

**UCC Library and UCC researchers have made this item openly available.  
 Please [let us know](#) how this has helped you. Thanks!**

<b>Title</b>	Artificial intelligence and training physicians to perform technical procedures
<b>Author(s)</b>	Shorten, George
<b>Publication date</b>	2019-08-02
<b>Original citation</b>	Shorten, G. (2019) 'Artificial Intelligence and Training Physicians to Perform Technical Procedures', JAMA Network Open, 2(8), e198375. (2pp.) DOI:10.1001/jamanetworkopen.2019.8375
<b>Type of publication</b>	Article (peer-reviewed)
<b>Link to publisher's version</b>	<a href="https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2740774">https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2740774</a> <a href="http://dx.doi.org/10.1001/jamanetworkopen.2019.8375">http://dx.doi.org/10.1001/jamanetworkopen.2019.8375</a> Access to the full text of the published version may require a subscription.
<b>Rights</b>	<b>This is an open access article distributed under the terms of the CC-BY License. © 2019 Shorten G. JAMA Network Open.</b> <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
<b>Item downloaded from</b>	<a href="http://hdl.handle.net/10468/8597">http://hdl.handle.net/10468/8597</a>

Downloaded on 2022-08-19T22:59:11Z



Invited Commentary | Medical Education

# Artificial Intelligence and Training Physicians to Perform Technical Procedures

George Shorten, MD, PhD

Winkler-Schwartz et al<sup>1</sup> have set out to determine if some combination of machine learning algorithms can differentiate participants according to their stage of practice (ie, neurosurgeon, fellow, senior or junior resident, or medical student) based on their performance of a complex simulated neurosurgical task. A total of 250 simulated surgical resections performed by 50 participants were studied using a prospective, observational case series design. The best-performing algorithm (K-nearest neighbor) had 90% accuracy for prediction and used 6 machine-selected metrics. Three of the 4 algorithms used in the study misclassified a medical student as a neurosurgeon.

The article addresses a very important question, using a valid approach, and presents credible and promising results. The authors' work prompts wider consideration of how to apply artificial intelligence to human behavior in medicine, particularly to the performance of technical tasks.

The most fundamental of these applications is the question of meaning. Artificial intelligence, of which machine learning is one advanced application, refers to the capacity of a computer to perform operations analogous to learning and decision-making in humans. The objective of this study was "to identify surgical and operative factors selected by a machine learning algorithm to accurately classify participants by level of expertise in a virtual reality surgical procedure."<sup>1</sup> In the absence of a standard criterion or objective measure (such as time taken to complete a 100-m race), machine learning offers unprecedented capacity to identify associations between different variables (in many combinations or forms) in a particular system. To put these discoveries to use, it is necessary to understand the significance of key variables. In this case, does participant role or title equate to level of expertise? Is a neurosurgeon's performance invariably more "expert" than that of a fellow or resident? If it is not, then perhaps some of the prediction "errors" were not erroneous. Metric-based assessment of consultant surgical performance consistently identifies a significant minority of inferior performing outliers (>2 SD from the mean).<sup>2</sup> In this study, concurrent application of expert-derived performance metrics<sup>3,4</sup> could have enabled discrimination between career stage and level of performance.

The use of machine learning to discriminate or predict is based on estimates of probability, with the estimates usually improving as the amount of data from which they are calculated increases. At first glance, this probabilistic view of the world appears to differ from that generally associated with the scientific method, in which a hypothesis is generated and then tested for a binary outcome, as if true or false. The difference between the 2 approaches is actually not large: the use of machine learning simply causes us to consider not just the selected outcome ("the participant is likely to be a neurosurgeon") but also the magnitude of that likelihood. There will be advantages to that degree of scrutiny of a result. When high-stakes decisions are to be made on the basis of a machine learning prediction, we may choose to turn up the "gain switch" (ie, the probability that is required to select or predict a particular outcome).

The authors insightfully point out the potential value of explainable artificial intelligence in the setting of training humans on technical skills. Intuitively, one can appreciate the value a trainee would attach to a movement or step that fits with his or her mental model of what the procedure requires. Although not part of the research question addressed in this study, one is led to consider how artificial intelligence-derived performance metrics might be applied to enhance a training program

## + Related article

Author affiliations and article information are listed at the end of this article.

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

or an individual's deliberate practice. How do these performance metrics (derived from raw data) compare, in terms of content and training value, with those elicited from expert practitioners (as per Angelo et al<sup>4</sup>)? The generation of performance metrics from raw data described in this study appears limited in scope. The selection of (derived) instrument velocity, acceleration, jerk, and tip separation are all reasonable, as are application speeding up, tips converging, and other metrics. But together the 270 metrics selected still represent a small subset of all possible motion or position metrics that might represent expert or novice performance.

The risks associated with misinterpreting or overinterpreting correlation values (or other forms of association) are not limited to artificial intelligence or machine learning. If specific surgical or operative factors can be used to consistently discriminate expert from nonexpert performance, it does not follow that these factors constitute or even contribute to expertise. Developing training interventions that cause trainees to exhibit these factors will not necessarily result in expert performance. Experts may move more rapidly at certain points of the procedures or take shortcuts that contribute nothing to the safety or effectiveness of a procedure. It is tempting to assume that expert performance of a particular procedure is uniform; experts may complement or compensate for their "native" characteristics and thereby achieve similar technical outcomes by quite different routes. Future studies of the type performed by Winkler-Schwartz et al<sup>1</sup> may take account of this possibility by measuring participants' psychomotor and visuospatial abilities and handedness using standard tests.

---

#### ARTICLE INFORMATION

**Published:** August 2, 2019. doi:[10.1001/jamanetworkopen.2019.8375](https://doi.org/10.1001/jamanetworkopen.2019.8375)

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2019 Shorten G. *JAMA Network Open*.

**Corresponding Author:** George Shorten, MD, PhD, Department of Anaesthesia and Intensive Care Medicine, Cork University Hospital, University College Cork, Cork T12DC4A, Ireland ([g.shorten@ucc.ie](mailto:g.shorten@ucc.ie)).

**Author Affiliation:** Department of Anaesthesia and Intensive Care Medicine, Cork University Hospital, University College Cork, Cork, Ireland.

**Conflict of Interest Disclosures:** None reported.

#### REFERENCES

1. 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. doi:[10.1001/jamanetworkopen.2019.8363](https://doi.org/10.1001/jamanetworkopen.2019.8363)
2. Gallagher AG, Henn PJ, Neary PC, et al. Outlier experienced surgeon's performances impact on benchmark for technical surgical skills training. *ANZ J Surg*. 2018;88(5):E412-E417. doi:[10.1111/ans.14474](https://doi.org/10.1111/ans.14474)
3. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. 2002;236(4):458-463. doi:[10.1097/0000658-200210000-00008](https://doi.org/10.1097/0000658-200210000-00008)
4. Angelo RL, Ryu RK, Pedowitz RA, et al. A proficiency-based progression training curriculum coupled with a model simulator results in the acquisition of a superior arthroscopic Bankart skill set. *Arthroscopy*. 2015;31(10):1854-1871. doi:[10.1016/j.arthro.2015.07.001](https://doi.org/10.1016/j.arthro.2015.07.001)