Copyright © 2019 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated

Topics in Training

Downloaded from https://journals.lww.com/jbjsjournal by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 12/06/2019

Artificial Intelligence Distinguishes Surgical Training Levels in a Virtual Reality Spinal Task

Vincent Bissonnette, MD*, Nykan Mirchi, BSc*, Nicole Ledwos, BA, Ghusn Alsidieri, MD, MSc, Alexander Winkler-Schwartz, MD, and Rolando F. Del Maestro, MD, PhD, on behalf of the Neurosurgical Simulation & Artificial Intelligence Learning Centre†

Investigation performed at the Neurosurgical Simulation & Artificial Intelligence Learning Centre, Department of Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada

Background: With the emergence of competency-based training, the current evaluation scheme of surgical skills is evolving to include newer methods of assessment and training. Artificial intelligence through machine learning algorithms can utilize extensive data sets to analyze operator performance. This study aimed to address 3 questions: (1) Can artificial intelligence uncover novel metrics of surgical performance? (2) Can support vector machine algorithms be trained to differentiate "senior" and "junior" participants who are executing a virtual reality hemilaminectomy? (3) Can other algorithms achieve a good classification performance?

Methods: Participants from 4 Canadian universities were divided into 2 groups according to their training level (senior and junior) and were asked to perform a virtual reality hemilaminectomy. The position, angle, and force application of the simulated burr and suction instruments, along with tissue volumes that were removed, were recorded at 20-ms intervals. Raw data were manipulated to create metrics to train machine learning algorithms. Five algorithms, including a support vector machine, were trained to predict whether the task was performed by a senior or junior participant. The accuracy of each algorithm was assessed through leave-one-out cross-validation.

Results: Forty-one individuals were enrolled (22 senior and 19 junior participants). Twelve metrics related to safety of the procedure, efficiency, motion of the tools, and coordination were selected. Following cross-validation, the support vector machine achieved a 97.6% accuracy. The other algorithms achieved accuracy of 92.7%, 87.8%, 70.7%, and 65.9%, respectively.

Conclusions: Artificial intelligence defined novel metrics of surgical performance and outlined training levels in a virtual reality spinal simulation procedure.

Clinical Relevance: The significance of these results lies in the potential of artificial intelligence to complement current educational paradigms and better prepare residents for surgical procedures.

*Vincent Bissonnette, MD, and Nykan Mirchi, BSc, contributed equally to this work.

†A list of the Neurosurgical Simulation & Artificial Intelligence Learning Centre group members is included as a note at the end of the article.

Disclosure: This work was supported by the Di Giovanni Foundation. Internal support was provided by the Montreal Neurological Institute and Hospital and the McGill Department of Orthopaedics. On the **Disclosure of Potential Conflicts of Interest** forms, *which are provided with the online version of the article,* one or more of the authors checked "yes" to indicate that the author had a relevant financial relationship in the biomedical arena outside the submitted work; and "yes" to indicate that the author had a patent and/or copyright, planned, pending, or issued, broadly relevant to this work (http://links.lww.com/JBJS/F488).

Copyright © 2019 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

With the shift toward competency-based curricula, surgical educational paradigms are evolving to include new methods of assessment and training. Whereas current assessments rely on subjective methods, new technologies offer the potential for more objective approaches to surgical skill evaluation[1]. Simulation has become important in surgical education, with many programs implementing courses involving animal models, cadavers, benchtop models, and virtual reality simulators[2]. Virtual reality simulators provide opportunities for repeat practice in risk-free environments and can quantify multiple aspects of psychomotor performance during surgical procedures[3]. The large amount of data collected from an individual's technical performance during a simulated task can be distilled into specific metrics. Metrics can be considered standards of reference to quantitate performance, efficiency, and progress[4,5]. Individual metrics often are incapable of effectively assessing surgical expertise since many procedures involve multiple complex psychomotor skills. The requirement of efficiently combining multiple metrics has resulted in the need to assess systems that are capable of analyzing extensive amounts of information from multivariate data sets.

Artificial intelligence employs machine learning algorithms, giving computers the ability to identify patterns and perform tasks without explicit programming when sufficient data are provided[6,7]. Different types of machine learning algorithms exist. Supervised algorithms, including support vector machines, are utilized most commonly. These algorithms are trained with examples of labeled data and learn patterns associated with each label, giving them the ability to label new data[7]. In surgical simulation, supervised algorithms could be trained utilizing sets of metrics labeled as senior or junior, thereby allowing them to classify new individuals' metrics as senior or junior. This is referred to as 2-class learning. One-class learning (training algorithms to identify individuals belonging to 1 group [e.g., experts]) and multiclass learning (training algorithms to classify individuals in ≥2 groups [e.g., junior residents, senior residents, and staff surgeons]) also could be employed but would require a large number of participants in each group to adequately train the algorithms. As such, these techniques have not been widely utilized to assess psychomotor skills in this context[8].

The purpose of this study was to evaluate the potential of artificial intelligence as an assessment tool in virtual reality spine surgery simulation. We aimed to provide a preliminary proof of concept that could act to introduce artificial intelligence as a mechanism to objectively assess surgical skill level. We addressed 3 questions in this investigation: (1) Can artificial intelligence uncover novel metrics of surgical performance that differentiate between 2 groups of different training levels? (2) Can support vector machine algorithms be trained to recognize whether an individual executing a virtual reality hemilaminectomy is

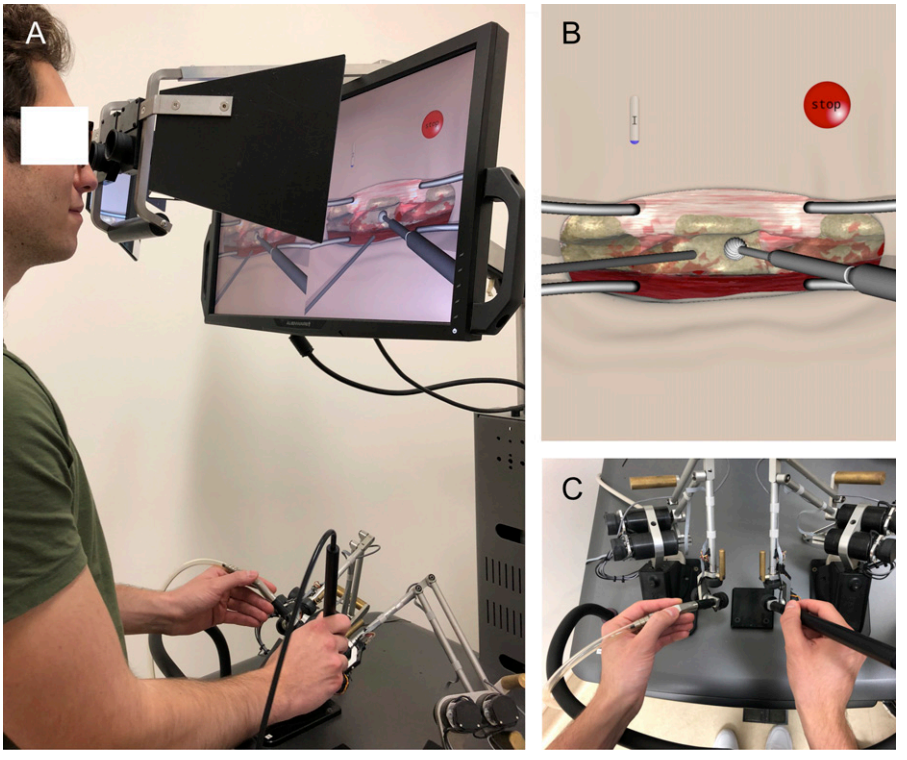

Fig. 1

Figs. 1-A, 1-B, and 1-C Demonstration of the NeuroVR platform. **Fig. 1-A** An individual performing the virtual hemilaminectomy scenario. **Fig. 1-B** Virtual tissues include L2, L3, and L4 vertebrae, interspinous ligaments, surrounding muscles, ligamentum flavum, intervertebral discs, and dura. **Fig. 1-C** The participant must hold the burr in his or her dominant hand and the suction instrument in the nondominant hand.

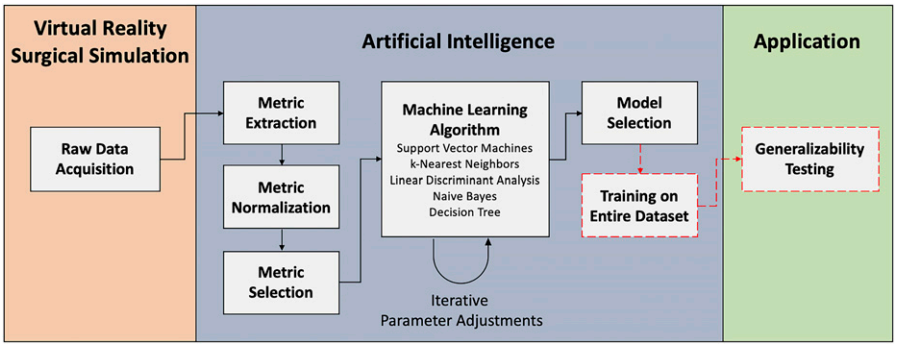

Fig. 2

A framework for integrating artificial intelligence in virtual reality surgical simulation. The virtual reality surgical simulation section involves raw data acquisition from the simulator. Machine learning methodology is followed by performing metric extraction, normalization, and selection. The selected metrics are fed to a collection of machine learning algorithms, and an iterative process of parameter adjustment is followed to optimize classification accuracy. This step uses cross-validation techniques to assess classification accuracy. Once the optimal algorithm and parameters are identified, a single model is trained using all of the data. This model can then be used for generalizability testing on new subjects.

of senior or junior level? (3) Can other algorithms achieve a good classification performance (accuracy >75%)?

Materials and Methods

Spine surgeons, spine fellows, orthopaedic and neurosurgery residents, and medical students from 4 Canadian universities were recruited. Because this investigation aimed to provide an initial proof of concept of the utility of machine learning as an assessment tool, we employed simple 2-class learning algorithms. Thus, 2 groups of different expertise level had to be defined a priori. Participants were divided into senior (postgraduate year [PGY]-4 and above) and junior (PGY-3 and below) groups because our group of surgeons considered that the simulated procedure required basic burr and suction instrument-handling skills that should be acquired by the fourth year of orthopaedic and neurosurgery training.

All participants signed a consent form that was approved by McGill University Health Centre Research Board before entering the study. The NeuroVR neurosurgical simulator (CAE Healthcare) virtual reality platform, which incorporates a microscopic view and haptic feedback, was employed to perform a left L3 hemilaminectomy[9]. This platform includes numerous simulated surgical scenarios that have been studied extensively[10-13]. As demonstrated in Video 1, the virtual hemilaminectomy required participants to remove the L3 lamina with a simulated burr in their dominant hand while controlling bleeding with a simulated suction instrument in their nondominant hand (Figs. 1-A, 1-B, and 1-C). Participants were given verbal and written instructions to remove the L3 lamina without damaging surrounding tissues. Subjects had 5 minutes to complete the task because this amount of time was found to be adequate in preliminary studies. Each participant performed the task once without prior practice. Individuals participated in the trial at a single time point without follow-up. The trial was conducted in an experimental setting that was void of distractions.

TABLE I Description of the Mechanisms of the 5 Employed Machine Learning Algorithms

Machine Learning Algorithm	Mechanism*
Support vector machine	Uses a hyperplane to separate data in ≥2 groups and maximizes the distance between the closest points from both groups and the hyperplane
Linear discriminant analysis	Projects multidimensional data (many metrics) on a single dimension to maximize the distance between the means of the groups and minimize the variance within each group
k-nearest neighbors	Uses distance functions such as the Euclidean distance to determine the closest neighbors to a point. A parameter (k) corresponds with the number of neighbors considered. The class of a participant is determined on the basis of its relationship with the nearest participants in a multidimensional space
Naive Bayes	Classifies participants on the basis of probabilities that the chosen metrics belong to experts or novice surgeons. It assumes that all of the chosen metrics are independent from each other
Decision tree	Classifies individuals by building a series of nodes whereby subjects are divided according to the value of a certain metric. The algorithm finds the optimal values to divide subjects in classes

*The mechanism of every algorithm is discussed further in the literature[7,17-21].

TABLE II Distribution of the Studied Sample Population in Regard to Training Level and Specialty*

Training Level	Orthopaedic Surgery *(no.)*	Neurosurgery *(no.)*	Total *(no.)*
Spine surgeons	1	5	6
Spine fellows	2	1	3
PGY-6	N/A	2	2
PGY-5	3	1	4
PGY-4	3	4	7
Total senior	8	8	22
PGY-3	1	1	2
PGY-2	3	2	5
PGY-1	2	2	4
Medical students	N/A	N/A	8
Total junior	6	5	19

*N/A = not applicable, and PGY = postgraduate year.

Artificial intelligence methodology was applied through a series of steps, including raw data acquisition, metric extraction, metric normalization, metric selection, machine learning algorithms, and model selection (Fig. 2). These methods follow guidelines to utilize machine learning algorithms to assess surgical expertise in simulation that previously had been established by our group[14].

Raw Data Acquisition

The position, angle, and force of both simulated instruments, along with the removed volume of all simulated tissues, were captured at 20-ms intervals and were exported to a file.

Metric Extraction

A metric is an input that is used to train a machine learning algorithm to predict whether a participant belongs to the senior or junior group. The accuracy of an algorithm can be defined as the number of good predictions out of the total number of predictions made. To obtain the best accuracy and to reduce computational cost, metrics given to algorithms must be carefully processed[15].

The raw variables provided by the NeuroVR can be combined to generate more complex metrics. For instance, by combining tooltip position and time, velocity can be assessed. A series of functions was developed to extract metrics from the raw data using MATLAB R2018a (MathWorks). Metrics were divided into 4 categories, including safety, efficiency, coordination, and motion[4,13]. Since metrics of varying scales were generated, data normalization was performed with z-scores.

Metric Selection

Metric selection is an important step in machine learning that attempts to find the combination of metrics that most accurately differentiates between the 2 groups[16]. This step is vital to prevent the algorithm from receiving irrelevant input, thereby avoiding the training of algorithms that are too closely "fitted" to a specific data set and tend to generalize poorly to new subjects[17].

In this study, metric selection was performed in 2 parts. First, to capture metrics that are clinically relevant, 2 spine surgeons selected metrics that they felt could differentiate between the 2 groups through a questionnaire (Appendix A). Second, since these metrics may not all adequately discriminate between the 2 groups in this scenario, a backward selection algorithm from PRtools (http://prtools.org/) was employed. This backward algorithm started with all of the metrics chosen by spine surgeons and removed them sequentially while iteratively training a machine learning algorithm and testing its accuracy using 10-fold cross-

TABLE III Number of Laminectomy Cases in Which Each Resident Assisted*

Participant Level	Number of Laminectomy Cases in Which the Resident Assisted
Junior orthopaedics	
PGY-1†	0
PGY-1†	0
PGY-2†	3
PGY-2†	6
PGY-2†	N/A
PGY-3†	25
Median	3
Junior neurosurgery	
PGY-1†	3
PGY-1†	15
PGY-2†	N/A
PGY-2†	N/A
PGY-3†	3
Median	3
Senior orthopaedics	
PGY-4†	4
PGY-4†	3
PGY-4†	20
PGY-5†	50
PGY-5†	10
PGY-5†	N/A
Median	20
Senior neurosurgery	
PGY-4‡	50
PGY-4‡	60
PGY-4‡	80
PGY-4‡	100
PGY-5‡	75
PGY-6†	30
PGY-6†	40
Median	60

*PGY = postgraduate year, and N/A = not applicable. †University A. ‡University B.

TABLE IV Initial Metrics Selected by 2 Spine Surgeons

Safety
Mean force applied on ligamentum flavum
Maximum force applied on ligamentum flavum
Mean force applied on dura
Maximum force applied on dura
Volume of ligamentum flavum removed
Number of times dura was touched with an active burr
Minimum and maximum position of the burr in the cephalad-caudad axis while removing L3
Minimum and maximum position of the burr in the medial-lateral axis while removing L3
Efficiency
Position of the burr when the first removal of L3 occurs
Idle time (amount of time no force is applied by any tool on any structure)
Total tip path length of the burr (sum of every change in position)
Total tip path length of the suction (sum of every change in position)
Amount of time spent removing L3/total time to completion
Time to completion
Coordination
Volume removed while simultaneously using the suction and the burr
Mean velocity of the suction while simultaneously using the burr
Number of times structures are touched with suction while using the burr
Amount of time spent while simultaneously using the suction and the burr
Mean distance between the tip of the burr and the tip of the suction
Motion of the tools
Variance of angles of the burr when removing L3
Consistency of movements (distance between 2 acceleration peaks for the burr when removing L3 in 3 different axes)
Consistency of movements (distance between 2 acceleration peaks for the suction instrument when removing L3 in 3 different axes)
Mean acceleration of the burr over the whole procedure in 3 different axes
Mean acceleration of the suction over the whole procedure in 3 different axes
Mean velocity of the burr when removing ligamentum flavum
Maximum velocity of the burr when removing ligamentum flavum

validation[16]. The backward algorithm stopped when a combination of metrics provided the highest accuracy of classifying senior individuals as senior and junior individuals as junior. Metrics that were not selected were not further analyzed.

Machine Learning Algorithms

Support vector machines are suited for small sample size and multivariate data that are necessary for global evaluation of surgical skill, thereby making them a prime candidate for virtual reality surgical simulation[7,17,18]. Furthermore, their decision-making process is describable. In a manner similar to the coefficients in a linear logistic regression, these algorithms attribute a weight to each metric and make their classification on the basis of an equation that considers every metric and its respective weight. This is interesting from an educational perspective because it could help juniors to understand what they need to improve to achieve the senior level. These factors led us to focus on this algorithm. Four other algorithms (k-nearest neighbors, linear discriminant analysis, naive Bayes, and decision tree) were also trained to assess whether the selected metrics could achieve a similar accuracy with diverse classification methods. The mechanism of each algorithm is explained in Table I. Additional information is available in the literature[7,17-21].

Because our sample size was relatively limited, leave-one-out cross-validation was employed to train and test the algorithms[19]. Leave-one-out cross-validation trains the algorithm with all but 1 of the participants, and subsequently tests the trained algorithm on the 1 participant who was left out of the training set. This process is repeated with every participant;

TABLE V Final Metrics Selected by Metric Selection Algorithm

Metric	Senior/Junior Ratio
Safety	
Maximum force applied on dura	0.56
Efficiency	
Amount of time spent removing L3/total time to completion	0.96
Coordination	
Amount of time spent while using suction and burr at the same time	1.73
Number of times structures are touched with suction while using the burr	2.18
Motion of the tools	
Distance between 2 acceleration peaks for the burr in the cephalad-caudad axis when removing L3 (consistency of movements of the burr)	1.48
Distance between 2 acceleration peaks for the suction in the medial-lateral axis when removing L3 (consistency of movements of the suction)	0.99
Mean acceleration of the burr in the anterior-posterior axis	0.61
Mean acceleration of the burr in the medial-lateral axis	0.73
Mean acceleration of the suction in the medial-lateral axis	0.46
Mean velocity of the burr when removing ligamentum flavum	1.16
Maximum velocity of the burr when removing ligamentum flavum	0.87
Variance of the pitch angle of the burr when removing L3	0.34

hence, in our case, the process was repeated 41 times. Because algorithms are built according to various parameters, these were adjusted in an iterative manner to optimize classification accuracy.

Metric Analysis

To analyze the performance of senior and junior participants, the ratio of the average metric score for senior and junior participants (the fold difference) was calculated for each metric.

Results

Twenty-two senior participants (6 spine surgeons, 3 spine fellows, and 13 senior residents) and 19 junior participants (11 junior residents and 8 medical students) were recruited. The distribution of the participants' training level and specialty is presented in Table II. The number of laminectomy cases in which each resident assisted is outlined in Table III. Forty-one metrics were generated. Of these, 36 metrics were selected by spine surgeons and are presented in Table IV. The backward algorithm identified 12 final metrics, which are listed in Table V. Eight metrics relate to motion of the tools, and 4 relate to safety, efficiency, and coordination. The maximum force applied on dura is lower in the senior group (fold difference: 0.56). The amount of time spent while simultaneously using the burr and suction instruments was higher for senior participants (fold difference: 1.73). The senior participants touched adjacent structures more with their suction instrument while removing L3 with the burr (fold difference: 2.18). The ratio of the amount of time spent removing L3 to the total time of the procedure was similar in both groups (fold difference: 0.96). Finally, senior participants displayed slower deceleration overall, showed higher delays between 2 consecutive accelerations while removing L3, and exhibited less variance in the pitch angle of the burr when they removed L3.

Using leave-one-out cross-validation, 5 algorithms were assessed. The support vector machine achieved the highest accuracy, at 97.6%. The k-nearest neighbors, linear discriminant analysis, decision tree, and naive Bayes had 92.7, 87.8, 70.7, and 65.9% accuracy, respectively (Fig. 3).

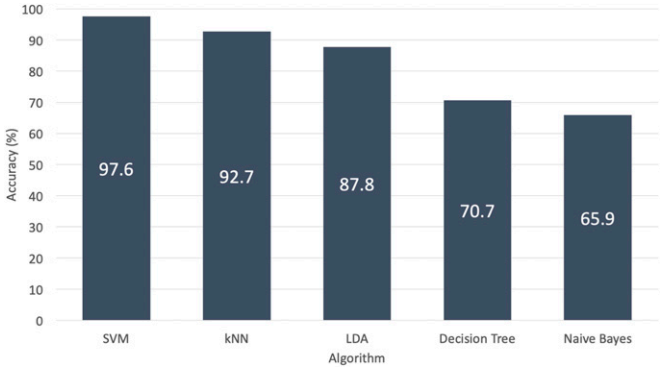

Fig. 3

The support vector machine (SVM) achieved the highest accuracy, at 97.6%, with use of leave-one-out cross-validation. The k-nearest neighbors (kNN) reached an accuracy of 92.7%. The linear discriminant analysis (LDA) achieved an accuracy of 87.8%. The decision tree had a 70.7% accuracy. The naive Bayes reached the lowest accuracy, at 65.9%.

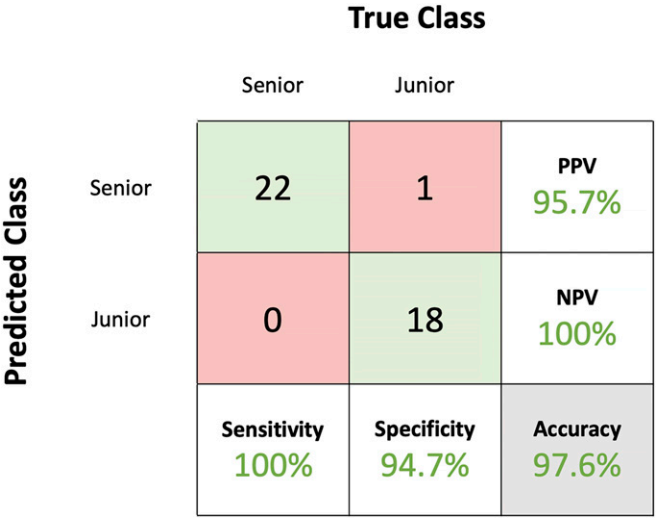

Fig. 4

Using leave-one-out cross-validation, the support vector machine classified senior participants with a sensitivity of 100% and junior participants with a specificity of 94.7%. The obtained positive predictive value (PPV) was 95.7%, and the negative predictive value (NPV) was 100%. The algorithm achieved an overall classification accuracy of 97.6%.

A confusion matrix was produced for the support vector machine algorithm (Fig. 4). Only 1 junior surgeon was misclassified.

Discussion

Machine learning algorithms have defined novel metrics of surgical performance in a virtual reality spinal task. This addresses our first research question.

The 4 areas of surgical skill that had been identified were represented in the 12 metrics that were selected. From a safety perspective, the senior group restricted the force applied on the dura. This is an important metric to teach, considering that applying high forces on the dura may increase the risk of dural tear. The senior participants also used their tools simultaneously more often than the junior participants. This shows the importance of the acquisition of bimanual skills in spine surgery. Furthermore, the senior participants displayed less angle variance with the burr when removing L3 and higher delays between 2 acceleration peaks, which provides new insights on the consistency of their movements. These results support that surgical skill is multifaceted and may benefit from teaching based on metrics that embody different aspects of surgical performance.

An automated feedback system was created with these metrics. Future participants will be able to see their scores on each of the metrics, as well as a global classification of the surgical training level (junior or senior). In addition, they will individually be guided to improve their skills through video-based and auditory feedback, which attempts to mimic current training in the operating room whereby surgeons explain what to improve and demonstrate how to do it.

We addressed the second research question by training a support vector machine algorithm with 12 metrics to classify senior and junior participants performing a virtual reality spine procedure. The advantage of applying machine learning to our multivariate data set is that it provides a more objective and holistic assessment of psychomotor performance.

As a support vector criterion was employed to select metrics, the final metrics were likely best performed with support vector machine algorithms. To evaluate the ability of these metrics to differentiate training level, other algorithms were trained with the same metrics. Two other algorithms (the k-nearest neighbors and the linear discriminant analysis) displayed accuracies of >75%, thereby addressing the third outlined research question.

The subject who was misclassified by the support vector machine algorithm was a PGY-2. Although we cannot be certain that this misclassification is attributed to a higher set of skills, we analyzed this individual's metrics to understand this result. This individual applied less force on the dura, spent more time using both tools simultaneously, and displayed more consistency with the burr (less variance in pitch angle and larger distance between 2 acceleration peaks) than other juniors. These results suggest that this individual's performance was more consistent with the expected performance of the senior group.

Participants were from multiple institutions and 2 specialties (neurosurgery and orthopaedics), making these data more representative of different training paradigms. Incorporating residents from both specialties allows the platform to have the potential to improve the standardization of spine training. However, this study was only an initial step to incorporate these technologies in residency training. It acts only as a proof of concept, and generalizability testing in a new population is required to ensure that the algorithm is not overfitted and to evaluate the platform's potential in training. This algorithm was trained according to residency training levels without explicit knowledge of surgical skill and has yet to be tested on an independent data set. Thus, it cannot be used to certify the proficiency of residents prior to independent practice, nor can it assess surgical skill level with certainty, but it may help with psychomotor skills acquisition.

There are limitations to employing machine learning in this simulated procedure. First, simulated burr and suction instruments are not representative of the many instruments and bimanual psychomotor skills that are employed during spine operations. Second, the visual and haptic complexities of the simulated procedure, the task duration, and the need to use a microscopic view may not adequately discriminate operator performance. More complex and realistic scenarios involving use of multiple instruments are currently being studied to address these issues. Third, although participants were asked to remove only the lamina, the lamina was not segmented separately from the spinous process and the facets. Therefore, the volume of lamina that was removed could not be determined. The new spine scenarios that are being developed are designed to segment all of the surrounding structures.

Defining participants' surgical skill level is difficult[22,23]. The number of surgical cases in which residents assist is often biased when reported by residents and may not reflect the skills acquired throughout their residency[24]. It was implied that senior residents had acquired the basic skills of using burr and suction instruments. Since spine training varies from 1 program to another and PGY-4 is a pivotal year in terms of surgical skill acquisition, efforts were made to understand whether the PGY-4 individuals should be included in the senior group. Thus, the study was repeated without incorporating the PGY-4 participants. The support vector machine algorithm achieved a 100% accuracy with 10 metrics, 6 of which are part of the 12 final metrics that have been described above. This is consistent with the concept that the psychomotor skills of PGY-4 participants in this study were more aligned with the senior group. However, assessment tools, such as the Objective Structured Assessment of Technical Skill, to evaluate residents' skills a priori may help to provide a better division of groups in the future[25]. Furthermore, if large numbers of spine surgeons are recruited, 1-class learning could be used to train algorithms to recognize expert performances and assess participants according to expert standards. This could provide a more robust evaluation of trainees' technical skill level.

To our knowledge, this is the first investigation employing machine learning to assess surgical expertise in a virtual reality spine procedure. Methods outlined in this study could be applied to any surgical simulation scenario provided that data on an individual's performance are collected. As virtual reality simulation becomes more realistic and more widely utilized, algorithms will become more robust. One could envision that once algorithms are rigorously validated to recognize expert surgeons, surgical accreditation bodies could employ these techniques to ensure their members' technical competency. The significance of this study lies in the potential of combining virtual reality simulation and artificial intelligence to provide safer training and objective assessment of surgical skills, which could lead to improved patient care.

Appendix

eA Supporting material provided by the authors is posted with the online version of this article as a data supplement at jbjs.org (http://links.lww.com/JBJS/F489). ■

NOTE: Members of the Neurosurgical Simulation & Artificial Intelligence Learning Centre group include Recai Yilmaz, MD, Samaneh Siyar, MSc, Hamed Azarnoush, PhD, Bekir Karlik, PhD, Robin Sawaya, MSc, Fahad E Alotaibi, MD, MSc, Abdulgadir Bugdadi, MD, MSc, Khalid Bajunaid, MD, MSc, MMgmt, Jean Ouellet, MD, and Greg Berry, MD, MSEd.

Vincent Bissonnette, MD[1,2]
Nykan Mirchi, BSc[1]
Nicole Ledwos, BA[1]
Ghusn Alsidieri, MD, MSc[1]
Alexander Winkler-Schwartz, MD[1]
Rolando F. Del Maestro, MD, PhD[1]

[1]Neurosurgical Simulation & Artificial Intelligence Learning Centre, Department of Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada

[2]Division of Orthopaedic Surgery, Montreal General Hospital, McGill University, Montreal, Quebec, Canada

Email address for V. Bissonnette: vincent.bissonnette@mail.mcgill.ca

ORCID iD for V. Bissonnette: 0000-0001-7448-2311
ORCID iD for N. Mirchi: 0000-0002-3742-4611
ORCID iD for N. Ledwos: 0000-0002-8604-3313
ORCID iD for G. Alsidieri: 0000-0002-5383-2105
ORCID iD for A. Winkler-Schwartz: 0000-0002-2110-9879
ORCID iD for R.F. Del Maestro: 0000-0003-3733-8921

References

1. Leong JJ, Leff DR, Das A, Aggarwal R, Reilly P, Atkinson HD, Emery RJ, Darzi AW. Validation of orthopaedic bench models for trauma surgery. J Bone Joint Surg Br. 2008 Jul;90(7):958-65.

2. Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. N Engl J Med. 2006 Dec 21;355(25):2664-9.

3. Bartlett JD, Lawrence JE, Stewart ME, Nakano N, Khanduja V. Does virtual reality simulation have a role in training trauma and orthopaedic surgeons? Bone Joint J. 2018 May 1;100-B(5):559-65.

4. Azarnoush H, Alzhrani G, Winkler-Schwartz A, Alotaibi F, Gelinas-Phaneuf N, Pazos V, Choudhury N, Fares J, DiRaddo R, Del Maestro RF. Neurosurgical virtual reality simulation metrics to assess psychomotor skills during brain tumor resection. Int J Comput Assist Radiol Surg. 2015 May;10(5):603-18. Epub 2014 Jun 27.

5. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann Surg. 2005 Feb;241(2):364-72.

6. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence. AI Mag. 2006 Dec;27(4):12-4.

7. Kotsiantis SB. Supervised machine learning: a review of classification techniques. Informatica. 2007;31:249-68.

8. Vedula SS, Ishii M, Hager GD. Objective assessment of surgical technical skill and competency in the operating room. Annu Rev Biomed Eng. 2017 Jun 21;19(1):301-25. Epub 2017 Mar 27.

9. Delorme S, Laroche D, DiRaddo R, Del Maestro RF. NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training. Neurosurgery. 2012 Sep;71(1)(Suppl Operative):32-42.

10. 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 Dec;22(6):636-42. Epub 2015 Apr 7.

11. Bajunaid K, Mullah MA, Winkler-Schwartz A, Alotaibi FE, Fares J, Baggiani M, Azarnoush H, Christie S, Al-Zhrani G, Marwa I, Sabbagh AJ, Werthner P, Del Maestro RF. Impact of acute stress on psychomotor bimanual performance during a simulated tumor resection task. J Neurosurg. 2017 Jan;126(1):71-80. Epub 2016 Mar 11.

12. Sawaya R, Alsideiri G, Bugdadi A, Winkler-Schwartz A, Azarnoush H, Bajunaid K, Sabbagh AJ, Del Maestro R. Development of a performance model for virtual reality tumor resections. J Neurosurg. 2018 Jul 1:1-9. Epub 2018 Jul 1.

13. Alotaibi FE, AlZhrani GA, Mullah MAS, Sabbagh AJ, Azarnoush H, Winkler-Schwartz A, Del Maestro RF. Assessing bimanual performance in brain tumor resection with NeuroTouch, a virtual reality simulator. Neurosurgery. 2015 Mar;11(1)(Suppl 2):89-98; discussion 98.

14. Winkler-Schwartz A, Bissonnette V, Mirchi N, Ponnudurai N, Yilmaz R, Ledwos N, Siyar S, Azarnoush H, Karlik B, Del Maestro RF. Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. J Surg Educ. 2019 Jun 13:S1931-7204(19)30106-0. Epub 2019 Jun 13.

15. Ding S, Zhu H, Jia W, Su C. A survey on feature extraction for pattern recognition. Artif Intell Rev. 2012;37(3):169-80.

16. Ladha L. Feature selection methods and algorithms. Int J Comput Sci Eng. 2011 Jan;3(5):1787-97.

17. Deo RC. Machine learning in medicine. Circulation. 2015 Nov 17;132(20):1920-30.

18. Noble WS. What is a support vector machine? Nat Biotechnol. 2006 Dec;24(12):1565-7.

19. Bishop CM. The curse of dimensionality. In: Pattern recognition and machine learning. 1st ed. Springer Science+Business Media; 2006. p 33-8.

20. Ye J, Janardan R, Li Q. Two-dimensional linear discriminant analysis. Adv Neural Inf Process Syst. 2005;17(60):1569-76.

21. McCallum A, Nigam K. A comparison of event models for naive Bayes text classification. AAAI/ICML-98 Workshop. 1998:41-48.

22. Gélinas-Phaneuf N, Del Maestro RF. Surgical expertise in neurosurgery: integrating theory into practice. Neurosurgery. 2013 Oct;73(4)(Suppl 1):30-8.

23. Bhatti NI, Cummings CW. Competency in surgical residency training: defining and raising the bar. Acad Med. 2007 Jun;82(6):569-73.

24. McPheeters MJ, Talcott RD, Hubbard ME, Haines SJ, Hunt MA. Assessing the accuracy of neurological surgery resident case logs at a single institution. Surg Neurol Int. 2017 Sep 6;8:206.

25. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997 Feb;84(2):273-8.