JAMA Network Open

Original Investigation | Medical Education

Machine Learning Identification of Surgical and Operative Factors Associated With Surgical Expertise in Virtual Reality Simulation

Alexander Winkler-Schwartz, MD; Recai Yilmaz, MD; Nykan Mirchi, BSc; Vincent Bissonnette, MD; Nicole Ledwos, BA; Samaneh Siyar, MSc; Hamed Azarnoush, PhD; Bekir Karlik, PhD; Rolando Del Maestro, PhD

Abstract

IMPORTANCE Despite advances in the assessment of technical skills in surgery, a clear understanding of the composites of technical expertise is lacking. Surgical simulation allows for the quantitation of psychomotor skills, generating data sets that can be analyzed using machine learning algorithms.

OBJECTIVE 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.

DESIGN, SETTING, AND PARTICIPANTS Fifty participants from a single university were recruited between March 1, 2015, and May 31, 2016, to participate in a case series study at McGill University Neurosurgical Simulation and Artificial Intelligence Learning Centre. Data were collected at a single time point and no follow-up data were collected. Individuals were classified a priori as expert (neurosurgery staff), seniors (neurosurgical fellows and senior residents), juniors (neurosurgical junior residents), and medical students, all of whom participated in 250 simulated tumor resections.

EXPOSURES All individuals participated in a virtual reality neurosurgical tumor resection scenario. Each scenario was repeated 5 times.

MAIN OUTCOMES AND MEASURES Through an iterative process, performance metrics associated with instrument movement and force, resection of tissues, and bleeding generated from the raw simulator data output were selected by K-nearest neighbor, naive Bayes, discriminant analysis, and support vector machine algorithms to most accurately determine group membership.

RESULTS A total of 50 individuals (9 women and 41 men; mean [SD] age, 33.6 [9.5] years; 14 neurosurgeons, 4 fellows, 10 senior residents, 10 junior residents, and 12 medical students) participated. Neurosurgeons were in practice between 1 and 25 years, with 9 (64%) involving a predominantly cranial practice. The K-nearest neighbor algorithm had an accuracy of 90% (45 of 50), the naive Bayes algorithm had an accuracy of 84% (42 of 50), the discriminant analysis algorithm had an accuracy of 78% (39 of 50), and the support vector machine algorithm had an accuracy of 76% (38 of 50). The K-nearest neighbor algorithm used 6 performance metrics to classify participants, the naive Bayes algorithm used 9 performance metrics, the discriminant analysis algorithm used 8 performance metrics, and the support vector machine algorithm used 8 performance metrics. Two neurosurgeons, 1 fellow or senior resident, 1 junior resident, and 1 medical student were misclassified.

CONCLUSIONS AND RELEVANCE In a virtual reality neurosurgical tumor resection study, a machine learning algorithm successfully classified participants into 4 levels of expertise with 90% accuracy.

(continued)

Key Points

Question Can a machine learning algorithm differentiate participants according to their stage of practice in a complex simulated neurosurgical task?

Findings In this case series study, 50 individuals (14 neurosurgeons, 4 fellows, 10 senior residents, 10 junior residents, and 12 medical students) participated in 250 simulated tumor resections. An accuracy of 90% was achieved using 6 performance features by a K-nearest neighbor algorithm and 2 neurosurgeons, 1 fellow or senior resident, 1 junior resident, and 1 medical student were misclassified.

Meaning The findings suggest that machine learning algorithms may be capable of classifying surgical expertise with greater granularity and precision than has been previously demonstrated in surgery.

+ Invited Commentary

+ Video

+ Supplemental content

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

Abstract (continued)

These findings suggest that algorithms may be capable of classifying surgical expertise with greater granularity and precision than has been previously demonstrated in surgery.

JAMA Network Open. 2019;2(8):e198363. doi:10.1001/jamanetworkopen.2019.8363

Introduction

Despite technological advances in artificial intelligence and machine learning, delivery of health care is mediated largely by direct interaction between physician and patient. This scenario is particularly true for surgical interventions, which carry substantive patient risks and increased costs to health care systems.[1] As a consequence, the burgeoning field of surgical data science represents efforts to improve interventional health care through increased data collection, quantification, and analysis.[2] Similarly, the use of virtual reality simulators has been explored as a means of providing objective assessment of technical ability in medicine, with the added benefit of retaining realism, pathology, and active bleeding states in a controlled laboratory setting. These systems generate vast data sets that quickly challenge traditional statistical methods. Artificial intelligence and machine learning systems lend themselves well to the analysis of large data sets generated in surgical procedures in 2 important ways: first, by uncovering previously unrecognized patterns, they can expand the understanding of the composites of technical expertise and surgical error, and second, by grouping participants according to technical ability, they offer novel avenues for training and feedback in health care.

We sought to study the operative factors selected by a series of machine learning algorithms to most accurately classify participants by level of expertise in a virtual reality surgery. Using an advanced high-fidelity neurosurgical simulator allows participants to conduct a complex open neurosurgical brain tumor resection task in a risk-free environment.[3,4] Our group has extensive experience in virtual reality surgical simulation; several studies have demonstrated that performance measures obtained from simulation can differentiate technical skills both between and within groups of expertise.[5-9] Given the task complexity and the abundance of data generated during the simulated operation, we hypothesized that machine learning algorithms could identify previously unrecognized performance measures, as well as differentiate participants according to their stage of medical practice.

Methods

Study Participants

All neurosurgeons, neurosurgical fellows, and neurosurgical residents from a single Canadian university were invited between March 1, 2015, and May 31, 2016, to participate in the trial. Medical students rotating on a neurosurgical service or having expressed interest in being contacted for trials were invited. Data were collected at a single time point and no follow-up data were collected. Participants were classified a priori as expert (neurosurgery staff), seniors (neurosurgery fellows and residents in years 4-6), juniors (neurosurgery residents in years 1-3), and medical students. All participants signed an approved Montreal Neurological Institute and Hospital Research Ethics Board consent form before trial participation. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Declaration of Helsinki.[10] The study received local ethics board approval at the Montreal Neurological Institute and Hospital. This report is structured according to guidelines for best practices in reporting studies on machine learning to assess surgical expertise in virtual reality simulation.[11,12]

Study Design

The Simulator

The NeuroVR (CAE Healthcare) is a high-fidelity neurosurgical simulator designed to recreate the visual and haptic experience of resecting a human brain tumor through an operative microscope. The platform was developed in 2012 by a team from the National Research Council of Canada in collaboration with an advisory network of surgeons from 23 Canadian and international teaching hospitals.[4] Care was taken to provide the most realistic sensory feedback for the user by incorporating physical properties of human primary brain tumors.[4] As such, the attention to detail and resources used in the creation of the NeuroVR make it one of the most advanced high-fidelity simulators available for neurosurgery.[3]

The Virtual Reality Tumor Resection Task

The trial was carried out at the McGill Neurosurgical Simulation and Artificial Intelligence Learning Centre in a controlled laboratory environment void of distractions. A human intrinsic subpial brain tumor resection task was designed by neurosurgeons with extensive experience in neuro-oncologic and epilepsy neurosurgery. The subpial technique is a challenging bimanual psychomotor skill acquired in neurosurgery and is primarily used in epilepsy and oncologic surgery, where preservation of adjacent eloquent structures is of paramount importance.[13] Participants were given written and verbal instructions that the goal of the scenario was removal of the cortical tumor using the ultrasonic aspirator without damaging adjacent normal brain tissue and vessels. A bipolar instrument could be used to lift and retract the pial membrane to gain access to the tumor and cauterize possible bleeding points. Participants performed the scenario 5 times; however, for the analysis these tasks were averaged and not treated separately. The duration of the resection procedure was limited to 3 minutes.[14] **Video 1** is a sample video of the task and **Video 2** is a 3-dimensional tumoral reconstruction.

Statistical Analysis

Raw Data Obtained From the Simulator

After each trial, the NeuroVR provides a comma separated value (CSV) file containing, in 20-millisecond increments, the activation, force applied, tip position, and angle of each instrument; the volume of tumor and surrounding healthy tissues removed; blood loss; and whether a given instrument was in contact with the tumor, a blood vessel, or healthy tissue. MATLAB, release 2018a (The MathWorks Inc) was used to process the data into operative performance metrics that can be used by a machine learning algorithm. Interpolation was used to render the data regular and fill occasional missing data points (due to slight fluctuations in computer processing). eFigure 1 in the Supplement has further examples.

Performance Metric Extraction

To begin, raw data were transformed into performance metrics to be used by the algorithm, with the intention of generating operative measurements that would be easily interpretable by teachers and students of surgery. This process includes transforming instrument movement from the original x, y, and z coordinates into 3-dimensional representations of velocity (first derivative of position), acceleration (first derivative of velocity), and jerk (first derivative of acceleration), as well as the separation between instrument tips. The acceleration and tip distance variables were further refined to reflect the rate of change while the instruments were speeding up and slowing down as well as converging and diverging. The rate of change in volume of tumor and healthy tissue, as well as the rate of change of bleeding, and the number of attempts to stop bleeding were generated. Next, the aforementioned variables were extracted during 3 operative conditions: during the course of the whole scenario, during the tumor resection (ie, only when the ultrasonic aspirator was activated with decreasing tumor volume), and during blood suctioning (ie, when the ultrasonic aspirator was not active and while blood in the operative view was decreasing). Finally, the mean, median, and

maximum values of all metrics in all conditions were obtained. **Table 1** lists all 270 metrics generated. Examples among the total 270 possible metrics generated include mean aspirator force while resecting the tumor, maximum rate of bleeding during the course of the whole scenario, and median tip distance while suctioning blood. Performance measures of the 5 scenarios were averaged together for each participant.

Metric Reduction and Normalization

Metrics failing to demonstrate a significant ($P < .05$) difference on a 2-sided t test between any 2 groups were excluded. No corrections for multiple tests were done, as the t tests were performed for data-reductive purposes. Subsequent inclusion of the metrics in the algorithm corrects for the probability of type I error at this stage. Metrics were normalized via z score transformation to ensure optimal algorithm functioning.

Iterative Loop

The following steps involve a repetitive process whereby algorithm optimization and final performance metric selection occur. The process is outlined in **Figure 1**. Forward (starting with 1 and increasing in number) metric selection was performed by randomly adding metrics and backward (starting with the maximum and decreasing in number) metric selection was performed by randomly removing metrics. Calculation of accuracy was accomplished by leave-1-out cross-validation. Leave-1-out validation involves training the machine learning algorithm on the entire participant data set except for 1 individual, whose group membership is then estimated. The process is repeated with different individuals excluded until all participants have been classified. The total number of correctly classified individuals represents the overall accuracy of a given algorithm. No external data set was used to obtain the algorithm accuracy.

Algorithms Used

Four classifier algorithms were used: K-nearest neighbor, naive Bayes, discriminant analysis, and support vector machine. Parameter optimizations were carried out using functions included in MATLAB, release 2018a, as well as code written by us.[15-19]

Results

Participant Characteristics

A total of 50 individuals (14 neurosurgeons, 4 fellows, 10 senior residents, 10 junior residents, and 12 medical students) participated in 250 simulated tumor resections. Demographic information is presented in **Table 2**. Consultant neurosurgeon subspecialization covered a wide breadth of practice, with most (9 [64%]) primarily involved in cranial surgery. A total of 7 neurosurgeons (50%), 10 senior residents (69%), 6 junior residents (60%), and 3 medical students (25%) indicated that they had used a surgical simulator previously.

Machine Learning Ability to Classify Participants

The K-nearest neighbor algorithm had an accuracy of 90% (45 of 50), the naive Bayes algorithm had an accuracy of 84% (42 of 50), the discriminant analysis algorithm had an accuracy of 78% (39 of 50), and the support vector machine algorithm had an accuracy of 76% (38 of 50). **Figure 2** presents details on individual misclassification. Although beyond the scope of the initial hypothesis, in response to misclassifications between medical students and neurosurgeons, the algorithm optimization process was repeated with an emphasis on preventing misclassification between neurosurgeons and medical students, with resulting accuracies ranging between 88% (44 of 50) and 72% (36 of 50). This was accomplished by allowing the algorithm optimization process to stop if no misclassifications between neurosurgeons and medical students occurred, in addition to attaining a

Table 1. Performance Metrics Generated From Raw Simulator Data

Metric No.	Measurement	Instrument	Performance Measure Associated With Movement, Force, Bleeding, or Tissue	Operative Condition
1	Maximum	Aspirator	Acceleration of instrument	Over whole scenario
2	Maximum	Aspirator	Force of instrument	Over whole scenario
3	Maximum	Aspirator	Change in force of instrument	Over whole scenario
4	Maximum	Aspirator	Jerk of instrument	Over whole scenario
5	Maximum	Aspirator	Rate of slowing down of instrument	Over whole scenario
6	Maximum	Aspirator	Rate of speeding up of instrument	Over whole scenario
7	Maximum	Aspirator	Velocity of instrument	Over whole scenario
8	Maximum	Bipolar	Acceleration of instrument	Over whole scenario
9[a,b]	Maximum	Bipolar	Force of instrument	Over whole scenario
10	Maximum	Bipolar	Change in force of instrument	Over whole scenario
11	Maximum	Bipolar	Jerk of instrument	Over whole scenario
12	Maximum	Bipolar	Rate of slowing down of instrument	Over whole scenario
13	Maximum	Bipolar	Rate of speeding up of instrument	Over whole scenario
14	Maximum	Bipolar	Velocity of instrument	Over whole scenario
15	Maximum	NA	Bleeding speed	Over whole scenario
16	Maximum	NA	Change in bleeding speed	Over whole scenario
17	Maximum	NA	Blood in view	Over whole scenario
18	Maximum	NA	Change in blood in view	Over whole scenario
19	Maximum	NA	Increase in bleeding speed	Over whole scenario
20	Maximum	NA	Converging of instrument tips	Over whole scenario
21	Maximum	NA	Diverging of instrument tips	Over whole scenario
22	Maximum	NA	Increased blood in view	Over whole scenario
23	Maximum	NA	Decreasing bleeding rate	Over whole scenario
24	Maximum	NA	Decrease in blood in view	Over whole scenario
25	Maximum	NA	Tip distance of instruments	Over whole scenario
26	Maximum	NA	Change in tip distance of instruments	Over whole scenario
27	Maximum	NA	Change in volume of brain tissue	Over whole scenario
28	Maximum	NA	Total blood emitted	Over whole scenario
29	Maximum	NA	Change in total blood emitted	Over whole scenario
30	Maximum	NA	Change in volume of tumor	Over whole scenario
31	Mean	Aspirator	Acceleration of instrument	Over whole scenario
32	Mean	Aspirator	Force of instrument	Over whole scenario
33	Mean	Aspirator	Change in force of instrument	Over whole scenario
34[a]	Mean	Aspirator	Jerk of instrument	Over whole scenario
35	Mean	Aspirator	Rate of slowing down of instrument	Over whole scenario
36	Mean	Aspirator	Rate of speeding up of instrument	Over whole scenario
37	Mean	Aspirator	Velocity of instrument	Over whole scenario
38	Mean	Bipolar	Acceleration of instrument	Over whole scenario
39[a]	Mean	Bipolar	Force of instrument	Over whole scenario
40	Mean	Bipolar	Change in force of instrument	Over whole scenario
41	Mean	Bipolar	Jerk of instrument	Over whole scenario
42	Mean	Bipolar	Rate of slowing down of instrument	Over whole scenario
43	Mean	Bipolar	Rate of speeding up of instrument	Over whole scenario
44[a]	Mean	Bipolar	Velocity of instrument	Over whole scenario
45[b]	Mean	NA	Bleeding speed	Over whole scenario
46	Mean	NA	Change in bleeding speed	Over whole scenario
47	Mean	NA	Blood in view	Over whole scenario
48[b]	Mean	NA	Change in blood in view	Over whole scenario
49	Mean	NA	Increase in bleeding speed	Over whole scenario
50	Mean	NA	Converging of instrument tips	Over whole scenario
51[a]	Mean	NA	Diverging of instrument tips	Over whole scenario

(continued)

Table 1. Performance Metrics Generated From Raw Simulator Data (continued)

Metric No.	Measurement	Instrument	Performance Measure Associated With Movement, Force, Bleeding, or Tissue	Operative Condition
52	Mean	NA	Increased blood in view	Over whole scenario
53	Mean	NA	Decreasing bleeding rate	Over whole scenario
54	Mean	NA	Decrease in blood in view	Over whole scenario
55[a,c]	Mean	NA	Tip distance of instruments	Over whole scenario
56	Mean	NA	Change in tip distance of instruments	Over whole scenario
57	Mean	NA	Change in volume of brain tissue	Over whole scenario
58	Mean	NA	Total blood emitted	Over whole scenario
59	Mean	NA	Change in total blood emitted	Over whole scenario
60[a,b,d]	Mean	NA	Change in volume of tumor	Over whole scenario
61	Median	Aspirator	Acceleration of instrument	Over whole scenario
62	Median	Aspirator	Force of instrument	Over whole scenario
63[c,d]	Median	Aspirator	Change in force of instrument	Over whole scenario
64	Median	Aspirator	Jerk of instrument	Over whole scenario
65	Median	Aspirator	Rate of slowing down of instrument	Over whole scenario
66	Median	Aspirator	Rate of speeding up of instrument	Over whole scenario
67[c]	Median	Aspirator	Velocity of instrument	Over whole scenario
68	Median	Bipolar	Acceleration of instrument	Over whole scenario
69[a]	Median	Bipolar	Force of instrument	Over whole scenario
70	Median	Bipolar	Change in force of instrument	Over whole scenario
71	Median	Bipolar	Jerk of instrument	Over whole scenario
72	Median	Bipolar	Rate of slowing down of instrument	Over whole scenario
73	Median	Bipolar	Rate of speeding up of instrument	Over whole scenario
74	Median	Bipolar	Velocity of instrument	Over whole scenario
75	Median	NA	Bleeding speed	Over whole scenario
76	Median	NA	Change in bleeding speed	Over whole scenario
77	Median	NA	Blood in view	Over whole scenario
78	Median	NA	Change in blood in view	Over whole scenario
79	Median	NA	Increase in bleeding speed	Over whole scenario
80	Median	NA	Converging of instrument tips	Over whole scenario
81	Median	NA	Diverging of instrument tips	Over whole scenario
82	Median	NA	Increased blood in view	Over whole scenario
83	Median	NA	Decreasing bleeding rate	Over whole scenario
84	Median	NA	Decrease in blood in view	Over whole scenario
85	Median	NA	Tip distance of instruments	Over whole scenario
86	Median	NA	Change in tip distance of instruments	Over whole scenario
87	Median	NA	Change in volume of brain tissue	Over whole scenario
88	Median	NA	Total blood emitted	Over whole scenario
89	Median	NA	Change in total blood emitted	Over whole scenario
90	Median	NA	Change in volume of tumor	Over whole scenario
91	Maximum	Aspirator	Acceleration of instrument	While removing tumor
92	Maximum	Aspirator	Force of instrument	While removing tumor
93	Maximum	Aspirator	Change in force of instrument	While removing tumor
94	Maximum	Aspirator	Jerk of instrument	While removing tumor
95	Maximum	Aspirator	Rate of slowing down of instrument	While removing tumor
96	Maximum	Aspirator	Rate of speeding up of instrument	While removing tumor
97	Maximum	Aspirator	Velocity of instrument	While removing tumor
98	Maximum	Bipolar	Acceleration of instrument	While removing tumor
99	Maximum	Bipolar	Force of instrument	While removing tumor
100	Maximum	Bipolar	Change in force of instrument	While removing tumor
101	Maximum	Bipolar	Jerk of instrument	While removing tumor
102	Maximum	Bipolar	Rate of slowing down of instrument	While removing tumor

(continued)

Table 1. Performance Metrics Generated From Raw Simulator Data (continued)

Metric No.	Measurement	Instrument	Performance Measure Associated With Movement, Force, Bleeding, or Tissue	Operative Condition
103	Maximum	Bipolar	Rate of speeding up of instrument	While removing tumor
104	Maximum	Bipolar	Velocity of instrument	While removing tumor
105	Maximum	NA	Bleeding speed	While removing tumor
106[b]	Maximum	NA	Change in bleeding speed	While removing tumor
107	Maximum	NA	Blood in view	While removing tumor
108	Maximum	NA	Change in blood in view	While removing tumor
109	Maximum	NA	Increase in bleeding speed	While removing tumor
110	Maximum	NA	Converging of instrument tips	While removing tumor
111	Maximum	NA	Diverging of instrument tips	While removing tumor
112	Maximum	NA	Increased blood in view	While removing tumor
113	Maximum	NA	Decreasing bleeding rate	While removing tumor
114	Maximum	NA	Decrease in blood in view	While removing tumor
115	Maximum	NA	Tip distance of instruments	While removing tumor
116	Maximum	NA	Change in tip distance of instruments	While removing tumor
117	Maximum	NA	Change in volume of brain tissue	While removing tumor
118	Maximum	NA	Total blood emitted	While removing tumor
119	Maximum	NA	Change in total blood emitted	While removing tumor
120	Maximum	NA	Change in volume of tumor	While removing tumor
121	Mean	Aspirator	Acceleration of instrument	While removing tumor
122	Mean	Aspirator	Force of instrument	While removing tumor
123[c]	Mean	Aspirator	Change in force of instrument	While removing tumor
124	Mean	Aspirator	Jerk of instrument	While removing tumor
125[c]	Mean	Aspirator	Rate of slowing down of instrument	While removing tumor
126	Mean	Aspirator	Rate of speeding up of instrument	While removing tumor
127	Mean	Aspirator	Velocity of instrument	While removing tumor
128	Mean	Bipolar	Acceleration of instrument	While removing tumor
129[a]	Mean	Bipolar	Force of instrument	While removing tumor
130	Mean	Bipolar	Change in force of instrument	While removing tumor
131	Mean	Bipolar	Jerk of instrument	While removing tumor
132	Mean	Bipolar	Rate of slowing down of instrument	While removing tumor
133	Mean	Bipolar	Rate of speeding up of instrument	While removing tumor
134	Mean	Bipolar	Velocity of instrument	While removing tumor
135	Mean	NA	Bleeding speed	While removing tumor
136	Mean	NA	Change in bleeding speed	While removing tumor
137	Mean	NA	Blood in view	While removing tumor
138	Mean	NA	Change in blood in view	While removing tumor
139	Mean	NA	Increase in bleeding speed	While removing tumor
140	Mean	NA	Converging of instrument tips	While removing tumor
141	Mean	NA	Diverging of instrument tips	While removing tumor
142	Mean	NA	Increased blood in view	While removing tumor
143	Mean	NA	Decreasing bleeding rate	While removing tumor
144	Mean	NA	Decrease in blood in view	While removing tumor
145	Mean	NA	Tip distance of instruments	While removing tumor
146	Mean	NA	Change in tip distance of instruments	While removing tumor
147	Mean	NA	Change in volume of brain tissue	While removing tumor
148	Mean	NA	Total blood emitted	While removing tumor
149	Mean	NA	Change in total blood emitted	While removing tumor
150	Mean	NA	Change in volume of tumor	While removing tumor
151	Median	Aspirator	Acceleration of instrument	While removing tumor
152	Median	Aspirator	Force of instrument	While removing tumor
153	Median	Aspirator	Change in force of instrument	While removing tumor

(continued)

Table 1. Performance Metrics Generated From Raw Simulator Data (continued)

Metric No.	Measurement	Instrument	Performance Measure Associated With Movement, Force, Bleeding, or Tissue	Operative Condition
154	Median	Aspirator	Jerk of instrument	While removing tumor
155	Median	Aspirator	Rate of slowing down of instrument	While removing tumor
156	Median	Aspirator	Rate of speeding up of instrument	While removing tumor
157[c]	Median	Aspirator	Velocity of instrument	While removing tumor
158	Median	Bipolar	Acceleration of instrument	While removing tumor
159[d]	Median	Bipolar	Force of instrument	While removing tumor
160	Median	Bipolar	Change in force of instrument	While removing tumor
161	Median	Bipolar	Jerk of instrument	While removing tumor
162	Median	Bipolar	Rate of slowing down of instrument	While removing tumor
163	Median	Bipolar	Rate of speeding up of instrument	While removing tumor
164	Median	Bipolar	Velocity of instrument	While removing tumor
165	Median	NA	Bleeding speed	While removing tumor
166	Median	NA	Change in bleeding speed	While removing tumor
167	Median	NA	Blood in view	While removing tumor
168[b,d]	Median	NA	Change in blood in view	While removing tumor
169	Median	NA	Increase in bleeding speed	While removing tumor
170	Median	NA	Converging of instrument tips	While removing tumor
171	Median	NA	Diverging of instrument tips	While removing tumor
172	Median	NA	Increased blood in view	While removing tumor
173	Median	NA	Decreasing bleeding rate	While removing tumor
174	Median	NA	Decrease in blood in view	While removing tumor
175	Median	NA	Tip distance of instruments	While removing tumor
176[b]	Median	NA	Change in tip distance of instruments	While removing tumor
177	Median	NA	Change in volume of brain tissue	While removing tumor
178	Median	NA	Total blood emitted	While removing tumor
179	Median	NA	Change in total blood emitted	While removing tumor
180	Median	NA	Change in volume of tumor	While removing tumor
181	Maximum	Aspirator	Acceleration of instrument	While suctioning blood
182	Maximum	Aspirator	Force of instrument	While suctioning blood
183	Maximum	Aspirator	Change in force of instrument	While suctioning blood
184	Maximum	Aspirator	Jerk of instrument	While suctioning blood
185	Maximum	Aspirator	Rate of slowing down of instrument	While suctioning blood
186	Maximum	Aspirator	Rate of speeding up of instrument	While suctioning blood
187	Maximum	Aspirator	Velocity of instrument	While suctioning blood
188	Maximum	Bipolar	Acceleration of instrument	While suctioning blood
189[d]	Maximum	Bipolar	Force of instrument	While suctioning blood
190	Maximum	Bipolar	Change in force of instrument	While suctioning blood
191	Maximum	Bipolar	Jerk of instrument	While suctioning blood
192	Maximum	Bipolar	Rate of slowing down of instrument	While suctioning blood
193	Maximum	Bipolar	Rate of speeding up of instrument	While suctioning blood
194[d]	Maximum	Bipolar	Velocity of instrument	While suctioning blood
195	Maximum	NA	Bleeding speed	While suctioning blood
196	Maximum	NA	Change in bleeding speed	While suctioning blood
197	Maximum	NA	Blood in view	While suctioning blood
198	Maximum	NA	Change in blood in view	While suctioning blood
199	Maximum	NA	Increase in bleeding speed	While suctioning blood
200	Maximum	NA	Converging of instrument tips	While suctioning blood
201	Maximum	NA	Diverging of instrument tips	While suctioning blood
202	Maximum	NA	Increased blood in view	While suctioning blood
203	Maximum	NA	Decreasing bleeding rate	While suctioning blood
204	Maximum	NA	Decrease in blood in view	While suctioning blood

(continued)

Table 1. Performance Metrics Generated From Raw Simulator Data (continued)

Metric No.	Measurement	Instrument	Performance Measure Associated With Movement, Force, Bleeding, or Tissue	Operative Condition
205	Maximum	NA	Tip distance of instruments	While suctioning blood
206	Maximum	NA	Change in tip distance of instruments	While suctioning blood
207	Maximum	NA	Change in volume of brain tissue	While suctioning blood
208	Maximum	NA	Total blood emitted	While suctioning blood
209	Maximum	NA	Change in total blood emitted	While suctioning blood
210	Maximum	NA	Change in volume of tumor	While suctioning blood
211	Mean	Aspirator	Acceleration of instrument	While suctioning blood
212	Mean	Aspirator	Force of instrument	While suctioning blood
213	Mean	Aspirator	Change in force of instrument	While suctioning blood
214	Mean	Aspirator	Jerk of instrument	While suctioning blood
215	Mean	Aspirator	Rate of slowing down of instrument	While suctioning blood
216	Mean	Aspirator	Rate of speeding up of instrument	While suctioning blood
217	Mean	Aspirator	Velocity of instrument	While suctioning blood
218	Mean	Bipolar	Acceleration of instrument	While suctioning blood
219	Mean	Bipolar	Force of instrument	While suctioning blood
220	Mean	Bipolar	Change in force of instrument	While suctioning blood
221	Mean	Bipolar	Jerk of instrument	While suctioning blood
222	Mean	Bipolar	Rate of slowing down of instrument	While suctioning blood
223	Mean	Bipolar	Rate of speeding up of instrument	While suctioning blood
224	Mean	Bipolar	Velocity of instrument	While suctioning blood
225	Mean	NA	Bleeding speed	While suctioning blood
226	Mean	NA	Change in bleeding speed	While suctioning blood
227	Mean	NA	Blood in view	While suctioning blood
228	Mean	NA	Change in blood in view	While suctioning blood
229	Mean	NA	Increase in bleeding speed	While suctioning blood
230	Mean	NA	Converging of instrument tips	While suctioning blood
231	Mean	NA	Diverging of instrument tips	While suctioning blood
232	Mean	NA	Increased blood in view	While suctioning blood
233	Mean	NA	Decreasing bleeding rate	While suctioning blood
234	Mean	NA	Decrease in blood in view	While suctioning blood
235[d]	Mean	NA	Tip distance of instruments	While suctioning blood
236	Mean	NA	Change in tip distance of instruments	While suctioning blood
237	Mean	NA	Change in volume of brain tissue	While suctioning blood
238	Mean	NA	Total blood emitted	While suctioning blood
239	Mean	NA	Change in total blood emitted	While suctioning blood
240	Mean	NA	Change in volume of tumor	While suctioning blood
241	Median	Aspirator	Acceleration of instrument	While suctioning blood
242	Median	Aspirator	Force of instrument	While suctioning blood
243	Median	Aspirator	Change in force of instrument	While suctioning blood
244	Median	Aspirator	Jerk of instrument	While suctioning blood
245	Median	Aspirator	Rate of slowing down of instrument	While suctioning blood
246	Median	Aspirator	Rate of speeding up of instrument	While suctioning blood
247	Median	Aspirator	Velocity of instrument	While suctioning blood
248	Median	Bipolar	Acceleration of instrument	While suctioning blood
249	Median	Bipolar	Force of instrument	While suctioning blood
250[d]	Median	Bipolar	Change in force of instrument	While suctioning blood
251	Median	Bipolar	Jerk of instrument	While suctioning blood
252	Median	Bipolar	Rate of slowing down of instrument	While suctioning blood
253	Median	Bipolar	Rate of speeding up of instrument	While suctioning blood
254	Median	Bipolar	Velocity of instrument	While suctioning blood
255	Median	NA	Bleeding speed	While suctioning blood

(continued)

Table 1. Performance Metrics Generated From Raw Simulator Data (continued)

Metric No.	Measurement	Instrument	Performance Measure Associated With Movement, Force, Bleeding, or Tissue	Operative Condition
256	Median	NA	Change in bleeding speed	While suctioning blood
257	Median	NA	Blood in view	While suctioning blood
258	Median	NA	Change in blood in view	While suctioning blood
259	Median	NA	Increase in bleeding speed	While suctioning blood
260	Median	NA	Converging of instrument tips	While suctioning blood
261	Median	NA	Diverging of instrument tips	While suctioning blood
262	Median	NA	Increased blood in view	While suctioning blood
263	Median	NA	Decreasing bleeding rate	While suctioning blood
264	Median	NA	Decrease in blood in view	While suctioning blood
265[b]	Median	NA	Tip distance of instruments	While suctioning blood
266	Median	NA	Change in tip distance of instruments	While suctioning blood
267	Median	NA	Change in volume of brain tissue	While suctioning blood
268	Median	NA	Total blood emitted	While suctioning blood
269	Median	NA	Change in total blood emitted	While suctioning blood
270	Median	NA	Change in volume of tumor	While suctioning blood

Abbreviation: NA, not applicable.

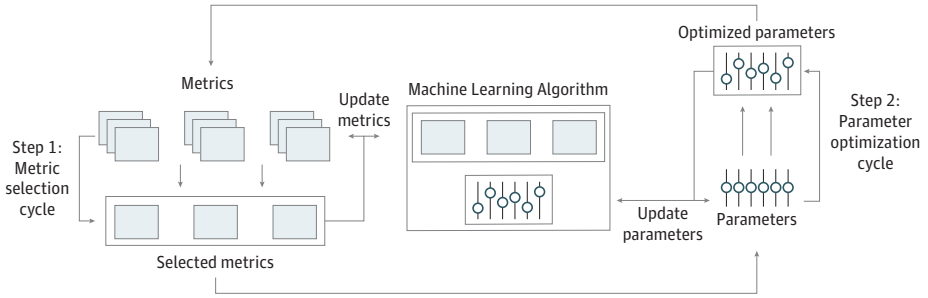
[a] Performance metric selected by naive Bayes algorithm.

[b] Performance metric selected by support vector machine algorithm.

[c] Performance metric selected by K-nearest neighbor algorithm.

[d] Performance metric selected by discriminant analysis algorithm.

Figure 1. The Process of Generating a Final Optimized Machine Learning Algorithm With a Set of Selected Metrics



For algorithm optimization, each machine learning algorithm has a defined set of parameters by which it functions, the adjustment of which will modify its overall performance. An analogy for these parameters is the statistical methods that underlie *P* value adjustments (eg, Bonferroni and Benjamini-Hochberg). MATLAB, release 2018a (MathWorks Inc) was used to modify the intrinsic properties of 4 machine learning algorithms (K-nearest neighbor, naive Bayes, discriminant analysis, and support vector machine).

desired accuracy. eFigure 2 in the Supplement has further information regarding the individual misclassifications of these algorithms.

Machine Learning Optimized Parameters

The final K-nearest neighbor algorithm used included 2 neighbors with a cosine distance calculation. Novel data points were classified into the more skilled group in cases when 2 neighbors were from differing groups.

The best-performing naive Bayes algorithm used gaussian (normal) kernel smoothing with a width of 0.31408. The final discriminant analysis algorithm used a δ value of 0.00068926 and a γ value of 0.99808 with a pseudo-linear discriminant type. The final support vector machine algorithm used a gaussian kernel function with the formula $G(xj, xk) = \exp(-||xj - xk||\chi^2)$. Box constraint was 0.12958 and kernel scale was 3.1667 using the 1-vs-all coding method (in which 1 group is compared with all others).

Performance Metrics Selected by Machine Learning Algorithm

Of the 270 performance metrics generated from raw data, 122 were selected after reduction and normalization. The K-nearest neighbor algorithm used 6 performance metrics to classify participants (55, 63, 67, 123, 125, and 157), the naive Bayes algorithm used 9 performance metrics (9, 34, 39, 44, 51, 55, 60, 69, and 129), the discriminant analysis algorithm used 8 performance metrics (60, 63, 159, 168, 189, 194, 235, and 250), and the support vector machine algorithm used 8 performance metrics

(9, 45, 48, 60, 106, 168, 176, and 265) (Table 1). Performance metrics selected by the algorithms spanned the following 4 principal domains: movement associated with a single instrument, both instruments used in concert, force applied by the instruments, and tissue removed or bleeding caused (**Figure 3**).

Discussion

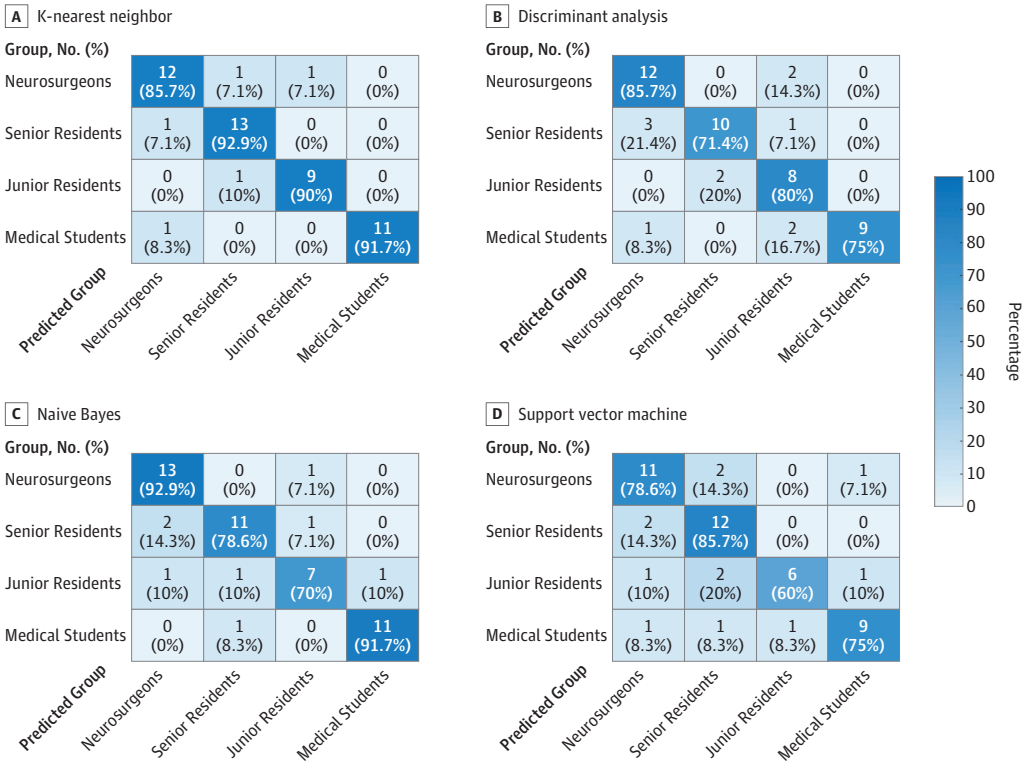
In this prospective study using a high-fidelity virtual reality simulated neurosurgical brain tumor resection procedure, we sought to assess whether machine learning algorithms could select

Table 2. Demographic Information of Participants

Characteristic	Staff Neurosurgeons (n = 14)	Fellows or Senior Residents (n = 14)	Junior Residents (n = 10)	Medical Students (n = 12)
Age, median (range), y	45 (33-59)	33 (29-35)	30 (27-38)	23 (23-26)
Sex, No. (%)				
Male	14 (100)	13 (93)	8 (80)	6 (50)
Female	0	1 (7)	2 (20)	6 (50)
Total No. of years of practice, median (range)	12.5 (1-25)	NA	NA	NA
Neurosurgical subspecialty, No. (%)				
Spine	5 (36)	NA	NA	NA
Oncology and epilepsy	4 (29)	NA	NA	NA
Skull base	2 (14)	NA	NA	NA
Pediatrics	2 (14)	NA	NA	NA
Cerebrovascular	1 (7)	NA	NA	NA

Abbreviation: NA, not applicable.

Figure 2. Individual Misclassifications by Machine Learning Algorithms



Matrix demonstrating actual vs estimated group memberships by 4 different machine learning algorithms. Percentages reflect the total among rows.

performance measures to classify participants according to their level of neurosurgical training. This study comes at a time of ever-increasing time pressure facing physician-educators to balance their commitment to patients and learners.[20] In parallel, in the United States the search continues for a reliable means of examining Part III of the Maintenance of Certification, namely, the assessment of knowledge, judgment, and skills unique to surgical and procedurally oriented medical specialties.[21,22] Both require an objective, consistent, transparent, and defendable means of summative and formative assessments of psychomotor ability.

Simulators, while affording learners the opportunity to safely develop technical skills during the particularly dangerous and error-prone early phases of skill acquisition, do not obviate the need for learner feedback, which is often given by skilled instructors.[23] Furthermore, although simulation has been incorporated into the certification process of the American Board of Surgery and the American Board of Anesthesiology, the former relies on human evaluators while the latter is meant only to stimulate self-reflection.[22,24] Simulation-based technical skills training informed by artificial intelligence feedback systems may offer a solution.

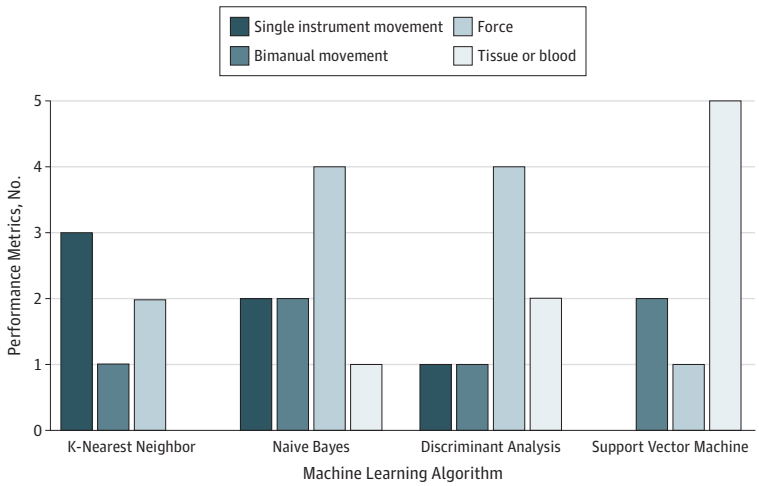
As innovations in artificial intelligence continue, so do the efforts to maintain human understanding of the algorithm classification process. This field has been termed *transparent* or *explainable artificial intelligence*.[25] By understanding the performance data used by the algorithm to render its decision, it is possible to design systems to deliver on-demand assessments at the convenience of the examinee and with minimal input from skilled instructors. Such systems may be subject to continuous improvement as increasing participant data are collected and integrated into the algorithm.

We found that the best-performing machine learning algorithm used as few as 6 performance metrics to successfully classify 45 of 50 participants into 1 of 4 groups of expertise. Although we chose to limit the performance measures to those that could be easily interpreted by a user, theoretically higher accuracies may be attained by including more abstruse metrics. Nevertheless, to our knowledge, no previous study using artificial intelligence to evaluate performance has demonstrated the ability to identify 4 groups in open surgery.[26-37]

Limitations

Insofar as technical skills measured on a simulator are reflective of operating skill in the real world, our findings outline a novel approach to understanding technical expertise in surgery. Although 4 different machine learning algorithms were used, there still exists the possibility that all algorithms are overfitted to our data set, limiting their performance when faced with novel data.[38] As such, these algorithms must be tested on an independent data set before making final conclusions about

Figure 3. Number of Performance Metrics Selected for By 4 Different Machine Learning Algorithms



Performance metrics are categorized as those involving movements of 1 or both instruments, force applied to the underlying structures and damage to underlying brain, blood loss, and quantity of tumor removed.

their accuracy. Furthermore, in 3 of 4 algorithms a single medical student was categorized as a neurosurgeon. In response to this misclassification, we sought to limit misclassifications between these 2 groups in the algorithm optimization process as a proof of concept. Although this modification came at a cost of reduced overall accuracy, explicitly preventing misclassifications between certain groups may be desirable in high-stakes certification examinations.

In addition, it is challenging to define populations of surgeons, fellows, and residents with equivalent skill to allow accurate classification. Neurosurgeon skill level was based on being a certified surgeon and resident skill level was based on their educational year, which does not adequately take into account subspecialization or other construct-validated objective assessments of skill sets. A more comprehensive evaluation of participants with an emphasis on demonstrated skills across assessment domains (eg, visual rating scales and training evaluations or assessment of visuospatial abilities) may result in improved algorithm performance.

Conclusions

Our study demonstrates the ability of machine learning algorithms to classify surgical expertise with greater granularity and precision than has been previously demonstrated. Although the task involved a complex neurosurgical tumor resection task, the protocol outlined can be applied to any digitized platform to assess performance in a setting in which technical skill is paramount.

ARTICLE INFORMATION

Accepted for Publication: June 10, 2019.

Published: August 2, 2019. doi:10.1001/jamanetworkopen.2019.8363

Corresponding Author: Alexander Winkler-Schwartz, MD, Neurosurgical Simulation and Artificial Intelligence Learning Centre, Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, 3801 University St, E2.89 Montreal, QC H3A 2B4, Canada (manuscriptinquiry@gmail.com).

Author Affiliations: Neurosurgical Simulation and Artificial Intelligence Learning Centre, Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada (Winkler-Schwartz, Yilmaz, Mirchi, Bissonnette, Ledwos, Siyar, Azarnoush, Karlik, Del Maestro); Department of Orthopedic Surgery, McGill University, Montreal, Quebec, Canada (Bissonnette); Department of Biomedical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran (Siyar, Azarnoush).

Author Contributions: Drs Winkler-Schwartz and Yilmaz had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Winkler-Schwartz and Yilmaz are co–first authors.

Concept and design: Winkler-Schwartz, Yilmaz, Mirchi, Bissonnette, Ledwos, Siyar, Azarnoush, Del Maestro.

Acquisition, analysis, or interpretation of data: Winkler-Schwartz, Yilmaz, Karlik.

Drafting of the manuscript: Winkler-Schwartz, Yilmaz, Siyar, Azarnoush.

Critical revision of the manuscript for important intellectual content: Winkler-Schwartz, Yilmaz, Mirchi, Bissonnette, Ledwos, Karlik, Del Maestro.

Statistical analysis: Winkler-Schwartz, Yilmaz, Bissonnette, Karlik.

Obtained funding: Del Maestro.

Administrative, technical, or material support: Del Maestro.

Supervision: Azarnoush, Del Maestro.

Conflict of Interest Disclosures: Dr Winkler-Schwartz reported receiving grants from Fonds de Recherche du Québec–Santé, Robert Maudsley Fellowship for the Royal College of Physicians and Surgeons of Canada, and Di Giovanni Foundation during the conduct of the study and having a patent pending to Method and System for Generating a Training Platform. Dr Yilmaz reported receiving grants from AO Foundation during the conduct of the study and having a patent pending to Method and System for Generating a Training Platform (2019; patent No.

05001770-843USPR). Mr Mirchi reported receiving grants from AO Foundation and Di Giovanni Foundation during the conduct of the study and having a patent pending to Method and System for Generating a Training Platform. Ms Ledwos reported receiving grants from AO Foundation and Di Giovanni Foundation during the conduct of the study and having a patent pending to Method and System for Generating a Training Platform. Dr Del Maestro reported receiving grants from AO Foundation, Di Giovanni Foundation, and Robert Maudsley Fellowship for Studies in Medical Education during the conduct of the study; receiving personal fees from AO Foundation and CAE outside the submitted work; having a patent pending to Methods and System for Generating a Training Platform; being a coauthor and neurosurgeon working with 3 engineers on a manuscript that was the first description of the NeuroTouch Simulator developed by the Medical Research Council of Canada, which was later taken over by CAE in 2016 who renamed the system NeuroVR; and being a visiting researcher at the Medical Research Council of Canada. No other disclosures were reported.

Funding/Support: This work was supported by the Di Giovanni Foundation, the Montreal English School Board, the Montreal Neurological Institute, the McGill Department of Orthopedics, 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. The Medical Research Council of Canada has provided a prototype of the NeuroTouch that was used in this study.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank all those who participated in this study.

REFERENCES

1. Anderson O, Davis R, Hanna GB, Vincent CA. Surgical adverse events: a systematic review. *Am J Surg*. 2013;206(2):253-262. doi:10.1016/j.amjsurg.2012.11.009

2. Maier-Hein L, Vedula SS, Speidel S, et al. Surgical data science for next-generation interventions. *Nat Biomed Eng*. 2017;1(9):691-696. doi:10.1038/s41551-017-0132-7

3. 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*. New York, NY: Springer New York; 2013:415-423. doi:10.1007/978-1-4614-5993-4_28

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

5. Bugdadi A, Sawaya R, Olwi D, et al. Automaticity of force application during simulated brain tumor resection: testing the Fitts and Posner model. *J Surg Educ*. 2018;75(1):104-115. doi:10.1016/j.jsurg.2017.06.018

6. Winkler-Schwartz A, Bajunaid K, Mullah MAS, et al. Bimanual psychomotor performance in neurosurgical resident applicants assessed using NeuroTouch, a virtual reality simulator. *J Surg Educ*. 2016;73(6):942-953. doi:10.1016/j.jsurg.2016.04.013

7. Bajunaid K, Mullah MA, Winkler-Schwartz A, et al. Impact of acute stress on psychomotor bimanual performance during a simulated tumor resection task. *J Neurosurg*. 2017;126(1):71-80. doi:10.3171/2015.5.JNS15558

8. AlZhrani G, Alotaibi F, Azarnoush H, et al. Proficiency performance benchmarks for removal of simulated brain tumors using a virtual reality simulator NeuroTouch. *J Surg Educ*. 2015;72(4):685-696. doi:10.1016/j.jsurg.2014.12.014

9. Alotaibi FE, AlZhrani GA, Mullah MA, et al. Assessing bimanual performance in brain tumor resection with NeuroTouch, a virtual reality simulator. *Neurosurgery*. 2015;11(1)(suppl 2):89-98.

10. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053

11. 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 [published online June 13, 2019]. *J Surg Educ*. doi:10.1016/j.jsurg.2019.05.015

12. Cheng A, Kessler D, Mackinnon R, et al; International Network for Simulation-based Pediatric Innovation, Research, and Education (INSPIRE) Reporting Guidelines Investigators. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Simul Healthc*. 2016;11(4):238-248. doi:10.1097/SIH.0000000000000150

13. Hebb AO, Yang T, Silbergeld DL. The sub-pial resection technique for intrinsic tumor surgery. *Surg Neurol Int*. 2011;2:180. doi:10.4103/2152-7806.90714

14. Bugdadi A, Sawaya R, Bajunaid K, et al. Is virtual reality surgical performance influenced by force feedback device utilized? *J Surg Educ*. 2019;76(1):262-273. doi:10.1016/j.jsurg.2018.06.012

15. MathWorks. fitcecoc: fit multiclass models for support vector machines or other classifiers. https://www.mathworks.com/help/stats/fitcecoc.html. Accessed January 9, 2019.

16. MathWorks. fitcsvm: Train support vector machine (SVM) classifier for one-class and binary classification. https://www.mathworks.com/help/stats/fitcsvm.html. Accessed January 9, 2019.

17. MathWorks. fitcdiscr: Fit discriminant analysis classifier. https://www.mathworks.com/help/stats/fitcdiscr.html. Accessed January 9, 2019.

18. MathWorks. fitcnb: Train multiclass naive Bayes model. https://www.mathworks.com/help/stats/fitcnb.html. Accessed January 9, 2019.

19. MathWorks. fitcknn: Fit *k*-nearest neighbor classifier. https://www.mathworks.com/help/stats/fitcknn.html. Accessed January 9, 2019.

20. Spencer J. Learning and teaching in the clinical environment. *BMJ*. 2003;326(7389):591-594. doi:10.1136/bmj.326.7389.591

21. American Board of Medical Specialties. Steps toward initial certification and MOC. https://www.abms.org/board-certification/steps-toward-initial-certification-and-moc/. Accessed January 23, 2019.

22. Ross BK, Metzner J. Simulation for maintenance of certification. *Surg Clin North Am*. 2015;95(4):893-905. doi:10.1016/j.suc.2015.04.010

23. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 2005;27(1):10-28. doi:10.1080/01421590500046924

24. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387-396. doi:10.1056/NEJMra054784

25. Abdul A, Vermeulen J, Wang D, Lim B, Kankanhalli M. Trends and trajectories for explainable, accountable and intelligible systems: an HCI research agenda. Presented at: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC.

26. Murphy TE, Vignes CM, Yuh DD, Okamura AM. Automatic motion recognition and skill evaluation for dynamic tasks. *Eurohaptics*. 2003;2003:363-373.

27. Megali G, Sinigaglia S, Tonet O, Dario P. Modelling and evaluation of surgical performance using hidden Markov models. *IEEE Trans Biomed Eng*. 2006;53(10):1911-1919. doi:10.1109/TBME.2006.881784

28. Hajshirmohammadi I, Payandeh S. Fuzzy set theory for performance evaluation in a surgical simulator. *Presence*. 2007;16(6):603-622. doi:10.1162/pres.16.6.603

29. Jog A, Itkowitz B, May L, et al. Towards integrating task information in skills assessment for dexterous tasks in surgery and simulation. Presented at: 2011 IEEE International Conference on Robotics and Automation; May 9-13, 2011; Shanghai, China.

30. Liang H, Shi MY. Surgical skill evaluation model for virtual surgical training. *Appl Mech Mater*. 2011;40-41:812-819. doi:10.4028/www.scientific.net/AMM.40-41.812

31. Loukas C, Georgiou E. Multivariate autoregressive modeling of hand kinematics for laparoscopic skills assessment of surgical trainees. *IEEE Trans Biomed Eng*. 2011;58(11):3289-3297. doi:10.1109/TBME.2011.2167324

32. Huang J, Payandeh S, Doris P, Hajshirmohammadi I. Fuzzy classification: towards evaluating performance on a surgical simulator. *Stud Health Technol Inform*. 2005;111:194-200.

33. 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. doi:10.3109/10929080801957712

34. Richstone L, Schwartz MJ, Seideman C, Cadeddu J, Marshall S, Kavoussi LR. Eye metrics as an objective assessment of surgical skill. *Ann Surg*. 2010;252(1):177-182. doi:10.1097/SLA.0b013e3181e464fb

35. Rhienmora P, Haddawy P, Suebnukarn S, Dailey MN. Intelligent dental training simulator with objective skill assessment and feedback. *Artif Intell Med*. 2011;52(2):115-121. doi:10.1016/j.artmed.2011.04.003

36. Kerwin T, Wiet G, Stredney D, Shen HW. Automatic scoring of virtual mastoidectomies using expert examples. *Int J Comput Assist Radiol Surg*. 2012;7(1):1-11. doi:10.1007/s11548-011-0566-4

37. Ershad M, Rege R, Fey AM. Meaningful assessment of robotic surgical style using the wisdom of crowds. *Int J Comput Assist Radiol Surg*. 2018;13(7):1037-1048. doi:10.1007/s11548-018-1738-2

38. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920-1930. doi:10.1161/CIRCULATIONAHA.115.001593

SUPPLEMENT.

eFigure 1. Screen Capture of the Comma Separated Value File Representing the Output of the Simulator

eFigure 2. Individual Misclassifications of Machine Learning Algorithms Emphasizing No Misclassifications Between Neurosurgeons and Medical Students