Artificial Neural Network Approach to Competency-Based Training

Aiden Reich, BEng
Experimental Surgery
McGill University, Montreal
July 2020

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Master of Science in Experimental Surgery

TABLE OF CONTENT

ABSTRACT

Objective: The methodology of assessment and training of surgical skills is evolving to deal with the emergence of competency-based training. Machine learning algorithms have been employed to assess surgical expertise during virtual reality surgical performance. Some of these approaches fail to outline the underlying reasons for classification and to quantify the relative importance of each metric utilized to train the model. Artificial neural networks, a branch of artificial intelligence, can utilize newly generated metrics not only for assessment but can quantitate individual metric contribution and provide new insights into surgical expertise. This study aims to outline the multiple educational utilities of employing an artificial neural network in the assessment and quantitation of surgical expertise. A virtual reality vertebral osteophyte removal during a simulated surgical spine procedure is utilized as a model to outline this methodology.

Design: Participants performed a simulated anterior cervical discectomy and fusion on the Sim-Ortho virtual reality simulator platform. Data was retrieved from the osteophyte removal component of the scenario, which involved utilizing a simulated drill. The data was manipulated to generate an initial 83 performance metrics spanning 3 categories (safety, efficiency, motion). The most relevant metrics were utilized to train and test the artificial neural network.

Setting: This study was carried out at the Neurosurgical Simulation and Artificial Intelligence Learning Centre at McGill University affiliated with the Montreal Neurological Institute and Hospital (Montreal, Canada).

Participants: Twenty-one participants performed a simulated anterior cervical discectomy and fusion. Participants were divided into 3 groups, including 9 post-residents, 5 senior and 7 junior residents.

Results: The artificial neural network model trained on the six most relevant metrics of performance, all involving safety, achieved 83.3% testing accuracy misclassifying only one participant.

Conclusions: This study outlines the potential utility of artificial neural networks which allows a deeper understanding of the composites of surgical expertise and may contribute to the paradigm shift towards competency-based surgical training.

RÉSUMÉ

Objectif : La méthodologie d'évaluation et de formation des compétences chirurgicales évolue pour faire face à l'émergence de formation par compétence. Les algorithmes d'apprentissage automatique ont été utilisés afin d'évaluer l'expertise chirurgicale lors de performances chirurgicales en réalité virtuelle. Certaines de ces approches ne parviennent pas à expliquer le raisonnement de la classification et à quantifier l'importance relative de chaque métrique utilisée pour entrainer le modèle. Les réseaux de neurones artificiels, une branche de l'intelligence artificielle, peuvent utiliser des métriques nouvellement générées non seulement en but d'évaluation, mais peuvent quantifier la contribution de métrique individuelle et fournir de nouvelles perspectives sur l'expertise chirurgicale. Cette étude vise à décrire l'utilité éducative de l'emploi d'un réseau neuronal artificiel quant à l'évaluation et la quantification de l'expertise chirurgicale. Une ablation d'ostéophytes vertébraux en réalité virtuelle au cours d'une procédure chirurgicale simulée de la colonne vertébrale est utilisée comme modèle afin de décrire cette méthodologie.

Conception : Les participants ont effectué une discectomie cervicale antérieure et fusion simulée sur le Sim-Ortho, simulateur de réalité virtuelle. Les données ont été récupérées à partir du composant d'ablation des ostéophytes du scénario, qui impliquait l'utilisation d'une perceuse simulée. Les données ont été manipulées afin de générer 83 métriques de performance initiales couvrant 3 catégories (sécurité, efficacité, mouvement). Les métriques les plus pertinentes ont été utilisées afin d'entrainer et tester le réseau neuronal artificiel.

Cadre : Cette étude a été réalisé au Neurosurgical Simulation and Artificial Intelligence Learning Centre à McGill University affilié à l'Institut-hôpital neurologique de Montréal (Montréal, Canada).

Participants : Vingt et un participants ont effectué une discectomie cervicale antérieure et fusion simulée. Les participants étaient divisés en 3 groupes, dont 9 «post-residents», 5 «seniors» et 7 «juniors».

Résultats : Le modèle de réseau de neurones artificiels entrainé à partir des six métriques les plus pertinentes, toutes reliées à la sécurité, a atteint 83.3% de précision à tort d'un seul participant.

Conclusions : Cette étude souligne l'utilité potentielle des réseaux de neurones artificiels permettant une compréhension plus approfondie des composites de l'expertise chirurgicale et peut contribuer au changement de paradigme vers une formation chirurgicale basée sur les compétences.

ACKNOWLEDGEMENT

First and foremost, I would like to express my utmost thanks to my supervisor, Dr. Rolando Del Maestro, without whom this project would not have been possible. Taking on the role of principle investigator, Dr. Del Maestro's transfer of knowledge related to the scientific research field, medicine and surgical education as well as his guidance and encouragement have served as pillars for my success and accomplishments throughout this whole process. Surpassing far more than what can be expected from a supervisor, he has also acted a mentor offering constant life advice and guidance and provided never-ending inspiration. It has been a great honour to have been his graduate student.

Immense gratitude is extended to all my colleagues at the Neurosurgical Simulation and Artificial Intelligence Learning Centre. Throughout our countless professional and personal discussions, you have helping me thrive and accomplish many goals. To Nykan Mirchi, Dr. Vincent Bissonnette, Nicole Ledwos and Dr. Alexander Winkler-Schwartz, thank you for all your help and advice providing me with more knowledge and critical ways of thinking. To Dr. Recai Yilmaz and Dan Huy Tran, I am beyond grateful for the time we spent together in or outside of the lab. From learning the basics of artificial intelligence to brewing delicious Turkish tea, your knowledge, support and friendship have been essential towards helping me succeed. Thank you for an enriching experience filled with many educational, encouraging and enjoyable moments. I could not have asked for better colleagues through this journey. To Dr. Berkir Karlik, Dr. Swajan Paul, and all other researchers who collaborated with the lab, thank you for providing continuously learning experiences throughout my time at the lab.

I would like to thank all the members of my research advisory committee: Dr Gregory Berry, Dr. Jacques Lapointe, Dr. Roy Dudley and Dr. Derek Rosenzweig who manifested their distinguished skills and knowledge through insightful comments and encouragements.

I would like to extend my sincere gratitude, firstly, to the AO Foundation who helped fund my research and, secondly, to the Franco Di Giovanni Foundation who have also done so and have been financially supporting the Neurosurgical Simulation and Artificial Intelligence Learning Centre for many years.

A special thank you to Léa Rafih who has been a strong moral support along the way. She has pushed me to succeed throughout my academic journey and inspire me to strive no matter the context.

Finally, to my parents, I owe all my success to them. For without their continuous love and support, I could not have achieved everything that I have. They have shown me that their advice and guidance is essential to any life goal I seek to accomplish. They are my role models and my biggest source of inspiration. I would not be where I am today without them. Thank you.

PREFACE AND AUTHOR CONTIBUTIONS

The thesis, in the following, is presented as a manuscript which has been submitted to the BMJ Simulation & Technology Enhanced Learning (BMJ STEL). Following the McGill University's guideline, this manuscript has been structured with addition insightful sections and comprehensive information. The data utilized in this study was previously collected by the Neurosurgical Simulation and Artificial Intelligence Learning Centre.

The candidate led the study throughout its entirety including manipulating the raw data, developing the artificial neural network methodology, interpreting the results and writing the manuscript.

Nykan Mirchi, Nicole Ledwos and Dr. Vincent Bissonnette were involved in the data collection by recruiting participants and running the trials the previous year.

Dr. Recai Yilmaz, Nykan Mirchi and Dr. Vincent Bissonnette offered their knowledge and know how related to artificial intelligence.

Dan Huy Tran and Dr. Swajan Paul contributed through providing their knowledge of surgical education and suggestions related to the development of metrics.

Dr. Alexander Winkler-Schwartz assisted with the recruitment of participants and shared his surgical perspective related to the development of the simulated scenario prior to this study. He also offered his insight on surgical training.

Dr. Bekir Karlik provided his expertise in the field of artificial intelligence and aided in the optimization of the model.

Dr. Rolando Del Maestro supervised the study from beginning to end including its planning and its direction. Prior to beginning this study, he was also involved in the recruitment process as well as the improvement of the simulator. Dr. Del Maestro offered his surgical expertise and extensive knowledge of surgical education for the interpretation of the results and future applications of the artificial neural network approach to competency-based training.

ABBREVIATIONS

AI: Artificial Intelligence

ACDF: Anterior Cervical Discectomy and Fusion

ANN: Artificial Neural Network

CWP: Connection Weight Product

THESIS INTRODUCTION

With the rapid emergence of artificial intelligence (AI) and constant development of this technology in the past few years, the medical field has recognized its potential and the introduction of this technology into practice to help solve a variety of clinical problems.[1-4] Artificial intelligence is defined as a computational method of making algorithm based decisions through input information such as performance metrics allowing educational training and objective assessments.[5,6] Also known as machine learning, artificial intelligence can be divided into three main categories: simple classifiers, artificial neural networks (ANN) and deep learning. Without being as complex as deep learning yet having more depth than simple classifiers, artificial neural networks are designed as a web of interconnected neurons (known as perceptrons) that interact between each other to resemble the structure of the human nervous system.[7] An artificial neural network identifies and reveals hidden patterns within a large dataset and, throughout its interconnected structure, attributes weights associated with each input value or performance metric in the case of this study.[8] The magnitude of the weights corresponds to the sensitivity of each metric on the algorithm's decision-making process. Returning to the neuronal connectivity analogy of the brain, weights of varying magnitudes throughout the network can be compared to polarizing or depolarizing signals through synapses. With its decision-making functionality, artificial neural networks can be employed as an educational tool to accompany the current shift towards competency-based training.[9,10]

A second highly evolving technology contributing to the paradigm shift towards competency-based surgical training is virtual reality (VR) simulation. The idea of the utilization of virtual reality simulation for surgical training actually became apparent years ago with the noticeable success of simulation-based training in aviation.[11] Providing reproducible and repeatable practice opportunities to improve surgical performance and to accelerate learning curves in risk-free environment, more and more virtual reality simulators are being developed in all fields of surgery.[9,12-15]

Commonly performed spine procedures such as an anterior cervical discectomy and fusion (ACDF) are complex enough to be considered as a good candidates for simulation training. Such multifaceted procedures involve the use of knowledge and technical skills to identify critical anatomical structures to develop a sense of the variable tolerances of surgical instruments interacting with a variety of structures.[16]

Virtual reality simulators being computer-based are able to constantly record, throughout the procedure, information related to the use of the virtual surgical instruments or the manipulations of the simulated anatomical structures. Our group at the Neurosurgical Simulation and Artificial Intelligence Learning Centre decided to apply artificial intelligence to the large datasets of information generated by the simulated anterior cervical discectomy and fusion. This was based on the potential usefulness of this technology in providing novel insights into complex questions. The integration of artificial intelligence within surgical simulation and its educational utility have previously been explored in other studies focusing primarily on establishing an assessment of surgical performance through the classification of individuals into their respective groups of varying expertise.[5,6,17-21] An article on the best practices for utilizing machine learning within virtual reality simulation to assess performance establishes a set of criteria entitled The Machine Learning to Assess Surgical Expertise (MLASE) checklist that serves as a guideline for researchers.[4] Despite the objective assessment that may allow trainees to learn in a more independent manner, these methods fail to outline the underlying reasons for classification and to quantify the relative importance of each performance metric utilized to train the model.

This study addresses multiple questions. Can an artificial neural network model provide an accurate assessment of surgical performance through the classification of individuals into groups of varying expertise level for a specific simulated spine procedure? How can a medical educator or trainee utilize this approach to gain additional information related to performance and a better understanding of expertise? Is this novel approach to competency-based training a potential tool to help shape the future of surgical training? Our hypotheses are that it will be possible to classify participants into either a junior resident, senior resident or post-resident level group according to their surgical performance of the vertebral osteophyte removal component of a simulated anterior cervical discectomy and fusion by introducing appropriate performance metrics related to motion, safety and efficiency into an artificial neural network and that a better understanding of surgical expertise will be provided through the analysis of the model's decision-making process revealing the underlying reasons for a specific classification. The objectives stemming from this hypothesis are:

1. To introduce an artificial neural networks methodology for assessing expertise in simulation-based training.
2. To outline the utility of artificial neural networks by demonstrating their usefulness in outlining novel metrics and the contributions of individual metrics to the composites of surgical performance utilizing a virtual reality spinal procedure model.

This study focuses on the educational utility of the artificial neural network approach to competency-based training. It outlines the potential of this method to reveal the composites of expertise alongside the quantifiable measures of their contributions and provides the necessary information to understand the interplay of these composites leading trainees to acquire the psychomotor and technical skills involved with expert surgical performance.

BACKGROUND

Past and Future of Surgical Training

Traditionally, surgical training and education has been taught through an apprenticeship model in surgical fields.[22,23] Due to constant changes in which surgery is practiced, the past models have substantial limitations.[24] There is a shift in the paradigm of surgical training from the apprenticeship model to a competency-based system.[10,25] Competency-based training in surgery can be defined as a method ensuring that graduating residents are competent, in parts, related to medical knowledge, technical skills and mindset allowing them to provide proper services to patients.[26,27] However, in order to comprehend the need for change, it is necessary to understand the limitations of past models in our current and future educational planning.

In the 19th century, physicians and surgeons in Europe would visit the leading hospitals of major cities such as London, Paris or Vienna to learn novel surgical methods and procedures.[24] With self-training and some form of apprenticeships being the accepted norm, no formal and standardized training programs existed before 1890 when William Steward Halstead introduced the first residency program and became the first chief of surgery at the Johns Hopkins Hospital in the United States.[24,28] By creating this formal training program with his "see one, do one, teach one" model whereby a trainee having observed a specific surgical procedure is expected to then have the capabilities to perform said procedure and finally teach the procedure to another trainee, Halstead envisioned a more efficient and effective way of transferring surgical knowledge and skills through apprenticeships.[28] The standardized apprenticeship model designed by Halstead is primarily based on two fundamental principles: a collaborative proximity of both the trainee and the expert to the surgical procedures and enough time spent doing so.[24] It is designed to have trainees gain increasing responsibilities as they learn throughout their years as residents and ultimately achieve virtually full independence.[28] Halstead, having trained in Germany, employed the concept of "graded responsibility" which was an integral part of the training system German residents would go through.[29] This model acted as the pillar for current residency programs by establishing a structure to surgical training and setting a new standard. Halstead's model has remained practically unchanged over the past century and is still regarded as the traditionally employed residency program of today.[24,28]

The apprenticeship model has a number of limitations.[25,29,30] Concerns regarding patient safety, more demanding educational expectations, and novel simulation technologies are becoming more apparent and have resulted in a reevaluation of the apprenticeship model of surgical training.[25] Furthermore, as this model is built upon the notion that a trainee will learn from an expert and eventually become an expert and teach a new trainee, the amount of time required is considerable and based on both time available and the expertise of the surgeon teacher. In 2003, the American Accreditation Council for Graduate Medical Education (ACGME) significantly reduced the average amount of hours permissible for a resident to work per week to 80 hours.[24,28] This regulation was implemented to address both residents' physical and mental health as well as patient safety.[31] Following this change, studies have failed to reach a consensus on the effects of the 80-hour workweek rule on the burnout rate of residents since study results have been inconsistent.[32-34] With no significant results, the only definite outcome from the regulation change is the reduced amount of time residents spend during surgical procedures. Therefore, considering exposure as one of the fundamentals principles that characterizes the current surgical training program for residents, new methods need to be considered to compensate for this lack of direct operative exposure.

As the future of surgical training pivots towards more competency-based training, more novel simulation technologies are available and trainees are looking for alternative ways to gain experience, surgical simulators are becoming increasingly utilized for assessment and evaluation.

Surgical Simulation

The military and aviation have at least one major commonality with surgery. Errors involving the technical skills of individuals are high-risk. It is expected that pilots have the knowledge required to fly a plane, but, even more so, to have the ability to actually fly one.[25] Surgical training is decades behind pilot training for airlines and the military,[24,25] in part, due to the late integration of simulation into surgical training. Simulation has the role of replicating various scenarios in realistic environments for assessment and feedback, thus, creating an ideal educational platform due its standardized, reproducible and safe environment features.[35] Pilots and surgeons alike has always been regarded as having high-risk professions. However, pilots require manual skills assessment through simulation in order to be certified for the airline industry,[36] whereas surgical trainees' skills and competencies are not reliably assessed.[35]

With the advancements of simulation technologies in the surgical field, it is important to assess the utilization of simulation in the training of high-risk scenarios. First, surgical simulation allows trainees to gain the multitude of critical skills in cognitive, technical, psychomotor and clinical domains.[35,37,38] Even experienced surgeons can reinforce acquired skills through repeated practice sessions performing high-risk simulated scenarios. The notion of repetition is a key element for the consolidation of skills and the development and maintenance of expertise.[35,37] Second, simulators replicate various surgical scenarios in realistic environments without actually being in the operating room or potential putting a patient at risk.[39,40] Third, with most surgical procedures being complex and requiring multiple steps, these simulators focus trainees on components of the procedure requiring improvement and development of technical skills by deconstructing the operative procedure into simple and convenient tasks.[41,42] Finally, a proper surgical simulator should provide educational information such as a quality assessment of performance and detailed feedback.[28,35,37,43] An objective assessment and structured feedback are integral components for the improvement of surgical performance through simulation-based training as the trainee's progress can be tracked and the learning is more effective.[44] However, not all simulators provide such assessments and feedback. Hence, many simulation-based training program have not been able to properly integrate these two important key elements.[44]

There are multiple types of surgical simulators available varying from low to high-fidelity which relates to the complexity and realism of the simulation.[37,45] Studies comparing the two types of simulators observed different results in regards to the superior model.[46] Depending on the task and the individual's level of expertise, a simulator with an appropriate fidelity must be utilized for training.[46] Low-fidelity simulators including primarily bench models and video box trainer are inexpensive, usually easy to build and portable, and use real surgical instruments.[37] However, these low-fidelity simulators provide limited feedback and require the presence of a medical instructor to observe and assess surgical performance.[37] Most current simulation-based training involves direct observation by an expert to provide feedback for proper learning.[43] One example is the Fundamentals of Laparoscopic Surgery video box trainer that is simple, low-cost and includes an optical system that mimics the real operative equipment utilized for laparoscopic procedures.[47] A set of simulations have been developed for this video box trainer with the purpose of training and assessing psychomotor skills.[47] Such a simulator fulfills its intended role in laparoscopic surgery, but other types of surgery or surgical tasks may require more immersive and realistic simulations that only a high-fidelity simulator can provide. These high-fidelity simulators can be divided into subcategories including virtual reality simulators, procedural simulators, mannekins and a hybrid of virtual reality and mannekin.[36,37] A major advantage of these simulators is that they can combine many different operative tasks in order to train for an entire surgical procedure,[36] whereas low-fidelity simulators focus on one specific task at a time. Furthermore, most of these simulators, being computer-based systems, compile immense amounts of information during a simulation. Therefore, simulators involving virtual reality technology reduce biased assessments and feedback in part due to the information generated providing a more quantitative and evidence-based understanding of learning progression.[48] A recurring criticism is the lack of realism in some virtual reality simulations.[49,50]. With computers becoming increasingly more powerful, this problem is being addressed. Continuum mechanics techniques such as finite elements have the potential to aid with the realism of virtual reality simulations.[50] A good example of a virtual reality simulator with high-quality realism is the NeuroVR.[51]

Originally developed as a collaboration between the National Research Council of Canada and the Neurosurgical Simulation and Artificial Intelligence Learning Centre under the name of NeuroTouch, this simulator is regarded as the most advanced and highly realistic in the field of neurosurgery.[9,51] Combined with the visual, auditory and haptic feedback, this virtual reality simulator utilizes a finite element model to create highly realistic interactions such as deformation in between anatomical tissues and the instruments being manipulated.[9] Designed for neurosurgical training, very few simulated scenarios on the NeuroVR are spine procedures.

The field of orthopedic surgery like many other surgical specialties has been slow in integration simulators into training programs.[12,52] Few simulators have been developed to address the needs associated with orthopedic training. The main challenge in regards to the development of orthopedic virtual reality simulators is to respect the many different anatomical structures in terms of morphology and tactile feeling.[53] Unlike other kinds of surgery, orthopedic surgery involves soft tissues such as muscles and ligaments as well as hard structures like bones. Therefore, the haptic feedback transmitted through the instruments utilized during an orthopedic simulation must be highly variable. Attempts to develop virtual reality simulators without haptic feedback became an option, however, the lack of tactile information limited their potential for surgical training.[53] It has been shown that haptic-based simulators provide a superior platform for surgical training.[54,55] A systematic review that investigated existing studies utilizing virtual reality simulators in spinal procedures revealed only 19 articles in which only 5 different simulators were mentioned.[40] Two noticeable aspects were highlighted by this review. First, the majority of the simulated scenarios on these simulators were simple operative tasks requiring few steps such as lumbar mass screw or pedicular screw placements, lumbar punctures or vertebroplasty.[40] There has been limited use of high-fidelity simulators. Another study investigating a mixed reality simulation of an artificial cervical disc replacement demonstrated that more complex scenarios involving spinal procedures are possible.[56] Second, augmented reality and mixed reality simulators were identified, alongside virtual reality simulators, in the previous review.[40] Mixed reality simulators are the combination of both a virtual platform of simulation and a physical one developed in an effort to address the issue of inaccurate haptic feedback associated with pure virtual reality simulators.[57,58] This systematic review outlined the underlying need to develop new multi-faceted scenarios for spinal procedures on virtual reality simulators with proper haptic feedback.

A spinal procedure to consider for such a scenario is the anterior cervical discectomy and fusion. This operative procedure can be separated into multiple components each involving different surgical instruments that interact with various anatomical structures ranging from soft tissues like the intervertebral disc and much harder structures like the actual vertebrae. Despite the challenging nature of developing a realistic virtual reality spinal procedure, a small startup company, OSSimTech™ (Montreal, Canada), has developed a virtual reality simulator which includes a variety of orthopedic scenarios including an anterior cervical discectomy and fusion.

Virtual Reality Spine Surgery Simulator

OSSimTech™ (Montreal, Canada) developed the Sim-Ortho Simulator in collaboration with the AO Foundation (Davos, Switzerland). Unlike the NeuroVR previously mentioned, this virtual reality simulator is driven by a gaming system and utilizes 3D glasses to provide realistic visual feedback. Furthermore, it is equipped with a single 5 degrees of freedom haptic system which appropriately replicates the variable applied forces by the instruments interacting with and deforming soft tissues and hard bones.[12,59] Finally, there is auditory feedback generated by the simulator which is produced while cutting or drilling into the different anatomical structures. Therefore, by combining these three appropriately true means of feedback with an anterior cervical discectomy and fusion scenario, the Sim-Ortho can be considered as a potentially useful virtual reality simulator for training of a complex surgical spine procedure.

Aside from sensory feedback systems, it was previously mentioned that virtual reality simulators built upon computer-based systems could provide an objective assessment of surgical performance as well as educational feedback while tracking progress through repeated training sessions. These features are possible through the collaborative effort of the haptic and computer-based systems by precisely tracking the instrument movements and force applications. Subsequently, advanced metrics related to psychomotor skills are available.[12] As previously mentioned, objective assessments aim to reduce bias by adopting a factual point of view and, therefore, do so through the analysis of these metrics which are the basis of objectively assessing the surgical performance and skills of an individual on a virtual reality simulator. Using common statistical methods such as t-tests or ANOVAs, analysis of performance metrics was demonstrated by many previous studies in a various fields of surgery.[60-64] These methods independently compare specific metrics of individuals at different levels of expertise and calculate if there is significant difference. However, with literature focusing on competency and defining it as a multifaceted set of skills culminating in desirable outcomes,[65] assessments of performance should be determined through a combination of metrics rather than the analysis of each metric independently. With the advancements of artificial intelligence and its known capability of processing large datasets, a novel methodology of analyzing data and metrics from virtual reality simulators should be considered.

Artificial Intelligence

Artificial intelligence, also known as machine learning, is a division of computer science that intends to mimic human intelligence and behavior within computers. Therefore, due to this artificial intelligence, computers may have abilities associated with human intelligence such as learning and problem solving. This advanced field of computer science can further be subdivided into 3 branches: simple classifiers, artificial neural networks and deep learning. Each of these approaches process and analyze large datasets in an attempt to recognize and reveal hidden patterns. Then, based on the learned information, the artificially intelligent algorithms can proceed with making a decisive action. This is main role of artificial intelligence, whereby it can learn on its own and then proceed to making decisions without being explicitly programmed to do so. Moreover, these algorithms have increasingly stronger capabilities from a simple classifier to deep learning as seen in Figure 1.

Many algorithms can be found within the simple classifier category such as Support Vectors Machines, Decision Trees, K-Nearest Neighbors or Naïve Bayes amongst others. Although each of these perform some variant of a similar statistical approach, previous studies have shown that the effectiveness and accuracy of a simple classifier may vary from one case to the other.[66-69]

Artificial neural networks also known as shallow neural networks are more complex. Designed to resemble the neuronal architecture of a brain with its programmed interconnected perceptrons (or neurons) separated into an input layer, a hidden layer and an output layer, an artificial neural network is an algorithm with similar learning capabilities as the previous simple classifiers. However, the interconnectivity of the perceptrons influences the decision-making process of the artificial neural network through specific weights attributed to each perceptronal connection. Once the artificial neural network has completed its learning (or training), certain inputs will generate different outputs through specific interconnected perceptronal paths of corresponding weights. As previously mentioned, the artificial neural network design functions in a similar manner to intertwined neurons transmitting and receiving synapses from one to the other through variant levels of stimuli whereby only certain stimuli will generate actual synapses. In this study, an artificial neural network was employed.

The third and most sophisticated subcategory of artificial intelligence is deep learning. As opposed to a shallow neural network, deep learning involves deep neural networks which are named after the 2 or more hidden layers in their design. These algorithms become increasingly more complex with each new layer of perceptrons added allowing stronger decision-making capabilities responsible for more subtle decisions.[70]

For each of these branches of artificial intelligence, the learning paradigms can be characterized as either supervised or unsupervised learning. Supervised learning is described to be a method of learning utilizing a training dataset where both the input data and the proper classification associated with that data are provided.[71]

Knowing both the input information as well as the correct classification assigned to that information, the artificially intelligent model can train itself to classify future undefined input datasets. Employed in this study, the supervised learning method used metrics derived from the surgical performance of individuals on the simulator as the input data alongside the predefined levels of expertise of each individual to train the artificial neural network. Therefore, by learning how sets of performance metrics relate to certain expertise groups, the model should be able to differentiate new individuals into the appropriate groups according to their surgical performances. Contrarily to supervised learning, unsupervised learning does not utilize input training data that has previously been labeled.[71] This method allows models to train without referring to the solution (classification labels), hence, obliging the model to discover hidden patterns in unlabeled input datasets.[71] This aspect of unsupervised learning allows the discovery of patterns that might not have been identified by a supervised learning model. However, the method of unsupervised learning requires a much larger dataset to function optimally. Had this study involved more participants and allowed for a greater data collection, unsupervised learning would have been possible. In such a case, the model being trained would not have known which input datasets of performance metrics corresponded to which expertise group allowing the model to classify the data into groups based on the similarities between groups as well as other hidden patterns identified.

With its multiple functionalities, machine learning can be considered as a useful tool when utilized in combination with virtual reality simulators for analyzing the large datasets generated. First, as opposed to traditional statistical methods that evaluate performance metrics individually, machine learning methodology takes a different approach by analyzing the complete data of intimately interconnected metrics which together characterize surgical performance. Second, this automated analysis allows the differentiation of individuals into two or more groups of various expertise levels. This classification serves as an objective assessment tool for surgical performance. Third, some machine learning algorithms have the capabilities of providing a better understanding of surgical expertise through the ranking of performance metrics relative to their importance for each surgical expertise group. This last feature provides insight into the contributions of each metric to expert performance. Therefore, out of the set of relevant performance metrics utilized to classify an individual, one may identify which of those metrics have a higher relative importance and, subsequently, have a greater impact on the decision-making process of the algorithm. Combining the usefulness and the educational utility of artificial intelligence and surgical simulation, it is reasonable to consider that the field of education would increasingly employ this technology.

Artificial Neural Networks Approach to Surgical Competency

The increasing use of artificial intelligence in the field of medical and surgical competency has resulted in many publications outlining a variety of applications dealing with these decision-making processes.[4,72] Out of 77 published articles identified in two separate systematic reviews, one conducted by Dias et al. and the other by the McGill Neurosurgical Simulation and Artificial Intelligence Learning Centre,[4,72] the vast majority of these studies utilized simple machine learning classifiers such as the support vector machine algorithm. This may be explained by the fact that the decision-making processes of simple classifier algorithms of this kind are usually well defined and understood unlike artificial neural networks. The complexity of the artificial neural network may be one of the reasons few studies employ this type of algorithm to resolve classification problems such the identification of a mass or lesions to aid with a diagnosis or the differentiation of skill levels to assess performance. Ten studies utilized artificial neural networks of which only 2 involved the assessment of surgical performance.

The first study utilized artificial neural networks to predict the academic performance in surgical training for residents within a variety of training programs in the United States.[73] Residents answered questions related to their own behavioural styles, motivators and acumen characteristics. Provided with this information as well as the American Board of Surgery In-Training Examination (ABSITE) scores of each resident which are used by program directors to evaluate the performance of residents in a standardize manner, the artificial neural network can be trained to classify and assess residents according to their psychosocial skills and characteristics outlined. Although these psychosocial skills are essential, surgical residents must develop skills in all six ACGME competencies which include the necessary psychomotor and technical skills associated with surgery and the Patient Care competency.[39,74] Training and development of such skills occur in the operating room, but may also involve the use of virtual reality simulation or other training device.[74,75] The second study involved the utilization of an artificial neural network to assess surgical performance of both simulated and live surgeries by tracking eye movements.[19] Despite having utilized more relevant metrics in regards to surgery, this study is limited by the fact that the metrics of performance that it assessed to differentiate levels of expertise are not at present easy to teach.

Having demonstrated that artificial neural networks are able to classify surgical performances, the next step would be for one to try to understand the decision-making process that led to a specific classification. However, out of the 77 studies in the previous systematic reviews, all fail to quantify the relative importance of each performance metric to outline the underlying factors for a classification. Many methods have proven to aid in the quantification of the contributions of metrics towards classification.[76,77] In light of this information, a study that focused on the discectomy component of the simulated anterior cervical discectomy and fusion on the Sim-Ortho platform served as a proof of concept for the integration of artificial neural networks within a virtual reality spine procedure.[8]

Building upon previous work, the following manuscript provides a perspective on the utilization of an artificial neural network approach to competency-based training by highlighting the educational utilities of this methodology utilizing a spinal procedure as a model. Not only serving as a proof of concept, this study elaborates on how information can be extracted from the artificial neural network approach and aid in the understanding of expertise and it adds new insights into surgical resident learning of complex spinal procedures.

RATIONALE FOR THE STUDY

This study aims to outline the relevance and potential of an artificial neural network approach to competency-based training through its utilization on a simulated spinal procedure. Based on previous work conducted on the discectomy portion of a simulated anterior cervical discectomy and fusion, this study focuses on the vertebral osteophyte removal components to inform surgical trainees and instructors concerning the educational utilities of this methodology.

STUDY

Artificial Neural Network Approach to Competency-Based Training

Aiden Reich BEng, Nykan Mirchi MSc, Recai Yilmaz MD, Nicole Ledwos MSc, Vincent Bissonnette MD, MSc, Dan Huy Tran BSc, Swajan Paul MBChB, Alexander Winkler-Schwartz MD, Bekir Karlik PhD, Rolando F. Del Maestro MD, PhD

The proceeding work has been amplified with supplementary material incorporated within the methods and results sections to fulfill the requirements of this thesis.

Manuscript submitted to the BMJ STEL (June 29th, 2020).

ABSTRACT

Objective. The methodology of assessment and training of surgical skills is evolving to deal with the emergence of competency-based training. Machine learning algorithms have been employed to assess surgical expertise during virtual reality surgical performance. Some of these approaches fail to outline the underlying reasons for classification and to quantify the relative importance of each metric utilized to train the model. Artificial neural networks, a branch of artificial intelligence, can utilize newly generated metrics not only for assessment but can quantitate individual metric contribution and provide new insights into surgical expertise. This study aims to outline the educational utility of employing an artificial neural network in the assessment and quantitation of surgical expertise. A virtual reality vertebral osteophyte removal during a simulated surgical spine procedure is utilized as a model to outline this methodology.

Design. Participants performed a simulated anterior cervical discectomy and fusion on the Sim-Ortho virtual reality simulator platform. Data was retrieved from the osteophyte removal component of the scenario, which involved utilizing a simulated burr. The data was manipulated to initially generate 83 performance metrics spanning 3 categories (safety, efficiency, motion). The most relevant metrics were utilized to train and test the artificial neural network.

Setting. This study was carried out at the Neurosurgical Simulation and Artificial Intelligence Learning Centre at McGill University affiliated with the Montreal Neurological Institute and Hospital (Montreal, Canada).

Participants. Twenty-one participants performed a simulated anterior cervical discectomy and fusion. Participants were divided into 3 groups, including 9 post-residents, 5 senior and 7 junior residents.

Results. The artificial neural network model trained on the six most relevant metrics of performance, all involving safety, achieved 83.3% testing accuracy misclassifying only one participant.

Conclusions. This study outlines the potential utility of artificial neural networks which allows a deeper understanding of the composites of surgical expertise and may contribute to the paradigm shift towards competency-based surgical training.

INTRODUCTION

Artificial intelligence is defined as a computational method of making algorithm-based decisions through novel metrics that can be used for assessment and educational training.[5,6] Machine learning, a subset of artificial intelligence, includes both simple classifier algorithms and artificial neural networks. These artificial neural networks function as a series of interconnected nodes that communicate between each other and assign weights, which correspond to the sensitivity of the algorithm's decision-making process. These systems can recognize and uncover hidden patterns in large datasets to build their connections and define the weights associated with each performance metric.[8] These artificial neural networks can be exploited by competency-based training systems to develop new paradigms for surgical education.[9,10]

Simulation-based training provides a safe environment for individuals to acquire the necessary surgical skillset to demonstrate competency in high-risk scenarios. A multifaceted operative procedure such as an anterior cervical discectomy and fusion allows trainees to develop a wide diversity of skills including knowledge, competence and technical proficiency.[16] Virtual reality simulators generate vast datasets of quantitative information relating to psychomotor skills.[5,6,8,78] Our group has explored the value of virtual reality simulation utilizing performance metrics derived from the raw data that can be employed by surgical educators to train individuals to improved levels of performance.[4,8,61,79,80] Studies including those focused on the educational utility of artificial intelligence techniques during surgical simulation have demonstrated that the classification of individuals into various groups is a valid method for assessing performance.[5,6,17,19,20,69] However, these approaches fail to outline the underlying reasons for classification and to quantify the relative importance of each performance metric utilized to train the model. An appreciation of the role that artificial neural networks play in classifying performance, defining novel metrics and quantitating the relative importance of each performance metric will enhance our understanding of the composites of surgical expertise and our ability to teach these composites.

The objectives of this study are: 1) to introduce an artificial neural network methodology for assessing the composites of expertise in simulation-based training, 2) to outline the utility of artificial neural networks by utilizing a virtual reality spinal procedure model. This study reveals the potential of virtual reality simulation combined with artificial neural networks to understand not only the essential composites of expertise, the contributions of each composite, but the interplay of specific composites to expert surgical performance.

MATERIAL AND METHODS

Participants

Participant data utilized in this study was collected from previous studies at the Neurosurgical Simulation and Artificial Intelligence Learning Centre.[8,81] Twenty-seven individuals were initially recruited. The simulator utilized in this study is optimized for right-handed participants resulting in the removal of 3 left-handed participants. Two neurosurgeons and a fellow were excluded since their practices and training were not spine focused. The remaining 21 participants performed the simulated anterior discectomy and fusion simulation on the Sim-Ortho virtual reality platform. Participants were divided into 3 groups, 9 post-residents, (4 practicing spinal neurosurgeons and orthopedic surgeons and 5 spine fellows), 5 senior and 7 junior residents. Table 1 outlines participant demographic information. This study was approved by the McGill University Health Centre Research Ethics Board, Neurosciences-Psychiatry. All participants whose data was used to train the artificial neural network model signed an approved written consent form.

Virtual Reality Simulator

This study utilized the Sim-Ortho virtual reality simulator platform developed through the collaborative work of OSSimTech™ (Montreal, Canada) and the AO Foundation (Davos, Switzerland). The system offers realistic 3D visuals of the anterior cervical discectomy and fusion surgical procedure using a combination of 3D glasses and graphics based on a gaming platform (Figure 2). The Sim-Ortho platform provides haptic feedback allowing individuals to feel the different anatomical structures with the various interchangeable instruments (Figure 2A to D) and auditory feedback generated by the utilized instruments (Figure 2E).

Simulated Operative Scenario

A simulated anterior discectomy and fusion operative scenario was employed in this study.[8,81] This simulated surgical procedure is deconstructed into 4 components: vertebral disc annulus incision, discectomy, vertebral osteophyte removal and posterior longitudinal ligament excision. These components are outlined in Video 1. The high face and construct validity of the simulated vertebral osteophyte removal component, using a simulated burr,[81] suggested that this simulation was well suited to assess our specific objectives of assessing the utility of an artificial neural network model in the assessment and quantitation of surgical expertise in this simulated operative procedure. Participants did not have prior experience performing the ACDF simulation on the Sim-Ortho platform and were limited to using only the burr, at one power setting, to perform the vertebral osteophyte removal. No limit was placed on the time allocated to complete this component of the simulation.

Integration of Artificial Neural Network within Virtual Reality Simulation

Data Collection and Metrics Generation

The process of integrating an artificial neural network to a virtual reality simulator is illustrated in Figure 3. Raw data including information on the position and angle of the surgical instruments, the force applied to anatomical structures and volume of tissue removed was constantly recorded as participants performed the simulation procedure. Once the simulation was completed, the raw data was manipulated and transformed into performance metrics serving as the features for the model. Features are defined as elements of information input into the artificial neural network which, in the case of this study, are the performance metrics. These metrics are standards of reference by which performance, efficacy and progress can be assessed and are essential to be able to provide effective feedback. Initially, 83 metrics were generated utilizing expert opinion, data from publications focused on burr performance associated with surgical performance[21,78] and novel metrics derived from the raw data.

Metrics Selection

Large sets of metrics are at risk of containing redundant and irrelevant features. Therefore, filtering this input data allows us to retain only useful and relevant information which will improve the artificial neural network.[82] In this study, an initial 83 performance metrics were generated of which 28 were removed for comprising null values throughout participants' performances. Following, a forward sequential feature selection algorithm was applied through the *sequentialfs* function in Matlab R2019b. This feature selection algorithm has proven to be useful in artificial neural network applications and picks the most important input data by adding one input at a time to ensure the best possible performance of the system.[83] The inputs, in this case, were the performance metrics and, thus, the forward sequential feature selection was able to determine the optimal combination of performance metrics to be utilized with the artificial neural network. After the utilization of a forward sequential feature selection algorithm, 6 important metrics of performance remained.

Artificial Neural Network Model

The final 6 metrics along with the data from the 21 participants were split into 2 sets. First, 70% was used for training the artificial neural network (15 participants: 6 post-residents, 4 seniors, and 5 juniors). Then, the remaining 30% (6 participants: 3 post-residents, 1 senior, and 2 juniors) was employed to test the model. Therefore, data from the training set was used to train the neural network in a supervised manner and the data from the testing set was used to test the model.

Model Optimization

The artificial neural network is comprised of 6 input layer nodes for each performance metric, 36 hidden layer nodes and 3 output layer nodes for each performance group as shown in Figure 4. The training algorithm utilized in this study was the Bayesian Regularization Backpropagation algorithm (*trainbr* function in Matlab R2019b). Due to its built-in regularization, this algorithm aids in avoiding an overfit model, which artificial neural networks are prone to.[84,85] The model was trained over 10 iterations, through a procedure known as early stopping, to minimize overfitting issues.[86] The *tansig* transfer function from Matlab R2019b was applied to the hidden and output layers. This method was chosen for its higher performance results previously demonstrated.[87] Finally, only two parameters were modified from their default values. The Marquardt adjustment parameter, also known as mu, and the decrease ratio of mu were set respectively at 2 and 0.95 for optimal results. These hyperparameters have been defined and optimized through random trial and error which has been proven to be an efficient method.[88] However, there is uncertainty in this process of random trial and error of hyperparameters as it may result in better artificial neural network models that are yet to be discovered for the exact same dataset.[89]

Relative Importance of Performance Metrics

Artificial neural networks are sometimes regarded as "black boxes" because the decision-making process of the algorithm remains hidden and requires supplementary methods to reveal insight into the importance of the metrics of performance utilized to train the model.[77,85,90] In this study, our group gains access to the valuable information hidden within the artificial neural network, also known as the relative importance of each performance metric for each group (post-resident, senior, and junior residents), through the utilization of the Connection Weights Algorithm, which is a method previously employed for analyzing the discectomy component of the simulated anterior cervical discectomy and fusion.[8,90] Briefly, the Connection Weights Algorithm takes advantage of the interconnected weighting system of the artificial neural network as illustrated in Figure 4 to rank each performance metric for each participant group according to their importance, thus, determining the importance of individual metrics on the network's decision-making process.[8]

While still referring to Figure 4, one will notice 216 interconnected weights ($w_{x,y}$) between the layer inputting the performance metrics and the hidden layer of the artificial neural network model as well as 108 interconnected weights ($v_{y,z}$) between the hidden layer and the output layer. Following the extraction of these weights from the artificial neural network, the Connection Weights algorithm is applied by summing the product of each input to hidden layer weight ($w_{x,y}$) and the hidden to output layer weight ($v_{y,z}$). This algorithm (Equation 1)[8,90] calculates the Connection Weight Product (CWP) which is a value for each metric of each performance group allowing to understand the significance of each metric.

$$CWP_x = \sum_{y=1}^{m} w_{x,y} \cdot v_{y,z}$$

Equation 1: Connection Weight Product

The relative importance of each performance metrics within a specific expertise group is derived from the Connection Weight Products. Furthermore, interpretation of the Connection Weight Product requires an understanding of the value's sign and magnitude. The Connection Weight Product specifies if the value of a specific metrics should be bigger or smaller in order to favor a specific classification. For example, Figure 6 shows that post-residents have a very negative Connection Weight Product, senior residents have a slightly negative Connection Weight Product and junior residents have a very positive Connection Weight Product for the force applied on the C5 vertebra metrics. Hence, a lower force applied on that vertebra implies a higher probability of being classified in a group of higher expertise. Both the post-resident group and the senior resident group had negative values for the Connection Weight Product of that metrics. However, a performance generating a more negative value for that metric would increase the likelihood of classification in the post-resident group.

RESULTS

Performance Metrics

Eighty-three performance metrics were created for every participant for the osteophyte removal component of the anterior cervical discectomy and fusion scenario. These were divided into three categories encompassing safety, efficiency and motion elements of simulation performance. Of the 83 metrics, 28 were removed for comprising zero values throughout participants' performances reducing the number to 55. Furthermore, a feature selection algorithm was employed reducing the final number of relevant metrics used to train the artificial neural network to 6. These metrics all originate from the safety category (Table 2) and outline the important role of safety in the osteophyte removal component of the anterior cervical discectomy and fusion procedure. These results are consistent with the findings from previous virtual reality spine studies and those from a virtual reality tumor resection model.[8,80]

Classification of Surgical Performance

Confusion matrices are tables which are commonly employed to outline the performance of a classification model (or "classifier") on datasets for which the true values are known. These matrices allow one to visualize the performance of the artificial neural network used. The confusion matrices of our data, separated into a training set (15 participants) and a testing set (6 participants), are outlined in Figure 5. Following the optimization of the artificial neural network's parameters and having the algorithm train over 10 iterations, the previously mentioned sets respectively achieved overall accuracies of 80% and 83.3%. The training set misclassified 3 participants (a junior resident, a senior resident and a post-resident) while the testing set only misclassified one participant (a senior resident).

Importance of Performance Metrics

The sensitivity of each performance metric for each group can vary allowing for different influences on the decision-making process of the artificial neural network. One can access the weights assigned to each metric within the artificial neural network by using the Connection Weights Algorithm, shown to be the most appropriate method of quantifying performance metric importance.[77,90] This algorithm takes advantage of the weights between layers of a neural network and forms the basis for calculating the importance of each performance metric for each participant group. The 6 performance metrics for post-residents, senior residents and junior residents along with their corresponding Connection Weight Products (CWP) and relative importance are represented respectively in Tables 3, 4 and 5. The magnitude of their Connection Weight Products ranged from +1.5 to -1.07 (Figure 6). Different metrics are important for each individual group and these values are represented in Tables 3 to 5. A number of different patterns are seen when one reviews the Connection Weight Products data and our group has reported these patterns previously.[8] The two most identifiable regularities outlined are a continuous learning pattern and a discontinuous learning pattern for technical skills. The former involves a progressive and sequential change in which learning occurs in incremental stages or follows a certain prescribed order when comparing performance as residents progress from junior to senior and then to the post-resident level of technical skills. The Connection Weight Products associated with average force applied on the C5 vertebra (post-residents CWP = -1.07; senior residents CWP = -0.08; junior residents CWP = +0.68) illustrates this continuous learning pattern. These values signify that there is an increased likelihood of being classified as a post-resident compared to a senior and junior resident if the participant applied less average force on the C5 vertebrae during the procedure and an increased likelihood of being classified as a senior resident compared to a junior resident if the average force being applied on the C5 vertebrae is less. The second pattern is associated with a variable and nonsequential modification of psychomotor skills learning as residents progress from junior to the post-resident level of performance while going through an intermediate senior resident level that performs incongruously. Examples of this discontinuous learning pattern are seen when a participant is much more likely to be classified as a senior resident than a post-resident or junior resident if that participant has increased numbers of contacts of an active burr on the spinal dura (post-residents CWP = -0.68; senior residents CWP = +0.32; junior residents CWP = -0.92) and if the participant has applied increased average force on the left posterior longitudinal ligament (post-residents CWP = -0.71; senior residents CWP = +0.94; junior residents CWP = -0.15). Interestingly, another pattern seen was that a participant was more likely to be classified as a post-resident if high average force was applied to the right posterior longitudinal ligament (post-residents CWP = +1.5; senior residents CWP = -1.07; junior residents CWP = -0.64). The percentage of relative importance for the 6 metrics defined by the artificial neural network provides new perspectives on critical aspects of expert performance during removal of vertebral osteophytes when utilizing a burr (Tables 3, 4 and 5).

DISCUSSION

Summary

The first objective of this study was to introduce artificial neural networks methodology for assessing expertise in simulation-based training. Using the vertebral osteophyte removal component of the anterior cervical discectomy and fusion virtual reality scenario developed on the Sim-Ortho virtual reality simulation platform, we determined 6 metrics which the artificial neural network utilized for assessing surgical performance. In the second objective, the utility of this artificial neural network was outlined through the determination of the individual contributions of these 6 metrics to the composites of surgical performance utilizing a virtual reality spinal procedure model. These results have improved our understanding of some of the critical composites of surgical expertise involved in this simulated procedure. The authors believe that an artificial neural network, as the one utilized in this study, could be employed in any medical or surgical procedure in which large datasets are available allowing to assist with competency-based learning.

Utility of Artificial Neural Networks

Our group has explored the utility of virtual reality simulators and machine learning classifiers to outline objective bimanual psychomotor assessments metrics in neurosurgical and spine procedures.[5,6] The high-fidelity simulators with haptic feedback retain instrument-tissue interactions and pathological realism allowing them to complement the present surgical training models. Learners have unrestricted freedom to practice in safe-to-fail environments without the limitations imposed by lack of supervision, the presence of patients or the use of OR resources.[8] Virtual reality simulators can deconstruct complex surgical procedures into manageable steps for learners to master allowing learners to bypass steps in which they are competent and focus on specific steps that require improvement.[41,42,91]

The Sim-Ortho anterior cervical discectomy and fusion simulation was deliberately designed in 4 specific components so that each component could be independently assessed and used for training. Utilizing the same participant data set and the artificial neural network employed in this study, the metrics and Connection Weight Products for the discectomy component of the procedure have also been published.[8] This allows a further analysis of the utility of the same artificial neural network in two very different components of the same simulation procedure (Table 6). Although the discectomy and the osteophyte drilling both take place within the disc cavity, there are major technical procedural differences between these two components of the anterior cervical discectomy and fusion. These are related to the complexity of each procedure. The discectomy involved the participants choosing between three possible instruments as opposed to the osteophyte removal where only the burr could be used. This difference provides insight in what the results of using the artificial neural network approach may be. The outlined new metrics and the role of individual metrics in determining composites of expertise are expected to be different. First, for complex operative procedures, this methodology may outline both increased numbers of metrics and more complicated metrics. In our previous study, 16 metrics, involving safety, efficiency, motion and cognition, were selected for the discectomy while only 6 safety metrics were selected for osteophyte removal. Second, the predominance in safety metrics for both components of the simulation suggests that, despite the differences in these two procedures, our methodology regards operative safety as a critical composite of surgical expertise. These results are consistent with the previous model suggested by our group.[80] Third, with the limited number of participants in the training set and the testing sets, lower classification accuracies such as 80% and 83.3% may be expected.[92] An artificial neural network model can be more robust by recruiting more individuals from various institutions.[86,93] Expanding both of the training and testing sets should help our model have a more representative picture of surgical expertise. Fourth, the Connection Weight Products also provide helpful insight to educators tasked with prioritizing specific metrics during training for both procedures. For the 6 safety metrics identified for osteophyte removal, the relative importance percentage of these metrics, as seen in Tables 3, 4 and 5, ranged from 33.31% to 5.36 % for the post-resident group, 37.97% to 2.88% for the senior resident group and 27.54% to 4.41% for the junior resident group. These percentages allow educators to rank the performance metrics according to which contributes more towards the classification of a specific expertise level. For example, a metric such as the average force applied on the C5 vertebra has more influence on classifying a participant as a post-resident or junior resident since that metric is ranked as second most important for both of these groups with respective relative importance values of 23.77% and 20.31%. However, this metric has far less impact on the classification of a participant in the senior resident group because it has a low relative importance (2.88%) and is ranked sixth for that group. Furthermore, the range in numbers are consistent with percentages seen in the discectomy data suggesting that surgical educators need to be aware of these differences in relative importance of metrics to maximize the surgical performance of trainees.[65] Considering the difference between the concepts of competence and expertise,[65] the improvement of the psychomotor skills involved with top tier metrics (high Connection Weight Products magnitude or relative importance) may benefit the overall performance of the trainee and aid in achieving competency. However, other metrics of lesser importance should not be neglected as they may also be important factors in attaining expertise.

Educational Patterns of Artificial Neural Networks

The ability of artificial neural networks to rank the significance of a specific metric in expertise assessment during virtual reality procedures allows surgical educators to investigate new concepts of teaching. Should metrics that predominately contribute to expertise take precedence in surgical training paradigms? Should surgical education programs focus on training junior residents to the senior resident or that of the post-resident levels? When looking at the average force applied to the spinal dura metric in junior residency training this is not an issue since senior resident or post-resident values are equivalent (Table 5). However, our results have identified two other patterns; the continuous learning pattern and the discontinuous learning pattern of technical skills. In the first pattern discussed previously, there is a progressive change seen when moving from the junior resident to the post-resident group. This pattern was observed with the incrementally decreasing average force applied on the C5 vertebrae metric when comparing the three groups. This continuous learning pattern was also prominent in the discectomy component of the procedure.[8] Several studies from our group have demonstrated that expertise is associated with decreasing force application and the findings with this metric are consistent with these results.[6,8,94,95] The question does arise as to whether junior residents should be trained to the senior or post-resident level of force application in this scenario to accelerate the acquisition of surgical expertise? Studies focused on this question are needed. The discontinuous learning pattern outlined involves the senior resident group performance being a nonsequential outlier compared to that of the junior and post-resident groups. This pattern was seen when the number of contacts of the active burr with the spinal dura and the average force applied to the left posterior longitudinal ligament metrics were analyzed. It may be reasonable to speculate that junior residents are hesitant to approach the spinal dura and the left posterior longitudinal ligament with the active burr, while senior residents are more aggressive in this activity. The post-resident group may have modulated this behavior with experience and their predominant focus on safety. Considering this pattern, should junior residents be trained to the senior resident level of performance which is potentially associated with increased risk? It would appear reasonable to also develop studies to address this issue. Both the continuous and discontinuous learning patterns are illustrated in Figure 7. A number of other patterns were found which can be seen in Figure 6. An unusual pattern was that of the average force on the right posterior longitudinal ligament where the post-residents applied more force than both senior and junior residents. This may be a result of instrument positioning and hand ergonomics in relation to the right posterior longitudinal ligament. One can speculate that post-residents were more aggressive when trying to completely remove osteophytes near the virtual patient's right posterior longitudinal ligament to decompress the adjacent cervical nerve. This action requires more wrist flexion and the muscle activation associated with wrist flexion may lead to greater forces applied.[80] If the Sim-Ortho platform could be optimized for left hand dominant individuals, one can hypothesize that this specific unusual pattern would be present for the force applied on the left posterior longitudinal ligament rather than the right side. Observing this reversal of patterns from right-handed to left-handed individuals would provide further insight concerning this unusual pattern and support the hand ergonomics hypothesis. Further studies involving not only a larger number of participants, but also left-handed participants will be required to investigate these less easily understood patterns. All these patterns allow medical educators and trainees to begin to better comprehend the composites associated with expertise and adjust the focus of the training paradigms.

Surgical Education Platform powered by Artificial Neural Networks

Adequate and reliable assessment tools that can effectively evaluate competency or expertise are crucial for competency-based training in medical education.[96] Appropriate technologies and the implementations of validated virtual reality training scenarios can help to determine a trainee's baseline skill level and provide them with critical information to ensure quality feedback.[97] Simple machine learning classifiers such as support vector machines, k-nearest neighbors, linear discrimination analysis and decision trees can be useful for the classification of skill level.[5,6,69] Artificial neural networks have the potential to help focus the training of residents into specific metrics which may be important in the development of expertise.

Simulation-based training may be automated by providing post-operative feedback to trainees using the information outputted by the artificial neural network. This may decrease the reliance on expert instructors to give feedback or modulate the way these educators provide feedback. This study amongst others have demonstrated that quality feedback can be provided by an artificial neural network.[8,98] However, questions remain concerning the best methods of optimizing these automatic feedback systems. Would an automated instructor be more effective than actual expert surgical trainers? What would be the optimal method for the artificial neural network-integrated feedback systems to present trainee information? Threshold expert benchmarks may be one methodology to present each important performance metric.[99] Clear goals set by educators through benchmarking aid in the creation of efficient competency-based curricula for trainees.[44] Video feedback is another method that has been shown to convey necessary information for medical trainees to acquire important surgical skills.[100] A comparison of various personal feedback model systems such as Technical Abilities Customized Training (TACT) and combinations of feedback technologies can aid in our understanding of their effectiveness and impact on trainees' learning performance.[64,101]

This study demonstrates how artificial neural networks can be utilized to develop training curricula molded to the specific needs of trainees. With the added benefits of a personalized and automated feedback system, a virtual operative assistant has been developed by our group in the Neurosurgical Simulation and Artificial Intelligence Learning Centre.[98] This virtual operative assistant utilizes a machine learning algorithm, support vector machine, to assess an individual's surgical performance during a simulated operative scenario and tailors its feedback based on performance metrics. This allows a personalized training experience with virtual reality simulation.[98] This type of virtual operative assistant could also be developed utilizing data obtained from artificial neural networks. Studies are planned to compare trainee performance on virtual reality simulators receiving personalized virtual operative assistant feedback to that of individuals receiving feedback from an expert instructor.

LIMITATIONS

Limitations of Artificial Neural Networks in Education

The artificial neural networks methods utilized in this study followed best practices to utilize machine learning algorithms to assess surgical expertise in simulation previously established by our group.[4] However, artificial neural networks have limitations regarding their use in surgical education and training. Employing artificial neural networks to classify individuals and assess performances may be difficult due to overfitting. There are a number of methods for resolving this issue.[85] The artificial neural network in this study benefited from regularization, limiting the number of hidden nodes as well as limiting the amount of training iterations. However, due to the process of optimizing the model's hyperparameters by random trial and error, a better artificial neural network model for the same data may remain undiscovered.[89] Training curricula differ depending on the surgical specialty and institution causing variable levels of resident expertise.[102,103] If an artificial neural network is trained with the data of individuals from one educational site, the classification may not perform well for individuals from other institutions. Hence, our model may be limited by the fact that all training data was obtained from surgeons and resident from a single institution which may not generalize well to other institutions. A more generalizable model would require data from a large number of individuals from multiple institutions.

Limitations of the Study

The incorporation of an advanced gaming engine into the Sim-Ortho virtual reality surgical simulator platform improved anatomical structure realism and the interactions of simulated instruments with these structures. However, this platform does not reproduce the complex ever-changing environment of a patient undergoing an anterior cervical discectomy and fusion procedure. A number of limitations need to be considered when assessing this study. First, to limit the variables associated with the task, only one specific burr, at one setting, was utilized in this simulation which is not consistent with the multiple burr options for surgeons completing a real

ACDF. Second, hand ergonomics have proved to be an important factor in simulated operative procedures.[80] The Sim-Ortho platform was designed for right-handed use limiting its usefulness in assessing and quantitating the composites of bimanual performance. Modifications of the Sim-Ortho platform are warranted to allow for a more comprehensive understanding of bimanual expertise. Third, the study involved a small sample size of a priori-defined participant groups from one institution which has previously been identified as a common limitation for artificial neural networks. Assessing the accuracy and generalizability of the utilization of artificial neural networks in competency-based training will require large prospective multi-institutional studies.

CONCLUSION

This study reveals the potential of virtual reality simulation combined with artificial neural networks to outline the important composites of surgical expertise, the contributions of each composite, and the interplay of specific composites which result in expert surgical performance. Our results demonstrate the educational potential of integrating artificial neural networks with virtual reality surgical simulation and automated feedback systems to develop personalized training curricula in competency-based education.

THESIS CONCLUSION

Summary

The present thesis demonstrates the educational utility of artificial neural networks using a simulated spine procedure as a model. The study objectives of introducing an artificial neural network methodology for assessing the composites of expertise in simulation-based training and outlining the utility of artificial neural networks by utilizing a virtual reality spinal procedure model were achieved. An initial 83 performance metrics involving safety, motion and efficiency were generated for the osteophyte removal portion of the anterior cervical discectomy and fusion. Following manipulation and selection of the most relevant 6 metrics, an artificial neural network was created and trained using participants' performances to classify participants into groups of varying expertise levels (post-residents, senior residents and junior residents). This model was then tested on the data of another set of participants, only misclassifying one individual, achieving an 83.3 % accuracy. The artificial neural network, often regarded as a black-box, was opened and analyzed to outline the relative importance of each performance metric and understand the composites of expertise as well as the interplay of these composites involved with training to higher levels of performance.

It is difficult to generalize this artificial neural network due to the small sample size. However, the model demonstrates its potential utility as an educational tool when combined with virtual reality simulation. First, the artificial neural network's ability of classification serves as a means of objective assessment of participant surgical performance. This assessment provides trainees with a baseline performance evaluation and allows them to track their progress as they train on the simulator and develop new psychomotor and technical skills. Second, if the artificial neural network presents the relative importance of each performance metric, trainees and surgical educators can acquire a better understanding of the composites of expertise. This information will allow surgical trainees and educators to adapt their training paradigms to focus either on the major skills necessary to be competent or other metrics skills which may also be essential for mastering performance. Third, the artificial neural network reveals hidden patterns related to the development of specific technical skills. Two noticeable patterns identified as the continuous learning pattern (progressive and incremental learning) and the discontinuous learning pattern (variable non sequential learning) illustrate very different methods of acquiring technical skills. These patterns can aid trainees and educators to better comprehend the composites of expertise and adapt the training accordingly. Hence, the artificial neural network approach can provide not only an objective assessment of surgical performance but also provide formative information related to how to structure a training curriculum specific to the needs of the surgical trainee.

Aside from the educational utility, one must also consider the social ramifications of the artificial neural network approach to competency-based training. There is concern that artificial intelligence may replace human involvement. Our group does not advocate the replacement of present educational paradigms by automated systems.[6,8,98] Human interaction is vital to learning. The integration and the wider availability of intelligent tutoring systems may complement present curricula. Intelligent tutoring systems can utilize different simulation platforms, increasing the surgical educators' armamentarium to help learners achieve mastery levels of surgical performance. The methodology suggested in this study is an incremental technological change that will continue to evolve and be integrated into the current apprenticeship model. Expert educators may not always be available. Therefore, intelligent tutoring systems based on artificial neural networks can augment the learning experience of trainees. This artificial neural network approach is an educational tool designed to provide an objective assessment and other formative information related to performance without replacing surgical educators and expert feedback.

This study outlines that the artificial neural network approach to competency-based training provides novel perspectives regarding surgical performance and further insights into the understanding of expertise and its composites.

Future Directions

This study suggests new questions related to the use of the learning patterns outlined suggesting possible modifications to current surgical training paradigms. Also, questions related to the generalizability of the artificial neural network model remain unanswered.

First, having identified and defined the continuous learning pattern and the discontinuous learning pattern involved in our investigation, further studies will be required to gain a better understanding of the significance of these two patterns. Investigating these patterns and the information they provide can help determine if a junior resident should train to a post-resident level of expertise either by only developing the surgical skillset of a post-resident or by needing to first develop an intermediate senior level of performance. Providing surgical educators and trainees with this information can improve current training curricula and competency-based training programs with the potential for more efficient learning and acquisition of psychomotor and technical skills which may result in safer patient outcomes.

Second, due to the small sample size utilized in this study, it is not possible to assess the generalizability of the artificial neural network model. The participants from this study all came from the same institution and their surgical performances data may not translate to individuals of other institutions. A future study will be required to assess the generalizability of the this trained artificial neural network by including new participants from multiple institutions. If the present model does not achieve good classification accuracy, this new data combined with the data from this study may be utilized to train a new artificial neural network which will be more robust in terms of generalizability.

Current surgical training relies on the feedback of expert educators. Other means of providing automated feedback involving intelligent tutoring systems are under development. The virtual operative assistant developed in the Neurosurgical Simulation and Artificial Intelligence Learning Centre is one such example.[98] Studies involving intelligent tutoring systems, such as the virtual operative assistant, are necessary to design optimal feedback methods to improve trainee learning curves. Investigations designed to assess the role of combining surgical educators with intelligent tutoring systems will provide further insights into how to optimize future surgical education paradigms.

REFERENCES

1. Buchlak QD, Esmaili N, Leveque J-C, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurgical review.* 2019:1-19 DOI: https://doi.org/10.1007/s10143-019-01163-8.
2. Fourcade A, Khonsari R. Deep learning in medical image analysis: A third eye for doctors. *Journal of stomatology, oral and maxillofacial surgery.* 2019;120(4):279-288 DOI: https://doi.org/10.1016/j.jormas.2019.06.002.
3. Lisboa PJ, Taktak AF. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural networks.* 2006;19(4):408-415 DOI: https://doi.org/10.1016/j.neunet.2005.10.007.
4. Winkler-Schwartz A, Bissonnette V, Mirchi N, et al. Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. *Journal of surgical education.* 2019;76(6):1681-1690 DOI: https://doi.org/10.1016/j.jsurg.2019.05.015.
5. Winkler-Schwartz A, Yilmaz R, Mirchi N, et al. Machine learning identification of surgical and operative factors associated with surgical expertise in virtual reality simulation. *JAMA network open.* 2019;2(8):e198363-e198363 DOI: https://doi.org/10.1001/jamanetworkopen.2019.8363.
6. Bissonnette V, Mirchi N, Ledwos N, Alsidieri G, Winkler-Schwartz A, Del Maestro RF. Artificial intelligence distinguishes surgical training levels in a virtual reality spinal task. *JBJS.* 2019;101(23):e127 DOI: https://doi.org/10.2106/JBJS.18.01197.
7. Ramesh A, Kambhampati C, Monson JR, Drew P. Artificial intelligence in medicine. *Annals of the Royal College of Surgeons of England.* 2004;86(5):334 DOI: https://doi.org/10.1308/147870804290.
8. Mirchi N, Bissonnette V, Ledwos N, et al. Artificial Neural Networks to Assess Virtual Reality Anterior Cervical Discectomy Performance. *Operative Neurosurgery.* 2019 DOI: https://doi.org/10.1093/ons/opz359.
9. Delorme S, Laroche D, DiRaddo R, Del Maestro RF. NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training. *Operative Neurosurgery.* 2012;71(suppl_1):ons32-ons42 DOI: https://doi.org/10.1227/NEU.0b013e318249c744.
10. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Medical teacher.* 2010;32(8):638-645 DOI: https://doi.org/10.3109/0142159X.2010.501190.
11. Satava RM. Virtual reality surgical simulator. *Surgical endoscopy.* 1993;7(3):203-205 DOI: https://doi.org/10.1007/BF00594110.
12. De Luca G, Choudhury N, Pagiatakis C, Laroche D. A Multi-procedural Virtual Reality Simulator for Orthopaedic Training. Paper presented at: International Conference on Human-Computer Interaction2019.

13. Wang H, Jiang S, Wu J. A Virtual Reality Based Simulator for Training Surgical Skills in Procedure of Catheter Ablation. Paper presented at: 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)2018 DOI: https://doi.org/10.1109/AIVR.2018.00057.
14. Wilson M, Middlebrook A, Sutton C, Stone R, McCloy R. MIST VR: a virtual reality trainer for laparoscopic surgery assesses performance. *Annals of the Royal College of Surgeons of England.* 1997;79(6):403.
15. Westwood JD. A networked haptic virtual environment for teaching temporal bone surgery. *Medicine Meets Virtual Reality 13: The Magical Next Becomes the Medical Now.* 2005;111:204.
16. Ray WZ, Ganju A, Harrop JS, Hoh DJ. Developing an anterior cervical diskectomy and fusion simulator for neurosurgical resident training. *Neurosurgery.* 2013;73(suppl_1):S100-S106 DOI: https://doi.org/10.1227/NEU.0000000000000088.
17. Watson RA. Use of a machine learning algorithm to classify expertise: analysis of hand motion patterns during a simulated surgical task. *Academic Medicine.* 2014;89(8):1163-1167 DOI: https://doi.org/10.1097/ACM.0000000000000316.
18. Vedula SS, Ishii M, Hager GD. Objective assessment of surgical technical skill and competency in the operating room. *Annual review of biomedical engineering.* 2017;19:301-325 DOI: https://doi.org/10.1146/annurev-bioeng-071516-044435.
19. Richstone L, Schwartz MJ, Seideman C, Cadeddu J, Marshall S, Kavoussi LR. Eye metrics as an objective assessment of surgical skill. *Annals of surgery.* 2010;252(1):177-182 DOI: https://doi.org/10.1097/SLA.0b013e3181e464fb.
20. Sewell C. *Automatic performance evaluation in surgical simulation* [PhD Thesis]. Stanford, CA: Stanford University; 2007 DOI: https://doi.org/10.5555/1293123.
21. Sewell C, Morris D, Blevins NH, et al. Providing metrics and performance feedback in a surgical simulator. *Computer Aided Surgery.* 2008;13(2):63-81 DOI: https://doi.org/10.3109/10929080801957712.
22. Harrop J, Rezai AR, Hoh DJ, Ghobrial GM, Sharan A. Neurosurgical training with a novel cervical spine simulator: posterior foraminotomy and laminectomy. *Neurosurgery.* 2013;73(suppl_1):S94-S99 DOI: https://doi.org/10.1227/NEU.0000000000000103.
23. Atesok K, Mabrey JD, Jazrawi LM, Egol KA. Surgical simulation in orthopaedic skills training. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons.* 2012;20(7):410-422 DOI: https://doi.org/10.5435/JAAOS-20-06-410.
24. Riles TS. Surgical training: the past, the present, and the future. *Annals of Vascular Surgery.* 2005;19(2):140-141 DOI: https://doi.org/10.1007/s10016-004-0186-3.
25. Pedowitz RA, Marsh LJ. Motor skills training in orthopaedic surgery: a paradigm shift toward a simulation-based educational curriculum. *JAAOS-*

Journal of the American Academy of Orthopaedic Surgeons. 2012;20(7):407-409 DOI: https://doi.org/10.5435/JAAOS-20-07-407.
26. Powell DE, Carraccio C. Toward competency-based medical education. *N Engl J Med.* 2018;378(1):3-5 DOI: https://doi.org/10.1056/NEJMp1712900.
27. Bhatti NI, Cummings CW. Competency in surgical residency training: defining and raising the bar. *Academic Medicine.* 2007;82(6):569-573 DOI: https://doi.org/10.1097/ACM.0b013e3180555bfb.
28. Kotsis SV, Chung KC. Application of see one, do one, teach one concept in surgical training. *Plastic and reconstructive surgery.* 2013;131(5):1194 DOI: https://doi.org/10.1097/PRS.0b013e318287a0b3.
29. Hamdorf J, Hall JC. Acquiring surgical skills. *British Journal of Surgery.* 2000;87(1):28-37 DOI: https://doi.org/10.1046/j.1365-2168.2000.01327.x.
30. Lee AG. Graduate medical education in ophthalmology: moving from the apprenticeship model to competency-based education. *Archives of ophthalmology.* 2008;126(9):1290-1291 DOI: https://doi.org/10.1001/archophthalmol.2008.3.
31. Steinbrook R. The debate over residents' work hours. In: Mass Medical Soc; 2002 DOI: https://doi.org/10.1056/NEJMhpr022383.
32. Gelfand DV, Podnos YD, Carmichael JC, Saltzman DJ, Wilson SE, Williams RA. Effect of the 80-hour workweek on resident burnout. *Archives of surgery.* 2004;139(9):933-940 DOI: https://doi.org/10.1001/archsurg.139.9.933.
33. Martini S, Arfken CL, Balon R. Comparison of burnout among medical residents before and after the implementation of work hours limits. *Academic Psychiatry.* 2006;30(4):352-355 DOI: https://doi.org/10.1176/appi.ap.30.4.352.
34. Thomas NK. Resident burnout. *Jama.* 2004;292(23):2880-2889 DOI: https://doi.org/10.1001/jama.292.23.2880.
35. Okuda Y, Bryson EO, DeMaria Jr S, et al. The utility of simulation in medical education: what is the evidence? *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine: A Journal of Translational and Personalized Medicine.* 2009;76(4):330-343 DOI: https://doi.org/10.1002/msj.20127.
36. Satava RM. Surgical education and surgical simulation. *World journal of surgery.* 2001;25(11):1484-1489 DOI: https://doi.org/10.1007/s00268-001-0134-0.
37. Hammoud MM, Nuthalapaty FS, Goepfert AR, et al. To the point: medical education review of the role of simulators in surgical training. *American journal of obstetrics and gynecology.* 2008;199(4):338-343 DOI: https://doi.org/10.1016/j.ajog.2008.05.002.
38. Schout BM, Hendrikx A, Scheele F, Bemelmans BL, Scherpbier A. Validation and implementation of surgical simulators: a critical review of present, past, and future. *Surgical endoscopy.* 2010;24(3):536-546 DOI: https://doi.org/10.1007/s00464-009-0634-9.
39. Alotaibi F, Al Zhrani G, Bajunaid K, Winkler-Schwartz A, Azarnoush H. Assessing Neurosurgical Psychomotor Performance: Role of Virtual Reality Simulators, Current and Future Potential. SOJ Neurol 2 (1), 1-7. *Assessing*

Neurosurgical Psychomotor Performance: Role of Virtual Reality Simulators, Current and Future Potential. 2015 DOI: https://doi.org/10.15226/2374-6858/2/1/00116.

40. Pfandler M, Lazarovici M, Stefan P, Wucherer P, Weigl M. Virtual reality-based simulators for spine surgery: a systematic review. *The Spine Journal.* 2017;17(9):1352-1363 DOI: https://doi.org/10.1016/j.spinee.2017.05.016.
41. Bartlett J, Lawrence J, Stewart M, Nakano N, Khanduja V. Does virtual reality simulation have a role in training trauma and orthopaedic surgeons? *Bone Joint J.* 2018;100(5):559-565 DOI: https://doi.org/10.1302/0301-620X.100B5.BJJ-2017-1439.
42. Malone HR, Syed ON, Downes MS, D'Ambrosio AL, Quest DO, Kaiser MG. Simulation in neurosurgery: a review of computer-based simulation environments and their surgical applications. *Neurosurgery.* 2010;67(4):1105-1116 DOI: https://doi.org/10.1227/NEU.0b013e3181ee46d0.
43. Palter VN, Grantcharov TP. Simulation in surgical education. *Cmaj.* 2010;182(11):1191-1196 DOI: https://doi.org/10.1503/cmaj.091743.
44. Raison N, Ahmed K, Fossati N, et al. Competency based training in robotic surgery: benchmark scores for virtual reality robotic simulation. *Bju international.* 2017;119(5):804-811 DOI: https://doi.org/10.1111/bju.13710.
45. Smith ML. Simulation and education in gynecologic surgery. *Obstetrics and Gynecology Clinics.* 2011;38(4):733-740 DOI: https://doi.org/10.1016/j.ogc.2011.09.007.
46. Munshi F, Lababidi H, Alyousef S. Low-versus high-fidelity simulations in teaching and assessing clinical skills. *Journal of Taibah University Medical Sciences.* 2015;10(1):12-15 DOI: https://doi.org/10.1016/j.jtumed.2015.01.008.
47. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery.* 2004;135(1):21-27 DOI: https://doi.org/10.1016/s0039-6060(03)00156-9.
48. Patel HR, Patel BP. Virtual reality surgical simulation in training. *Expert review of anticancer therapy.* 2012;12(4):417-420 DOI: https://doi.org/10.1586/era.12.23.
49. Blyth P, Anderson IA, Stott NS. Virtual reality simulators in orthopedic surgery: What do the surgeons think? *Journal of Surgical Research.* 2006;131(1):133-139 DOI: https://doi.org/10.1016/j.jss.2005.08.027.
50. Wang P, Becker AA, Jones IA, et al. A virtual reality surgery simulation of cutting and retraction in neurosurgery with force-feedback. *Computer methods and programs in biomedicine.* 2006;84(1):11-18 DOI: https://doi.org/10.1016/j.cmpb.2006.07.006.
51. Brunozzi D, McGuire LS, Alaraj A. NeuroVR™ Simulator in Neurosurgical Training. In: *Comprehensive Healthcare Simulation: Neurosurgery.* Springer; 2018:211-218.

52. Mabrey JD, Reinig KD, Cannon WD. Virtual reality in orthopaedics: is it a reality? *Clinical Orthopaedics and Related Research®.* 2010;468(10):2586-2591 DOI: https://doi.org/10.1007/s11999-010-1426-1.
53. Tsai M-D, Hsieh M-S, Jou S-B. Virtual reality orthopedic surgery simulator. *Computers in biology and medicine.* 2001;31(5):333-351 DOI: https://doi.org/10.1016/s0010-4825(01)00014-2.
54. Morris D, Tan H, Barbagli F, Chang T, Salisbury K. Haptic feedback enhances force skill learning. Paper presented at: Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)2007 DOI: https://doi.org/10.1109/WHC.2007.65.
55. Panait L, Akkary E, Bell RL, Roberts KE, Dudrick SJ, Duffy AJ. The role of haptic feedback in laparoscopic simulation training. *Journal of Surgical Research.* 2009;156(2):312-316 DOI: https://doi.org/10.1016/j.jss.2009.04.018.
56. Halic T, Kockara S, Bayrak C, Rowe R. Mixed reality simulation of rasping procedure in artificial cervical disc replacement (ACDR) surgery. Paper presented at: BMC bioinformatics2010 DOI: https://doi.org/10.1186/1471-2105-11-S6-S11.
57. Bova FJ, Rajon DA, Friedman WA, et al. Mixed-reality simulation for neurosurgical procedures. *Neurosurgery.* 2013;73(suppl_1):S138-S145 DOI: https://doi.org/10.1227/NEU.0000000000000113.
58. Lacey G, Ryan D, Cassidy D, Young D. Mixed-reality simulation of minimally invasive surgeries. *Ieee Multimedia.* 2007;14(4):76-87 DOI: https://doi.org/10.1109/MMUL.2007.79
59. Sharma DK, Khera A, Singh D. Using Artificial Intelligence to Bring Accurate Real-Time Simulation to Virtual Reality. In: *Advanced Computational Intelligence Techniques for Virtual Reality in Healthcare.* Springer; 2020:141-163.
60. Gélinas-Phaneuf N, Choudhury N, Al-Habib AR, et al. Assessing performance in brain tumor resection using a novel virtual reality simulator. *International journal of computer assisted radiology and surgery.* 2014;9(1):1-9 DOI: https://doi.org/10.1007/s11548-013-0905-8.
61. Alotaibi FE, AlZhrani GA, Mullah MA, et al. Assessing bimanual performance in brain tumor resection with NeuroTouch, a virtual reality simulator. *Operative Neurosurgery.* 2015;11(1):89-98 DOI: https://doi.org/10.1227/NEU.0000000000000631.
62. Gallagher AG, Smith CD, Bowers SP, et al. Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures. *Journal of the American College of Surgeons.* 2003;197(3):479-488 DOI: https://doi.org/10.1016/S1072-7515(03)00535-0.
63. Gomoll AH, O'toole RV, Czarnecki J, Warner JJ. Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. *The American journal of sports medicine.* 2007;35(6):883-888 DOI: https://doi.org/10.1177/0363546506296521.

64. Winkler-Schwartz A, Bajunaid K, Mullah MA, et al. Bimanual psychomotor performance in neurosurgical resident applicants assessed using NeuroTouch, a virtual reality simulator. *Journal of surgical education.* 2016;73(6):942-953 DOI: https://doi.org/10.1016/j.jsurg.2016.04.013.
65. Gélinas-Phaneuf N, Del Maestro RF. Surgical expertise in neurosurgery: integrating theory into practice. *Neurosurgery.* 2013;73(suppl_1):S30-S38 DOI: https://doi.org/10.1227/NEU.0000000000000115.
66. Aboalayon KA, Almuhammadi WS, Faezipour M. A comparison of different machine learning algorithms using single channel EEG signal for classifying human sleep stages. Paper presented at: 2015 Long Island Systems, Applications and Technology2015 DOI: https://doi.org/10.1109/LISAT.2015.7160185.
67. Kumar A, Sushil R, Tiwari AK. Machine Learning Based Approaches for Cancer Prediction: A Survey. *Available at SSRN 3350294.* 2019 DOI: https://doi.org/10.2139/ssrn.3350294.
68. Hassan CAU, Khan MS, Shah MA. Comparison of machine learning algorithms in data classification. Paper presented at: 2018 24th International Conference on Automation and Computing (ICAC)2018 DOI: https://doi.org/10.23919/IConAC.2018.8748995.
69. Siyar S, Azarnoush H, Rashidi S, et al. Machine Learning Distinguishes Neurosurgical Skill Levels in a Virtual Reality Tumor Resection Task. *arXiv preprint arXiv:181108159.* 2018 DOI: https://doi.org/10.1007/s11517-020-02155-3.
70. Nielsen MA. *Neural networks and deep learning.* Vol 2018: Determination press San Francisco, CA, USA:; 2015.
71. Sathya R, Abraham A. Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence.* 2013;2(2):34-38 DOI: https://doi.org./10.14569/IJARAI.2013.020206.
72. Dias RD, Gupta A, Yule SJ. Using machine learning to assess physician competence: A systematic review. *Academic Medicine.* 2019;94(3):427-439 DOI: https://doi.org/10.1097/ACM.0000000000002414.
73. Yost MJ, Gardner J, Bell RM, et al. Predicting academic performance in surgical training. *Journal of surgical education.* 2015;72(3):491-499 DOI: https://doi.org/10.1016/j.jsurg.2014.11.013.
74. Mery CM, Greenberg JA, Patel A, Jaik NP. Teaching and assessing the ACGME competencies in surgical residency. *BULLETIN-AMERICAN COLLEGE OF SURGEONS.* 2008;93(7):39.
75. Dunnington GL, Williams RG. Addressing the new competencies for residents' surgical training. *Academic Medicine.* 2003;78(1):14-21 DOI: https://doi.org/10.1097/00001888-200301000-00005.
76. Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling.* 2003;160(3):249-264 DOI: https://doi.org/10.1016/S0304-3800(02)00257-0.

77. Olden JD, Joy MK, Death RG. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological modelling.* 2004;178(3-4):389-397 DOI: https://doi.org/10.1016/j.ecolmodel.2004.03.013.
78. Sewell C, Morris D, Blevins NH, et al. Validating metrics for a mastoidectomy simulator. Paper presented at: MMVR2007.
79. Azarnoush H, Alzhrani G, Winkler-Schwartz A, et al. Neurosurgical virtual reality simulation metrics to assess psychomotor skills during brain tumor resection. *International journal of computer assisted radiology and surgery.* 2015;10(5):603-618 DOI: https://doi.org/10.1007/s11548-014-1091-z.
80. Sawaya R, Alsideiri G, Bugdadi A, et al. Development of a performance model for virtual reality tumor resections. *Journal of neurosurgery.* 2018;1(aop):1-9 DOI: https://doi.org/10.3171/2018.2.JNS172327.
81. Ledwos N, Mirchi N, Bissonnette V, Winkler-Schwartz A, Yilmaz R, Del Maestro RF. Virtual Reality Anterior Cervical Discectomy and Fusion Simulation on the Novel Sim-Ortho Platform: Validation Studies. *Operative Neurosurgery.* In Press.
82. Marcano-Cedeno A, Quintanilla-Domínguez J, Cortina-Januchs M, Andina D. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. Paper presented at: IECON 2010-36th annual conference on IEEE industrial electronics society2010 DOI: https://doi.org/10.1109/IECON.2010.5675075.
83. Kavzoglu T, Mather PM. Using feature selection techniques to produce smaller neural networks with better generalisation capabilities. Paper presented at: IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment.2000 DOI: https://doi.org/10.1109/IGARSS.2000.860339.
84. Burden F, Winkler D. Bayesian regularization of neural networks. In: *Artificial neural networks.* Springer; 2008:23-42.
85. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology.* 1996;49(11):1225-1231 DOI: https://doi.org/10.1016/S0895-4356(96)00002-9.
86. Pasini A. Artificial neural networks for small dataset analysis. *Journal of thoracic disease.* 2015;7(5):953 DOI: https://doi.org/10.3978/j.issn.2072-1439.2015.04.61.
87. Karlik B, Olgac AV. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems.* 2011;1(4):111-122.
88. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *Journal of machine learning research.* 2012;13(Feb):281-305.
89. Naik B, Ragothaman S. Using neural networks to predict MBA student success. *College Student Journal.* 2004;38(1):143-150.
90. Olden JD, Jackson DA. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks.

Ecological modelling. 2002;154(1-2):135-150 DOI: https://doi.org/10.1016/S0304-3800(02)00064-9.
91. Grantcharov TP, Reznick RK. Teaching procedural skills. *Bmj.* 2008;336(7653):1129-1131 DOI: https://doi.org/10.1136/bmj.39517.686956.47.
92. Chang FM. Characteristics analysis for small data set learning and the comparison of classification methods. Paper presented at: 7th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases2008.
93. Novak R, Bahri Y, Abolafia DA, Pennington J, Sohl-Dickstein J. Sensitivity and generalization in neural networks: an empirical study. Paper presented at: International Conference on Learning Representations2018.
94. Azarnoush H, Siar S, Sawaya R, et al. The force pyramid: a spatial analysis of force application during virtual reality brain tumor resection. *Journal of neurosurgery.* 2016;127(1):171-181 DOI: https://doi.org/10.3171/2016.7.JNS16322.
95. Siyar S, Azarnoush H, Rashidi S, Del Maestro RF. Tremor Assessment during Virtual Reality Brain Tumor Resection. *Journal of Surgical Education.* 2019 DOI: https://doi.org/10.1016/j.jsurg.2019.11.011.
96. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. *Academic medicine.* 2002;77(5):361-367 DOI: https://doi.org/10.1097/00001888-200205000-00003.
97. Bing-You RG, Trowbridge RL. Why medical educators may be failing at feedback. *Jama.* 2009;302(12):1330-1331 DOI: https://doi.org/10.1001/jama.2009.1393.
98. Mirchi N, Bissonnette V, Yilmaz R, Ledwos N, Winkler-Schwartz A, Del Maestro RF. The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PloS one.* 2020;15(2):e0229596 DOI: https://doi.org/10.1371/journal.pone.0229596.
99. AlZhrani G, Alotaibi F, Azarnoush H, et al. Proficiency performance benchmarks for removal of simulated brain tumors using a virtual reality simulator NeuroTouch. *Journal of surgical education.* 2015;72(4):685-696 DOI: https://doi.org/10.1016/j.jsurg.2014.12.014.
100. Shippey SH, Chen TL, Chou B, Knoepp LR, Bowen CW, Handa VL. Teaching subcuticular suturing to medical students: video versus expert instructor feedback. *Journal of Surgical Education.* 2011;68(5):397-402 DOI: https://doi.org/10.1016/j.jsurg.2011.04.006.
101. Hawkins S, Osborne A, Schofield S, Pournaras D, Chester J. Improving the accuracy of self-assessment of practical clinical skills using video feedback–the importance of including benchmarks. *Medical teacher.* 2012;34(4):279-284 DOI: https://doi.org/10.3109/0142159X.2012.658897.
102. Dvorak MF, Collins JB, Murnaghan L, et al. Confidence in spine training among senior neurosurgical and orthopedic residents. *Spine.*

2006;31(7):831-837 DOI: https://doi.org/10.1097/01.brs.0000207238.48446.ce.

103. Arnold PM, Brodke DS, Rampersaud YR. Differences between neurosurgeons and orthopedic surgeons in classifying cervical dislocation injuries and making assessment and treatment decisions: a multicenter reliability study. *Am J Orthop (Belle Mead NJ).* 2009;38(10):E156-E161.

APPENDIX

TABLES

Table 1: Demographics information related to each group of participants for this study.

	Junior Residents	Senior Residents	Post-Residents	
Number of individuals	7	5	9	
Age (years)				
Mean ± SD	27.4±1.4	30.6±2.3	44.2±13.2	
Sex				
Male	5	4	9	
Female	2	1	0	
Level of training / Surgical Specialty	PGY 1-3	PGY 4-6	Fellows	Consultants
Neurosurgery	3	3	2	2
Orthopedic Surgery	4	2	3	2

Table 2: List of performance metrics used to train artificial neural network.

Category of Performance Metrics	Performance Metrics
Safety	Average force applied on right posterior longitudinal ligament
Safety	Average force applied on C5 vertebra
Safety	Average force applied on left posterior longitudinal ligament
Safety	Number of contacts on the spinal dura with active burr
Safety	Average force applied on spinal dura
Safety	Average force applied on right vertebral artery

Table 3: List of ranked performance metrics with their corresponding weights and relative importance for post-residents. all six of these metrics were associated with safety.

Rank	Performance Metrics	Connection Weight Product	Relative Importance (%)
1	Average force applied on right posterior longitudinal ligament	1.5	33.31
2	Average force applied on C5 vertebra	-1.07	23.77
3	Average force applied on left posterior longitudinal ligament	-0.71	15.78
4	Number of contacts on the spinal dura with active burr	-0.68	15.06
5	Average force applied on spinal dura	-0.30	6.72
6	Average force applied on right vertebral artery	0.24	5.36

Table 4: List of ranked performance metrics with their corresponding weights and relative importance for senior residents. all six of these metrics were associated with safety.

Rank	Performance Metrics	Connection Weight Product	Relative Importance (%)
1	Average force applied on right posterior longitudinal ligament	-1.07	37.97
2	Average force applied on left posterior longitudinal ligament	0.94	33.44
3	Number of contacts on the spinal dura with active burr	0.32	11.27
4	Average force applied on spinal dura	-0.30	10.56
5	Average force applied on right vertebral artery	0.11	3.88
6	Average force applied on C5 vertebra	-0.08	2.88

Table 5: List of ranked performance metrics with their corresponding weights and relative importance for junior residents. all six of these metrics were associated with safety.

Rank	Performance Metrics	Connection Weight Product	Relative Importance (%)
1	Number of contacts on the spinal dura with active burr	-0.92	27.54
2	Average force applied on C5 vertebra	0.68	20.31
3	Average force applied on right posterior longitudinal ligament	-0.64	19.04
4	Average force applied on spinal dura	0.63	18.75
5	Average force applied on right vertebral artery	0.33	9.95
6	Average force applied on left posterior longitudinal ligament	-0.15	4.41

Table 6: Comparison between discectomy and osteophyte removal components of the simulated anterior cervical discectomy and fusion.

	Discectomy	Osteophyte Removal
Number of instruments utilized	3 (bone curette, pituitary rongeur and disc rongeur)	1 (burr)
Number of metrics	16 metrics	6 metrics
Metrics categories	Safety, Cognition, Efficiency & Motion	Safety
Most important category of metrics	Safety (makes up more than 50% of the metrics)	Safety (makes up 100% of the metrics)
Testing accuracy	83.3%	83.3%
Connection Weight Product signs	Positive & Negative	Positive & Negative
Highest magnitude of Connection Weight Product	5.24	1.5
Lowest magnitude of Connection Weight Product	0.02	0.08
Hidden patterns	Continuous learning & discontinuous learning	Continuous learning & discontinuous learning

FIGURES

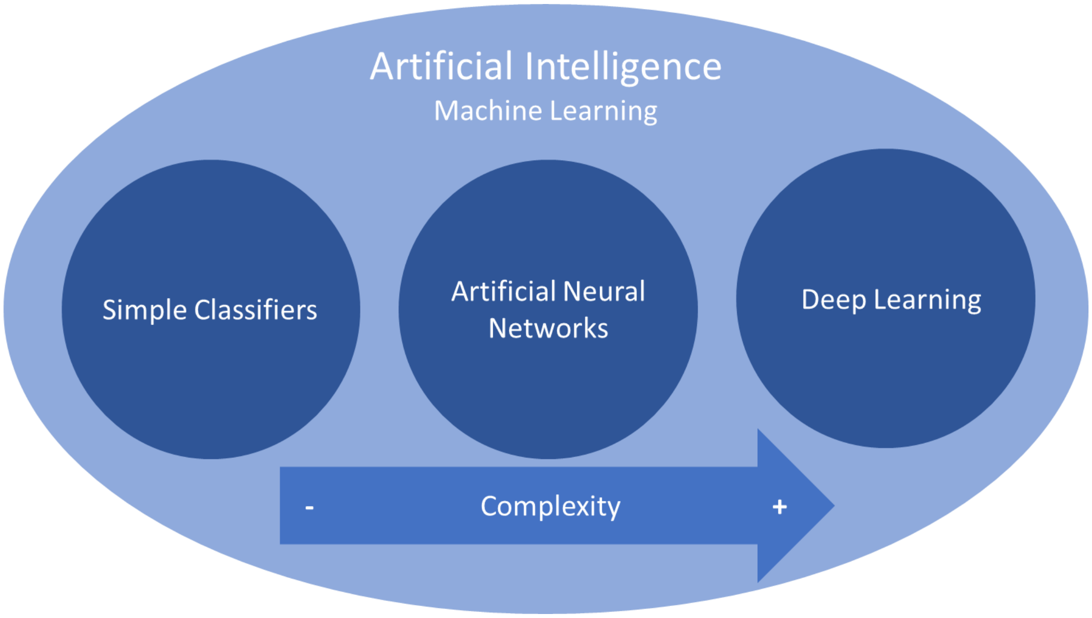

Figure 1: The 3 branches of artificial intelligence in order of complexity related to decision-making capabilities.

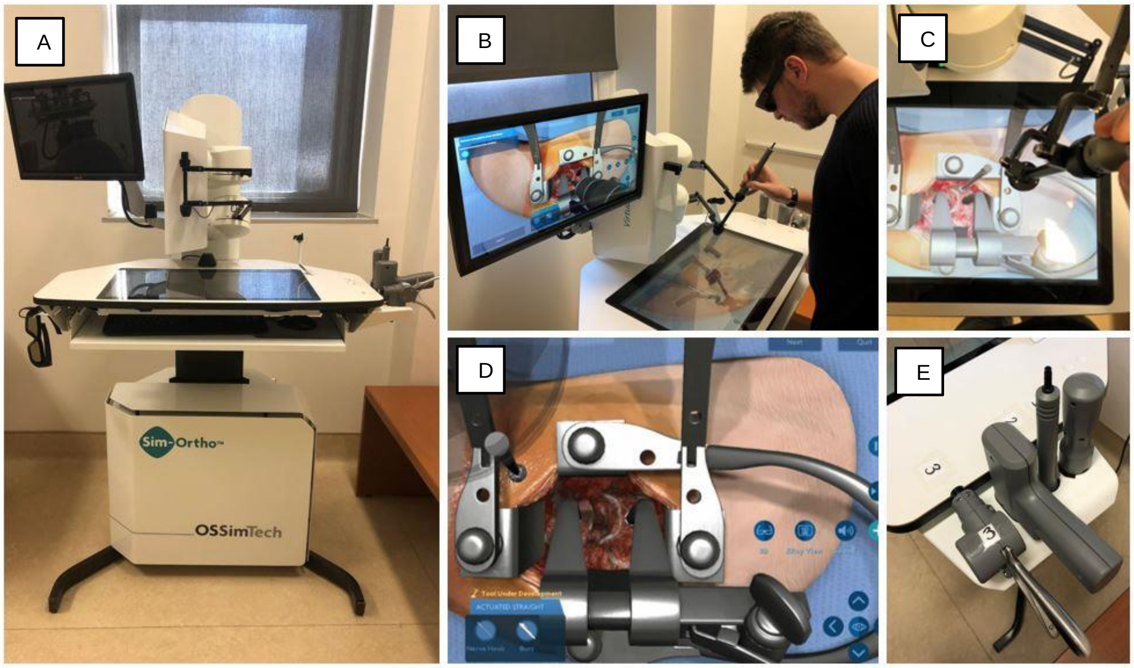

Figure 2: The virtual reality simulator utilized for this study
The Sim-Ortho virtual reality simulator developed by OSSimTech™ (A) Utilization of 3D glasses and the haptic feedback arm. (B) Participant's point of view on the Sim-Ortho Virtual Reality Simulator. (C) Beginning of the osteophyte removal component of the anterior cervical discectomy and fusion. (D) Surgical instrument handles utilized on the Sim-Ortho Virtual Reality Simulator. (E)

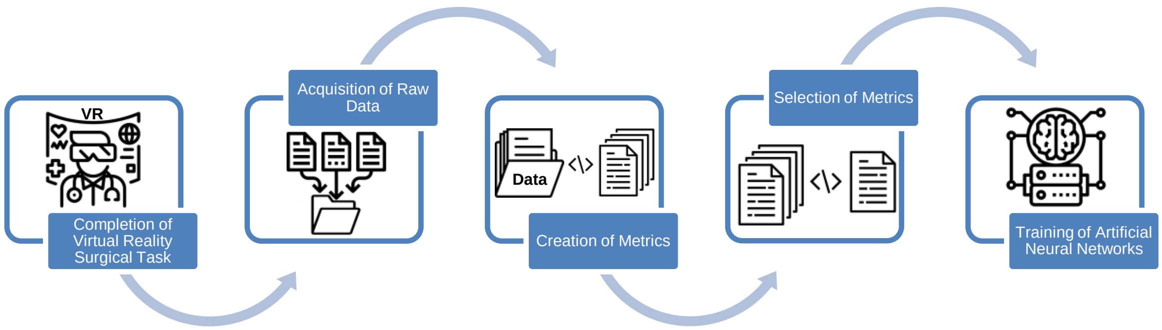

Figure 3: Process of integrating artificial neural networks within virtual reality simulation.
Participants perform the simulated scenario which generates a large dataset of information related to their performance. The data is collected and transformed into quantifiable performance metrics. A feature selection is employed to reduce the metrics to only the most relevant ones. These final metrics are used to train the artificial neural network.

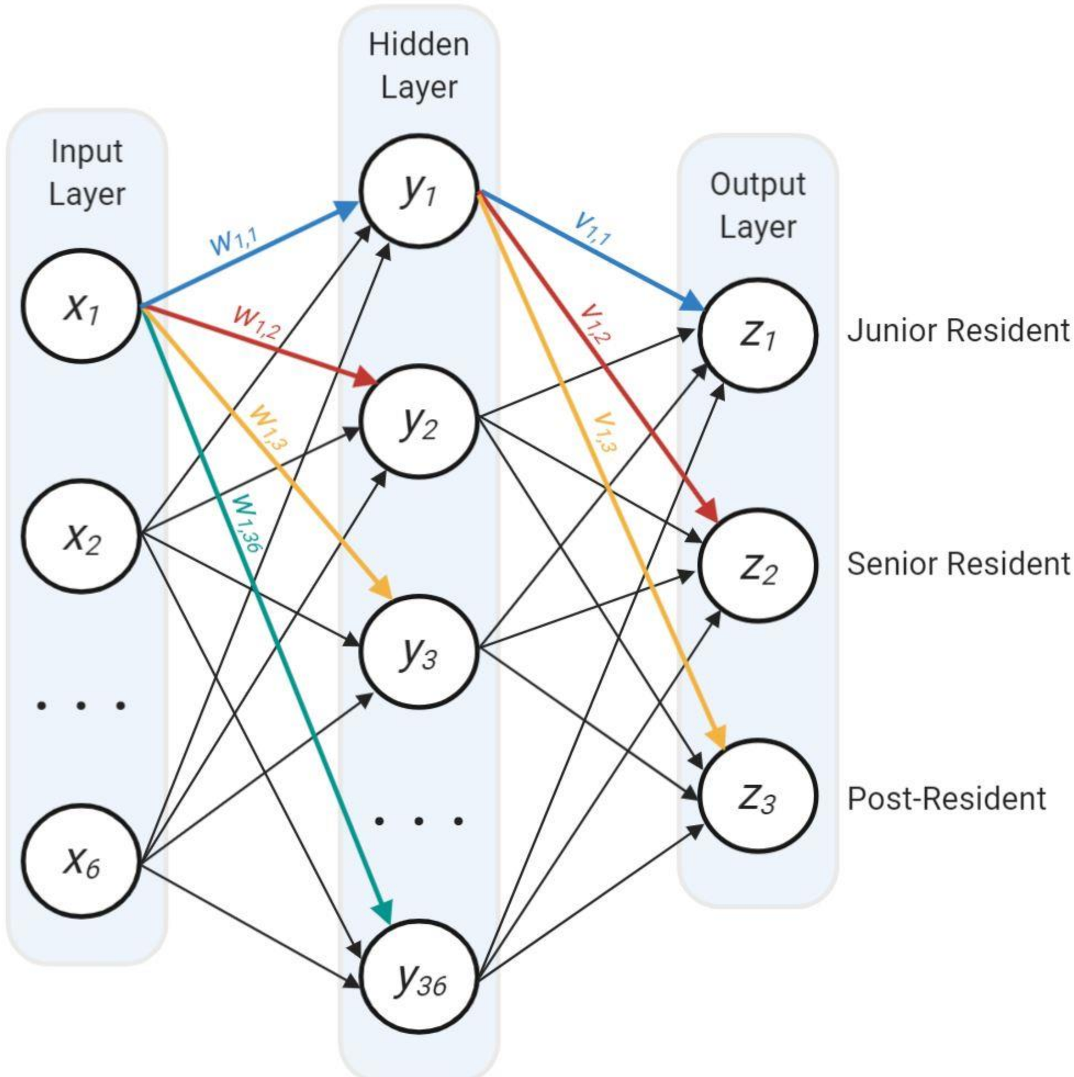

Figure 4: Interconnectivity of the artificial neural network.
The artificial neural network is composed of three layers. The input layer (left) comprises the performance metrics (Xx). Each performance metric is connected to every hidden layer (middle)perceptron (Yx). The perceptrons of the hidden layer are themselves connected to each of the perceptrons (Zx) in the output layer (right) representing each group of variant expertise level. Both the input-hidden layer connections and the hidden-output layer connections have weights (respectively Wx,y and Vx,y) attributed to them. These weights correspond to the sensitivity of each input on the algorithm's decision-making process.

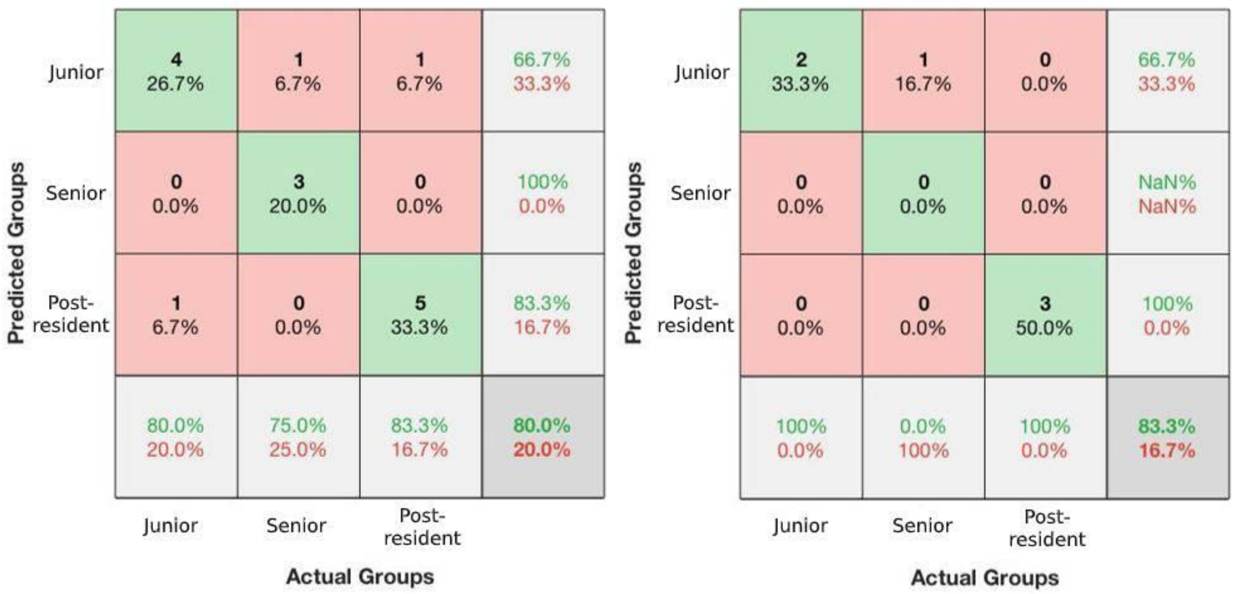

Figure 5: Classification of the model.
The confusion matrix of the training set (left) correctly classified 12 out of 15 participants reaching an 80% accuracy. The confusion matrix of testing set (right) correctly classified 5 out 6 participants reaching an 83.3% accuracy.

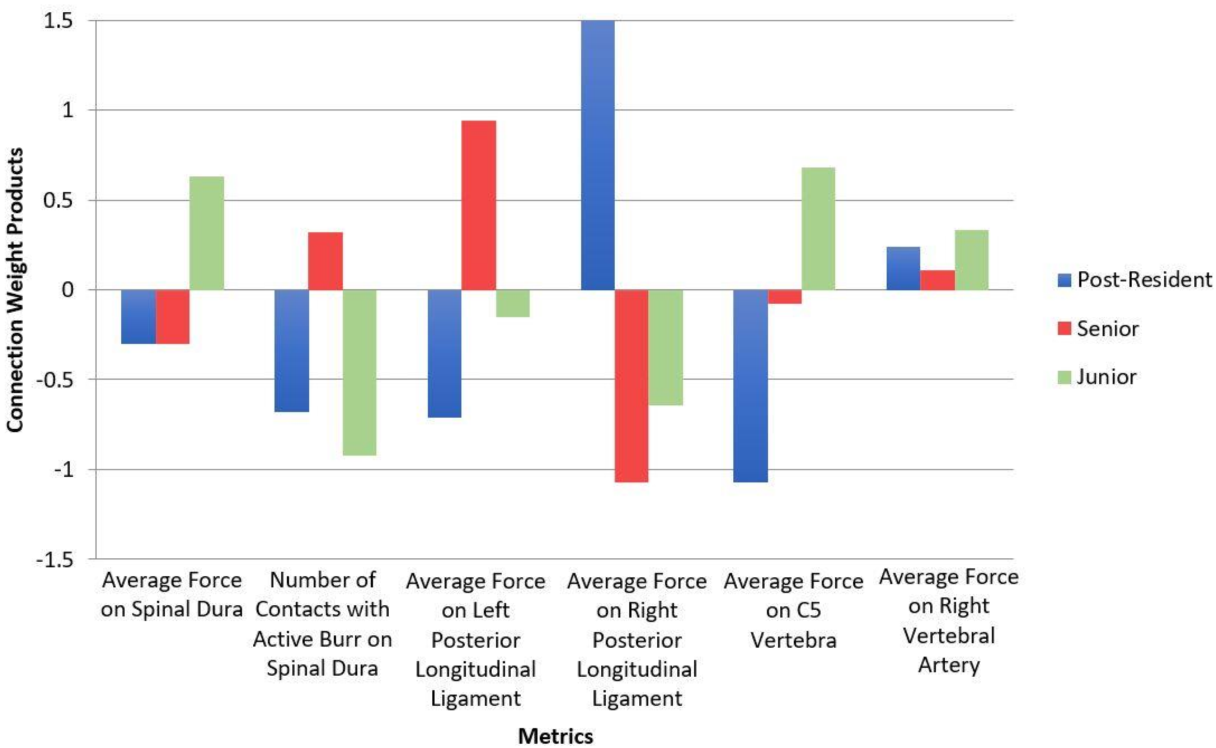

Figure 6: Representation of performance metric Connection Weight Products relative to each level of expertise.
Each relative importance percentage corresponds to the magnitude of the Connection Weight Product of a specific performance metric over the sum of all Connection Weight Product magnitudes within the same performance level group of either post-resident (blue), senior resident (red) or junior resident (green). Hence, the relative importance represents the influence of each metric on the classification of a participant into a specific group.

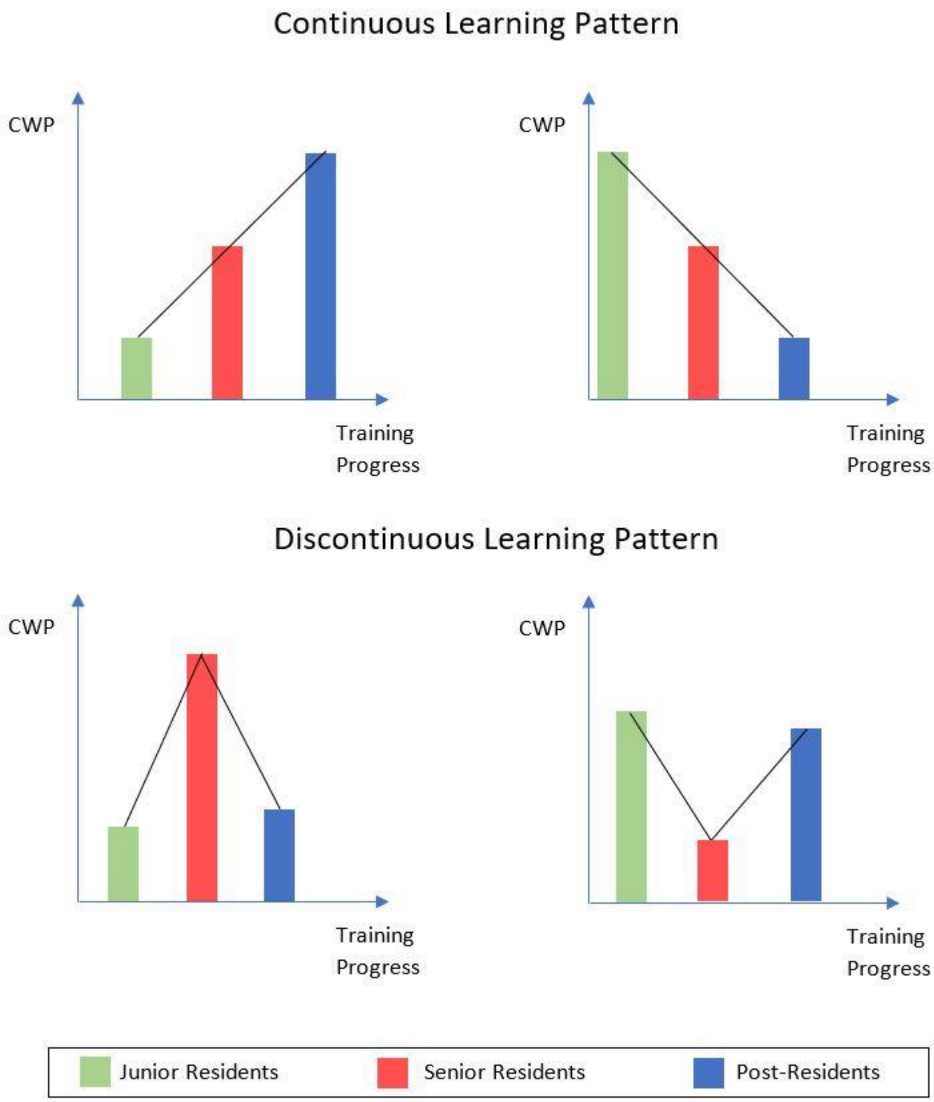

Figure 7: Illustration of continuous and discontinuous learning patterns.
With the X-axis representing the time variable training progress of an individual and the Y-axis representing the Connection Weight Product, the progressive and incremental learning curve associated with the continuous learning pattern and variable non sequential learning curve of the discontinuous learning pattern are illustrated.