Design and Development of LapBot: An Interactive Mobile Game for Mastering Safe Laparoscopic Cholecystectomy

Mohammad Noroozi

A Thesis

in

The Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Computer Science () at

Concordia University

Montréal, Québec, Canada

September 2023

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Mohammad Noroozi

Entitled: Design and Development of LapBot: An Interactive Mobile Game for Mastering Safe Laparoscopic Cholecystectomy

and submitted in partial fulfillment of the requirements for the degree of

Master of Computer Science ()

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

______________________________ Chair
Dr. Tiberiu Popa

______________________________ External Examiner
Dr. Tiberiu Popa

______________________________ Examiner
Dr. Sudhir Mudur

______________________________ Supervisor
Dr. Marta Kersten-Oertel

Approved by ______________________________
Dr. Joey Paquet, Chair
Department of Computer Science and Software Engineering

__________ 2023 ______________________________
Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Design and Development of LapBot: An Interactive Mobile Game for Mastering Safe Laparoscopic Cholecystectomy

Mohammad Noroozi

Major bile duct injuries during laparoscopic cholecystectomy (LC) are a significant source of morbidity, mortality, disability, and healthcare costs. These injuries are primarily due to errors in surgical judgment and visual misperception of critical anatomy and tissue planes. To facilitate learning of safe LC we designed and developed LapBot Safe Chole, a novel mobile game integrating artificial intelligence (AI) feedback to enhance intraoperative decision-making during LC training.

LapBot Safe Chole offers an engaging learning experience through short video clips of LC scenarios. Users identify optimal dissection zones, with real-time AI-generated annotations delivering accuracy scores and immediate feedback. The game comprises five progressively challenging levels aligned with the Parkland grading scale. Progression to the next level necessitates over 50% accuracy across five consecutive responses.

Beta-testing (n = 22) results indicate improvement in game scores with each round, with attendings and senior trainees reaching top-scores earlier than junior residents per level. Our testing also showed that candidates can be distinguished by their learning curves and learning progression which can facilitate a competency-based curriculum. A statistically significant correlation (p=0.003) between user experience and score was observed. Furthermore, user feedback highlighted the game's ease of use (80% agreement) and its effectiveness in making learning enjoyable (100% agreement).

LapBot Safe Chole introduces and reinforces safe LC principles through an easily accessible and free gaming platform. Positive beta-testing outcomes suggest its potential adoption among surgical trainees. Future directions involve broader validation.

Acknowledgments

In the realm of acknowledgments, I extend my deepest gratitude to Dr. Marta Kersten-Oertel, my supervisor, whose unwavering support and guidance have been instrumental throughout my master's journey. Her mentorship not only encompassed the technical aspects of my thesis project but also extended to navigating the challenges that arose, turning her into not just a supervisor, but a compassionate friend.

I am profoundly indebted to Dr. Amin Madani for his invaluable insights and surgical expertise, which greatly contributed to the development of this project. The foundation laid by him and his accomplished team's data proved pivotal. I extend my thanks to Caterina Masino for her steadfast commitment to project management, ensuring its seamless progression. I am also indebted to Dr. Ace St. John for his instrumental role in data analysis, without which the insights gained would not have been as comprehensive.

I'm grateful to Andrey Titov for his invaluable assistance on the project's server side, and I extend my thanks to all AP Lab members for their support. Special mention goes to Donya Meshgin, Nadine El-Mufti, and David Xie for their collaborative efforts in crafting the game's user-friendly and engaging interface. On a more personal note, I offer heartfelt appreciation to my family. Their unwavering financial and emotional support enabled my pursuit of this academic journey abroad, a decision that was both challenging and fulfilling. To my friends, your steadfast encouragement and support over the past two years have been truly appreciated.

To all who have touched my journey, your contributions have been invaluable, and I am sincerely thankful for your roles in making this endeavor possible.

Contents

List of Figures vii

List of Tables ix

1 Introduction 2

1.1 Motivation 3

1.2 Contributions 3

1.3 Organization of Thesis 5

2 Background 6

2.1 Laparoscopic cholecystectomy 6

2.2 Bile Duct Injuries (BDIs) 7

2.3 The Role of Training and Simulation 8

2.4 Artificial Intelligence in Laparoscopic Cholecystectomy 11

2.5 AI in Surgical Education 14

2.6 Augmented Reality in Surgery 15

2.7 Think Like a Surgeon Application 16

2.8 Gamification 17

2.9 Summary 21

3 LapBot: Safe Chole 22

3.1 Introduction 22

3.1.1 Surgical Artificial Intelligence in Laparoscopic Cholecystectomy 23

3.1.2 Games in Surgical Education 23
3.2 Methods: LapBot Serious Game Design 25
3.2.1 Onboarding 25
3.2.2 Game Play 25
3.2.3 In-app Survey 30
3.3 Preliminary User Study 30
3.3.1 Study Participants 31
3.3.2 Game Level Evaluation 32
3.3.3 Player Confidence 33
3.3.4 Scores by Experience 34
3.3.5 Video watching rate 37
3.3.6 In-app survey results 38
3.3.7 External Survey 39
3.4 Conclusions 40

4 Conclusions 41
4.1 Summary 41
4.2 Future Work 42

Bibliography 44

List of Figures

Figure 2.1 Pre-operative incision sites (left) and the intra-operative laparoscopic view during cholecystectomy (right) [20],(Miguel A. Burch). 6

Figure 2.2 Photographs from laparoscopic cholecystectomy procedures [40]. 7

Figure 2.3 The Bismuth classification of Bile Duct Injuries (BDIs) categorizes them based on their distance from the bile duct bifurcation (Type 1–4) and includes individual right sectoral bile duct injury (type 5) [26]. 8

Figure 2.4 RoSS is a robotic surgery simulator with haptic feedback that offers an immersive experience with full-length surgical procedures in 3D [36]. 9

Figure 2.5 Simedo simulator: (A) Exercise "Sort the Rings," (B) exercise "Stretch and Transfer," (C) exercise "Ring and Rope," (D) Simendo VR laparoscopy simulator, (E) exercise "Balance," and (F) Exercise "Puzzle." [45]. 10

Figure 2.6 Comparing model predictions of GoNoGoNet with expert assessments, we provide an illustrative example of the model's predictions contrasted with the original frame (a). These comparisons are visually represented as overlays (b), wherein the Go zone is highlighted in green, and the No-Go zone is marked in red. Additionally, probability heat maps depicting the likelihood of Go (c) and No-Go zones (d) are presented for the model's predictions. To provide context, we include expert annotation heat maps for Go (e) and No-Go zones (f) as well. [24]. 12

Figure 2.7 Multi-Stage Temporal Convolution Network (MS-TCN) for LC stage recognition [17] 14

Figure 2.8 iSurgeon system 17

Figure 2.9 Screenshot of the "Think Like a Surgeon" platform is an interactive scenario where users are prompted to mark the point of initiation for dissecting the hepatocystic triangle. This exercise enables a comparative analysis between the annotations provided by users (depicted in green) and the annotations offered by experts (displayed as a heat map). [28] . . . 18
Figure 2.10 The Laparoscopic Dexterity Game Box . . . 19
Figure 2.11 Captured images from the Touch Surgery mobile platform on Laparoscopic Cholecystectomy, illustrating both the 'Learn' phase (A) and the 'Test' phase (B) [6] 21
Figure 3.1 LapBot Onboarding Screens . . . 26
Figure 3.2 LapBot GamePlay Screens . . . 27
Figure 3.3 The frame that facilitates the calculation of the user's score is shown. The left frame displays three distinctive colors, green, red, and blue, which denote the Go zones, No-Go zones, and neutral zones, respectively. The right frame provides a clearer visualization of these zones in a surgical setting. . . . 28
Figure 3.4 The LapBot app features three leader boards. . . . 29
Figure 3.5 LapBot GamePlay Screens . . . 30
Figure 3.6 Participants' Level of Training . . . 32
Figure 3.7 Levels difficulty chart . . . 33
Figure 3.8 Confidence rates chart . . . 34
Figure 3.9 Academic performance levels . . . 35
Figure 3.10 Learning Curve chart . . . 36
Figure 3.11 Similar pattern in expertise . . . 37
Figure 3.12 Analyzing user scores in descending order and the corresponding number of feedback videos watched in the game. . . . 38
Figure 3.13 survey graphs . . . 39

List of Tables

Contributions

I am the first author of an abstract accepted to and presented at the SAGES (Society of American Gastrointestinal and Endoscopic Surgeons) annual meeting in 2023 entitled: "LapBot Safe Chole": An Artificial Intelligence-Powered Mobile Game App to Teach Safe Cholecystectomy Amongst Surgical Trainees

- Authors: Mohammad Noroozi, Ace St. John, Caterina Masino, Amin Madani, Marta Kersten-Oertel.
- Contributions: Guarantors of integrity of the study: all authors; study and design concepts: all authors and M.K.-O.; software development: M. N.; data collection: M. N.; data prepara-tion and analysis: M.N. and A. St J.; supervision: M.K.-O.; manuscript preparation: M.N.; manuscript revision: all authors; editing and final version: all authors.

I am also the first author of the manuscript included in Chapter 3, "There's an app for that: Design and feasibility study of an education game app for laparoscopic cholecystectomy surgical trainees" To be submitted to Journal of Medical Internet Research (JMRI) Serious Games. The contributions of all authors are described below:

- Authors: Mohammad Noroozi, Ace St. John, Caterina Masino, Simon Laplante, Jaryd Hunter, Michael Brudno, Amin Madani, Marta Kersten-Oertel.
- Contributions: Guarantors of integrity of the study: all authors; study and design concepts: M.N., A. St J., C. M., A.M., and M.K.-O.; software development: M. N.; data collection: M. N.; AI data preparation S.L., J. H., M. B. and A. M.; game data preparation and analysis: M.N. and A. St J.; supervision: M.K.-O.; manuscript preparation: M.N.; manuscript revision: M.N., A. St J., C. M., A.M., and M.K.-O.; editing and final version: M.N., A. St J., C. M., A.M., and M.K.-O..

Chapter 1

Introduction

Laparoscopic cholecystectomy (LC), a minimally invasive procedure for removing the gallbladder, is one of the most commonly performed procedures in the United States and the world, with between 750,000 to 1 million cases per year in the United States alone [8]. Bile duct (i.e. a series of thin tubes that go from the liver to the small intestine) injuries or BDIs following LC are estimated to occur in about 1 in 1000 cases [21, 31]. BDIs are a significant cause of morbidity and short-term mortality. These injuries can have significant health-related consequences and substantially reduce patient quality of life.

Most complications and BDIs following LC result from perception errors, such as misinterpreting anatomical structures or misjudging safe planes of dissection [49]. For surgeons and trainees to attain a “critical view of safety”, they must undergo extensive dissections and apply complex knowledge and skills that can be integrated and adapted to specific clinical and patient scenarios. This can be particularly challenging for surgical trainees with limited practical experience or extensive exposure to diverse cases in the operating room. The traditional way for residents, especially those in their first years of training, to gain an understanding of surgery is to watch videos, read textbooks, and look at images. This method of teaching lacks experiential learning and has the disadvantage that learners are not challenged to view and interact with the operating field in a way that facilitates the development of a correct perception of safe dissection planes, which may have poor translation to real-life learning experiences [22].

1.1 Motivation

Surgical education and training face unique challenges to ensure a surgical trainee's procedural competency while guaranteeing patient safety. Many surgical trainees have limited hands-on experience particularly during their junior years. There is also limited opportunity for risk-taking during real-life exposure, which may affect a resident's understanding of the operative field. As LC is among the most common operations performed and one of the first surgeries performed by surgical trainees, quality-improvement initiatives are needed to address complications from bile duct issues [2, 3].

Recently, there has been a trend to include simulators and mobile games into medical curricula to enable experiential low risk learning. Using game-based learning can enhance student involvement, facilitate the willingness to take risks, and foster collaborative learning [34]. In surgical education in particular, serious gaming have been shown to be effective in enhancing intraoperative skills and decision-making.

Motivated by the high number of complications in LC and the shear volume of cases of this type of surgery, and the effectiveness of educational games in the clinical domain, our goal was to develop an innovative and interactive teaching tool that provided the means for deliberate practice using experiential learning and immediate feedback based on AI annotated videos.

To meet this goal we developed "LabBot: Safe Chole" or "LapBot" for short, as a tool to enhance trainees' understanding of dissection planes and anatomy in LC. Utilizing AI-generated feedback, the game establishes an interactive learning environment that reinforces safe dissection practices, thereby improving surgical decision-making skills. The integration of AI aims to provide expert surgical guidance and offers consistent feedback, elevating the learning experience.

1.2 Contributions

In the following thesis, we aimed to address the need to create an interactive teaching tool that allows for experiential learning. To do this we developed an innovative and fun mobile game that provides players with immediate AI-generated feedback to improve learning and retention of safe and unsafe dissection areas in laparoscopic cholecystectomy. Specifically, the thesis led to the following contributions:

(1) Design and development of a cross-platform (iOS and Android) mobile game for learning safe laparoscopic cholecystectomy

The integration of gamification elements within the LapBot: Safe Chole application presents an advancement in LC surgical education. By incorporating features such as scoring, leaderboards, animations, an intuitive interface, and effortless navigation, we aimed to provide users with an immersive and enjoyable learning experience. As users progress through various levels, they receive scores and are able to compete against other players. This scoring system, coupled with competitive leaderboards, encourages players to improve their performance and increases engagement. The app's user-friendly design, combined with real-world surgical scenarios, encourages active decision-making, while the integration of AI annotations as feedback enhances the learning process. In essence, the integration of game elements in LapBot aims to provide an exciting and interactive experiential learning environment.

(2) Novel Integration of AI-Generated Feedback in a Surgical Training Game

One of the contributions of this research lies in the integration of AI-generated feedback into a surgical training game. The utilization of AI-generated annotations as real-time feedback within the LapBot game aims to enhance learning. This integration helps trainees to practice in a dynamic, risk-free environment that mirrors real surgical scenarios, offering expert insights that contribute to safer and more proficient surgical practices. The AI models' ability to identify safe and unsafe zones draws from a large data set of expert annotations. The game serves as a bridge that translates AI insights into actionable guidance for trainees, promoting safer and more effective surgical techniques.

(3) Initial beta-testing of the LapBot game with medical students, residents and surgeons

We conducted a beta testing phase for the LapBot application, which is accessible on both iOS and Android platforms. Based on feedback and input from users including medical students, postgraduate students, and surgeons, minor adjustments were made to enhance the application, preparing it for its final version. The results of this testing showed the validity of the designed game play elements, interface choices, and AI feedback. Based on the development of the game and the initial testing, the project received funding from SAGES (Society of American Gastrointestinal and Endoscopic Surgeons) to conduct a comprehensive testing of the application with a larger number of medical students, residents, and surgeons. At the time of writing this thesis, the game has been played by more than 700 participants across 61 countries. Additionally, the research findings were presented as a poster at the SAGES conference and are to be submitted to the JMIR (Journal of Medical Internet Research) Serious Games for further dissemination.

1.3 Organization of Thesis

The subsequent sections of the thesis are structured as follows. In Chapter 2, we give background about laparoscopic cholecystectomy surgery, and describe surgical training for students, the significance of simulators in surgical education, recent advances of artificial intelligence integration in surgical education, and an exploration of gamification in surgical education. In Chapter 3, we describe the design of the "LapBot Safe Chole" game as well as the results of our beta-testing. Finally, in Chapter 4, we conclude the thesis and describe avenues of future work, specifically in regard to improving and adding features to the developed game.

Chapter 2

Background

2.1 Laparoscopic cholecystectomy

Laparoscopic cholecystectomy, a minimally invasive surgical procedure to remove the gallbladder, is one of the most commonly performed procedures in the United States and across the world [8]. The surgical procedure begins with the administration of general anesthesia to the patient. Subsequently, a few small incisions are made in the abdomen, through which specialized surgical instruments, including a laparoscope (i.e. a thin tube with a camera), are inserted. The laparoscope offers a magnified view of the surgical area, enabling the surgeon to navigate (by looking at the video on a monitor) and perform the procedure accurately (see Figure 2.1).

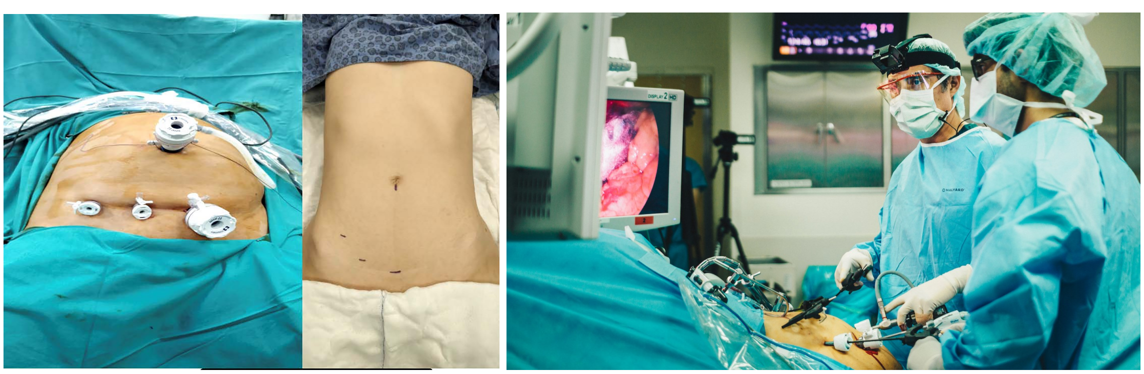


Figure 2.1: Pre-operative incision sites (left) and the intra-operative laparoscopic view during cholecystectomy (right) [20],(Miguel A. Burch).

Using these instruments, the surgeon dissects and separates the gallbladder from the surrounding tissues and structures while ensuring the preservation of vital organs such as the bile ducts and blood vessels. Once the gallbladder is completely freed, it is extracted through one of the small incisions. In certain cases, a drain may be placed near the surgical site to aid in fluid drainage as shown in Figure 2.2) [25].

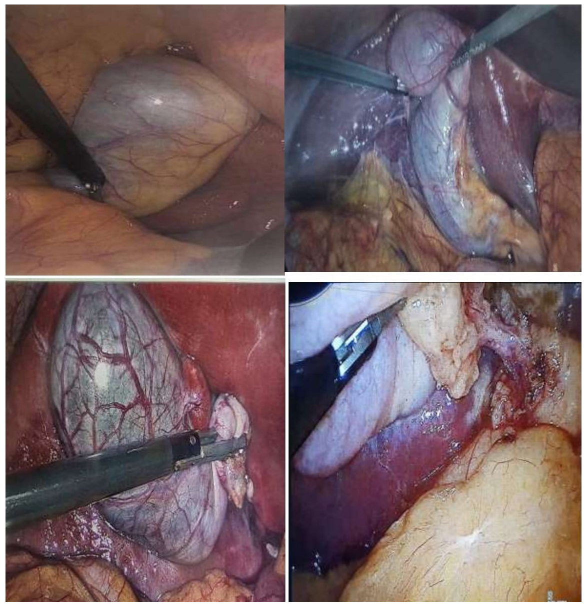


Figure 2.2: Photographs from laparoscopic cholecystectomy procedures [40].

2.2 Bile Duct Injuries (BDIs)

Laparoscopic cholecystectomy is now the preferred approach for managing symptomatic cholecystolithiasis. Although it offers advantages such as reduced postoperative complications and mortality rates, it is essential to be aware of the potential occurrence of bile duct injuries (see Figure 2.3), which are more severe and prevalent (0–2.7%) in laparoscopic procedures when compared to open cholecystectomy (0.2–0.5%). These injuries may involve bile leaks, strictures, partial or complete transection, and even removal of a portion of the bile duct, sometimes accompanied by vascular damage. Post-cholecystectomy bile duct injury is linked to considerable perioperative complications and mortality, diminished long-term survival and quality of life [7].

Several factors contribute to the occurrence of bile duct injuries, including challenges in identifying biliary tract anatomy due to conditions like acute cholecystitis, and the presence of large impacted stones, or a short cystic duct. Additionally, anatomical variations and technical errors leading to bleeding with subsequent clipping and coagulation trauma can also contribute to such injuries. In other words, BDIs during laparoscopic cholecystectomy can often be attributed to errors in surgical judgment and visual misperception of critical anatomy [35].

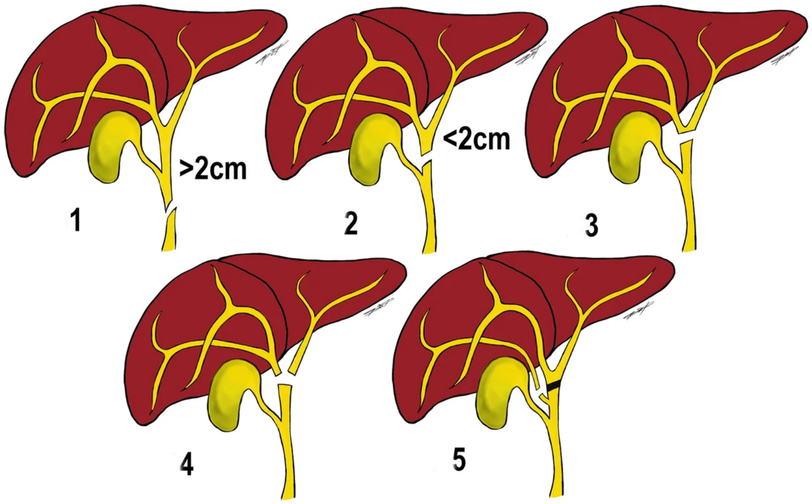


Figure 2.3: The Bismuth classification of Bile Duct Injuries (BDIs) categorizes them based on their distance from the bile duct bifurcation (Type 1–4) and includes individual right sectoral bile duct injury (type 5) [26].

Thus, the laparoscopic approach, while offering numerous advantages, presents unique challenges in terms of visual perception and depth perception. The limited field of view, two-dimensional imaging, and lack of haptic feedback can contribute to the misinterpretation of anatomical structures and increase the risk of inadvertent injury to the bile duct. Studies have highlighted the significance of errors in surgical judgment, including misidentification of the cystic duct and common bile duct, misperception of anatomical landmarks, and failure to recognize anatomical variations [9].

2.3 The Role of Training and Simulation

Historically, surgical training relied on lengthy apprenticeships, where trainees honed their skills on patients under the guidance of mentors in the workplace. However, contemporary factors, such as the implementation of the European Working Time Directive, shortened training duration, and reduced patient admission times, have reduced the availability of apprenticeship training [13].

The decreased time for workplace training has posed a challenge to the traditional apprenticeship model in medical education. The simulation presents a solution by providing a safe and controlled environment for trainees to engage in repeated practice tailored to their individual needs. Recent technological advancements have given rise to diverse surgical simulators, which have already found their place in surgical training. The success of simulation-based training depends on the proper validation and integration of simulators into the training curriculum. Within a surgical curriculum, trainees should be granted protected time for simulation-based training which should be overseen by supervisors. This approach to surgical education through simulation should foster the appropriate practice of technical skills while adhering to principles of safety and effectiveness [13].

As an example, the robotic surgery simulator (RoSS) is a portable and stand-alone simulator, that can be used for instructing both novice and advanced surgeons in the essential motor and cognitive skills involved in operating the da Vinci surgical robot (Figure 2.4). Leveraging virtual reality, RoSS immerses users in the fundamentals of robot-assisted surgery, providing a platform for skill development. The simulator offers a multi-level curriculum, has been designed to accommodate varying levels of difficulty and has been shown to be an effective training tool [36].

Figure 2.4: RoSS is a robotic surgery simulator with haptic feedback that offers an immersive experience with full-length surgical procedures in 3D [36].

Virtual reality (VR) simulators have also gained significant recognition in laparoscopic surgery training due to their ability to provide a controlled, risk-free environment for skill honing. A prominent example is the Simendo® VR Simulator, developed by Simendo B.V., Rotterdam, Netherlands.

The Simedo system aims to help trainees improve spatial awareness, hand-eye coordination, precise instrument manipulation, and coordinated bimanual techniques, however, this is done with simulated non-anatomical scenarios [45] as can be seen in Figure 2.5.

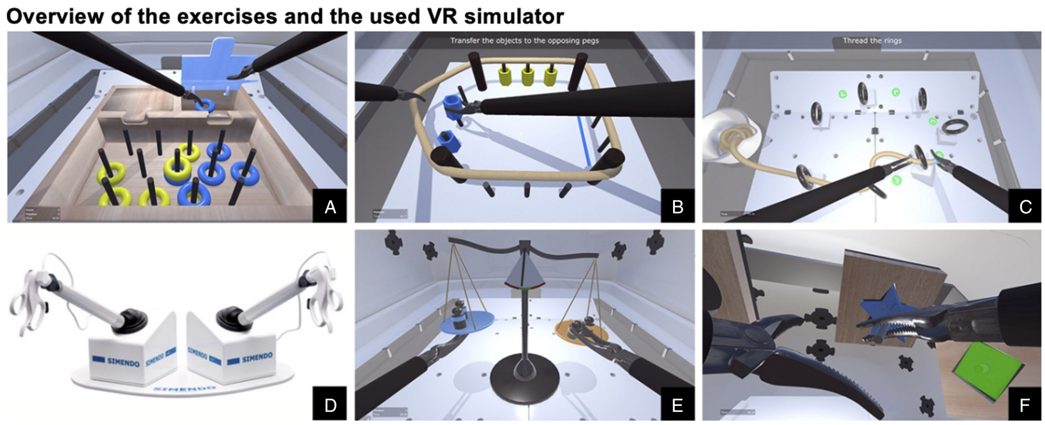


Figure 2.5: Simedo simulator: (A) Exercise "Sort the Rings," (B) exercise "Stretch and Transfer," (C) exercise "Ring and Rope," (D) Simendo VR laparoscopy simulator, (E) exercise "Balance," and (F) Exercise "Puzzle." [45].

In a recent publication, focused on the evaluation of Simendo's Virtual Reality Laparoscopy Simulator effectiveness, the objective was to evaluate the validity of the 'Bimanual Fundamentals' curriculum [45]. This curriculum was designed to augment psychomotor proficiencies of surgical residents which are essential for laparoscopic surgery. The study involved the recruitment of 49 participants, encompassing medical students, residents, and medical specialists specialized in gynecology, general surgery, and urology, drawn from three prominent tertiary medical centers in the Netherlands. Each participant completed the comprehensive curriculum, featuring five distinct exercises, on the simulator, executing each exercise three times. Notably, face validity, i.e., the simulator's fidelity to authentic laparoscopic procedures, had positive feedback. Participants appreciated the simulator's realism and its effectiveness in achieving the designated training objectives. The results also established construct validity, as performance disparities were found between novices, intermediates, and experts across various performance parameters. This ability of the simulator to stratify individuals based on their skill levels was a pivotal finding. The study also showed areas of potential improvement particularly related to depth perception and haptic feedback. Overall, Simendo showed improvements in psychomotor skills among surgical residents and thus has the potential to improve surgical proficiency [45].

2.4 Artificial Intelligence in Laparoscopic Cholecystectomy

Artificial intelligence (AI) and computer vision have opened up new avenues of research in laparoscopic surgery, with a specific focus on laparoscopic cholecystectomy (LC) [24]. Recent research has shown the potential of AI to augment various facets of this surgical procedure, addressing challenges tied to visual perception, decision-making, and skill acquisition [24, 5, 17, 47].

In the context of laparoscopic cholecystectomy (LC), AI plays a crucial role in providing real-time decision support, specifically in identifying safe and hazardous areas within the surgical field. The GoNoGoNet AI model represents a promising solution in this regard (see Figure 2.6). An analysis was conducted on 25 LC videos collected from 5 countries, involving 9 surgeons and 7 institutions. These videos were carefully selected according to predefined criteria, resulting in a representative sample of cases, including those with acute or chronic cholecystitis and various surgical complexities. The study assessed GoNoGoNet's performance in recognizing critical anatomical structures within the videos. Results showed high accuracy, specificity, and negative predictive value (NPV) for both "Go" and "No-Go" zones. Although sensitivity for the "Go" zone was slightly lower than that for the "No-Go" zone, the model's performance remained consistent across different expert consensus definitions, with the best outcomes achieved when consensus included pixels with over 50% agreement, indicating substantial spatial alignment between the AI model and expert annotations. This underscores the potential of AI in offering real-time guidance during LC, which could contribute to reducing adverse events [24].

One significant area where AI is applied in laparoscopic cholecystectomy (LC) is the recognition and analysis of surgical phases. One study gathered 163 LC videos from four medical centers and utilized deep learning models to accurately identify various phases of the procedure, aligning well with expert surgeon annotations [5]. This study also discusses the creation of a substantial LC dataset and the development of a CNN- and LSTM-based model for surgical phase recognition. The results showed an overall accuracy of 91.05% in determining surgical phase. On the other hand, the study acknowledges lower accuracy in some specific phases (e.g. adhesion lysis and clear

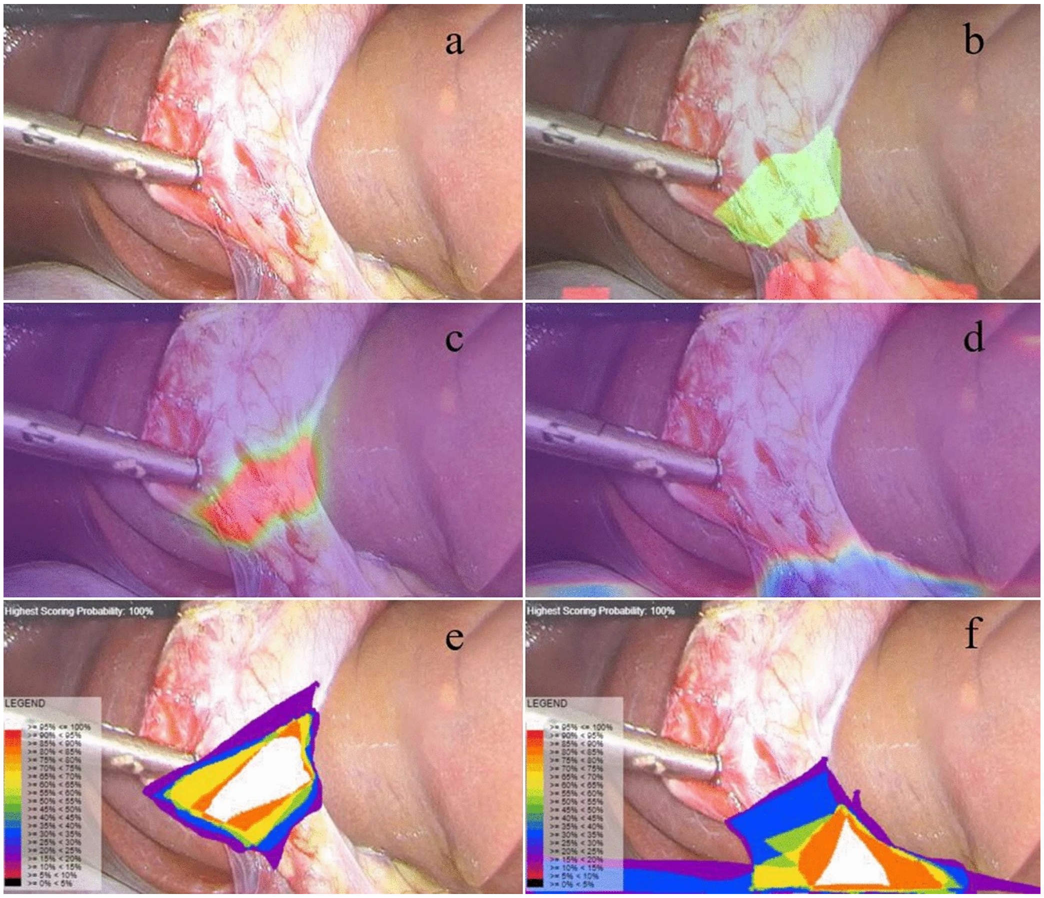


Figure 2.6: Comparing model predictions of GoNoGoNet with expert assessments, we provide an illustrative example of the model's predictions contrasted with the original frame (a). These comparisons are visually represented as overlays (b), wherein the Go zone is highlighted in green, and the No-Go zone is marked in red. Additionally, probability heat maps depicting the likelihood of Go (c) and No-Go zones (d) are presented for the model's predictions. To provide context, we include expert annotation heat maps for Go (e) and No-Go zones (f) as well. [24].

the operative region) due to limited training frames for these phases, underscoring the need for a more extensive dataset to enhance performance. Additionally, the study had a relatively small testing set (63 cases), which may restrict the generalizability of the findings. A forthcoming large-scale prospective study with over 10,000 cases is planned to address this limitation, along with efforts to recognize the critical view of safety (CVS) for reducing the risk of biliary duct injury (BDI). Overall, the study demonstrated the potential of AI for surgical phase recognition in LC but emphasizes the importance of larger datasets, improved accuracy in specific phases, and ongoing research into recognizing critical safety views and surgical actions.

In the domain of surgical phase recognition, another study has also yielded promising results. Golany et al. [17] used an LC video dataset, encompassing cases with adverse events, where the AI algorithm consistently exhibited a high level of accuracy in phase recognition. The methodology employed a two-stage model comprising Resnet50 and MS-TCN (Figure 2.7). The initial stage achieved an overall classification accuracy of 78%, while the subsequent stage, incorporating temporal information, achieved a higher accuracy of 89% on the test set. Notably, the model excelled in identifying critical phases, such as the Calot triangle dissection (92% accuracy), clipping and cutting (82% accuracy), and gallbladder dissection (96% accuracy). However, occasional misclassification of frames from the preparation phase as part of the Calot triangle dissection phase was observed. Furthermore, a comparison of the AI model's performance with expert surgeons revealed that for straightforward LC procedures, the AI model's recognition capability was comparable to that of surgeons, but for complex procedures, surgeons surpassed the AI model. The model's accuracy was also influenced by adverse events during LC procedures, resulting in reduced accuracy in such cases. Additionally, the study evaluated the model's generalizability across different hospitals, revealing variations in accuracy but overall strong performance when trained on data from four hospitals and tested on the fifth, achieving an accuracy of 87% [17].

In another laparoscopic cholecystectomy study, gallbladder inflammation was found to have varying impacts on surgical procedures and outcomes. Slightly inflamed gallbladders added 2.7 minutes to an operation, whereas severely inflamed gallbladders increased operation duration significantly, with an average of approximately 16.9 minutes and up to 31.3 minutes for the most affected surgeons. Gallbladder injuries occurred in 25% of cases, with a slight increase observed with higher inflammation levels, reaching up to a 28% higher probability of injury during severely inflamed cases for certain surgeons. Interestingly, gallbladder inflammation did not substantially affect a surgeon's ability to achieve the critical view of safety. Moreover, an AI model effectively quantified gallbladder inflammation, showing promising potential for optimizing operating room workflows and providing targeted feedback to surgical teams. Overall, the study found that gallbladder inflammation plays a notable role in operation duration and the likelihood of injury, but surgeons consistently achieved the critical view of safety regardless of inflammation levels [47].

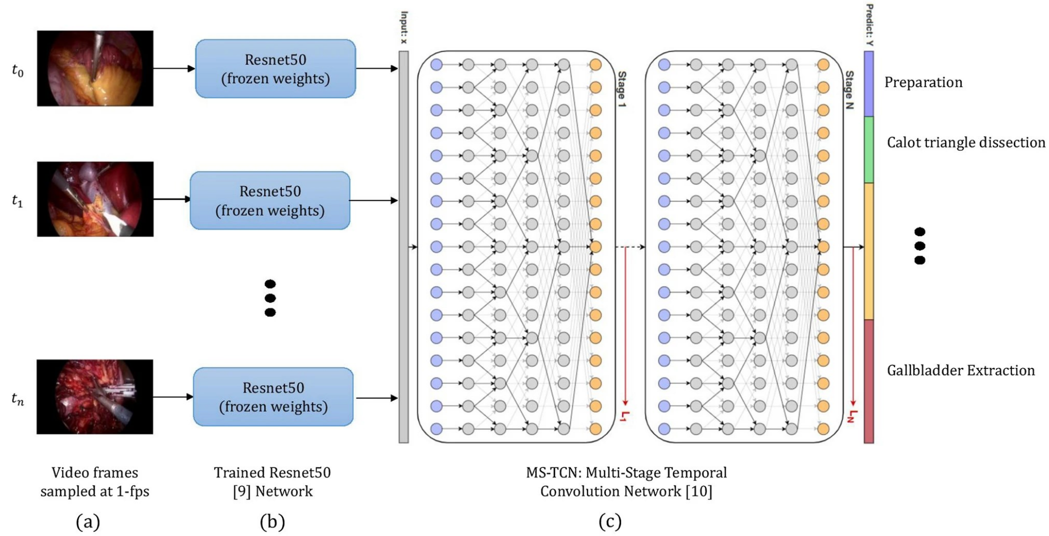


Figure 2.7: Multi-Stage Temporal Convolution Network for LC stage recognition [17]: (a) The laparoscopic cholecystectomy (LC) video is processed at a rate of 1 frame per second (fps). (b) Each individual frame is input into a deep convolutional neural network (Resnet50). The Resnet50 model is initially trained to classify the surgical phase associated with each frame independently. After training, the final prediction layer of Resnet50 is removed, and all network parameters become fixed (non-trainable). The Resnet50 then generates a feature vector for each frame, condensing its visual content into a lower-dimensional numerical representation. (c) All feature vectors from the input LC video are aggregated to create a sequence of feature vectors that represent the entire video. This sequence is then fed into the MS-TCN model, which comprises temporal convolutional layers with progressively increasing dilation rates. These temporal convolutional layers capture temporal relationships, and the expanding dilation pattern allows for the modeling of longer-term temporal dependencies. The final layer of the MS-TCN model produces surgical phase predictions for each frame within the video [17].

2.5 AI in Surgical Education

The consensus among surgeons leans towards the view that, as well as psychomotor skills, cognitive skills play a substantial role in shaping a good surgeon. This sentiment finds support in the surgical education literature, which underscores the significance of intraoperative judgment and decision-making as primary influencers of surgical proficiency and, consequently, surgical outcomes [27]. An illustrative example of this notion is the potential consequences of human visual pattern recognition errors, which can lead to the misinterpretation of surgical anatomy and result in critical structure damage, as seen in the case of bile duct injuries during laparoscopic cholecystectomy [49].

A study with the aim of using machine learning (ML) to identify surgical and operative variables that could accurately categorize participants based on their level of expertise in a virtual reality surgical procedure was done by [51]. In the study, fifty participants drawn from a single university were categorized a priori into four expertise tiers: experts (neurosurgery staff), seniors (neurosurgical fellows and senior residents), juniors (neurosurgical junior residents), and medical students. This diverse cohort engaged in 250 simulated tumor resections within a virtual reality neurosurgical tumor resection scenario, each scenario being replicated five times. An iterative process, guided by ML algorithms including K-nearest neighbor, naive Bayes, discriminant analysis, and support vector machine (SVMs), used metrics derived from instrument movement, force, tissue resection, and bleeding data output by the simulator to determine performance metrics and classify surgical expertise classification [51]. The results from this study found that the K-nearest neighbor algorithm had a 90% accuracy in classifying participants, naive Bayes algorithm 84%, the discriminant analysis algorithm had 78% accuracy, and SVM 76%. Notably, these findings underscore the potential of ML in delineating surgical expertise. This study shows the potential of simulators and ML to assess surgical skills, holding the promise of enhancing surgical education and training in the future [51].

2.6 Augmented Reality in Surgery

Augmented reality (AR) offers the ability to overlay images onto anatomical structures in real-time during surgical procedures, whether they are open or minimally invasive, to help guide the surgeon. For example, in liver surgery, AR has found application in hepatectomy for liver tumors. These AR systems function as guidance tools, providing surgeons with a live view of the tumor's location in relation to vital intra-parenchymal vascular structures [33]. While initially limited to 2D representation, advancements have led to the development of 3D AR intra-operative images, facilitated by techniques such as stereoscopic surface reconstruction and semi-automatic registration, often powered by deep learning algorithms (e.g., SmartLiver an image guidance system for liver surgery) [39]. Although several research groups have demonstrated the feasibility of AR-assisted procedures, it is important to note that the AR images still require verification through conventional methods, such as intra-operative ultrasound and cross-sectional imaging, to ensure accuracy and alignment with the surgical reality [4].

An example of an application of AR in LC is the iSurgeon system, which was designed to address the communication and guidance challenges frequently encountered in minimally invasive surgery (MIS) [50]. In MIS, traditional intraoperative guidance often relies on verbal communication, which can sometimes lead to misinterpretations and errors in identifying anatomical structures on the surgical screen. The iSurgeon system was developed to provide real-time visual guidance through telestration, defined as the process where the virtual hand of an experienced surgeon is captured in the sterile field and displayed in real-time on the operating screen using AR technology (Figure 2.8). To assess the impact of this technology, a study involving 60 laparoscopic novices who were randomly divided into two groups was done. In a cross-over design, one group initially trained with verbal guidance and then with additional telestration with AR, while the other group followed the reverse order. The training program included fundamental laparoscopic training and a specialized course that involved performing a porcine laparoscopic cholecystectomy (LC). The results found that training with telestration using AR technology led to significantly shorter total training times (1163 ± 275 seconds vs. 1658 ± 375 seconds, $p < 0.001$), reduced error rates, improved performance in LC procedures (GOALS 21 ± 5 vs. 18 ± 4, $p < 0.007$, and OSATS 67 ± 11 vs. 61 ± 8, $p < 0.015$), and a lower rate of complications (13.3% vs. 40%, $p < 0.020$). Additionally, participants reported reduced subjective workload and stress during training with AR (33.6 ± 12.0 vs. 30.6 ± 12.9, $p < 0.022$) [50]. These findings underscore the substantial advantages of using AR in surgical education and practice.

2.7 Think Like a Surgeon Application

Madani et al. [28] introduced the concept of the visual concordance test (VCT) as an innovative approach to establishing expert consensus within the surgical field. The VCT involves surgeons annotating frames extracted from surgical videos while simultaneously referencing the video itself. Accumulating annotations from an expert panel enables the creation of a “heat map,” visually representing the extent of concordance among these experts (Figure 2.9). Although the precise set of annotated pixels may not be identical across experts, the convergence area of these annotations is

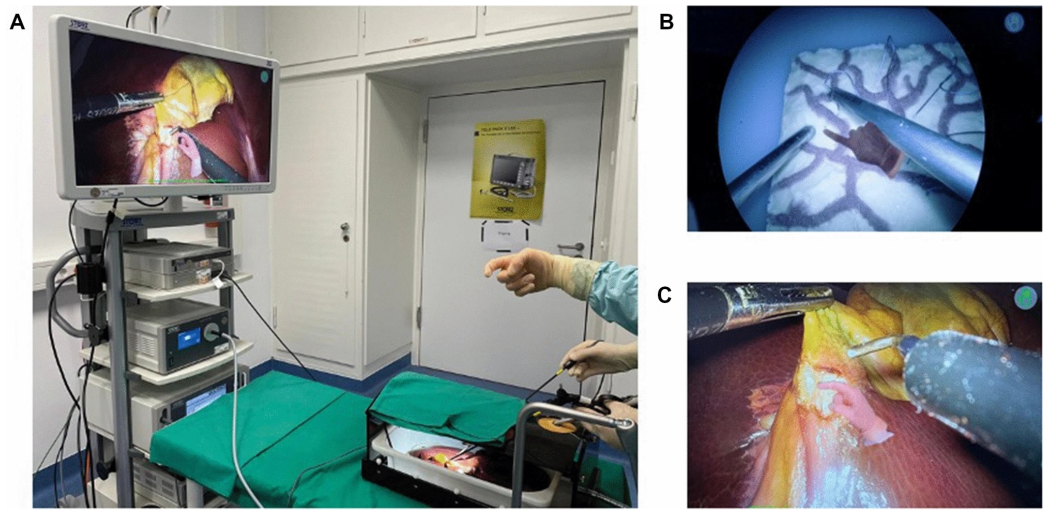


Figure 2.8: To illustrate the experimental setup for laparoscopic cholecystectomy on a porcine liver: A) The porcine liver was securely placed inside a plastic enclosure within the training box. B) An intriguing feature of this system was the ability to capture the virtual representation of the experienced surgeon's hand within the sterile field. This virtual hand was then dynamically displayed in real-time on the surgical screen. C) The iSurgeon system was successfully applied during a laparoscopic cholecystectomy simulation within the training box. Notably, the transparency level of the virtual hand was adjustable, allowing for fine-tuning, as demonstrated here with an 80% transparency setting [50].

deemed a point of consensus among the panel. Subsequently, these pixels have been employed to train AI models, such as identifying the optimal dissection location for a specific scene. Incorporating real-time AR overlays of Go and No-Go zones offers invaluable guidance and feedback to surgeons engaging in learning new procedures, aiming to enhance their performance, or navigating particularly intricate surgeries. The AI models trained on these data were used in our "LapBot" game as feedback.

2.8 Gamification

Gamification can be defined as the process of adding game play elements to a non-game context to make a boring or mundane task more interesting [1]. An advantage of gamification lies in the ability to motivate users to excel (e.g., by competing with others or trying to get top scores), which can enhance the quality of collected data and lead to more participants taking part in a study. Nonetheless, challenges exist; not every study or goal lends itself to gamification that's enjoyable

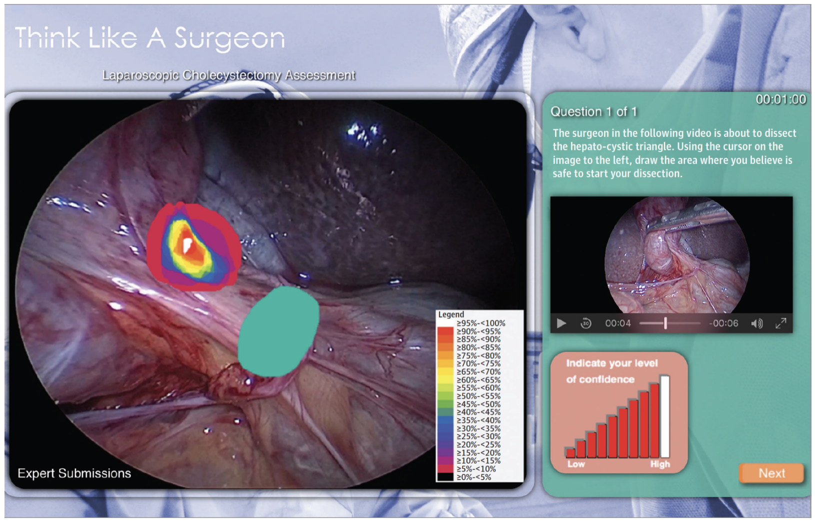


Figure 2.9: Screenshot of the “Think Like a Surgeon” platform is an interactive scenario where users are prompted to mark the point of initiation for dissecting the hepatocystic triangle. This exercise enables a comparative analysis between the annotations provided by users (depicted in green) and the annotations offered by experts (displayed as a heat map). [28]

to play. Additionally, crafting and releasing a game often demands more time and effort then a simple experimental user study or teaching scenario. Ultimately, the researcher’s ability to devise captivating game mechanics in line with the principles of game design contributes to the success of this approach [1].

Serious games or games that are designed with a primary purpose beyond mere entertainment, in the medical domain specifically, have shown to heighten learner motivation, enjoyment, and overall performance. Upon analyzing six distinct studies centered around gamification in general medical education [19, 46, 41, 38, 32, 37], the implementation of gamification consistently exhibited a propensity to elevate motivation and engagement levels. Furthermore, it exhibited an ability to effectively reinforce the accomplishment of learning objectives, thereby improving educational outcomes. This innovative approach not only resulted in more gratifying and positively perceived learning experiences compared to conventional teaching methods but also delivered additional advantages. The use of gamification correlated with an increase in examination scores, improvedd trainee confidence, and the better team dynamics, thereby affording a number of benefits to medical learners [43].

In laparoscopic surgery, a number of studies have looked at the impact of gamification elements on students' performance, training duration, and their willingness to continue training voluntarily.

Zwimpfer et al [52] integrated gamification into laparoscopic training through the introduction of a Laparoscopic Dexterity Game Box (LDGB)(Figure 2.10). Sixty medical students, all without prior laparoscopy experience, were divided into three groups. Each group performed four runs of the validated Laparoscopic Skills Training and Testing 3 (LASTT 3) exercise (Figure 2.10). Group A completed 40 runs with the LDGB, Group B completed 20 runs, and Group C had no LDGB runs. Task completion time and the number of errors were recorded and analyzed. The results showed that students who used the LDGB demonstrated significant improvement in their performance. They completed the LASTT 3 exercise more quickly, with a mean difference of 40 seconds between Group A (185s) and Group C(225s). Additionally, 86.4% of participants found the LDGB to be intuitive. This study suggests that gamification, as demonstrated by the LDGB, can enhance laparoscopic performance and offers a promising avenue for improvement [52].

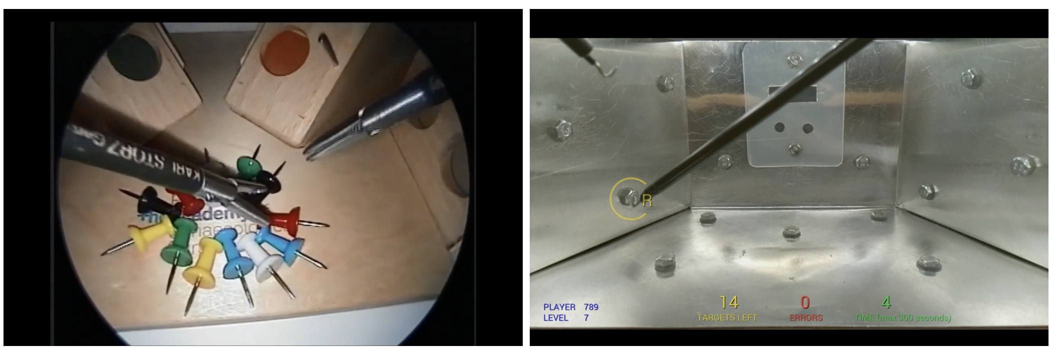


Figure 2.10: Left: Practical test to assess the competence level of individuals in basic laparoscopic psychomotor skills using the Laparoscopic Skills Training and Testing 3 (LASTT 3) examination. Right: The Laparoscopic Dexterity Game Box (LDGB), specially designed for this study, offers a gamified approach to laparoscopic training and skill development [52].

In another study [11], the authors aimed to assess the impact of a serious game skills competition on the voluntary usage of a laparoscopic simulator within the standard simulation curriculum for first-year surgical residents. The competition cohort experienced the introduction of leaderboards displaying the top performers' peg transfer times alongside the standard curriculum. Surprisingly, 76% of participants reported that they were not primarily motivated by competition. Nonetheless, the competition cohort exhibited significantly increased simulator usage and completed their course requirements earlier than the control cohort. Moreover, a higher proportion of residents from the competition cohort continued to use the simulator voluntarily after completing their requirements(44% vs. 4%). Their performance in peg transfer tasks also showed significant improvement. This study highlights that even when competition may not be the primary motivator, a well-designed serious game skills competition can boost simulator engagement and enhance laparoscopic skills development [11].

In summary, both of these studies underscore the potential of gamification elements to positively influence laparoscopic training, improving performance and encouraging ongoing voluntary participation in simulation activities.

An example of a mobile game platform for surgical education, is the Touch Surgery mobile application (Figure 2.11). This is an interactive platform designed to offer realistic surgical simulations with a diverse range of surgical specialties, the application breaks up each surgical intervention into their sequential and technique-specific steps. Users can engage in tutorials and test modules, enabling them to familiarize themselves with surgical anatomy, relevant tools, medications, and the specific surgical steps involved in a specific procedure. The application, however, does not delve into disease pathology or the practical surgical skills required for treatment, rather it aims to equip users with an understanding of the critical facets of a particular procedure [44].

A study to ascertain the app's validity, determine whether performance corresponds with the participants' expertise levels, analyze performance improvements over practice, and garner user feedback was done with medical students, residents, and attending hand surgeons, using the Carpal Tunnel Release module. The study revealed a consistent improvement in performance for each cohort with repetitive use, aligned with their respective training levels. Notably, users' confidence in the app's authenticity and utility declined as their training levels increased [44].

It was also observed that the simulator's performance displayed a consistent progression through practice, aligning with the principles outlined by Gallagher [16] that simulator training can lead to heightened focus on particular elements of a surgical task. Moreover, the study findings indicated a positive correlation between surgical experience levels and performance on the app's module. Although this does not necessarily imply that the simulator refines surgical skill, it does underline the validity of the app, substantiating a connection between surgical experience and simulator performance [44].

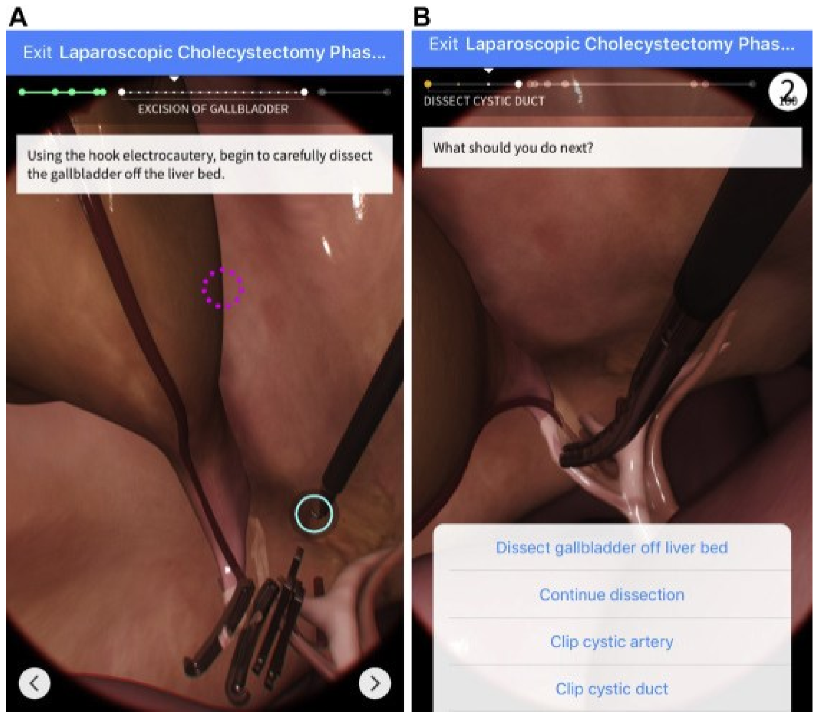


Figure 2.11: Captured images from the Touch Surgery mobile platform on Laparoscopic Cholecystectomy, illustrating both the 'Learn' phase (A) and the 'Test' phase (B) [6]

.

2.9 Summary

Studying the use of serious games in laparoscopic cholecystectomy offers a wealth of benefits that has the potential to mitigate some of the issues of the "see one, do one, teach one" surgical learning paradigm. Serious games provide an innovative and interactive platform for surgeons-in-training to practice complex procedures in a risk-free virtual environment, allowing them to hone their skills, improve hand-eye coordination, and enhance decision-making abilities. By simulating realistic scenarios and complications, serious games can facilitate a dynamic learning experience that bridges the gap between theory and practice. This approach not only reduces the learning curve for novice surgeons but also offers experienced surgeons an avenue to refine their techniques and adapt to new advancements. Moreover, integrating serious games into laparoscopic cholecystectomy training can potentially reduce the occurrence of errors during real-life surgeries, leading to safer procedures and better patient care overall. With this in mind, in the next Chapter, we describe the design and development of LapBot a serious game for learning safe LC.

Chapter 3

LapBot: Safe Chole

The following chapter is based on a manuscript to be submitted to Journal of Medical Internet Research (JMIR) Serious Games "There's an app for that: Design and feasibility study of an education game app for laparoscopic cholecystectomy surgical trainees".

3.1 Introduction

Laparoscopic cholecystectomy (i.e., gallbladder removal) is one of the most commonly performed procedures in the United States and across the world [29]. A common complication of laparoscopic cholecystectomy (LC) are bile duct injuries (BDIs), which are a significant source of morbidity, mortality, disability, and in turn healthcare costs [48, 42, 14, 10]. These injuries are often due to errors in surgical judgment and visual mis-perceptions of critical anatomy and tissue planes [48]. The goal of the following work is to develop a game that may improve trainees' perception and mental models of dissection planes and anatomy in the context of LC in order to enhance their performance during LC and help avoid injuries. Specifically, we have developed a game, "LapBot Safe Chole" to study the impact of feedback from artificial intelligence (AI) models on learning safe and not safe dissection zones (Go/No-Go) in laparoscopic cholecystectomy.

3.1.1 Surgical Artificial Intelligence in Laparoscopic Cholecystectomy

In the surgical domain, AI models can be trained to perform computer vision tasks (e.g., identify anatomical structure, interpret scenes in the surgical field), and therefore potentially provide guidance, navigational data and decision support to surgeons to reduce errors that lead to complications. Madani et al. [29] developed deep neural networks (DNN) models to identify Go (safe) and No-Go (unsafe) zones of dissection, as well as, identification of the liver, gallbladder, and hepatocystic triangle during laparoscopic cholecystectomy. In their work, deep learning was used to develop an AI model to accurately identify Go and No-Go zones. The model was trained on annotations of 2627 random frames from 290 LC videos, procured from 37 countries, 136 institutions, and 153 surgeons. The mean IOU (intersection over union), F1 score (uses the harmonic mean of a classifier to combine precision and recall into a single metric, accuracy, sensitivity, and specificity for the AI to identify Go zones was 0.53 (±0.24), 0.70 (±0.28), 0.94 (±0.05), 0.69 (±0.20). and 0.94 (±0.03), respectively. For No-Go zones, these metrics were 0.71 (±0.29), 0.83 (±0.31), 0.95 (±0.06), 0.80 (±0.21), and 0.98 (±0.05), respectively.

The predicted Go and No-Go zones can be visualized on surgical scenes using augmented reality (AR) for training or surgical guidance. With augmented reality virtual elements, in this case the visual representation of the AI model results of where it is safe and unsafe to dissect, are overlaid on the real surgical scenes. These annotated videos are used as feedback in our game application to determine the impact of (AI-generated) feedback on learning curves.

3.1.2 Games in Surgical Education

Games have proven to be an effective educational tool in numerous domains and across varying skill levels. Game-based learning can promote student engagement, allow for risk-taking, and promote collaborative learning [34]. In surgical education, serious games have been shown to improve intraoperative skills and decision-making and can be formally incorporated into the official curriculum for individual residency program [18]. Using games in surgical education is essential for promoting deliberate practice [15], i.e. repeatedly performing a specific activity in low-pressure conditions to better use the skill in high-stake situations, such as surgical decision-making in the operating room. Such practice is critical to achieving expert proficiency, particularly in performance-based disciplines [12]. A critical element of deliberate practice is expert feedback and coaching. Modern technological advancements, including the introduction of AI for coaching, have potentially made gaming even more beneficial. In this way, if an algorithm can be trained to replicate the mental model of expert surgeons, then it can be leveraged to provide feedback anywhere, at any time, and reach many individuals simultaneously. Ultimately, this would overcome one of the greatest obstacles to surgical coaching, which is the availability of faculty for consistent and timely feedback.

The field of educational game applications has witnessed significant advancements in recent years, with various studies exploring the effectiveness of innovative tools for training and assessment purposes. The integration of serious gaming in healthcare, such as VR training, has also shown promising results in increasing intrinsic motivation. One noteworthy research endeavor in the realm of educational game applications, akin to our developed game “LapBot: Safe Chole”, is the renowned Touch Surgery™ (TS) gaming application. Touch Surgery™ is a mobile platform that offers training and self-assessment modules for various procedures which are incorporated into a game-like structure. The application provides step-by-step guidance in surgical procedures and assesses cognitive decision-making through multiple-choice questions. A study aimed to establish the validity of the laparoscopic cholecystectomy (LC) module within Touch Surgery showed the importance of multimedia learning tools and proficiency-based training in LC surgical education [23]. Additionally, a pilot trial investigated the potential synergy between Touch Surgery and virtual reality (VR) trainers in reducing learning curves. The findings showed the need for automated feedback and evaluation systems to enhance training efficiency and the study advocated for a multimodal training approach that combines Touch Surgery with VR trainers to optimize laparoscopy training. Overall, the study concluded that the cost free Touch Surgery application is an innovative tool for training and assessment.

The purpose of the research presented in this paper was to assess the feasibility of developing an AI-generated feedback game, “LapBot Safe Chole” (or LapBot for short), to facilitate learning of safe LC. The specific focus of this paper is to describe the motivation behind the game as well as the development of the game including game play elements such as levels and scoring. Furthermore, our study provides the Lapbot Game with its first evaluation.

3.2 Methods: LapBot Serious Game Design

The LapBot: Safe Chole app was developed using React Native[1], an open-source UI (User Interface) software framework created by Meta Platforms. Both the front-end and back-end sides of LapBot were developed with React and the server side was developed using PHP. All the requests from the app to the server are sent by the Fetch API (a recent version of XMLHttpRequest). The first version was developed for iOS but we have recently ported the application to Android.

3.2.1 Onboarding

After downloading the game, at the first launch of the app, users will see the onboarding screens of LapBot (see Figure 3.1). These screens introduce users to the app and describe the main goals. Prior to game play, the users are required to create an account, consent to data collection, and fill out a demographic survey which includes questions as to gender, institution or hospital of affiliation, and country and city of residence. The demographic survey also has questions pertaining to the participant's academic role and experience. Specifically, their current role, i.e. medical student, postgraduate in years 1 to 5 (PGY 1-5), surgical fellow or attending surgeon, and total number of cholecystectomy surgeries performed. Additionally, the survey asks about the number of hours on average per week that the participant plays video games, thereby gauging the individual's level of game-playing experience.

3.2.2 Game Play

The game is divided into a practice level for learning about the game play mechanics, followed by 5 levels with increasing difficulty. The level of difficulty corresponds to the Parkland grading scale for cholecystitis [30]. The Parkland grading of each of the surgical scenarios was done by two surgeons that divided the rounds into five different levels of the game. Each level contains 20

[1] React Native website: https://reactnative.dev/
(Last accessed on December 21, 2022)

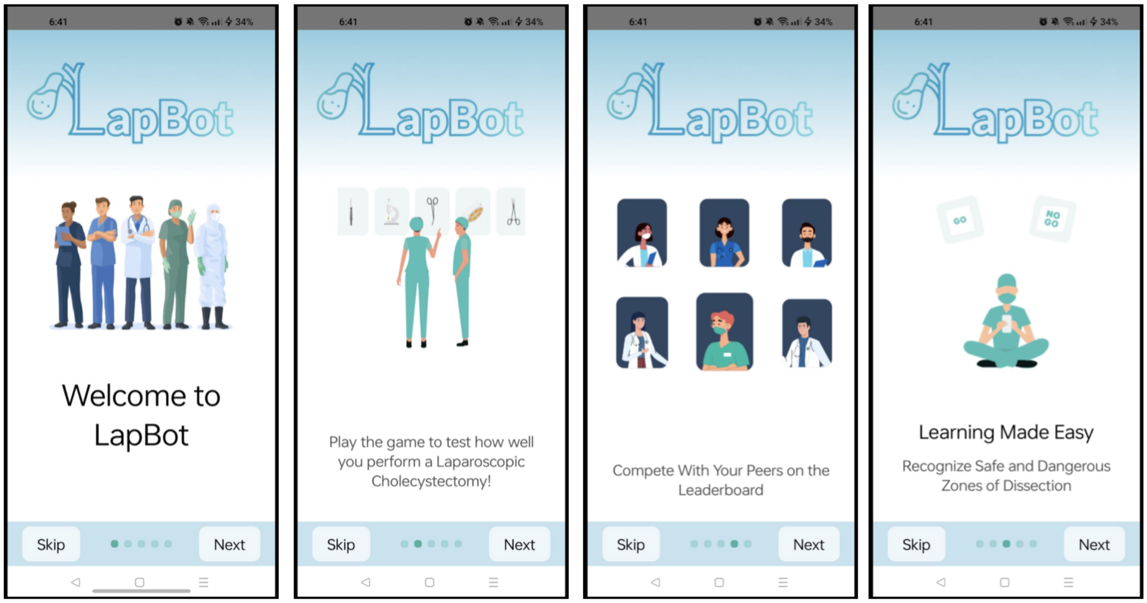


Figure 3.1: LapBot: Safe Chole onboarding screens which explain the functionality of the application.

rounds, however, to pass a level the player needs only to pass 5 consecutive rounds with a minimum threshold score of greater than 50%.

In each round of the game the player is presented with a surgical frame/video from a real surgery and their task is to choose where they would dissect next with consideration of a safe dissection zone (Figure 3.2). To understand the surgical context of the frame the player is presented with, the player can watch up to 5 seconds of the surgical video prior to the given frame by clicking on the "Video" button. After the target is confirmed by the user (by clicking the "Confirm" button), the player is a presented with a pop-up asking how confident they are with their answer (Figure 3.5).

After the confidence survey pop-up, users are presented with feedback that depicts their chosen target point with respect to the AI annotations of safe (Go) and dangerous (No-Go) zones of dissection given the surgical frame (Figure 3.2). The user can also choose to watch the video with the AI annotation overlaid in an AR view on all of the frames to get additional feedback by clicking on the "Feedback" button.

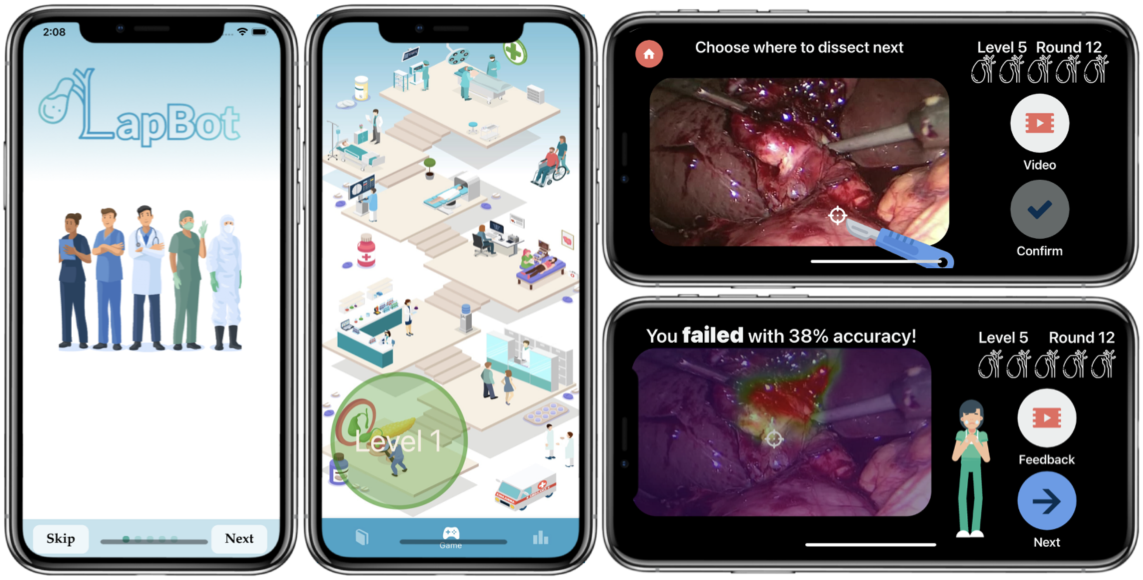
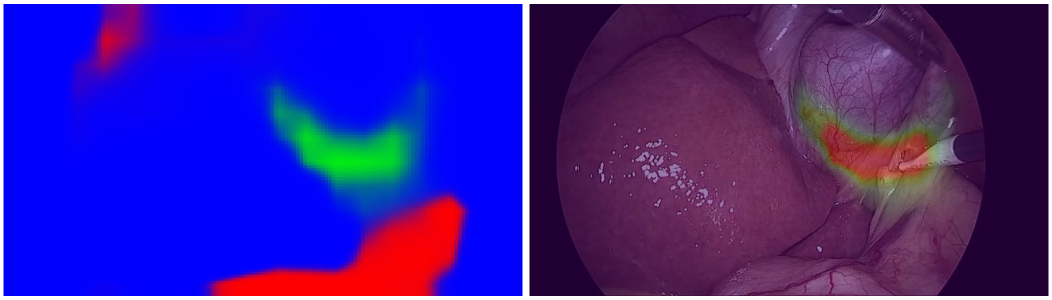


Figure 3.2: Screenshots of LapBot. From left to right: Home Screen, Level board Screen, game (top: the player chooses a location to dissect next and bottom: receives feedback on accuracy based on the AI model feedback).

Scoring

As described above, each level contains 20 individual rounds, however, to pass a level the player needs only to pass 5 consecutive rounds with a minimum threshold score of greater 50%. There are 5 gallbladder icons in the top right of the screen indicating how many consecutive rounds the player has passed in the current level. If at any point a player fails an individual round, the gallbladder counter resets to zero and the player must restart that level.

The player receives a score based on their answer, i.e. where they would dissect next. This score is calculated using the chosen target's location with respect to the correct answer. More specifically, each pixel in the surgical image frame has a score between -100 to +100 (calculated as the probability of the pixel being part of the Go zone or No-Go zone. This is shown in Figure 3.3a) where the green color shows the Go zones, the red color shows the No-Go zone and the blue color is neutral.



(a) Go (green) and No-Go (red) zones.

(b) AR heatmap showing the Go zone.

Figure 3.3: The frame that facilitates the calculation of the user's score is shown. The left frame displays three distinctive colors, green, red, and blue, which denote the Go zones, No-Go zones, and neutral zones, respectively. The right frame provides a clearer visualization of these zones in a surgical setting.

When a player chooses a dissection target, the score is calculated as follows. First, for all pixels in the target zone (where each pixel is defined by their RGB colour), we calculate the pixel score as follows:

$$\text{Pixel score} = \frac{pixel_{green}}{255} \times 100 - \frac{pixel_{red}}{255} \times 100 \tag{1}$$

Next, the round score is calculated as a sum of the pixels that fall within the target:

$$\text{Roundscore} = \frac{\sum \text{Pixel score}}{\text{Number of pixels}} \tag{2}$$

The round score is shown on the screen and the round is passed if it is above the threshold (50%). Upon completing a level, the final score for that level is the sum of the scores obtained in all rounds played within that level divided by the number of rounds played.

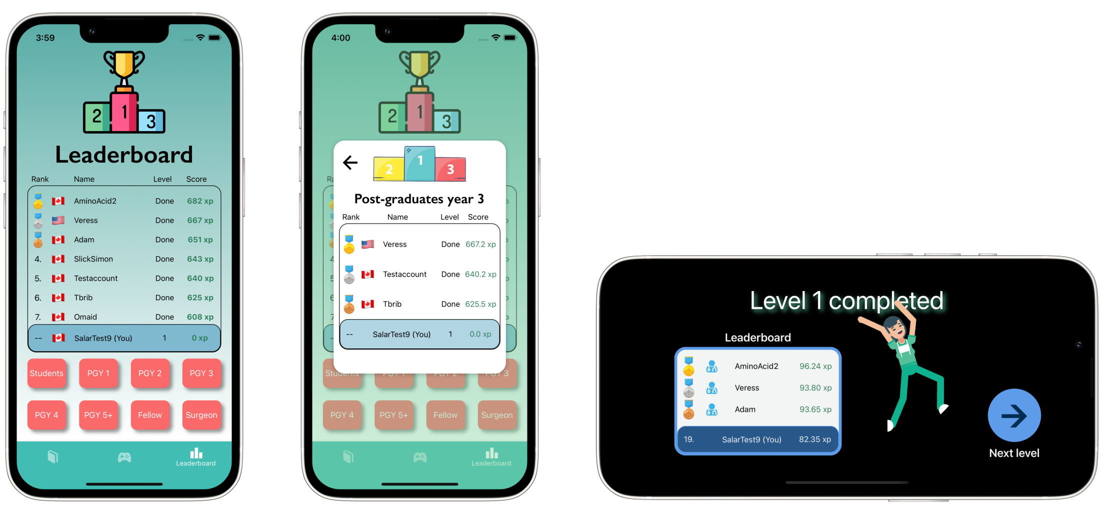
$$\text{Level score} = \frac{\sum \text{Round score}}{\text{Number of played rounds}} \tag{3}$$

Lastly, the total score is the sum of all scores earned across all the played levels. However, as the difficulty of the levels increases, we assign a varying increasing weight to each level's score. The score from level 5 carries the greatest weight, while the score from level 1 has the least.

$$\text{Total score} = (\text{First level score} \times 1.1) + (\text{Second level score} \times 1.2) + (\text{Third level score} \times 1.3) + (\text{Forth level score} \times 1.4) + (\text{Fifth level score} \times 1.5) \quad (4)$$

Leader boards

There are three different leader boards available in LapBot: (1) general, (2) experience-based and (3) level-by-level. All three are generated in real-time by fetching the latest data from the server as a JSON file which contains all the played rounds data across all the players. The general leader board (Figure 3.4a) displays the top 20 players with the highest total score among all players and it shows their names, current total score, and most recently played level. The experience-based leader board (Figure 3.4b) filter users based on their level of experience: medical students, post graduate year (PGY) 1 through 5, surgical fellows, and attending surgeons. Lastly, the level-by-level leader board (Figure 3.4c) ranks the players based on players' performance in that particular level. This leaderboard is only available after finishing the level and shows the top three players of that level.



(a) General (b) Experience-based (c) Level-by-level

Figure 3.4: The LapBot app features three leader boards.

3.2.3 In-app Survey

To gather more information about the ease of use, usability and/or visualization design, and the educational value we have included a pop-up survey that appears after finishing the second level. The survey presents three statements for users to rate on a scale of one to five, indicating their level of agreement with each statement:

- LapBot is a fun way to learn about safe cholecystectomy.
- The feedback (videos/images) are helpful for learning.
- Game elements (scores, leaderboards, consecutive gallbladders, etc) encourage me to play more.

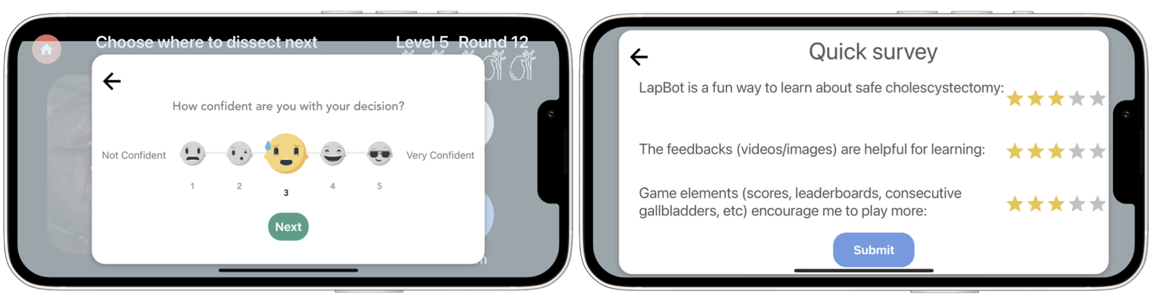


Figure 3.5: Screenshots of two quick in-app survey: The topper screen appears after each rounds and ask the users how confident they are with their chosen targets. The bottom survey screen only appears once after finishing level two and asks the users about their experience using LapBot.

3.3 Preliminary User Study

This retrospective study was based on existing data from the beta testing over a period of a month (November 2022 to December 2022). The mobile app was offered free of charge. The study was approved by the Concordia University Research Ethics Board. A convenience sample was employed for the recruitment of medical students, resident trainees, fellows and surgeon consultants. Participants were recruited by advertisements during workshops, social media, or word of mouth.

Data management and statistical analyses were performed using SAS 9.4 software. For qualitative variables, responses were grouped into congruent themes, and the corresponding response rates were computed and expressed as percentages. Quantitative variables were carefully examined through univariate analysis, and the outcomes were likewise reported in percentage terms. Any data incompleteness was thoroughly evaluated, with the extent of missing answers considered negligible and thus not impacting the conclusions. For statistical comparisons, the Mann-Whitney U test was applied to categorical variables, while ANOVA was employed to compare continuous variables between more than two groups. Furthermore, multivariate linear regression was utilized to estimate the adjusted relationships between the end score, postgraduate year, and participants' self-perceived confidence in their abilities. Throughout our analysis, a $p < 0.05$ was adhered to as the threshold for statistical significance.

The goal of our preliminary user study was to answer following research questions:

- Has the game been designed in order to: allow for gradual difficulty progression across levels, to be easy to navigate, fun to play and encourage users to play more.
- Do participants with more experience (staff surgeons and senior residents) will demonstrate greater game scores (total score and scores for each level) compared to cohorts with less experience (intern and junior residents), with staff surgeons performing the best.
- Do more inexperienced trainees (interns and junior residents) will demonstrate the greatest learning curve with the largest improvement in scores throughout the course of the game, compared to cohorts with greater experience (staff surgeons and senior residents)
- Does AI-generated feedback for deliberate practice of assessing safe planes of dissection for LC is a practical and valuable educational tool for surgical training.

3.3.1 Study Participants

A total of 22 participants participated in the usability testing phase of which fourteen (see Figure 3.6) (63.6%) completed all 5 levels of the game. We found no association between training level and likelihood for completing the game ($p = 0.59$). Of the 14 participants that completed the game, 2 were medical students, 2 were PGY1 (Postgraduate Year 1) trainees, 4 were PGY3 trainees, 1 was a PGY4 trainee, 2 were clinical fellow surgeons, and there were 3 attending surgeons. This diverse cohort of participants provided valuable insights into the usability and potential of the game, facilitating a comprehensive assessment of its effectiveness.

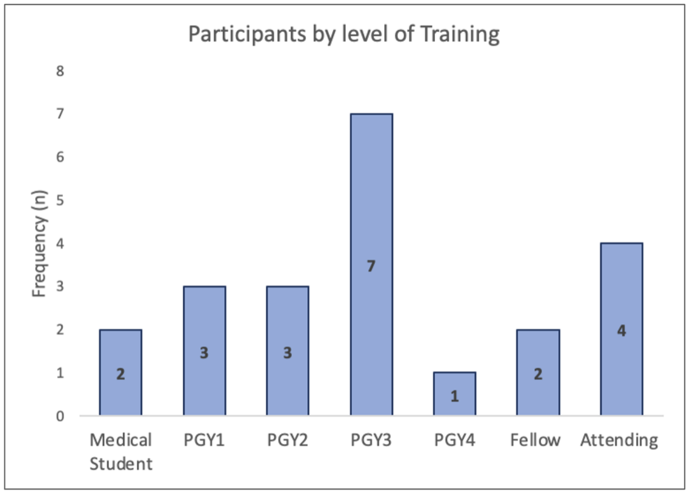


Figure 3.6: Participant count by level of training.

3.3.2 Game Level Evaluation

The intention behind the game design was to increase the difficulty as users progressed further, ultimately making higher levels more challenging. In order to accomplish this, the ordering of the levels within the game was carried out by expert surgeons to ensure that the content progression was tailored to reflect the real-world expertise and skill development of surgeons. As can be seen in Figure 3.7, players indeed encountered greater difficulties in passing the higher levels, aligning with the intended design. The graph shows the average number of attempts users needed to successfully pass each level, with a clear upward trend from level one to level five.

We also found that as the level increased, there was a step-wise decline in the average score, accompanied by a corresponding decrease in participants' average confidence (see Figure 3.8 right). This effectively reflects the progressively challenging levels (as described above each level is designed to correspond to the Parkland grading scale for cholecystitis (Grades 1-5). This outcome relating to our first research question indicates a successful implementation of the game's progression structure, achieving the desired goal of presenting users with escalating challenges as they advanced.

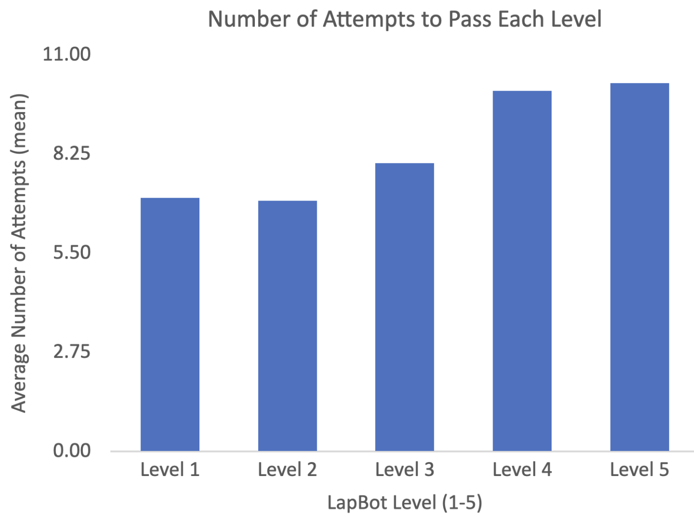


Figure 3.7: An analysis of player performance in terms of rounds needed to pass a level across varying difficulty levels in the LapBot game.

3.3.3 Player Confidence

In terms of player confidence, captured through the pop-up question after each round, as expected we found a positive and significant correlation between self assessed confidence level and score ($p < 0.001$). Users who expressed higher levels of confidence in their answers achieved better scores, while those who displayed lower confidence obtained comparatively lower scores (see Figure 3.8 left). In other words users were able to adequately self-assess predicted performance.

It should be noted however, that the confidence questions were initialized with a default value of 3 on the 1 to 5 confidence scale. As depicted in the chart, a majority of users maintained this initial value throughout the game, resulting in a notable concentration of confidence data around the value of 3. This suggests that many may have just skipped answering the question. In the next version of the application, we will ensure that we do not initialize the confidence scale.

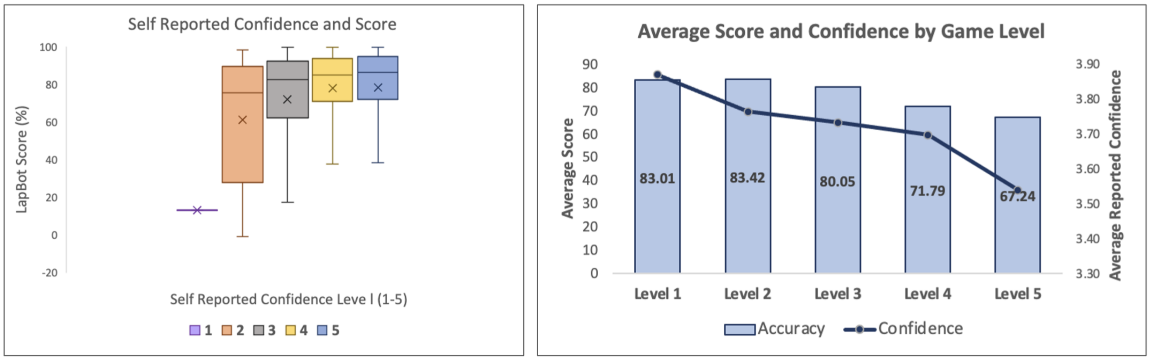


Figure 3.8: Left: Users who expressed higher levels of confidence in their answers achieved better scores. Right: As the level increased the average confidence decreased.

3.3.4 Scores by Experience

In looking at the relationship between academic proficiency and corresponding scores achieved in LapBot, we found a consistent pattern where users with higher levels of academic expertise consistently outperform those with less experience (see Figure 3.9). Notably, the scores obtained by medical students are lower compared to those of postgraduate year 1 (PGY1) residents, and this trend continues with increasing academic expertise with attending surgeons achieving the highest scores. These initial results show that participants with more experience (staff surgeons and senior residents) demonstrating greater game scores (total score and scores for each level) compared to cohorts with less experience (intern and junior residents), with staff surgeons performing the best, would be confirmed with a larger number of subjects.

This result validates the design of the LapBot application, showing its alignment with academic performance standards. In Figure 3.9, on the left, the progressive increase in scores as users' academic years advance demonstrates a positive relationship between experience and performance. On the right we see distinct lines representing different academic expertise grades across various levels, demonstrating higher scores for more experienced academic grades and lower scores for those with less expertise. These preliminary findings confirm the effectiveness of the LapBot app in accurately reflecting and aligning with academic proficiency levels, serving as a valuable tool for assessing and reinforcing surgical knowledge and skills.

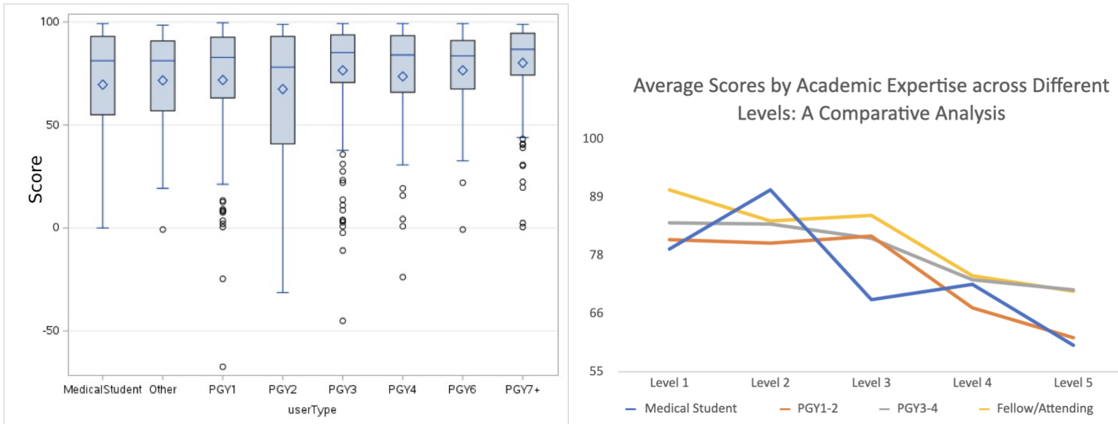


Figure 3.9: Comparison of Average Scores by Academic Expertise: Analysis of User Performance in Two Distinct Charts, highlighting the average scores achieved by users based on their academic expertise across different levels.

We also looked at the players' average scores as they progress through the game's levels. This progression is particularly evident when analyzing the data for each level separately, as the difficulty increases with each level. Notably, users with lower academic expertise begin with lower scores but exhibit a significant improvement in their performance as they continue playing and learning from the game. On the other hand, more experienced players, such as attending surgeons, start with higher scores from the outset and maintain a consistently high level of performance throughout the level (see Figure 3.10). These results suggest that inexperienced trainees (interns and junior residents) demonstrate the greatest learning curve with the largest improvement in scores throughout the course of the game, compared to cohorts with greater experience (staff surgeons and senior residents) as per our third research question.

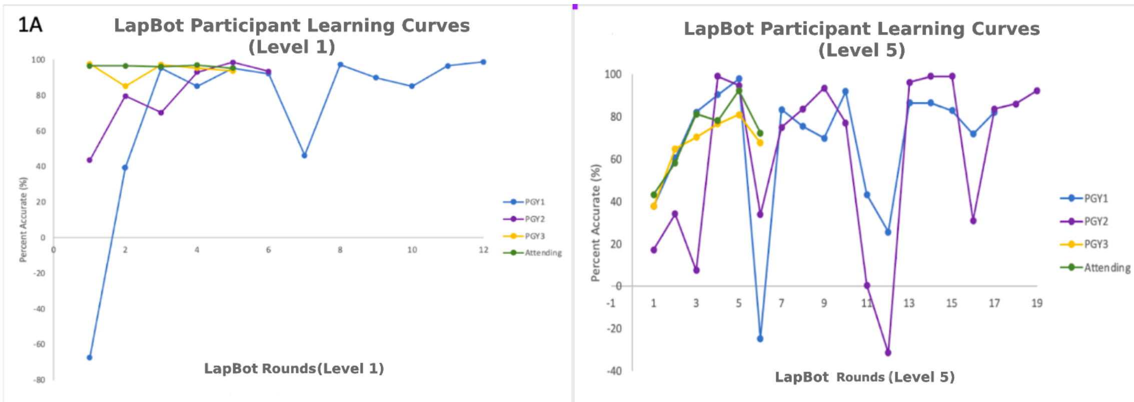


Figure 3.10: Learning Curves Across Levels: Comparative analysis of average scores in the LapBot Game, illustrating the improvement in scores as players progress further within each level, indicating the learning and skill development experienced by players.

In Figure 3.11, we see a relationship such that users with similar academic skills and grades exhibit comparable performance in passing levels and achieving scores. There was also no statistically significant difference observed in the overall performance of participants within each group: medical students ($p = 0.35$), PGY1&2 residents ($p = 0.43$), PGY3&4 residents ($p = 0.95$), and fellows/attendings ($p = 0.37$). This finding suggests that LapBot can effectively measure skill levels appropriate to each stage of training, and any significant or pre-defined deviations from the expected performance could be utilized in a competency-based curriculum to tailor instruction and assessment.

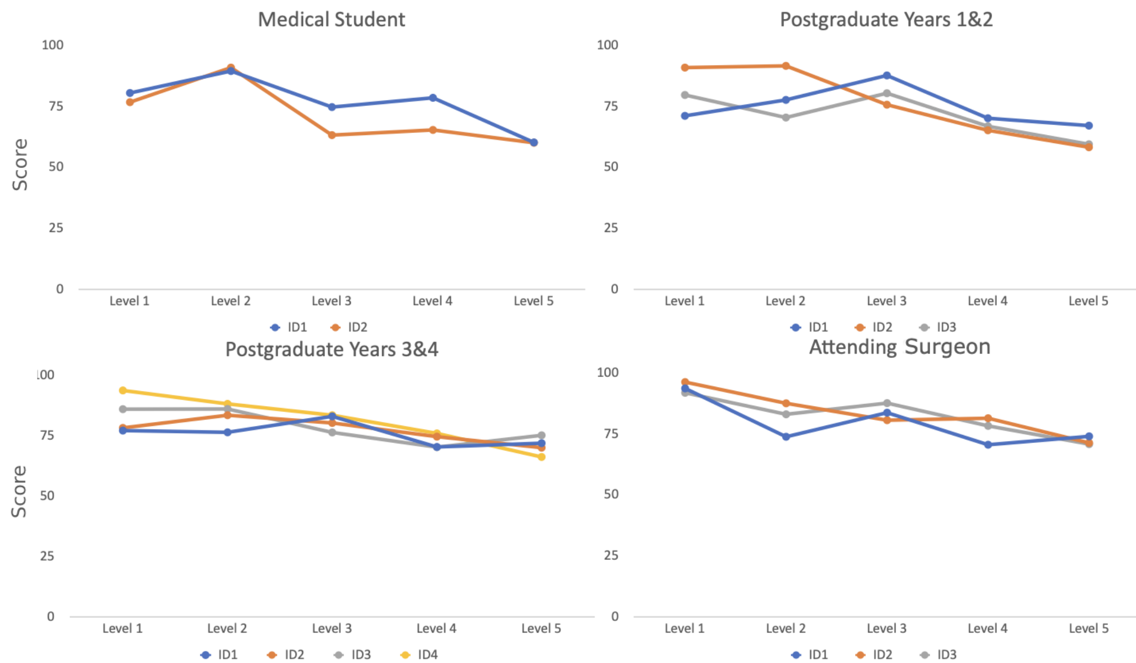


Figure 3.11: Comparative Analysis of Academic Expertise in LapBot Game: Examination of Average Scores across Four Academic Expertise Years (Medical Student, Postgraduate Year 1&2, Postgraduate Year 3&4, and Attending Surgeon), showcasing distinct patterns and similarities in game-play performance within each expertise group, highlighting the progression and consistency of scores.

3.3.5 Video watching rate

In order to understand the impact of the feedback videos, or how often they are used we looked at how often players looked at the videos. In Figure 3.12, users are sorted based on their scores in descending order, while the corresponding number of auxiliary videos they watched is displayed alongside each bar. A clear trend emerges from the chart, revealing a relationship between users' achieved scores and the utilization of feedback videos. As the average scores decrease, the number of feedback videos viewed also decreases. This finding emphasizes the significance of watching these videos, as they provide users with valuable insights and guidance to enhance their decision-making process and ultimately improve their scores. This preliminary finding underscores the importance of encouraging users to engage with the brief, AI-generated videos, as they offer important information about the Go and No-Go zones in the surgical context. By watching the videos, players can boost their learning process, develop a comprehensive understanding of surgical scenarios, and optimize their performance in the LapBot game. This result begins to suggest that AI-generated feedback for deliberate practice of assessing safe planes of dissection for LC is a practical and valuable educational tool for surgical training as per our fourth research question.

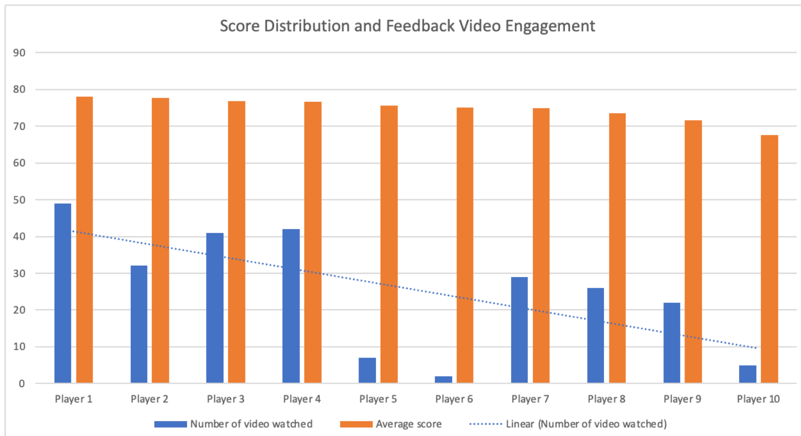


Figure 3.12: Analyzing user scores in descending order and the corresponding number of feedback videos watched in the game.

3.3.6 In-app survey results

A total of 12 users participated in the in-app survey, rating their level of agreement with three statements on a scale of 1 to 5. The first statement, "LapBot is a fun way to learn about safe cholecystectomy," received an average score of 4.41 out of 5, indicating a high level of agreement among users. The majority of users found the app to be an enjoyable and an effective way to learn about surgical procedures. The second statement, "The feedback (videos/images) are helpful for learning," received an average score of 3.91 out of 5. This result suggests that while most users found the feedback helpful, some users felt that either the videos could be improved or that they personally did not need the videos. The third statement, "Game elements (scores, leader boards, consecutive gallbladders, etc) encourage me to play more," received an average score of 4.33 out of 5. This result indicates that users find the game elements motivating and fun. These qualitative results suggest that users generally have a positive view of LapBot as a fun and effective way to learn about safe cholecystectomy answering our first research question. However, some improvements in the feedback and game elements could enhance the overall user experience.

3.3.7 External Survey

We also asked users to fill out an external survey after completing the game. We describe the key take-away points below. In terms of the non-annotated surgical videos players could watch prior to choosing a target, players appreciated the use of videos as it helped them to contextualize the information. In terms of the AI feedback, users appreciated the feedback provided, particularly in the form of videos and the augmented reality overlay of the AI, which helped guide their learning.

The design of the app, and the interface was described as easy to use, intuitive, and visually appealing and many users found the application to be addictive and enjoyable, with a fun interface and increasing difficulty level. Furthermore, the addition of leader boards was well-received, with users enjoying the competitive aspect of the application.

Additionally, the external survey data produced results consistent with the in-app survey. As shown in Figure 3.13, the majority of respondents (4.8 out of 5) found LapBot to be a fun way to learn laparoscopic cholecystectomy, while 4.6 out of 5 agreed that the app is easy to use and navigate. Furthermore, 3.8 out of 5 respondents agreed that the feedback provided in the app was helpful for their learning process and 4 out of 5 agreed that gamification elements motivated them to play more.

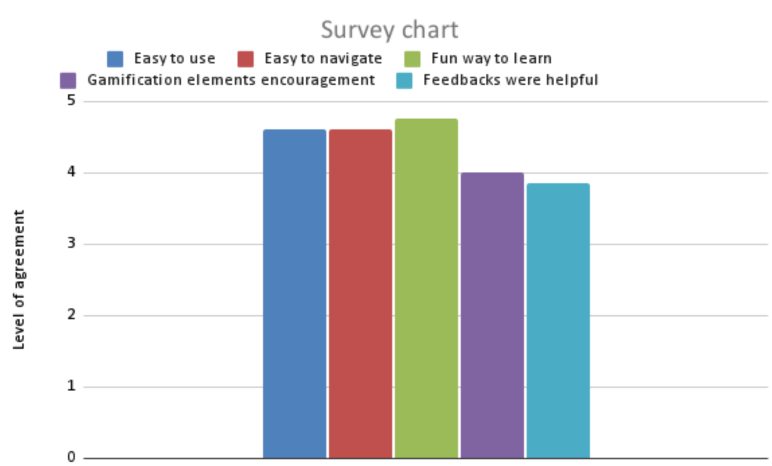


Figure 3.13: A chart representing data from the external survey

3.4 Conclusions

In the surgical domain, adverse events often occur due to errors in visual perception and judgments that lead to misinterpretation of anatomy [29]. The preliminary results from the game-play scoring data suggest that LapBot could be useful in evaluating students' progress and identifying individuals who may require additional lessons and practice before advancing to the next academic year. As part of our future work, further testing of the game with a larger user population will enable us to refine and establish more precise patterns based on users' experience. This iterative process will allow us to develop robust and reliable measures that effectively distinguish players based on their achieving scores, thereby facilitating the implementation of a competency-based curriculum.

In summary, the findings of our testing of the first prototype of Lapbot: Safe Chole, underscore the effectiveness of LapBot as an engaging and efficient tool for exposing users to a wide range of surgical cases and facilitating skill development. By providing a fun and interactive platform for users to engage with surgical scenarios, LapBot offers a valuable opportunity for users to enhance their surgical knowledge, gain practical insights, and improve their skills over time.

Chapter 4

Conclusions

4.1 Summary

Major bile duct injuries during laparoscopic cholecystectomy (LC) are a significant source of morbidity, mortality, disability, and healthcare costs. These injuries are primarily due to errors in surgical judgment and visual misperception of critical anatomy and tissue planes. In surgical education, serious games have been shown to improve intraoperative skills and decision-making. To facilitate learning of safe LC we developed an engaging and educational mobile game that provides coaching and feedback from an artificial intelligence (AI) model to improve intraoperative decision-making during LC.

LapBot Safe Chole was designed to be an innovative, free, accessible and interactive simulated learning experience with real-time AI-generated feedback for deliberate practice. After downloading the game on their smartphone device, users are presented with intraoperative LC scenarios (short video clips) and tasked with identifying the ideal target zone of dissection. After submitting a response, users are provided with an accuracy score according to the pre-annotated safe and unsafe zones of dissection. Annotations come from a validated AI algorithm that provides real-time scoring and feedback overlay on the video clips and avoids the need for faculty availability. There are 5 levels in the game, with increasing difficulty based on the Parkland grading scale for cholecystitis. The game application generates a total score for each level as well as a global total score, each of which can be viewed on a leaderboard. In the study version of the game, a user is required to pass

(> 50% accuracy score) five consecutive questions in a row before proceeding to the next level.

Beta-testing (n = 22) results indicated improvement in game scores with each round, with attendings and senior trainees reaching top-scores earlier than junior residents per level. Players could be distinguished by their learning curves and learning progression which can facilitate a competency-based curriculum. Furthermore, we found a positive and statistically significant association between the level of the user and the accuracy of their scores (p=0.003). Specifically, the more advanced users tend to achieve higher scores on average. Qualitative feedback from our end-users was highly positive; 80% of users reported LapBot was easy to use (20% were neutral), 60% agreed it was easy to navigate (40% strongly agreed), and 100% of users reported LapBot was a very fun way to learn about safe LC.

In summary, LapBot: Safe Chole is a new and innovative mobile application game that integrates safe LC principles into an educational, accessible and free game using AI-generated feedback. Initial beta-testing suggests validity evidence in the assessment score and that this has a high likelihood of adoption and engagement by surgical trainees.

4.2 Future Work

Future directions include additional validation with a larger sample size. In fact, in late August 2023 we released the game live for our larger study and to date have 700 international participants (comprised of medical students, residents and attendings) from over 60 countries. Analysis of this new data will be done in early September.

Our preliminary survey results suggests that in order to improve the application, we should focus on enhancing the feedback section to provide users with more comprehensive feedback that is specifically tailored to their decision-making targets. The comment section of the survey revealed that some users were having difficulty understanding why their answers were incorrect, and they would like to receive more detailed explanations for their failed rounds. By implementing a more detailed feedback system, users will be able to better understand their mistakes and improve their skills more efficiently. This may ultimately result in a more effective learning curve for users.

In addition, this app has the potential to expand beyond cholecystectomy and include other types of laparoscopic surgeries. With advancements in machine learning technology, it is possible to extract an augmented reality layer version (Go and No-Go zones) of any well-recorded surgery, which can be integrated into the app. This approach can help students and medical residents to enhance their decision-making skills on various types of surgeries, not just limited to cholecystectomy. By incorporating multiple surgeries, the app can provide a more comprehensive learning experience and help users to develop a more diverse skillset in laparoscopic surgery.

The integration of additional gamification elements, such as music and animated visuals like progressively changing gallbladder images, could enhance user engagement and enjoyment. Incorporating music can evoke emotions and make the app more entertaining, while dynamic visual elements can visually represent the increasing difficulty levels and encourage users to strive for better performance. These additions could contribute to a more immersive and motivating user experience, fostering increased interaction with the app.

As a means of further enriching the training capabilities of the app, expanding the number of rounds and levels could offer users exposure to a broader array of surgical scenarios. Increasing the variety of surgical cases would provide users with a more comprehensive learning experience, allowing them to encounter diverse challenges and build a deeper understanding of safe dissection zones. This expanded content would ensure that users receive ample practice and preparation across a range of surgical contexts, enhancing their overall surgical education.

Further, we could explore the implementation of the game concept within virtual reality (VR) to create a more lifelike and immersive surgical training environment. Utilizing touch controllers as laparoscopic instruments could provide users with a more realistic sensation of performing surgery, enhancing the training experience by closely simulating the actual surgical procedure. This extension into VR could attract a wider range of users, including those who seek hands-on experience in a controlled, immersive setting. However, would be more limited in terms of available hardware to users.

Bibliography

[1] N. Ahmed and K. Mueller. Gamification as a paradigm for the evaluation of visual analytics systems. In Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization, pages 78–86, 2014.

[2] B. Alkhaffaf and B. Decadt. 15 years of litigation following laparoscopic cholecystectomy in england. Annals of surgery, 251(4):682–685, 2010.

[3] G. Berci, J. Hunter, L. Morgenstern, M. Arregui, M. Brunt, B. Carroll, M. Edye, D. Fermelia, G. Ferzli, F. Greene, et al. Laparoscopic cholecystectomy: first, do no harm; second, take care of bile duct stones, 2013.

[4] L. R. Bertrand, M. Abdallah, Y. Espinel, L. Calvet, B. Pereira, E. Ozgur, D. Pezet, E. Buc, and A. Bartoli. A case series study of augmented reality in laparoscopic liver resection with a deformable preoperative model. Surgical endoscopy, 34:5642–5648, 2020.

[5] K. Cheng, J. You, S. Wu, Z. Chen, Z. Zhou, J. Guan, B. Peng, and X. Wang. Artificial intelligence-based automated laparoscopic cholecystectomy surgical phase recognition and analysis. Surgical endoscopy, 36(5):3160–3168, 2022.

[6] S. Chidambaram, S. Erridge, D. Leff, and S. Purkayastha. A randomized controlled trial of skills transfer: from touch surgery to laparoscopic cholecystectomy. Journal of Surgical Research, 234:217–223, 2019.

[7] S. Connor and O. Garden. Bile duct injury in the era of laparoscopic cholecystectomy. British journal of surgery, 93(2):158–168, 2006.

[8] N. G. Csikesz, A. Singla, M. M. Murphy, J. F. Tseng, and S. A. Shah. Surgeon volume metrics in laparoscopic cholecystectomy. Digestive diseases and sciences, 55:2398–2405, 2010.

[9] S. W. Dekker and T. B. Hugh. Laparoscopic bile duct injury: understanding the psychology and heuristics of the error. ANZ journal of surgery, 78(12):1109–1114, 2008.

[10] A. Ejaz, G. Spolverato, Y. Kim, R. Dodson, J. K. Sicklick, H. A. Pitt, K. D. Lillemoe, J. L. Cameron, and T. M. Pawlik. Long-term health-related quality of life after iatrogenic bile duct injury repair. Journal of the American College of Surgeons, 219(5):923–932, 2014.

[11] M. El-Beheiry, G. McCreery, and C. M. Schlachta. A serious game skills competition increases voluntary usage and proficiency of a virtual reality laparoscopic simulator during first-year surgical residents' simulation curriculum. Surgical endoscopy, 31:1643–1650, 2017.

[12] K. A. Ericsson. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Academic medicine, 79(10):S70–S81, 2004.

[13] E. Evgeniou and P. Loizou. Simulation-based surgical education. ANZ journal of surgery, 83 (9):619–623, 2013.

[14] D. R. Flum, A. Cheadle, C. Prela, E. P. Dellinger, and L. Chan. Bile duct injury during cholecystectomy and survival in medicare beneficiaries. Jama, 290(16):2168–2173, 2003.

[15] D. R. Flum, E. P. Dellinger, A. Cheadle, L. Chan, and T. Koepsell. Intraoperative cholangiography and risk of common bile duct injury during cholecystectomy. Jama, 289(13):1639–1644, 2003.

[16] A. G. Gallagher, E. M. Ritter, H. Champion, G. Higgins, M. P. Fried, G. Moses, C. D. Smith, and R. M. Satava. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Annals of surgery, 241(2):364, 2005.

[17] T. Golany, A. Aides, D. Freedman, N. Rabani, Y. Liu, E. Rivlin, G. S. Corrado, Y. Matias, W. Khoury, H. Kashtan, et al. Artificial intelligence for phase recognition in complex laparoscopic cholecystectomy. Surgical Endoscopy, 36(12):9215–9223, 2022.

[18] M. Graafland, W. A. Bemelman, and M. P. Schijven. Game-based training improves the surgeon's situational awareness in the operation room: a randomized controlled trial. Surgical endoscopy, 31(10):4093–4101, 2017.

[19] J. Guckian, A. Sridhar, and S. Meggitt. Exploring the perspectives of dermatology undergraduates with an escape room game. Clinical and experimental dermatology, 45(2):153–158, 2020.

[20] N. Gulaydin. Laparoscopic cholecystectomy by the modified bikini line approach as a simple and safe technique. Revista da Associação Médica Brasileira, 67:1172–1176, 2021.

[21] K. M. Harboe and L. Bardram. The quality of cholecystectomy in denmark: outcome and risk factors for 20,307 patients from the national database. Surgical endoscopy, 25:1630–1641, 2011.

[22] M. Hu, D. Wattchow, and D. de Fontgalland. From ancient to avant-garde: A review of traditional and modern multimodal approaches to surgical anatomy education. ANZ journal of surgery, 88(3):146–151, 2018.

[23] K.-F. Kowalewski, J. D. Hendrie, M. W. Schmidt, T. Proctor, S. Paul, C. R. Garrow, H. G. Kenngott, B. P. Müller-Stich, and F. Nickel. Validation of the mobile serious game application touch surgery™ for cognitive training and assessment of laparoscopic cholecystectomy. Surgical endoscopy, 31:4058–4066, 2017.

[24] S. Laplante, B. Namazi, P. Kiani, D. A. Hashimoto, A. Alseidi, M. Pasten, L. M. Brunt, S. Gill, B. Davis, M. Bloom, et al. Validation of an artificial intelligence platform for the guidance of safe laparoscopic cholecystectomy. Surgical endoscopy, 37(3):2260–2268, 2023.

[25] D. E. Litwin and M. A. Cahan. Laparoscopic cholecystectomy. Surgical Clinics of North America, 88(6):1295–1313, 2008.

[26] J. Lubikowski, B. Piotuch, A. Stadnik, M. Przedniczek, P. Remiszewski, P. Milkiewicz, M. A. Silva, and M. Wojcicki. Difficult iatrogenic bile duct injuries following different types of

upper abdominal surgery: report of three cases and review of literature. BMC surgery, 19(1): 1–9, 2019.

[27] A. Madani, M. C. Vassiliou, Y. Watanabe, B. Al-Halabi, M. S. Al-Rowais, D. L. Deckelbaum, G. M. Fried, and L. S. Feldman. What are the principles that guide behaviors in the operating room? Annals of Surgery, 265(2):255–267, 2017.

[28] A. Madani, K. Grover, and Y. Watanabe. Measuring and teaching intraoperative decision-making using the visual concordance test: deliberate practice of advanced cognitive skills. JAMA surgery, 155(1):78–79, 2020.

[29] A. Madani, B. Namazi, M. S. Altieri, D. A. Hashimoto, A. M. Rivera, P. H. Pucher, A. Navarrete-Welton, G. Sankaranarayanan, L. M. Brunt, A. Okrainec, et al. Artificial intelligence for intraoperative guidance: using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. Annals of surgery, 2022.

[30] T. D. Madni, D. E. Leshikar, C. T. Minshall, P. A. Nakonezny, C. C. Cornelius, J. B. Imran, A. T. Clark, B. H. Williams, A. L. Eastman, J. P. Minei, et al. The parkland grading scale for cholecystitis. The American Journal of Surgery, 215(4):625–630, 2018.

[31] B. Navez, F. Ungureanu, M. Michiels, D. Claeys, F. Muysoms, C. Hubert, M. Vanderveken, O. Detry, B. Detroz, J. Closset, et al. Surgical management of acute cholecystitis: results of a 2-year prospective multicenter survey in belgium. Surgical endoscopy, 26:2436–2445, 2012.

[32] D. Neureiter, E. Klieser, B. Neumayer, P. Winkelmann, R. Urbas, and T. Kiesslich. Feasibility of kahoot! as a real-time assessment tool in (histo-) pathology classroom teaching. Advances in Medical Education and Practice, pages 695–705, 2020.

[33] P. Phutane, E. Buc, K. Poirot, E. Ozgur, D. Pezet, A. Bartoli, and B. Le Roy. Preliminary trial of augmented reality performed on a laparoscopic left hepatectomy. Surgical Endoscopy, 32: 514–515, 2018.

[34] J. L. Plass, B. D. Homer, and C. K. Kinzer. Foundations of game-based learning. Educational psychologist, 50(4):258–283, 2015.

[35] E. Rauws and D. Gouma. Endoscopic and surgical management of bile duct injury after laparoscopic cholecystectomy. Best Practice & Research Clinical Gastroenterology, 18(5): 829–846, 2004.

[36] S. Rehman, S. J. Raza, A. P. Stegemann, K. Zeeck, R. Din, A. Llewellyn, L. Dio, M. Trznadel, Y. W. Seo, A. J. Chowriappa, et al. Simulation-based robot-assisted surgical training: a health economic evaluation. International Journal of Surgery, 11(9):841–846, 2013.

[37] C. C. Roche, N. P. Wingo, and J. H. Willig. Kaizen: An innovative team learning experience for first-semester nursing students. Journal of Nursing Education, 56(2):124–124, 2017.

[38] E. F. Schlegel and N. J. Selfridge. Fun, collaboration and formative assessment: Skinquizition, a class wide gaming competition in a medical school with a large class. Medical teacher, 36 (5):447–449, 2014.

[39] C. Schneider, S. Thompson, J. Totz, Y. Song, M. Allam, M. Sodergren, A. Desjardins, D. Barratt, S. Ourselin, K. Gurusamy, et al. Comparison of manual and semi-automatic registration in augmented reality image-guided liver surgery: a clinical feasibility study. Surgical endoscopy, 34:4702–4711, 2020.

[40] M. Y. Shah, U. Somasundaram, T. Wilkinson, and N. Wasnik. Feasibility and safety of three-port laparoscopic cholecystectomy compared to four-port laparoscopic cholecystectomy. Cureus, 13(11), 2021.

[41] R. Sheppeard, A. Samuels, and P. Davies. i-dermify: does using practical illustration and verbal description in a game format improve recognition and confidence in describing common skin diseases? Clinical and Experimental Dermatology, 45(5):610–611, 2020.

[42] D. Stefanidis, P. Montero, D. R. Urbach, A. Qureshi, K. Perry, S. L. Bachman, A. Madan, R. Petersen, and A. D. Pryor. Sages research agenda in gastrointestinal and endoscopic surgery: updated results of a delphi study. Surgical endoscopy, 28(10):2763–2771, 2014.

[43] M. D. Szeto, D. Strock, J. Anderson, T. E. Sivesind, V. M. Vorwald, H. R. Rietcheck, G. S.

Weintraub, and R. P. Dellavalle. Gamification and game-based strategies for dermatology education: Narrative review. JMIR Dermatology, 4(2):e30325, 2021.

[44] J. Tulipan, A. Miller, A. G. Park, J. T. Labrum IV, and A. M. Ilyas. Touch surgery: analysis and assessment of validity of a hand surgery simulation “app”. Hand, 14(3):311–316, 2019.

[45] M. P. van Ginkel, M. P. Schijven, W. M. van Grevenstein, and H. W. Schreuder. Bimanual fundamentals: validation of a new curriculum for virtual reality training of laparoscopic skills. Surgical Innovation, 27(5):523–533, 2020.

[46] R. Ventre, C. Pardoe, S. Singhal, D. Cripps, and J. Hough. Gamification of dermatology: Stud2ybuddy, a novel game to facilitate dermatology revision for final-year medical students. Future Healthc J, 6(Suppl 2):22–22, 2019.

[47] T. M. Ward, D. A. Hashimoto, Y. Ban, G. Rosman, and O. R. Meireles. Artificial intelligence prediction of cholecystectomy operative course from automated identification of gallbladder inflammation. Surgical Endoscopy, 36(9):6832–6840, 2022.

[48] L. W. Way, L. Stewart, W. Gantert, K. Liu, C. M. Lee, K. Whang, and J. G. Hunter. Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. Annals of surgery, 237(4):460, 2003.

[49] L. W. Way, L. Stewart, W. Gantert, K. Liu, C. M. Lee, K. Whang, and J. G. Hunter. Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. Annals of surgery, 237(4):460, 2003.

[50] C. Wild, F. Lang, A. Gerhäuser, M. Schmidt, K. Kowalewski, J. Petersen, H. Kenngott, B. Müller-Stich, and F. Nickel. Telestration with augmented reality for visual presentation of intraoperative target structures in minimally invasive surgery: a randomized controlled study. Surgical Endoscopy, 36(10):7453–7461, 2022.

[51] A. Winkler-Schwartz, R. Yilmaz, N. Mirchi, V. Bissonnette, N. Ledwos, S. Siyar, H. Azarnoush, B. Karlik, and R. Del Maestro. Machine learning identification of surgical

and operative factors associated with surgical expertise in virtual reality simulation. JAMA network open, 2(8):e198363–e198363, 2019.

[52] T. A. Zwimpfer, C. Wismer, F. Geissler, R. Oehler, J. Geiger, A. Schötzau, B. Fellmann-Fischer, and V. Heinzelmann-Schwarz. Gamification in laparoscopic training: a randomized, controlled study. 2022.