Artificial Neural Networks to Assess Virtual Reality Anterior Cervical Discectomy Performance

Nykan Mirchi, MSc, BSc[*]
Vincent Bissonnette, MD[*‡]
Nicole Ledwos, MSc, BA[*]
Alexander Winkler-Schwartz, MD[*]
Recai Yilmaz, MD[*]
Bekir Karlik, PhD[*]
Rolando F. Del Maestro, MD, PhD[*]

[*]Neurosurgical Simulation and Artificial Intelligence Learning Centre, Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Canada; [‡]Division of Orthopaedic Surgery, Montreal General Hospital, McGill University, Montreal, Canada

Correspondence:
Nykan Mirchi, MSc, BSc,
Neurosurgical Simulation and Artificial Intelligence Learning Centre,
Department of Neurology and Neurosurgery,
Montreal Neurological Institute and Hospital,
McGill University,
3801 University St, Room E2.89,
Montreal H3A 2B4, Canada.
Email: nykan.mirchi@mail.mcgill.ca
Twitter: @nykanm

Received, April 30, 2019.
Accepted, September 4, 2019.

BACKGROUND: Virtual reality surgical simulators provide a safe environment for trainees to practice specific surgical scenarios and allow for self-guided learning. Artificial intelligence technology, including artificial neural networks, offers the potential to manipulate large datasets from simulators to gain insight into the importance of specific performance metrics during simulated operative tasks.

OBJECTIVE: To distinguish performance in a virtual reality-simulated anterior cervical discectomy scenario, uncover novel performance metrics, and gain insight into the relative importance of each metric using artificial neural networks.

METHODS: Twenty-one participants performed a simulated anterior cervical discectomy on the novel virtual reality Sim-Ortho simulator. Participants were divided into 3 groups, including 9 post-resident, 5 senior, and 7 junior participants. This study focused on the discectomy portion of the task. Data were recorded and manipulated to calculate metrics of performance for each participant. Neural networks were trained and tested and the relative importance of each metric was calculated.

RESULTS: A total of 369 metrics spanning 4 categories (safety, efficiency, motion, and cognition) were generated. An artificial neural network was trained on 16 selected metrics and tested, achieving a training accuracy of 100% and a testing accuracy of 83.3%. Network analysis identified safety metrics, including the number of contacts on spinal dura, as highly important.

CONCLUSION: Artificial neural networks classified 3 groups of participants based on expertise allowing insight into the relative importance of specific metrics of performance. This novel methodology aids in the understanding of which components of surgical performance predominantly contribute to expertise.

KEY WORDS: Anterior cervical discectomy, Artificial intelligence, Artificial neural networks, Machine learning, Simulation, Surgical training, Virtual reality

Operative Neurosurgery 0:1–11, 2019 DOI: 10.1093/ons/opz359

Virtual reality surgical simulation is an evolving field of research that has the potential to complement more objective and competency-based surgical training methods.[1-3] The incidence, complexity, and implications of the anterior cervical discectomy and fusion (ACDF) makes this procedure a good candidate for simulation-based training.[4,5] The ACDF procedure requires proficiency in multiple areas, including an understanding of the critical anatomic structures, along with gaining an appreciation of how different structures react to manipulations and instrument usage.[4] Virtual reality simulators record an enormous amount of data concerning psychomotor performance during a simulated task. Studies in surgical simulation have developed methodologies to exploit these large datasets to develop validated metrics of performance, which can be used by surgical educators to enhance performance.[6,7]

ABBREVIATION: ACDF, anterior cervical discectomy and fusion

Supplemental digital content is available for this article at www.operativeneurosurgery-online.com.

Artificial intelligence is a broad term used to describe a set of algorithms that can make seemingly intelligent decisions.[8,9] Machine learning is a subset of artificial intelligence, in which algorithms are able to identify and learn from hidden patterns in multivariate datasets without the need for explicit programming.[9] Artificial neural networks are a deeper subset of machine learning, inspired from the brain's neuronal connectivity.[10] They are sets of interconnected nodes (referred to as neurons in this paper) that can communicate with each other through connections of different weights. These weights are essentially analogous to neuromodulatory signals that influence how neurons communicate. When combined with virtual reality simulation, artificial neural networks can be designed to classify participants and discover specific metrics that differentiate surgical performance. This information can provide an objective assessment of surgical psychomotor performance, providing insight into the components that underpin surgical expertise.

Artificial intelligence has been utilized to assess surgical expertise during virtual reality performance, but the majority of these studies limit their analysis to the classification of different participant groups.[11-13] These systems fail to explore the underlying reasons for classification by investigating the relative importance of the individual metrics of performance.[14]

The 3 objectives of this study were the following: (1) to develop metrics of performance for a novel virtual reality ACDF simulation; (2) to employ artificial neural networks to classify participants' expertise based on their performance in the simulated task; and (3) to examine the ability of our neural network to outline the relative weights of specific metrics in the determination of expert performance in this virtual reality spinal procedure. This novel methodology has the potential to aid in the understanding of components of surgical expertise and contribute to the paradigm shift towards competency-based surgical training. To our knowledge, this is the first study to employ artificial neural networks to gain insight into the relative weights of teachable performance metrics in a virtual reality surgical simulation.

TABLE 1. Demographics Information for 3 Groups of Participants Performing the Virtual Reality Surgical Task

	Post-Resident (9)	Senior (5)	Junior (7)
Age (years)			
Mean, SD	44.2 ± 13.2	30.6 ± 2.3	27.4 ± 1.4
Sex			
Male	9	4	5
Female	0	1	2
Level of training			
Neurosurgery residents			
PGY 1	–	–	0
PGY 2	–	–	2
PGY 3	–	–	1
PGY 4	–	1	–
PGY 5	–	0	–
PGY 6	–	2	–
Orthopaedic residents			
PGY 1	–	–	0
PGY 2	–	–	3
PGY 3	–	–	1
PGY 4	–	0	–
PGY 5	–	2	–
Spine fellows			
Neurosurgical	3	–	–
Orthopaedic	2	–	–
Spine surgeons			
Neurosurgeons (years of practice)	2 (3, 35)	–	–
Orthopaedic surgeons (years of practice)	2 (18, 28)	–	–
Surgical knowledge of an ACDF (self-rated, Likert scale, 1-5)			
Median (range)	5 (4-5)	3 (3-4)	3 (1-3)

METHODS

Participants

Twenty-seven participants were recruited to perform a virtual reality ACDF utilizing the Sim-Ortho platform. No participants had previous experience with this ACDF scenario on the Sim-Ortho platform. This simulator is only optimized for right-handed users, which excluded 3 left-handed participants. One fellow and 2 neurosurgeons were also excluded, as their practice was not primarily spine focused. The demographics of the 21 remaining participants are outlined in Table 1. The participants were divided into the following 3 groups: post-resident, senior, and junior. All participants signed consent forms approved by the Montreal Neurological Institute and Hospital Research Ethics Board before trial participation.

Virtual Reality Surgical Simulator

The virtual reality simulator employed is the Sim-Ortho platform (Figure 1A) codeveloped by OSSimTech™ (Montreal, Canada) and the AO Foundation (Bienne, Switzerland).[15] The platform offers a variety of tool handles (Figure 1B), each used to simulate different surgical instruments utilized by participants for the simulated procedure (Figure 1C). The platform relies on voxel-based gaming graphics to create a hyper-realistic 3D intraoperative environment mimicking real surgical procedures (Figure 1D). The participant wears 3D glasses, experiencing visual and auditory feedback when employing instruments (Figure 1E); whereas, the haptic feedback allows multiple tissue manipulations by the instruments utilized.

Simulated Surgical Scenario

The scenario simulated in this study was an ACDF. The simulation was divided into 4 steps as follows: cutting the disc annulus to gain disc access, cervical discectomy with excision of disc annulus and nucleus, removal of posterior osteophytes, and excision of the posterior longitudinal ligament. Prior to the start of the scenario, participants were asked to provide information concerning their knowledge of the ACDF procedure on a 5-point Likert scale. No time limit was imposed, and the simulated task was performed in an environment devoid of distractions. Each step of the simulated scenario is described in **Methods, Supplemental Digital Contents 1-5** and **Figure, Supplemental Digital Content 6**.

The focus of this manuscript is the discectomy step, as demonstrated in Video. This component of the ACDF was chosen because of its complex nature, requiring participants to choose between 3 different

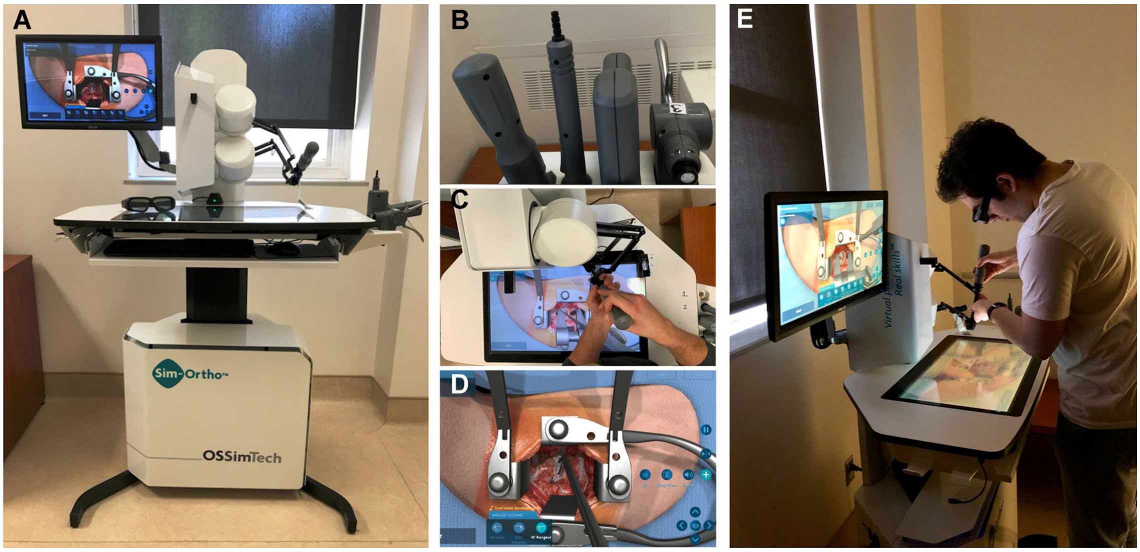
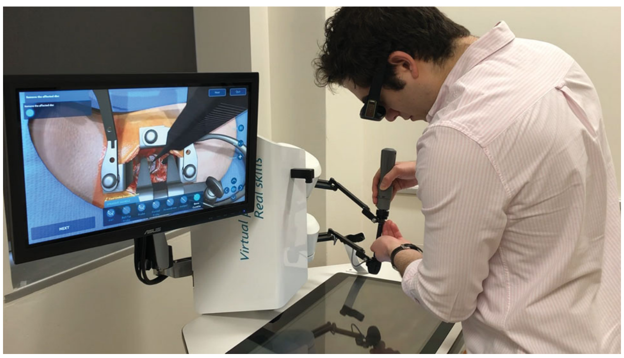


FIGURE 1. *The virtual reality platform used to simulate an anterior cervical discectomy.* **A**, *The Sim-Ortho platform, codeveloped by OSSimTech^{TM} and the AO Foundation, is designed to simulate a number of surgical procedures.* **B**, *A variety of instruments are available accompanied by a variety of different handles to simulate the feel of each instrument.* **C**, *The participant holds the instrument in their dominant hand, receiving haptic feedback when interacting with anatomic structures.* **D**, *The platform is built on a gaming engine, providing very realistic 3D graphics.* **E**, *The participant wears 3D glasses while interacting with the platform.*



VIDEO. *User performing a virtual reality simulation of an anterior cervical discectomy on the Sim-Ortho platform. The user wears 3D glasses, and they can select between 3 simulated instruments: bone curette, pituitary rongeur, or disc rongeur to perform the discectomy while receiving haptic feedback.*

instruments to adequately perform the discectomy. This allowed for a comprehensive evaluation of both psychomotor performance and cognition decision-making.

Performance Metrics Generation

The study methodology is illustrated in Figure 2. This method follows guidelines established by our group to utilize machine learning algorithms to assess surgical expertise in simulation.[16] The simulator recorded a series of data pertaining to participant use of individual instruments throughout the procedure. This resulting information was divided into 66 variables for each tool, including time, position and angles of the simulated instruments, forces applied on specific anatomic structures, and volume of any anatomic structure removed. Metrics of performance were developed by combining available raw data to develop a smaller and more understandable set of metrics. Metrics were generated in 3 ways. (1) Consultation with expert spine surgeons to identify the components of the discectomy surgery they believed important in performing a safe procedure. The engineers involved with the development of the Sim-Ortho platform also consulted with spine surgeons to decide which raw data could be adequately provided on the platform. (2) Metrics were derived from published work involving lumbar discectomy.[17] (3) Novel metrics were created by the authors based on different components of surgical skill.

Machine Learning and Artificial Neural Network Design

The data consisting of 21 participants with the final metrics were split into 2 sets, in which 70% was used for training (15 participants: 6 post-residents, 4 seniors, and 5 juniors) and 30% for testing (6 participants: 3 post-residents, 1 senior, and 2 juniors). The training group was used to train the neural network (Figure 3) in a supervised manner, and the remaining data were used to test the model. The training and testing process is more explicitly explained in Figure 4. The Connection Weights Algorithm was used to determine the relative importance of each metric of performance for each group (post-resident, senior, and junior).[18] An extensive description of the metric selection, artificial

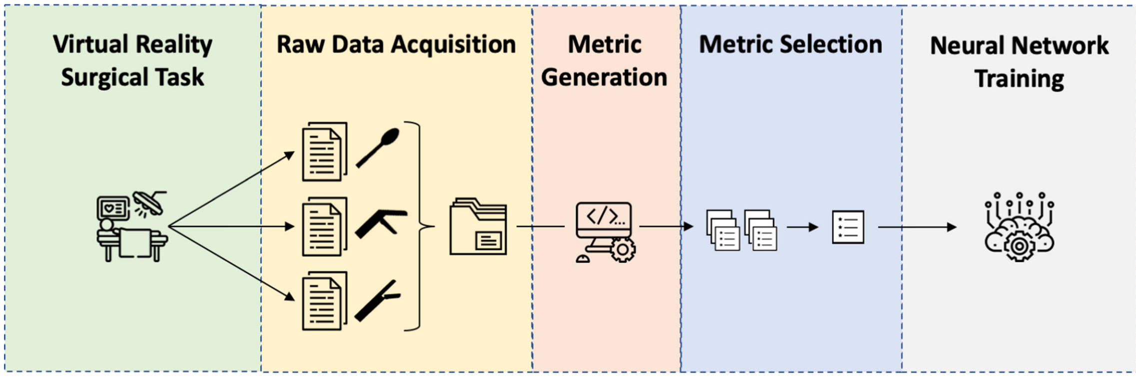


FIGURE 2. *Methodology for the use of neural networks to assess expertise in a virtual reality surgical simulator. Users begin by performing the surgical task on the virtual reality platform. Raw data acquisition occurs as the platform creates large datasets for each instrument employed. All instrument datasets are combined into a single dataset. The large dataset can be used to generate metrics of performance for each participant. The new set of metrics can then undergo metric selection to narrow down a group of metrics able to differentiate levels of surgical expertise. The final metrics are then fed to the neural network for training and testing.*

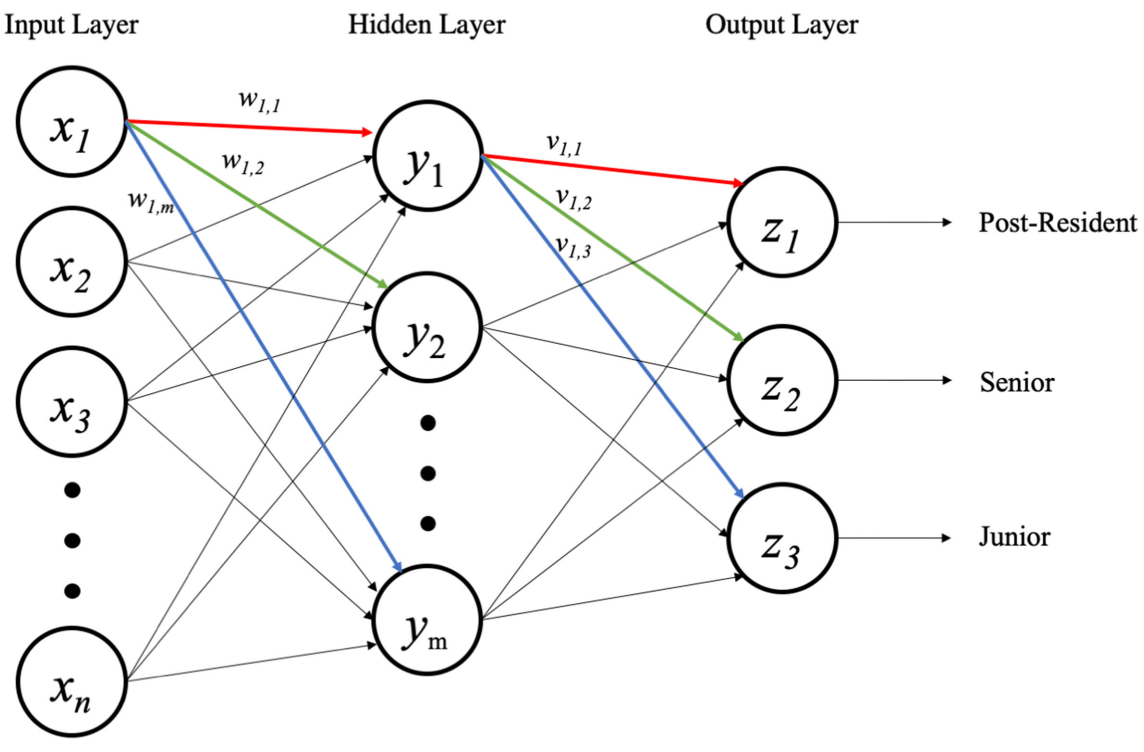


FIGURE 3. *Simplified illustration of multilayered perceptron used by the artificial neural network. Metrics of performance are inputs of the neural network represented by x. Each and every input is connected to a number of neurons in the hidden (middle) layer of the network, represented by y. The connection between each input and each neuron is supported by a weight (w), in which a large magnitude of the weight means that the hidden neuron will be more sensitive to alterations in this specific input (x). Each neuron of the hidden layer is then connected to the 3 possible outputs, post-resident, senior, and junior, represented by z1, z2, and z3, respectively. Similarly, to the input layer, each neuron of the hidden layer is assigned a weight (v) to influence the output neurons. The output neuron with the value closest to 1 will be the neural network's final decision.*

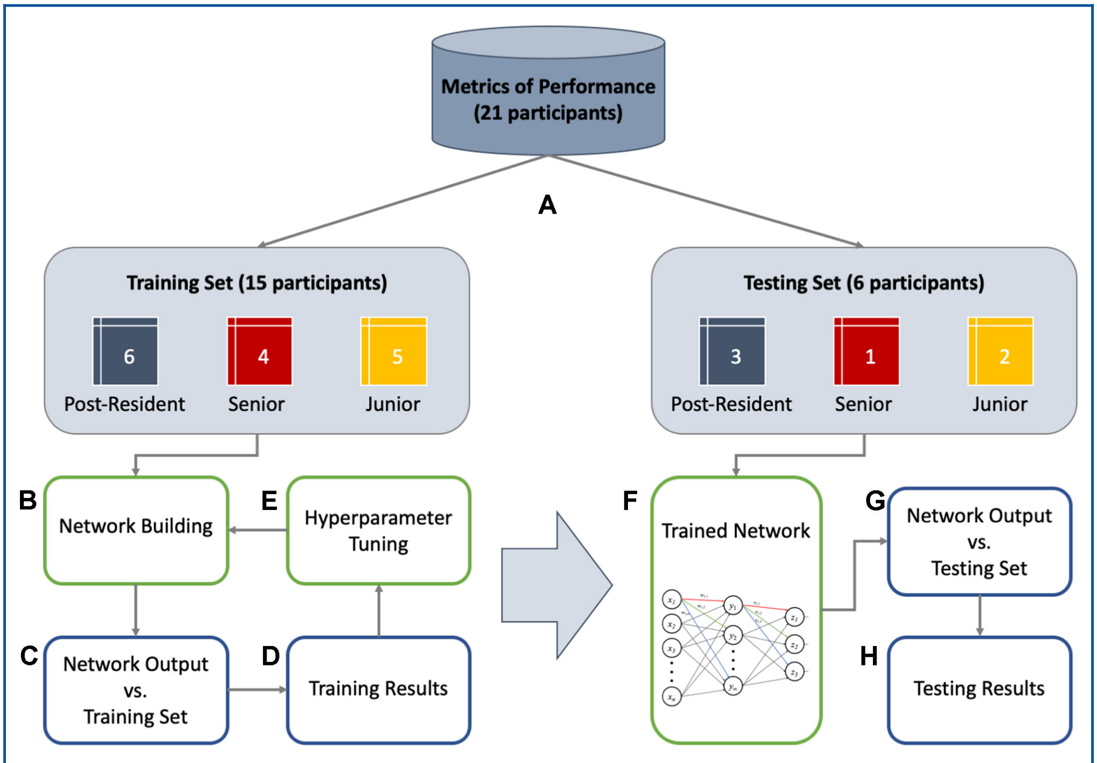


FIGURE 4. *Training and testing of artificial neural network.* **A**, *The final set of metrics of performance from 21 participants is split into a training and testing set.* **B**, *The training set is used to build an initial neural network.* **C**, *The network predicts a group for each participant, and this prediction is compared to the true grouping of each respective participant.* **D**, *A training accuracy is calculated based on how well the network output resembles the true training set grouping. This is based on calculating a cost function for the network, in which a high cost represents poor performance.* **E**, *If the neural network performed poorly, its hyperparameters are tuned according to a set of predefined rules in an attempt to reduce cost. This process repeats until the network's cost reaches a minimum.* **F**, *Following the iterative training process, the final trained network can now be exported and tested. The testing set, containing previously unseen data, is fed to the network.* **G**, *The network predicts a group for each participant in the testing set, and this prediction is compared to the true grouping of each respective participant.* **H**, *The testing accuracy is calculated based on how well the network's output resembles the true testing test grouping. Unlike training, this process is only done once.*

neural network design, and metric importance calculation can be found in **Methods, Supplemental Digital Contents 1-5** and **Figure, Supplemental Digital Content 6**.

RESULTS

Metrics of Performance

Performance metrics developed for the discectomy components of the simulated ACDF were divided into 4 categories: safety, efficiency, motion, and cognition, which are outlined in Table 2. A total of 369 metrics were calculated for each participant. Following removal of metrics that contained a value of zero for all participants, 333 remained. Following metric selection with the stepwisefit function, 13 significant metrics were selected, and they are described in Table 3. Some of these metrics are specific to certain instruments used for the discectomy. The authors deemed it important for the model to consider the tool choice of participants, as this would influence one's metric for this particular instrument. Hence, 3 binary metrics were added, 1 for each tool (bone curette, pituitary rongeur, and disc rongeur), in which a value of 1 corresponds to the use of an instrument, and 0 corresponds to no use of the respective instrument, for a total of 16 metrics.

Surgical Performance Classification

The dataset was divided into a training set (15 participants and 16 metrics) and testing set (6 participants and 16 metrics). The artificial neural network was then optimized (Marquardt adjustment parameter $\mu = 0.01$, μ decrease ratio $= 0.95$, and μ increase ratio $= 10$) and trained over 10 000 iterations. A training accuracy of 100% and a testing accuracy of 83.3% were achieved.

TABLE 2. Metrics of Performance for Virtual Reality Simulation of Anterior Cervical Discectomy Scenario

Metric category	Metric list
Safety	Number of voxels (volume) removed from an anatomic structure.
	Average force applied on anatomic structure.
	Maximum force applied on anatomic structure.
Motion	Velocity of instrument while in contact with the disc nucleus and disc annulus.
	Acceleration of instrument while in contact with the disc nucleus and disc annulus.
	Angles of instruments while in contact with the disc nucleus and disc annulus.
	Angular velocity of instruments while in contact with the disc nucleus and disc annulus.
Efficiency	Number of instrument contacts on an anatomic structure.
	Number of instances a volume or an anatomic structure is removed with an instrument.
	Amount of time an instrument spends in contact with an anatomic structure.
	Total path travelled by an instrument while in contact with an anatomic structure.
Cognition	Instrument choice.

A breakdown of the networks training and testing performance are displayed in confusion matrices in Figure 5A and 5B, respectively.

Metric Importance

The decision-making process of the artificial neural network is more sensitive to alterations in certain metrics of performance for each group. Using the Connection Weights Algorithm, the relative importance of each metric of performance was calculated. The ranked metrics for the post-resident, senior, and junior groups are displayed in Tables 4-6, respectively. Figure 6 illustrates a visual comparison of the relative importance of each metric. Additional visual interpretations of the Connection Weight Products can also be found in **Figure, Supplemental Digital Content 6**. Interestingly, the number of contacts with the spinal dura is in the top 3 most important metrics for all 3 groups. Following the signs of the connection weight product for this metric (post-resident CWP = −3.01; senior CWP = 4.83; junior CWP = −2.71), a decrease in the contacts with the spinal dura increases the likelihood of being classified in the post-resident group, whereas an increase in the number of contacts on the spinal dura increases the likelihood of classification in the senior group. The maximum amount of force applied on the left posterior longitudinal ligament (post-resident CWP = 1.91; senior CWP = 4.17; junior CWP = -5.24) is also highly ranked across all groups. A larger force application increases the likelihood of post-resident or senior classification; whereas, a lower force application increases the likelihood of junior group classification.

DISCUSSION

Patterns in Relative Metric Importance

An important study finding is the ability of the network to rank the importance of a specific metric in the final assessment of expertise in a virtual reality procedure. Generating this data allows surgical educators to address a number of new questions. Should surgical educational paradigms predominantly focus on making sure that specific metrics that contribute extensively to expertise take precedence in any surgical training system? The analysis of the neural network uncovered some general patterns in some performance metrics. For example, our network would preferentially classify a new participant in the senior group as opposed to the post-resident or junior groups if they had large numbers of instrument contacts with the dura. The senior group contacted the dura more frequently than either the junior or the post-resident group with the post-resident group having the least number of dural contacts. One explanation for these findings may be that lack of ACDF experience in the junior group caused more hesitation when approaching the dura or other structures with instruments, thus explaining their low number of dural contacts. The post-resident group, possibly associated with their greater appreciation for this safety component of ACDF procedures, appears to have modulated their behavior after completing residency, resulting in decreased instrument dural contact when compared to the senior group. A different pattern was observed with the maximum force applied to the left posterior longitudinal ligament. For this metric, the network associates higher forces with the senior and post-resident groups. Maximum force application for the post-resident group lies in an intermediate range compared to the senior and junior groups. This could also be due to the post-resident group altering their behavior before residency, resulting in decreased instrument force application in the posterior longitudinal ligament region associated with safety concerns. The results of this study appear to be consistent with previous findings with a virtual reality tumor resection model, in which both safety and efficiency were found to be hallmarks of expert performance.[19]

TABLE 3. Selected Metrics of Performance for Simulated Discectomy

Category	Label	Description
Safety	Contacts_Dura	Number of contacts with the spinal dura during discectomy.
	VolumeRemoved_PLL_Right	Volume of right posterior longitudinal ligament removed.
	ForceMax_PLL_Left	Maximum force applied on the left posterior longitudinal ligament.
	ForceMax_DiscAnnulus_BoneCurette	Maximum force applied on the disc annulus by the bone curette.
	ForceMax_LVA_PitRongeur	Maximum force applied on the left vertebral artery region by the pituitary rongeur.
Motion	VelocityMean_DiscAnnulus_PitRongeur	Average velocity of the pituitary rongeur while in contact with the disc annulus.
	AccelerationNumZ_DiscAnnulus_PitRongeur	Number of accelerations of the pituitary rongeur along the anterior-posterior axis while in contact with the disc annulus.
	AccelerationMaxY_DiscNucleus_BoneCurette	Maximum acceleration of the bone curette along the anterior-posterior axis while in contact with the disc nucleus.
	PitchMean_DiscNucleus_BoneCurette	Average pitch of the bone curette while in contact with the disc nucleus. Pitch is the rotation of the curette in up and down (scooping) motion.
Efficiency	ContactNumber_C5	Number of contacts on the C5 vertebra over the entire procedure.
	CuttingNumber_C5_BoneCurette	Number of contacts on the C5 vertebra using the bone curette.
	ContactTime_LVA	Amount of time spent in contact with the left vertebral artery region.
	TTPLperStrokeY_DiscNucleus_DiscRongeur	Total length of individual strokes along the medial-lateral axis with disc rongeur while in contact with the disc nucleus.

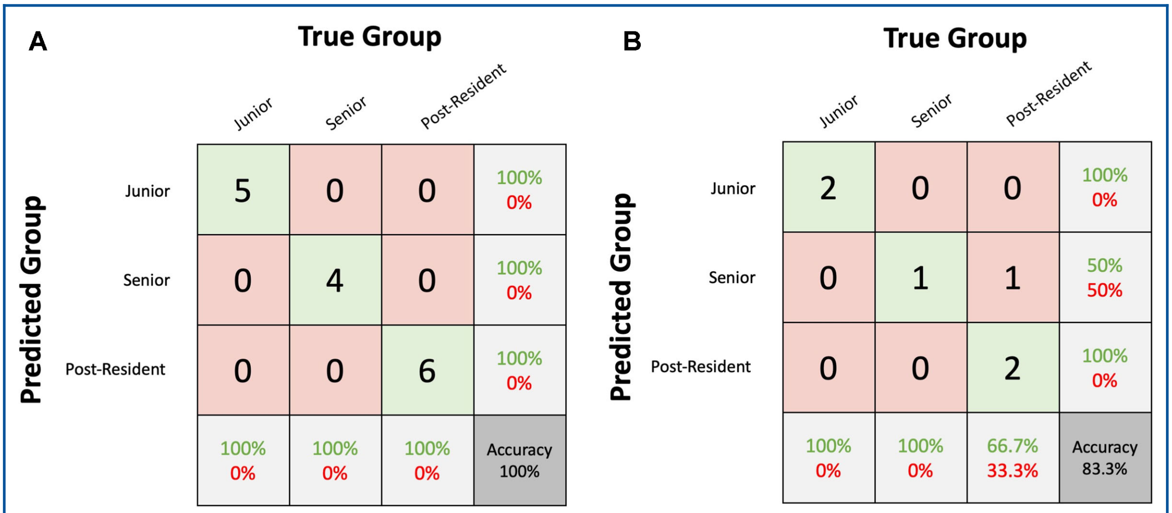


FIGURE 5. *Confusion matrices of the artificial neural network.* **A**, *The training group correctly classified all participants into their respective groups, reaching a 100% accuracy.* **B**, *The testing group correctly classified 5 of the 6 participants, reaching an accuracy of 83.3%. One participant belonging to the post-resident group was incorrectly classified in the senior group.*

TABLE 4. Metrics of Performance Ranked by Their Relative Importance for the Post-resident Group

Rank	Category	Metric	Connection weight product	Relative importance (%)
1	Safety	Contacts_Dura	– 3.01	17.18
2	Safety	ForceMax_DiscAnnulus_BoneCurette	2.56	14.60
3	Cognition	ToolChoice_BoneCurette	1.92	10.97
4	Safety	ForceMax_PLL_Left	1.91	10.93
5	Efficiency	TTPLperStrokeY_DiscNucleus_DiscRongeur	1.14	6.51
6	Efficiency	ContactTime_LVA	1.02	5.84
7	Cognition	ToolChoice_PituitaryRongeur	0.89	5.11
8	Motion	VelocityMean_DiscAnnulus_PitRongeur	0.87	4.98
9	Efficiency	ContactNumber_C5	– 0.81	4.63
10	Motion	AccelerationNumZ_DiscAnnulus_PitRongeur	0.78	4.43
11	Cognition	ToolChoice_DiscRongeur	0.71	4.03
12	Efficiency	CuttingNumber_C5_BoneCurette	– 0.66	3.79
13	Motion	AccelerationMaxY_DiscNucleus_BoneCurette	– 0.38	2.16
14	Safety	VolumeRemoved_PLL_Right	– 0.35	1.98
15	Motion	PitchMean_DiscNucleus_BoneCurette	0.34	1.96
16	Safety	ForceMax_LVA_PitRongeur	0.16	0.89

TABLE 5. Metrics of Performance Ranked by Their Relative Importance for the Senior Group

Rank	Category	Metric	Connection weight product	Relative importance (%)
1	Safety	Contacts_Dura	4.83	15.90
2	Motion	PitchMean_DiscNucleus_BoneCurette	– 4.65	15.31
3	Safety	ForceMax_PLL_Left	4.17	13.74
4	Efficiency	TTPLperStrokeY_DiscNucleus_DiscRongeur	– 3.16	10.39
5	Cognition	ToolChoice_BoneCurette	– 3.03	9.97
6	Safety	ForceMax_DiscAnnulus_BoneCurette	– 2.68	8.82
7	Cognition	ToolChoice_PituitaryRongeur	– 2.06	6.77
8	Efficiency	ContactTime_LVA	– 1.90	6.25
9	Efficiency	CuttingNumber_C5_BoneCurette	– 1.42	4.68
10	Motion	AccelerationNumZ_DiscAnnulus_PitRongeur	– 1.23	4.04
11	Efficiency	ContactNumber_C5	– 0.42	1.39
12	Motion	AccelerationMaxY_DiscNucleus_BoneCurette	– 0.28	0.93
13	Cognition	ToolChoice_DiscRongeur	0.26	0.86
14	Safety	ForceMax_LVA_PitRongeur	0.14	0.46
15	Safety	VolumeRemoved_PLL_Right	– 0.13	0.42
16	Motion	VelocityMean_DiscAnnulus_PitRongeur	– 0.02	0.07

The authors do not believe that virtual reality surgical training combined with neural networks replaces present methods of training for an anterior cervical discectomy. However, the information from this study may provide surgical educators with new perspectives on critical aspects of expert performance during cervical discectomies as well as providing newer metrics for self-guided learning through an artificial intelligence-powered feedback platform.

Application of Neural Network in Education

Education for complex tasks has become a growing application of interest for artificial neural networks.[20] Attention is focused on the development of tools that employ neural networks to break down and better understand the factors that differentiate learner performance. Unlike traditional teaching methods, which may weigh all components of a task relatively equally, the neural networks allow for each component (or metric of

TABLE 6. Metrics of Performance Ranked by Their Relative Importance for the Junior Group

Rank	Category	Metric	Connection weight product	Relative Importance (%)
1	Safety	ForceMax_PLL_Left	− 5.24	21.55
2	Motion	PitchMean_DiscNucleus_BoneCurette	5.14	21.13
3	Safety	Contacts_Dura	− 2.71	11.14
4	Efficiency	TTPLperStrokeY_DiscNucleus_DiscRongeur	1.75	7.18
5	Cognition	ToolChoice_PituitaryRongeur	1.56	6.41
6	Cognition	ToolChoice_BoneCurette	1.33	5.46
7	Efficiency	ContactNumber_C5	1.31	5.40
8	Cognition	ToolChoice_DiscRongeur	− 1.06	4.36
9	Motion	AccelerationMaxY_DiscNucleus_BoneCurette	− 0.89	3.65
10	Efficiency	CuttingNumber_C5_BoneCurette	0.80	3.29
11	Safety	VolumeRemoved_PLL_Right	− 0.60	2.47
12	Efficiency	ContactTime_LVA	0.54	2.19
13	Safety	ForceMax_LVA_PitRongeur	− 0.48	1.98
14	Safety	ForceMax_DiscAnnulus_BoneCurette	0.34	1.41
15	Motion	VelocityMean_DiscAnnulus_PitRongeur	− 0.33	1.35
16	Motion	AccelerationNumZ_DiscAnnulus_PitRongeur	− 0.25	1.02

performance) of a task to be weighed individually, offering a more holistic understanding of expertise. However, some questions remain unanswered. Should the training of the junior residents performing the ACDF scenario on this simulator be focused on training to the senior level of performance, or that of the post-resident group? Should significant time be spent on training all metrics, including those which are less important (ie, less likely to influence a participant being classified into a particular group as defined by the neural network), or should training follow best practices in adult learning theory, such as cognition load theory, and focus only on a small set of critical metrics (those most likely to influence participant classification as defined by the neural network) at any given time?[21] More research is needed to address these important questions.

In the future, the network presented in this study will be employed to develop an automated and more personalized feedback platform for virtual reality surgical training. Once this platform is in place, we will be able to determine whether feedback on the selected metrics has the ability to determine and truly improve performance.

Limitations

The Sim-Ortho virtual reality surgical simulator incorporates an advanced gaming engine, but fails to represent the continually changing operating room environment. First, this study was only focused on the cervical discectomy component of the ACDF. This was done to develop and assess the potential of artificial neural networks before applying them to other components of the ACDF procedure. Ongoing studies are now investigating the other 3 components of the ACDF. Second, the simulator is one-handed, which does not allow for the quantification of bimanual skills, which have been shown to be important in differentiating expertise level in previous studies and an important component of a proposed model for virtual reality surgical performance.[22,23] The simulator is also only applicable for right-handed participants. Left- and right-handed ergonomics have been shown to be different in virtual reality trials.[20] Hence, modifications to the Sim-Ortho platform are encouraged in order to allow participants to use both hands and capture a more holistic understanding of bimanual expertise. Third, multiple variables were controlled to simplify the interpretation of participants' surgical performance, including the specific instruments to be used in each component of the scenario. Fourth, the study involved a small a priori-defined sample size from a single institution. Hence, it is difficult to confidently extend our results to larger populations. Prospective testing of the neural network with a large sample size from multiple institutions is required to assess its accuracy and generalizability.

CONCLUSION

This study achieved all 3 of our objectives: to develop performance metrics, to employ artificial neural networks to classify participants' expertise, and to outline the relative weights of specific metrics. We employed a new virtual reality simulator to develop novel performance metrics for an ACDF procedure. A robust artificial neural network was designed to classify 3 groups of participants based on expertise with 83.3% accuracy. Insight into the relative importance of specific metrics of performance was outlined. The novel methodology and results presented have

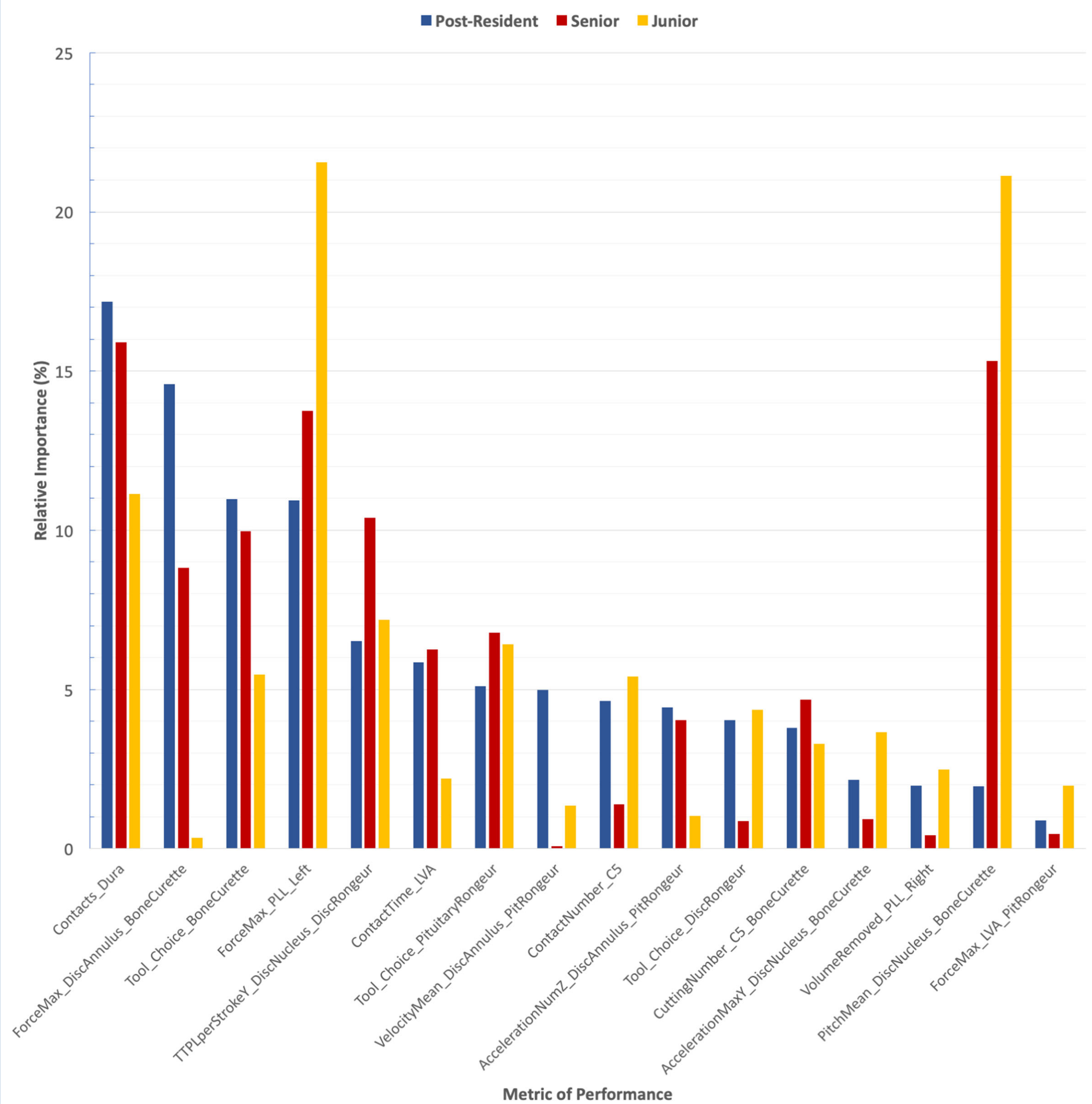


FIGURE 6. *Relative importance of each metric of performance in the classification of participants between 3 groups. The relative metric importance corresponds to the magnitude of the connection weight product for each metric of performance and its corresponding group: post-resident (blue), senior (red), and junior (yellow). A higher importance indicates that fluctuations in the respective metric have a larger influence on the neural network's decision to classify a participant in the respective group. The metrics are ranked from left to right according to descending importance for the post-resident group.*

the potential to aid in the understanding of components of surgical expertise and contribute to the paradigm shift towards competency-based training for surgery.

Disclosures

This work was supported by the AO Foundation, the Di Giovanni Foundation, the Montreal Neurologic Institute and Hospital, and the Department of Orthopaedic Surgery at McGill University. The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg*. 2005;241(2):364.

2. Delorme S, Laroche D, DiRaddo R, Del Maestro RF. NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training. *Oper Neurosurg*. 2012;71(suppl_1):ons32-ons42.
3. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638-645.
4. 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.
5. Fountas KN, Kapsalaki EZ, Nikolakakos LG, et al. Anterior cervical discectomy and fusion associated complications. *Spine*. 2007;32(21):2310-2317.
6. Azarnoush H, Alzhrani G, Winkler-Schwartz A, et al. Neurosurgical virtual reality simulation metrics to assess psychomotor skills during brain tumor resection. *Int J CARS*. 2015;10(5):603-618.
7. Alotaibi FE, AlZhrani GA, Sabbagh AJ, Azarnoush H, Winkler-Schwartz A, Del Maestro RF. Neurosurgical assessment of metrics including judgment and dexterity using the virtual reality simulator NeuroTouch (NAJD Metrics). *Surg Innov*. 2015;22(6):636-642.
8. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955. *AI mag*. 2006;27(4):12.
9. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: a review of classification techniques. *Emerg artificial intel appl comput eng*. 2007;160:3-24.
10. Haykin S. *Neural Networks*. Vol 2. New York: Prentice hall; 1994.
11. Vedula SS, Ishii M, Hager GD. Objective assessment of surgical technical skill and competency in the operating room. *Annu Rev Biomed Eng*. 2017;19:301-325.
12. Bissonnette V, Mirchi N, Ledwos N, et al. Artificial intelligence distinguishes surgical training levels in a virtual reality spinal task. *J Bone Joint Surg Am*. published online: 2019 (doi:10.2106/JBJS.18.01197).
13. 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 Netw Open*. 2019;2(8):e198363.
14. Richstone L, Schwartz MJ, Seideman C, Cadeddu J, Marshall S, Kavoussi LR. Eye metrics as an objective assessment of surgical skill. *Ann Surg*. 2010;252(1):177-182.
15. Ruikar DD, Hegadi RS, Santosh K. A systematic review on orthopedic simulators for psycho-motor skill and surgical procedure training. *J Med Syst*. 2018;42(9):168.
16. 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. *J Surg Educ*. published online: 2019 (doi: 10.1016/j.jsurg.2019.05.015).
17. Riffaud L, Neumuth T, Morandi X, et al. Recording of surgical processes: a study comparing senior and junior neurosurgeons during lumbar disc herniation surgery. *Oper Neurosurg*. 2010;67(suppl_2):ons325-ons332.
18. Olden JD, Jackson DA. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecol Model*. 2002;154(1-2):135-150.
19. AlZhrani G, Alotaibi F, Azarnoush H, et al. Proficiency performance benchmarks for removal of simulated brain tumors using a virtual reality simulator NeuroTouch. *J Surg Educ*. 2015;72(4):685-696.
20. Livieris IE, Drakopoulou K, Pintelas P. Predicting students' performance using artificial neural networks. Paper Presented at: 8th PanHellenic Conference with International Participation Information and Communication Technologies in Education, Volos, Greece; 2012.
21. Sweller J. Cognitive load theory, learning difficulty, and instructional design. *Learn Instr*. 1994;4(4):295-312.
22. Azarnoush H, Siar S, Sawaya R, et al. The force pyramid: a spatial analysis of force application during virtual reality brain tumor resection. *J Neurosurg*. 2017;127(1):171-181.
23. Sawaya R, Alsideiri G, Bugdadi A, et al. Development of a performance model for virtual reality tumor resections. *J Neurosurg*. 2018;1(aop):1-9.

Acknowledgments

We thank all the neurosurgical and orthopaedic residents, spine fellows, and spine surgeons from the Montreal Neurological Institute and Hospital and McGill University who participated in this study. The authors would like to acknowledge the contributions of OSSimTech™ for ongoing improvement to their Sim-Ortho simulator.

Supplemental digital content *is available for this article at www.operativeneurosurgery-online.com.*

Supplemental Digital Content 1. Methods. Simulated anterior cervical discectomy scenario.
Supplemental Digital Content 2. Methods. Metric selection.
Supplemental Digital Content 3. Methods. Artificial neural network design.
Supplemental Digital Content 4. Methods. Neural network optimization.
Supplemental Digital Content 5. Methods. Metric importance calculation.
Supplemental Digital Content 6. Figure. Connection weight products of each group for metrics of performance for the virtual reality surgical simulation. The magnitude of the connection weight product represents the relative importance of the corresponding metric in the neural network's decision to classify as participant in the corresponding group: post-resident (blue), senior (red), and junior (yellow). The sign of the connection weight product indicates whether a metric's z-score value should be positive (if sign if positive) or negative (if sign is negative) to increase the likelihood of classification in the corresponding group.