsection 4.1.2). Additionally, we included metrics of human-robot collaboration fluency . Collaboration fluency refers to the capability of a human and robot to work as a team proficiently towards a common goal [129]. We included the following fluency metrics. (1) Human idle time, measured by the idle time of the main instrument arms (PSM1 and PSM2). (2) Robot idle time, measured as the percentage of time that the autonomous assistant (PSM3) was idle. (3) Concurrent activity, measured as the percentage of time that both the autonomy and human were simultaneously active. Low human and robot idle time and high concurrent activity time are indicators of fluency in human-robot collaboration. To objectively evaluate the users' cognitive workload, the average workload prediction was calculated in both experimental conditions. This workload index was generated in real-time by the proposed framework

4.3.3 Results

Objective metrics results are presented in table 4.6. The mean and standard deviation were calculated for each metric to compare between the autonomous (*AS*) and manual suction modalities (*MS*). Completion time under the *AS* (371 ± 133) was 162 seconds lower compared to the *MS* modality (533 ± 210). These differences were found statistically significant. In terms of collaboration fluency, *AS* lead to a 23% increase of the concurrent activity. Additionally, *AS* reduced the PSM1 and PSM2 idle time respectively by 28% and

Table 4.6. Objective performance metrics results. Statistically significant results ($\alpha = 0.05$) were highlighted in bold

Type	Name	Mean (std), N=8		T-test (alpha=0.05)	
		Autonomy	Manual	T-statistic	p-value
time	clutching time (s)	3.69 (7.17)	17.94 (10.48)	-3.357	**0.012**
	completion time (s)	370.86 (133.16)	532.93 (210.51)	-3.953	**0.006**
percentage	concurrent activity (%)	0.23 (0.08)	0 (0)	7.730	**0.000**
	psm1 idle time (%)	0.21 (0.09)	0.49 (0.08)	-8.553	**0.000**
	psm2 idle time (%)	0.39 (0.16)	0.58 (0.13)	-7.984	**0.000**
	psm3 idle time (%)	0.71 (0.09)	0.81 (0.05)	-2.229	0.061
motion	psm1 velocity (cm/s)	1.01 (0.26)	0.73 (0.21)	4.729	**0.002**
	psm2 velocity (cm/s)	0.71 (0.18)	0.51 (0.18)	8.861	**0.000**
	psm3 velocity (cm/s)	0.65 (0.19)	0.39 (0.09)	3.013	**0.020**
events	tool changing events	1 (1.31)	13.13 (8.56)	-4.392	**0.003**
	clutching events	0 (0)	4.25 (2.82)	-4.269	**0.004**
blood	percentage blood (%)	0.13 (0.03)	0.12 (0.05)	0.686	0.515

18%. There were no statistical differences in the autonomous assistant (PSM3) idle time. These results imply an increment of the users' attention on the main task in the *AS*. The velocity of the main instruments (PSM1 and PSM2) was found significantly higher in the *AS* modality. The respective velocity increments of PSM1 and PSM2 were 0.28cm/s and 0.20cm/s. The PSM3 velocity was not statistically different in both conditions. Finally, no statistically significant differences were found in the percentage of the detected blood accumulations.

Cognitive metrics are given in table 4.7. These metrics were divided between the subjective metrics from the NASA-TLX questionnaires and objective metrics from the cognitive workload sensing framework. Similarly, mean and standard deviation were calculated to compare between experimental conditions. All components of NASA-TLX were found significantly lower in the *AS* modality. Additionally, the overall workload score in *AS* (23 ± 11) was reduced 15 points compared to *MS* (35 ± 12). Both these results indicate, the user experiencing lower cognitive demands when working with the autonomous system. The average cognitive index was not found statistically different between the two conditions.

Table 4.7. Nasa-tlx results and measured cognitive workload. Statistically significant results (α=0.05) were highlighted in bold.

Type	Component	Mean (std), N=8		T-test	
		autonomy	manual	t-statistic	p-value
Nasa-TLX	Mental demand	4.38 (3.10)	6.12 (2.60)	-4.041	**0.005**
	Physical demand	3.50 (2.35)	6.06 (2.29)	-4.432	**0.003**
	Temporal demand	4.94 (2.69)	6.81 (2.33)	-4.356	**0.003**
	Performance	2.19 (1.91)	3.38 (2.33)	-2.967	**0.021**
	Effort	4.50 (2.63)	6.31 (3.33)	-3.200	**0.015**
	Frustration	4.12 (2.53)	6.81 (2.15)	-5.480	**0.001**
	Workload score	23.62 (13.62)	35.50 (12.68)	-4.799	**0.002**
Workload prediction	Cognitive index	0.527 (0.254)	0.585 (0.230)	-1.666	0.14

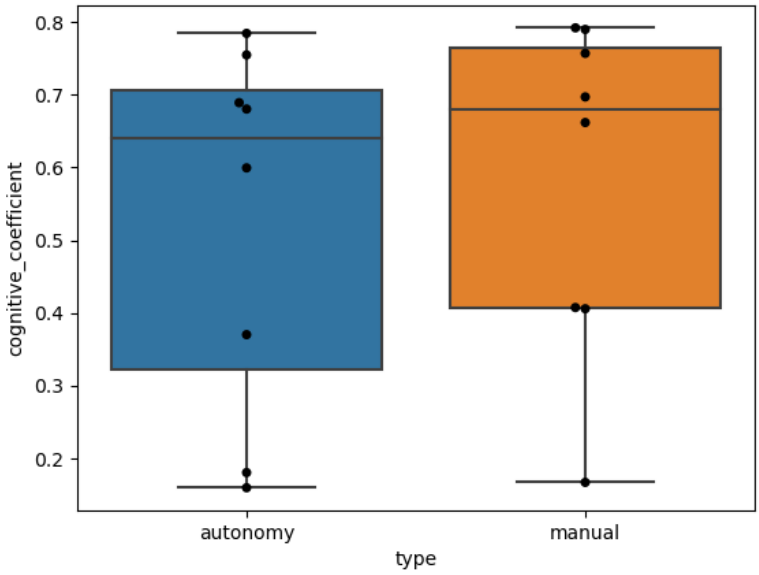


Figure 4.8. Cognitive module average predictions box plot

4.3.4 Discussions

In this section, a cognitive-triggered robotic assistant was presented. In this system, the user's cognitive workload is monitored in real-time and used to trigger the actions of the automated assistant. In this regard, autonomous assistance can be viewed as a feedback mechanism to alleviate the surgeon's cognitive demands during a procedure. A user study comparing the cognitive triggered autonomous system against manual teleoperation of the suction tool was conducted. Objective surgical performance metrics and cognitive metrics were calculated to validate the benefits of our system. In terms of surgical performance, the cognitive autonomy allowed the user to complete the suturing exercise 162 seconds faster than manual teleoperation. This result could be attributed to the improvements in the collaboration fluency between the robot and the user. On the autonomous modality, the idle time of the main instruments was reduced by a 32%. This result implies that the surgeon was able to dedicate more time to the main task. Additionally, it is highlighted that the suction tool was active the same percentage of time during both conditions (PSM3 idle time

showed no statistically significant differences). This result suggests that the motions of the suction tool were more efficient when the autonomous algorithm oversaw the teleoperation.

In terms of cognitive demands, NASA-TLX scores indicate users felt the surgical exercise was less demanding while working with autonomy. These results have two important implications. First, it indicates the importance of allowing novice users to fully concentrate on the main surgical exercise to improve surgical performance. Second, it highlights the effectiveness of our system in handling bleeding events. On the other hand, the cognitive index showed no statistically significant differences between the experimental conditions. This finding could be explained by additional challenges created by the autonomous system, such as occlusion of the surgical view. Certain positions inside the cavity resulted in unavoidable occlusions from the suction tool. This most likely resulted in some additional mental demands that were not present in the manual teleoperation. This problem can be alleviated by providing autonomy with information on the current surgical step performed by the user. In this regard, the autonomy can plan suctions in locations where occlusion is unavoidable at better timings. Overall, the experiments validate the hypothesis **H3** that performance and workload metrics can be improved using an automated suction system triggered by a cognitive workload sensing framework.

4.4 Summary

This chapter presents three evaluation experiments for our autonomous suction assistant and cognitive workload sensing module. In experiment 1, it was evaluated whether surgical performance could be improved by collaborating with a robotic assistant. In the second experiment, the performance of a proposed cognitive workload prediction module was assessed in a single-user and multi-user setting. Finally, a closed-loop system was evaluated where the autonomy was triggered by high cognitive workload as detected with our sensing framework. This final experiment aimed to validate whether a cognitive triggered autonomy could lead to improved performance and lower mental demands.

5. CONCLUSIONS AND FUTURE WORK

The introduction of autonomy in the OR has the potential to transform surgical care by reducing the number of surgical assistants and improving the overall surgical performance [130]. However, autonomous robots are not capable of completing surgical procedures without the supervision of human surgeons [54]. The deployment of autonomous technology is even more limited in minimally invasive procedures involving soft tissue interaction as they are challenging environments to model and manipulate. These environments will require significant breakthroughs both at the hardware and at the algorithm level before autonomous systems are reliable enough to be used with patients.

An initial effort towards AI systems in surgery is developing autonomous systems that can take over specific surgical subtasks rather than the full task, and thereby alleviating the cognitive and physical demands on the surgeons and surgical assistants. In this thesis, an autonomous blood suction robotic assistant was developed and deployed on a DVRK robot. The goal of this system was to alleviate the cognitive demands created by hemorrhage events during a procedure. Our framework leveraged state-of-the-art semantic segmentation algorithms to localize blood pooling events in endoscopic images. Then, using the segmented images, spatial coordinates were calculated and used by the robot to draw the blood accumulations, improving the visibility of the surgical field, and reducing the potential for infections.

To create a synergistic human-robot collaboration, a mechanism that monitor the operator's cognitive load was developed and an AR module was created to visualize imminent robot actions. These two developments enable mutual grounding: the surgeon and the robot can understand the state and intentions of each other. First, a real-time cognitive estimation system was developed, that allowed the robot to adapt its behavior based on the surgeon's current mental state. Cognitive estimation was achieved by continuously monitoring EEG and eye-tracker physiological markers. Second, an AR annotations module was developed to indicate the target location for the next action of the robotic assistant. These annotations were displayed in the console of the surgical robot and superimposed over the patient's anatomy.

To evaluate the proposed framework, a computational experiment and two user studies were conducted. The first experiment was a user study to evaluate the effectiveness of the autonomous robotic assistant to control a simulated bleeding event. For this experiment, the autonomous system provided suction immediately after the computer system detected blood accumulations, while the user performed a suturing surgical exercise. The autonomous system was compared against manual teleoperation of the suction tool to assess the system's benefits in terms of surgical performance and workload demands.

The study's main finding was a reduction of the completion time, and the reported workload demands when using the autonomous system. In terms of completion time, users working with the autonomous system finished the task 162 seconds faster than when manually teleoperating the suction tool. Furthermore, there was a significant reduction in the number of clutching events and the percentage of detected blood. These results indicate that the autonomous system can effectively perform the blood suction without affecting the user's work on the main task. Last, a reduced number of clutching events imply a more effective use of the console space, which has been associated with surgical expertise [75]. These results address the initial hypothesis H1 that automating surgical sub-tasks in RMIS can improve performances and lower mental demands.

For the second study, a computational experiment was conducted to assess the prediction performance of the proposed cognitive workload detection system. For this, two different datasets of physiological signals were collected. For the first dataset, users were asked to perform a peg transfer task on multiple days while EEG data was recorded. In this study, the task's difficulty was increased to elicit states of high cognitive demands on the users by modifying the teleoperation of the surgical robot. Using a recurrent neural network and a set of spectral features from the EEG, the developed framework obtained a 78% session-to-session prediction accuracy with a sample of 170 seconds of EEG data. This prediction accuracy was superior to the baseline classification accuracy provided by a convolutional LSTM model and a feedforward neural network.

Additionally, the effectiveness of the model to predict data from new unseen users was evaluated. In this regard, a user-to-user classification accuracy of 78.9% was obtained when using a calibration set of 90 minutes. Using fine-tunning techniques, the calibration set was

reduced to 11.25 minutes (12.5% of the original size) while maintaining a calibration accuracy of 78.9%. The two main findings from these experiments are that there the following. (1) There exists a positive correlation between the size of the EEG sample and the model's classification accuracy. In other words, more accurate prediction accuracies can be obtained at the expense of delays in the predictions. (2) Using fine-tuning techniques on EEG spectral features allow workload assessments to unseen users with a minimized calibration set of 11.25 minutes of EEG data. Last, spectral features to scalp plots were generated to analyze changes in the brain regions when increasing the cognitive workload. On the resulting plot, a significant increase of theta and alpha oscillations on fronto-temporal channels wa s observed when increasing the task difficulty.

The second dataset of physiological signals included EEG and eye-tracking. This dataset comprises a single session of multiple users teleoperating the surgical robot to perform a surgical exercise during a bleeding scenario. Models trained with this dataset were evaluated under two different modalities: single-user models and multi-user models. In the single-user modality, an average classification accuracy of 76% was achieved. This result shows how to effectively use EEG and eye-tracker features to predict cognitive states in RMIS procedures. In the multi-user models, which is a more challenging setting, an average classification accuracy of 58% was obtained. Results from both datasets validate the hypothesis H2 that cognitive workload can be measured accurately through physiological analysis.

The third study evaluated the integration of the autonomous system with the cognitive workload framework. In this setting, the robotic assistant would only act when the user's mental state was classified as high cognitive workload. To achieve the study's goal, the system was assessed against suction tool manual teleoperation. The main results for this study are reduction of the completion time and improved human-robot collaboration fluency, i.e., the user was more dedicated to the main task as the robot effectively handled the bleeding events. In terms of completion time, users completed the proposed task 161 seconds faster than compared to manual teleoperation. These results can be attributed to the improved human-robot collaboration fluency. Concurrent activity increased by a 23% in the autonomous condition. On the other hand, the idle time of the robot's main instruments dropped from 50% in the manual to 21% in the autonomous condition. Both these metrics indicate a

more efficient use of the robotic instruments. Overall, the experiments' results highlight the benefits of automating surgical subtasks and real-time cognitive workload monitoring in RMIS.

5.1 Limitations and Future Work

In the following section, the limitations of this work and possible future work solutions are presented. Additionally, some ideas are presented regarding the use of cognitive workload sensing and possible enhancements to improve the reliability and interpretability of the prediction models.

First, the DVRK's teleoperation software is limited compared to the teleoperation of commercial da Vinci surgical systems. Notably, the software version 1.7 of the DVRK presented delays when the user attempted to switch between the main instruments and the suction tools and tried to move the robot from an idle position. Another limitation is that the users involved in the experiments were college students and not medical personnel. Even though participants were allowed to practice the surgical tasks thoroughly before the experiments, the surgical technique was poor when compared to medical students. These issues indicate that the findings presented may slightly differ when applied to a setting outside the laboratory environment. Future work should focus on conducting newer user studies with training residents or surgeons using a clinical da Vinci system. Provided that commercials systems do not offer programmable APIs, future experiments would have the surgeon teleoperating a clinical system and the autonomous algorithm deployed using a single PSM of the DVRK.

On a different note, and in terms of the usability of the system, some participants noted that the lack of control of the suction tool led to increased frustration levels while working with the autonomy framework. This could be due to a lack of trust and familiarization with the autonomous system. Previous works in human-robot collaboration point out that trust can be generated by improving the predictability and *legibility* of the robotic system [131], [132]. In future work, the AR annotations can be enhanced using predictive displays [133], i.e., advanced augmented reality graphics to convey the users with the subsequent

autonomous actions before they occur. Utilizing a graphics engine such as OpenGL, an alpha-blended CAD model of the suction tool could be rendered in the console screen, providing better information on how the robotic assistant would move in the surgical field. Other potential solutions to improve collaboration can include adding voice and gesture control mechanisms to enhance the interaction between the surgeon and the autonomy.

Another limitation of our framework concerns the simulation setting. Providing suction in some regions inside the anatomical cavities led to partial visual occlusion, affecting users' performance. In future work, this issue can be addressed by enhancing the AI algorithm to select target locations and trajectories that minimize the occlusion in the image space [134]. In addition, when occlusion is unavoidable, the AI algorithm could leverage contextual information about the surgical steps performed by the surgeon. In this regard, the autonomy would plan suction actions in target locations where occlusion is unavoidable during simple operator tasks.

Our cognitive workload sensing module was only tested on bleeding events. While this is one of the most critical stressors in surgery, it is not the only one. Therefore, there is the question about how well our results generalize to other surgical stressors. In future work, user studies can be conducted in additional surgical settings to verify the generality of our framework. Additionally, the prediction models can be enhanced to include other input modalities, such as the surgeon's hand motions, position and stance of the surgical staff, position of the instruments, etc. Dalveren *et al.* demonstrated that hand state, i.e., performing with dominant, non-dominant, or both hands, influenced gaze patterns during simulated laparoscopic training [135]. These results indicate that hand kinematic data provides meaningful information about the mental state of the surgeon. Our cognitive workload sensing module could be easily enhanced to include kinematic data, as it is provided by the robot API, to improve the robustness of the predictions.

5.2 Research Questions

In the first chapter, three research questions were introduced and were answered by the proposed methodology and experiments. Below a summary of the primary outcomes and answers to each question is presented.

Research question 1 (RQ1): How to develop computer vision algorithms and control policies to automate the blood suction task in the context of RMIS?

To answer this question, a semantic segmentation algorithm to localize blood accumulations on the robot's endoscopic images was developed. To perform the automatic segmentation, a fully convolutional adaptation of the VGG-16 network was used. For training purposes, a dataset of 180 images was collected and labelled. This dataset was used to train a model to localize and segment relevant surgical objects such as the blood accumulations, vessels, and the robotic instruments in the images. Then, using geometric computer vision methods, the robotic instruments to the camera reference frame were registered. This process allowed to transform the pixel location of the objects of interest into spatial coordinates that the robot could use. Last, a simple path planning algorithm was developed to allow the surgical robot to reach and clear the detected blood accumulations.

Research question 2 (RQ2): How to predict cognitive load from multimodal physiological signals in real-time?

To answer this question, a multimodal framework based on EEG and eye-tracking to predict the cognitive workload in real-time was proposed. First, a dataset of physiological signals was collected, where the difficulty level of the surgical exercises was used to label the data ground truth values. Then, low dimensional features from the sensors' raw data were extracted. From the EEG, frequency domain features were built, which have been previously correlated with mental workload demands. From the eye tracker, features from the pupil diameter and the points where the user fixated his gaze were calculated. Using a combined feature vector from both sensors, a classification accuracy of 75% was obtained when training and testing our models on data from the same user. Finally, a synchronization

framework was developed to calculate the physiological features in real time. Results from the evaluation experiments validate hypothesis H2 that real-time can be measured in real-time with a combination of EEG and eye tracking signals.

Research question 3 (RQ3): Does the autonomy-controlled suction action improve the surgeon's task performance and cognitive demands?

To answer this question, a within-user study with the DVRK robot was conducted. In this experiment, the participants were required to complete a suturing exercise using the main robotic manipulators. In addition, a custom-made suction tool was attached to the auxiliary arm of the robot. To measure the benefits of our system, users completed a surgical exercise under two conditions: (1) manual teleoperation of the suction tool and (2) autonomous teleoperation of the suction tool. Surgical performance metrics and cognitive metrics were recorded during both conditions. Surgical performance metrics included completion time, idle time of the users, concurrent activity, the robotic instruments' velocity, and the average percentage of blood in the surgical field. To quantify cognitive demands, the predictions of the cognitive prediction framework and responses from NASA-TLX questionnaires were studied. The main findings from this study are that the autonomous system led to faster completion of the surgical exercise and lower reported mental demands, and those results were statistically significant. Results from the user studies validate the hypothesis H1 that surgical performance can be improved and that cognitive demands can be reduced by using an automated system for blood suction.

5.3 Summary

This chapter presents the main results and findings from our work. The limitations of the proposed algorithms were discussed, and some future working directions are presented. Finally, the three research questions that guided this dissertation with their corresponding answers are presented.

REFERENCES

[1] C. D'Ettorre, A. Mariani, A. Stilli, F. R. y. Baena, P. Valdastri, A. Deguet, P. Kazanzides, R. H. Taylor, G. S. Fischer, S. P. DiMaio, A. Menciassi, and D. Stoyanov, "Accelerating surgical robotics research: Reviewing 10 years of research with the dVRK," *arXiv:2104.09869 [cs]*, May 13, 2021. arXiv: 2104.09869. [Online]. Available: http://arxiv.org/abs/2104.09869.

[2] *Intuitive surgical sustainability Report 2020*.

[3] *Surgical Robots Market | by Product & Service, Application & End User | MarketsandMarkets*. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/surgical-robots-market-256618532.html.

[4] O. Sgarbura and C. Vasilescu, "The decisive role of the patient-side surgeon in robotic surgery," en, *Surgical Endoscopy*, vol. 24, no. 12, pp. 3149–3155, Dec. 2010, ISSN: 1432-2218. DOI: 10.1007/s00464-010-1108-9. [Online]. Available: https://doi.org/10.1007/s00464-010-1108-9.

[5] L. Zhang, J. Ma, L. Zang, F. Dong, A. Lu, B. Feng, Z. He, H. Hong, and M. Zheng, "Prevention and management of hemorrhage during a laparoscopic colorectal surgery," en, *Annals of Laparoscopic and Endoscopic Surgery*, vol. 1, no. 7, Dec. 2016, Number: 7 Publisher: AME Publishing Company, ISSN: 2518-6973. DOI: 10.21037/3615. [Online]. Available: https://ales.amegroups.com/article/view/3615.

[6] C. G. Ball and E. Dixon, Eds., *Treatment of Ongoing Hemorrhage*, en. Cham: Springer International Publishing, 2018, ISBN: 978-3-319-63494-4 978-3-319-63495-1. DOI: 10.1007/978-3-319-63495-1. [Online]. Available: http://link.springer.com/10.1007/978-3-319-63495-1.

[7] D. Papandria, T. Lardaro, D. Rhee, G. Ortega, A. Gorgy, M. A. Makary, and F. Abdullah, "Risk factors for conversion from laparoscopic to open surgery: Analysis of 2138 converted operations in the american college of surgeons national surgical quality improvement program," *The American Surgeon*, vol. 79, no. 9, pp. 914–921, Sep. 2013, ISSN: 1555-9823.

[8] R. Kumar and A. K. Hemal, "The 'scrubbed surgeon' in robotic surgery," en, *World Journal of Urology*, vol. 24, no. 2, pp. 144–147, Jun. 2006, ISSN: 1433-8726. DOI: 10.1007/s00345-006-0068-0. [Online]. Available: https://doi.org/10.1007/s00345-006-0068-0.

[9] B. Yuh, "The Bedside Assistant in Robotic Surgery – Keys to Success," en, *UROLOGIC NURSING*, vol. 33, no. 1, p. 4,

[10] O. E. Martinez, C. Murphy, A. Bsatee, D. H. Brown, and R. Abaza, "Impact of Surgeon-Controlled Suction during Robotic Prostatectomy to Reduce Dependence on Bedside Assistance," *Journal of Endourology*, Feb. 2021, Publisher: Mary Ann Liebert, Inc., publishers, ISSN: 0892-7790. DOI: 10.1089/end.2020.1059. [Online]. Available: https://www.liebertpub.com/doi/10.1089/end.2020.1059.

[11] M. Stifelman and A. Mass, *#WCE2014 - The use of a robotic remotely operated suction/irrigation (ROSI) system may lead to less post-operative blood loss after robotic pyeloplasty - Interview*. [Online]. Available: https://www.urotoday.com/conference-highlights/global-congress-on-bladder-cancer-2018/1622-conferences/wce-2014/wce-2014-trauma-reconstruction/75097-wce2014-the-use-of-a-robotic-remotely-operated-suction-irrigation-rosi-system-may-lead-to-less-post-operat.

[12] R. D. Dias, M. C. Ngo-Howard, M. T. Boskovski, M. A. Zenati, and S. J. Yule, "Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload," *BJS (British Journal of Surgery)*, vol. 105, no. 5, pp. 491–501, 2018, ISSN: 1365-2168. DOI: 10.1002/bjs.10795. [Online]. Available: https://bjssjournals.onlinelibrary.wiley.com/doi/abs/10.1002/bjs.10795.

[13] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *arXiv:1605.06211 [cs]*, May 2016, arXiv: 1605.06211. [Online]. Available: http://arxiv.org/abs/1605.06211.

[14] R. Smith, V. Patel, and R. Satava, "Fundamentals of robotic surgery: A course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 10, no. 3, pp. 379–384, 2014, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rcs.1559, ISSN: 1478-596X. DOI: https://doi.org/10.1002/rcs.1559. [Online]. Available: http://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.1559.

[15] L. Rozo, H. B. Amor, S. Calinon, A. Dragan, and D. Lee, "Special issue on learning for human–robot collaboration," *Autonomous Robots*, vol. 42, no. 5, pp. 953–956, Jun. 1, 2018, ISSN: 1573-7527. DOI: 10.1007/s10514-018-9756-z. [Online]. Available: https://doi.org/10.1007/s10514-018-9756-z.

[16] S. Hajibandeh, S. Hajibandeh, A. Kelly, J. Shah, R. M. A. Khan, N. Panda, M. Mansour, S. Malik, and S. Dalmia, "Irrigation versus suction alone in laparoscopic appendectomy: Is dilution the solution to pollution? a systematic review and meta-analysis," *Surgical Innovation*, vol. 25, no. 2, pp. 174–182, Apr. 2018, ISSN: 1553-3514. DOI: 10.1177/1553350617753244.

[17] E. P. W.-v. d. Putten, R. H. M. Goossens, J. J. Jakimowicz, and J. Dankelman, "Haptics in minimally invasive surgery – a review," *Minimally Invasive Therapy & Allied Technologies*, vol. 17, no. 1, pp. 3–16, Jan. 1, 2008, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13645700701820242, ISSN: 1364-5706. DOI: 10.1080/13645700701820242. [Online]. Available: https://doi.org/10.1080/13645700701820242.

[18] A. E. Abdelaal, P. Mathur, and S. E. Salcudean, "Robotics in vivo: A perspective on human–robot interaction in surgical robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, pp. 221–242, 2020, _eprint: https://doi.org/10.1146/annurev-control-091219-013437. DOI: 10.1146/annurev-control-091219-013437. [Online]. Available: https://doi.org/10.1146/annurev-control-091219-013437.

[19] M. H. L. Liow, P. L. Chin, H. N. Pang, D. K.-J. Tay, and S.-J. Yeo, "THINK surgical TSolution-one® (robodoc) total knee arthroplasty," *SICOT-J*, vol. 3, ISSN: 2426-8887. DOI: 10.1051/sicotj/2017052. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5663203/.

[20] B. S. Peters, P. R. Armijo, C. Krause, S. A. Choudhury, and D. Oleynikov, "Review of emerging surgical robotic technology," *Surgical Endoscopy*, vol. 32, no. 4, pp. 1636–1655, Apr. 1, 2018, ISSN: 1432-2218. DOI: 10.1007/s00464-018-6079-2. [Online]. Available: https://doi.org/10.1007/s00464-018-6079-2.

[21] E. Ackerman. (). "New da vinci xi surgical robot is optimized for complex procedures - IEEE spectrum," IEEE Spectrum: Technology, Engineering, and Science News, [Online]. Available: https://spectrum.ieee.org/automaton/robotics/medical-robots/new-da-vinci-xi-surgical-robot.

[22] K. Khandalavala, T. Shimon, L. Flores, P. R. Armijo, and D. Oleynikov, "Emerging surgical robotic technology: A progression toward microbots," *Annals of Laparoscopic and Endoscopic Surgery*, vol. 5, no. 0, Jan. 20, 2020, Number: 0 Publisher: AME Publishing Company, ISSN: 2518-6973. DOI: 10.21037/ales.2019.10.02. [Online]. Available: https://ales.amegroups.com/article/view/5499.

[23] J. A. Buza, C. R. Good, R. A. Lehman, J. Pollina, R. V. Chua, A. L. Buchholz, and J. L. Gum, "Robotic-assisted cortical bone trajectory (CBT) screws using the mazor x stealth edition (MXSE) system: Workflow and technical tips for safe and efficient use," *Journal of Robotic Surgery*, vol. 15, no. 1, pp. 13–23, Feb. 1, 2021, ISSN: 1863-2491. DOI: 10.1007/s11701-020-01147-7. [Online]. Available: https://doi.org/10.1007/s11701-020-01147-7.

[24] Y.-H. Su, A. Munawar, A. Deguet, A. Lewis, K. Lindgren, Y. Li, R. H. Taylor, G. S. Fischer, B. Hannaford, and P. Kazanzides, "Collaborative robotics toolkit (CRTK): Open software framework for surgical robotics research," in *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, Nov. 2020, pp. 48–55. DOI: 10.1109/IRC.2020.00014.

[25] B. Hannaford, J. Rosen, D. W. Friedman, H. King, P. Roan, L. Cheng, D. Glozman, J. Ma, S. N. Kosari, and L. White, "Raven-II: An open platform for surgical robotics research," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 954–959, Apr. 2013, Conference Name: IEEE Transactions on Biomedical Engineering, ISSN: 1558-2531. DOI: 10.1109/TBME.2012.2228858.

[26] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da Vinci® Surgical System," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, ISSN: 1050-4729, May 2014, pp. 6434–6439. DOI: 10.1109/ICRA.2014.6907809.

[27] N. Madapana, M. M. Rahman, N. Sanchez-Tamayo, M. V. Balakuntala, G. Gonzalez, J. P. Bindu, L. N. V. Venkatesh, X. Zhang, J. B. Noguera, T. Low, R. Voyles, Y. Xue, and J. Wachs, "DESK: A robotic activity dataset for dexterous surgical skills transfer to medical robots," *arXiv:1903.00959 [cs]*, Mar. 3, 2019. arXiv: 1903.00959. [Online]. Available: http://arxiv.org/abs/1903.00959.

[28] G. Ra and A. A, "An approach to modeling closed-loop kinematic chain mechanisms, applied to simulations of the da vinci surgical system," *Acta Polytechnica Hungarica*, vol. 16, no. 8, Sep. 2019. DOI: 10.12700/APH.16.8.2019.8.3. [Online]. Available: https://par.nsf.gov/biblio/10113819-approach-modeling-closed-loop-kinematic-chain-mechanisms-applied-simulations-da-vinci-surgical-system.

[29] G. A. Fontanelli, M. Selvaggio, M. Ferro, F. Ficuciello, M. Vendittelli, and B. Siciliano, "A v-REP simulator for the da vinci research kit robotic platform," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, ISSN: 2155-1782, Aug. 2018, pp. 1056–1061. DOI: 10.1109/BIOROB.2018.8487187.

[30] A. Munawar, Y. Wang, R. Gondokaryono, and G. S. Fischer, "A real-time dynamic simulator and an associated front-end representation format for simulating complex robots and environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, ISSN: 2153-0866, Nov. 2019, pp. 1875–1882. DOI: 10.1109/IROS40897.2019.8968568.

[31] F. Chadebecq, F. Vasconcelos, E. Mazomenos, and D. Stoyanov, "Computer vision in the surgical operating room," *Visceral Medicine*, vol. 36, no. 6, pp. 456–462, 2020, Publisher: Karger Publishers, ISSN: 2297-4725, 2297-475X. DOI: 10.1159/000511934. [Online]. Available: https://www.karger.com/Article/FullText/511934.

[32] B. Varadarajan, C. Reiley, H. Lin, S. Khudanpur, and G. Hager, "Data-derived models for segmentation with application to surgical assessment and training," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, Eds., ser. Lecture Notes in Computer Science, Berlin, Heidelberg: Springer, 2009, pp. 426–434, ISBN: 978-3-642-04268-3. DOI: 10.1007/978-3-642-04268-3_53.

[33] J. Y. K. Chan, F. C. Holsinger, S. Liu, J. M. Sorger, M. Azizian, and R. K. Y. Tsang, "Augmented reality for image guidance in transoral robotic surgery," *Journal of Robotic Surgery*, vol. 14, no. 4, pp. 579–583, Aug. 1, 2020, ISSN: 1863-2491. DOI: 10.1007/s11701-019-01030-0. [Online]. Available: https://doi.org/10.1007/s11701-019-01030-0.

[34] M. Allan, S. Kondo, S. Bodenstedt, S. Leger, R. Kadkhodamohammadi, I. Luengo, F. Fuentes, E. Flouty, A. Mohammed, M. Pedersen, A. Kori, V. Alex, G. Krishnamurthi, D. Rauber, R. Mendel, C. Palm, S. Bano, G. Saibro, C.-S. Shih, H.-A. Chiang, J. Zhuang, J. Yang, V. Iglovikov, A. Dobrenkii, M. Reddiboina, A. Reddy, X. Liu, C. Gao, M. Unberath, M. Kim, C. Kim, C. Kim, H. Kim, G. Lee, I. Ullah, M. Luna, S. H. Park, M. Azizian, D. Stoyanov, L. Maier-Hein, and S. Speidel, "2018 robotic scene segmentation challenge," *arXiv:2001.11190 [cs]*, Aug. 2, 2020. arXiv: 2001.11190. [Online]. Available: http://arxiv.org/abs/2001.11190.

[35] R. Zhu, R. Zhang, and D. Xue, "Lesion detection of endoscopy images based on convolutional neural network features," in *2015 8th International Congress on Image and Signal Processing (CISP)*, Oct. 2015, pp. 372–376. DOI: 10.1109/CISP.2015.7407907.

[36] A. Attanasio, B. Scaglioni, M. Leonetti, A. F. Frangi, W. Cross, C. S. Biyani, and P. Valdastri, "Autonomous Tissue Retraction in Robotic Assisted Minimally Invasive Surgery – A Feasibility Study," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6528–6535, Oct. 2020, Conference Name: IEEE Robotics and Automation Letters, ISSN: 2377-3766. DOI: 10.1109/LRA.2020.3013914.

[37] D. Pakhomov, V. Premachandran, M. Allan, M. Azizian, and N. Navab, "Deep residual learning for instrument segmentation in robotic surgery," in *Machine Learning in Medical Imaging*, H.-I. Suk, M. Liu, P. Yan, and C. Lian, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 566–573, ISBN: 978-3-030-32692-0. DOI: 10.1007/978-3-030-32692-0_65.

[38] L. C. Garcia-Peraza-Herrera, L. Fidon, C. D'Ettorre, D. Stoyanov, T. Vercauteren, and S. Ourselin, "Image compositing for segmentation of surgical tools without manual annotations," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1450–1460, May 2021, Conference Name: IEEE Transactions on Medical Imaging, ISSN: 1558-254X. DOI: 10.1109/TMI.2021.3057884.

[39] M. S. Nosrati, R. Abugharbieh, J.-M. Peyrat, J. Abinahed, O. Al-Alao, A. Al-Ansari, and G. Hamarneh, "Simultaneous multi-structure segmentation and 3d nonrigid pose estimation in image-guided robotic surgery," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 1–12, Jan. 2016, Conference Name: IEEE Transactions on Medical Imaging, ISSN: 1558-254X. DOI: 10.1109/TMI.2015.2452907.

[40] Y. Fu, W. Zhang, M. Mandal, and M. Q.-H. Meng, "Computer-aided bleeding detection in WCE video," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 2, pp. 636–642, Mar. 2014, Conference Name: IEEE Journal of Biomedical and Health Informatics, ISSN: 2168-2208. DOI: 10.1109/JBHI.2013.2257819.

[41] T. Okamoto, T. Ohnishi, H. Kawahira, O. Dergachyava, P. Jannin, and H. Haneishi, "Real-time identification of blood regions for hemostasis support in laparoscopic surgery," *Signal, Image and Video Processing*, vol. 13, no. 2, pp. 405–412, Mar. 1, 2019, ISSN: 1863-1711. DOI: 10.1007/s11760-018-1369-7. [Online]. Available: https://doi.org/10.1007/s11760-018-1369-7.

[42] F. Richter, S. Shen, F. Liu, J. Huang, E. K. Funk, R. K. Orosco, and M. C. Yip, "Autonomous Robotic Suction to Clear the Surgical Field for Hemostasis using Image-based Blood Flow Detection," *arXiv:2010.08441 [cs]*, Jan. 2021, arXiv: 2010.08441. [Online]. Available: http://arxiv.org/abs/2010.08441.

[43] C. Schenck and D. Fox, "Perceiving and reasoning about liquids using fully convolutional networks," *The International Journal of Robotics Research*, vol. 37, no. 4, pp. 452–471, Apr. 1, 2018, Publisher: SAGE Publications Ltd STM, ISSN: 0278-3649. DOI: 10.1177/0278364917734052. [Online]. Available: https://doi.org/10.1177/0278364917734052.

[44] A. Yamaguchi and C. G. Atkeson, "Stereo vision of liquid and particle flow for robot pouring," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, ISSN: 2164-0580, Nov. 2016, pp. 1173–1180. DOI: 10.1109/HUMANOIDS.2016.7803419.

[45] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, Jun. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[46] G.-Z. Yang, J. Cambias, K. Cleary, E. Daimler, J. Drake, P. E. Dupont, N. Hata, P. Kazanzides, S. Martel, R. V. Patel, V. J. Santos, and R. H. Taylor, "Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy," *Science Robotics*, vol. 2, no. 4, Mar. 15, 2017, Publisher: Science Robotics Section: Editorial, ISSN: 2470-9476. DOI: 10.1126/scirobotics.aam8638. [Online]. Available: https://robotics-sciencemag-org.ezproxy.lib.purdue.edu/content/2/4/eaam8638.

[47] M. G. Jacob, Y.-T. Li, G. A. Akingba, and J. P. Wachs, "Collaboration with a robotic scrub nurse," *Communications of the ACM*, vol. 56, no. 5, pp. 68–75, May 1, 2013, ISSN: 0001-0782. DOI: 10.1145/2447976.2447993. [Online]. Available: http://doi.org/10.1145/2447976.2447993.

[48] M. Yip and N. Das, "Robot autonomy for surgery," *arXiv:1707.03080 [cs]*, Jul. 10, 2017. arXiv: 1707.03080. [Online]. Available: http://arxiv.org/abs/1707.03080.

[49] M. M. Rahman, M. V. Balakuntala, G. Gonzalez, M. Agarwal, U. Kaur, V. L. N. Venkatesh, N. Sanchez-Tamayo, Y. Xue, R. M. Voyles, V. Aggarwal, and J. Wachs, "SARTRES: A semi-autonomous robot teleoperation environment for surgery," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 0, no. 0, pp. 1–8, Nov. 5, 2020, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21681163.2020.1834878, ISSN: 2168-1163. DOI: 10.1080/21681163.2020.1834878. [Online]. Available: https://doi.org/10.1080/21681163.2020.1834878.

[50] S. Panesar, Y. Cagle, D. Chander, J. Morey, J. Fernandez-Miranda, and M. Kliot, "Artificial intelligence and the future of surgical robotics," *Annals of Surgery*, vol. 270, no. 2, pp. 223–226, Aug. 2019, ISSN: 0003-4932. DOI: 10.1097/SLA.0000000000003262. [Online]. Available: https://journals.lww.com/annalsofsurgery/Fulltext/2019/08000/Artificial_Intelligence_and_the_Future_of_Surgical.7.aspx?casa_token=GmGJM22ETLwAAAAA:JQ_xRVkouKEImK_risrnXkSvc2_xFJSZfwHssc2U0LJO4lVpq8sV

[51] R. Moccia, C. Iacono, B. Siciliano, and F. Ficuciello, "Vision-based dynamic virtual fixtures for tools collision avoidance in robotic surgery," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1650–1655, Apr. 2020, Conference Name: IEEE Robotics and Automation Letters, ISSN: 2377-3766. DOI: 10.1109/LRA.2020.2969941.

[52] L. Qian, J. Y. Wu, S. P. DiMaio, N. Navab, and P. Kazanzides, "A review of augmented reality in robotic-assisted surgery," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 1, pp. 1–16, Feb. 2020, Conference Name: IEEE Transactions on Medical Robotics and Bionics, ISSN: 2576-3202. DOI: 10.1109/TMRB.2019.2957061.

[53] I. E. Rassi and J.-M. E. Rassi, "A review of haptic feedback in tele-operated robotic surgery," *Journal of Medical Engineering & Technology*, vol. 44, no. 5, pp. 247–254, Jul. 3, 2020, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/03091902.2020.17723 ISSN: 0309-1902. DOI: 10.1080/03091902.2020.1772391. [Online]. Available: https://doi.org/10.1080/03091902.2020.1772391.

[54] A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, and P. Valdastri, "Autonomy in surgical robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 651–679, 2021, _eprint: https://doi.org/10.1146/annurev-control-062420-090543. DOI: 10.1146/annurev-control-062420-090543. [Online]. Available: https://doi.org/10.1146/annurev-control-062420-090543.

[55] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, "Imitation learning: A survey of learning methods," *ACM Computing Surveys*, vol. 50, no. 2, 21:1–21:35, Apr. 6, 2017, ISSN: 0360-0300. DOI: 10.1145/3054912. [Online]. Available: http://doi.org/10.1145/3054912.

[56] M. M. Rahman, N. Sanchez-Tamayo, G. Gonzalez, M. Agarwal, V. Aggarwal, R. M. Voyles, Y. Xue, and J. Wachs, "Transferring dexterous surgical skill knowledge between robots for semi-autonomous teleoperation," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, ISSN: 1944-9437, Oct. 2019, pp. 1–6. DOI: 10.1109/RO-MAN46459.2019.8956396.

[57] M. Hwang, B. Thananjeyan, D. Seita, J. Ichnowski, S. Paradis, D. Fer, T. Low, and K. Goldberg, "Superhuman Surgical Peg Transfer Using Depth-Sensing and Deep Recurrent Neural Networks," *arXiv:2012.12844 [cs]*, Dec. 2020, arXiv: 2012.12844. [Online]. Available: http://arxiv.org/abs/2012.12844.

[58] C. D'Ettorre, G. Dwyer, X. Du, F. Chadebecq, F. Vasconcelos, E. De Momi, and D. Stoyanov, "Automated pick-up of suturing needles for robotic surgical assistance," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, ISSN: 2577-087X, May 2018, pp. 1370–1377. DOI: 10.1109/ICRA.2018.8461200.

[59] P. Sundaresan, B. Thananjeyan, J. Chiu, D. Fer, and K. Goldberg, "Automated extraction of surgical needles from tissue phantoms," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, ISSN: 2161-8089, Aug. 2019, pp. 170–177. DOI: 10.1109/COASE.2019.8843089.

[60] T. D. Nagy, M. Takács, I. J. Rudas, and T. Haidegger, "Surgical subtask automation—soft tissue retraction," in *2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, IEEE, 2018, pp. 000 055–000 060.

[61] R. Elek, T. D. Nagy, D. Á. Nagy, T. Garamvölgyi, B. Takács, P. Galambos, J. K. Tar, I. J. Rudas, and T. Haidegger, "Towards surgical subtask automation — blunt dissection," in *2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES)*, Oct. 2017, pp. 000 253–000 258. DOI: 10.1109/INES.2017.8118565.

[62] W. J. Hyung and Y. Woo, "TilePro," in *Robotics in General Surgery*, K. C. Kim, Ed., New York, NY: Springer, 2014, pp. 457–460, ISBN: 978-1-4614-8739-5. DOI: 10.1007/978-1-4614-8739-5_35. [Online]. Available: https://doi.org/10.1007/978-1-4614-8739-5_35.

[63] O. Mohareri, G. Nir, J. Lobo, R. Savdie, P. Black, and S. Salcudean, "A system for MR-ultrasound guidance during robot-assisted laparoscopic radical prostatectomy," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2015, pp. 497–504, ISBN: 978-3-319-24553-9. DOI: 10.1007/978-3-319-24553-9_61.

[64] L. Qian, J. Y. Wu, S. P. DiMaio, N. Navab, and P. Kazanzides, "A review of augmented reality in robotic-assisted surgery," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 1, pp. 1–16, Feb. 2020, Conference Name: IEEE Transactions on Medical Robotics and Bionics, ISSN: 2576-3202. DOI: 10.1109/TMRB.2019.2957061.

[65] D. Lee, H. W. Yu, S. Kim, J. Yoon, K. Lee, Y. J. Chai, J. Y. Choi, H.-J. Kong, K. E. Lee, H. S. Cho, and H. C. Kim, "Vision-based tracking system for augmented reality to localize recurrent laryngeal nerve during robotic thyroid surgery," *Scientific Reports*, vol. 10, no. 1, p. 8437, May 21, 2020, Number: 1 Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10.1038/s41598-020-65439-6. [Online]. Available: https://www.nature.com/articles/s41598-020-65439-6.

[66] P. Gilling, R. Reuther, A. Kahokehr, and M. Fraundorfer, "Aquablation – image-guided robot-assisted waterjet ablation of the prostate: Initial clinical experience," *BJU International*, vol. 117, no. 6, pp. 923–929, Jun. 1, 2016, Publisher: John Wiley & Sons, Ltd, ISSN: 1464-4096. DOI: 10.1111/bju.13358. [Online]. Available: http://bjui.journals.onlinelibrary.wiley.com/doi/full/10.1111/bju.13358.

[67] K. E. Kaplan, K. A. Nichols, and A. M. Okamura, "Toward human-robot collaboration in surgery: Performance assessment of human and robotic agents in an inclusion segmentation task," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 723–729. DOI: 10.1109/ICRA.2016.7487199.

[68] N. Padoy and G. D. Hager, "Human-machine collaborative surgery using learned models," in *2011 IEEE International Conference on Robotics and Automation*, ISSN: 1050-4729, May 2011, pp. 5285–5292. DOI: 10.1109/ICRA.2011.5980250.

[69] S. V. Kotsis and K. C. Chung, "Application of see one, do one, teach one concept in surgical training," *Plastic and reconstructive surgery*, vol. 131, no. 5, pp. 1194–1201, May 2013, ISSN: 0032-1052. DOI: 10.1097/PRS.0b013e318287a0b3. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4785880/.

[70] J. Vozenilek, J. S. Huff, M. Reznek, and J. A. Gordon, "See one, do one, teach one: Advanced technology in medical education," *Academic Emergency Medicine*, vol. 11, no. 11, pp. 1149–1154, 2004, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1197/j.aem.20 ISSN: 1553-2712. DOI: 10.1197/j.aem.2004.08.003. [Online]. Available: http://onlinelibrary.wiley.com/doi/abs/10.1197/j.aem.2004.08.003.

[71] J. M. Rodriguez-Paz, M. Kennedy, E. Salas, A. W. Wu, J. B. Sexton, E. A. Hunt, and P. J. Pronovost, "Beyond "see one, do one, teach one": Toward a different training paradigm," *BMJ Quality & Safety*, vol. 18, no. 1, pp. 63–68, Feb. 1, 2009, Publisher: BMJ Publishing Group Ltd Section: Education and training, ISSN: 2044-5415, 2044-5423. DOI: 10.1136/qshc.2007.023903. [Online]. Available: https://qualitysafety.bmj.com/content/18/1/63.

[72] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surgical Endoscopy*, vol. 25, no. 2, pp. 356–366, Feb. 1, 2011, ISSN: 1432-2218. DOI: 10.1007/s00464-010-1190-z. [Online]. Available: https://doi.org/10.1007/s00464-010-1190-z.

[73] A. Darzi, S. Smith, and N. Taffinder, "Assessing operative skill: Needs to become more objective," *BMJ*, vol. 318, no. 7188, pp. 887–888, Apr. 3, 1999, Publisher: British Medical Journal Publishing Group Section: Editorial, ISSN: 0959-8138, 1468-5833. DOI: 10.1136/bmj.318.7188.887. [Online]. Available: https://www.bmj.com/content/318/7188/887.

[74] J. A. Martin, G. Regehr, R. Reznick, H. Macrae, J. Murnaghan, C. Hutchison, and M. Brown, "Objective structured assessment of technical skill (OSATS) for surgical residents: OBJECTIVE STRUCTURED ASSESSMENT OF TECHNICAL SKILL," *British Journal of Surgery*, vol. 84, no. 2, pp. 273–278, Feb. 1997, ISSN: 00071323. DOI: 10.1046/j.1365-2168.1997.02502.x. [Online]. Available: http://doi.wiley.com/10.1046/j.1365-2168.1997.02502.x.

[75] R. Kumar, A. Jog, B. Vagvolgyi, H. Nguyen, G. Hager, C. C. G. Chen, and D. Yuh, "Objective measures for longitudinal assessment of robotic surgery training," en, *The Journal of Thoracic and Cardiovascular Surgery*, vol. 143, no. 3, pp. 528–534, Mar. 2012, ISSN: 0022-5223. DOI: 10.1016/j.jtcvs.2011.11.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022522311012748.

[76] C. R. Wottawa, B. Genovese, B. N. Nowroozi, S. D. Hart, J. W. Bisley, W. S. Grundfest, and E. P. Dutson, "Evaluating tactile feedback in robotic surgery for potential clinical application using an animal model," *Surgical Endoscopy*, vol. 30, no. 8, pp. 3198–3209, Aug. 1, 2016, ISSN: 1432-2218. DOI: 10.1007/s00464-015-4602-2. [Online]. Available: https://doi.org/10.1007/s00464-015-4602-2.

[77] F. A. Haji, D. Rojas, R. Childs, S. d. Ribaupierre, and A. Dubrowski, "Measuring cognitive load: Performance, mental effort and simulation task complexity," *Medical Education*, vol. 49, no. 8, pp. 815–827, 2015, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/medu.12773, ISSN: 1365-2923. DOI: 10.1111/medu.12773. [Online]. Available: http://onlinelibrary.wiley.com/doi/abs/10.1111/medu.12773.

[78] Y. Y. Yurko, M. W. Scerbo, A. S. Prabhu, C. E. Acker, and D. Stefanidis, "Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool," *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, vol. 5, no. 5, pp. 267–271, Oct. 2010, ISSN: 1559-713X. DOI: 10.1097/SIH.0b013e3181e3f329.

[79] H. A. Maior, M. L. Wilson, and S. Sharples, "Workload alerts—using physiological measures of mental workload to provide feedback during tasks," *ACM Transactions on Computer-Human Interaction*, vol. 25, no. 2, 9:1–9:30, Apr. 11, 2018, ISSN: 1073-0516. DOI: 10.1145/3173380. [Online]. Available: https://doi.org/10.1145/3173380.

[80] B. R. Lowndes, K. L. Forsyth, R. C. Blocker, P. G. Dean, M. J. Truty, S. F. Heller, S. Blackmon, M. S. Hallbeck, and H. Nelson, "NASA-TLX assessment of surgeon workload variation across specialties," *Annals of Surgery*, vol. 271, no. 4, pp. 686–692, Apr. 2020, ISSN: 0003-4932. DOI: 10.1097/SLA.0000000000003058. [Online]. Available: https://journals.lww.com/annalsofsurgery/Fulltext/2020/04000/NASA_TLX_Assessment_of_Surgeon_Workload_Variation.16.aspx?context=LatestArticles.

[81] C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau, G. Davis, V. T. Zivkovic, R. E. Olmstead, P. D. Tremoulet, and P. L. Craven. (May 2007). "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks." Library Catalog: www-ingentaconnect-com.ezproxy.lib.purdue.edu Publisher: Aerospace Medical Association, [Online]. Available: https://www-ingentaconnect-com.ezproxy.lib.purdue.edu/content/asma/asem/2007/00000078/A00105s1/art00032.

[82] K. A. Guru, E. T. Esfahani, S. J. Raza, R. Bhat, K. Wang, Y. Hammond, G. Wilding, J. O. Peabody, and A. J. Chowriappa, "Cognitive skills assessment during robot-assisted surgery: Separating the wheat from the chaff," *BJU International*, vol. 115, no. 1, pp. 166–174, 2015, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bju.12657, ISSN: 1464-410X. DOI: 10.1111/bju.12657. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/bju.12657.

[83] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, 2012, Publisher: Frontiers, ISSN: 1662-453X. DOI: 10.3389/fnins.2012.00039. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2012.00039/full.

[84] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, Feb. 2016, Conference Name: IEEE Computational Intelligence Magazine, ISSN: 1556-6048. DOI: 10.1109/MCI.2015.2501545.

[85] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "A subject-transfer framework for obviating inter- and intra-subject variability in EEG-based drowsiness detection," *NeuroImage*, vol. 174, pp. 407–419, Jul. 1, 2018, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2018.03.032. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811918302428.

[86] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *Journal of Neural Engineering*, vol. 16, no. 3, p. 031 001, Apr. 2019, Publisher: IOP Publishing, ISSN: 1741-2552. DOI: 10.1088/1741-2552/ab0ab5. [Online]. Available: https://doi.org/10.1088/1741-2552/ab0ab5.

[87] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017, ISSN: 10659471. DOI: 10.1002/hbm.23730. arXiv: 1703.05051. [Online]. Available: http://arxiv.org/abs/1703.05051.

[88] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056 013, Jul. 2018, ISSN: 1741-2552. DOI: 10.1088/1741-2552/aace8c. [Online]. Available: https://doi.org/10.1088%2F1741-2552%2Faace8c.

[89] R. L. Charles and J. Nixon, "Measuring mental workload using physiological measures: A systematic review," *Applied Ergonomics*, vol. 74, pp. 221–232, Jan. 2019, ISSN: 00036870. DOI: 10.1016/j.apergo.2018.08.028. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0003687018303430.

[90] R. Hefron, B. Borghetti, C. Schubert Kabban, J. Christensen, and J. Estepp, "Cross-participant EEG-based assessment of cognitive workload using multi-path convolutional recurrent neural networks," *Sensors*, vol. 18, no. 5, p. 1339, May 2018, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute. DOI: 10.3390/s18051339. [Online]. Available: https://www.mdpi.com/1424-8220/18/5/1339.

[91] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time fourier transformation and deep neural networks for motor imagery brain computer interface recognition," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 23, e4413, 2018, ISSN: 1532-0634. DOI: 10.1002/cpe.4413. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.4413.

[92] Z. Jiao, X. Gao, Y. Wang, J. Li, and H. Xu, "Deep convolutional neural networks for mental load classification based on EEG data," *Pattern Recognition*, vol. 76, pp. 582–595, Apr. 1, 2018, ISSN: 0031-3203. DOI: 10.1016/j.patcog.2017.12.002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320317304879.

[93] P. Antonenko, F. Paas, R. Grabner, and T. van Gog, "Using electroencephalography to measure cognitive load," *Educational Psychology Review*, vol. 22, no. 4, pp. 425–438, Dec. 1, 2010, ISSN: 1573-336X. DOI: 10.1007/s10648-010-9130-y. [Online]. Available: https://doi.org/10.1007/s10648-010-9130-y.

[94] M. Uemura, M. Yamashita, M. Tomikawa, S. Obata, R. Souzaki, S. Ieiri, K. Ohuchida, N. Matsuoka, T. Katayama, and M. Hashizume, "Objective assessment of the suture ligature method for the laparoscopic intestinal anastomosis model using a new computerized system," en, *Surgical Endoscopy*, vol. 29, no. 2, pp. 444–452, Feb. 2015, ISSN: 1432-2218. DOI: 10.1007/s00464-014-3681-9. [Online]. Available: https://doi.org/10.1007/s00464-014-3681-9.

[95] M. Shah, "Solving the Robot-World/Hand-Eye Calibration Problem Using the Kronecker Product," *Journal of Mechanisms and Robotics*, vol. 5, no. 031007, Jun. 2013, ISSN: 1942-4302. DOI: 10.1115/1.4024473. [Online]. Available: https://doi.org/10.1115/1.4024473.

[96] J. A. Marvel and K. V. Wyk, "Simplified framework for robot coordinate registration for manufacturing applications," in *2016 IEEE International Symposium on Assembly and Manufacturing (ISAM)*, Aug. 2016, pp. 56–63. DOI: 10.1109/ISAM.2016.7750718.

[97] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556 [cs]*, Apr. 2015, arXiv: 1409.1556. [Online]. Available: http://arxiv.org/abs/1409.1556.

[98] (2020). "SuperAnnotate," SuperAnnotate, [Online]. Available: https://superannotate.com/download.

[99] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 6, 2019, ISSN: 2196-1115. DOI: 10.1186/s40537-019-0197-0. [Online]. Available: https://doi.org/10.1186/s40537-019-0197-0.

[100] A. Jung, *Aleju/imgaug*, original-date: 2015-07-10T20:31:33Z, May 29, 2021. [Online]. Available: https://github.com/aleju/imgaug.

[101] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[102] M. Attia, I. Hettiarachchi, S. Mohamed, M. Hossny, and S. Nahavandi, "A frequency domain classifier of steady-state visual evoked potentials using deep separable convolutional neural networks," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, ISSN: 2577-1655, Oct. 2018, pp. 2134–2139. DOI: 10.1109/SMC.2018.00367.

[103] C. P. Amaral, M. A. Simões, S. Mouga, J. Andrade, and M. Castelo-Branco, "A novel brain computer interface for classification of social joint attention in autism and comparison of 3 experimental setups: A feasibility study," *Journal of Neuroscience Methods*, vol. 290, pp. 105–115, Oct. 1, 2017, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2017.07.029. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027017302728.

[104] A. Chan, C. E. Early, S. Subedi, Yuezhe Li, and H. Lin, "Systematic analysis of machine learning algorithms on EEG data for brain state intelligence," in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Nov. 2015, pp. 793–799. DOI: 10.1109/BIBM.2015.7359788.

[105] Z. Yin and J. Zhang, "Cross-session classification of mental workload levels using EEG and an adaptive deep learning model," *Biomedical Signal Processing and Control*, vol. 33, pp. 30–47, Mar. 1, 2017, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2016.11.013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809416302014.

[106] P. Maggi and F. Di Nocera, "Sensitivity of the spatial distribution of fixations to variations in the type of task demand and its relation to visual entropy," *Frontiers in Human Neuroscience*, vol. 15, 2021, Publisher: Frontiers, ISSN: 1662-5161. DOI: 10.3389/fnhum.2021.642535. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2021.642535/full.

[107] P. Qvarfordt and S. Zhai. (2009). "Gaze-aided human-computer and human-human dialogue," Handbook of Research on Socio-Technical Design and Social Networking Systems. ISBN: 9781605662640 Pages: 529-543 Publisher: IGI Global, [Online]. Available: www.igi-global.com/chapter/gaze-aided-human-computer-human/21431.

[108] C. Kothe, *Sccn/labstreaminglayer*, original-date: 2018-02-28T10:50:12Z, Jun. 12, 2021. [Online]. Available: https://github.com/sccn/labstreaminglayer.

[109] F. Chollet *et al.*, *Keras*. 2015. [Online]. Available: https://keras.io.

[110] S. Hochreiter and J. Schmidhuber, "LSTM can solve hard long time lag problems," in *Advances in Neural Information Processing Systems 9*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds., MIT Press, 1997, pp. 473–479. [Online]. Available: http://papers.nips.cc/paper/1215-lstm-can-solve-hard-long-time-lag-problems.pdf.

[111] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980 [cs]*, Jan. 29, 2017. arXiv: 1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980.

[112] J. Barragan, D. Chanci, D. Yu, and J. Wachs, "SACHETS: Semi-autonomous cognitive hybrid emergency teleoperated suctio," presented at the International Symposium on Robot and Human Interactive Communication.

[113] J. H. Nguyen, J. Chen, S. P. Marshall, S. Ghodoussipour, A. Chen, I. S. Gill, and A. J. Hung, "Using objective robotic automated performance metrics and task-evoked pupillary response to distinguish surgeon expertise," en, *World Journal of Urology*, vol. 38, no. 7, pp. 1599–1605, Jul. 2020, ISSN: 1433-8726. DOI: 10.1007/s00345-019-02881-w. [Online]. Available: https://doi.org/10.1007/s00345-019-02881-w.

[114] A. J. Hung, P. J. Oh, J. Chen, S. Ghodoussipour, C. Lane, A. Jarc, and I. S. Gill, "Experts vs super-experts: Differences in automated performance metrics and clinical outcomes for robot-assisted radical prostatectomy," eng, *BJU international*, vol. 123, no. 5, pp. 861–868, May 2019, ISSN: 1464-410X. DOI: 10.1111/bju.14599.

[115] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," en, in *Advances in Psychology*, ser. Human Mental Workload, P. A. Hancock and N. Meshkati, Eds., vol. 52, North-Holland, Jan. 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166411508623869.

[116] T. Zhou, J. S. Cha, G. Gonzalez, J. P. Wachs, C. P. Sundaram, and D. Yu, "Multimodal physiological signals for workload prediction in robot-assisted surgery," *ACM Transactions on Human-Robot Interaction*, vol. 9, no. 2, 12:1–12:26, Jan. 25, 2020. DOI: 10.1145/3368589. [Online]. Available: http://doi.org/10.1145/3368589.

[117] C. Wu, J. S. Cha, J. Sulek, T. Zhou, C. P. Sundaram, and D. Yu, "Analysis of eye behavior: Mental workload assessment in robotic surgery training," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1, pp. 1920–1921, Nov. 1, 2019, Publisher: SAGE Publications Inc, ISSN: 2169-5067. DOI: 10.1177/1071181319631408. [Online]. Available: https://doi.org/10.1177/1071181319631408.

[118] I. R. Crothers, A. G. Gallagher, N. McClure, D. T. D. James, and J. McGuigan, "Experienced laparoscopic surgeons are automated to the "fulcrum effect": An ergonomic demonstration," *Endoscopy*, vol. 31, no. 5, pp. 365–369, Jun. 1999, Publisher: Georg Thieme Verlag Stuttgart · New York, ISSN: 0013-726X, 1438-8812. DOI: 10.1055/s-1999-26. [Online]. Available: http://www.thieme-connect.de/DOI/DOI?10.1055/s-1999-26.

[119] R. K. Lange, B. Godde, and C. Braun, "EEG correlates of coordinate processing during intermanual transfer," *Experimental Brain Research*, vol. 159, no. 2, pp. 161–171, Nov. 1, 2004, ISSN: 1432-1106. DOI: 10.1007/s00221-004-1942-x. [Online]. Available: https://doi.org/10.1007/s00221-004-1942-x.

[120] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," *arXiv:1511.06448 [cs]*, Feb. 29, 2016. arXiv: 1511.06448. [Online]. Available: http://arxiv.org/abs/1511.06448.

[121] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Reviews Neuroscience*, vol. 3, no. 3, pp. 201–215, Mar. 2002, Bandiera_abtest: a Cg_type: Nature Research Journals Number: 3 Primary_atype: Reviews Publisher: Nature Publishing Group, ISSN: 1471-0048. DOI: 10.1038/nrn755. [Online]. Available: https://www.nature.com/articles/nrn755.

[122] T. Harmony, "The functional significance of delta oscillations in cognitive processing," *Frontiers in Integrative Neuroscience*, vol. 7, Dec. 5, 2013, ISSN: 1662-5145. DOI: 10.3389/fnint.2013.00083. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3851789/.

[123] P. Zarjam, J. Epps, N. H. Lovell, and F. Chen, "Characterization of memory load in an arithmetic task using non-linear analysis of EEG signals," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, ISSN: 1558-4615, Aug. 2012, pp. 3519–3522. DOI: 10.1109/EMBC.2012.6346725.

[124] P. Zarjam, J. Epps, and Fang Chen, "Characterizing working memory load using EEG delta activity," in *2011 19th European Signal Processing Conference*, ISSN: 2076-1465, Aug. 2011, pp. 1554–1558.

[125] P. Sauseng, B. Griesmayr, R. Freunberger, and W. Klimesch, "Control mechanisms in working memory: A possible function of EEG theta oscillations," *Neuroscience & Biobehavioral Reviews*, Binding Processes: Neurodynamics and Functional Role in Memory and Action, vol. 34, no. 7, pp. 1015–1022, Jun. 1, 2010, ISSN: 0149-7634. DOI: 10.1016/j.neubiorev.2009.12.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149763409001973.

[126] S. Puma, N. Matton, P.-V. Paubel, É. Raufaste, and R. El-Yagoubi, "Using theta and alpha band power to assess cognitive workload in multitasking environments," *International Journal of Psychophysiology*, vol. 123, pp. 111–120, Jan. 1, 2018, ISSN: 0167-8760. DOI: 10.1016/j.ijpsycho.2017.10.004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167876016307930.

[127] J. J. Moles, P. E. Connelly, E. E. Sarti, and S. Baredes, "Establishing a training program for residents in robotic surgery," *The Laryngoscope*, vol. 119, no. 10, pp. 1927–1931, 2009, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/lary.20508, ISSN: 1531-4995. DOI: 10.1002/lary.20508. [Online]. Available: http://onlinelibrary.wiley.com/doi/abs/10.1002/lary.20508.

[128] S. Roy, A. Siarohin, E. Sangineto, S. R. Bulo, N. Sebe, and E. Ricci, "Unsupervised domain adaptation using feature-whitening and consensus loss," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9471–9480. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2019/html/Roy_Unsupervised_Domain_Adaptation_Using_Feature-Whitening_and_Consensus_Loss_CVPR_2019_paper.html.

[129] G. Hoffman, "Evaluating fluency in human–robot collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, Jun. 2019, ISSN: 2168-2291, 2168-2305. DOI: 10.1109/THMS.2019.2904558. [Online]. Available: https://ieeexplore.ieee.org/document/8678448/.

[130] F. Setti, E. Oleari, A. Leporini, D. Trojaniello, A. Sanna, U. Capitanio, F. Montorsi, A. Salonia, and R. Muradore, "A multirobots teleoperated platform for artificial intelligence training data collection in minimally invasive surgery," in *2019 International Symposium on Medical Robotics (ISMR)*, Apr. 2019, pp. 1–7. DOI: 10.1109/ISMR.2019.8710209.

[131] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," en, *The International Journal of Robotics Research*, p. 0 278 364 919 842 925, Apr. 2019, ISSN: 0278-3649. DOI: 10.1177/0278364919842925. [Online]. Available: https://doi.org/10.1177/0278364919842925.

[132] T. Sanders, K. E. Oleson, D. Billings, J. Y. Chen, and P. Hancock, "A Model of Human-Robot Trust: Theoretical Model Development," en, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 55, no. 1, pp. 1432–1436, Sep. 2011, Publisher: SAGE Publications Inc, ISSN: 2169-5067. DOI: 10.1177/1071181311551298. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/1071181311551298.

[133] F. Richter, Y. Zhang, Y. Zhi, R. K. Orosco, and M. C. Yip, "Augmented Reality Predictive Displays to Help Mitigate the Effects of Delayed Telesurgery," *arXiv:1809.08627 [cs]*, Feb. 2019, arXiv: 1809.08627. [Online]. Available: http://arxiv.org/abs/1809.08627.

[134] D. Nicolis, M. Palumbo, A. M. Zanchettin, and P. Rocco, "Occlusion-Free Visual Servoing for the Shared Autonomy Teleoperation of Dual-Arm Robots," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 796–803, Apr. 2018, Conference Name: IEEE Robotics and Automation Letters, ISSN: 2377-3766. DOI: 10.1109/LRA.2018.2792143.

[135] G. G. M. Dalveren and N. E. Cagiltay, "Insights from surgeons' eye-movement data in a virtual simulation surgical training environment: Effect of experience level and hand conditions," *Behaviour & Information Technology*, vol. 37, no. 5, pp. 517–537, May 4, 2018, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0144929X.2018.1460399, ISSN: 0144-929X. DOI: 10.1080/0144929X.2018.1460399. [Online]. Available: https://doi.org/10.1080/0144929X.2018.1460399.

VITA

Juan Antonio Barragan Noguera
School of Industrial Engineering, Purdue University

Education

B.S., Electronic Engineering, 2013, Universidad Nacional de Colombia, Colombia.
M.S., Industrial Engineering, 2021, Purdue University, West Lafayette IN, USA

Research Interests

Human-robotic interaction, Robotic-assisted surgery, Autonomous systems, Computer vision

PUBLICATION(S)

Barragan, J., Chanci, D., Yu, D., & Wachs, J. (n.d.). SACHETS: Semi-Autonomous Cognitive Hybrid Emergency Teleoperated Suction. International Symposium on Robot and Human Interactive Communication.

Lin, C., Rojas-Muñoz, E., Cabrera, M. E., Sanchez-Tamayo, N., Andersen, D., Popescu, V., **Barragan, J. A.**, Zarzaur, B., Murphy, P., Anderson, K., Douglas, T., Griffis, C., & Wachs, J. (2020). How About the Mentor? Effective Workspace Visualization in AR Telementoring. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 212–220. https://doi.org/10.1109/VR46266.2020.00040

Lin, C., Rojas-Muñoz, E., Cabrera, M. E., Sanchez-Tamayo, N., Andersen, D., Popescu, V., **Barragan, J. A.**, Zarzaur, B., Murphy, P., Anderson, K., Douglas, T., Griffis, C., & Wachs, J. (2019). Robust High-Level Video Stabilization for Effective AR Telementoring. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 1038–1039. https://doi.org/10.1109/VR.2019.8798331

Madapana, N., Rahman, M. M., Sanchez-Tamayo, N., Balakuntala, M. V., Gonzalez, G., Bindu, J. P., Vishnunandan Venkatesh, L. N., Zhang, X., **Barragan, J. A.**, Low, T., Voyles, R. M., Xue, Y., & Wachs, J. (2019). DESK: A Robotic Activity Dataset for Dexterous Surgical Skills Transfer to Medical Robots. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 6928–6934. https://doi.org/10.1109/IROS40897.2019.8967760

Rojas-Muñoz, E., Lin, C., Sanchez-Tamayo, N., Cabrera, M. E., Andersen, D., Popescu, V., **Barragan, J. A.**, Zarzaur, B., Murphy, P., Anderson, K., Douglas, T., Griffis, C., McKee, J., Kirkpatrick, A. W., & Wachs, J. P. (2020). Evaluation of an augmented reality platform for austere surgical telementoring: A randomized controlled crossover study in cricothyroidotomies. Npj Digital Medicine, 3(1), 1–9. https://doi.org/10.1038/s41746-020-0284-9