JNS

CLINICAL ARTICLE

J Neurosurg 137:1160–1171, 2022

Assessment of learning curves on a simulated neurosurgical task using metrics selected by artificial intelligence

Nicole Ledwos, MSc,[1] Nykan Mirchi, MSc,[1] Recai Yilmaz, MD,[1] Alexander Winkler-Schwartz, MD,[1,3] Anika Sawni, BSc,[1] Ali M. Fazlollahi, MSc,[1] Vincent Bissonnette, MD, MSc,[1,2] Khalid Bajunaid, MD, MSc,[6] Abdulrahman J. Sabbagh, MBChB,[4,5] and Rolando F. Del Maestro, MD, PhD[1,3]

[1]Neurosurgical Simulation and Artificial Intelligence Learning Centre, Department of Neurology & Neurosurgery, Montreal Neurological Institute and Hospital, McGill University; [2]Division of Orthopaedic Surgery, Montreal General Hospital, McGill University; [3]Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada; [4]Division of Neurosurgery, Department of Surgery, College of Medicine, King Abdulaziz University; [5]Clinical Skills and Simulation Center, King Abdulaziz University; and [6]Department of Surgery, College of Medicine, University of Jeddah, Jeddah, Saudi Arabia

OBJECTIVE Understanding the variation of learning curves of experts and trainees for a given surgical procedure is important in implementing formative learning paradigms to accelerate mastery. The study objectives were to use artificial intelligence (AI)–derived metrics to determine the learning curves of participants in 4 groups with different expertise levels who performed a series of identical virtual reality (VR) subpial resection tasks and to identify learning curve differences among the 4 groups.

METHODS A total of 50 individuals participated, 14 neurosurgeons, 4 neurosurgical fellows and 10 senior residents (seniors), 10 junior residents (juniors), and 12 medical students. All participants performed 5 repetitions of a subpial tumor resection on the NeuroVR (CAE Healthcare) platform, and 6 a priori–derived metrics selected using the K-nearest neighbors machine learning algorithm were used to assess participant learning curves. Group learning curves were plotted over the 5 trials for each metric. A mixed, repeated-measures ANOVA was performed between the first and fifth trial. For significant interactions ($p < 0.05$), post hoc Tukey's HSD analysis was conducted to determine the location of the significance.

RESULTS Overall, 5 of the 6 metrics assessed had a significant interaction ($p < 0.05$). The 4 groups, neurosurgeons, seniors, juniors, and medical students, showed an improvement between the first and fifth trial on at least one of the 6 metrics evaluated.

CONCLUSIONS Learning curves generated using AI-derived metrics provided novel insights into technical skill acquisition, based on expertise level, during repeated VR-simulated subpial tumor resections, which will allow educators to develop more focused formative educational paradigms for neurosurgical trainees.

https://thejns.org/doi/abs/10.3171/2021.12.JNS211563

KEYWORDS neurosurgical simulation; surgical education; learning curves; virtual reality; machine learning; artificial intelligence

The subpial resection of brain tumors that are adjacent to critical cortical structures is a challenging operative procedure and one in which neurosurgical trainees are expected to acquire proficiency.[1] Technical errors in this complex bimanual skill include injury to adjacent normal tissues and hemorrhage from subpial vessels, which can result in significant patient morbidity.[1] Our group developed complex and realistic virtual reality (VR) tumor resection tasks to aid learners in mastering this skill.[2,3] We also used artificial intelligence (AI) to quantitate the multiple components of the bimanual psychomotor skills utilized to successfully perform complex

ABBREVIATIONS AI = artificial intelligence; VR = virtual reality.

SUBMITTED June 28, 2021. ACCEPTED December 9, 2021.

INCLUDE WHEN CITING Published online February 4, 2022; DOI: 10.3171/2021.12.JNS211563.

©AANS 2022, except where prohibited by US copyright law

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

tumor resection tasks.[4,5] VR and AI technology can be employed to further our understanding of how repetition aids trainees in improving surgical performance.[6] By plotting the process of learning over time on specific metrics, a series of learning curves can be developed that demonstrate a trainee's improvement.[7–9] Surgical learning curves are graphical representations of the relationship between how proficient surgeons are at an operative task, such as subpial resection, and the amount of experience they have in that designated bimanual performance. Trainee learning curves, at various stages of skill acquisition, may be useful to surgical educators in developing formative learning strategies since they can help deconstruct which skill components may be most useful to the rate of improvement.

VR simulator performance data provide an opportunity to assess individual learning curves and group learning curves for different levels of expertise in a standardized, no-risk environment.[8,9] VR surgical simulators can record immense data sets concerning surgical procedures, such as the subpial resection, and this technology has been used to develop new quantitative assessment metrics.[5,10–14] Our group has applied machine learning algorithms to these data sets to uncover specific surgical and operative factors associated with surgical performance.[4,5,10,11] A K-nearest neighbors algorithm achieved an accuracy of 90% using 6 performance features (metrics), allowing classification of surgical expertise with greater granularity and precision.[5] The determination of these 6 AI-derived metrics through machine learning methodologies provided a unique opportunity to develop and assess individual learning curves for each of the 4 groups assessed in this study: neurosurgeons, senior and junior neurosurgical trainees, and medical students.

We sought to test two hypotheses: that all groups studied would show improvement in their learning curves, generated using AI-derived metrics during a repeated VR subpial resection task, and that junior neurosurgical trainees and medical students would demonstrate more improvement than the other groups. The specific objectives were 1) to determine the learning curves of groups with different levels of expertise using the 6 metrics previously outlined by the K-nearest neighbors algorithm, and 2) to identify the specific differences in the learning curves among the 4 groups. To our knowledge, this is the first time that metrics for differentiating surgical expertise with a machine learning algorithm have been used to generate learning curves to assess expertise on a VR surgical simulator. The results from this study provide novel information that may contribute to an increased understanding of the continuum of learning through all levels of expertise.

Methods

Study Participants

Neurosurgeons, neurosurgical fellows and residents, and medical students who expressed interest in neurosurgery or who were rotating on the neurosurgical service at one Canadian university were invited to participate. Demographic data regarding participants' age, handedness, level of expertise, and prior use of surgical simulators were collected before trial participation. Data were recorded at a single time point; no follow-up was carried out. All procedures were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Declaration of Helsinki.[15] All participants signed a consent form before trial participation that was approved by the McGill University Health Centre Research Ethics Board, Neurosciences-Psychiatry. This study follows CONSORT-AI guidelines and the best practices for Machine Learning to Assess Surgical Experience (MLASE) reporting guidelines.[16,17]

NeuroVR Simulator and Simulation Scenario

The NeuroVR platform (CAE Healthcare), a high-fidelity VR neurosurgical simulator, provides 3D visual operative experience with haptic feedback (Fig. 1A and B).[3,18,19] The VR-simulated subpial practice scenario consists of a yellow band representing the cortical tumor, with an underlying blood vessel, covered with a simulated pial membrane incorporated into simulated normal white matter. Utilizing a simulated ultrasonic aspirator in the dominant hand and simulated bipolar forceps in the nondominant hand, participants carried out the subpial tumor resection 5 times (Fig. 1C). A demonstration of the task is featured in Video 1.

VIDEO 1. A trainee performs the simulated subpial resection on the NeuroVR simulator using bipolar forceps in the nondominant hand and a simulated ultrasonic aspirator in the dominant hand. Copyright Nicole Ledwos. Published with permission. Click here to view.

Participants were given 3 minutes to complete each simulated resection. These tasks were specifically designed to model a subpial brain tumor resection procedure as it would be performed in patients (Fig. 1D and E).[2] Participants were asked to complete a tumor resection while minimizing bleeding and injury to the surrounding simulated normal tissue (Fig. 1F). They were provided with written and verbal instructions. The NeuroVR platform recorded data, in 20-msec increments, on the utilization of the two simulated instruments, including instrument activation, instrument tip coordinates, instrument force applied, and tumor and surrounding tissue contact. The platform also recorded operative factors such as the amount of tumor resected, bleeding, and healthy tissue removal.

Performance Metrics

Six performance metrics utilized by the K-nearest neighbors algorithm to classify individuals into 4 levels of expertise with a 90% accuracy were employed in this study.[10] Although multiple algorithms were tested and several metrics were described in our previous paper, the authors chose the metrics employed by the most accurate machine learning model, the K-nearest neighbors algorithm. These 6 performance metrics were subcategorized into 1) metrics applicable to the entire scenario and 2) metrics specific to the tumor resection. Metrics applicable to the entire scenario included the mean instrument tip distance, median change in force of the aspirator, and

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

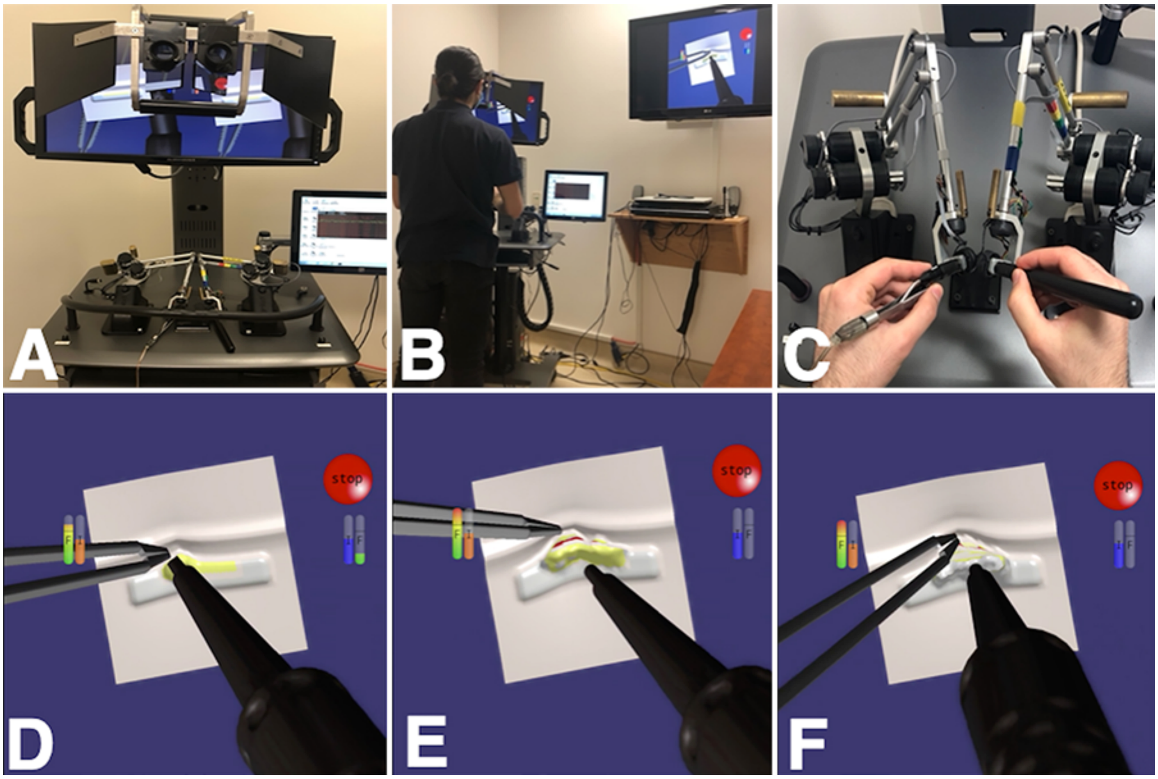

FIG. 1. A–C: Photographs of the NeuroVR simulator (A), a participant performing a simulated task on the NeuroVR (B), and the handles used for the subpial resection on the NeuroVR (C). **D–F:** Images showing the operative view of the simulated subpial scenario, including a participant elevating the pia mater with the bipolar and resecting tumor (D), the subpial scenario exposing the hidden simulated vessel behind the tumor (E), and a participant resecting simulated tumor (F). Panels D–F: Copyright Nicole Ledwos. Published with permission. Figure is available in color online only.

median velocity of the aspirator. Metrics specific to the tumor resection included the mean change in force of the aspirator, median velocity of the aspirator, and mean deceleration of the aspirator. These metrics were employed in this study to outline the learning curves of each of the 4 groups. Performances on trial 1 and trial 5 were compared between expertise groups.

Statistical Analysis

The first trial was considered the participants' baseline level of performance. Learning curves were generated and assessed by comparing the performance for each group between the first and the fifth trial for each of the 6 metrics. The Shapiro-Wilk test was conducted for each metric to check for normality of data distribution ($p > 0.05$). Significance was assessed using a mixed repeated-measures analysis of variance (mixed ANOVA) using IBM SPSS Statistics version 26 (IBM Corp.). For metrics with a significant interaction ($p < 0.05$), post hoc Tukey's HSD analysis was conducted in Matlab R2018a (The MathWorks, Inc.) to determine if there were differences within each group and between each group on trial 1 and trial 5. The Kruskal-Wallis nonparametric test was conducted to determine age differences between groups. In addition, the chi-square test was used to determine the association between groups and prior experience using a VR simulator. Both tests were run using IBM SPSS Statistics.

Results

Participants

Fifty participants were classified a priori into the 4 expertise groups: 14 neurosurgeons, 4 neurosurgical fellows and 10 senior neurosurgical residents in training years 4 to 6 (seniors), 10 junior neurosurgical residents in training years 1 to 3 (juniors), and 12 medical students. One neurosurgeon was excluded for not completing all trials. Participant demographic data regarding age, handedness, and level of expertise are displayed in Table 1. The neurosurgeon group was significantly older than all other groups, and the medical student group was significantly younger than seniors. Neurosurgeon subspecialization covered a wide range of practice, with most participants (8, 62%) primarily involved in cranial surgery (Table 1). A total of 25 individuals (51%) had previous experience using a surgical simulator. However, there was no association between groups and prior use of a simulator. The number of complete subpial tumor resections performed previously varied significantly between groups, with the neurosurgeon group having completed more subpial resections compared with juniors ($p = 0.008$) and seniors ($p = 0.002$).

Performance Metrics

The results of the mixed repeated-measures ANOVA indicated that 5 of the 6 metrics assessed had a significant

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

TABLE 1. Demographic data

	Neurosurgeons, n = 13	Seniors (neurosurgical fellows & residents, yrs 4–6), n = 14	Juniors (neurosurgical residents, yrs 1–3), n = 10	Medical Students, n = 12
Median age, yrs (range)	44 (33–59)	33 (29–35)	30 (27–38)	23 (23–26)
Handedness, n (%)				
Rt	12 (92)	12 (86)	9 (90)	10 (83)
Lt	1 (8)	2 (14)	1 (10)	2 (17)
Sex, n (%)				
M	13 (100)	13 (93)	8 (80)	6 (50)
F	0	1 (7)	2 (20)	6 (50)
Median no. of yrs of practice (range)	12 (1–25)	NA	NA	NA
Neurosurgical subspecialty, n (%)				
Spine	5 (38)	NA	NA	NA
Oncology & epilepsy	3 (23)	NA	NA	NA
Skull base	2 (15)	NA	NA	NA
Pediatrics	2 (15)	NA	NA	NA
Cerebrovascular	1 (8)	NA	NA	NA
Mean no. of subpial resections self-reported (SD)				
Partial	27.8 (82.3)	24.5 (37.8)	6 (14.2)	NA
Complete	213.1 (283.3)	11.2 (14.7)	1 (2.3)	NA
Previous experience, n (%)				
w/ a VR simulator	6 (46)	10 (71)	6 (60)	3 (25)
w/ the NeuroVR	6 (46)	9 (64)	6 (60)	1 (8)

NA = not applicable.

interaction term ($p < 0.05$), and post hoc testing indicated significant changes between groups in the same 5 metrics.

The mean instrument tip distance over the whole scenario (Fig. 2) demonstrated a significant interaction ($p = 0.019$). Post hoc testing revealed that only the medical student group had a significantly decreased mean instrument tip distance between trials 1 and 5 ($p = 0.014$). Neurosurgeon, senior, and junior groups had no significant differences between trial 1 and trial 5, indicating relatively stable tip distances across the 5 trials. On the first trial, the medical student group had a significantly larger mean instrument tip distance compared with the neurosurgeon ($p < 0.001$) and senior ($p < 0.001$) groups. On the fifth trial, the neurosurgeon group had a significantly smaller mean instrument tip distance compared with the junior ($p = 0.023$) and medical student ($p = 0.026$) groups.

The second metric assessed the median change in force of the aspirator over the whole scenario (Fig. 3) and showed a significant interaction ($p = 0.003$). Post hoc analysis demonstrated that the neurosurgeon ($p < 0.001$), senior ($p < 0.001$), and junior ($p = 0.012$) groups showed a significant increase in median aspirator change in force between trial 1 and trial 5. The neurosurgeon group had a significantly higher median aspirator change in force compared with all other groups on trial 5 (senior, $p = 0.001$; junior, $p = 0.033$; and medical student, $p < 0.001$).

The median aspirator velocity over the whole scenario (Fig. 4) had a significant interaction ($p = 0.018$). Post hoc analysis revealed that the neurosurgeon ($p = 0.028$) and junior ($p = 0.026$) groups had a significantly higher aspirator median velocity when comparing trial 1 and trial 5. The neurosurgeon group also had a significantly higher median velocity compared with the senior group on trial 1 ($p = 0.013$) and trial 5 ($p = 0.011$). The neurosurgeon group had a significantly higher median velocity compared with the medical student group on trial 5 ($p = 0.016$). The junior group also had a higher median velocity on trial 5 compared with both the senior ($p = 0.037$) and medical student ($p = 0.047$) groups.

The remaining 3 metrics were focused on aspirator force, velocity, and deceleration during the tumor resection procedure. The mean change in aspirator force while resecting tumor (Fig. 5) demonstrated a significant interaction ($p = 0.016$). Post hoc analysis revealed that the neurosurgeon ($p = 0.003$), senior ($p = 0.015$), and junior ($p = 0.033$) groups showed a significant increase between trial 1 and trial 5. Between-group analysis demonstrated that only the senior and medical student groups were significantly different on trial 1 ($p = 0.028$).

The final metric with a significant interaction was the median aspirator velocity while resecting tumor ($p = 0.003$) (Fig. 6). Both the neurosurgeon ($p = 0.049$) and junior ($p = 0.030$) groups showed an increase in the mean aspirator velocity between trial 1 and trial 5. On trial 1 and trial 5, the neurosurgeon group had a significantly higher median aspirator velocity compared with the senior ($p = 0.024$) and medical student ($p = 0.020$) groups. On trial 5, the neurosurgeon group had a higher median aspirator velocity compared with the senior ($p = 0.006$) and medical student ($p < 0.001$) groups.

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

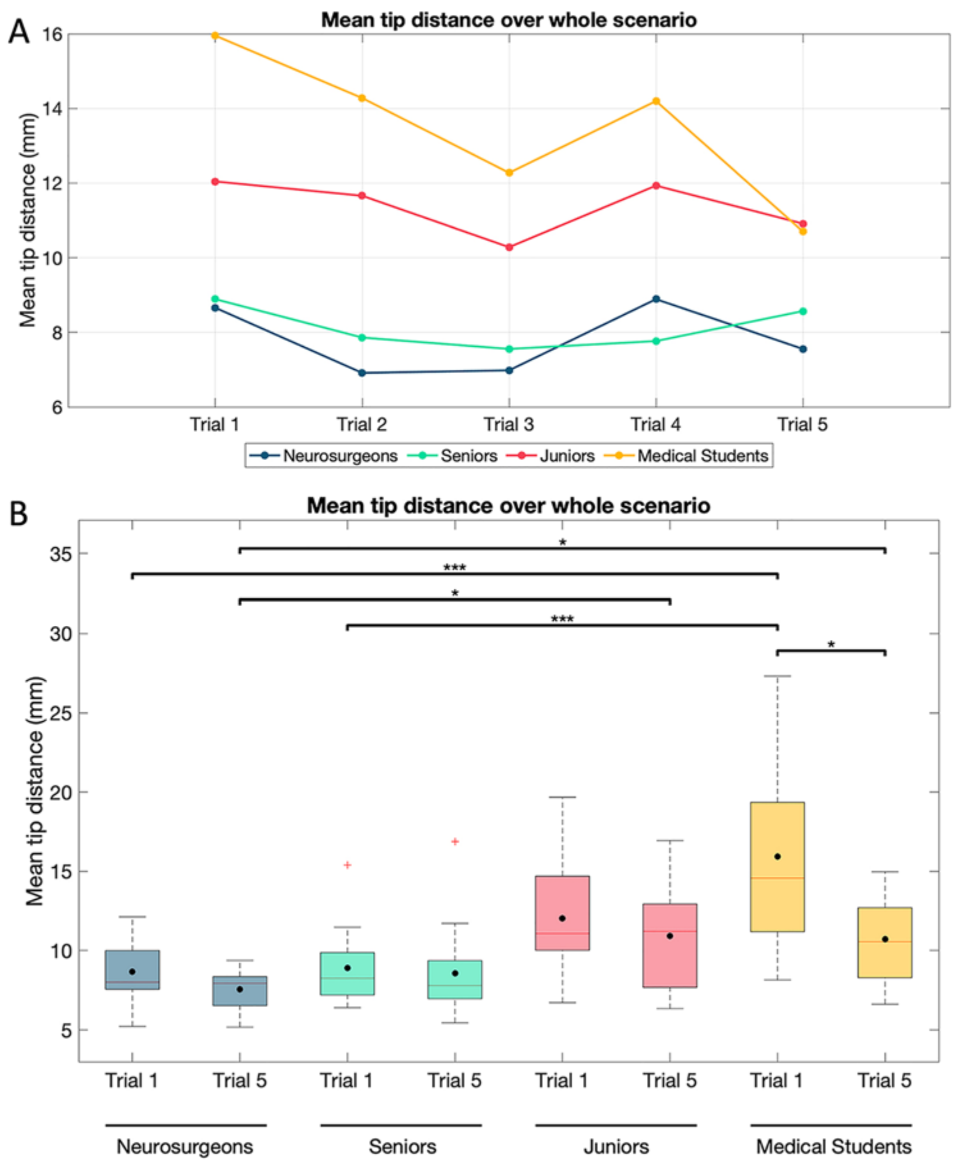

FIG. 2. A: Line graph showing the learning curves for each expertise level over 5 trials for the mean instrument tip distance over the whole scenario. **B:** Boxplot showing between-trial analysis and between-group analysis for the mean instrument tip distance over the whole scenario for each group: neurosurgeons (n = 13), seniors (n = 14), juniors (n = 10), and medical students (n = 12). The *plus sign* (+) represents an outlier; *black dots* indicate the mean and *red lines* indicate the median. $^{*}p < 0.05$; $^{***}p < 0.001$. Figure is available in color online only.

The mean aspirator deceleration while resecting tumor did not demonstrate a significant interaction term, and post hoc testing did not indicate significant changes between groups for this metric (Fig. 7).

Discussion

Our study confirms that previously employed metrics used by the K-nearest neighbors machine learning algorithm for expertise classification can be utilized to develop and assess the learning curves of individuals performing a VR-simulated neurosurgical task. Each of the 4 groups studied demonstrated a significant change in learning curve performance between trial 1 and trial 5 for at least one of the 6 stated metrics. These results support our hypothesis that all groups would show improvement during repeated performance of a VR subpial resection task using the AI-derived metrics. However, contrary to our second hypothesis, the neurosurgeon and junior groups demonstrated a significant change in learning curve performance between trial 1 and trial 5 in 4 of the 6 metrics. The senior group demonstrated improvement in 2 metrics. Although there was a trend to an increased median change in aspirator velocity over both the whole scenario and during tumor resection, this was not significant. The medical student group showed a significant change in performance in only 1 metric.

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

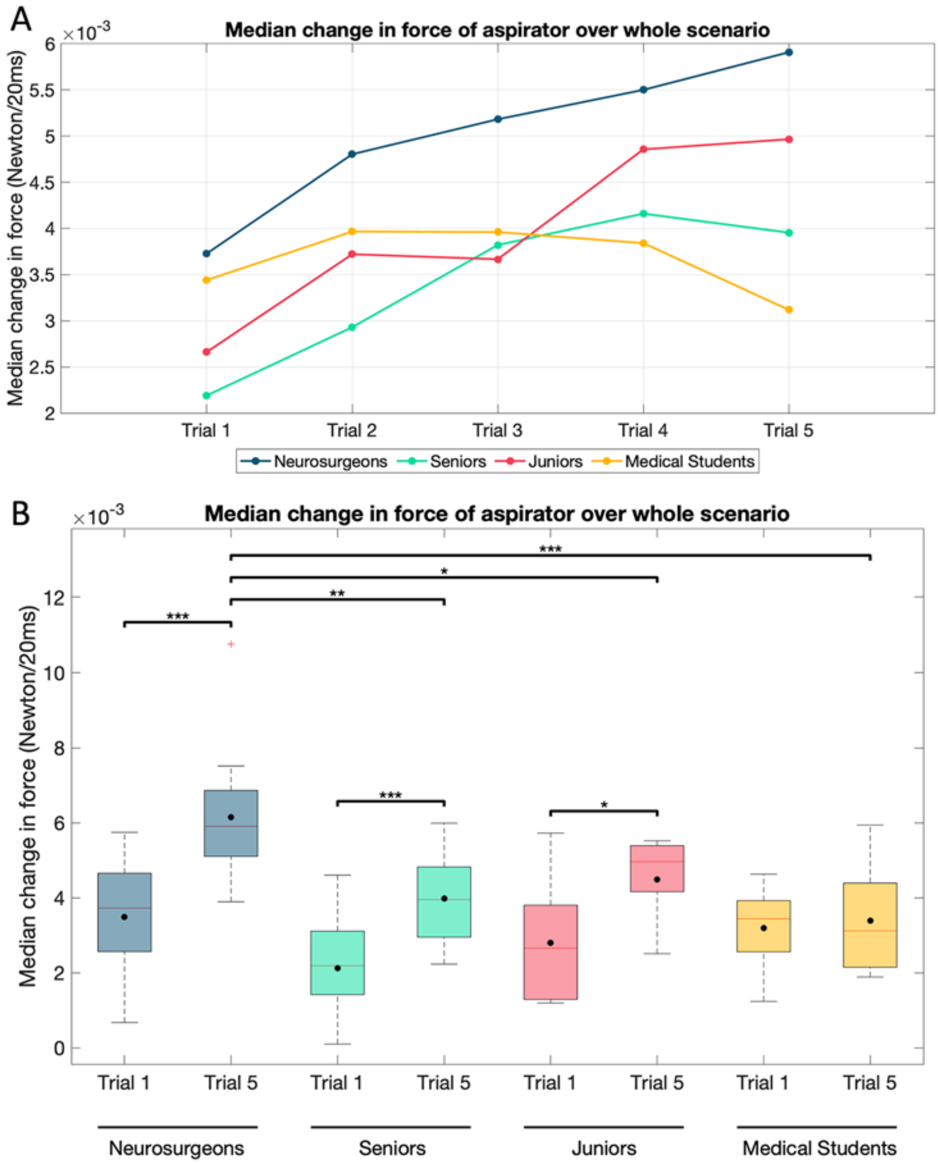

FIG. 3. A: Line graph showing the learning curves for each expertise level over 5 trials for the median aspirator change in force over the whole scenario. **B:** Boxplot showing between-trial analysis and between-group analysis for the median aspirator change in force over the whole scenario for each group: neurosurgeons (n = 13), seniors (n = 14), juniors (n = 10), and medical students (n = 12). The *plus sign* (+) represents an outlier; *black dots* indicate the mean and *red lines* indicate the median. *p < 0.05; **p < 0.01; ***p < 0.001. Figure is available in color online only.

The mean instrument tip distance over the whole scenario remained quite stable for the neurosurgeon, senior, and junior groups. Since this metric demonstrated stability across these three expertise levels, it supports the concept that a small mean instrument tip distance is an important component of expertise for this simulated task and may be acquired early during training. The medical student group had a significantly reduced mean instrument tip distance during self-directed learning between trial 1 and trial 5, consistent with the implied important role for this metric task performance. This suggests that this metric may be useful in the training of bimanual technical skills for novice and less-skilled individuals, and that the smaller mean instrument tip values seen in the neurosurgeon and senior groups appear to be optimal (Fig. 2).

Neurosurgeon, senior, and junior groups demonstrated a significant mean change in the force of the aspirator metric during tumor resection over the whole scenario. For these 3 groups, a significant increase in aspirator force was found between trial 1 and trial 5, which supports the concept that an increasing mean aspirator force, as a dynamic psychomotor skill change, is associated with expertise and self-directed learning during this task. For this metric, focusing neurosurgical trainee self-learning on increasing the mean forces to that of the neurosurgical group values appears appropriate. A similar pattern was demonstrated

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

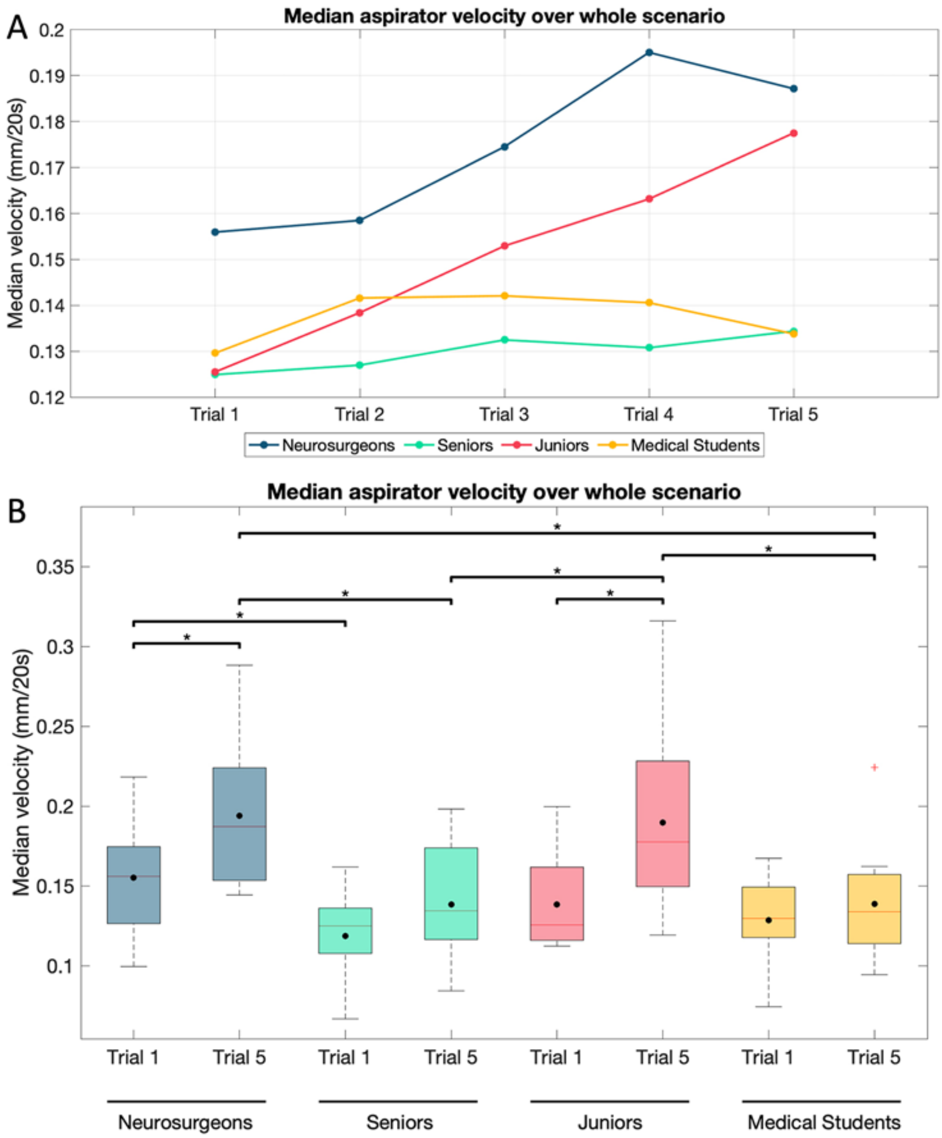

FIG. 4. A: Line graph showing the learning curves for each expertise level over 5 trials for the median aspirator velocity over the whole scenario. **B:** Boxplot showing between-trial analysis and between-group analysis for the median aspirator velocity over the whole scenario for each group: neurosurgeons (n = 13), seniors (n = 14), juniors (n = 10), and medical students (n = 12). The *plus sign* (+) represents an outlier; *black dots* indicate the mean and *red lines* indicate the median. *p < 0.05. Figure is available in color online only.

for the median change in aspirator velocity while resecting tumor. This metric increased between trial 1 and trial 5 in the neurosurgeon, senior, and junior groups, suggesting that this metric was also important for expertise in this task and that neurosurgical trainees should strive to achieve neurosurgeon group values. For the medical student group, there were no significant changes for these 2 metrics, suggesting that for this group's level of ability, self-directed learning for these metrics was not prominent. Medical student self-directed learning should also focus on increasing the values in these 2 metrics. The median aspirator velocities over the whole scenario and during tumor resection were found to significantly increase for both the neurosurgeon and junior groups, but not for the senior group. The reason for this divergence between the senior group and neurosurgeon and junior (resident) groups in these 2 metrics is unclear. This discontinuous learning pattern has been reported previously by our group and needs further investigation.[11] However, focusing all neurosurgical learners on achieving neurosurgical group values seems reasonable.

Although most of the metrics demonstrated a significant difference between trials, the mean aspirator deceleration while resecting tumor did not. The results of this metric have demonstrated numerous outliers in each group, and this large variation in the data may have contributed to the lack of group differences. Furthermore, the K-nearest neighbors algorithm used in the previous study did not investigate the change between trials or the interaction of groups and time.[10] Rather, the algorithm as-

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

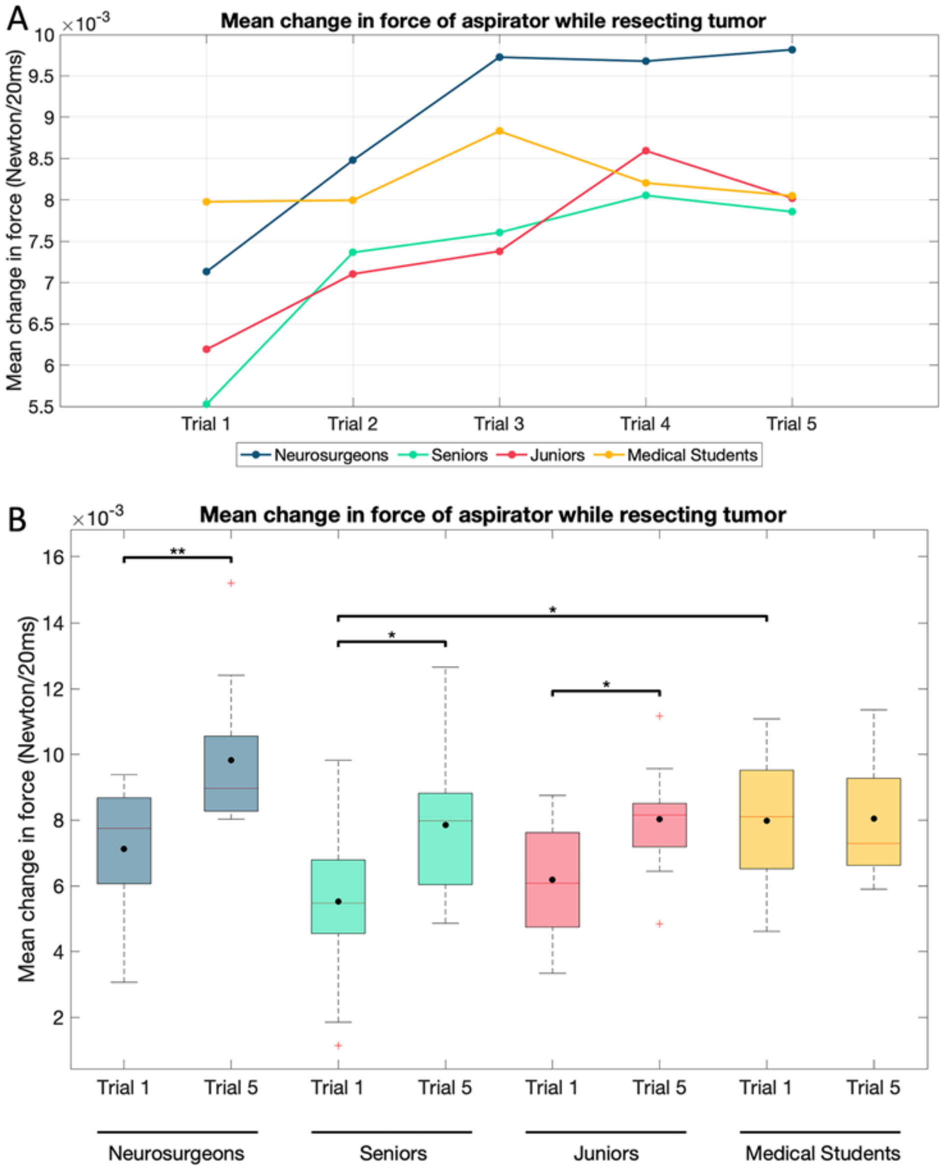

FIG. 5. A: Line graph showing the learning curves for each expertise level over the 5 trials for the mean aspirator change in force while resecting tumor. **B:** Boxplot showing between-trial analysis and between-group analysis for the mean aspirator change in force while resecting tumor for each group: neurosurgeons (n = 13), seniors (n = 14), juniors (n = 10), and medical students (n = 12). The *plus sign* (+) represents an outlier; *black dots* indicate the mean and *red lines* indicate the median. *p < 0.05; **p < 0.01. Figure is available in color online only.

sessed the interactions among multiple specific differentiating factors between groups to help accurately classify individuals.

The assessment of learning curves using VR simulation presents a unique opportunity to understand the acquisition of surgical psychomotor skills outside the operative environment. The COVID-19 pandemic has resulted in a reduction of in-person learning opportunities and the cancellation of elective cases.[20,21] Using innovative methodologies that combine machine learning and VR to assess learning can help mitigate some of these consequences by allowing technical skills surgical training to continue, even in the midst of a pandemic.

Implications for AI-Based Teaching

Our results have implications for the development of AI-powered tutoring systems for VR surgical simulation. Previous studies have outlined the ability of machine learning algorithms to automate surgical performance classification.[5,10,11] However, these results have failed to demonstrate the ability of these derived metrics to improve surgical performance. Our results highlight that neurosurgeons, senior and junior neurosurgical trainees, and medical students can improve their learning curve performance on some metrics that were used for expertise classification by the K-nearest neighbors machine learning algorithm. More generally, these results indicate that metrics derived

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

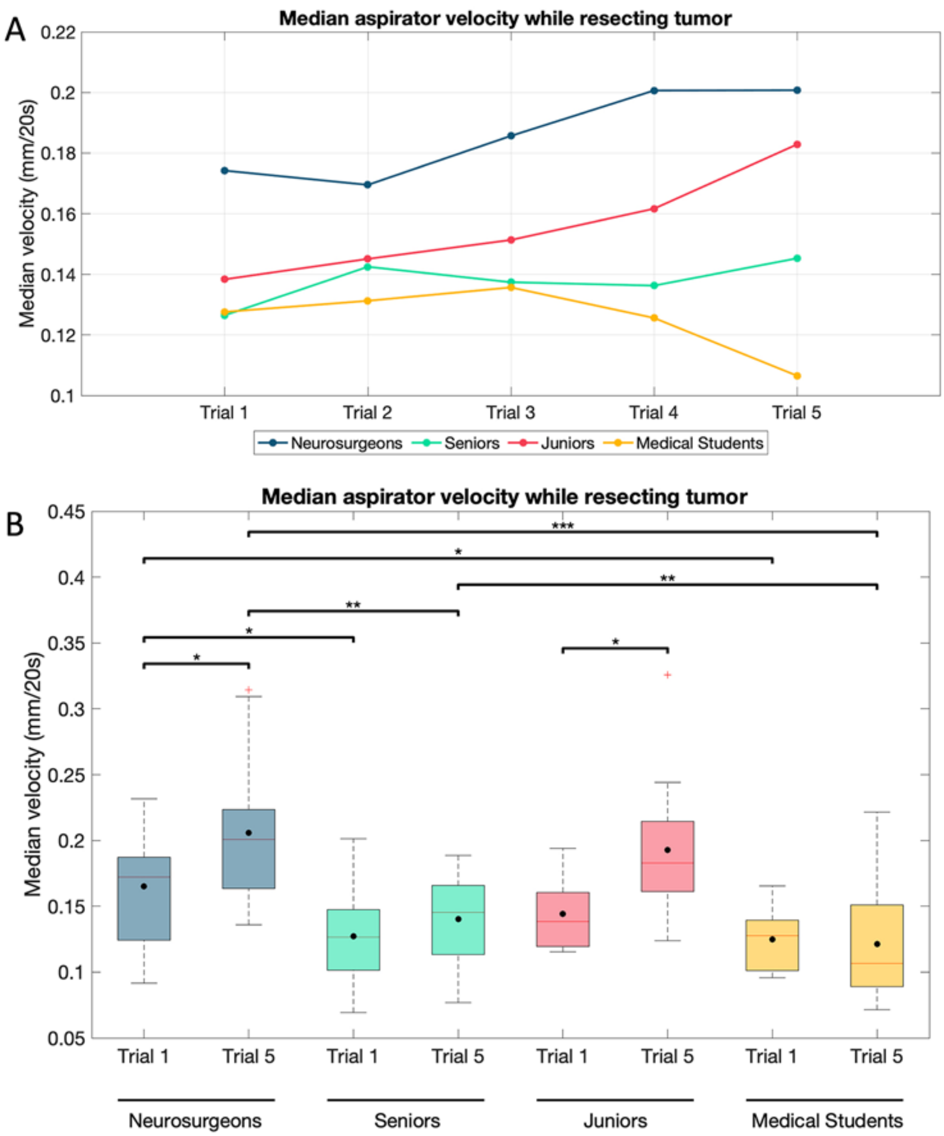

FIG. 6. A: Line graph showing the learning curves for each expertise level over 5 trials for the median aspirator velocity while resecting tumor. **B:** Boxplot showing between-trial analysis and between-group analysis for the median aspirator velocity while resecting tumor for each group: neurosurgeons (n = 13), seniors (n = 14), juniors (n = 10), and medical students (n = 12). The *plus sign* (+) represents an outlier; *black dots* indicate the mean and *red lines* indicate the median. *p < 0.05; **p < 0.01; ***p < 0.001. Figure is available in color online only.

from AI for classification purposes can also be used to improve surgical performance. Although this has only been investigated in the context of a neurosurgical task, it is reasonable to suggest that these results could be generalized to other surgical fields that have used AI metrics for classification purposes. Additionally, learning curve performance improved for these metrics even without specific coaching or knowledge of which metrics were being assessed. This suggests that it is reasonable to develop randomized controlled trials comparing participant learning curves on VR simulators using metrics selected by AI algorithms with instructor-based systems. The results of this study further underline the importance of selecting metrics that are appropriate for different learners because the type and quantity of metrics, as well as the type of instructional support, may vary by expertise level. Using defined metrics, as outlined in this study, surgical educators can assess trainee learning curves and provide personalized feedback to improve performance. Learners can compare their individual learning curves with those of expert groups and focus on specific metrics that need improvement.

The Virtual Operative Assistant (VOA) is an AI-powered tutoring platform that relies on repetitive practice on a VR surgical simulator to achieve performance mastery on specific instrument motion and safety metrics.[22] The combination of personalized AI-powered feedback and learning curve assessment could play an important role in helping surgical educators and trainees better understand

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

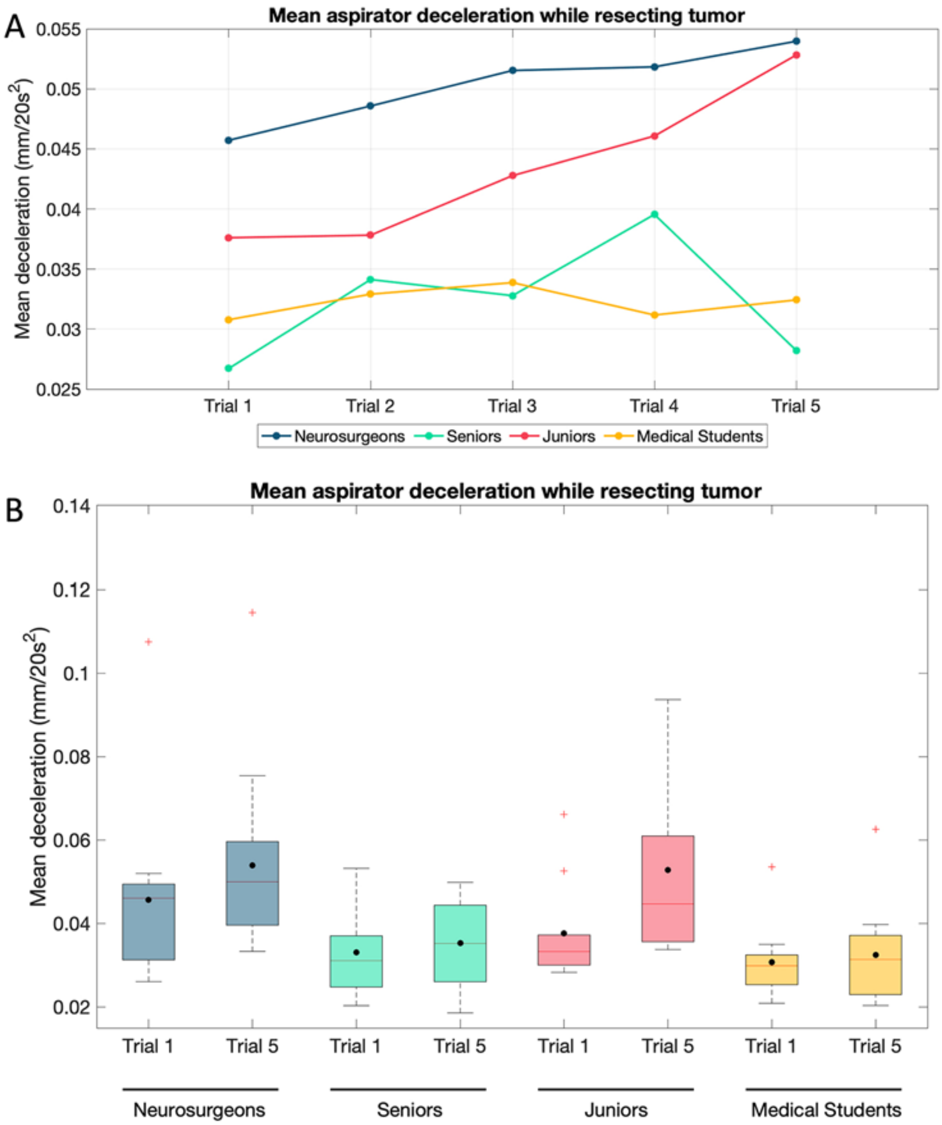

FIG. 7. A: Line graph showing the learning curves for each expertise level over 5 trials for the mean aspirator deceleration while resecting tumor. **B:** Boxplot showing between-trial analysis and between-group analysis for the mean aspirator deceleration while resecting tumor for each group: neurosurgeons (n = 13), seniors (n = 14), juniors (n = 10), and medical students (n = 12). The *plus sign* (+) represents an outlier; *black dots* indicate the mean and *red lines* indicate the median. Figure is available in color online only.

how to improve surgical skills. To address this issue, our group is conducting a randomized controlled trial to assess if VOA intelligent tutoring systems improve participant performance compared with expert educator instruction (ClinicalTrials.gov registration no. NCT04700384).

Limitations

There are limitations associated with this study. First, the decision to use the 6 metrics selected by the K-nearest neighbors algorithm in the present investigation was based on their ability to classify expertise with greater precision in the same simulated VR study.[5] Two of these metrics were also identified by 2 of the 3 different machine learning algorithms used in our previous research as important in expertise classification.[5] However, some metrics that were not selected by the K-nearest neighbors algorithm were identified as useful by the other algorithms used.[5] Further research is required to explore the learning curves associated with different machine learning models. Second, factors such as age and clinical experience were not controlled for in this study. Results indicated that the neurosurgeon group was significantly older and had more self-reported complete subpial resections compared with juniors and seniors. There was no association between groups and prior use of a simulator in this trial. However, some participants had more previous experience using VR simulators and this may have influenced their baseline performance. Studies with larger sample sizes are necessary to investigate how factors such as age, clinical experience, and previous exposure to surgical simulation may

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

influence initial performance and learning curves. Third, participants were not provided feedback during the trials, and feedback may be an important factor in reducing the learning curve for performance improvement, especially in novices.[9] A randomized controlled trial using a similar VR model is presently being conducted to obtain more clarity on the role personalized feedback plays in learning. Fourth, all individuals who participated in this study were from a single institution, therefore limiting the generalization of results. Fifth, this study demonstrated learning curve development on a VR simulator; however, further studies are needed to determine how these results translate to learning surgical skills in a clinical setting and in other surgical fields. Finally, participants were assigned to a group based on their level of training. A more comprehensive method to classify the level of individual expertise using quantitative assessment across a defined series of operative skills may improve the accuracy of learning curve assessment.

Conclusions

Learning curves created using AI-derived metrics, which enhance the precision of expertise classification during repeated VR-simulated subpial tumor resections, provide novel insights into technical skill acquisition. This insight will allow surgical educators to develop more focused formative educational paradigms to expedite neurosurgical trainee progress.

Acknowledgments

We thank all the neurosurgeons, neurosurgical fellows, senior and junior residents, and medical students who participated in this study. We also thank Dr. Robert DiRaddo, group leader of the Simulation section, Life Sciences Division, National Research Council of Canada at Boucherville, Quebec, Canada, and his team, for their assistance in developing the scenario used in this study, and the National Research Council of Canada for providing a prototype of the NeuroVR. We acknowledge the National Neurosciences Institute, King Fahad Medical City, Riyadh, Saudi Arabia, for helping to fund the creation of the scenario used. Finally, we thank Dr. Jason Harley from McGill University for their contributions to this publication.

This work was supported by the Franco Di Giovanni Foundation, the Montreal English School Board, the Montreal Neurological Institute and Hospital, and the Department of Orthopaedic Surgery at McGill University. Nicole Ledwos is a recipient of the Christian Gaeda Brain Tumour Research Studentship from the Montreal Neurological Institute at McGill University, and Alexander Winkler-Schwartz was supported by the Fonds de recherche du Québec–Santé and a Robert Maudsley Fellowship for Studies in Medical Education from the Royal College of Physicians and Surgeons of Canada.

References

1. Hebb AO, Yang T, Silbergeld DL. The sub-pial resection technique for intrinsic tumor surgery. *Surg Neurol Int*. 2011; 2:180.
2. Sabbagh AJ, Bajunaid KM, Alarifi N, et al. Roadmap for developing complex virtual reality simulation scenarios: subpial neurosurgical tumor resection model. *World Neurosurg*. 2020;139:e220-e229.
3. Delorme S, Laroche D, DiRaddo R, Del Maestro RF. NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training. *Neurosurgery*. Sep 2012;71(1 Suppl Operative):32-42.
4. Siyar S, Azarnoush H, Rashidi S, Del Maestro RF. Tremor assessment during virtual reality brain tumor resection. *J Surg Educ*. 2020;77(3):643-651.
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 Netw Open*. 2019;2(8):e198363.
6. Chan J, Pangal DJ, Cardinal T, et al. A systematic review of virtual reality for the assessment of technical skills in neurosurgery. *Neurosurg Focus*. 2021;51(2):E15.
7. Bartlett JD, Lawrence JE, Yan M, et al. The learning curves of a validated virtual reality hip arthroscopy simulator. *Arch Orthop Trauma Surg*. 2020;140(6):761-767.
8. Aggarwal R, Ward J, Balasundaram I, Sains P, Athanasiou T, Darzi A. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Ann Surg*. 2007; 246(5):771-779.
9. Janse JA, Goedegebuure RS, Veersema S, Broekmans FJ, Schreuder HW. Hysteroscopic sterilization using a virtual reality simulator: assessment of learning curve. *J Minim Invasive Gynecol*. 2013;20(6):775-782.
10. 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. *J Bone Joint Surg Am*. 2019;101(23):e127.
11. Mirchi N, Bissonnette V, Ledwos N, et al. Artificial neural networks to assess virtual reality anterior cervical discectomy performance. *Oper Neurosurg (Hagerstown)*. 2020;19(1):65-75.
12. Sewell C, Morris D, Blevins NH, et al. Providing metrics and performance feedback in a surgical simulator. *Comput Aided Surg*. 2008;13(2):63-81.
13. 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.
14. Sawaya R, Bugdadi A, Azarnoush H, et al. Virtual reality tumor resection: the force pyramid approach. *Oper Neurosurg (Hagerstown)*. 2018;14(6):686-696.
15. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194.
16. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health*. 2020;2(10):e537-e548.
17. 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*. 2019;76(6):1681-1690.
18. Brunozzi D, McGuire LS, Alaraj A. NeuroVR Simulator in Neurosurgical Training. In: Alaraj A, ed. *Comprehensive Healthcare Simulation: Neurosurgery*. Springer International Publishing; 2018:211-218.
19. Alaraj A, Tobin MK, Birk DM, Charbel FT. Simulation in neurosurgery and neurosurgical procedures. In: Levine AI, DeMaria S, Schwartz AD, Sim AJ, eds. *The Comprehensive Textbook of Healthcare Simulation*. Springer New York; 2013:415-423.
20. Mirchi N, Ledwos N, Del Maestro RF. Intelligent tutoring systems: re-envisioning surgical education in response to COVID-19. *Can J Neurol Sci*. 2021;48(2):198-200.
21. Tomlinson SB, Hendricks BK, Cohen-Gadol AA. Editorial. Innovations in neurosurgical education during the COVID-19 pandemic: is it time to reexamine our neurosurgical training models? *J Neurosurg*. 2020;133(1):14-15.
22. 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-

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC

based training in surgery and medicine. *PLoS One*. 2020; 15(2):e0229596.

Disclosures

Ms. Ledwos, Mr. Mirchi, Dr. Yilmaz, Dr. Winkler-Schwartz, Dr. Bissonnette, and Dr. Del Maestro are patent holders for a system for generating a training platform.

Author Contributions

Conception and design: Ledwos, Mirchi, Yilmaz, Winkler-Schwartz, Sawni, Bissonnette, Del Maestro. Acquisition of data: Winkler-Schwartz, Bajunaid, Del Maestro. Analysis and interpretation of data: Ledwos, Mirchi, Sawni, Del Maestro. Drafting the article: Ledwos, Mirchi, Fazlollahi, Del Maestro. Critically revising the article: Ledwos, Mirchi, Yilmaz, Fazlollahi, Bissonnette, Del Maestro. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Ledwos. Statistical analysis: Ledwos, Mirchi, Winkler-Schwartz. Administrative/technical/material support: Sabbagh, Del Maestro. Study supervision: Del Maestro.

Supplemental Information

Videos

Video 1. https://vimeo.com/656963987.

Correspondence

Nicole Ledwos: Neurosurgical Simulation and Artificial Intelligence Learning Centre, McGill University, Montreal, QC, Canada. nicole.ledwos@mail.mcgill.ca.

Unauthenticated | Downloaded 07/06/23 06:57 PM UTC