Machine learning and neurosurgical outcome prediction: a systematic review

Senders, J.T.; Staples, P.C.; Karhade, A.V.; Zaki, M.M.; Gormley, W.B.; Broekman, M.L.D.; ... ; Arnaout, O.

Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review

Joeky T. Senders[1,2], Patrick C. Staples[3], Aditya V. Karhade[2], Mark M. Zaki[2], William B. Gormley[2], Marike L.D. Broekman[1,2], Timothy R. Smith[2], Omar Arnaout[2]

Key words
- Artificial intelligence
- Machine learning
- Neurosurgery
- Prediction

Abbreviations and Acronyms
AUC: Area under the receiver operating curve
AVM: Arteriovenous malformation
EEG: Electroencephalography
ML: Machine learning
MRI: Magnetic resonance imaging
UPDRS: Unified Parkinson's Disease Rating Scale

From the [1]Department of Neurosurgery, University Medical Center Utrecht, Utrecht, The Netherlands; [2]Computational Neurosciences Outcomes Center, Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; and [3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA

To whom correspondence should be addressed: Omar Arnaout, M.D. [E-mail: oarnaout@bwh.harvard.edu]

Supplementary digital content available online.

Citation: World Neurosurg. (2018) 109:476-486.
https://doi.org/10.1016/j.wneu.2017.09.149
Journal homepage: www.WORLDNEUROSURGERY.org

Available online: www.sciencedirect.com

OBJECTIVE: **Accurate measurement of surgical outcomes is highly desirable to optimize surgical decision-making. An important element of surgical decision making is identification of the patient cohort that will benefit from surgery before the intervention. Machine learning (ML) enables computers to learn from previous data to make accurate predictions on new data. In this systematic review, we evaluate the potential of ML for neurosurgical outcome prediction.**

METHODS: **A systematic search in the PubMed and Embase databases was performed to identify all potential relevant studies up to January 1, 2017.**

RESULTS: **Thirty studies were identified that evaluated ML algorithms used as prediction models for survival, recurrence, symptom improvement, and adverse events in patients undergoing surgery for epilepsy, brain tumor, spinal lesions, neurovascular disease, movement disorders, traumatic brain injury, and hydrocephalus. Depending on the specific prediction task evaluated and the type of input features included, ML models predicted outcomes after neurosurgery with a median accuracy and area under the receiver operating curve of 94.5% and 0.83, respectively. Compared with logistic regression, ML models performed significantly better and showed a median absolute improvement in accuracy and area under the receiver operating curve of 15% and 0.06, respectively. Some studies also demonstrated a better performance in ML models compared with established prognostic indices and clinical experts.**

CONCLUSIONS: **In the research setting, ML has been studied extensively, demonstrating an excellent performance in outcome prediction for a wide range of neurosurgical conditions. However, future studies should investigate how ML can be implemented as a practical tool supporting neurosurgical care.**

INTRODUCTION

"The decision is more important than the incision."[1] This statement is meant to highlight the cardinal importance of individual risk/benefit analysis that should be made in each individual patient. For neurosurgical interventions, the hope of a positive outcome is always intimately tied to the risk of an unfavorable consequence. Furthermore, benefits by one measure can come at the cost of another. Extending survival by means of an aggressive tumor resection, for example, could come at the cost of impaired functional neurologic status.

Clinical expertise can predict surgical outcome with a high degree of accuracy based on experience and available evidence.[2,3] The results from clinical studies, however, are averaged estimates of patient cohorts but do not directly apply to each individual patient. Also, subjectivity can influence clinical judgement. This subjectivity is inherently translated to the patient's expectations, and the hierarchical relation between doctors and patients can make patients incline toward the personal preferences of the surgeon.

To date, no adequate tools have been developed that accurately predict surgical outcomes in the individual patient, which can be used to aid patients and physicians in the process of surgical decision making. Prognostic indices are easy to apply in clinical practice; however, this applicability goes at the cost of their predictive performance.[4,5] To calculate prognosis, numerical values often are simplified as categorical variables, the weight given to predictive factors is rounded up to integers, and only a limited subset of variables can be included in the prognostic index.

Machine learning (ML) is a branch of artificial intelligence and is entering the

realm of clinical research at an increasing pace. ML enables computer algorithms to learn from experience, without explicitly being programmed.[6,7] ML is driven by a data explosion combined with increasing computational power, and classical epidemiology is now incorporating newer data science techniques to harness the power within population data.[8] By studying large data sets, these tools seek to approximate the clinically meaningful relationships between input and output parameters. The learning aspect makes them very powerful prediction algorithms that can model previously unknown relationships in large, complex data sets and adapt to dynamic data environments.

The complex diagnostic and therapeutic modalities used in neurosurgery provide a vast array of multidimensional and variegated data and therefore an opportune framework for the creation of ML models. This suggests a vast potential for the application of ML in neurosurgical care and supports a growing trend toward precision medicine in which therapy is tailored to the individual patient. Previously, we have compared the performance of ML with clinical judgment across the wide spectrum of neurosurgical care.[9] This suggested a great potential for neurosurgical outcome prediction. Since prediction tasks lie at the core of most ML approaches,[6] we have performed an in-depth analysis of ML used for neurosurgical outcome prediction. The aim of this systematic review is to provide a brief introduction into the theoretical concepts of ML and to evaluate its usefulness to aid neurosurgical decision-making. It also evaluates the performance of ML compared with prognostic indices, traditional statistical models, and clinical experts.

MACHINE LEARNING

Within the field of ML, a broad distinction can be made between supervised and unsupervised learning. Supervised learning algorithms learn from "labeled" training data to produce a model that can make predictions on previously unseen data.[10] The desired output for these training data is known; therefore, it is referred to as labeled. This learning aspect indicates the difference with traditional programming. In traditional programming, a programmer manually writes a set of instructions—"the program"—to generate a desired output from a given set of input variables. In ML, the input is provided together with the desired output, and computer algorithms are asked to derive the "rules" from the labeled training data. The product of this process is, therefore, not the desired output but a model that can predict the output in previously unseen data (Figure 1). The automated learning process is an efficient way of analyzing large quantities of data, modeling hidden relationships in complex data sets, and adapting to changing environments.[11-13] In the learning process, algorithms try to find the optimal combination of input variables (features) and weights given to these features in the model, thereby minimizing the difference between the predicted and actual outcomes. The mathematical structures of the algorithms most frequently used in neurosurgical outcome prediction tasks are briefly outlined in Figure 2.

A medical application of a ML model could be survival prediction in patients with brain tumor. Supervised learning algorithms are trained on historical patients for whom the length of survival is known. Meaningful features in the prediction model could, for example, be age, tumor grading, and functional neurologic status. If the model is too complex, such as containing too many features relative to the number of cases, the model could overfit the data characterized by predicting noise. In our example, overfitting can mean including clinically irrelevant features such as hair color or shoe size. This reduces the prediction error in the training set but at the cost of a reduced generalizability to previously unseen data.

To overcome this problem, an ML model should include a validation procedure. A basic technique is to use part of the data for training and judging the ML model's performance on the remainder, or *test set*. There are 2 competing concerns with this approach. Using a very high proportion of data for training results in more stable parameter estimates and typically increases the performance of the chosen measure of accuracy yet increases the chance that the ML model is tested on a nonrepresentative selection of cases. In contrast, allocating too much data for testing can leave the model poorly trained. A common allocation tradeoff used in practice is a proportion of 2:1 for training and testing. Another technique, used especially in situations in which data are limited, is cross-validation. Cross-validation divides data into multiple partitions (or *folds*), where each fold in turn is held out treated as the test set and the rest are

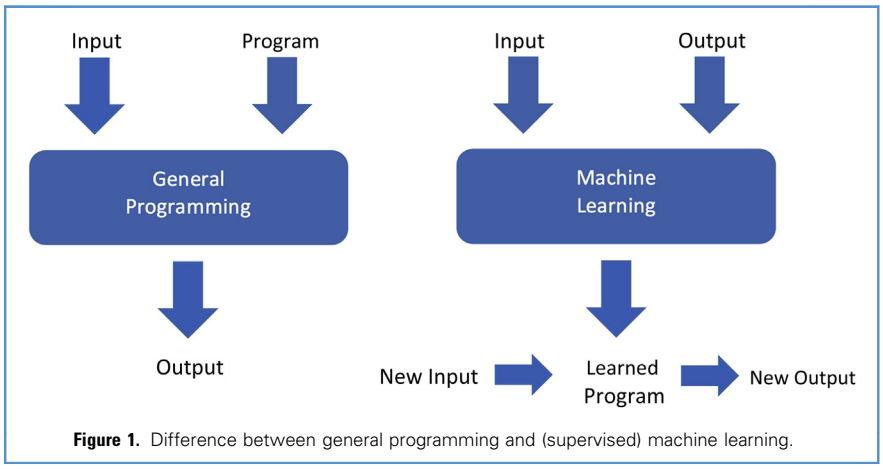

Figure 1. Difference between general programming and (supervised) machine learning.

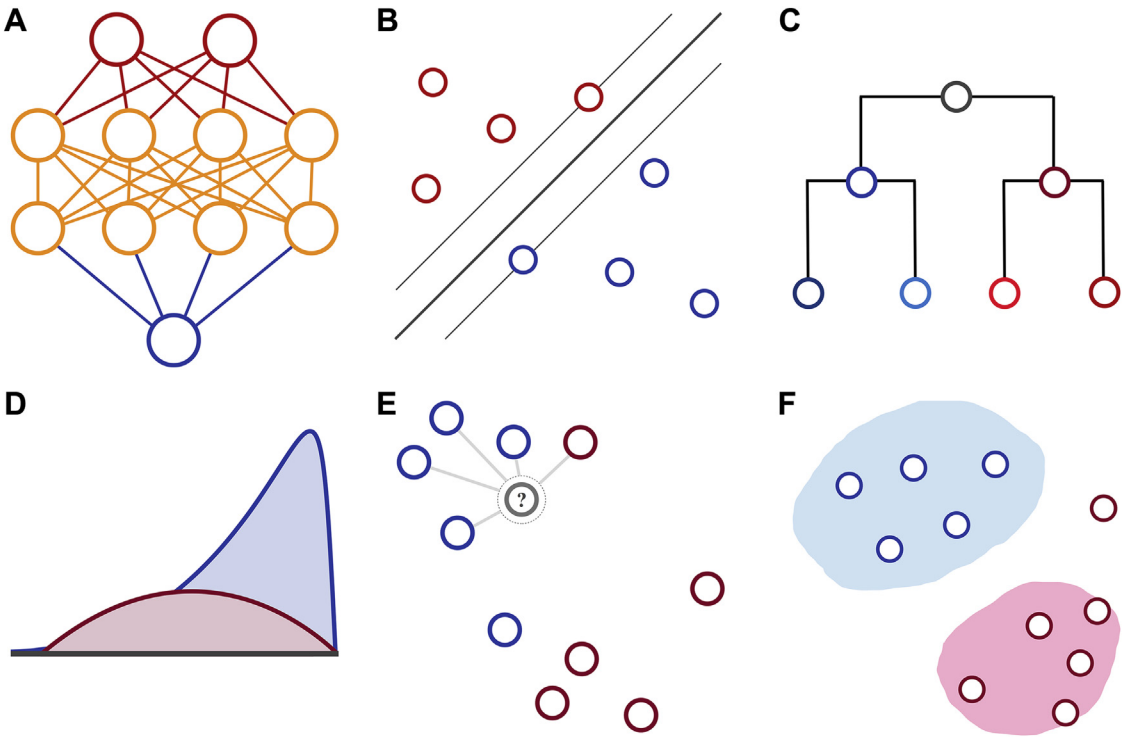

Figure 2. Explanation of the most frequently used machine learning prediction algorithms. (**A**) Artificial neural networks are inspired on the neural networks in the brain and organized in layers of interconnected nodes. The nodes in the upper layer (*red*) represent the input features, and the node in the lower layer (*blue*) represents the distinct output. The nodes in the "hidden" (*orange*) and output layers (*blue*) base the value of their output on the total input they receive. Although artificial neural networks can model very complex relationships between in- and output features, it remains difficult to interpret these explicit relationships. (**B**) Support vector machines classify data points by selecting the "separating hyperplane" that maximize the distance from the 2 closest points on either side to increase the generalizability to unseen data. (**C**) Decision trees make predictions or classifications based on several input features with the use of bifurcating the feature space. These algorithms try to find the optimal features at which a split is made and the optimal value in case of a continuous feature. Random forests is an ensemble learning method that takes the mode of classes or the mean predictions of the individual trees, to avoid overfitting of a single tree. (**D**) Naïve Bayes calculates the most likely outcome (*blue*) as a product of the a priori chance (*red*) and the conditional probabilities given by the individual features, which is often not strictly true, but typically is quickly computed and provides competitive prediction in practice. (**E**) The K-Nearest Neighbors compares a data point with unknown class to its K nearest neighbors, and determines its class as the most common class of its neighbors. For K = 1, the algorithm assigns the class of a data point to the class of the single closest neighbor. (**F**) Fuzzy C-Means is an unsupervised learning algorithm that clusters data points based on their input features without having a desired output. The "fuzzy" aspect gives the algorithm the flexibility to classify a data point to each cluster to a certain degree relating to the likelihood of belonging to that cluster.

used for training, and performance results are averaged across each test fold. Common used partitions are 5-fold, 10-fold, and leave-one-out cross-validation. Testing and cross-validation are not mutually exclusive, and both may be used to increase model robustness.

For unsupervised learning techniques, on the other hand, only unlabeled data are available and the algorithm looks to find similarities and patterns. Expanding on the previous example, unsupervised algorithms can take large sets of unlabeled genomic data as input and would identify previously unknown clusters. These algorithms can be powerful tools for detecting previously unknown patterns in multidimensional data that may not be prima facie detectable by humans[11] and also can be used to generate labels to subsequently train a supervised model.

METHODS

A systematic search in the PubMed and Embase databases has been performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines to identify all

potential relevant studies up to January 1, 2017. The search syntax was built with the guidance of a professional librarian using search terms related to "Machine learning" and "Neurosurgery." The exact search syntax for the PubMed and Embase databases is provided in **Supplementary Table 1**.

Studies were included if they evaluated ML models for application in neurosurgical outcome prediction. Exclusion criteria were conference abstracts, lack of full text, languages other than English and Dutch, animal studies, and studies in which not all participants were surgically treated at baseline. Study selection was done by 2 independent authors (J.S., A.K.), and data extraction was done by 2 independent authors (J.S., M.Z.). Disagreements were solved by discussion including a third author (O.A.).

Data obtained from each study were 1) year of publication; 2) disease condition; 3) type of operation; 4) prediction task; 5) level of measurement of output; 6) number of output classes; 7) ML model used; 8) input features; 9) size of training set; 10) size validation set/validation method; 11) size test set; 12) follow-up time; 13) statistical measures used for evaluation; and 14) prediction performance.

We considered a quantitative synthesis to be inappropriate due to the heterogeneity in neurosurgical applications. A qualitative synthesis of results and assessment of risk of bias on outcome, study, and review level is provided by means of a narrative approach. However, to summarize the findings in some quantitative form, the median accuracy and area under the receiver operating curve (AUC) of the prediction performance were calculated for all studies, and the median absolute improvement in performance was calculated for all studies comparing ML models versus traditional logistic regression. Accuracy refers to the proportion of correct predictions among the total number of predictions, and the AUC corresponds to the probability that a binary classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

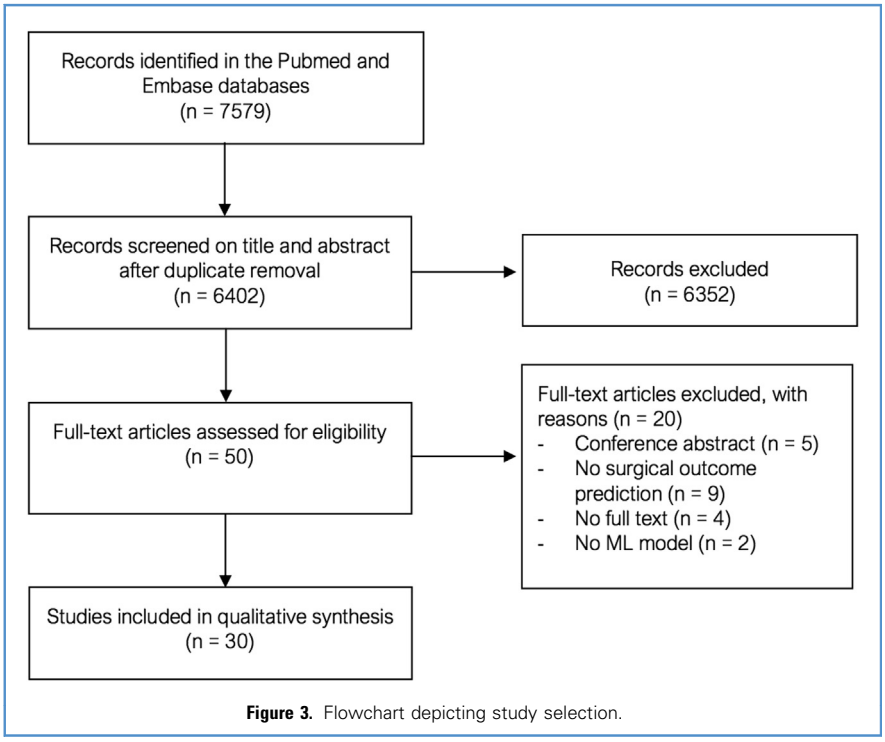

Figure 3. Flowchart depicting study selection.

RESULTS

After removal of duplicates, a total of 6402 citations in the PubMed and Embase databases were identified. Fifty potential relevant studies were selected by title/abstract screening, of which 30 remained after full-text screening (**Figure 3**).

ML models were used to predict outcomes after neurosurgery for epilepsy,[14-22] brain tumors,[2,3,23-26] spinal disease,[27-31] Parkinson disease,[32,33] aneurysmal subarachnoid hemorrhage,[34,35] traumatic brain injury,[36] chronic subdural hemorrhage,[37] arteriovenous malformation (AVM),[5,38] and patients with hydrocephalus (**Table 1**).[27,39] Most frequently used algorithms were artificial neural networks[4,15,19,20,26,27,30,31,34-39] and support vector machines.[3,5,14,18,20,21,24,28,33,38,40] Other algorithms used were random forests,[5,22,32,33] naïve bayes,[16,20,33] linear discriminant analysis,[17,20] K-means,[17] deep learning,[21] k-nearest neighbor,[16] Chi-square Automatic Interaction Detector,[25] fuzzy C-means,[2] boosting,[5] random trees,[5] principal component analysis,[23] and fuzzy inference system.[29] Input features used by the algorithms were magnetic resonance imaging (MRI),[15,17-21,38] computed tomography,[38] clinical features (age, sex, symptoms, signs, disease history, family history, medicine usage etc.),[15,16,20,26,30,31,35,38,39] intracranial electroencephalography (EEG),[14,15,19,25] normal EEG,[15,19] pathology,[15,16,25] angiogram,[35] neuropsychological testing,[19] hospital volume (cases per year),[36] surgeon's volume (cases per years),[36] microelectrode recordings during deep brain stimulation,[32] and surgical factors (surgery time, emergency status, intra- or postoperative complications).[29,38]

Neurosurgical Outcome Prediction

For epilepsy surgery, 9 studies used ML models to predict seizure freedom after resection for temporal lobe epilepsy.[14-22] Input features used were clinical,[15,16,20]

Table 1. Studies Evaluating Machine Learning Algorithms Used for Neurosurgical Outcome Prediction

First Author, Year of Publication, Reference	Disease Condition	Operation	Input Features	Output	Machine Learning Model	Size Training Set	Validation Method	Size Test Set	No. of Output Classes	Prediction Performance (CA/AUC)	Minimal Follow-Up in months
Epilepsy											
Bernhardt et al., 2015[17]	TLE	Resection nos	MRI	Free of seizures	LDA, K-Means	79	LOOCV	27	2	92%/—	49*
Memarian et al., 2015[20]	TLE	ATL	Clin, EEG, MRI	Free of seizures	ANN, SVM, LDA, NB	20	LOOCV	—	2	95%/—	54*
Munsell et al., 2015[21]	TLE	ATL	MRI	Free of seizures	ANN, SVM	70	10-FCV	—	2	70%/—	12
Yankam et al., 2015[22]	TLE	Resection nos	^{18}F-PET/^{11}C-PET	Free of seizures	RF	16	LOOCV	—	2	88%/0.63	12
Feis et al., 2013[18]	TLE	Resection nos	MRI	Free of seizures	SVM	49	LOOCV	—	2	95%/0.94	NA
Antony et al., 2013[14]	TLE	ATL	(i)EEG	Free of seizures	SVM	23	LOOCV	—	2	87%/—	12
Armañanzas et al., 2013[16]	TLE	Resection nos	Clin, PA,	Free of seizures	LR, KNN, NB	23	LOOCV	—	2	89%/0.93	24
Arle et al., 1999[15]	TLE	Resection nos	Clin, NP, (i)EEG, MRI, PA	Free of seizures	ANN	80	5-FCV	—	2	98%/—	6
Grigsby et al., 1998[19]	TLE	ATL	NP, (i)EEG, MRI	Free of seizures	ANN	65	—	22	2	95%/—	12
Neuro-oncology											
Akbari et al., 2016[23]	GBM	GTR	MRI	Recurrence	PCA, SVM	31	LOOCV	34	2	91%/0.84	7*
Azimi et al., 2015[31]	Metastasis	SRS	Clin	Recurrence	ANN	96	48	48	2	95.3%/0.88	6
Knoll et al., 2016[25]	Metastasis	SRS	Clin, (i)EEG, PA	Survival	CHAID	NA	—	NA	2	—/0.80	12
Emblem et al., 2015[3]	GBM	B, STR, GTR	MRI	Survival	SVM	101	—	134	2	—/0.76—0.85	6
Emblem et al., 2014[24]	Glioma	B, STR, GTR	MRI	Survival	SVM	47	—	47	2	—/0.76—0.82	12
Emblem et al., 2009[2]	Glioma	Resection, B	MRI	Survival	FCM	—	—	50	2	—/—	24
Spine											
Azimi et al., 2016[30]	LDH	Discectomy	Clin	Success	ANN	101	51	51	4	95.8%/0.82	24
Azimi et al., 2015[31]	LDH	Discectomy	Clin	Recurrence	ANN	201	101	100	4	94.1%/0.83	12
Hoffman et al., 2015[28]	CSM	Laminectomy, ACD	Clin, MRI	ODI score	SVM	20	—	NA	—†	—/—	6
Azimi et al., 2014[27]	LSS	Laminectomy	Clin, MRI	Patient satisfaction	ANN	84	42	42	2	97%/0.80	24

Shamim et al., 2009[29]	LDH	Discectomy	Clin, Sur, MRI	Failed disk surgery‡	FIS	501	—	NA	2	87%/—	6
Vascular											
Asadi et al., 2016[38]	AVM	Embolization	Clin, CT, MRI, Sur	ICH, stroke, mortality	ANN, SVM	199	LOOCV	—	2	98%/—	63*
Oermann et al., 2016[5]	AVM	SRS	Clin, MRI	Success/ complication§	SVM, Bo, RF, RT	1442	CV (nos)	—	3	74%/0.71	24
Lo et al., 2013[34]	aSAH	Clipping	Clin	GOS	ANN	2130	—	1420	5	—/0.85	3
Dumont et al., 2011[25]	aSAH	Clipping/ endovasc	Clin, angiogram	Cerebral vasospasm	ANN	91	—	22	2	—/0.96	NA
Abouzari et al., 2009[37]	cSDH	Burr-hole	Clin	GOS	ANN	150	75	75	5	73%/0.77	3
Other											
Habibi et al., 2016[39]	Hydro	VPS	Clin	VPS infection	ANN	126	—	22	2	—/0.83	6
Kostoglou et al., 2017[32]	Parkinson	DBS	MERs	Motor improvement‖	RF	20	NA	NA	2/—†	95%/—	24
Shamir et al., 2015[33]	Parkinson	DBS	Clin	Motor improvement¶	SVM, NB, RF	10	—	14	3	86%/—	9
Azimi et al., 2014[27]	Hydro	ETV	Clin	Successful ETV#	ANN	84	42	42	2	95%/0.87	6
Shi et al., 2013[36]	TBI	DC, HE	Clin, HF	In-hospital mortality	ANN	11304	—	5652	2	95%/0.89	1*

CA, classification accuracy; AUC, area under the receiver operating curve; TLE, temporal lobe epilepsy; nos, not otherwise specified; MRI, magnetic resonance imaging; LDA, linear discriminant analysis; K-means, the K is a parameter that refers to the number of clusters; LOOCV, leave-one-out cross validation; ATL, anterior temporal lobectomy; Clin, clinical; EEG, electroencephalography; ANN, artificial neural networks; SVM, support vector machine; FCV, fold cross validation; ^{18}F-PET, fluorodeoxyglucose-positron emission tomography; ^{11}C-PET, flumazenil—positron emission tomography; RF, random forests; (i)EEG, (intra-cranial) electroencephalography; PA, pathology; LR, logistic regression; KNN, K-nearest neighbors; NB, naïve Bayes; —, not applicable; NP, neuropsychological factors; GBM, glioblastoma multiforme; GTR, gross-total resection; PCA, principal component analysis; SRS, stereotactic radiosurgery; CHAID, Chi-square Automatic Interaction Detector; NA, not available; B, biopsy; STR, subtotal resection; FCM, fuzzy C-means; LDH, lumbar disk hernia; CSM, cervical spondylotic myelopathy; ACD, anterior cervical discectomy; ODI, Oswestry Disability Index; LSS, lumbar spinal stenosis; Sur, surgical factors; FIS, fuzzy inference system; AVM, arteriovenous malformation; CT, computed tomography; ICH, intracranial hemorrhage; Bo, boosting; RT, random trees; CV, cross-validation; aSAH, aneurysmal subarachnoid hemorrhage; GOS, Glasgow Outcome Score; cSDH, chronic subdural hematoma; VPS, ventricular peritoneal shunt; DBS, deep brain stimulation; MERs, micro-electrode recordings; Hydro, hydrocephalus; ETV, endoscopic third ventriculostomy; TBI, traumatic brain injury; DC, decompressive craniectomy; HE, hematoma evacuation; HF, hospital-related factors (surgeon's volume, hospital's volume).

*No minimal follow-up was provided. The median follow-up was provided instead.

†Continuous outcome.

‡Defined as 1) no improvement/worsening of symptoms versus 2) complete/partial improvement.

§Defined as 1) obliteration of AVM; 2) no obliteration of AVM and no unfavorable outcome; 3) unfavorable outcome: postradiosurgery hemorrhage or permanently symptomatic radiation-induced changes.

‖Binary classification by dichotomized Unified Parkinson's Disease Rating Scale (UPDRS) and quantified improvement by normalized mean squared error of UPDRS improvement.

¶Trinary classification categorized as 1) nonresponsive (<35% UPDRS improvement); 2) moderate response (35%—65% UPDRS improvement); 3) high response (>65% UPDRS improvement).

#Binary classification categorized as 1) absence of ETV failure versus 2) ETV failure defined as surgical intervention for definitive cerebrospinal fluid diversion or death due to hydrocephalus treatment.

neuropsychological,[19] EEG,[15,19,20] intracranial-EEG,[14,15,19] MRI,[15,17-21] proton emission tomography,[22] and pathology.[15,16] In all 9 studies, a binary classification model was used classifying patients in seizure freedom (Engel's Class I) or no seizure freedom (Engel's Class II-IV), with a minimal follow up ranging between 6 and 24 months (median 12 months). The median classification accuracy in all 9 studies was 92% (range 70%–98%). The AUC ranged between 0.93 and 0.95 in 2 studies.[16,18]

For brain tumors, 4 studies used ML models to predict survival after gross-total resection, subtotal resection, biopsy, or stereotactic radiosurgery of metastasis,[25] glioblastoma multiforme,[3] or glioma.[2,24] The AUC ranged between 0.76 and 0.85, depending on the minimal follow-up time (median 21 months; range 6–48 months).[3,24,25] One study predicted recurrence and location of recurrence after gross-total resection for glioblastomas with an accuracy of 91% and AUC of 0.84.[23] Another study predicted recurrence in patients with metastasis after Gamma knife radiosurgery with an accuracy of 95% and AUC of 0.88.[26] Input features were MRI, clinical, intracranial EEG, and/or pathology.[2,3,23-26]

Five studies predicted outcome after discectomy and/or laminectomy in spine patients based on clinical factors, MRI, and intraoperative factors.[27-31] Hoffman et al.[28] predicted the Oswestry Disability Index version 2.0 (score between 0 and 1) in patients 6 months after operation for cervical spondylotic myelopathy with a coefficient of determination (r^2) of 0.93 and a mean absolute difference of 0.028.[28] Azimi et al.[27] predicted patient satisfaction 24 months after operation for lumbar spinal stenosis with an accuracy and AUC of 97% and 0.80, respectively. In 3 studies, Azimi et al. and Shamim et al. predicted symptom improvement, surgical success, and recurrence after surgery for lumbar disk hernia with accuracies ranging between 87% and 96% and AUCs ranging between 0.82 and 0.83.[29-31]

Two studies predicted Glasgow Outcome Score[34] and cerebral vasospasm score[35] after aneurysmal subarachnoid hemorrhage with an AUC of 0.85 and 0.96, respectively, using clinical and angiogram data. Glasgow Outcome Score also was predicted after burr hole surgery for chronic subdural hemorrhage (accuracy 73%, AUC 0.77).[37] Oermann et al.[5] predicted favorable (obliteration), neutral (no obliteration, no unfavorable outcome), and unfavorable outcome (postradiosurgery hemorrhage or permanently symptomatic radiation-induced changes) 2 until 8 years after stereotactic radiosurgery for AVMs based on clinical input and MRI, with an AUC of 0.70–0.71.[5] Asadi et al.[38] predicted complications and mortality after endovascular embolization of AVMs and achieved an accuracy of 98% for mortality, including type of complications as most predictive input features.

Two studies predicted motor improvement after deep brain stimulation for Parkinson disease.[32,33] Based on microelectrode recordings, Kostoglou et al. predicted improvement 24 months postoperatively with an accuracy of 95% and quantified this improvement on the Unified Parkinson's Disease Rating Scale (UPDRS) with a normalized means square error of 3.4%.[32,33] Based on clinical factors, Shamir et al.[33] predicted improvement 9 months postoperatively on the UPDRS in a trinary classification (nonresponsive <35% UPDRS improvement; moderate response = 35%–65% UPDRS improvement; high response >65% UPDRS improvement) with an accuracy of 86%.[33]

Azimi et al.[4] predicted successful endoscopic third ventriculostomy 6 months postoperatively with an accuracy and AUC of 95.1 and 0.87, respectively, based on clinical factors alone. Another study on patients with hydrocephalus predicted ventricular peritoneal shunt infections with an AUC of 0.83.[39] Shi et al.[36] predicted in-hospital mortality after decompressive craniectomy or hematoma evacuation in patients with traumatic brain injury with an accuracy and AUC of 95% and 0.89, respectively.

ML Compared with Other Predictive Methods

Nine studies compared the prediction performance of ML models to other modalities including prognostic indices,[4,5] classical statistical models,[4,5,27,28,35,36,38,39] and clinical experts.[2,3] First, 7 studies compared ML models with logistic regression models.[5,27,35-39] ML statistically significantly outperformed logistic regression in the prediction of successful endoscopic third ventriculotomy (accuracy 95% vs. 85%; AUC 0.87 vs. 0.80; both $P < 0.001$),[4] postoperative ventricular peritoneal shunt infection (AUC 0.83 vs. 0.56; $P < 0.001$),[39] mortality after embolization of AVMs (accuracy 98% vs. 43%; $P < 0.001$),[38] patient satisfaction after laminectomy for lumbar spinal stenosis (accuracy 97% vs. 82%; AUC 0.80 vs. 0.77; both $P < 0.001$),[27] in-hospital mortality in patients with traumatic brain injury (accuracy 95% vs. 88%; AUC 0.80 vs. 0.76; both $P < 0.001$),[36] cerebral vasospasm after aneurysmal subarachnoid hemorrhage (AUC 0.96 vs. 0.92, $P = 0.010$),[35] and outcome after burr-hole for a chronic subdural hematoma (accuracy 73% vs. 51%, P-value not reported; AUC 0.77 vs. 0.59; $P < 0.001$).[37] One study demonstrated that ML models were more accurate in predicting postoperative Oswestry Disability Index scores than linear regression but did not provide a P value (r^2 0.93 vs. 0.45; mean absolute deviation 0.03 vs. 0.09).[28]

In addition, 2 studies demonstrated that ML outperformed established prognostic indices.[4,5] Azimi et al. compared an artificial neural network against 2 prognostic indices for the prediction of successful endoscopic third ventriculostomy 6 months postoperatively.[4] Accuracy and AUC were statistically significantly greater for the artificial neural network compared with the Children's Hospital of Uganda endoscopic third ventriculostomy Success Score (accuracy 95% vs. 84%; AUC 0.87 vs. 0.78; all $P < 0.001$) and the endoscopic third ventriculostomy Success Score (accuracy 95% vs. 82%; AUC 0.87 vs. 0.76; $P < 0.001$). Oerman et al.[5] compared several ML algorithms against 3 prognostic indices for outcome

prediction after stereotactic radiosurgery for AVM. AUC was greater in all ML models compared with the Spetzler-Martin scale, radiosurgery-based AVM score, and Virginia radiosurgery AVM scale (AUC 0.70–0.71 vs. 0.57–0.69 depending on the specific index evaluated; no P values provided).

Lastly, 2 studies compared ML methods against clinical experts.[2,3] Emblem et al.[2] demonstrated that fuzzy C-means achieved a similar log-rank value and AUC in predicting survival compared to four neuroradiologists (log-rank value 14.4 vs. 10.7; AUC 0.89 vs. 0.88–0.91 dependent on the individual neuroradiologist; no P-values provided). Emblem et al.[2] showed in another study that support vector machine predicted survival in patients with a glioblastoma better compared with neuroradiologists (AUC 0.76–0.85 vs. 0.50–0.66 dependent on follow-up time; all $P < 0.01$).[3]

Summary of Results

Dependent on the specific prediction task and the type of input features included, ML models predicted neurosurgical outcome with a median accuracy of 94.5% (interquartile range [IQR] 87%–95%; range 63%–98%) and an AUC of 0.84 (IQR 0.82–0.88; range 0.71–0.96). In the 7 studies that compared ML models with classical logistic regression, ML models performed significantly better. The median absolute improvement in accuracy and AUC was 15% (IQR: 10%–22%; range 7%–55%) and 0.06 (IQR: 0.04–0.15; range 0.03–0.27).

DISCUSSION

ML models are being explored as tools for neurosurgical outcome prediction across a wide range of fields encompassing epilepsy, brain tumor, spine, neurovascular, Parkinson disease, traumatic brain injury, and patients with hydrocephalus. Moreover, some ML models have even been demonstrated to outperform prognostic indices and classical statistical models, performing similar or better than clinical experts under certain conditions.

Reviews on Predictive ML Models in Other Fields of Medicine

ML already has been well studied in clinical and preclinical medical research. Several reviews have evaluated the application of ML for prediction tasks. Sousa et al.[41] described the use of ML models for prediction of survival and/or rejection after transplantation surgery varying from stem-cell transplantation to heart transplantation, and 5 reviews focused on prognostic applications of ML in cancer research.[42-46] Cruz and Wishart[44] estimated that ML substantially improves the accuracy (10%–25% absolute improvement) of predicting cancer susceptibility, recurrence, and mortality.

ML models could also help in our understanding of cancer development and progression.[44] Abbod et al.[42] concluded that ML allows a more individualized prediction of disease behavior in patients with urological cancers compared with traditional regression statistics. This review addresses the lack of transparency of some ML algorithms but also the potential of other ML algorithms to reveal unused patterns and relations among the data. In 2 reviews, ML is explored for its potential on outcome prediction in patients with colon cancer[43] or patients who had undergone radiation therapy.[45] Previously, we have compared the performance of ML with clinical judgement and reviewed the potential applications of ML across the wide spectrum of neurosurgical care (J. T. Senders, M. M. Zaki, A. V. Karhade, et al., unpublished data, 2017).[9] Some included studies demonstrated astonishing results for neurosurgical outcome prediction due to the powerful prediction algorithms ML has to offer. The current review provides an in-depth analysis on ML used for neurosurgical outcome prediction.

Limitations

A few limitations of this review should be mentioned. First, the studies are very heterogeneous in nature. The median prediction accuracy and AUC represents, therefore, a summary of performance overall. The actual performance is still dependent on patient characteristics of the training set, specific ML algorithm used, input features used in the model, size training/test set, follow-up time, type of outcome measure (nominal, ordinal, or continuous), and the number of classes in a nominal/ordinal classification. Most studies (23/30) used a binary outcome measure, allowing evaluation by means of the AUC statistic. Continuous outcomes can be dichotomized into binary outcomes measures as well. For example, postoperative survival can be dichotomized into either less than or more than 12 months. By doing this, the algorithm might perform better according to the performance statistic used for evaluation; however, the set of predicted outcomes is less granular and might be less clinically useful compared with more granular outcomes (e.g., more than 2 ordinal classes or continuous outcomes). Due to this heterogeneity in methodology across all studies, we used a qualitative rather than a quantitative synthesis of the results. Second, publication and/or outcome bias could be present in this review. Studies and outcome measures demonstrating high-performing ML models might be published or reported more often. Finally, in this paper, we've summarized the accuracies and AUCs because these were the most frequently reported statistical measures and provide a general idea of the overall prediction performance. However, many statistical error metrics are often reported. Depending on the specific application and context, the right error metric should be chosen to evaluate the prediction performance. Despite these limitations, our study provides valuable insight into the applicability and performance of ML models in neurosurgical outcome prediction.

Implications and Future Challenges

By generating robust predictive algorithms, ML can guide health care providers, patients, and their families by enhancing the conversation regarding the odds of a successful outcome, occurrence of an adverse event, or quantify the benefit patients will have from surgery before the intervention. ML models including pre-,

intra-, and postoperative features can help predict long-term outcomes. Therapy could thus be more tailored to the individual patient as opposed to solely relying on population-based studies. ML also can provide a better understanding of relevant factors influencing outcomes in neurosurgical patients, and therefore help optimize that outcome. Furthermore, ML fits in a growing trend toward precision medicine. By combining clinical, genetic, pathologic, and radiologic data streams, clinicians could predict the individual course of the disease and estimate the effect of therapy. Although such models may supplement medical decision making, we do not foresee or endorse that algorithms replace human clinical decision-making. Rather, ML has great potential as complementary source of information, that can help guide the process of surgical and medical decision making.

Even though ML models show very promising results, many practical and ethical issues must be overcome to bridge the gap between research and clinical practice. First, these algorithms are sometimes referred to as "black box" techniques because the internal mechanisms are sometimes very difficult to interpret. Trusting medical decisions to these techniques can be uncomfortable to say the least. Furthermore, it is unclear who would be responsible if machines made a mistake. Errors made by computers also would be less tolerated, even if the error rate is lower compared with humans. Lastly, ML models are created and tested in their own environment; therefore, high performance in research settings does not guarantee high performance in clinical settings. Models are often trained in a sterile research setting in which complete and uncomplicated patient data are available. These models may not have the same performance when they are created or applied in settings in which patient data are lacking or coded differently. Furthermore, the use of patients for whom complete data are available could introduce selection bias, as these patients are not necessarily representative of the overall patient population.

Several solutions can help overcome these hurdles associated with the implementation of ML in clinical care. We deem open-source coding as the key element in validating promising ML models before implementation in clinical care, and we advocate that authors make their code accessible on established code repositories such as GitHub for others to view, learn from, and apply to their own data. In addition, literate programming packages, such as the R Notebook or Jupyter for Python, facilitate the sharing of code with explanatory text, equations, and visualizations to enhance the reproducibility of even the most complex data science projects. Software defects such as bugs or omitting best practices can then be detected and solved sooner. Open-source coding as a requirement for publication in scientific journals would further incentivize and promote this transparency.

ML models can have a major impact on clinical decision-making and effectively on patient outcomes and patient safety. Malfunctioning of these models can, therefore, have detrimental consequences. Before approval for clinical use, an external validation study should be done to the confirm their safety and effectiveness in the clinical realm. To assure the quality of validation studies and avoid publication bias, we suggest that their protocol should be pre-registered and approved before executing the study. Based on the results of external validation studies, ML models can be approved to be suitable for clinical use, comparable with the approval of pharmaceutical drugs by the Food and Drug Administration.

After clinical approval, ML models cannot be incorporated into the process clinical decision-making immediately but have to adapt to the environment in which they will be applied. This is similar to the speech recognition in smartphones that adapts to the user's voice after purchase. Models should operate parallel to clinical experts and be implemented once their performance and error margin are sufficient and acceptable, respectively. Because ML models are dynamic in nature, their performance is not a static fact either and can be influenced by changes in clinical practice or quality of the data. The performance of ML models should, therefore, be monitored continuously to ensure that deviation of performance can be detected and evaluated in a timely manner by clinicians as well as data scientists.

In the research setting, ML has been studied extensively demonstrating an excellent performance in outcome prediction for a wide range of neurosurgical conditions. However, it remains to be elucidated how ML can be implemented as a practical tool to improve clinical care in the hands of clinicians. Future studies should investigate how clinicians can benefit most from the powerful analyses ML can offer. To enhance the implementation of ML in clinical care, ethical and legal frameworks should be created that support collection of training data, validation of ML models on heterogeneous test sets before deployment, and regulation of the ML performance after deployment in clinical care.

CONCLUSIONS

ML models have great potential for improving neurosurgical outcome prediction. They can be a valuable aid for physicians, patients, and their families in the process of surgical and medical decision-making. Future studies should explore the hurdles associated with the creation, validation, and deployment of ML models in clinical care parallel to the development of these methods, as well as ethical and societal implications of their adoption.

REFERENCES

1. Moisi MD, Page J, Gahramanov S, Oskouian RJ. Bullet fragment of the lumbar spine: the decision is more important than the incision. *Global Spine J*. 2015;5:523-526.

2. Emblem KE, Nedregaard B, Hald JK, Nome T, Due-Tonnessen P, Bjornerud A. Automatic glioma characterization from dynamic susceptibility contrast imaging: brain tumor segmentation using knowledge-based fuzzy clustering. *J Magn Reson Imaging*. 2009;30:1-10.

3. Emblem KE, Pinho MC, Zöllner FG, Due-Tonnessen P, Hald JK, Schad LR, et al. A generic support vector machine model for preoperative glioma survival associations. *Radiology*. 2015;275:228-234.

4. Azimi P, Mohammadi HR. Predicting endoscopic third ventriculostomy success in childhood hydrocephalus: an artificial neural network analysis. *J Neurosurg Pediatr*. 2014;13:426-432.

5. Oermann EK, Rubinsteyn A, Ding D, Mascitelli J, Starke RM, Bederson JB, et al. Using a machine

learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations. *Sci Rep*. 2016;6:21161.

6. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216-1219.

7. Breiman L. StatisticalModeling: the two cultures. *Stat Sci*. 2001;16:199-231.

8. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: new opportunities, familiar challenges. *Clin Epidemiol*. 2017;9:245-250.

9. Senders JT, Arnaout O, Karhade AV, Dasenbrock HH, Gormley WB, Broekman ML, et al. Natural and artificial intelligence in neurosurgery: a systematic review [e-pub ahead of print]. *Neurosurgery*. 2017. https://doi.org/10.1093/neuros/nyx384.

10. Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S, Montazeri A. Artificial neural networks in neurosurgery. *J Neurol Neurosurg Psychiatry*. 2015;86:251-256.

11. Deo RC. Machine learning in medicine. *Circulation*. 2015;132:1920-1930.

12. Ghahramani Z. Probabilistic machine learning and artificial intelligence. *Nature*. 2015;521:452-459.

13. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015;349:255-260.

14. Antony AR, Alexopoulos AV, Gonzalez-Martinez JA, Mosher JC, Jehi L, Burgess C, et al. Functional connectivity estimated from intracranial EEG predicts surgical outcome in intractable temporal lobe epilepsy. *PLoS One*. 2013;8:e77916.

15. Arle JE, Perrine K, Devinsky O, Doyle WK. Neural network analysis of preoperative variables and outcome in epilepsy surgery. *J Neurosurg*. 1999;90:998-1004.

16. Armañanzas R, Alonso-Nanclares L, DeFelipe-Oroquieta J, Kastanauskaite A, de Sola RG, Defelipe J, et al. Machine learning approach for the outcome prediction of temporal lobe epilepsy surgery. *PLoS One*. 2013;8:e62819.

17. Bernhardt BC, Hong SJ, Bernasconi A, Bernasconi N. Magnetic resonance imaging pattern learning in temporal lobe epilepsy: classification and prognostics. *Ann Neurol*. 2015;77:436-446.

18. Feis DL, Schoene-Bake JC, Elger C, Wagner J, Tittgemeyer M, Weber B. Prediction of post-surgical seizure outcome in left mesial temporal lobe epilepsy. *Neuroimage Clin*. 2013;2:903-911.

19. Grigsby J, Kramer RE, Schneiders JL, Gates JR, Brewster Smith W. Predicting outcome of anterior temporal lobectomy using simulated neural networks. *Epilepsia*. 1998;39:61-66.

20. Memarian N, Kim S, Dewar S, Engel J, Staba RJ. Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. *Comput Biol Med*. 2015;64:67-78.

21. Munsell BC, Wee CY, Keller SS, Weber B, Elger C, da Silva LA, et al. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. *Neuroimage*. 2015;118:219-230.

22. Yankam Njiwa J, Gray KR, Costes N, Mauguiere F, Ryvlin P, Hammers A. Advanced [(18)F]FDG and [(11)C]flumazenil PET analysis for individual outcome prediction after temporal lobe epilepsy surgery for hippocampal sclerosis. *Neuroimage Clin*. 2015;7:122-131.

23. Akbari H, Macyszyn L, Da X, Bilello M, Wolf RL, Martinez-Lage M, et al. Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma. *Neurosurgery*. 2016;78:572-580.

24. Emblem KE, Due-Tonnessen P, Hald JK, Bjornerud A, Pinho MC, Scheie D, et al. Machine learning in preoperative glioma MRI: survival associations by perfusion-based support vector machine outperforms traditional MRI. *J Magn Reson Imaging*. 2014;40:47-54.

25. Knoll MA, Oermann EK, Yang AI, Paydar I, Steinberger J, Collins B, et al. Survival of patients with multiple intracranial metastases treated with stereotactic radiosurgery: does the number of tumors matter [e-pub ahead of print]? *Am J Clin Oncol*. https://doi.org/10.1097/COC.0000000000000299, accessed September 20, 2017.

26. Azimi P, Shahzadi S, Sadeghi S. Use of artificial neural networks to predict the probability of developing new cerebral metastases after radiosurgery alone [e-pub ahead of print]. *J Neurosurg Sci* accessed September 20, 2017.

27. Azimi P, Benzel EC, Shahzadi S, Azhari S, Mohammadi HR. Use of artificial neural networks to predict surgical satisfaction in patients with lumbar spinal canal stenosis: clinical article. *J Neurosurg Spine*. 2014;20:300-305.

28. Hoffman H, Lee SI, Garst JH, Lu DS, Li CH, Nagasawa DT, et al. Use of multivariate linear regression and support vector regression to predict functional outcome after surgery for cervical spondylotic myelopathy. *J Clin Neurosci*. 2015;22:1444-1449.

29. Shamim MS, Enam SA, Qidwai U. Fuzzy Logic in neurosurgery: predicting poor outcomes after lumbar disk surgery in 501 consecutive patients. *Surg Neurol*. 2009;72:565-572 [discussion: 572].

30. Azimi P, Benzel EC, Shahzadi S, Azhari S, Mohammadi HR. The prediction of successful surgery outcome in lumbar disc herniation based on artificial neural networks. *J Neurosurg Sci*. 2016;60:173-177.

31. Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S. Use of artificial neural networks to predict recurrent lumbar disk herniation. *J Spinal Disord Tech*. 2015;28:E161-E165.

32. Kostoglou K, Michmizos KP, Stathis P, Sakas D, Nikita KS, Mitsis GD. Classification and prediction of clinical improvement in deep brain stimulation from intraoperative microelectrode recordings. *IEEE Trans Biomed Eng*. 2017;64:1123-1130.

33. Shamir RR, Dolber T, Noecker AM, Walter BL, McIntyre CC. Machine learning approach to optimizing combined stimulation and medication therapies for Parkinson's disease. *Brain Stimul*. 2015;8:1025-1032.

34. Lo BW, Macdonald RL, Baker A, Levine MA. Clinical outcome prediction in aneurysmal subarachnoid hemorrhage using Bayesian neural networks with fuzzy logic inferences. *Comput Math Methods Med*. 2013;2013:904860.

35. Dumont TM, Rughani AI, Tranmer BI. Prediction of symptomatic cerebral vasospasm after aneurysmal subarachnoid hemorrhage with an artificial neural network: feasibility and comparison with logistic regression models. *World Neurosurg*. 2011;75:57-63 [discussion: 25-58].

36. Shi HY, Hwang SL, Lee KT, Lin CL. In-hospital mortality after traumatic brain injury surgery: a nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. *J Neurosurg*. 2013;118:746-752.

37. Abouzari M, Rashidi A, Zandi-Toghani M, Behzadi M, Asadollahi M. Chronic subdural hematoma outcome prediction using logistic regression and an artificial neural network. *Neurosurg Rev*. 2009;32:479-484.

38. Asadi H, Kok HK, Looby S, Brennan P, O'Hare A, Thornton J. Outcomes and complications after endovascular treatment of brain arteriovenous malformations: a prognostication attempt using artificial intelligence. *World Neurosurg*. 2016;96:562-569.e561.

39. Habibi Z, Ertiaei A, Nikdad MS, Mirmohseni AS, Afarideh M, Heidari V, et al. Predicting ventriculoperitoneal shunt infection in children with hydrocephalus using artificial neural network. *Child's Nerv Syst*. 2016;32:2143-2151.

40. Akbari H, Macyszyn L, Da X, Wolf RL, Bilello M, Verma R, et al. Pattern analysis of dynamic susceptibility contrast-enhanced MR imaging demonstrates peritumoral tissue heterogeneity. *Radiology*. 2014;273:502-510.

41. Sousa FS, Hummel AD, Maciel RF, Cohrs FM, Falcão AE, Teixeira F, et al. Application of the intelligent techniques in transplantation

databases: a review of articles published in 2009 and 2010. *Transplant Proc*. 2011;43:1340-1342.

42. Abbod MF, Catto JW, Linkens DA, Hamdy FC. Application of artificial intelligence to the management of urological cancer. *J Urol*. 2007;178:1150-1156.

43. Ahmed FE. Artificial neural networks for diagnosis and survival prediction in colon cancer. *Mol Cancer*. 2005;4:29.

44. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform*. 2007;2:59-77.

45. Kang J, Schwartz R, Flickinger J, Beriwal S. Machine learning approaches for predicting radiation therapy outcomes: a clinician's perspective. *Int J Radiat Oncol Biol Phys*. 2015;93:1127-1135.

46. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8-17.

Conflict of interest statement: The authors declare that the article content was composed in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received 16 June 2017; accepted 21 September 2017

Citation: World Neurosurg. (2018) 109:476-486.
https://doi.org/10.1016/j.wneu.2017.09.149

*Journal homepage: **www.WORLDNEUROSURGERY.org***

*Available online: **www.sciencedirect.com***

SUPPLEMENTARY DATA

Supplementary Table 1. Search Syntaxes for the PubMed and Embase Databases

PubMed search: 1—1-2017	((((("Machine Learning"[Mesh] OR "Artificial Intelligence"[Mesh] OR "Natural Language Processing"[Mesh] OR "Neural Networks (Computer)"[Mesh] OR "Support Vector Machine"[Mesh] OR Machine learning[Title/Abstract] OR Artificial Intelligence[Title/Abstract] OR Naive Bayes[Title/Abstract] OR bayesian learning[Title/Abstract] OR Neural network[Title/Abstract] OR Neural networks[Title/Abstract] OR Natural language processing[Title/Abstract] OR support vector*[Title/Abstract] OR random forest*[Title/Abstract] OR boosting[Title/Abstract] OR deep learning[Title/Abstract] OR machine intelligence[Title/Abstract] OR computational intelligence[Title/Abstract] OR computer reasoning[Title/Abstract])) AND ((("Neurosurgery"[Mesh] OR "Neurosurgical Procedures"[Mesh] OR "Central Nervous System Neoplasms"[Mesh] OR "Intervertebral Disc Displacement"[Mesh] OR "Spinal Stenosis"[Mesh] OR "Nerve Sheath Neoplasms"[Mesh] OR "Skull Base Neoplasms"[Mesh] OR "Neoplasms, Neuroepithelial"[Mesh] OR "Craniocerebral Trauma"[Mesh] OR "Intracranial Hemorrhages"[Mesh] OR "Deep Brain Stimulation"[Mesh] OR "Hydrocephalus"[Mesh] OR "Trigeminal Neuralgia"[Mesh] OR "Diskectomy"[Mesh] OR "Epilepsy"[Mesh] OR "Intracranial Pressure"[Mesh] OR neurosurgery[Title/Abstract] OR neurosurgeries[Title/Abstract] OR neurosurgical[Title/Abstract] OR neurosurgically[Title/Abstract] OR radiosurgery[Title/Abstract] OR Brain tumor[Title/Abstract] OR Brain tumour[Title/Abstract] OR glioma[Title/Abstract] OR glioblastoma[Title/Abstract] OR brain metastases[Title/Abstract] OR Brain metastasis[Title/Abstract] OR Epilepsy[Title/Abstract] OR Deep brain stimulation[Title/Abstract] OR traumatic brain injury[Title/Abstract] OR head injury[Title/Abstract])) OR (("Brain"[Mesh] OR "Spine"[Mesh] OR brain[Title/Abstract] OR spine[Title/Abstract] OR spinal [Title/Abstract] OR lumbar[Title/Abstract] OR cerebr*[Title/Abstract] OR cerebell*[Title/Abstract]) AND ("Surgical Procedures, Operative"[Mesh] OR "Postoperative Complications"[Mesh] OR "surgery" [Subheading] OR "Postoperative Period"[Mesh] OR "Perioperative Period"[Mesh] OR "Preoperative Period"[Mesh] OR surgery[Title/Abstract] OR surgeries[Title/Abstract] OR surgical[Title/Abstract] OR postoperative*[Title/Abstract] OR post operative*[Title/Abstract] OR preoperative*[Title/Abstract] OR pre operative*[Title/Abstract] OR perioperative*[Title/Abstract] OR peri operative*[Title/Abstract] OR operative procedure*[Title/Abstract])))) NOT (Comment[Publication Type] OR editorial[Publication Type] OR letter[Publication Type] OR case reports[Publication Type])
Embase search: 1—1-2017	'neurosurgery'/exp OR 'brain tumor'/exp OR 'brain cancer'/exp OR 'brain metastasis'/exp OR 'posterior cranial fossa tumor'/exp OR 'brain stem tumor'/exp OR 'intervertebral disk hernia'/exp OR 'vertebral canal stenosis'/exp OR 'nerve sheath tumor'/exp OR 'head injury'/exp OR 'brain depth stimulation'/exp OR 'hydrocephalus'/exp OR 'epilepsy'/exp OR 'trigeminus neuralgia'/exp OR 'intervertebral diskectomy'/exp OR 'intracranial pressure'/exp OR 'radiosurgery'/exp OR neurosurgery:ab,ti OR neurosurgeries:ab,ti OR neurosurgical:ab,ti OR neurosurgically:ab,ti OR radiosurgery:ab,ti OR 'brain tumor':ab,ti OR 'brain tumour':ab,ti OR glioma:ab,ti OR glioblastoma:ab,ti OR 'brain metastases':ab,ti OR 'brain metastasis':ab,ti OR epilepsy:ab,ti OR 'deep brain stimulation':ab,ti OR 'traumatic brain injury':ab,ti OR 'head injury':ab,ti OR ('brain'/exp OR 'spine'/exp OR brain:ab,ti OR spine:ab,ti OR spinal:ab,ti OR lumbar:ab,ti OR cerebr*:ab,ti OR cerebell*:ab,ti AND ('surgical technique'/exp OR 'postoperative complication'/exp OR 'surgery':lnk OR 'postoperative period'/exp OR 'perioperative period'/exp OR 'preoperative period'/exp OR surgery:ab,ti OR surgeries:ab,ti OR surgical:ab,ti OR postoperative*:ab,ti OR 'post-operative*':ab,ti OR preoperative*:ab,ti OR 'pre-operative*':ab,ti OR perioperative*:ab,ti OR 'peri-operative*':ab,ti OR 'operative procedure*':ab,ti)) AND ('machine learning'/exp OR 'artificial intelligence'/exp OR 'natural language processing'/exp OR 'artificial neural network'/exp OR 'support vector machine'/exp OR 'bayesian learning'/exp OR 'random forest'/exp OR 'machine learning':ab,ti OR 'artificial intelligence':ab,ti OR 'neural network':ab,ti OR 'neural networks':ab,ti OR 'natural language processing':ab,ti OR 'support vector*':ab,ti OR boosting:ab,ti OR 'deep learning':ab,ti OR 'random forest*':ab,ti OR 'naive bayes:ab,ti' OR 'bayesian learning':ab,ti OR 'machine intelligence':ab,ti OR 'computational intelligence':ab,ti OR 'computer reasoning':ab,ti) NOT (comment:it OR editorial:it OR letter:it OR 'case reports':it)