
Emrah CELTIKCI
University of Pittsburgh Medical Center, Department of Neurological Surgery, Pittsburgh, PA, USA

ABSTRACT

Current practice of neurosurgery depends on clinical practice guidelines and evidence-based research publications that derive results using statistical methods. However, statistical analysis methods have some limitations such as the inability to analyze non-linear variables, requiring setting a level of significance, being impractical for analyzing large amounts of data and the possibility of human bias. Machine learning is an emerging method for analyzing massive amounts of complex data which relies on algorithms that allow computers to learn and make accurate predictions. During the past decade, machine learning has been increasingly implemented in medical research as well as neurosurgical publications. This systematic review aimed to assemble the current neurosurgical literature that machine learning has been utilized, and to inform neurosurgeons on this novel method of data analysis.

KEYWORDS: Bayesian network, Logistic regression analysis, Machine learning, Neural network, Neurosurgery, Support vector machine

INTRODUCTION

Current decision-making process in neurosurgery is based on clinical practice guidelines that cite wide sample clinical trials and case series publications. Common statistical methods are utilized in the majority of these publications such as: t-test, cox-hazard model and chi-square test which are also well-known to the neurosurgeons. Statistical tests calculate the correlation between variables and require assumptions as in determining level of significance, e.g. setting a p-value. Also, the presence of a correlation between variables does not always prove a causation (2). Most of the statistical tests have some kind limitations such as being unable to analyze non-linear variables or difficulty in finding correlations when data sets are massive.

In the past decades, machine learning (ML) has been proposed as a novel computational data analysis method in order to overcome the limitations of traditional scientific statistical methods and has become increasingly popular amongst medical sciences. ML is a subfield of computer science, and focuses on creating algorithms that allow computers to learn from data. Its ability to perform comprehensive analyses even with massive amounts of non-linear data makes it favorable in medical decision-making. There are different types of learning and different mathematical algorithms applied in ML.

Principally, there are two types of learning: supervised and unsupervised learning. In supervised learning, the aim is to predict the output. In order to achieve the capability of prediction, ML requires a training period, when all inputs and their readily available outputs must be registered. Following the training, the algorithm should be tested to observe the quality of predictions. In this step, the researcher inspects if the predictions of the computer are as accurate as known outputs. If the algorithm performs with a good percentage of correct predictions, it can be employed for making predictions where outputs are not known. For example, in order to build an ML algorithm which will perform predictions on lumbar spine surgery, pre-surgical clinical data and post-surgical outcome data of patients must be registered as the first step. Then, the researcher must test the algorithm if predictions are in line...
with real patient outcomes. Provided that the predictions are satisfactory, the algorithm can be utilized to predict unknown outcomes. In unsupervised learning, the aim is to predict certain patterns in the data rather than an outcome. Pattern analysis is implemented especially for defining pathophysiological mechanisms of diseases. For example; defining the relation patterns between variables such as imaging characteristics, molecular markers, performance score and amount of cerebral edema in glioma patients requires an unsupervised learning process where predicting an outcome is not the ultimate goal.

As mentioned earlier, different mathematical algorithms have been defined for ML. In logistic regression, inputs and outputs come additively and linearly. However, once established, this method does not allow further addition of alternative variables into the analysis. Decision tree allows performing predictions from different variables even if they are not occurring together. For example, a patient with a motor deficit can have a cerebral mass or subdural hematoma or spinal burst fracture or a herniated disc or a non-neurosurgical condition. Decision tree analysis reveals any possible correlation between these clinical entities. Neural networks have the flexibility of changing input features and adding new types of inputs during data collection. Bayesian networks can predict conditional dependencies of random variables. When symptoms are provided, this algorithm can calculate the probability of various diseases in a subject. Linear discriminant analysis aims to find a linear combination of variables that characterizes or classifies two or more types of objects or events. For example, by using inputs as magnetic resonance imaging (MRI) scans, this algorithm can perform automatized diagnosis of brain tumors. Support vector machines are supervised models that require a set of training data, and have the flexibility of using inputs either the data is linear or non-linear. Support vector machines recently demonstrated the capability of analyzing massive amount of diverse data, e.g., predicting prognosis of stroke patients by analyzing diffusion MRI scans (14).

Recently, ML related studies are progressively increasing in the literature due to its advanced data analyzing capabilities. The aim of this systematical review is to compile the current neurosurgical literature that utilized different ML techniques, and to provide neurosurgeons with an up-to-date overview of this emerging data analysis method.

■ MATERIAL and METHODS

‘Guidelines of Preferred Reporting Items for Systematic Reviews and Meta-analyses: the PRISMA Statement’ was followed while performing this systematic review (37). MEDLINE, and the Cochrane Database of Systematic Reviews was queried using combinations of the following keywords: neurosurgery, machine learning, glioma, spine, skull base, prediction, support vector machine, Bayesian network, decision tree, data mining, neural network. Articles in English language, published from 1 January 1900 to 30 January 2017 were evaluated.

Studies were screened by title and abstract in order to identify relevant articles. Inclusion criteria were; studies that utilized ML 1) in preoperative planning, 2) in predicting outcomes of either interventions or diseases. Flowchart of article selection is illustrated in Figure 1. Studies 1) that are not related to ML algorithms or neurological sciences, 2) concerning psychiatry, radiology, neurology or specialties other than neurosurgery, 3) on human computer interface, 4) written in a language other than English, 5) with no available full-text were excluded. Also, animal studies, reviews, conference papers, poster presentations, editorials, letters to editor were not included. Full-texts of all included articles were retrieved for further analysis. References of all included full-text articles were reviewed to find any related manuscripts.

■ RESULTS

Literature Search

Following elimination of duplicates, a total of 9098 citations were identified. Following exclusion, 37 full-text articles were examined, and consequently 7 articles were excluded due to unrelated study aim. Five of the excluded studies were focused on automatized tumor detection or anatomical structure segmentation, whereas 2 studies were about automatized prediction of pathological or genetic characteristics without considering outcome or survival. After checking references and related articles of full-text articles, 21 additional articles were found eligible and included into systematic review. Included studies were classified according to topics; hydrocephalus, deep brain stimulation, neurovascular, epilepsy, glioma, radiosurgery, spine, traumatic brain injury. Remaining topics were summarized under ‘other’. Distributions of studies by topics were summarized in Table I.

Algorithms

Amongst 51 studies, following algorithms were identified; neural network (n=17), Bayesian network (n=11), support vector machine (n=16), decision tree (n=5), logistic regression (n=12) and discriminant analysis (n=2). Six studies had employed more than one ML algorithm. Most common algorithms were neural network (27%) and support vector machine (25%).

Hydrocephalus

There were 2 studies on hydrocephalus. Azimi and Mohammadi used neural network analysis for evaluating endoscopic third ventriculostomy (ETV) success in childhood hydrocephalus (6). Habibi et al. also used neural network analysis to predict the risk of ventriculoperitoneal shunt infection in children with hydrocephalus (23).

Deep Brain Stimulation

There were 6 studies that utilized ML algorithms in deep brain stimulation studies. The majority of these studies focused on automatized localization of the subthalamic nucleus (STN). Two studies used Bayesian network algorithms for STN targeting (38,52). In the study of Rajpurohit et al., logistic regression was employed to define STN (45), and Wong et al. applied an unsupervised ML technique in order to localize STN (53). Baumgarten et al. used neural network analysis to predict any pyramidal tract side effects during deep brain stimulation (8).
In an investigation by Muniz et al., three different techniques of ML; neural network analysis, support vector machine and logistic regression analysis were employed in order to evaluate the effect of STN stimulation on ground reaction force during gait (39).

**Neurovascular**

There were 6 studies that applied ML techniques in analysis of neurovascular pathologies. The study of Asadi et al. investigated the outcome of arteriovenous malformations (AVMs) following endovascular treatment (4). In the separate studies of Dumont (16) and Dumont et al. (17), the authors tried the same algorithm, which is a neural network model, on different patient populations to predict cerebral vasospasm. In the study of Lo et al. authors combined two ML methods; neural networks with Bayesian network algorithms, to predict clinical outcome of patients following aneurysmal subarachnoid hemorrhage (SAH) (31). One other study employed decision tree analysis to reveal factors that increase the risk of developing aggressive behavior in dural arteriovenous fistulas. They reported the presence of cortical venous drainage as the main risk factor (48).

**Epilepsy**

In the literature, there were numerous studies that utilized ML methods in the field of epilepsy. Also, there were related studies that applied ML techniques to analyze data achieved from electroencephalography recordings, intraoperative surface electrode recordings or intracortical electrode recordings, in order to define neural activity signal patterns of language, seizure and face recognition. Six publications were identified as neurosurgical. Three of them were epilepsy surgery studies.

---

**Figure 1:** Flow-chart showing selection of relevant articles.
Radiosurgery for brain metastases (10,29,43). Oermann et al. used ML to predict patient outcomes of AVM radiosurgery (43). Another study implemented ML methods to predict the risk of hydrocephalus following gamma knife radiosurgery for intracranial schwannoma (30).

**Spine**

There were 7 studies identified, with 6 of them investigated prediction of spinal surgical outcomes via ML (5,7,22,24,34,35). One study applied support vector machine to predict the risk of progression in adolescent idiopathic scoliosis patients (1).

**Traumatic Brain Injury**

Publications on traumatic brain injury that used ML techniques had the greatest number of publications (n=11) with wider sample sizes compared to other topics in this review. These studies employed various ML algorithms such as neural networks, logistic regression model, support vector machine

### Table 1: Distribution of Studies Included in Systematic Review According to Topics

<table>
<thead>
<tr>
<th>Topic</th>
<th>Aim of study</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrocephalus</td>
<td>Surgical outcome prediction</td>
<td>2</td>
</tr>
<tr>
<td>Deep brain stimulation</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Neurovascular</td>
<td>STN localization</td>
<td>4</td>
</tr>
<tr>
<td>Endovascular treatment outcome prediction</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Vasospasm following SAH prediction</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Clinical outcome prediction</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Prediction of developing aggressive DAVF behavior</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Epilepsy</td>
<td>Surgery candidate prediction</td>
<td>1</td>
</tr>
<tr>
<td>Identification of epileptogenic zone for surgery</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pre-operative identification of language dominance</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Surgical outcome prediction</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Glioma</td>
<td>Survival prediction</td>
<td>4</td>
</tr>
<tr>
<td>Pre-operative prediction of intraoperative MRI use</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Survival prediction of intracranial metastatic disease</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Radiosurgery</td>
<td>Prediction of hydrocephalus development following radiosurgery for schwannoma</td>
<td>1</td>
</tr>
<tr>
<td>Outcome prediction of AVM radiosurgery</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Spine</td>
<td>Predicting risk of progression adolescent idiopathic scoliosis</td>
<td>1</td>
</tr>
<tr>
<td>Surgical outcome prediction</td>
<td>6*</td>
<td></td>
</tr>
<tr>
<td>Traumatic Brain Injury</td>
<td>Outcome prediction</td>
<td>10</td>
</tr>
<tr>
<td>Other</td>
<td>Outcome prediction after Chiari malformation surgery</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ICP prediction from intracranial pressure signal morphology</td>
<td>4</td>
</tr>
</tbody>
</table>

*Two studies out of 6 were about cervical spine, whereas 4 studies were on lumbar spine.


outcome prediction studies (3,9,36). Cohen et al. tested if ML method can identify surgery candidates among intractable temporal lobe epilepsy patients as well as physicians (11). The remaining studies investigated pre-operative identification of the epileptic focus (15), and language dominance (21).

### Glioma

Five studies implemented ML techniques, and 4 of them investigated factors affecting survival in glioma patients (19,32,33,56). One study employed an ML algorithm in order to identify in which patients intraoperative MRI should be used (50). The ML algorithm was given a combined data of surgeon’s preoperative opinions and patient’s clinical characteristics to make the decision.

### Radiosurgery

Five studies were identified with 3 of them using ML algorithms to predict survival in patients that underwent stereotactic radiosurgery for brain metastases (10,29,43). Oermann et al. used ML to predict patient outcomes of AVM radiosurgery (43). Another study implemented ML methods to predict the risk of hydrocephalus following gamma knife radiosurgery for intracranial schwannoma (30).
and linear discriminant function. All studies, without exception, were designed to investigate outcome prediction (18,20,25,40,41,44,47,51,55,57,58).

Other

Other examples of publications that implemented ML methods for neurosurgical practice were; predicting intracranial pressure (ICP) increase in intensive care unit patients from ICP monitor waveform morphology (26,27,42,49) and predicting the resolution of syringomyelia following Chiari malformation type 1 decompression (13).

DISCUSSION

To the best of our knowledge, the current review is the most comprehensive systematic review on neurosurgical literature that employed various ML methods for data analysis. The author believes that this systematic review would provide up-to-date information on neurosurgical publications that utilized ML methods while informing neurosurgeons on the subject briefly.

In the past decade, ML was proposed as an alternative method of data analysis, in order to overcome the limitations of linear statistical methods while analyzing and interpreting massive amounts of data. Although ML is not artificial intelligence (AI), it is derived from AI studies which automate analytical building. ML gives computers the ability to learn patterns even when they are non-linear (3). In other fields, ML is employed for tasks like optical character recognition, spam filtering, face recognition, and in search engines. The use of ML in biological sciences has increased over time due to the limitations of classical statistical methods. Also, statistical methods may cause bias in scientific studies because they require excessive human intervention like choosing the statistical test or assuming the level of significance.

Researchers tend to limit the variables in a study in order to simplify statistical analysis. In contrast, ML techniques can provide high-value predictions without human intervention that may guide better decisions in real time. Furthermore, ML techniques provide convenience for setting many clinical features, as inputs, and have the ability to analyze the data as a whole. Hoffman et al. compared multivariate linear regression and support vector machine in predicting functional outcome after surgery for cervical spondylotic myelopathy and demonstrated that the ML method was superior to traditional statistical analysis (24). The algorithm used in the study of Oermann et al. was able to provide best predictions of AVM radiosurgery to date (43). Another study by Cohen et al. demonstrated that ML could select epilepsy surgery candidates as good as physicians, with faster evaluation and decision-making (11).

Despite the number of neurosurgical studies using ML are still limited, the increasing trend in this field promises more research publications implementing this technique into decision-making process. This means that some of the clinical guidelines and indications could change in the near future. The study of Knoll et al. provides a clue as to how ML methods could modify our decision-making process (29). The authors retrospectively analyzed 507 patients with intracranial metastases that were treated with radiosurgery in order to determine predictive factors of overall survival. Previously, multiple studies had suggested that the number of metastases (>4) is predictive for overall survival (1,9,24,25,35). However, the ML technique they employed revealed that performance status and systemic disease status of patients were predictive for overall survival, not the number of intracranial metastases (29). Asadi et al. retrospectively analyzed records of 199 AVM patients and utilized a supervised ML method (Levenberg-Marquardt algorithm). They reported that ML predicted patient outcome with an accuracy of 97.5% and identified the presence or absence of nidal fistulae as the most important factor (4).

This systematic review only included studies on prediction of patient outcome and survival or pre-operative planning. Also, there are other neurosurgery related studies that implemented ML and were not included in this review. Radiogenomics studies investigate automatized MRI-based diagnosis of gliomas by establishing the correlation between genetic properties and imaging characteristics of the tumor (28,46,54). Human-computer interface studies like the publication of Collinger et al. is another example, which demonstrated the possibility of a neuroprosthetic arm controlled by intracortical microelectrode implants in the motor cortex (12). In their study, ML algorithms were not used for predicting any outcome, but employed in order to analyze the complex data acquired from motor cortex recordings.

CONCLUSION

ML provides accurate and fast interpretation of complex data in large amounts, overcoming possible human error and/or bias. This progressively developing method has the ability to learn and gain experience, and would continue to increase its success in accurate decision-making in the future. Wider application of ML in the medical and neurosurgical studies might improve clinical and surgical management, ultimately the quality of patient care and outcome.

ACKNOWLEDGEMENTS

I gratefully acknowledge the contributions of Pinar Celtikci, M.D. in critical reviewing, proof-reading and editing of the manuscript. I also thank David T. Fernandes-Cabral M.D. for editing the manuscript and Vijayakrishna Rowthu Ph.D. for providing general advice on the field.

REFERENCES


18. Eftekhar B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E: Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Med Inform Decis Mak 5: 3, 2005


