Artificial Intelligence Assessment of Expertise in Virtual Reality Spine Pedicle Screw Insertion:

Case Series Study

Puja Bimalbhai Pachchigar

Master Of Science (MSc.) Experimental Surgery (Thesis): Surgical Education

McGill University, Montreal

Aug 2024

This thesis is submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Experimental Surgery (thesis)

Table of Contents

Abstract:

IMPORTANCE: Our understanding of the composites of technical expertise during spinal procedures including the insertion of pedicle screws is incomplete. Datasets generated from surgical simulation allows the quantitation of psychomotor skills, which can be analyzed using machine learning algorithms which allows a more complete understanding of surgical performance.

OBJECTIVE: The primary aim of this study was to identify important features distinguishing skilled and less skilled levels of expertise during simulated pedicle screw insertion. The secondary aim was to benchmark the classification accuracy of surgical performance through the implementation of machine learning algorithms.

DESIGN: Participants from four universities were recruited between July 15, 2022, and May 31, 2023, to participate in a case-series study. Data were collected over a single time point and no follow-up data were collected. Participants were classified a priori as either skilled or less skilled based on their experience in performing human pedicle screw insertion procedures.

SETTING: McGill University Neurosurgical Simulation and Artificial Intelligence Learning Centre.

PARTICIPANTS: Forty-three neurosurgery and orthopedic spine surgeons, spine fellows, and neurosurgery and orthopedic residents.

INTERVENTION: Insertion of bilateral L5 and L4 pedicle screw insertions on a virtual reality platform resulting in 172 inserted screws for analysis. These 172 datapoints were divided into training set (70% - 121 data points) and testing set (30% -51 data points) for algorithm's training and testing. We used 5-fold cross validation to validate the algorithm.

EXPOSURES: All participants performed a simulated virtual reality L5-L4 bilateral pedicle screw insertion during which they each inserted 4 screws.

MAIN OUTCOMES AND MEASURES The main outcomes and measures were determined through an iterative process, wherein features related to instrument movement, force application, and tissue resection were chosen from the raw simulator data output. This selection was achieved through a combination of four feature selection methods, wrapper-based, embedded, filter-based, and weight-based, in conjunction with Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (KNN), and Artificial Neural Network (ANN) models. The objective was to accurately assess the skill levels of participants in simulated pedicle screw insertion.

RESULTS A cohort of 43 participants, including 5 women and 38 men with a mean age of 33.6 years (SD 9.5), was evaluated. Machine learning models demonstrated varying accuracies on the test set: SVM achieved 78%, Random Forest 80%, KNN 82.3%, and ANN 82.3%. Analysis revealed 24 common features across Random Forest, KNN, and ANN, each achieving a classification accuracy of over 80%.

CONCLUSIONS AND RELEVANCE By employing machine learning algorithms, our study identified key features that may determine components of expertise during simulated pedicle screw insertion. We introduced a combined approach for feature selection that could enhance the accuracy of classifying skilled versus less skilled performance in future experiments. This method may prove valuable in the assessment and training of various surgical procedures.

RÉSUMÉ

IMPORTANCE : Notre compréhension des composants de l'expertise technique lors des procédures spinales, y compris l'insertion de vis pédiculaires, est incomplète. Les ensembles de données générés par la simulation chirurgicale permettent la quantification des compétences psychomotrices, qui peuvent être analysées à l'aide d'algorithmes d'apprentissage automatique, offrant ainsi une compréhension plus complète des performances chirurgicales.

OBJECTIF : L'objectif principal de cette étude était d'identifier les caractéristiques importantes distinguant les niveaux de compétence élevés et faibles lors de l'insertion simulée de vis pédiculaires. L'objectif secondaire était de mesurer la précision de classification des performances chirurgicales grâce à l'implémentation d'algorithmes d'apprentissage automatique.

CONCEPTION : Les participants de quatre universités ont été recrutés entre le 15 juillet 2022 et le 31 mai 2023 pour participer à une étude de série de cas. Les données ont été collectées à un seul moment et aucune donnée de suivi n'a été collectée. Les participants ont été classés a priori comme soit compétents, soit moins compétents, en fonction de leur expérience dans la réalisation de procédures d'insertion de vis pédiculaires humaines.

LIEU : Centre de simulation neurochirurgicale et d'apprentissage de l'intelligence artificielle de l'Université McGill.

PARTICIPANTS : Quarante-trois chirurgiens en neurochirurgie et en orthopédie rachidienne, boursiers en chirurgie rachidienne et résidents en neurochirurgie et en orthopédie.

INTERVENTION : Les participants ont effectué des insertions de vis pédiculaires bilatérales L5 et L4 sur une plateforme de réalité virtuelle, résultant en 172 (points de données) vis insérées pour analyse. Ces 172 points de données ont été divisés en un ensemble d'entraînement (70% - 121 points de données) et un ensemble de test (30% - 51 points de données) pour l'entraînement et le test de l'algorithme. De plus, une validation croisée à 5 plis a été employée pour valider l'algorithme.

EXPOSITIONS : Tous les participants ont réalisé une insertion simulée de vis pédiculaires bilatérales L5-L4 en réalité virtuelle, au cours de laquelle chacun a inséré 4 vis. PRINCIPAUX **RÉSULTATS ET MESURES :** Les principaux résultats et mesures ont été déterminés par un processus itératif, dans lequel des caractéristiques liées au mouvement des instruments, à l'application de force et à la résection des tissus ont été choisies à partir des données brutes du simulateur. Cette sélection a été réalisée grâce à une combinaison de quatre méthodes de sélection de caractéristiques : basées sur des enveloppes, intégrées, basées sur des filtres et basées sur des poids, en conjonction avec les modèles Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (KNN) et Artificial Neural Network (ANN). L'objectif était d'évaluer avec précision les niveaux de compétence des participants lors de l'insertion simulée de vis pédiculaires. **RÉSULTATS :** Un cohort de 43 participants, dont 5 femmes et 38 hommes avec un âge moyen de 33,6 ans (SD 9,5), a été évalué. Les modèles d'apprentissage automatique ont démontré des précisions variables sur l'ensemble de test : SVM a atteint 78 %, Random Forest 80 %, KNN 82,3 % et ANN 82,3 %. L'analyse a révélé 24 caractéristiques communes à Random Forest, KNN et ANN, chacune atteignant une précision de classification de plus de 80 %. CONCLUSIONS ET **CONCLUSIONS ET PERTINENCE** En utilisant des algorithmes d'apprentissage automatique, notre étude a identifié des caractéristiques clés qui pourraient déterminer les composantes de l'expertise lors de l'insertion simulée de vis pédiculaires. Nous avons introduit une approche combinée de la sélection des caractéristiques qui pourrait améliorer la précision de la classification des performances qualifiées par rapport aux performances moins qualifiées dans de futures

expériences. Cette méthode pourrait s'avérer précieuse dans l'évaluation et la formation pour diverses procédures chirurgicales.

सार

महत्व: रीढ़ की प्रक्रियाओं के दौरान तकनीकी विशेषज्ञता के घटकों की हमारी समझ, जिसमें पेडिकल स्क्रू का सम्मिलन शामिल है, अधूरी है। शल्य चिकित्सा सिमुलेशन से उत्पन्न डेटासेट साइकोमोटर कौशलों की मात्रात्मकता की अनुमति देते हैं, जिन्हें मशीन लर्निंग एल्गोरिदम का उपयोग करके विश्लेषित किया जा सकता है, जो शल्य चिकित्सा प्रदर्शन की एक अधिक पूर्ण समझ की अनुमति देता है।

उद्देश्य: इस अध्ययन का मुख्य उद्देश्य पेडिकल स्क्रू सम्मिलन के दौरान कुशल और कम कुशल स्तरों की विशेषज्ञता को अलग करने वाले महत्वपूर्ण विशेषताओं की पहचान करना था। द्वितीयक उद्देश्य शल्य चिकित्सा प्रदर्शन की वर्गीकरण सटीकता को मापने के लिए मशीन लर्निंग एल्गोरिदम के कार्यान्वयन के माध्यम से मापन करना था।

डिजाइन: चार विश्वविद्यालयों से प्रतिभागियों को 15 जुलाई 2022 और 31 मई 2023 के बीच एक केस-सीरीज अध्ययन में भाग लेने के लिए भर्ती किया गया था। डेटा को एक ही समय बिंदु पर एकत्र किया गया था और कोई अनुवर्ती डेटा एकत्र नहीं किया गया था। प्रतिभागियों को उनके मानव पेडिकल स्क्रू सम्मिलन प्रक्रियाओं को करने के अनुभव के आधार पर पहले से ही कुशल या कम कुशल के रूप में वर्गीकृत किया गया था।

सेटिंग: मैकगिल विश्वविद्यालय न्यूरोसर्जिकल सिमुलेशन और आर्टिफिशियल इंटेलिजेंस लर्निंग सेंटर।

प्रतिभागी: तैंतालीस न्यूरोसर्जरी और ऑर्थोपेडिक स्पाइन सर्जन, स्पाइन फेलो और न्यूरोसर्जरी और ऑर्थोपेडिक निवासी।

हस्तक्षेप: प्रतिभागियों ने वर्चुअल रियलिटी प्लेटफॉर्म पर द्विपक्षीय L5 और L4 पेडिकल स्क्रू इंसर्शन किए, जिसके परिणामस्वरूप विश्लेषण के लिए 172 (डेटा पॉइंट्स) स्क्रू डाले गए। इन 172 डेटा पॉइंट्स को एल्गोरिदम प्रशिक्षण और परीक्षण के लिए प्रशिक्षण सेट (70% - 121 डेटा पॉइंट्स) और परीक्षण सेट (30% - 51 डेटा पॉइंट्स) में विभाजित किया गया। इसके अतिरिक्त, एल्गोरिदम को मान्य करने के लिए 5-गुना क्रॉस-वैलिडेशन का उपयोग किया गया।

एक्सपोजर: सभी प्रतिभागियों ने एक सिम्युलेटेड वर्चुअल रियलिटी L5-L4 द्विपक्षीय पेडिकल स्क्रू सम्मिलन किया, जिसके दौरान उन्होंने प्रत्येक ने 4 स्क्रू सम्मिलित किए।

मुख्य परिणाम और उपाय: मुख्य परिणाम और उपाय एक पुनरावृत्त प्रक्रिया के माध्यम से निर्धारित किए गए थे, जिसमें कच्चे सिम्युलेटर डेटा आउटपुट से उपकरण आंदोलन, बल आवेदन, और ऊतक पुनःप्राप्ति से संबंधित विशेषताओं को चुना गया था। यह चयन चार विशेषता चयन विधियों, रैपर-आधारित, एंबेडेड, फिल्टर-आधारित, और वेट-आधारित, के संयोजन के माध्यम से प्राप्त किया गया था, जो समर्थन वेक्टर मशीन (SVM), रैंडम फॉरेस्ट, K-निकटतम पड़ोसी (KNN), और कृत्रिम तंत्रिका नेटवर्क (ANN) मॉडलों के साथ किया गया था। उद्देश्य प्रतिभागियों के कौशल स्तरों को सिम्युलेटेड पेडिकल स्क्रू सम्मिलन में सटीक रूप से आकलन करना था।

परिणाम: एक 43 प्रतिभागियों का समूह, जिसमें 5 महिलाएं और 38 पुरुष थे, जिनकी औसत आयु 33.6 वर्ष (SD 9.5) थी, का मूल्यांकन किया गया। मशीन लर्निंग मॉडल ने टेस्ट सेट पर विभिन्न सटीकताएं प्रदर्शित कीं: SVM ने 78%, रैंडम फॉरेस्ट ने 80%, KNN ने 82.3%, और ANN ने 82.3% हासिल किया। विश्लेषण ने रैंडम फॉरेस्ट, KNN, और ANN में 24 सामान्य विशेषताओं का खुलासा किया, प्रत्येक ने 80% से अधिक वर्गीकरण सटीकता हासिल की।

निष्कर्ष और प्रासंगिकता मशीन लर्निंग एल्गोरिदम का उपयोग करके, हमारे अध्ययन ने महत्वपूर्ण विशेषताओं की पहचान की जो सिम्युलेटेड पेड़िकल स्क्रू इनसर्शन के दौरान विशेषज्ञता के घटकों को निर्धारित कर सकती हैं। हमने फीचर चयन के लिए एक संयुक्त दृष्टिकोण प्रस्तुत किया है जो भविष्य के प्रयोगों में कुशल और कम कुशल प्रदर्शन को वर्गीकृत करने की सटीकता को बढ़ा सकता है। यह विधि विभिन्न शल्य चिकित्सा प्रक्रियाओं के मूल्यांकन और प्रशिक्षण में उपयोगी साबित हो सकती है।

ACKNOWLEDGEMENTS:

I extend my deepest appreciation to a group of extraordinary individuals whose guidance, support, and inspiration have been pivotal throughout my journey.

Firstly, I owe an immense debt of gratitude to Dr. Rolando Del Maestro, who entered my life as a guiding light. I consider myself one of the fortunate few to have him as a supervisor. His profound understanding of people, gained from his vast experiences, was apparent from our very first meeting, where he inquired about my life goals and steered me toward experimental surgery—a path perfectly aligned with my aspirations. Dr. Del Maestro is a veritable ocean of knowledge and lived experience, whose continuous motivation and vibrant energy have inspired me to persevere through all challenges. His life exemplifies meaningful living, setting a standard that continually inspires me. His lessons in patience, motivation, compassion, and diligence have profoundly shaped my personal and professional growth.

A heartwarming appreciation goes to our dearest Pam Del Maestro, whose love and positivity radiate through our lives. Her care has often come in the form of delicious treats, from muffins to festive Christmas delights, sustaining not just our bodies but our spirits. Dr. Del Maestro is the heart of our community, and Pam, you are the lifeblood that keeps it beating.

A heartfelt thank you to Dr. Ahmed Aoude, who not only allowed me to shadow him in the operating theater, providing a firsthand look at the critical nature of our study, but also supported our research by providing lab space. His mentorship, motivation, and support have been invaluable, and I am deeply grateful to have him as my research chair. His achievements and unwavering support have been a continuous source of inspiration.

I would also thank the RAC Committee members, including Dr Houssem Gueziri and Dr. Mark Driscoll for their input and suggestions on improving my thesis.

My deepest gratitude extends to Dr. Recai Yilmaz, whose achievements and life journey have both astonished and inspired me. Our profound conversations have opened new vistas of perception in my life, and his friendship and mentorship have urged me to set higher standards of hard work and vision.

I am extremely fortunate to have worked with Dr. Bilal Tarabay on this project. His insights have been the bedrock of our study, and our shared conversations—creative and wild—will remain with me as cherished memories. His approach to life, marked by humor and kindness, has taught me to face life's challenges with a light heart.

To my dearest friend and project partner, Trisha Tee, whose unwavering support and wisdom have been crucial in navigating the toughest phases of our research. Her friendship has been a lesson in patience, sincerity, and warmth, and I am truly blessed to have shared this journey with her.

I also want to extend my gratitude to Dr Abdulmajeed Albeloushi and Dr. Mohamed Albantobi for their guidance and expertise in spine surgery and supporting me through out the project in terms of technical expertise as well as for moral support. My special thanks to Dr. Zhi Wang, Dr. Sung-Joo Yuh, Dr. Ahmed Aoude, Dr. Lucy Luo, Dr. Ahmad Alsayegh, Dr. Mohamad Bakhaidan, Dr. Carlo Santaguida, Nour Abou Hamdan, and Dr. Abdulrahman Almansouri for their help recruiting trial participants and supporting me in my journey of research in experimental surgery.

Thanks to Ali Fazlollahi, Nour Abou Hamdan, and Chinyelum Agu, the lifelines of our lab. Your positive energy has been a beacon for us all, helping us to maintain our momentum and focus.

I am grateful to Neevya Balasubramaniam, Bianca Giglio, and Vanya Davidovic for their invaluable assistance in conducting vital tests for our study. Each moment spent with you has brought new learning and perspectives that have enriched our work.

Special thanks to my family—my brother, my mother, and in loving memory of my father. Your unwavering faith in me during the darkest times has been my stronghold. The sacrifices you have made are the foundation of all that I have achieved. No words can fully express my gratitude for your love and dedication.

I must extend a special word of thanks to Pranay, my dearest friend, whose unwavering support and patience have been my anchors in tumultuous times. Pranay, your deep understanding, faith in my abilities, and steadfast support have been instrumental in my journey, helping me to navigate challenges with resilience and dedication. Your presence in my life is a gift for which I am profoundly grateful. Thank you for being my confidant, my motivator, and a continuous source of inspiration.

Lastly, I must acknowledge my partner, Karan, my anchor and my wings. Karan, you have helped me realize my potential, encouraged me to follow my dreams, and supported me through every high and low. You are a pillar of strength that has supported me through every step of this endeavor. Your patience during my worst times and your presence in moments of joy have meant the world to me. I am incredibly fortunate to have you by my side.

To all of you, thank you from the bottom of my heart. Your support has not only made this journey possible but has also made it worthwhile.

CHAPTER 1

Preface And Contribution of Authors:

This thesis, an exemplar of original research, is presented in a manuscript format intended for academic publication. The manuscript, detailing my research, will be submitted to JAMA Surgery for peer review.

As the primary investigator, my involvement in this research was comprehensive and multi-faceted, encompassing the initial conceptualization of the trial, meticulous design of the trial protocol, rigorous participant recruitment processes, execution of the trial, advanced data curation practices, the innovative design of performance matrix extraction architecture, and the development and implementation of machine learning algorithms for the improved classification of expertise levels among surgical trainees. Moreover, I have undertaken the task of coding these algorithms, conducting an analysis of the outcomes, and our findings, thereby ensuring the integrity and reliability of the data collected. This work stands as a testament to my commitment to advancing the field of surgical education and artificial intelligence.

Trisha Tee's, expertise greatly enhanced the conceptual foundation of the trial. Her role was instrumental in designing the trial protocols, implementing the methodology, recruiting participants, and the hands-on management of the trial, reflecting a dedication that has significantly shaped the study's success.

Dr. Bilal Tarabay, whose insights into the realm of spine surgery provided a critical lens through which the trial was conceptualized. From the establishment of trial protocols to the identification of pivotal performance matrices that distinguish skilled from less skilled performances, Dr. Tarabay contributions have been indispensable. His provision of surgical expertise and resources has been a cornerstone in the development of this project.

Dr. Recai Yilmaz, whose expertise in the domains of data curation techniques and algorithmic benchmarking has profoundly influenced the methodological rigour of this study. Dr. Yilmaz's involvement ensured that our approaches to data analysis remained both innovative and grounded in empirical evidence.

The expertise in spine surgery and surgical knowledge of Dr Abdulmajeed Albeloushi, and Dr. Mohamed Albantobi, has provided constant support in understanding nuances of performance matrices to understand and validate the machine learning algorithm’s prediction with real life scenario.

Dr. Rolando Del Maestro, whose guidance and mentorship have been the guiding light of this research journey. Contributing to every facet of the project—from conceptualization and methodology to validation, administration, and funding acquisition—Dr. Del Maestro's supervision has been both a privilege and a profound learning experience. His expertise in neurosurgery and unwavering support have been instrumental in the project's development, execution, and interpretation of results.

This collaborative effort embodies a shared vision for the advancement of surgical training through the integration of artificial intelligence, marking a significant contribution to the field. Our collective endeavor reflects not only a commitment to scientific inquiry but also a dedication to improving surgical outcomes and education.

Abbreviations & Acronyms:

VR: Virtual Reality

AI: Artificial Intelligence

OSATS: Objective Structured Assessment of Technical Skills

PGY: Post Graduate Year

ANN: Artificial Neural Network

SVM: Support Vector Machine

KNN: K-Nearest Neighbours

REF: Recursive Feature Elimination

RFC: Random Forest Classifier

SHAP: Shapley Additive Explanation

ICEMS: Intelligent Continuous Expertise Monitoring System

Thesis Introduction:

Traditionally, surgical training has adhered to an apprenticeship model rather than a competency-based framework, posing challenges in objectively evaluating a trainee's competency (Van Heest et al., 2022). This method relies primarily on subjective assessments, lacking an objective quantitative methodology necessary for accurately gauging a trainee's performance, particularly in complex spine surgeries like pedicle screw insertions, known for their steep learning curves (Franzese & Stringer, 2007). The need to assess and quantify the complex aspects of surgical skill development is thus increasingly recognized, especially with the rapid evolution of technology offering insights into the composites of surgical tasks (Rogers et al., 2021).

Equally, the emphasis on patient safety during surgical training cannot be overstated. Historically, surgical trainees have learned procedures directly on patients, carrying significant risks (Coelho et al., 2014; Tang et al., 2005). Instances of patient harm due to trainee errors underscore the urgency for safer training modalities. Advances in simulation technology are revolutionizing the concept of surgical training, with medical schools increasingly integrating such technologies into their curricula to mitigate risks (Steadman et al., 2016).

The field of medical training is undergoing significant transformation with the integration of artificial intelligence (AI), particularly in surgical education (Guerrero et al., 2022). Exploratory methods spanning in vivo, ex-vivo, cadaveric studies, as well as augmented reality and virtual reality (VR), are being investigated (Patel et al., 2021; Yilmaz et al., 2022). AI's ability in processing vast datasets and identifying key learning parameters from simulator-generated data is paving the way for enhanced training methodologies (Yilmaz et al., 2022).

Previous initiatives at the Neurosurgical Simulation and Artificial Intelligence Learning Centre have laid the groundwork by developing and validating surgical simulators, alongside systems for analyzing extensive simulator data to provide real-time feedback on neurosurgical tasks (Yilmaz et al., 2022). This study aims to extend this innovative approach by validating the TSYM simulator, developed by the Montreal-based startup Symgery, designed for spine procedures (Symgery, 2023). Through AI, we intend to discern critical performance attributes in pedicle screw insertion by using simulated operations that distinguish skilled and less skilled participants performance.

CHAPTER 2

Background

History of thoracolumbar spinal instrumentation

Spinal procedures are critical interventions for managing a variety of debilitating conditions, particularly low back pain, which has become a significant public health issue in the Western world. It is estimated that over 80% of individuals will experience low back pain at some point during their lives, making it not only a leading cause of activity limitation but also the third most frequent indication for surgical intervention in the United States (Abumi et al., 1989; Dickman et al., 1992). This condition has led to a noticeable increase in the number of spinal operations performed in recent decades, reflecting its substantial clinical and economic impact.

The revolution in spinal surgery began with the introduction of internal fixation by Paul Harrington in 1975. Originally designed for deformity correction, the Harrington rod system was subsequently adapted for broader applications across various spinal conditions, significantly enhancing the surgical management of spinal disorders (Andén et al., 1980; Livingston & Perrin, 1978; Sundaresan et al., 1984; Wang et al., 1979). This innovation marked a seminal shift in the approach to spinal surgery, offering new therapeutic possibilities for patients.

Further advancements were made around 1975, when Eduardo Luque enhanced the Harrington system by integrating sublaminar wires, which improved spinal stabilization and expanded the system's clinical utility (Luque, 1982). The evolution of spinal fixation technologies continued into the 1980s with the introduction of the Cotrel Dubousset (CD) system. This system incorporated sophisticated hook-rod mechanisms specifically designed to manage complex spinal deformities, thereby broadening the surgical options available to clinicians (Cotrel & Dubousset, 1984).

During the same period, Roy-Camille made significant contributions to spinal instrumentation by pioneering the use of the pedicle for segmental fixation. His development of pedicle screws offered superior biomechanical stability and versatility, allowing for their application in diverse spinal segments and conditions (Abumi et al., 1989; Dickman et al., 1992; Roy-Camille et al., 1970). This innovation was further refined by Magerl in 1977, who introduced the "fixateur externe", an external fixation system utilizing pedicle screws. This concept was subsequently modified by Walter Dick, who developed the "fixateur interne", integrating the rods inside the body to enhance internal fixation strategies (Dick et al., 1985).

The adoption and popularization of pedicle screws in the United States were notably advanced by Arthur Steffee in 1984. His developments led to the preference for rod-based systems due to their flexibility and potential to facilitate spinal fusion, significantly advancing the capability for performing complex spine surgeries (Esses & Bednar, 1989).

These sequential innovations have fundamentally transformed the landscape of spinal surgery, offering enhanced surgical approaches and improved patient outcomes. Each development has built upon the previous, cumulatively enriching the surgical techniques that have become essential in the modern management of spine-related ailments.

Risk associated with pedicle screw insertion procedure

The intricate nature of spinal procedures is primarily due to the spine's proximity to critical neurological structures, significantly increasing the risk of complications. Surgical errors during these procedures can lead to a wide array of adverse outcomes, including neural damage, pulmonary embolism, neurological impairments, and surgical site infections. Furthermore, some patients may experience chronic back pain postoperatively, necessitating additional interventions in more severe cases (Dickman et al., 1992).

A fundamental technique in spinal surgery is the insertion of pedicle screws, which is essential for stabilizing and achieving fusion in the thoracolumbar spine. This technique is particularly crucial in treating conditions associated with axial instability, such as degenerative, neoplastic, and infectious diseases (Bono & Lee, 2004; Dick et al., 1985; Esses & Bednar, 1989). Despite technological advancements in surgical methods, the placement of pedicle screws involves significant risks of complications, underscoring the importance of surgeons and trainees mastering this procedure to minimize the likelihood of acute neurological deficits and the subsequent need for revision surgeries (Kim et al., 2004).

Research indicates that the rate of pedicle screw misplacement is notably variable, reported between 15.7% and 41% (Baird et al., 2017; Gelalis et al., 2012; Gonzalvo et al., 2009; Hicks et al., 2010). This variability emphasizes the critical need for enhanced training protocols and sustained proficiency among surgical staff.

The precision of pedicle screw placement is a focal point in spinal surgery, essential for avoiding complications and optimizing patient outcomes. Improper placement of screws can jeopardize the structural integrity of the vertebrae, posing significant risks to neural, vascular, and visceral structures, and potentially leading to severe clinical complications (de Kater et al., 2022; Gautschi et al., 2011; Kim et al., 2004; Sarwahi et al., 2016). While minor cortical breaches are often asymptomatic, they can lead to hardware failure, instability, diminished fusion rates, and accelerated adjacent level degeneration (Amaral et al., 2021; Aoude et al., 2015; Aoude et al., 2018; de Kater et al., 2022; Sarwahi et al., 2016; Gautschi et al., 2011).

Reviews of the literature reveal that the incidence of pedicle screw misplacement ranges from 4.2% to 8.7 %, influenced by patient demographics and specific surgical details (Gautschi et al., 2011; Hicks et al., 2010). A detailed study by Hicks et al. found that among these mispositioned screw, the most frequent breaches occurred in the lateral cortex (53%), followed by medial (24%), inferior (14%), superior (8%), and anterior (1%) breaches (Hicks et al., 2010).

Despite the relatively low frequency of critical complications in non-deformity cases—less than 0.5%—the consequences of misplacement can be dire, including nerve damage, cerebrospinal fluid leaks, instability, pseudarthrosis, and the need for revision surgeries, not to mention the potential for malpractice litigation (Adamski et al., 2023; Aoude et al., 2015; Gautschi et al., 2011; Sankey et al., 2020). Additionally, the literature notes rare but serious complications such as intraoperative pedicle fractures, screw loosening or pullout, and pulmonary effusion, highlighting the necessity for meticulous surgical techniques and vigilant postoperative monitoring (Gautschi et al., 2011).

Current Surgical Education, Challenges

Surgical education has historically been structured around an apprenticeship model, where trainees observe and learn from experienced surgeons before gradually performing procedures under supervision and eventually independently (Polavarapu et al., 2013; Grillo, 2018). This traditional approach allows knowledge to be passed down through generations of surgeons. However, it relies heavily on subjective assessments of skill by mentors, which may not capture the full complexity of surgical proficiency. For example, direct observation does not provide quantitative data on essential technical skills such as the amount of force applied to tissues during a procedure, which can be precisely measured using specialized equipment (Asad et al., 2014; McCall, 2016).

Despite its evolution, with more informed surgical educators utilizing advanced protocols, the apprenticeship model still faces significant limitations, particularly in the realm of objective performance assessment. The reliance on cadaveric training illustrates another challenge. While cadaveric dissections play a crucial role in surgical education, they do not fully replicate the dynamic nature of living tissues, nor do they present the same complications such as bleeding or post-operative issues (Asad et al., 2014; McCall, 2016). Additionally, there is a notable discrepancy between the demand for and supply of cadavers, with over 23,000 needed annually in the United States alone. Supplies frequently fall short, particularly impacting medical schools with less access to resources (Asad et al., 2014; McCall, 2016).

Moreover, the risk to patients during surgical training remains a significant concern. Trainees often do not receive sufficient hands-on experience before transitioning to live patient operations, raising the potential for errors that could result in patient harm (Polavarapu et al., 2013). This gap underscores the necessity for improved training methods that can safely bridge the transition from theoretical learning to practical application.

The response to these educational challenges has been a gradual shift from informal apprenticeships to structured competency-based training models. This transition was notably influenced by the seminal contributions of Sir William Osler, who advocated for early clinical exposure at McGill University and Johns Hopkins Medical School, as well as those of William S. Halsted, who introduced a progressive surgical training model emphasizing supervised training and gradual autonomy at Johns Hopkins Hospital (Grillo, 2018; Mueller, 2010; Waugh & Bailey, 2013). These developments laid the groundwork for more formalized and systematic surgical training programs.

The Flexner Report in the early 20th century further catalyzed the standardization of medical education by exposing deficiencies across U.S. and Canadian medical schools, leading to the establishment of the American College of Surgeons in 1912 with the aim of enhancing training standards (Polavarapu et al., 2013; Waugh & Bailey, 2013). This period also saw the introduction of nationwide examinations for medical graduates, fostering uniformity in medical education and ensuring a baseline competency among physicians (Cooke et al., 2006; Grillo, 2018; Mueller, 2010).

In recent decades, the growing complexity of surgical procedures and the imperative for patient safety have driven the adoption of Competency-Based Medical Education (CBME). This educational paradigm focuses on ensuring that surgical trainees develop specific competencies, assessed through structured frameworks such as the CanMEDS roles, which were first introduced by the Royal College of Physicians and Surgeons of Canada in 1996 and updated in 2015 (Frank et al., 2015; Iobst et al., 2010; Harris et al., 2020). Furthermore, the Competence by Design (CBD) program by the Royal College aims to overhaul postgraduate medical education by delineating clear competencies for each stage of training and employing Entrustable Professional Activities (EPAs) to monitor progress (Frank et al., 2015; Frank et al., 2017; Harris et al., 2020; Stockley et al., 2020).

The shift towards CBME represents a significant paradigm shift in surgical education, emphasizing outcome-focused learning and skill mastery over time-based progression. This approach is designed to ensure that all trainees are adequately prepared for independent practice, enhancing patient safety and the overall quality of care. The American Board of Surgery's integration of EPAs reflects a broader commitment to advancing competency-based surgical training across various specialties, aiming to standardize the assessment and certification of surgical competencies globally (Patel et al., 2020; Sonnadara et al., 2014; Frank et al., 2017; Frank et al., 2017; Stockley et al., 2020).

Surgical Educations and Simulators

Surgical education has experienced a significant evolution with the advent of simulation-based training, an approach that has transformed traditional methodologies. Originating from the aviation industry's model, simulation training in surgery emphasizes the acquisition of skills through repetitive practice and reflection. Resnick and MacRae have categorized the development of surgical skills into cognitive, integrative, and autonomous phases, a framework that closely aligns with pilot training programs focused on skill development through pattern recognition (Reznick & MacRae, 2006).

Initially adopted in the 1960s, medical simulators have progressively embraced advanced computerized VR platforms since the late 1990s (Badash et al., 2016). These technologies enable surgical trainees to engage in immersive experiences that closely replicate real-world scenarios, thereby enhancing their procedural skills in a risk-free environment. Simulation laboratories provide a controlled setting where trainees can practice various surgical techniques, from basic suture tying to complex robotic surgeries, without the immediate pressures of the operating room (Ray et al., 2013).

One of the key benefits of simulation is the ability to objectively assess and improve surgical skills using standardized tools such as the Objective Structured Assessment of Technical Skills (OSATS) (Martin et al., 1997; Hatala et al., 2015). This allows for structured feedback and quantifiable benchmarks of trainee performance, facilitating a smoother transition to operating on real patients with increased confidence and proficiency (Yilmaz et al., 2022).

VR technology has particularly revolutionized training for minimally invasive procedures like laparoscopic and endoscopic surgeries, offering high-fidelity simulations that include realistic haptic feedback mechanisms (Alaraj et al., 2015; Chan et al., 2013). The Minimally Invasive Surgery Trainer – Virtual Reality (MIST-VR) is an example of how VR can significantly enhance surgical performance and reduce procedural errors (Gallagher et al., 2004; Wilson et al., 1997).

However, the application of VR in spine surgical training has not progressed as rapidly as in other surgical specialties such as laparoscopic or robotic surgery. Most commercially available spine simulators focus on less complex procedures like vertebroplasty and pedicle screw placement, with limited access to simulators that cover more advanced interventions such as anterior cervical discectomy and fusion or scoliosis surgery (Pfandler et al., 2017). This gap is attributed to various challenges, including the difficulty of replicating the detailed anatomical structures and the different force dynamics required between soft tissues and bone, as well as the high costs associated with developing these comprehensive simulation platforms (Pfandler et al., 2017).

Moreover, spine surgery simulations demand a high degree of anatomical accuracy and realistic tactile feedback, essential for tasks that require precise manipulation such as bone drilling (Ray et al., 2013; Vaughan et al., 2016). Although technological barriers such as the limitations of haptic devices and the slow response rates of simulated tools exist, ongoing advancements are improving the fidelity and responsiveness of VR platforms.

The NeuroVR simulator platform consists of a stereoscope through which a three-dimensional image is projected for the participant to interact with Figure 2.1. The user manipulates the simulation object through a bimanual haptic system using both hands together, mimicking the bimanual tasks faced in a live operating environment. The finite element method numerically models the structures in the scenario and applies a physics-based approach permitting mechanical characterization of the simulated structures. Although more computationally demanding the finite element method for a series of brain tumor and spine related procedures (Delorme et al.,2012).

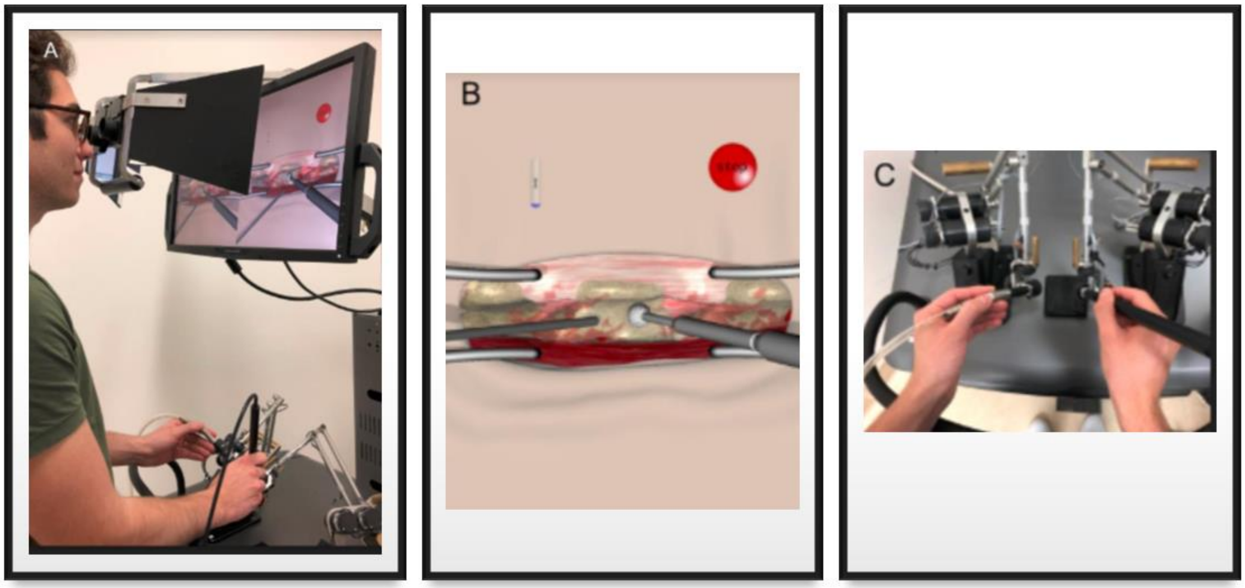

Figure 2.1 The NeuroVR simulation platform. A. Participant performing a laminectomy on the NeuroVR Platform. B: Simulated laminectomy procedure. C. Bimanual force feedback handles.

As VR technology continues to evolve, it is anticipated that more sophisticated simulators capable of facilitating more complex spine procedures will become available. This advancement holds the potential to mitigate risks associated with surgical errors and enhance the training outcomes, ultimately ensuring better patient safety and more competent surgical practitioners.

Review of existing simulators for spine surgery

In the evolving landscape of spinal surgery education, VR simulators have emerged as pivotal tools, offering an immersive, interactive, and detailed platform for surgeons to hone their skills safely. These simulators blend advanced technology with practical surgical training, providing a dynamic environment that mimics real-life scenarios.

Immersive Touch has advanced visualization and haptic feedback capabilities. It utilizes head and hand tracking via robotic arms to dynamically adjust the user's perspective, simulating a wide range of spinal surgery scenarios, including pedicle screw placement, vertebroplasty, and lumbar puncture (Alaraj et al., 2013; Luciano et al., 2005; Luciano et al., 2011; Luciano et al., 2013). This system's strengths lie in its immersive experience, integration of patient-specific imaging, and versatility across multiple procedures. Despite these advantages, the system does not provide audio feedback, which could otherwise enhance the immersive experience. Moreover, there is a notable gap in validation studies specifically focused on the accuracy and effectiveness of the simulated procedures, limiting its empirical endorsement (Alaraj et al., 2013; Gasco et al., 2014; Luciano et al., 2005; Luciano et al., 2011; Luciano et al., 2013; Roitberg et al., 2013).

The Virtual Surgical Training System (VSTS) offers targeted training with its screen-based VR environment and a robotic arm for tactile feedback, focusing on cervical spine drilling and thoracic pedicle screw placement (Shi et al., 2018). The simulator's realistic anatomical models derived from actual human spines provide an accurate base for training. However, its use of a 2D screen to display 3D scenarios may impair depth perception and immersion. Furthermore, while initial validation efforts have been made, the system lacks comprehensive validation across critical metrics like face, content, and construct validity, which are essential for confirming the simulator's educational efficacy (Hou et al., 2018; Shi et al., 2018).

The Immersive Virtual Reality Surgical Simulator for Pedicle Screws Placement (IVRSS-PSP) is tailored specifically for pedicle screw placement. It features a heads-up display and a robotic arm that collectively enhance the realism of the operative environment. This simulator promises a highly immersive training experience with its realistic spine models and 3D-printed tools that mimic actual surgical instruments. Despite these features, the simulator's focus remains narrowly confined to pedicle screw placement, and it lacks extensive validation studies that would substantiate its effectiveness and reliability in broader surgical training context (Xin et al., 2019; (Xin et al., 2020).

Sim-Ortho® provides a comprehensive VR training environment with voxel-based simulation technology, stereoscopic 3D glasses, and haptic plus auditory feedback (Figure 2.2). It supports a wide array of spinal procedures, making it a versatile tool for training in complex surgeries. While some procedures like anterior cervical discectomy and fusion (ACDF) have undergone validation, others within its range have not, raising questions about the overall reliability and effectiveness of the simulator across its full range of applications (Bakhaidar et al., 2023; Ledwos et al., 2021; Reich et al., 2022).

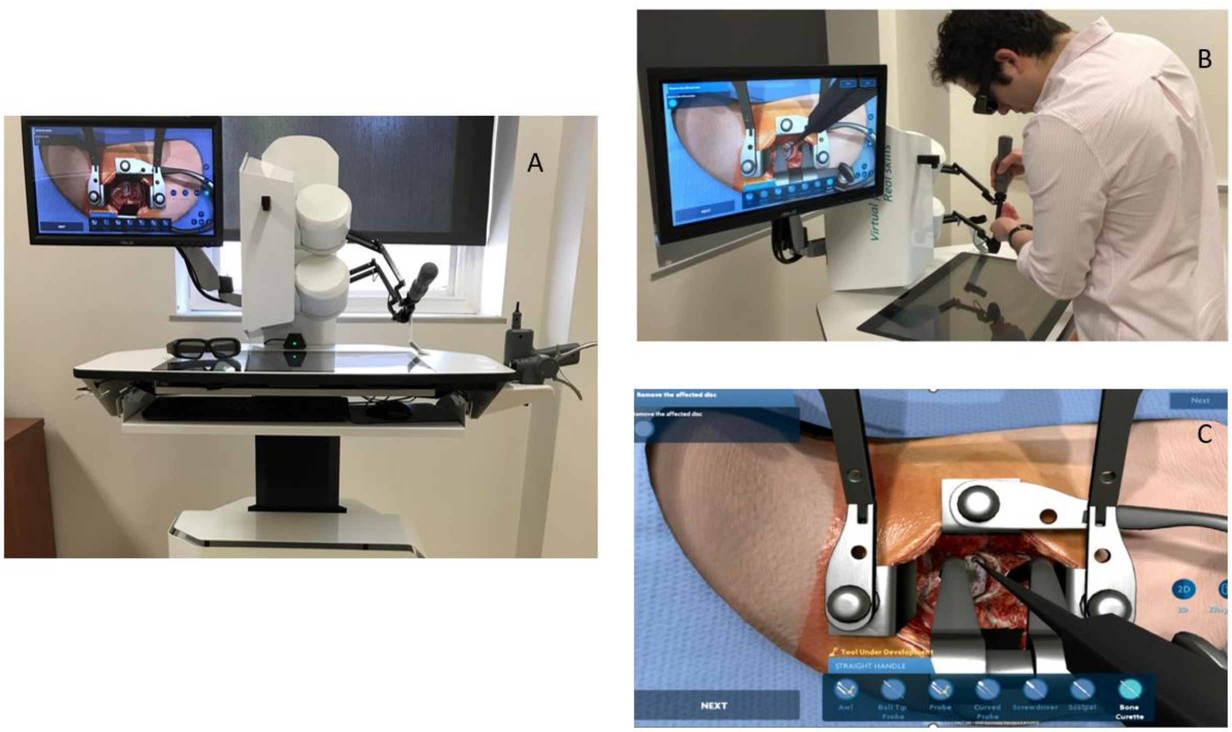

Figure 2.2 The Sim-Ortho® virtual reality simulator platform. **A.** Sim-Ortho® virtual reality

simulator. **B.** Participant performing a simulated anterior cervical discectomy and fusion procedure on the Sim-Ortho® platform. **C.** Immersive 3D view of the simulated VR discectomy component.

A Custom VR Simulator provides a personalized training approach with its HUD headset and interactive controllers, focusing primarily on pedicle screw placement and drilling. The use of patient-specific models enhances the realism of the training scenarios. However, this simulator suffers from a lack of comprehensive validation and a narrow focus, which may limit its utility in broader surgical education (Chen et al., 2021).

TSYM Spine Simulator

The TSYM Symgery Virtual Reality Surgical Simulator, developed by Cedarome Canada Inc. known as Symgery, in Montreal, Canada, represents an advanced improvement in the domain of spinal surgery training. This state-of-the-art VR simulator leverages a voxel-based system to construct a detailed three-dimensional (3D) representation of the intraoperative surgical environment, facilitating a highly realistic and non-immersive training experience (Figure 2.3)

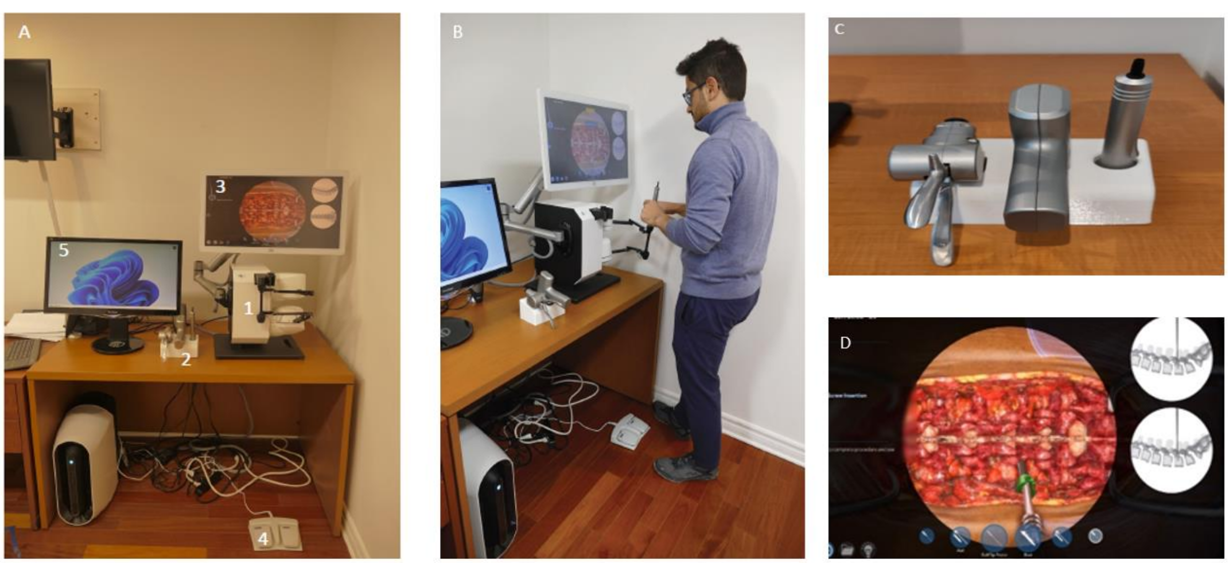

Figure 2.3: The virtual reality platform used to simulate the L5-L4 bilateral pedicle screw insertion. **A.** The TSYM simulator set up, showing the (1) robotic arm that uses and provide advanced haptic feedback technology, (2) the different tool handles that can be used in the simulated scenario, (3) 3D monitor, (4) pedals for activating fluoroscopy and (5) secondary monitor. **B.** The participant interacting with the platform. **C.** A variety of instruments are available accompanied by different handles to simulate each instrument haptics. **D.** The platform provides very realistic 3D graphics along with appropriate Xray images. Utilizing sophisticated haptic feedback technology, the TSYM Symgery simulator allows participants to experience tactile sensations and realistic tissue handling akin to actual surgical conditions. This system enhances the learning process by enabling trainees to develop essential surgical skills through interaction with virtual surgical instruments that mimic the look and feel of real tools used in spinal procedures (Figure 2.3 C and D).

The simulator supports a variety of spinal surgical procedures, including complex tasks such as laminectomy and pedicle screw placement. Participants engage with these procedures using the simulated tool handles, gaining practical experience in a controlled environment. Complementing the tactile feedback, the TSYM Symgery simulator also offers comprehensive auditory and visual feedback, including sounds from patient cardiac monitoring and surgical instruments, which serve to deepen the immersion and realism of the training experience.

Furthermore, the simulator is equipped with capabilities to record detailed performance metrics at ever 2 microseconds. This data collection includes metrics such as force applied, instrument tracking, tissue removal rates, velocity, and acceleration. Such data is crucial for the thorough assessment and feedback on participants' performance. After completing a task, the simulator generates a three-dimensional vertebral body structure, highlighting the final positions of inserted pedicle screws, thus providing feedback that can be used as an educational tool to visually demonstrate the outcomes of surgical actions.

Considering these capabilities, the integration of the TSYM Symgery into neurosurgery and orthopedics residency programs could significantly enhance both the training and assessment of surgical skills, promoting a deeper comprehension of surgical expertise related to spinal procedures. Despite its potential, there is a notable lack of published data on its effectiveness or reliability, which is crucial for establishing its credibility as a training tool. At our lab another team is working on validating face, content, and construct validity of this simulator.

Overall, TSYM spine simulator along with other VR simulators collectively represent a significant advancement in surgical training, offering diverse benefits from immersive experiences to accurate anatomical modeling. However, they also share common challenges, such as insufficient validation and limited procedural coverage. As these technologies continue to develop, ongoing research and evaluation are essential to ensure they meet educational standards and effectively enhance surgical competency in the field of spine surgery.

Objective Assessment of Technical Skills using Machine Learning Techniques:

The integration of VR simulation platforms in surgical education has significantly advanced the field by enabling detailed data collection from various aspects of surgical procedures. These platforms capture extensive data on instrument utilization, force exerted, instrument activation, and interactions with tissues, which are crucial for objectively assessing technical skills in surgery (Sawaya et al., 2017; Sugiyama et al., 2018). This rich dataset is instrumental in evaluating surgical proficiency, particularly in neurosurgical training where precision and the gentle handling of delicate tissues are paramount.

Excessive force application by surgical instruments, often linked to suboptimal outcomes, is a critical factor assessed during training. Innovative methodologies leveraging VR simulation data, such as force pyramid and force heatmap models, have been developed to provide both two-dimensional and three-dimensional visualizations of force distribution (Sawaya et al., 2017; Azarnoush et al., 2016). These models are vital for imparting insights into effective surgical techniques and optimal hand postures, thereby enhancing surgical skill sets (Azarnoush et al., 2016; Sawaya et al., 2017). Furthermore, these tools have demonstrated efficacy in differentiating between the skill levels of surgeons and delivering targeted feedback for improvements (Sawaya et al., 2018).

To further enhance the objectivity in skill assessment, performance metrics have been established as a standard quantitative measure. These metrics assess key operational aspects such as safety, efficiency, quality, bimanual dexterity, and instrument maneuverability during simulated procedures (Alotaibi et al., 2015; AlZhrani et al., 2015; Bissonnette et al., 2019). For instance, the NeuroVR platform, capable of recording data at a rate of 50 points per second, analyzes approximately 6600 performance metrics from a single tumor resection task, highlighting the granularity and depth of the evaluation process (Winkler-Schwartz et al., 2019).

In this context, AI is emerging as a transformative force in surgical training, introducing 'intelligent systems' that can process extensive datasets and provide high-fidelity assessments (Nagi et al., 2023). These AI-driven systems deliver sophisticated feedback, akin to a human instructor, and play a pivotal role as decision-makers in educational settings. Despite initial overextensions in the use of 'intelligence' in surgical training applications, the path towards fully integrated intelligent systems is evident and holds great promise for the future of surgical education.

At the Neurosurgical Simulation and Artificial Intelligence Learning Centre, the Intelligent Continuous Expertise Monitoring System (ICEMS) exemplifies the application of AI in enhancing training outcomes (Yilmaz et al., 2022). This system assesses the bimanual performance of neurosurgical trainees by continuously analyzing surgical performance at intervals of 0.2 seconds and provides real-time feedback at the tool level. This intelligent tutoring system has shown proficiency in categorizing skill levels across different stages of surgical training—from students to neurosurgeons—and can predictively validate technical skills throughout a surgical residency. In a randomized controlled trial training with real-time ICEMS feedback resulted in significantly better performance outcomes compared to no real-time feedback and in-person instruction and similar OSATS ratings compared to in-person training with expert instruction (Yilmaz et al, 2024). Intelligent tutoring systems may help improve the methods that bimanual operating skills are assessed and taught, providing tailored, quantifiable feedback along with actionable instructions in real-time**.** The ICEMS is one of the few systems capable of real-time performance assessment using technologies like machine learning algorithms to analyze tool-generated data (Nagi et al., 2023).

The ongoing advancements in AI and VR technologies are set to revolutionize surgical training, promising more effective learning experiences and higher competency among future surgeons, thereby ensuring enhanced patient safety and care quality.

Machine Learning Algorithms and Feature Selection Methods:

Features Selection:

We used a combination of feature selection techniques to iterate the most important features to distinguish skilled versus less skilled performance of the screw. The rationale behind using multiple feature selection techniques is to leverage on various strong points of each technique like immunity against noise in data, robustness and generalizability. We applied four primary methods for feature selection which are Wrapper Based, Embedded based, Filter Based and Weight Based.

1. **Wrapper Based Recursive Feature Elimination (REF) using an SVM model:** The fundamental concept behind RFE is the iterative training of the model, eliminating the least important features until the optimal features are identified. This wrapper-based technique filters features based on their performance in predicting the output (classification) (Jeon & Oh, 2020). RFE can be used with various machine learning models, including Support Vector Machines (SVM), to rank features based on their importance.
2. **Embedded-Based Feature Selection Using Random Forest:** This is a popular ensemble learning technique that is also an embedded method for feature selection. It is robust to overfitting and noise, providing stable feature importance rankings (Louppe, 2014). Random Forest can capture non-linear relationships between features and the target variable, making it a powerful tool for identifying and ranking important features in a dataset.
3. **Filter-Based Feature Selection:** SelectKBest, is a filter-based method approach that utilizes statistical test scores to evaluate the relationship between each feature and the target variable. This method selects the top K features based on their scores from these tests, such as chi-square tests for classification tasks or ANOVA F-value for regression tasks (Fitriani

et al., 2022). The strength of SelectKBest lies in its simplicity and computational efficiency, making it suitable for high-dimensional datasets. By filtering features based on statistical relevance, SelectKBest effectively reduces dimensionality, helping to alleviate overfitting and enhance model performance (Fitriani et al., 2022). To determine the optimal value for K, various values were tested, and cross-validation was employed to identify the most influential features. This ensures that the selected features contribute significantly to the predictive power of the model.

4. **Weight Based:** Weight-based feature selection in a multi-layer perceptron (MLP) involves identifying and prioritizing input features based on the weights assigned during the training process. In this methodology, each connection between neurons across all layers has an associated weight that determines the strength of the connection (Sun et al., 2013). During training, these weights are adjusted to minimize the error between predicted and actual outputs using the backpropagation algorithm. The significance of a feature is deduced from the magnitude of its associated weights; larger weights indicate a more pronounced influence on the network's output (Sun et al., 2013). This involves analyzing the combined effect of weights from the input layer through all hidden layers to the output layer. Features are ranked based on the absolute values of their weights across all layers, considering the product of weights along the paths from the input features to the output. By focusing on the features with the most significant cumulative weights, this method reduces noise and enhances the model's performance, providing a clear understanding of which features are most important for the network's decisions.

Machine Learning Algorithms

A comprehensive benchmarking analysis can be used to assess the performance of diverse machine learning algorithms. The selected algorithms encompass a range of methodologies, including Support Vector Machine (SVM), Random Forest, K-Nearest Neighbours KNN), and ANN can be utilized as outlined below.

1. **Support Vector Machine (SVM):** A supervised machine learning algorithm, SVM can be employed for classification and regression tasks. It operates by identifying an optimal hyperplane that effectively separates data points within the input space. The crucial "support vectors" represent the data points closest to the decision boundary, influencing the hyperplane's position and orientation (Hearst et al., 1998). SVM strives to maximize the margin between classes in the training data, which contributes to a robust decision boundary.
2. **Random Forest:** Random Forest stands as an ensemble learning algorithm that orchestrates the construction of numerous decision trees during training and derives predictions by aggregating individual tree outputs, resulting in either the mode for classification or mean prediction for regression (Paul et al., 2018). In this process, each decision tree in the forest is trained on a random subset of the training data, introducing variability through bootstrapping or bagging. Additionally, randomness is injected by considering a random subset of features at each split in the decision tree.
3. **K- Nearest Neighbours:** K-Nearest Neighbors (KNN) stands as a straightforward and intuitive supervised machine learning algorithm, adept at handling both classification and regression tasks. In KNN, predictions hinge on either the majority class (for classification) or the average of neighboring data points (for regression) among the K nearest neighbors to a given input data point (Peterson, 2009). The algorithm efficiently stores the entire training

dataset in memory and, during prediction, identifies the K data points closest to the input point using a distance metric, commonly the Euclidean distance. The predicted output is then determined by the majority class or the average value of the K neighbors. Exploration involves tuning the model by experimenting with various parameter combinations.

4. **Artificial Neural Network:** An Artificial Neural Network (ANN) stands as a computational model inspired by the intricate structure and functioning of the human brain. Comprising interconnected nodes (neurons) organized into layers (input, hidden, and output), ANNs undergo a learning process by adjusting the weights of connections between nodes (Krenker et al., 2011). This adjustment often guided by backpropagation, enables ANNs to discern patterns, along with relationships and representations within data.

Chapter 3

Preface:

Study Rational and Objectives:

Rationale:

The current landscape of spinal surgical education, particularly in the realm of pedicle screw insertion, faces significant challenges that are yet to be fully addressed by existing training methodologies. Despite the critical importance of pedicle screw insertion in spinal surgeries—a procedure known for its steep learning curve—there is a conspicuous gap in the application of advanced data-driven technologies to enhance surgical training. Current spine surgical simulators, while collecting vast amounts of operational data, have not been effectively utilized to dissect and improve the nuances of surgical performance, especially in tasks as complex and delicate as pedicle screw insertion (Sugiyama et al., 2018).

With the rapid evolution of technology, particularly AI, there exists an unprecedented opportunity to revolutionize surgical training. AI's capability to process extensive datasets and generate meaningful insights into surgical performance can significantly advance training methodologies. This study aims to harness AI to analyze the extensive data captured by VR simulation platforms, specifically focusing on identifying the subtle patterns and critical parameters that distinguish skilled from novice performance in pedicle screw insertion.

Integrating AI-driven assessments into the TSYM simulator promises to not only refine the evaluation and training of surgical skills but also to enhance the scope and quality of surgical education in spine surgery. By developing and applying sophisticated models this study will provide a more structured and objective assessment of surgical proficiency, moving beyond the limitations of direct observation and subjective judgment (Azarnoush et al., 2016; Sawaya et al., 2017).

This approach will also address the significant risks associated with pedicle screw insertion, such as neural damage and other post-operative complications, by providing a platform for repetitive, focused training. This will enable less skilled trainees to master the necessary skills in a risk-free environment before performing human procedures, potentially decreasing the incidence of complications associated with less skilled errors.

Overall, this study seeks to fill the critical gap in current surgical training for pedicle screw insertion by leveraging advanced AI tools to analyze performance data from VR simulators. This integration aims to elevate the training protocols, ensure a higher degree of precision and safety in surgical procedures, and ultimately enhance patient outcomes by fostering a deeper and more scientifically grounded understanding of the complexities involved in spine surgery.

Objectives:

The objectives of this case series study were to:

1) employ a combined feature selection process to identify the most important features that differentiate skilled and less skilled surgical performance for simulated pedicle screw insertion on the TSYM platform.

2) Benchmark various machine learning algorithms to classify skilled versus less skilled performance of simulated pedicle screw insertion.

Manuscript

Title: Artificial Intelligence Assessment of Expertise in Virtual Reality Spine Pedicle Screw Insertion

Puja Pachchigar, MSc[1,3], Trisha Tee, BSc[1,3], Bilal Tarabay, MD[1,3], Recai Yilmaz, MD,PhD[1,3], Abdulmajeed Abeloushi, MD[1,2,4], Mohamed Alhantoobi, MD MSc[1,3,5,6], Rolando F. Del Maestro, MD, PhD[1,2]

[1] Neurosurgical Simulation and Artificial Intelligence Learning Centre, Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, 300 Rue Léo Pariseau, Suite 2210, H2X 4B3, Montreal, Quebec, Canada

[2] Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, 3801 Rue University, Montreal, Quebec, Canada, H3A 2B4

[3] Faculty of Medicine and Health Sciences, Department of Experimental Surgery, McGill University, Montreal, Canada

[4]Department of Neurosurgery, Ibn Sina Hospital, Ministry of Health, Kuwait.

[5] Department of Neurosurgery, Hamilton General Hospital, McMaster University Medical Centre, 237 Barton St E., Hamilton, ON L8L 2X2, Ontario, Canada

[6] Department of Neurosurgery, Zayed Military Hospital, Abu Dhabi, United Arab Emirates

*** Corresponding Author**

Puja Pachchigar MSc[1,2]

Neurosurgical Simulation and Artificial Intelligence Learning Centre,

Montreal Neurological Institute and Hospital,

300 Rue Léo Pariseau, Suite 2210, Montreal, Quebec, Canada, H2X 4B3.

E-mail : puja.pachchigar@mail.mcgill.ca

Funding This study was supported by Mitacs Accelerate Grant, grants from Brain Tumour Foundation of Canada Brain Tumour Research Grant, a Medical Education Research Grant from the Royal College of Physicians, the Franco Di Giovanni Foundation and the Montreal Neurological Institute and Hospital, McGill University. Cedarome Canada Inc. dba Symgery supplied the TSYM Symgery virtual reality non-immersive simulator platform utilized for these investigations.

Disclosures Puja Pachchigar, Trisha Tee, and Bilal Tarabay are supported by a Mitacs Accelerate Internship Grant. Trisha Tee also received support from a Masters-CIHR. Dr. Recai Yilmaz was supported by a Brain Tumour Foundation of Canada Brain Tumour Research Grant, a Medical Education Research Grant from the Royal College of Physicians, a Max Binz Fellowship from McGill University Internal Studentships, and a grant from the Fonds de recherche du Quebec–Santé. The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

Acknowledgments

The authors would like to thank all the neurosurgeons and orthopedic spine surgeons, fellows along with neurosurgical and orthopedic residents who participated in this study. Special thanks to Dr. Ahmed Aoude for allowing the authors to use the Orthopedic Research Laboratory, Montreal General Hospital for these studies and Dr. Greg Berry for helping organize the use of orthopedic facilities and recruiting orthopedics residents for the study. The authors also thank Dr. Zhi Wang, Dr. Sung-Joo Yuh, Dr. Ahmed Aoude, Dr. Lucy Luo, Dr. Ahmad Alsayegh, Dr. Mohamad Bakhaidan, Dr. Carlo Santaguida, Nour Abou Hamdan, and Dr. Abdulrahman Almansouri for their help recruiting trial participants.

ABSTRACT

IMPORTANCE: Our understanding of the composites of technical expertise during spinal procedures including the insertion of pedicle screws is incomplete. Datasets generated from surgical simulation allows the quantitation of psychomotor skills, which can be analyzed using machine learning algorithms which allows a more complete understanding of surgical performance.

OBJECTIVE: The primary aim of this study was to identify important features distinguishing skilled and less skilled levels of expertise during simulated pedicle screw insertion. The secondary aim was to benchmark the classification accuracy of surgical performance through the implementation of machine learning algorithms.

DESIGN: Participants from four universities were recruited between July 15, 2022, and May 31, 2023, to participate in a case series study. Data were collected over a single time point and no follow-up data were collected. Participants were classified a priori as either skilled or less skilled based on their experience in performing human pedicle screw insertion procedures.

SETTING: McGill University Neurosurgical Simulation and Artificial Intelligence Learning Centre.

PARTICIPANTS: Forty-three neurosurgery and orthopedic spine surgeons, spine fellows, and neurosurgery and orthopedic residents.

INTERVENTION: Insertion of bilateral L5 and L4 pedicle screw insertions on a virtual reality platform resulting in 172 inserted screws for analysis. These 172 datapoints were divided into training set (70% - 121 data points) and testing set (30% -51 data points) for algorithm's training and testing. We used 5-fold cross validation to validate the algorithm.

EXPOSURES: All participants performed a simulated virtual reality L5-L4 bilateral pedicle screw insertion during which they each inserted 4 screws.

MAIN OUTCOMES AND MEASURES The main outcomes and measures were determined through an iterative process, wherein features related to instrument movement, force application, and tissue resection were chosen from the raw simulator data output. This selection was achieved through a combination of four feature selection methods, wrapper-based, embedded, filter-based, and weight-based, in conjunction with Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (KNN), and Artificial Neural Network (ANN) models. The objective was to accurately assess the skill levels of participants in simulated pedicle screw insertion.

RESULTS A cohort of 43 participants, including 5 women and 38 men with a mean age of 33.6 years (SD 9.5), was evaluated. Machine learning models demonstrated varying accuracies on the test set: SVM achieved 78%, Random Forest 80%, KNN 82.3%, and ANN 82.3%. Analysis revealed 24 common features across Random Forest, KNN, and ANN, each achieving a classification accuracy of over 80%.

CONCLUSIONS AND RELEVANCE By employing machine learning algorithms, our study identified key features that may determine components of expertise during simulated pedicle screw insertion. We introduced a combined approach for feature selection that could enhance the accuracy of classifying skilled versus less skilled performance in future experiments. This method may prove valuable in the assessment and training of various surgical procedures.

Introduction

Surgical education is undergoing a transformative phase with the integration of advanced technologies aimed at enhancing training efficacy and patient safety (Varghese et al., 2024). Spinal procedures, including pedicle screw insertion, are complex due to the possible resulting injury to surrounding anatomical structures. (Ansorge et al., 2023; Elmi-Terander & Skulason, 2023; Kim et al., 2004). The traditional apprenticeship model relies on subjective assessments and direct observation, which does not capture all composites of expertise of surgical performance (Hashimoto et al., 2018; Shahzad & Anwar, 2021). This has necessitated the exploration of more objective and quantitative methods to evaluate and enhance surgical skills, particularly in procedures known for their steep learning curves and high risk of complications, such as pedicle screw insertion (Andersen et al., 2023; Halvorsen et al., 2022).

The advent of virtual reality (VR) simulation platforms has provided a forum for new research involving surgical training, allowing the simulation of real-world scenarios within controlled environments (Grillo, 2018; Polavarapu et al., 2013; Alaraj et al., 2015; Luciano et al., 2005; Luciano et al., 2011; Luciano et al., 2013). These simulators can derive large amounts of data involving complex surgical tasks and can provide real-time feedback to the learner (Mirchi et al., 2020; Reich et al., 2022; Yilmaz et al., 2022). Despite their capabilities, there remains a significant gap in their utilization for psychomotor performance analysis, particularly involving the integration of Artificial Intelligence (AI) (Alaraj et al., 2015; Luciano et al., 2013). The capability of AI to process and analyze vast datasets generated from VR simulations provides an opportunity to improve surgical training by increasing our understanding of the critical metrics involved in surgical expertise (Winkler-Schwartz et al., 2019). The relative importance of each metric identified can be assessed by utilizing artificial neural networks (ANN) (Alkadri et al., 2021; Bakhaidar et al., 2023). By extracting detailed psychomotor performance data from simulation operative data, AI can provide insights into the surgical composites of expertise that differentiate less skilled and skilled operative performance (Dickman et al., 1992; Sundaresan et al., 1984).

Despite these advancements, the application of AI in the assessment of surgical performance for pedicle screw insertion has not been extensively explored (Jia et al., 2023; Ma et al., 2022). Existing research predominantly focuses on other types of simulated spinal interventions (Alkadri et al., 2021; Ledwos et al., 2021). Pedicle screw insertion is technically demanding and associated with significant variability in clinical outcomes, often contingent on the surgeon's skill and experience (Manbachi et al., 2014; McGaghie, 2015). Misplacement of pedicle screw can lead to severe complications, including neurological damage and structural instability, which underscores the critical need for comprehensive training and assessment (Gang et al., 2012; Gonzalvo et al., 2009).

The TSYM Symgery VR platform is a non immersive VR simulator, employs a robotic arm and various tool handles which utilizes advanced haptic feedback technology to provide a realistic operative experience. This system can simulate, record, and allows a detailed analysis of complicated spine procedures, such as lumbar pedicle screw insertion.

The objectives of this case series study were 1) employing a novel performance metrics selection process identify the most important performance metrics that differentiate skilled and less skilled surgical performance for simulated pedicle screw insertion on the TSYM platform. 2) utilizing various machine learning algorithm improve classification of skilled vs less skilled performance of simulated pedicle screw insertion applying selected performance metrics found through novel approach.

Method

Participants

Forty-three neurosurgical and orthopedic residents, fellows, and neurosurgical and orthopedic spine surgeons participated in this case series study. Participants were categorized a priori into two groups: skilled and less skilled. The skilled group included individuals who had inserted at least one pedicle screw, either supervised or independently, during a human spinal procedure. The less skilled group consisted of participants with no such experience The skilled group included individuals who had inserted at least one pedicle screw, either under supervision or independently, during a human spinal procedure. The less skilled group consisted of participants with no such experience. This criterion of at least one pedicle screw insertion was chosen to ensure a balanced distribution of participants in each category. By setting the threshold at one screw, we aimed to achieve a more equitable representation between the skilled and less skilled groups. Additionally, the decision to create a-priori groups based on actual procedural experience, rather than traditional training levels, was made to gain deeper insights into skill levels. This approach allowed for a more nuanced understanding of expertise beyond just the amount of training. An informed consent approved by the Neurosciences-Psychiatry McGill University Health Center Research Ethics Board was signed by all participants. A demographic questionnaire was completed after consent and standardized written and verbal instructions regarding the steps and instruments available to complete the simulated L4-L5 bilateral pedicle screw insertion on the TSYM simulator were provided. A dry lab and an L2 simulated laminectomy procedure were used to acquaint participants with the TSYM simulator and simulated tools and their functions (Supplementary Digital Content). When these two tasks were completed, a simulated L4-L5 bilateral pedicle screw insertion was performed. Each step was dependent, and once completed, required participant confirmation before proceeding. No time limit was imposed. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Declaration of Helsinki (World Medical Association, 2013). This report is structured according to guidelines for Best Practices in reporting studies on machine learning to assess surgical expertise in virtual reality simulation and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines (Cheng et al., 2016; Winkler-Schwartz et al., 2019).

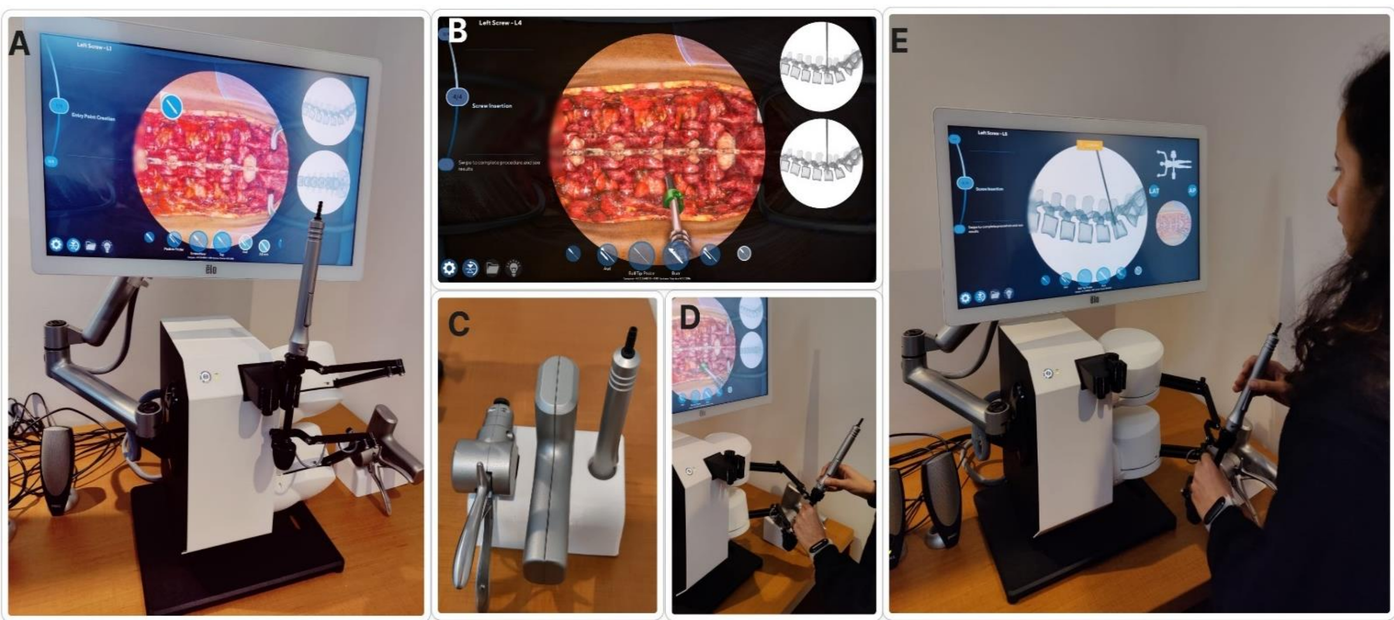

Figure 3.1 The virtual reality platform used to simulate the L5-L4 bilateral pedicle screw insertion. **A.** The TSYM simulator is designed to simulate various scenario. **B.** The platform provides very realistic 3D graphics along with appropriate Xray images. **C.** A variety of instruments are available accompanied by different handles to simulate each instrument haptics. **D.** Participant holds the instrument in their dominant hand, receiving haptic feedback when interacting with anatomic structure. **E.** The participant interacting with the platform.

Virtual Reality Simulator Platform

In this study the TSYM Symgery simulation platform, developed by developed by Cedarome Canada Inc. dba Symgery, (Montreal, Canada), was utilized (Figure 3.1 A). The three-dimensional (3D) intraoperative spinal surgical procedures present in this simulator rely on a voxel-based system. In a voxel-based system, the anatomy is represented by small, three-dimensional cubes called voxels, which are the 3D equivalent of pixels in a 2D image (Ledwos et al., 2021). Each voxel contains specific information about the tissue it represents, such as density and composition, allowing for detailed and accurate simulation of surgical procedures. This approach enables the simulator to create a highly realistic and interactive environment for training and assessment (Figure 3.1 B) (Ledwos et al., 2021). The simulator consists of a single haptic arm that provides continuous tactile feedback during operator manipulation of the surgical instruments employed to complete the task (Figure 3.1 C) and generates appropriate auditory and visual information for each tool employed. This system is equipped with pre-programmed surgical tools and captures multiple performance metrics, enabling a comprehensive surgical performance analysis. The pedicle screw insertion simulation task consists of one animated step and four deconstructed interactive steps. The animated step, where the animated lumber spine is exposed, is skipped in this study as the training for exposing the lumber spine is not within the scope. The focus is directly on the pedicle insertion task and its four interactive steps. The series of 4 deconstructive steps were repeated for each screw. For standardization purposes, users inserted 6.5 x 45 mm pedicle screws in a predetermined order left L5, left L4, right L5, right L4, inserted using a predefined magnification (Supplemental Digital Content Simulated L4 & L5 pedicle screw placement scenario). Participants could utilize simulated live X-ray fluoroscopy at any time during the procedure to verify the entry point and angles for pedicle canulation and confirm inserted screw accuracy (Figure 3.1 D and E).

Simulated Surgical Scenario

There are four steps that participants need to perform in simulated L5-L4 bilateral pedicle screw insertion study. 1) pedicle cannulation with an Awl and pedicle finder, 2) breach verification using a ball tip, 3) pedicle pre-threading with the tap, and 4) pedicle screw insertion. Detailed step descriptions are available in Supplemental Digital Content Simulated L4 & L5 pedicle screw placement scenario. Collected data from all steps were used to assess and classify participants as skilled or less skilled.

Data Processing

Data Clean up

In this study, we are considering tool-level data to assess the skill level of the participants. The TSYM simulator collects tool-level data every 2 microseconds while a participant performs the procedure on the simulator. During the pedicle insertion procedure, five tools are used: 1) Awl, 2) Pedicle Finder, 3) Ball Tip, 4) Tap Screw, and 5) Screwdriver. The data for each of these tools is stored in a .csv file within each participant's folder as they perform the procedure. To manage the vast amount of data collected every 2 microseconds, the data is averaged and reduced to one record per participant for each tool. Once each participant's data for each tool is gathered, it is arranged vertically with the data from the other tools for that participant. Ultimately, we have 65 features for each participant, with 13 features for each of the 5 tools, excluding participant ID, screw number, and label. Participants placed four screws during their trial, and the data for each screw is treated as an independent data point. This results in a total of 172 data points.

The alignment of the number of data points with the number of independent samples helps address the challenge of overdetermination, where the risk of having more data points than independent samples could lead to overlearning. To further mitigate these risks, we implement feature selection, reducing the dataset to 65 features—fewer than the number of independent samples. Additionally, we employ cross-validation to ensure that our models generalize well to new data, minimizing the risk of overfitting and enhancing the robustness of our findings. The architecture of the study can be found in Figure 3.2.

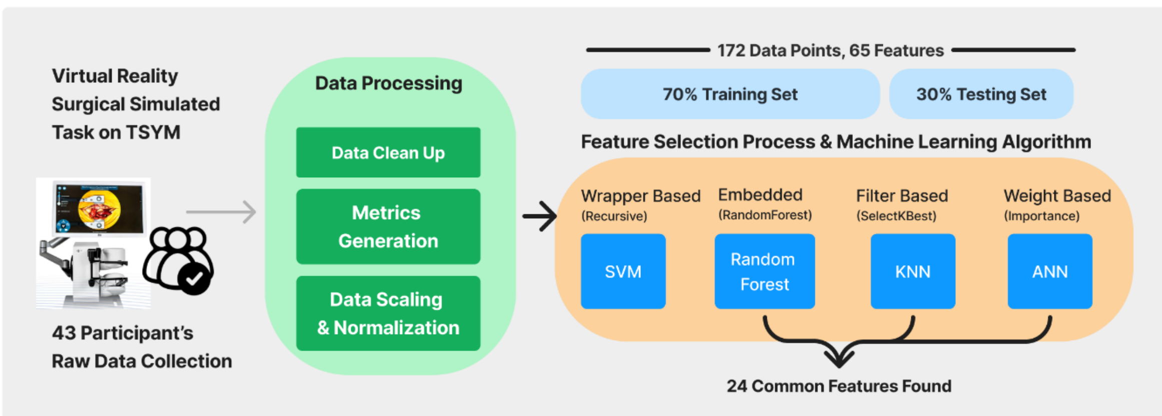

Figure 3.2 Methodology for the use of machine learning algorithms to determine the optimal features and find out the classification accuracy. Users begin by performing the surgical task on the virtual reality platform. Raw data acquisition occurs as the platform creates large datasets for each instrument employed. All instrument datasets are combined into a single dataset of 172 data points. This dataset is then divided into training set (70%) and test set (30%). Using training set four different algorithms 1) SVM, 2) Random Forest, 3) KNN 4) ANN are trained using training set with various feature selection techniques. Each models' performance was evaluated using test set. Twenty-four common important features were found amongst high performing algorithms Random Forest, KNN, and ANN.

Feature Generation and Refinement

The simulation initially captured 37 variables for each tool, encompassing metrics such as time, position (X, Y, Z coordinates), instrument angles, forces exerted on anatomical structures (haptic force in X, Y, Z components), and the volume of structure removed (excised voxels for specific tissue). For more details, please refer to the Supplemental Digital Content Tool Data. Features with all zero values were removed, and new metrics, including 3D Velocity, 3D Acceleration, and 3D Force over time, were calculated and averaged to create a single participant record. This process reduced the original 37 features per tool to 13, resulting in a total of 65 distinct metrics (13 features for each of the 5 tools arranged horizontally) used for the feature selection process. Developed through expert consultations and innovative feature design, these metrics ensured relevance and applicability to surgical proficiency assessment.

Data Preparation: The dataset included 172 data points, each representing data from four screws for each of the 43 participants. These data points were randomly divided into a training set (70%, 121 data points) and a test set (30%, 51 data points), with the test set preserved exclusively for evaluation purposes. Each screw's data was treated as an independent data point, and the division was performed on a data point basis rather than participant-wise to ensure a more randomized distribution. Data normalization was performed using the ‘StandardScaler’ from the ‘sklearn’ library. Each machine learning algorithm was trained separately on the training dataset and evaluated on the test dataset. For each algorithm, 5-fold cross-validation (cv=5) was employed on the training set, ensuring robust performance evaluation. The test set was not used in any way during the training process, including during cross-validation, to ensure an unbiased evaluation of the model's performance.

Feature Selection and Machine Learning Algorithms

This study uses four different feature selection methods: 1) Wrapper Based 2) Embedded based ,3) Filter Based and 4) Weight Based.

1) **Wrapper-Based Recursive Feature Elimination (RFE)**: RFE iteratively trains the model and eliminates the least important features until the optimal feature set is identified. This wrapper-based technique filters features based on their performance in predicting the output (Jeon & Oh, 2020). It provides a robust and tailored feature selection, directly optimizing model performance by evaluating feature subsets during training.
2) **Embedded-Based Feature Selection Using Random Forest**: Random Forest is an ensemble learning technique that ranks features based on their importance, as determined by their contribution to the decision trees within the forest (Louppe, 2014). It is robust to overfitting and noise, captures non-linear relationships, and provides stable and interpretable feature importance rankings. Because Random Forest performs feature selection as an inherent part of its training process, it is classified as an embedded method. This embedded feature selection ensures that the model identifies the most significant features in the dataset, enhancing the model's predictive performance.
3) **Filter-Based (SelectKBest) Feature Selection:** SelectKBest evaluates each feature based on statistical test scores and selects the top K features with the highest scores. It is simple, computationally efficient, and suitable for high-dimensional datasets, effectively reducing dimensionality and enhancing model performance by selecting statistically relevant features (Fitriani et al., 2022).
4) **Weight-Based Feature Selection in a Multi-Layer Perceptron (MLP)**: This method identifies and prioritizes input features based on the weights assigned during the training process of an MLP, considering the cumulative weights from input to output layers. It provides

a clear understanding of feature importance within the neural network, reduces noise, and enhances model performance by focusing on features with significant cumulative weights (Sun et al., 2013).

These four feature selection methods were used for the advantages they bring. These feature selection methods are used with four different machine learning algorithms 1) SVM (Wrapper based feature selection) 2) Random Forest (it has its own embedded feature selection)3) K-Nearest Neighbours (Filter based- SelectKBest) and 4) Artificial Neural Network (Multi-Layer Perceptron) (Weight Based).

1. **Support Vector Machine with Wrapper based Recursive Feature Elimination feature selection:** In this implementation, we utilized Recursive Feature Elimination (RFE) with an SVM classifier to identify the most important features and find out the classification accuracy. The SVM classifier, C (0.01), Kernel (linear), Gamma (scale), Degree (2), was employed as the estimator within the RFE model. By analyzing the ranking provided by RFE, we identified the top 31 features that had the greatest impact on the classification task. In the ranking, we selected features that had more than a 50% impact on classifying skilled and less skilled performance.
2. **Random Forest with embedded Feature selection:** In this implementation, we utilized a Random Forest classifier for feature selection and classification. A Random Forest classifier with 100 estimators and a random state of 0 was created. For feature selection, we employed ‘SelectFromModel’ with the Random Forest classifier. This method ranks features based on their importance, as determined by the Random Forest model. The threshold parameter was set to "median", meaning that features with importance above the median value were selected. This approach allowed us to leverage the Random Forest’s

embedded feature importance mechanism to identify and retain the most relevant features. By doing so, we improved the model's interpretability and ensured that it focused on features that significantly contributed to the classification task.

3. **K- Nearest Neighbours with Filter based SelectKBest:** In this implementation, we utilized a K-Nearest Neighbors (KNN) classifier for feature selection and classification, employing the SelectKBest method to identify the most important features. Initially, we specified a range of k values to determine the optimal number of neighbors for the KNN classifier. Using SelectKBest, we selected the top k features based on their scores from statistical tests. This method ranks features according to their importance, and only the top k features are selected. Through our analysis, we determined that k=35 yielded the best classification accuracy, and therefore, we selected the top 35 features based on their rank and importance.

4. **Artificial Neural Network (Multi-Layer Perceptron) with Weight Based Feature Selection:** In this implementation, we utilized an Artificial Neural Network (ANN) to perform classification and identify important features. We experimented with different hyperparameters to find the best-performing model. Specifically, we tested 70 learning rates (lr_values), 13 epochs (epochs_values), and 5 batch sizes (batch_sizes), resulting in a total of 4,550 models. For each combination of these hyperparameters, we built an ANN model with two hidden layers (64 and 32 neurons, respectively) and a sigmoid output layer for binary classification. The Adam optimizer was used to compile the model, with the loss function set to binary cross-entropy.

   The model was trained on the training set for each combination of hyperparameters, and its performance was evaluated on the test set. Predictions were converted to binary values by setting a threshold of 0.5. The accuracy of the model on the test set was calculated, and the combination of hyperparameters (learning rate of 0.098 and 350 epochs) yielding the highest accuracy was identified as the best model.

To extract important features, we examined the weights of the first hidden layer in the best-performing model. The absolute values of these weights were averaged across all neurons to determine the feature importance scores. These scores indicate the contribution of each input feature to the network's decision-making process. By analyzing these scores, we identified 26 features that contributed more than 70% to classify skilled performance from less skilled performance.

This method allowed us to systematically search for the optimal hyperparameters, build an accurate ANN model, and derive feature importance scores that highlight the influential features in the dataset.

Results:

Demographic data and relevant information concerning the two groups in this case series study are presented in Table 1. A total of 43 participants from four universities were included in this investigation. The skilled group (n=24) reported a mean of 1470 pedicle screws (SD = 675) inserted independently prior to the experiment while the less skilled group (n=19) reported a mean of 0 pedicle screws (SD = 0) inserted independently. The difference between the two groups was statistically significant, ($P < .001$). Since each participant inserted 4 screws, a total of 172 simulated screws were inserted.

Table 1: Demographics of 2 Groups of Participants Performing the virtual Reality Surgical Task		
	Skilled	**Less Skilled**
Age(Years)		
Mean, SD	36.8±7.3	29±2.7
Sex		
Male	23	15
Female	1	4
Total Number of participants: 43		
	24	19
Level of Training		
Neurosurgery Resident	-	-
PGY1	-	7
PGY2	-	2
PGY3	-	3
PGY4	1	2
PGY5	2	1
PGY6	3	1
Orthopedic Resident	-	-
PGY1	-	1
PGY2	-	1
PGY3	1	1
PGY4	2	-
PGY5	-	-
Spine Fellow		
Neurosurgical	5	-
Orthopedic	3	-
Spine Surgeon		
Neurosurgical	3	-
Orthopedic	4	-
Number of Reported Screws Inserted		
Mean, SD	1470±675	0±0
Median (Range)	40(2-3000)	(0-0)

Table 1 Demographic Data of Participants

Machine Learning Algorithm & Feature Selection:

Using Recursive Feature Elimination (RFE) with an SVM model, we identified an optimized list of 31 impactful metrics and achieved an accuracy of 78% (40 out of 51) on the test set. For the Random Forest Classifier (RFC), we also identified 31 impactful features, demonstrating an accuracy of 80.3% (41 out of 51) on the test set. Using SelectKBest with K-Nearest Neighbors

(KNN), we iterated through values of K from 1 to 65 and identified the optimal configuration with a peak accuracy of 82.3% (42 out of 51) on the test set at K=35. This method yielded 35 most impactful features for classifying skill levels. In weight-based metric selection with an Artificial Neural Network (ANN), we achieved a peak accuracy of 82.3% (42 out of 51) on the test set and list of 26 features that contributed more than 70% in classification. Refer to Figure 3.3 for the confusion matrix of each algorithm and Table 2 for a detailed comparison between these algorithms.

Algorithm Name	Sensitivity (Recall)	Specificity	TestSet Accuracy	F1 Score	AU ROC
SVM	0.62	0.9	78	0.45	0.79
Random Forest	0.64	0.93	80.3	0.48	0.82
KNN	0.64	1	82.3	0.5	0.87
ANN with All Features	0.65	0.96	82.3	0.51	0.85

Table 2 Sensitivity, Specificity, Test Set Accuracy, F1 Score and AU ROC comparisons among the Algorithms Assessed in this Study

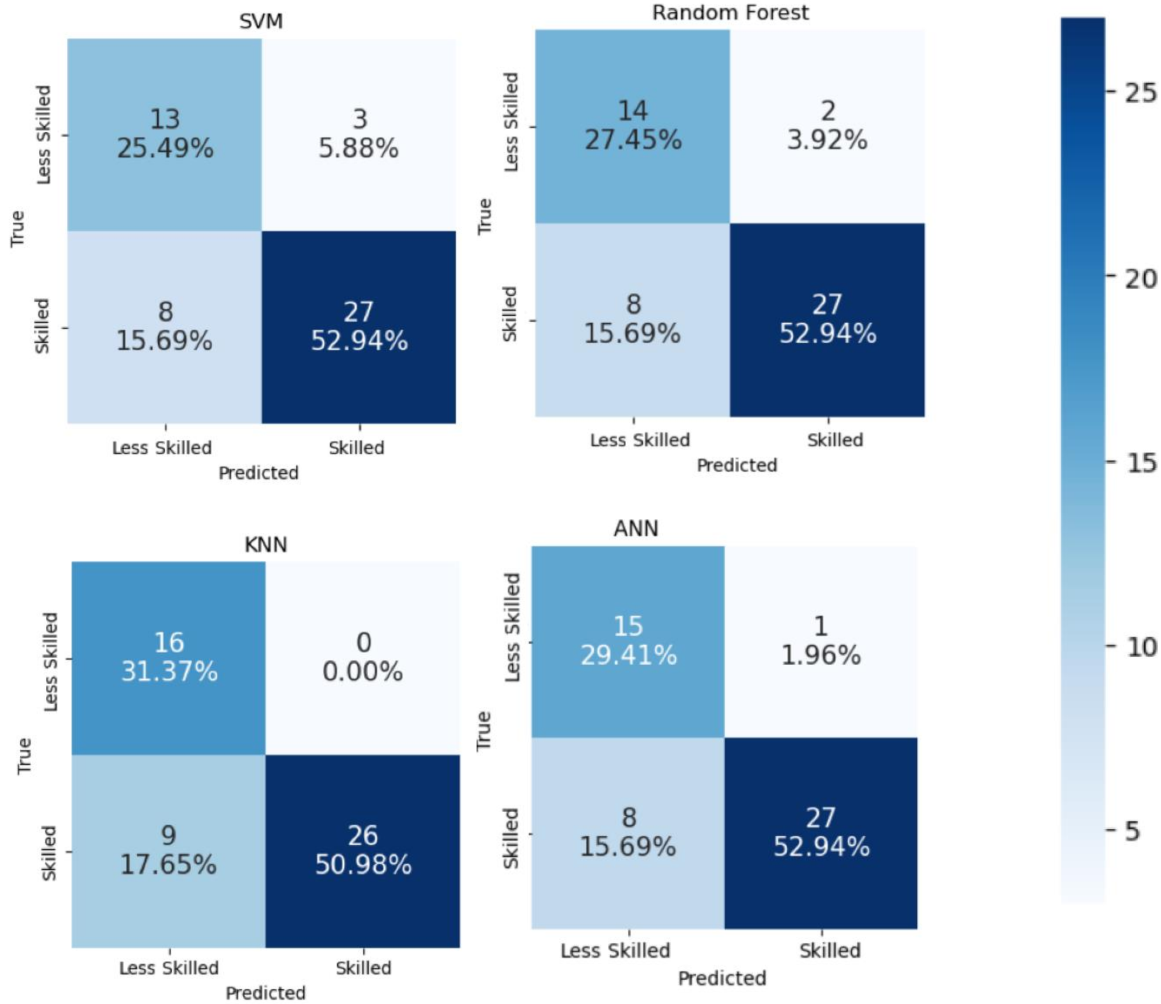

Figure 3.3 Confusion matrices of machine learning algorithms including SVM, Random Forest, KNN, ANN

Machine learning algorithms—Random Forest, K-Nearest Neighbors (KNN), and Artificial Neural Network (ANN) (Multi-Layer Perceptron)—successfully classified skilled versus less-skilled performance with an accuracy of over 80%. Because of their high accuracy, we used these algorithms to identify the key features contributing to accurately classifying skilled versus less-skilled performance. By analyzing the results from all three algorithms, we identified 24 common features that were consistently deemed important, as detailed in Table 3. The relative importance of these features, along with a comparison across different machine learning algorithms, is presented in Figure 3.4. This comprehensive analysis underscores the significance of these 24 features in evaluating skill levels. These 24 common features can be categorised in three categories which are safety, motion and efficiency as shown in Table 3.

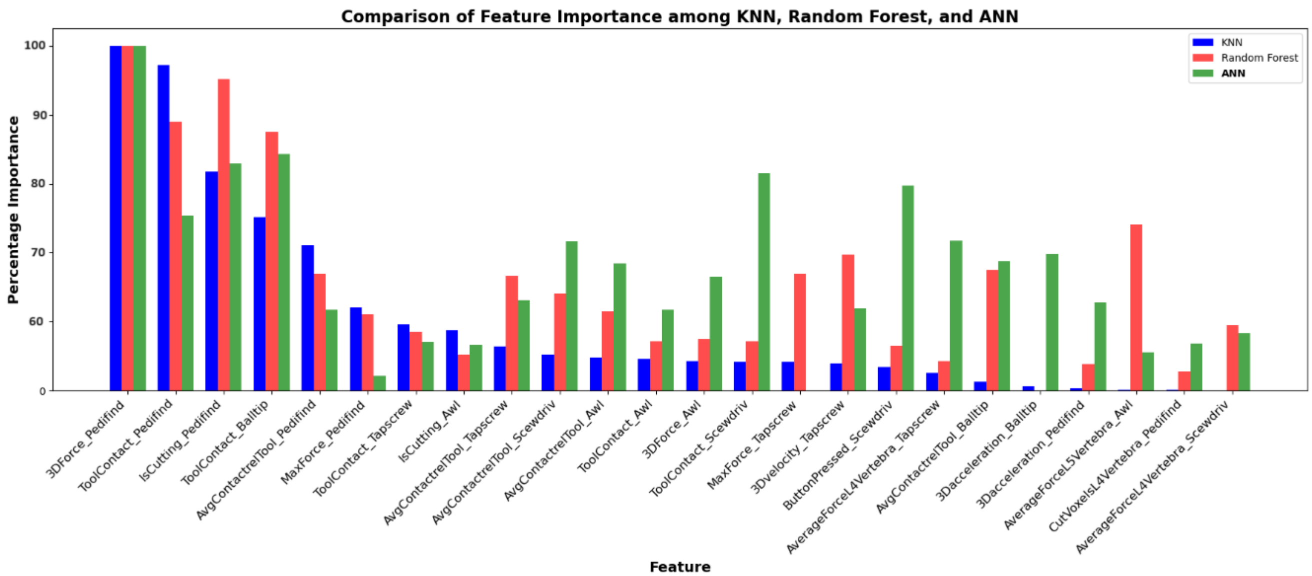

Figure 3.4: Relative importance of each feature in comparison to other algorithms

Category	Feature Name	Description
Safety: This category assesses the potential risks and impact on anatomical structures during the simulated surgical procedure. It includes measurements related to the volume removed, forces applied, and indications of cutting or removing tissue.	**CutVoxelsL4Vertebra_Pedifind**	Measures the extent of tissues removal (volume) from the L4 vertebra using Pedicle Finder Tool, indicating potential impact on surrounding structures.
	AverageForceL4Vertebra_Tapscrew	Represents the average force applied to the L4 vertebra during the use of the Tap Screw Tool, indicating the stress exerted on the bone.
	AverageForceL4Vertebra_Scewdriv	Reflects the average force applied to the L4 vertebra using the Screwdriver Tool, offering insights into the forces involved in the procedure.
	AverageForceL5Vertebra_Awl	Indicates the average force applied to the L5 vertebra with the Awl Tool, providing information on the applied pressure during the procedure.
	3DForce_Pedifind	Represents the force applied using the Pedicle Finder Tool, indicating the stress points in the procedure.
	MaxForce_Pedifind	Represents the maximum force applied using the Pedicle Finder Tool, indicating the highest stress points in the procedure.
	MaxForce_Tapscrew	Indicates the maximum force applied using the Tap Screw Tool, providing information on the peak force applied during the procedure.

	IsCutting_Awl	Binary indication of whether the Awl Tool is actively cutting or removing bone or soft tissue.
	IsCutting_Pedifind	Binary indication of whether the Pedicle Finder Tool is actively cutting or removing bone or soft tissue.
	3DForce_Awl	Represents the force applied using the Awl Tool, indicating the force applied during the procedure.
Motion: Focused on the movement dynamics of the simulated surgical tools, this category provides insights into the speed and acceleration of instruments during the procedure.	**3Dvelocity_Tapscrew**	Offering insights into the velocity of the Tap Screw Tool movement.
	3Dacceleration_Balltip	Providing information on the Ball Tip Tool acceleration during the procedure.
	3Dacceleration_Pedifind	Offering insights into the Pedicle Finder Tool acceleration during the procedure.
Efficiency: Examines the efficiency of the simulated surgical procedure by considering factors such as precision, frequency of contact, instances of structure removal, duration of tool contact, and the overall path traversed by instruments.	**AvgContactrelTool_Pedifind**	Represents the average contact points between the Pedicle Finder Tool and the vertebra, offering details on the precision and frequency of contact.
	AvgContactrelTool_Awl	Represents the average contact points between the Awl Tool and the vertebra, providing information on the precision and frequency of contact.
	AvgContactrelTool_Balltip	Represents the average contact points between the Ball Tip Tool and the vertebra, offering insights into the precision and frequency of contact.
	AvgContactrelTool_Scewdriv	Represents the average contact points between the Screwdriver Tool and the vertebra, providing details on the precision and frequency of contact.
	AvgContactrelTool_Tapscrew	Represents the average contact points between the Tap Screw Tool and the vertebra, offering insights into the precision and frequency of contact.
	ToolContact_Awl	Represents the contact points between the Awl Tool and the vertebra, providing information on the precision and frequency of contact.
	ToolContact_Balltip	Represents the contact points between the Ball Tip Tool and the vertebra, providing information on the precision and frequency of contact.
	ToolContact_Pedifind	Represents the contact points between the Pedicle Finder Tool and the vertebra, providing information on the precision and frequency of contact.
	ToolContact_Scewdriv	Represents the contact points between the Screwdriver Tool and the vertebra, providing information on the precision and frequency of contact.
	ButtonPressed_Scewdriv	Represents the efficiency of placing the screw using Screwdriver Tool.
	ToolContact_Tapscrew	Represents the contact points between the Tap Screw Tool and the vertebra, providing information on the precision and frequency of contact.

Table 3: Final selected metrics utilized data from the initial determination of KNN, Random Forest, and ANN with Safety, Movement and Efficiency categories.

Discussion

In this case series study, we employed a feature selection approach to determine the most effective performance metrics that distinguish skilled and less skilled performance in a pedicle screw insertion simulation utilizing the TSYM platform. The use of single feature selection techniques with the corresponding machine learning algorithms to determine important metrics may fail to identify key metrics that underly surgical expertise (Pudjihartono et al., 2022). Using combination of various feature selection method might help to address this issue. We found the 24 common features among three algorithms, which shows the importance of those features to classify performance. This resulted in outlining more holistic features which are robust, generalizable, and precise which might help ensuring comprehensive data coverage and minimize biases inherent in single-method analyses. The next step of this study could be to gather more data from more participants and feed only selected 24 features to further analyse the potential of these 24 features to classify skill level of participants for screw's performance

In previous investigations by our group assessing resident and fellow psychomotor skill levels by an intelligent tutor called the Intelligent Continuous expertise Monitoring System (ICEMS) trainee average performance score correlated with post graduate year of training, however, there was a wide variation in individual learner performance (Yilmaz et al.,2023). Categorizing participant data solely based on their training level could lead to incorrect data labeling resulting in inaccurate classification in machine learning models (Luciano et al., 2013). In this investigation actual experience with inserting pedicle screws during human operative procedures was utilized to a priori categorize expertise rather than trainee post graduate year. This categorization was based on the principle that if surgical educators allowed trainees to insert a pedicle screw into a patient's spine essential components of the anatomy, procedural steps, and possible complications were appreciated by the learner. Physically inserting the pedicle screw either under supervision or independently would provide further tactile and procedural cognitive information and enhanced support for that trainee being categorized as having more skills to perform the simulation. Trainees with no such clinical exposure were considered less skilled.

Artificial Intelligence in Surgical Education

The ability to utilize the large data sets generated from spine simulation models, will allow AI algorithms to further increase the precision and granularity of classification of surgical expertise utilizing VR spine platforms (Reich et al., 2022). However, it also risks issue of over determination, which can limit the generalization of the outcome, which needs to be tackled by cross verifications and data pruning which can remove noise of the data to an extent. Despite the risk, incorporating virtual reality simulation for performing complex procedures into the spine surgery learning curriculum may particularly benefit less skilled learners and be useful as potential formative and summative educational tools. Artificial intelligence methodologies with their ability to generate large amounts of data may identify new features and rank their importance allowing surgical instructors to focus on teaching these important features related to surgical skills. Employing deep learning algorithms and novice and expert data intelligent tutoring systems can be developed and tested (Yilmaz et al., 2022; Reich et al., 2022). Human educator input care is required in developing these programs to prevent unintended outcomes (Fazlollahi et al., 2023).

Limitations

The TSYM simulation platform has limitations. The pedicle screw insertion simulation does not capture the dynamic human intraoperative environment where continuous surgical educator personalized feedback to the learner is crucial. The simulated procedure is designed with one animated and four deconstructed steps in a linear, unidirectional sequence, which does not reflect the flexible and adaptive approach required during human pedicle screw insertion. Unrealistic details of the simulation might affect the performance of an expert, especially in the initial phase of adjusting to the simulator’s feel and haptic. The TSYM platform offers a range of screw sizes and lengths, to standardize the procedure, a fixed-size screw was utilized which does not allow exploration of the learners' ability to appreciate the importance of screw dimensions in spinal surgery. Each pedicle screw insertion was evaluated individually but other studies are needed to assess the impact of repeated pedicle screw insertions on participant learning curves. The simulator platform consists of a single-handed robotic arm, not reproducing the bimanual technical skills employed in patient spinal procedures (Mirchi et al., 2020; Yilmaz et al., 2022; Alotaibi et al., 2015; Reich et al., 2022; Anderson et al., 2016; Wongworawat, 2016). Enhancing the TSYM platform to allow for bimanual usage would provide a more comprehensive assessment of surgical expertise. Measurement errors inherent to the simulator architecture may be present and employing other pedicle screw insertion simulators to validate these results is important. Studies to benchmark the performance of various algorithm and derive common important features among those algorithms involving different pedicle screw insertion simulators will be important. Further analysis is needed to conclude the importance of the individual features that this study found important. This underscores the need for independent testing before drawing definitive conclusions about the accuracy of our findings (Handelman et al., 2018; Wongworawat, 2016).

Conclusion

In this case series study, we benchmarked the performance of various machine learning algorithms and identified key features that may represent expertise during simulated pedicle screw insertion. The combination of feature selection techniques outlined here demonstrated the potential for improved accuracy in future studies, enhancing the classification of skilled and less-skilled performance. These findings have significant implications for the understanding, assessment, and training of all surgical procedures.

References

1. Alaraj, A., Charbel, F. T., Birk, D., et al. (2013). Role of cranial and spinal virtual and augmented reality simulation using immersive touch modules in neurosurgical training. *Neurosurgery, 72*(suppl 1).
2. Alaraj, A., Luciano, C. J., Bailey, D. P., et al. (2015). Virtual reality cerebral aneurysm clipping simulation with real-time haptic feedback. *Operative Neurosurgery, 11*(1), 52-58.
3. Alkadri, S., Ledwos, N., Mirchi, N., et al. (2021). Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure. *Computers in Biology and Medicine, 136*, 104770. https://doi.org/10.1016/j.compbiomed.2021.104770
4. Alotaibi, F., Al Zhrani, G., Bajunaid, K., Winkler-Schwartz, A., Azarnoush, H., & Mullah, M. (2015). Assessing neurosurgical psychomotor performance: Role of virtual reality simulators, current and future potential. *SOJ Neurology, 2*(1), 1-7.
5. Andersen, D. S., Konge, L., & Sørensen, M. S. (2023). Advances in the objective assessment of technical skills in surgery. *Annals of Surgery, 277*(4), 654-661. https://doi.org/10.1097/SLA.0000000000005817
6. Anderson, D. D., Long, S., Thomas, G. W., Putnam, M. D., Bechtold, J. E., & Karam, M. D. (2016). Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively. *Clinical Orthopaedics and Related Research, 474*(4), 874-881. https://doi.org/10.1007/s11999-015-4603-4
7. Ansorge, A., Sarwahi, V., Bazin, L., Vazquez, O., De Marco, G., & Dayer, R. (2023). Accuracy and safety of pedicle screw placement for treating adolescent idiopathic scoliosis: A narrative review comparing available techniques. *Diagnostics, 13*(14), 2402. https://doi.org/10.3390/diagnostics13142402

8. Bakhaidar M., Alsayegh A., Yilmaz R, et al. (2023) Performance in a Simulated Virtual Reality Anterior Cervical Discectomy and Fusion Task: Disc Residual, Rate of Removal, and Efficiency Analyses. Oper Neurosurg (Hagerstown). 2023 Oct 1;25(4):e196-e205. doi: 10.1227/ons.0000000000000813. Epub 2023 Jul 13.
9. Cheng, A., Kessler, D., Mackinnon, R., et al. (2016). Reporting guidelines for health care simulation research: Extensions to the CONSORT and STROBE statements. *Simulation in Healthcare, 11*(4), 238-248. https://doi.org/10.1097/SIH.0000000000000150
10. Deo, R. C. (2015). Machine learning in medicine. *Circulation, 132*(20), 1920-1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593
11. Dickman, C. A., Fessler, R. G., MacMillan, M., et al. (1992). Transpedicular screw-rod fixation of the lumbar spine: Operative technique and outcome in 104 cases. *Journal of Neurosurgery, 77*, 860-870.
12. Elmi-Terander, A., & Skulason, H. (2023). Pedicle screw placement accuracy in robot-assisted versus image-guided freehand surgery of thoracolumbar spine: A randomized controlled trial. *Trials, 24*, 387. https://doi.org/10.1186/s13063-023-07309-7
13. Fazlollahi, A. M., Yilmaz, R., Winkler-Schwartz, A., et al. (2023). AI in surgical curriculum design and unintended outcomes for technical competencies in simulation training. *JAMA Network Open, 6*(9). https://doi.org/10.1001/jamanetworkopen.2023.34658
14. Fitriani, S. A., Astuti, Y., & Wulandari, I. R. (2022). Least absolute shrinkage and selection operator (LASSO) and k-nearest neighbors (k-NN) algorithm analysis based on feature selection for diamond price prediction. In *Proceedings of the 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)* (pp. 135-139). Jakarta, Indonesia. https://doi.org/10.1109/ISMODE53584.2022.9742936

15. Gang, C., Haibo, L., Fancai, L., Weishan, C., & Qixin, C. (2012). Learning curve of thoracic pedicle screw placement using the free-hand technique in scoliosis: How many screws needed for an apprentice? *European Spine Journal, 21*(6), 1151-1156. https://doi.org/10.1007/s00586-011-2065-2
16. Gonzalvo, A., Fitt, G., Liew, S., et al. (2009). The learning curve of pedicle screw placement: How many screws are enough? *Spine (Phila Pa 1976), 34*(21). https://doi.org/10.1097/BRS.0b013e3181b2f928
17. Grillo, H. C. (2018). To impart this art: The development of graduate surgical education in the United States. *Surgery, 125*(1), 1-14.
18. Halvorsen, F. H., Elle, O. J., et al. (2022). The role of artificial intelligence in the objective assessment of surgical skills. *International Journal of Medical Robotics and Computer Assisted Surgery, 18*(1). https://doi.org/10.1002/rcs.2354
19. Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Lee, M. J., & Asadi, H. (2018). eDoctor: Machine learning and the future of medicine. *Journal of Internal Medicine, 284*(6), 603-619. https://doi.org/10.1111/joim.12822
20. Hashimoto, D. A., Rosman, G., Rus, D., & Meireles, O. R. (2018). Artificial intelligence in surgery: Promises and perils. *Annals of Surgery, 268*(1), 70-76. https://doi.org/10.1097/SLA.0000000000002693
21. Jeon, H., & Oh, S. (2020). Hybrid-recursive feature elimination for efficient feature selection. *Applied Sciences, 10*(9), 3211. https://doi.org/10.3390/app10093211
22. Jia, S., Weng, Y., Wang, K., et al. (2023). Performance evaluation of an AI-based preoperative planning software application for automatic selection of pedicle screws based on computed

tomography images. *Frontiers in Surgery, 10*, 1247527. https://doi.org/10.3389/fsurg.2023.1247527

23. Kim, Y. J., Lenke, L. G., Bridwell, K. H., Cho, Y. S., & Riew, K. D. (2004). Free hand pedicle screw placement in the thoracic spine: Is it safe? *Spine (Phila Pa 1976), 29*(3), 333-342. https://doi.org/10.1097/01.BRS.0000109983.12113.9B

24. Ledwos, N., Mirchi, N., Bissonnette, V., Winkler-Schwartz, A., Yilmaz, R., & Del Maestro, R. F. (2021). Virtual reality anterior cervical discectomy and fusion simulation on the novel sim-ortho platform: Validation studies. *Operative Neurosurgery, 20*(1), 74-82. https://doi.org/10.1093/ons/opaa269

25. Louppe, G. (2014). Understanding random forests: From theory to practice. arXiv. Published online July 28, 2014. https://arxiv.org/abs/1407.7502

26. Luciano, C., Banerjee, P., Florea, L., & Dawe, G. (2005). Design of the immersivetouch: A high-performance haptic augmented virtual reality system. Presented at the *11th International Conference on Human-Computer Interaction*, Las Vegas, NV.

27. Luciano, C. J., Banerjee, P. P., Bellotte, B., et al. (2011). Learning retention of thoracic pedicle screw placement using a high-resolution augmented reality simulator with haptic feedback. *Operative Neurosurgery, 69*(suppl 1).

28. Luciano, C. J., Banerjee, P. P., Sorenson, J. M., et al. (2013). Percutaneous spinal fixation simulation with virtual reality and haptics. *Neurosurgery, 72*(suppl 1), 89-96.

29. Ma, C., Zou, D., Qi, H., et al. (2022). A novel surgical planning system using an AI model to optimize planning of pedicle screw trajectories with highest bone mineral density and strongest pull-out force. *Neurosurgical Focus, 52*(4). https://doi.org/10.3171/2022.1.FOCUS21721

30. Manbachi, A., Cobbold, R. S., & Ginsberg, H. J. (2014). Guided pedicle screw insertion: Techniques and training. *The Spine Journal, 14*(1), 165-179. https://doi.org/10.1016/j.spinee.2013.03.029

31. McGaghie, W. C. (2015). Mastery learning: It is time for medical education to join the 21st century. *Academic Medicine, 90*(11), 1438-1441. https://doi.org/10.1097/ACM.0000000000000911

32. Mirchi, N., Bissonnette, V., Yilmaz, R., Ledwos, N., et al. (2020). The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PLoS One, 15*(2). https://doi.org/10.1371/journal.pone.0229596

33. Polavarapu, H. V., Kulaylat, A. N., Sun, S., & Hamed, O. H. (2013). 100 years of surgical education: The past, present, and future. *Bulletin of the American College of Surgeons, 98*(7), 22-27.

34. Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics, 2*, 927312. https://doi.org/10.3389/fbinf.2022.927312

35. Reich, A., Mirchi, N., Yilmaz, R., et al. (2022). Artificial neural network approach to competency-based training using a virtual reality neurosurgical simulation. *Operative Neurosurgery, 23*(1), 31-39. https://doi.org/10.1227/ons.0000000000000173

36. Shahzad, S., & Anwar, M. I. (2021). Apprenticeship model in 21st century's surgical education: Should it perish? *Archives of Surgical Research, 2*(3), 1-3. https://doi.org/10.48111/2021.03.01

37. Sun, X., Liu, Y., Xu, M., et al. (2013). Feature selection using dynamic weights for classification. *Knowledge-Based Systems, 37*, 541-549. https://doi.org/10.1016/j.knosys.2012.10.001

38. Symgery. (2023). Document compilation. Retrieved from https://symgery.com/wp-content/uploads/2023/10/T-SYM_Doc_Comp.pdf

39. Sundaresan, N., Galicich, J. H., & Lane, J. M. (1984). Harrington rod stabilization for pathological fractures of the spine. *Journal of Neurosurgery, 60*, 282-286.

40. Varghese, C., Harrison, E. M., O'Grady, G., et al. (2024). Artificial intelligence in surgery. *Nature Medicine, 30*(10), 1257-1268. https://doi.org/10.1038/s41591-024-02970-3

41. Winkler-Schwartz, A., Yilmaz, R., Mirchi, N., et al. (2019). Machine learning identification of surgical and operative factors associated with surgical expertise in virtual reality simulation. *JAMA Network Open, 2*(8). https://doi.org/10.1001/jamanetworkopen.2019.8363

42. Wongworawat, M. D. (2016). Editor's spotlight/take 5: Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively. *Clinical Orthopaedics and Related Research, 474*(4), 871-873. https://doi.org/10.1007/s11999-015-4677-z

43. World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*(20), 2191-2194. https://doi.org/10.1001/jama.2013.281053

44. Yilmaz, R., Winkler-Schwartz, A., Mirchi, N., et al. (2022). Continuous monitoring of surgical bimanual expertise using deep neural networks in virtual reality simulation. *NPJ Digital Medicine, 5*(1), 54. https://doi.org/10.1038/s41746-022-00596-8

Chapter 4

Thesis Summary

This thesis investigates the application of machine learning utilizing VR surgical simulators, with a specific focus on pedicle screw insertion—a complex and critical technique in spine surgery. The study utilizes a high-fidelity VR simulator to evaluate the efficacy of machine learning algorithms in accurately classifying surgical expertise among participants. This research addresses the need for objective and reliable assessment tools in surgical education, particularly under the constraints faced by physician-educators balancing clinical responsibilities and educational demands.

Utilizing four different feature selection techniques employed in the study include Recursive Feature Elimination, Embedded with Random Forest Classifier (RFC), Filter based, and Weight-Based methods. These methodologies are utilized to identify and prioritize features that effectively differentiate between less skilled and skilled participants. The integration of these diverse feature selection methods enhances the analysis by capturing complex, non-linear relationships, minimizing overfitting, and thereby improving the predictive accuracy of the models.

Another key outcome of this thesis is the benchmarked performance of various machine learning algorithms, including SVM, Random Forest, KNN, and ANN, three of which demonstrated the ability to classify surgical expertise with over 80% accuracy. Additionally, we identified 24 common features among three of these algorithms, highlighting the critical factors in assessing surgical performance involving L4-L5 pedicle screw insertion on the TSYM simulation platform. These results outline the potential of AI-enhanced VR simulators to advance training and assessment in surgical education, particularly for intricate procedures such as pedicle screw insertion.

By focusing on selected common features, the AI model aligns its findings with practical surgical knowledge and the nuanced skills that distinguish skilled individuals from less skilled learners. For instance, metrics related to the handling of the Awl and Pedicle Finder tools reflect the techniques used to identify correct entry points and apply the appropriate forces necessary for successful pedicle screw placement—skills important in minimizing tissue damage and enhancing surgical outcomes.

In conclusion, this case series study by employing machine learning algorithms outlines important features which determine novel composites of expertise during simulated pedicle screw insertion. This investigation introduces a combined feature selection approach which might help in achieving improved accuracy in classifying skilled and less skilled performance which may have utility in the assessment and training of all surgical procedures.

Future Direction

Building on the results of this study, which established benchmarked data for distinguishing between skilled and less skilled trainee performance levels in pedicle screw insertion, several promising avenues for future research have been identified.

Further Improve the Classification Accuracy by Employing Selected 24 Features: The next step of this study is to recruit more participants to gather additional data points, enabling further training and testing of the algorithms using the selected 24 features. This will provide a more robust understanding of these features as key indicators of surgical skill. The insights gained can then be used to enhance training programs by focusing on the skills identified as most important.

Development of the ICEMS System for Spine Surgery: On acquiring a more robust understanding of the features and their impact on classification, the next logical step is to extend the capabilities of the Intelligent Continuous Expertise Monitoring System (ICEMS) specifically for spine surgery tasks such as pedicle screw insertion. Leveraging the benchmarked data generated in this study, the proposed ICEMS for spine pedicle screw insertion surgery will aim to provide real-time, personalized feedback and assessment tailored to the requirements of individual components of these spinal procedures. This advancement will facilitate more targeted and effective training, potentially reducing the learning curve for trainees and enhancing overall surgical proficiency.

Investigation of Learning Curves: A component of this study was the requirement for participants to place four screws during the simulation. Future research could analyze the data collected to determine whether a learning curve is evident across these repetitive tasks. By assessing performance improvements after each screw placement, researchers can gain insights into the efficacy of simulation-based training and the potential for accelerated skill acquisition through repetitive practice.

Comprehensive Performance Assessment System: To create a more robust system for evaluating surgical performance, it is proposed to integrate tool data with OSATS and visual subjective classification of pedicle screw placement. By combining these diverse data sources, a comprehensive assessment framework can be developed that not only evaluates technical proficiency but also considers the qualitative aspects of surgical tasks. Such a multidimensional approach may lead to increased understanding of surgical competence and could lead to the development of a more robust AI-driven assessment model.

These future directions underscore the potential to significantly advance surgical training and assessment methodologies through the integration of AI and VR technologies. By deepening the analysis of learning processes and enhancing the accuracy of performance assessments, these initiatives will contribute to the refinement of surgical education and, ultimately, to improvements in patient outcomes in spine surgery.

References:

1. Abumi, K., Panjabi, M. M., & Duranceau, J. (1989). Biomechanical evaluation of spinal fixation devices. Part III. Stability provided by six spinal fixation devices and interbody bone graft. *Spine (Phila Pa 1976), 14*, 1249-1255.

2. Adamski, S., Stogowski, P., Rocławski, M., Pankowski, R., & Kloc, W. (2023). Review of currently used classifications for pedicle screw position grading in cervical, thoracic and lumbar spine. *Chirurgia Narządów Ruchu i Ortopedia Polska, 88*(4), 165-171. https://doi.org/10.31139/chnriop.2023.88.4.2

3. Alaraj, A., Charbel, F. T., Birk, D., et al. (2013). Role of cranial and spinal virtual and augmented reality simulation using immersive touch modules in neurosurgical training. *Neurosurgery, 72*(suppl 1).

4. Alaraj, A., Luciano, C. J., Bailey, D. P., et al. (2015). Virtual reality cerebral aneurysm clipping simulation with real-time haptic feedback. *Operative Neurosurgery, 11*(1), 52-58.

5. Alkadri, S., Ledwos, N., Mirchi, N., et al. (2021). Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure. *Computers in Biology and Medicine, 136*, 104770. https://doi.org/10.1016/j.compbiomed.2021.104770

6. Alotaibi, F. E., et al. (2015). Neurosurgical assessment of metrics including judgment and dexterity using the virtual reality simulator NeuroTouch (NAJD Metrics). *Surgical Innovation, 22*(6), 636-642.

7. Alotaibi, F., Al Zhrani, G., Bajunaid, K., Winkler-Schwartz, A., Azarnoush, H., & Mullah, M. (2015). Assessing neurosurgical psychomotor performance: Role of virtual reality simulators, current and future potential. *SOJ Neurology, 2*(1), 1-7.

8. Amaral, T. D., Hasan, S., Galina, J., & Sarwahi, V. (2021). Screw malposition: Are there long-term repercussions to malposition of pedicle screws? *Journal of Pediatric Orthopaedics, 41*(Suppl 1). https://doi.org/10.1097/BPO.0000000000001828

9. Andersen, D. S., Konge, L., & Sørensen, M. S. (2023). Advances in the objective assessment of technical skills in surgery. *Annals of Surgery, 277*(4), 654-661. https://doi.org/10.1097/SLA.0000000000005817

10. Anderson, D. D., Long, S., Thomas, G. W., Putnam, M. D., Bechtold, J. E., & Karam, M. D. (2016). Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively. *Clinical Orthopaedics and Related Research, 474*(4), 874-881. https://doi.org/10.1007/s11999-015-4603-4

11. Andén, U., Lake, A., & Nordwall, A. (1980). The role of the anterior longitudinal ligament in Harrington rod fixation of unstable thoracolumbar spinal fractures. *Spine (Phila Pa 1976), 5*, 23-25.

12. Ansorge, A., Sarwahi, V., Bazin, L., Vazquez, O., De Marco, G., & Dayer, R. (2023). Accuracy and safety of pedicle screw placement for treating adolescent idiopathic scoliosis: A narrative review comparing available techniques. *Diagnostics, 13*(14), 2402. https://doi.org/10.3390/diagnostics13142402

13. Asad, A. L., Anteby, M., & Garip, F. (2014). Who donates their bodies to science? The combined role of gender and migration status among California whole-body donors. *Social Science & Medicine, 106*, 53-58. https://doi.org/10.1016/j.socscimed.2014.01.041

14. Aoude, A. A., Fortin, M., Figueiredo, R., Jarzem, P., Ouellet, J., & Weber, M. H. (2015). Methods to determine pedicle screw placement accuracy in spine surgery: A systematic review. *European Spine Journal, 24*(5), 990-1004. https://doi.org/10.1007/s00586-015-3853-x

15. Aoude, A., Ghadakzadeh, S., Alhamzah, H., et al. (2018). Postoperative assessment of pedicle screws and management of breaches: A survey among Canadian spine surgeons and a new scoring system. *Asian Spine Journal, 12*(1), 37-46. https://doi.org/10.4184/asj.2018.12.1.37

16. Azarnoush, H., et al. (2016). The force pyramid: A spatial analysis of force application during virtual reality brain tumor resection. *Journal of Neurosurgery, 127*(1), 171.

17. Badash, I., Burtt, K., Solorzano, C. A., & Carey, J. N. (2016). Innovations in surgery simulation: A review of past, current and future techniques. *Annals of Translational Medicine, 4*(23).

18. Bakhaidar M., Alsayegh A., Yilmaz R, et al. (2023) Performance in a Simulated Virtual Reality Anterior Cervical Discectomy and Fusion Task: Disc Residual, Rate of Removal, and Efficiency Analyses. Oper Neurosurg (Hagerstown). 2023 Oct 1;25(4):e196-e205. doi: 10.1227/ons.0000000000000813. Epub 2023 Jul 13.

19. Baird, E., McAnany, S., Overley, S., Skovrlj, B., Guzman, J., & Qureshi, S. (2017). Accuracy of percutaneous pedicle screw placement: Does training level matter? *Clinical Spine Surgery, 30*(6), 7748-7753.

20. Bissonnette, V., et al. (2019). Artificial intelligence distinguishes surgical training levels in a virtual reality spinal task. *Journal of Bone and Joint Surgery, 101*(23).

21. Bono, C. M., & Lee, C. K. (2004). Critical analysis of trends in fusion for degenerative disc disease over the past 20 years: Influence of technique on fusion rate and clinical outcome. *Spine (Phila Pa 1976), 29*, 455-463.

22. Chan, S., Conti, F., Salisbury, K., & Blevins, N. H. (2013). Virtual reality simulation in neurosurgery: Technologies and evolution. *Neurosurgery, 72*(suppl 1).

23. Cheng, A., Kessler, D., Mackinnon, R., et al. (2016). Reporting guidelines for health care simulation research: Extensions to the CONSORT and STROBE statements. *Simulation in Healthcare, 11*(4), 238-248. https://doi.org/10.1097/SIH.0000000000000150

24. Chen, T., Zhang, Y., Ding, C., et al. (2021). Virtual reality as a learning tool in spinal anatomy and surgical techniques. *North American Spine Society Journal, 6*, 100063. https://doi.org/10.1016/j.xnsj.2021.100063

25. Coelho, G., Zanon, N., & Warf, B. (2014). The role of simulation in neurosurgery. *Child's Nervous System, 30*(12), 1997-2000. https://doi.org/10.1007/s00381-014-2548-7

26. Cooke, M., Irby, D. M., Sullivan, W., & Ludmerer, K. M. (2006). American medical education 100 years after the Flexner report. *New England Journal of Medicine, 355*(13), 1339-1344.

27. Cotrel, Y., & Dubousset, J. (1984). A new technique for segmental spinal osteosynthesis using the posterior approach. *Revue de Chirurgie Orthopédique et Réparatrice de l'Appareil Moteur, 70*, 489-494.

28. de Kater, E. P., Sakes, A., Edström, E., Elmi-Terander, A., Kraan, G., & Breedveld, P. (2022). Beyond the pedicle screw–A patent review. *European Spine Journal, 31*(6), 1553-1565. https://doi.org/10.1007/s00586-022-07193-z

29. Delorme, S., Laroche, D., DiRaddo, R., & Del Maestro, R. F. (2012). NeuroTouch: A physics-based virtual simulator for cranial microneurosurgery training. *Neurosurgery, 71*, 32-42.

30. Deo, R. C. (2015). Machine learning in medicine. *Circulation, 132*(20), 1920-1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

31. Dick, W., Kluger, P., Magerl, F., Woersdörfer, O., & Zäch, G. (1985). A new device for internal fixation of thoracolumbar and lumbar spine fractures: The 'fixateur interne.' *Spinal Cord, 23*(4), 225-232. https://doi.org/10.1038/sc.1985.38

32. Dickman, C. A., Fessler, R. G., MacMillan, M., et al. (1992). Transpedicular screw-rod fixation of the lumbar spine: Operative technique and outcome in 104 cases. *Journal of Neurosurgery, 77*, 860-870.

33. Elmi-Terander, A., & Skulason, H. (2023). Pedicle screw placement accuracy in robot-assisted versus image-guided freehand surgery of thoracolumbar spine: A randomized controlled trial. *Trials, 24*, 387. https://doi.org/10.1186/s13063-023-07309-7

34. Esses, S. T., & Bednar, D. A. (1989). The spinal pedicle screw: Techniques and systems. *Orthopedic Review, 18*, 676-682.

35. Fazlollahi, A. M., Yilmaz, R., Winkler-Schwartz, A., et al. (2023). AI in surgical curriculum design and unintended outcomes for technical competencies in simulation training. *JAMA Network Open, 6*(9). https://doi.org/10.1001/jamanetworkopen.2023.34658

36. Fitriani, S. A., Astuti, Y., & Wulandari, I. R. (2022). Least absolute shrinkage and selection operator (LASSO) and k-nearest neighbors (k-NN) algorithm analysis based on feature selection for diamond price prediction. In *Proceedings of the 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)* (pp. 135-139). Jakarta, Indonesia. https://doi.org/10.1109/ISMODE53584.2022.9742936

37. Frank, J. R., Snell, L., & Sherbino, J. (2015). CanMEDS 2015 Physician Competency Framework. *Royal College of Physicians and Surgeons of Canada*.

38. Frank, J. R., Snell, L., Englander, R., Holmboe, E. S., & Collaborators, I. (2017). Implementing competency-based medical education: Moving forward. *Medical Teacher, 39*(6), 568-573.

39. Franzese, C. B., & Stringer, S. P. (2007). The evolution of surgical training: Perspectives on educational models from the past to the future. *Otolaryngologic Clinics of North America, 40*(6), 1227-1235, vii. https://doi.org/10.1016/j.otc.2007.07.004

40. Gang, C., Haibo, L., Fancai, L., Weishan, C., & Qixin, C. (2012). Learning curve of thoracic pedicle screw placement using the free-hand technique in scoliosis: How many screws needed for an apprentice? *European Spine Journal, 21*(6), 1151-1156. https://doi.org/10.1007/s00586-011-2065-2

41. Gasco, J., Patel, A., Ortega-Barnett, J., et al. (2014). Virtual reality spine surgery simulation: An empirical study of its usefulness. *Neurological Research, 36*, 968-973.

42. Gautschi, O. P., Schatlo, B., Schaller, K., & Tessitore, E. (2011). Clinically relevant complications related to pedicle screw placement in thoracolumbar surgery and their management: A literature review of 35,630 pedicle screws. *Neurosurgical Focus, 31*(4). https://doi.org/10.3171/2011.7.FOCUS11168

43. Gelalis, I., Paschos, N., Pakos, E., et al. (2012). Accuracy of pedicle screw placement: A systematic review of prospective in vivo studies comparing free hand, fluoroscopy guidance and navigation techniques. *European Spine Journal, 21*(2), 247-255.

44. Gonzalvo, A., Fitt, G., Liew, S., et al. (2009). The learning curve of pedicle screw placement: How many screws are enough? *Spine (Phila Pa 1976), 34*, 761-765.

45. Grillo, H. C. (2018). To impart this art: The development of graduate surgical education in the United States. *Surgery, 125*(1), 1-14.

46. Guerrero, D. T., Asaad, M., Rajesh, A., Hassan, A., & Butler, C. E. (2022). Advancing surgical education: The use of artificial intelligence in surgical training. *The American Surgeon*. Advance online publication. https://doi.org/10.1177/00031348221101503

47. Halvorsen, F. H., Elle, O. J., et al. (2022). The role of artificial intelligence in the objective assessment of surgical skills. *International Journal of Medical Robotics and Computer Assisted Surgery, 18*(1). https://doi.org/10.1002/rcs.2354

48. Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Lee, M. J., & Asadi, H. (2018). eDoctor: Machine learning and the future of medicine. *Journal of Internal Medicine, 284*(6), 603-619. https://doi.org/10.1111/joim.12822

49. Harris, K. A., Nousiainen, M. T., & Reznick, R. (2020). Competency-based resident education—The Canadian perspective. *Surgery, 167*(4), 681-684

50. Hashimoto, D. A., Rosman, G., Rus, D., & Meireles, O. R. (2018). Artificial intelligence in surgery: Promises and perils. *Annals of Surgery, 268*(1), 70-76. https://doi.org/10.1097/SLA.0000000000002693

51. Hatala, R., Cook, D. A., Brydges, R., & Hawkins, R. (2015). Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): A systematic review of validity evidence. *Advances in Health Sciences Education, 20*(5), 1149-1175.

52. Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications, 13*(4), 18-28. https://doi.org/10.1109/5254.708428

53. Hicks, J. M., Singla, A., Shen, F. H., & Arlet, V. (2010). Complications of pedicle screw fixation in scoliosis surgery. *Spine (Phila Pa 1976), 35*(11). https://doi.org/10.1097/BRS.0b013e3181d1021a

54. Hou, Y., Lin, Y., Shi, J., Chen, H., & Yuan, W. (2018). Effectiveness of the thoracic pedicle screw placement using the virtual surgical training system: A cadaver study. *Operative Neurosurgery (Hagerstown), 15*, 677-685.

55. Iobst, W. F., Sherbino, J., Cate, O. T., et al. (2010). Competency-based medical education in postgraduate medical education. *Medical Teacher, 32*(8), 651-656.

56. Jeon, H., & Oh, S. (2020). Hybrid-recursive feature elimination for efficient feature selection. *Applied Sciences, 10*(9), 3211. https://doi.org/10.3390/app10093211

57. Jia, S., Weng, Y., Wang, K., et al. (2023). Performance evaluation of an AI-based preoperative planning software application for automatic selection of pedicle screws based on computed tomography images. *Frontiers in Surgery, 10*, 1247527. https://doi.org/10.3389/fsurg.2023.1247527

58. Kim, Y. J., Lenke, L. G., Bridwell, K. H., Cho, Y. S., & Riew, K. D. (2004). Free hand pedicle screw placement in the thoracic spine: Is it safe? *Spine (Phila Pa 1976), 29*(3), 333-342. https://doi.org/10.1097/01.BRS.0000109983.12113.9B

59. Kobayashi, K., Ando, K., Nishida, Y., Ishiguro, N., & Imagama, S. (2018). Epidemiological trends in spine surgery over 10 years in a multicenter database. *European Spine Journal, 27*(8), 1698-1703. https://doi.org/10.1007/s00586-018-5513-4

60. Krenker, A., Bešter, J., & Kos, A. (2011). Introduction to the artificial neural networks. In K. Suzuki (Ed.), *Artificial neural networks: Methodological advances and biomedical applications* (pp. 1-18). InTech.

61. Ledwos, N., Mirchi, N., Bissonnette, V., Winkler-Schwartz, A., Yilmaz, R., & Del Maestro, R. F. (2021). Virtual reality anterior cervical discectomy and fusion simulation on the novel sim-ortho platform: Validation studies. *Operative Neurosurgery, 20*(1), 74-82. https://doi.org/10.1093/ons/opaa269

62. Livingston, K. E., & Perrin, R. G. (1978). The neurosurgical management of spinal metastases causing cord and cauda equina compression. *Journal of Neurosurgery, 49*, 839-843.

63. Louppe, G. (2014). Understanding random forests: From theory to practice. *arXiv*. Published online July 28, 2014. https://arxiv.org/abs/1407.7502

64. Lu, J., Cuff, R. F., & Mansour, M. A. (2021). Simulation in surgical education. *American Journal of Surgery, 221*(3), 509-514.

65. Luciano, C., Banerjee, P., Florea, L., & Dawe, G. (2005). Design of the immersivetouch: A high-performance haptic augmented virtual reality system. Presented at the 11th International Conference on Human-Computer Interaction, Las Vegas, NV.

66. Luciano, C. J., Banerjee, P. P., Bellotte, B., et al. (2011). Learning retention of thoracic pedicle screw placement using a high-resolution augmented reality simulator with haptic feedback. *Operative Neurosurgery, 69*(suppl 1).

67. Luciano, C. J., Banerjee, P. P., Sorenson, J. M., et al. (2013). Percutaneous spinal fixation simulation with virtual reality and haptics. *Neurosurgery, 72*(suppl 1), 89-96.

68. Ludmerer, K. M. (1985). Learning to heal: The development of American medical education. Published online.

69. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765-4774.

70. Luque, E. R. (1982). Segmental spinal instrumentation for correction of scoliosis. *Clinical Orthopaedics and Related Research, 163*, 192-198.

71. Ma, C., Zou, D., Qi, H., et al. (2022). A novel surgical planning system using an AI model to optimize planning of pedicle screw trajectories with highest bone mineral density and strongest pull-out force. *Neurosurgical Focus, 52*(4). https://doi.org/10.3171/2022.1.FOCUS21721

72. Manbachi, A., Cobbold, R. S., & Ginsberg, H. J. (2014). Guided pedicle screw insertion: Techniques and training. *The Spine Journal, 14*(1), 165-179. https://doi.org/10.1016/j.spinee.2013.03.029

73. Martin, J., Regehr, G., Reznick, R., et al. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *Journal of British Surgery, 84*(2), 273-278.

74. McCall, V. (2016, July 29). Body donation cadavers anatomy medical education. *National Geographic*. https://www.nationalgeographic.com/science/article/body-donation-cadavers-anatomy-medical-education

75. McGaghie, W. C. (2015). Mastery learning: It is time for medical education to join the 21st century. *Academic Medicine, 90*(11), 1438-1441. https://doi.org/10.1097/ACM.0000000000000911

76. Mirchi, N., Bissonnette, V., Yilmaz, R., Ledwos, N., et al. (2020). The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PLoS One, 15*(2). https://doi.org/10.1371/journal.pone.0229596

77. Mueller, P. S. (2010). AM last page: Sir William Osler's major contributions to medical education. *Academic Medicine, 85*(7), 1260.

78. Nagi, F., et al. (2023). Applications of artificial intelligence (AI) in medical education: A scoping review. *Studies in Health Technology and Informatics, 305*, 648-651.

79. Patel, E. A., Aydin, A., Cearns, M., Dasgupta, P., & Ahmed, K. (2020). A systematic review of simulation-based training in neurosurgery, part 2: Spinal and pediatric surgery, neurointerventional radiology, and nontechnical skills. *World Neurosurgery, 133*, 874-892.

80. Patel, U. K., Anwar, A., Saleem, S., et al. (2021). Artificial intelligence as an emerging technology in the current care of neurological disorders. *Journal of Neurology, 268*(5), 1623-1642. https://doi.org/10.1007/s00415-019-09518-3

81. Paul, A., Mukherjee, D. P., Das, P., et al. (2018). Improved random forest for classification. *IEEE Transactions on Image Processing, 27*(8), 4012-4024. https://doi.org/10.1109/TIP.2018.2834830

82. Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia, 4*(2), 1883.

83. Pfandler, M., Lazarovici, M., Stefan, P., Wucherer, P., & Weigl, M. J. (2017). Virtual reality-based simulators for spine surgery: A systematic review. *Spine Journal, 17*(9), 1352-1363.

84. Polavarapu, H. V., Kulaylat, A. N., Sun, S., & Hamed, O. H. (2013). 100 years of surgical education: The past, present, and future. *Bulletin of the American College of Surgeons, 98*(7), 22-27.

85. Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics, 2*, 927312. https://doi.org/10.3389/fbinf.2022.927312

86. Ray, W. Z., Ganju, A., Harrop, J. S., & Hoh, D. J. (2013). Developing an anterior cervical diskectomy and fusion simulator for neurosurgical resident training. *Neurosurgery, 73*(suppl 1).

87. Reich, A., Mirchi, N., Yilmaz, R., et al. (2022). Artificial neural network approach to competency-based training using a virtual reality neurosurgical simulation. *Operative Neurosurgery, 23*(1), 31-39. https://doi.org/10.1227/ons.0000000000000173

88. Reznick, R. K., & MacRae, H. (2006). Teaching surgical skills—Changes in the wind. *New England Journal of Medicine, 355*(25), 2664-2669.

89. Rogers, M. P., DeSantis, A. J., Janjua, H., Barry, T. M., & Kuo, P. C. (2021). The future surgical training paradigm: Virtual reality and machine learning in surgical education. *Surgery, 169*(5), 1250-1252. https://doi.org/10.1016/j.surg.2020.09.040

90. Roitberg, B., Banerjee, P., Luciano, C., et al. (2013). Sensory and motor skill testing in neurosurgery applicants: A pilot study using a virtual reality haptic neurosurgical simulator. *Neurosurgery, 73*(suppl 1).

91. Roy-Camille, R., Roy-Camille, M., & Demeulenaere, C. (1970). Osteosynthesis of dorsal, lumbar, and lumbosacral spine with metallic plates screwed into vertebral pedicles and articular apophyses. *Presse Médicale, 78*, 1447-1448.

92. Sankey, E. W., Mehta, V. A., Wang, T. Y., et al. (2020). The medicolegal impact of misplaced pedicle and lateral mass screws on spine surgery in the United States. *Neurosurgical Focus, 49*(5). https://doi.org/10.3171/2020.8.FOCUS20600

93. Sarwahi, V., Wendolowski, S. F., Gecelter, R. C., et al. (2016). Are we underestimating the significance of pedicle screw misplacement? *Spine (Phila Pa 1976), 41*(9). https://doi.org/10.1097/BRS.0000000000001318

94. Sawaya, R., et al. (2017). Virtual reality tumor resection: The force pyramid approach. *Operative Neurosurgery, 14*(6), 686-696.

95. Sawaya, R., et al. (2018). Development of a performance model for virtual reality tumor resections. *Journal of Neurosurgery, 131*(1), 192-200.

96. Shi, J., Hou, Y., Lin, Y., Chen, H., & Yuan, W. (2018). Role of Visuohaptic Surgical Training Simulator in resident education of orthopedic surgery. *World Neurosurgery, 111*. https://doi.org/10.1016/j.wneu.2017.12.015

97. Shahzad, S., & Anwar, M. I. (2021). Apprenticeship model in 21st century's surgical education: Should it perish? *Archives of Surgical Research, 2*(3), 1-3. https://doi.org/10.48111/2021.03.01

98. Sonnadara, R. R., Mui, C., McQueen, S., et al. (2014). Reflections on competency-based education and training for surgical residents. *Journal of Surgical Education, 71*(1), 151-158.

99. Steadman, R. H., Rudolph, M. D., Myo-Bui, C. C., & Matevosian, R. (2016). Incorporating simulation into the medical school curriculum. In R. H. Riley (Ed.), *Manual of simulation in healthcare* (2nd ed.). Oxford Academic. https://doi.org/10.1093/med/9780198717621.003.0010

100. Stockley, D., Egan, R., Van Wylick, R., et al. (2020). A systems approach for institutional CBME adoption at Queen's University. *Medical Teacher, 42*(8), 916-921.

101. Sugiyama, T., et al. (2018). Forces of tool-tissue interaction to assess surgical skill level. *JAMA Surgery, 153*(3), 234-242.

102. Sun, X., Liu, Y., Xu, M., et al. (2013). Feature selection using dynamic weights for classification. *Knowledge-Based Systems, 37*, 541-549. https://doi.org/10.1016/j.knosys.2012.10.001

103. Symgery. (2023). Document compilation. Retrieved from https://symgery.com/wp-content/uploads/2023/10/T-SYM_Doc_Comp.pdf

104. Tang, B., Hanna, G. B., & Cuschieri, A. (2005). Analysis of errors enacted by surgical trainees during skills training courses. *Surgery, 138*(1), 14-20. https://doi.org/10.1016/j.surg.2005.02.014

105. Van Heest, A. E., Armstrong, A. D., Bednar, M. S., et al. (2022). American Board of Orthopaedic Surgery's initiatives toward competency-based education. *JBJS Open Access, 7*(2), e00150. https://doi.org/10.2106/JBJS.OA.21.00150

106. Varghese, C., Harrison, E. M., O'Grady, G., et al. (2024). Artificial intelligence in surgery. *Nature Medicine, 30*(10), 1257-1268. https://doi.org/10.1038/s41591-024-02970-3

107. Vaughan, N., Dubey, V. N., Wainwright, T. W., & Middleton, R. G. (2016). A review of virtual reality based training simulators for orthopaedic surgery. *Medical Engineering & Physics, 38*(2), 59-71.

108. Wang, G. J., Whitehill, R., Stamp, W. G., et al. (1979). The treatment of fracture dislocations of the thoracolumbar spine with halofemoral traction and Harrington rod instrumentation. *Clinical Orthopaedics and Related Research*, 168-175.

109. Waugh, D., & Bailey, P. G. (2013). Medical education. *The Canadian Encyclopedia*. Published online.

110. Wilson, M. S., Middlebrook, A., Sutton, C., Stone, R., & McCloy, R. F. (1997). MIST VR: A virtual reality trainer for laparoscopic surgery assesses performance. *Annals of the Royal College of Surgeons of England, 79*, 403-404.

111. Winkler-Schwartz, A., Yilmaz, R., Mirchi, N., et al. (2019). Machine learning identification of surgical and operative factors associated with surgical expertise in virtual reality simulation. *JAMA Network Open, 2*(8). https://doi.org/10.1001/jamanetworkopen.2019.8363

112. World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*(20), 2191-2194. https://doi.org/10.1001/jama.2013.281053

113. Yilmaz, R., Winkler-Schwartz, A., Mirchi, N., et al. (2022). Continuous monitoring of surgical bimanual expertise using deep neural networks in virtual reality simulation. *NPJ Digital Medicine, 5*(1), 54. https://doi.org/10.1038/s41746-022-00596-8

114. Yilmaz R, Bakhaidar M, Alsayegh A, Hamdan NA, Fazlollahi AM, Tee T, Langleben I,

Winkler-Schwartz A, Laroche D, Santaguida C, Del Maestro RF. (2024) Real-Time multifaceted artificial intelligence vs In-Person instruction in teaching surgical technical skills: a randomized controlled trial. *Sci Rep* **14**, 15130 (2024). https://doi.org/10.1038/s41598-024-65716-8

Supplemental Digital Content

Methods.

Simulated L4 & L5 pedicle screw placement scenario

Dry Lab:

Before starting the simulation, you will go through the dry lab. The dry lab will help you get acquainted with the instruments and the haptic handle. Follow the on-screen instructions to perform this step. This part of the simulation will not be used for evaluation.

With the straight handle: (Please note that you will need to detach the handle after you finish using an instrument in the dry lab only).

1. Using **the Awl**, **<u>create one hole</u>** in the simulated wooden object. You will feel resistance from the haptic handle once you reach the object's surface.

- Press  at bottom left and exit the simulation.

2. Using **the burr**, **<u>drill any one sphere</u>**. You can move the screen to appreciate the shape of the sphere.

- Once done, press  at bottom left and exit the simulation.

3. Using **the pedicle finder**, **<u>create one trajectory</u>** in the simulated object.

- Once done, press confirm and Exit**.**

With the Kerrison handle:

1. Using **the Kerrison**, remove **<u>only three bites</u>** of the simulated object.

- Zoom in as per your convenience.
- Whenever you see the pantry dish at the side of the simulation, make sure you empty your Kerrison.

- Once done, press  at bottom left and exit the simulation.

After the Dry Lab is over, **call the trial attendee**.

Instructions for Pedicle Screw Placement

Objective: You will perform an L5 and L4 pedicle screw placement using four pedicle screws.

The simulation starts at the stage where the lumbar spine is completely exposed.

Note:

1. The magnification is standardized for the whole procedure. Please avoid changing it.
2. The current step of the procedure will appear in bold on the left of the screen.
3. The simulated anatomical model/ the working area will appear at the center of the screen.
4. You can see fluoroscopy images during the procedure, which will appear in the screen's top right corner.
5. The tools that you can use during the procedure are displayed at the bottom of the screen.
6. Perform all the steps of screw placement of each pedicle screw before going to the next screw.

Steps of the Procedure:

Step 1 – Entry Point Creation (Refer to Figure 5)

- Using the **Awl**, choose the entry point of the left L5 pedicle screw. You can verify your entry point using fluoroscopy.
    - To initiate the fluoroscopy, press the right pedal of the foot switch. After that, with the left pedal, you can observe the real-time position of the tools. You can change the fluoroscopic view by pressing LAT (lateral view) or AP (anterior/posterior) on the right side.
    - To go back to the surgery, press the right pedal.
- Once you are satisfied with the entry point, press 

to go to the next step.

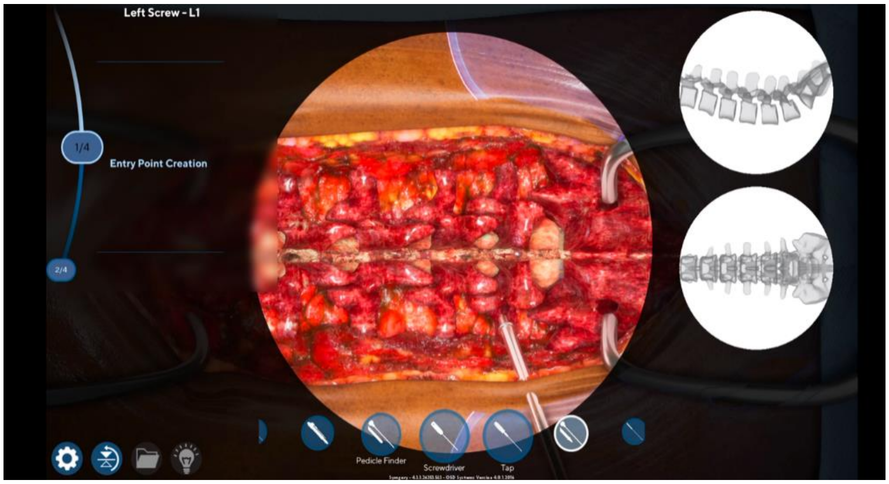

Figure 5: Entry Point Creation

Step 2 – Entry Point Creation with Pedicle Finder (Refer to Figure 6)

- Using the **pedicle finder**, create a channel in the pedicle.

- You can verify the trajectory using fluoroscopy.
  - To initiate the fluoroscopy, press the right pedal of the foot switch. After that, with the left pedal, you can observe the real-time position of the tools. You can change the fluoroscopic view by pressing LAT (lateral view) or AP (anterior/posterior) on the right side.
  - To go back to the surgery, press the right pedal.

- As shown in figure 6, the depth of the instrument will appear at the top center of the screen.

- Once you are satisfied with the depth, press 

to go to the next step.

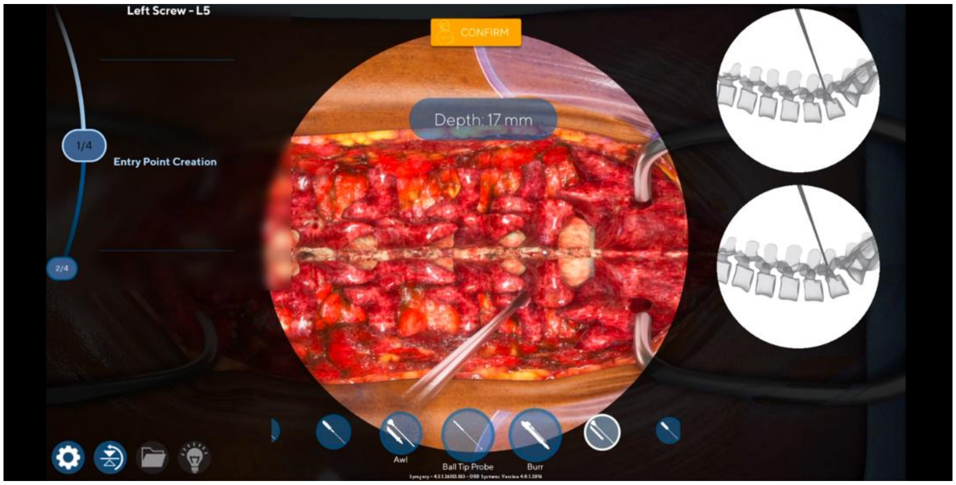

Figure 6: Entry Point Creation Using a Pedicle Finder

Step 3 – Channel Breach Verification (Refer to Figure 7)

- Using the **2 mm ball tip probe**, check for the presence or absence of a pedicle breach.
- As shown in Figure 7, choose whether you felt that you had created a pedicle breach or not, and then select 
.

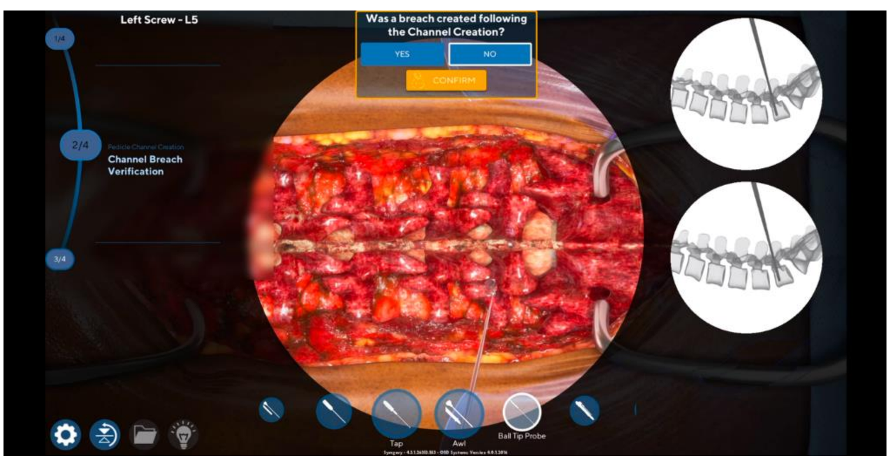

Figure 7: Channel Breach Verification

Step 4 – Tap Insertion (Refer to Figure 8)

- Using the **5.5 mm tap**, tap the previously created channel in the pedicle.

- You can verify the tap insertion using fluoroscopy.
  - To initiate the fluoroscopy, press the right pedal of the foot switch. After that, with the left pedal, you can observe the real-time position of the tools. You can change the fluoroscopic view by pressing LAT (lateral view) or AP (anterior/posterior) on the right side.
  - To go back to the surgery, press the right pedal.

- Once you are satisfied with the tap, press 

to go to the next step.

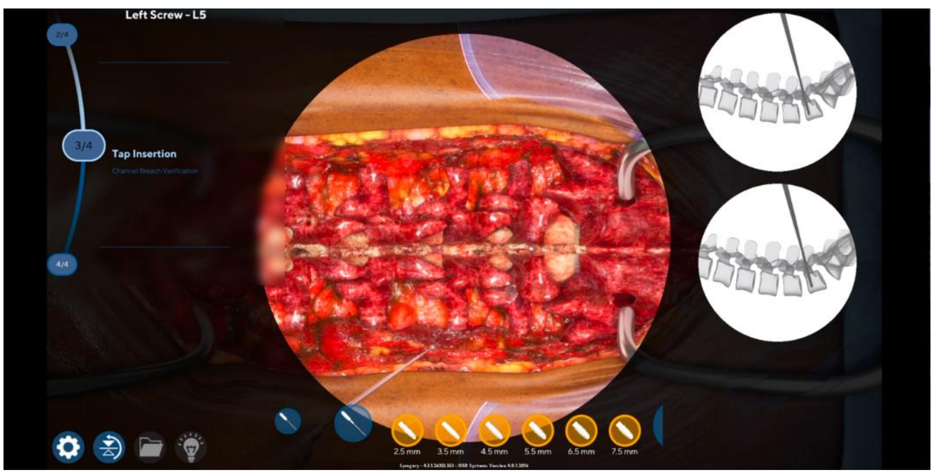

Figure 8: Tap Insertion

Step 5 – Pedicle Breach Verification

- Using a **2mm ball tip probe**, check for the presence or absence of a pedicle breach.
- Choose whether you felt that you created a pedicle breach or not and select 

.

Step 6 – Screw Insertion (Refer to Figure 9)

- Select the **screwdriver** from the instrument list.
- A screen will appear asking you to choose the diameter and length of the screw. Select **6.5mm for the screw diameter and 45mm for the screw length**.
- Insert the selected screw by rotating the screwdriver. To release the screw from the screwdriver, press the button on the handle.
- You can verify the screw insertion using fluoroscopy.
    - To initiate the fluoroscopy, press the right pedal of the foot switch. After that, with the left pedal, you can observe the real-time position of the tools. You can change the fluoroscopic view by pressing LAT (lateral view) or AP (anterior/posterior) on the right side.
    - To go back to the surgery, press the right pedal.

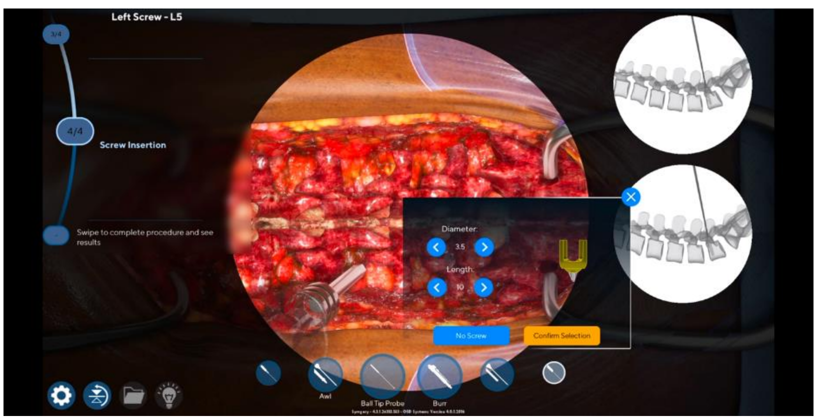

Figure 9: Screw Insertion

- **Step 7:** Repeat Steps 1-6 for the left L4 pedicle.
- After finishing the left L4 pedicle, you can click the orientation button in the bottom left to switch to the right side of the patient.

- **Step 8:** Repeat Steps 1-6 for the right L5 pedicle.

- **Step 9:** Repeat Steps 1-6 for the right L4 pedicle.

- After completing all pedicle screw placements, swipe up at the left side of the screen and select **"Yes, See Results."**

- **The Pedicle Screw Placement scenario concludes now. Please call for the trial attendee to guide you in the next phase.**

List of 65 Metrics Generated by the TSYM Simulator Platform

No	Features	Description
1	3Dacceleration_Awl	Acceleration using Awl Tool
2	**3Dacceleration_Balltip**	**Acceleration using Ball Tip Tool**
3	**3Dacceleration_Pedifind**	**Acceleration using Pedicle Finder Tool**
4	3Dacceleration_Scewdriv	Acceleration using Screwdriver Tool
5	3Dacceleration_Tapscrew	Acceleration using Tap Screw Tool
6	**3DForce_Awl**	**Haptic Force applied using Awl Tool**
7	3DForce_Balltip	Haptic Force applied using Ball Tip Tool
8	**3DForce_Pedifind**	**Haptic Force applied using Pedicle Finder Tool**
9	3DForce_Scewdriv	Haptic Force applied using Screwdriver Tool
10	3DForce_Tapscrew	Haptic Force applied using Tap Screw Tool
11	3Dvelocity_Awl	Velocity while using Awl Tool
12	3Dvelocity_Balltip	Velocity while using Ball Tip Tool
13	3Dvelocity_Pedifind	Velocity while using Pedicle Finder Tool
14	3Dvelocity_Scewdriv	Velocity while using Screwdriver Tool
15	**3Dvelocity_Tapscrew**	**Velocity while using Tap Screw Tool**
16	AverageForceL4Vertebra_Awl	Force applied on L4 vertebra voxels with Awl Tool

17	AverageForceL4Vertebra_Balltip	Force applied on L4 vertebra voxels with Ball Tip Tool
18	AverageForceL4Vertebra_Pedifind	Force applied on L4 vertebra voxels with Pedicle Finder Tool
19	**AverageForceL4Vertebra_Scewdriv**	**Force applied on L4 vertebra voxels with Screwdriver Tool**
20	**AverageForceL4Vertebra_Tapscrew**	**Force applied on L4 vertebra voxels with Tap Screw Tool**
21	**AverageForceL5Vertebra_Awl**	**Force applied on L5 vertebra voxels with Awl Tool**
22	AverageForceL5Vertebra_Balltip	Force applied on L5 vertebra voxels with Ball Tip Tool
23	AverageForceL5Vertebra_Pedifind	Force applied on L5 vertebra voxels with Pedicle Finder Tool
24	AverageForceL5Vertebra_Scewdriv	Force applied on L5 vertebra voxels with Screwdriver Tool
25	AverageForceL5Vertebra_Tapscrew	Force applied on L5 vertebra voxels with Tap Screw Tool
26	**AvgContactrelTool_Awl**	**Position of average contact point between tool and vertebra**
27	**AvgContactrelTool_Balltip**	**Position of average contact point between tool and vertebra**
28	**AvgContactrelTool_Pedifind**	**Position of average contact point between tool and vertebra**
29	**AvgContactrelTool_Scewdriv**	**Position of average contact point between tool and vertebra**
30	**AvgContactrelTool_Tapscrew**	**Position of average contact point between tool and vertebra**
31	**ButtonPressed_Scewdriv**	**Button pressed on Screwdriver Tool**
32	ButtonPressed_Tapscrew	Button pressed on Tap Screw Tool
33	ContactVoxelsL4Vertebra_Awl	Number of L4 vertebra voxels in contact with Awl tool
34	ContactVoxelsL4Vertebra_Balltip	Number of L4 vertebra voxels in contact with Ball Tip Tool
35	ContactVoxelsL4Vertebra_Pedifind	Number of L4 vertebra voxels in contact with Pedicle Finder Tool
36	ContactVoxelsL4Vertebra_Scewdriv	Number of L4 vertebra voxels in contact with Screwdriver Tool
37	ContactVoxelsL4Vertebra_Tapscrew	Number of L4 vertebra voxels in contact with Tap Screw Tool
38	ContactVoxelsL5Vertebra_Awl	Number of L5 vertebra voxels in contact with Awl Tool
39	ContactVoxelsL5Vertebra_Balltip	Number of L5 vertebra voxels in contact with Ball Tip Tool
40	ContactVoxelsL5Vertebra_Pedifind	Number of L5 vertebra voxels in contact with Pedicle Finder Tool
41	ContactVoxelsL5Vertebra_Scewdriv	Number of L5 vertebra voxels in contact with Screwdriver Tool

42	ContactVoxelsL5Vertebra_Tapscrew	Number of L5 vertebra voxels in contact with Tap Screw Tool
43	CutVoxelsL4Vertebra_Awl	Number of L4 vertebra voxels removed using Awl Tool
44	**CutVoxelsL4Vertebra_Pedifind**	**Number of L4 vertebra voxels removed using Pedicle Finder Tool**
45	CutVoxelsL4Vertebra_Scewdriv	Number of L4 vertebra voxels removed using Screwdriver Tool
46	CutVoxelsL4Vertebra_Tapscrew	Number of L4 vertebra voxels removed using Tap Screw Tool
47	CutVoxelsL5Vertebra_Awl	Number of L5 vertebra voxels removed using Awl
48	CutVoxelsL5Vertebra_Pedifind	Number of L5 vertebra voxels removed using Pedicle Finder Tool
49	CutVoxelsL5Vertebra_Scewdriv	Number of L5 vertebra voxels removed using Screwdriver Tool
50	CutVoxelsL5Vertebra_Tapscrew	Number of L5 vertebra voxels removed using Tap Screw Tool
51	**IsCutting_Awl**	**If bone or soft tissue is being cut or removed**
52	**IsCutting_Pedifind**	**If bone or soft tissue is being cut or removed**
53	IsCutting_Scewdriv	If bone or soft tissue is being cut or removed
54	IsCutting_Balltip	If bone or soft tissue is being cut or removed
55	IsCutting_Tapscrew	If bone or soft tissue is being cut or removed
56	MaxForce_Awl	Maximum Haptic Force applied using Awl Tool
57	MaxForce_Balltip	Maximum Haptic Force applied using Ball Tip Tool
58	**MaxForce_Pedifind**	**Maximum Haptic Force applied using Pedicle Finder Tool**
59	MaxForce_Scewdriv	Maximum Haptic Force applied using Screwdriver Tool
60	**MaxForce_Tapscrew**	**Maximum Haptic Force applied using Tap Screw Tool**
61	**ToolContact_Awl**	**Awl Tool contact with vertebra. It returns binary output**
62	**ToolContact_Balltip**	**Ball Tip Tool contact with vertebra**
63	**ToolContact_Pedifind**	**Pedicle Finder Tool contact with patient**
64	**ToolContact_Scewdriv**	**Screwdriver Tool contact with vertebra**
65	**ToolContact_Tapscrew**	**Tap Screw Tool contact with vertebra**

Table 1: List of 65 metrics including a description.