Commentary: Artificial Neural Networks to Assess Virtual Reality Anterior Cervical Discectomy Performance

Ifije E. Ohiorhenuan, MD, PhD
Juan S. Uribe, MD

Department of Neurosurgery, Barrow-Neurological Institute, St. Joseph's Hospital and Medical Center, Phoenix, Arizona

Correspondence:
Juan S. Uribe, MD,
c/o Neuroscience Publications,
Barrow Neurological Institute,
St. Joseph's Hospital and Medical Center,
350 W. Thomas Rd.,
Phoenix, AZ 85013, USA.
Email: Neuropub@barrowneuro.org

Received, November 18, 2019.
Accepted, December 1, 2019.

We commend the authors[1] for their provocative study using artificial neural networks to assess the performance of anterior cervical discectomy (ACDF) using a virtual reality simulation. In their study, the authors[1] found that the number of times instruments contact the dura and the force applied to the posterior longitudinal ligament (PLL) are significant differentiators between levels of expertise. Surprisingly, the authors found a nonlinear relationship between these surgical skill metrics and expertise level. Specifically, senior residents contacted the dura more often and applied more force to the PLL than either postresidents or junior residents. This finding underscores the point that the training on surgical procedures involves a nuanced learning curve.

The authors[1] measured the instrument position using a virtual reality simulator, and they correlated these measurements with the expertise level using an artificial neural network (ANN). An understanding of each of these elements is helpful for appreciating the conclusions and limitations of the study. The virtual reality simulator consisted of 3D glasses and instrument handles capable of providing haptic feedback to the user. Throughout the ACDF procedure, a number of variables were recorded, including instrument choice, position, force, and volume of tissue removed. The study group consisted of 21 participants in 3 groups: junior residents, senior residents, and postresidents. Once recorded, the raw data were combined into a set of performance metrics. These data could be roughly categorized into 4 types: safety (eg, maximum force applied on an anatomic structure), motion (eg, velocity of instrument while in contact with the disc), efficiency (eg, number of instrument contacts on an anatomic structure), and cognitive (ie, instrument choice). For each participant, a total of 333 metrics were obtained. Stepwise linear regression was used to narrow this dataset down to 16 metrics.

Next, the authors[1] used an ANN to identify which performance metrics were most strongly associated with which expertise level. ANNs are a form of classification model, which have been particularly successful in the classification of high-dimensional data or data with nonlinear interactions. Typically, ANNs are trained on a subset of data and tested on a hold-out dataset. In this case, the authors trained the ANN on ~70% of the participants (15) and tested its accuracy on the remaining 6.

They found a training accuracy of 100% and a testing accuracy of 83%, meaning that 5 of the 6 users in the hold-out dataset were correctly identified as either junior, senior, or postresident. This level of accuracy is impressive, although the sample size is small. In addition, the authors found that as the number of contacts with the dura increased, the odds of being classified as a senior resident increased, while as the number of contacts with the dura decreased, the odds of being classified as a postresident increased. Also, they found that senior residents applied higher forces to the PLL than either postresidents or junior residents.

The primary limitation of this study is the small sample size. A larger sample size may have uncovered more performance metrics that were associated with the level of expertise. In addition, it is unclear how similar the virtual reality ACDF simulator is to an actual ACDF, so adeptness in the simulator may not reflect performance in an actual ACDF case (and vice versa). Finally, experience level is taken as a de facto measure of expertise, which may or may not be valid.

Despite these limitations, the study provides a lot of food for thought. In an era of resident work hour restrictions and increased supervision, surgical simulators may be useful for training adjuncts, allowing trainees to accumulate training repetitions "offline." Similar training simulators have been useful for the management of complications, such as carotid artery injury during endoscopic endonasal surgery[2] and dural repair during minimally invasive spine surgery.[3] More broadly, this study demonstrates that a small subset of key maneuvers can be identified that separate surgeons with proficiency in a common neurosurgical task from those with expertise in that task. By focusing on these key maneuvers, the efficiency of training could be improved. Similarly to flight simulators for pilots, this method has the potential to raise the bar of proficiency and reduce variability across surgical trainees.

Disclosures

The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Mirchi N, Bissonnette V, Ledwos N, et al. Artificial neural networks to assess virtual reality anterior cervical discectomy performance. *Oper Neurosurg*. published online: 2019 (doi:10.1093/ons/opz359).
2. Pham M, Kale A, Marquez Y, et al. A perfusion-based human cadaveric model for management of carotid artery injury during endoscopic endonasal skull base surgery. *J Neurol Surg B Skull Base*. 2014;75(5):309-313.
3. Bakhsheshian J, Strickland BA, Patel NN, et al. The use of a novel perfusion-based cadaveric simulation model with cerebrospinal fluid reconstitution comparing dural repair techniques: a pilot study. *Spine J*. 2017;17(9):1335-1341.

Acknowledgments

The authors thank the staff of Neuroscience Publications at Barrow Neurological Institute for assistance with manuscript preparation.