


Artificial Learning and Machine Learning Applications in Spine Surgery: A Systematic Review

Global Spine Journal
2022, Vol. 12(7) 1561–1572
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/21925682211049164
journals.sagepub.com/home/gsj



Cesar D. Lopez, BS¹, Venkat Boddapati, MD¹ , Joseph M. Lombardi, MD¹, Nathan J. Lee, MD¹, Justin Mathew, MD¹ , Nicholas C. Danford, MD, MA¹ , Rajiv R. Iyer, MD¹, Marc D. Dyrszka, MD¹, Zeeshan M. Sardar, MD¹, Lawrence G. Lenke, MD¹, and Ronald A. Lehman, MD¹

Abstract

Objectives: This current systematic review sought to identify and evaluate all current research-based spine surgery applications of AI/ML in optimizing preoperative patient selection, as well as predicting and managing postoperative outcomes and complications.

Methods: A comprehensive search of publications was conducted through the EMBASE, Medline, and PubMed databases using relevant keywords to maximize the sensitivity of the search. No limits were placed on level of evidence or timing of the study. Findings were reported according to the PRISMA guidelines

Results: After application of inclusion and exclusion criteria, 41 studies were included in this review. Bayesian networks had the highest average AUC (.80), and neural networks had the best accuracy (83.0%), sensitivity (81.5%), and specificity (71.8%). Preoperative planning/cost prediction models (.89,82.2%) and discharge/length of stay models (.80,78.0%) each reported significantly higher average AUC and accuracy compared to readmissions/reoperation prediction models (.67,70.2%) ($P < .001$, $P = .005$, respectively). Model performance also significantly varied across postoperative management applications for average AUC and accuracy values ($P < .001$, $P < .027$, respectively).

Conclusions: Generally, authors of the reviewed studies concluded that AI/ML offers a potentially beneficial tool for providers to optimize patient care and improve cost-efficiency. More specifically, AI/ML models performed best, on average, when optimizing preoperative patient selection and planning and predicting costs, hospital discharge, and length of stay. However, models were not as accurate in predicting postoperative complications, adverse events, and readmissions and reoperations. An understanding of AI/ML-based applications is becoming increasingly important, particularly in spine surgery, as the volume of reported literature, technology accessibility, and clinical applications continue to rapidly expand.

Keywords

machine learning, artificial intelligence, deep learning, predictive modeling, spine surgery, orthopedic surgery

Introduction

Machine learning (ML) is increasingly reported on in health care, including orthopedics, especially for its applications in predictive analytics. ML is a form of artificial intelligence (AI) that employs the use of algorithms and mathematical models that can learn from data, identify patterns and complex

¹Department of Orthopaedic Surgery, The Spine Hospital, New York-Presbyterian/Columbia University Irving Medical Center, New York, NY, USA

Corresponding Author:

Venkat Boddapati, MD, Columbia University Irving Medical Center, 622 W. 168th St., PH-11, New York, NY 10032, USA.
Email: venkatboddapati1@gmail.com



Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

relationships, and make automated decisions—oftentimes with minimal human intervention.^{1,2} These algorithms are able to find patterns in the data and apply those patterns to new challenges in the future. Algorithms include artificial neural networks (ANN), decision trees (DT), boosting/ensemble learning models (BEL), Bayesian networks (BN), logistic regression (LR), and support vector machines (SVM). Neural networks are modeled on neurons in the brain, and they use artificial intelligence to untangle and break down extremely complex relationships. Across various medical specialties, AI/ML has been shown to be beneficial in guiding clinical decision-making, and artificial neural networks used as outcome prediction models have been applied in diagnosing various medical conditions.¹⁻⁴

Within orthopedics, demonstrated applications of AI/ML include surgical risk stratification and optimization,⁵ clinical outcome prediction and diagnostics,⁶ cost-efficiency analyses, and in total joint arthroplasty literature it has been used for proposed risk-adjusted insurance reimbursement models.⁷ Spine surgery, in particular, is a field that involves high-risk procedures and is continually seeking to improve surgical planning, outcomes, and to reduce complications. With its powerful predictive capabilities, AI/ML has the potential to be used in new and innovative applications that may improve the safety of spine surgery and improve outcomes.

The use of AI/ML is rapidly expanding in health care and has the potential to improve surgical care and reduce costs, especially for high-cost and complex spine surgery procedures. As such, it is important for spine surgeons to better understand the current applications of AI/ML, especially in light of the burgeoning literature regarding this topic in recent years. The purpose of this review is to identify and evaluate all current research-based spine surgery applications of AI/ML, namely, in optimizing preoperative patient selection, as well as predicting and managing postoperative outcomes and complications.

Materials and Methods

Search Strategy

A comprehensive search of publications, up to February 2020, was conducted using the EMBASE, Medline, and PubMed databases in accordance with PRISMA guidelines. Sample search query keywords and MeSH terms are provided in [Supplementary Table 1](#). Screening of reference lists of retrieved articles also yielded additional studies.

Eligibility Criteria

Inclusion criteria consisted of original clinical studies, including studies which evaluate spine surgery applications of AI/ML in guiding clinical decision-making. Exclusion criteria consisted of studies that did not evaluate spine surgery applications of AI/ML, studies involving oncologic spine

surgery or infectious etiologies, studies involving applications for design and development of hardware or implants, medical imaging analysis studies without explicit reference or application to spine surgery, studies with non-human subjects, non-English-language studies, inaccessible articles, conference abstracts, reviews, and editorials. No limits were placed on level of evidence or timing of the study since the majority of the reviewed studies were published within the last 10 years.

Study Selection

Article titles and abstracts were screened initially by two reviewers, and full-text articles were subsequently screened based on the selection criteria. The studies were rated by their level of evidence, based on the Oxford Centre for Evidence-based Medicine Levels of Evidence.⁸ Two authors reviewed each individual article that was included. Discrepancies in inclusion studies were discussed and resolved by consensus.

Data Extraction and Categorization

A database was generated from all included studies which consisted of the journal of publication, publication year, country of origin, study design, level of evidence, study duration, blinding of the study, number of involved institutions, AI/ML methods and clinical applications, surgical domain, data sources, input variables and output variables, sample size, average patient age, percent female patients, and any additional pertinent findings from the study. The reviewed articles were sorted into different, non-mutually exclusive categories based on AI/ML clinical application. AI/ML clinical applications were divided into two major groups: (1) administrative and clinical decision support and (2) postoperative prediction and management of complications and outcomes. The former group contained the following prediction and optimization sub-categories: preoperative planning and cost prediction, hospital discharge and length of stay (LOS), readmissions, and reoperations. The other group included postoperative cardiovascular complications, other complications, mortality, and functional and clinical outcomes.

Data Analysis

Descriptive statistics were employed to summarize important findings and results from the selected articles and to describe trends in AI/ML techniques, clinical applications, and relevant findings associated with its use. Summary data were presented using simple means, frequencies, standard deviations (for normally distributed data on a decimal scale), and proportions. AI/ML model performance within the reviewed studies were summarized using various metrics, including the area under the curve (AUC) of receiver operating

characteristic (ROC) curves, accuracy (%), sensitivity (%), and specificity (%). AUC is a measure of a ML model's discriminative ability (i.e., accurately predicting true positives and negatives while identifying false positive or negative cases).^{9,10} AUC values range from .50 to 1 and measure a prediction models' discriminative ability, with a higher AUC value signifying better predictive ability of the model correctly placing a patient into an outcome category. A model with an AUC of 1.0 is a perfect discriminator, .90 to .99 is considered excellent, .80 to .89 is good, .70 to .79 is fair, and .51 to .69 is considered poor.¹¹ AUC measures a model's discriminative ability in accurately selecting true positives and negatives, while minimizing false positives and false negatives. Accuracy is simply a measure of a model's ability to correctly predict true positives and true negatives, without accounting for identifying false positives/negatives. Reported model performance metrics for each AI/ML algorithm type and for each clinical application category were aggregated across the reviewed studies. A formal bias assessment for each study was performed based on the Cochrane Handbook for Systematic Reviews methodology (Supplementary Table 1).¹² One-way ANOVA with post hoc Tukey tests were performed, with statistical significance set at $P < .05$. All statistical analysis was performed using Stata (version 16.1, Stata Corporation—College Station, Texas, USA).

Results

Search Results and Study Selection

Using our pre-defined search terms resulted in 335 articles, of which 67 duplicate articles were removed. The remaining 268 articles were screened by title and abstract according to inclusion and exclusion criteria. Ultimately, there were 44 articles included for full review, of which 41 met full inclusion and exclusion criteria. (Figure 1) Over 83% of studies had level of evidence III, and the median number of patients in each study was 964 (mean 2784, standard deviation [SD] 3122). Although there were no limitations on publication dates in the selection process, the majority of studies (77.5%) were published during the last 2 years (2018–2020) (Figure 2) AUC was the most frequently reported performance metric, appearing in 37 out of the 41 total reviewed studies (90.2%). In comparison, accuracy was reported less frequently (16 studies, 39.0%), as were sensitivity and specificity (11 studies, or 26.8%).

Administrative and Clinical Decision Support Applications

A total of 18 reviewed studies (43.9%) evaluated the use of AI/ML applications in optimizing preoperative patient selection or projecting surgical costs, through prediction of hospital

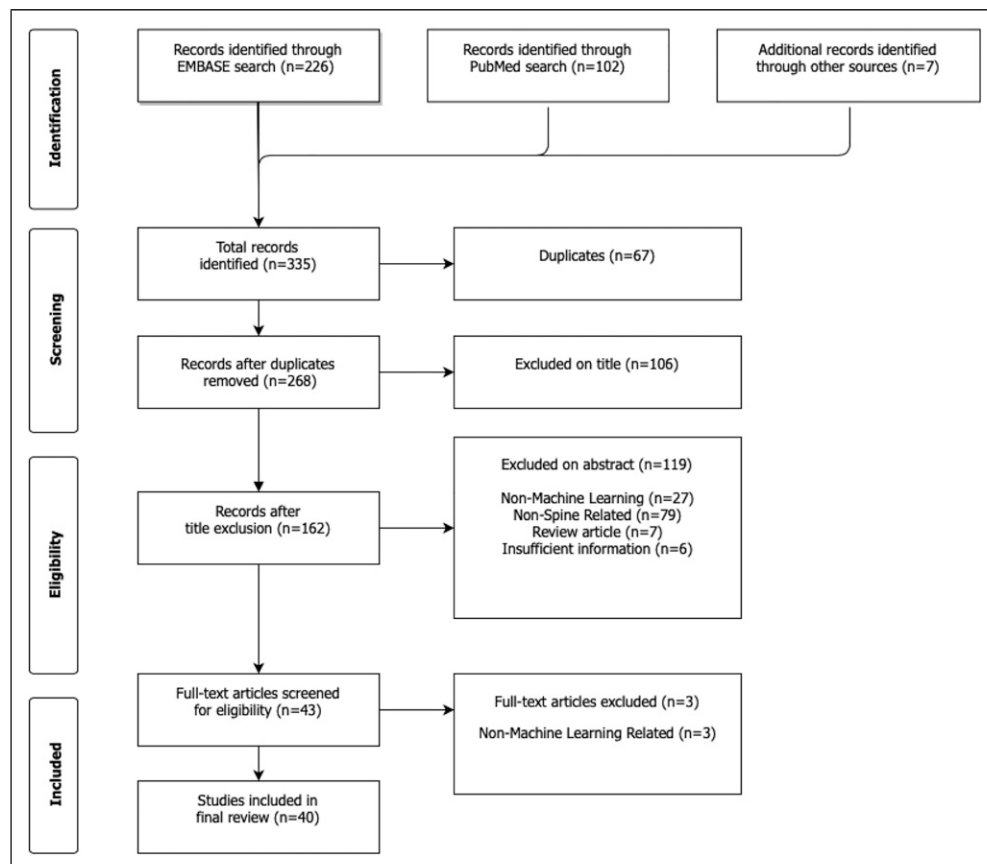


Figure 1. PRISMA flowchart showing systematic review search strategy.

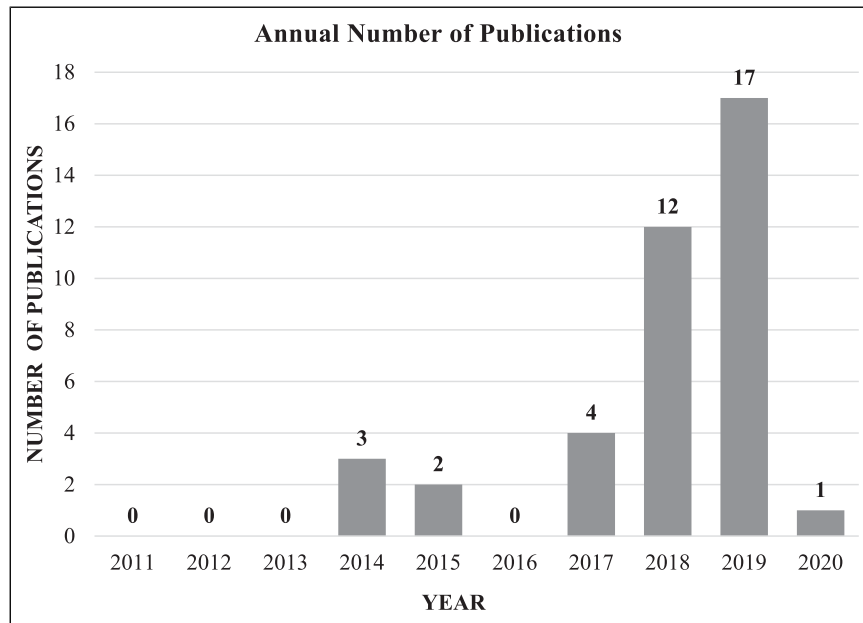


Figure 2. Trends in the annual number of AI/ML publications in spine surgery (2011 to 2020).

length of stay, discharges, readmissions, and other cost-contributing factors (Table 1, Supplementary Table 3). Eleven studies (26.8%) evaluated AI/ML applications in accurately predicting patient reoperations, operating time, hospital length of stay, discharges, readmissions, or surgical and inpatient costs.¹³⁻²³ Four studies (9.8%) used patients' preoperative risk factors and other patient-specific variables to optimize the patient selection and surgical planning process through the use of AI/ML-based predictions of surgical outcomes and postoperative complications.²⁴⁻²⁷ Four studies (9.8%) investigated the use of AI/ML in improving preoperative planning, through accurate identification of previously implanted anterior cervical spinal implants and a decision support system for spine fusion surgery that enhances surgical planning through prediction of pedicle screw pullout strength for any combination of patient-specific factors.^{24,28-30} The authors of 17 studies mentioned the potential of AI/ML applications in bringing down the costs of spine surgery and optimizing cost-efficient, value-based care delivery.

The majority of the decision support studies evaluated AI/ML model performance using ROC/AUC, accuracy, sensitivity, and specificity. Two studies did not test model performance, but instead optimized preoperative patient selection using cluster analysis to classify patients based on preoperative risk factors and other variables. Four studies each evaluated different AI/ML-based predictive models of readmissions and reoperations, with an average AUC of .67 (SD .08) across 11 models and five different ML methods (Table 2). Predictive models of LOS and discharges were used in eight studies, with an average AUC of .80 (SD .08) across 23 models and six different AI/ML methods. Applications of preoperative patient selection/planning and cost prediction were used in 11 models across nine studies, reporting

an average AUC of .89 (SD .08). ANOVA testing found statistically significant variability in model AUC, accuracy, and specificity across the different decision support applications ($P < .001$, $P = .005$, $P < .001$, respectively), and preoperative planning/cost prediction models and discharge/LOS models each reported significantly higher average accuracy (82.2% and 78.0%, respectively) compared to readmissions/reoperation prediction models ($P = .009$, $P = .019$, respectively). The same relationships were confirmed in comparisons of model specificity by Tukey post hoc testing. There were no significant differences in model sensitivity between the applications (Table 2).

Prediction and Management of Postoperative Outcomes and Complications

A total of 24 reviewed studies used various AI/ML models to predict outcomes, complications, and adverse events (Supplementary Table 2, Table 3).^{19,23,31-52} Average AUC and accuracy values significantly varied ($P < .001$ and $P < .027$, respectively) (Table 4). AUC for predicting postoperative cardiovascular complications averaged .69 (SD .12, 21 models), other postoperative complications averaged .68 (SD .12, 31 models), postoperative mortality models averaged .82 (SD .08, 30 models), and postoperative functional and clinical outcome models averaged .75 (SD .09, 30 models). Tukey post hoc testing found statistically significant differences between postoperative mortality models (average AUC of .82) and each of the other prediction models (Table 4). Average accuracy was also found to be significantly different between other postoperative complications and postoperative functional and clinical outcomes, (85.8% vs 72.2%, respectively) ($P = .027$) and there was no significant variation in reported sensitivity and specificity values (Table 4).

Table 1. Reviewed Studies of Preoperative Patient Selection and Planning in Spine Surgery.

Author, year	Pathology or procedure	ML algorithms	Prediction outputs	Number of patients	Avg. age	% female	Data source
Kalagara, 2018	Lumbar laminectomy	Boosting/ensemble learning	Readmissions/reoperations	4030	63	–	ACS-NSQIP database
Stopa, 2019	Spine fusion	Deep Learning/ANN	Discharge/LOS	144	50	45.10	Single center
Ames, 2019	ASD	Cluster analysis	Pre-op selection/planning	570	56.8	78.80	Multicenter ASD databases
Goyal, 2019	Spine fusion	Regression analysis, boosting/ensemble learning, deep Learning/ANN, decision tree, and Bayesian networks	Discharge/LOS and readmissions/reoperations	8872	57	48.50	ACS-NSQIP database
Ogink, 2019	Spondylolisthesis	Deep learning/ANN, SVM, decision tree, and Bayesian networks	Discharge/LOS	1868	63	63.00	ACS-NSQIP database
Kuo, 2018	Spinal fusion	Regression analysis, SVM, decision tree, and Bayesian networks	Cost prediction	532	62.4	58.60	Single center
Lerner, 2019	Posterior lumbar spinal fusion	Cluster analysis	Pre-op selection/planning	18770	51.3	56.10	IBM MarketScan® commercial database
Siccoli, 2019	Lumbar decompression	Boosting/ensemble learning and decision tree	Discharge/LOS and readmissions/reoperations	635	62	48.00	Single center
Chia, 2017	Cerebral palsy	Deep learning/ANN	Pre-op selection/planning	242	–	–	Single center
Huang, 2019	ACDF	Bayesian networks, SVM, and regression analysis	Pre-op selection/planning	321	–	–	Single center
Varghese, 2018	Spinal fusion	Decision tree and regression analysis	Pre-op selection/planning	–	–	–	Single center
Karhade, 2018	Lumbar degeneration	Deep learning/ANN, decision tree, SVM, and Bayesian networks	Discharge/LOS	5273	53	46.90	ACS-NSQIP database
Hopkins, 2019	Posterior lumbar spinal fusion	Deep learning/ANN	Readmissions/reoperations	5816	–	–	ACS-NSQIP database
Ogink, 2019	Lumbar spinal stenosis	Deep Learning/ANN, decision tree, SVM, and Bayesian networks	Discharge/LOS	9338	67	47.30	ACS-NSQIP database
Karnuta, 2019	Spinal fusion	Bayesian networks	Discharge/LOS and cost prediction	3807	–	57.80	New York state SPARCS database
Khatrri, 2019	Spinal fusion	Decision tree	Pre-op selection/planning	–	–	–	Single center
Bekelis, 2014	ACDF	Regression analysis	Discharge/LOS and readmissions/reoperations	2732	55.7	46.30	ACS-NSQIP database
Assi, 2014	Scoliosis	Regression analysis	Pre-op selection/planning	141	–	–	Single center

Abbreviations: ASD, adult spinal deformity; ACDF, anterior cervical discectomy and fusion; ANN, artificial neural network; SVM, support vector machine; LOS, length of stay; ACS-NSQIP, American College of Surgery-National Surgical Quality Improvement Program; SPARCS, Statewide Planning and Research Co-operative System.

Comparison of AI/ML Algorithms

The most commonly applied AI/ML algorithms in the reviewed studies were logistic regression (24 studies, 58.5%), while cluster analysis was only used in 3 studies (7.3%)

(Supplementary Table 2, Table 5). When comparing AI/ML model performance across various algorithm types, there was statistically significant variation confirmed by one-way ANOVA testing. Bayesian networks had the highest average AUC (.80, SD .09, 13 models), while support vector

Table 2. Statistical Comparisons of Reported Model Performance Metrics, by Administrative/Clinical Decision Support Application.

Administrative or clinical decision support applications	Performance metrics: Mean (SD, N)			
	AUC	Accuracy	Sensitivity	Specificity
Preoperative planning and cost prediction	.89 (.08, 11)	82.2 (4.8, 7)	70.5 (10.9, 6)	87.7 (5.1, 6)
Discharge, LOS	.80 (.08, 23)	78.0 (7.7, 9)	69.1 (19.8, 7)	76.6 (7.8, 7)
Readmissions and reoperations	.67 (.08, 11)	70.2 (11.8, 8)	56.0 (16.5, 7)	59.0 (16.5, 7)
ANOVA	$P < .001$	$P = .005$	$P = .472$	$P < .001$
Tukey post hoc tests	1 vs 2 ($P = .005$)	1 vs 3 ($P = .009$)	–	1 vs 3 ($P < .001$)
	1 vs 3 ($P < .001$)	2 vs 3 ($P = .019$)	–	2 vs 3 ($P = .002$)
	2 vs 3 ($P < .001$)	–	–	–

Abbreviations: AUC, area under the curve; SD, standard deviation; N, number of models; LOS, length of stay.

Table 3. Reviewed Studies of Postoperative Outcome Prediction in Spine Surgery.

Author/ year	Pathology or procedure	ML algorithms	Prediction outputs	Number of patients	Avg age	% Female	Data source
Arvind, 2018	ACDF	Deep learning/ANN, regression analysis, and SVM	Cardiac, VTE, wound infection, and 30-day mortality	6264	53	52	Multi-center database
Kim, 2018	Lumbar decompression	Deep learning/ANN and regression analysis	Cardiac, VTE, wound infection, and 30-day mortality	6789	60	55	ACS-NSQIP database
Kim, 2018	ASD	Deep learning/ANN and regression analysis	Cardiac, VTE, wound infection, and 30-day mortality	1746	60	59	ACS-NSQIP database
Karhade, 2019	ACDF	Deep learning/ANN, SVM, decision tree, regression analysis, and boosting/ ensemble learning	Sustained opioid use	2737	51	53	Multi-center database
Han, 2019	General	Regression analysis	Adverse events, cardiac/ CHF, neurologic, pulmonary/pneumonia, and overall medical/ surgical complication	331870	63	54	IBM MarketScan, CMS Medicaid, Medicare databases
Durand, 2018	ASD	Decision tree	Postoperative blood transfusion	205	54	66	ACS-NSQIP database
Karhade, 2019	Spinal metastatic disease	Deep learning/ANN, regression analysis, SVM, decision tree, and boosting/ ensemble learning	90-day mortality, and 1- year mortality	732	61	42	Single center
Scheer, 2017	ASD	Decision tree	Adverse events and major complications	557	58	79	Multi-center ASD databases
Janssen, 2018	Thoracolumbar spine surgery	Regression analysis	Wound infection	898	52	51	NCI SEER registry
Karhade, 2019	Spinal epidural abscess	Deep Learning/ANN, regression analysis, SVM, decision tree, and boosting/ ensemble learning	Mortality: 90-day	1053	59	39	Multi-center database
Karhade, 2019	Lumbar spine surgery	Regression analysis	Sustained opioid use	8435	60	46	Multi-center database
Ryu, 2018	Spinal ependymoma	Decision tree and regression analysis	5-year mortality and 10- year mortality	2822	–	47	NCI SEER registry

(continued)

Table 3. (continued)

Author/ year	Pathology or procedure	ML algorithms	Prediction outputs	Number of patients	Avg age	% Female	Data source
Khan, 2020	Degenerative cervical myelopathy (DCM)	Decision tree, regression analysis, SVM, and boosting/ensemble learning	PRO/functional outcomes (SF-36 MCS, PCS)	193	52	35	Multi-center AOSpine CSM clinical trials
Staatjes, 2019	Single-level tubular microdiscectomy for lumbar disc herniation	Deep learning/ANN and regression analysis	Clinical improvement (leg pain, back pain, and functional disability)	422	49	49	Single center
Hoffman, 2015	Cervical spondylotic myelopathy	SVM and regression analysis	PRO/functional outcomes (post-op ODI)	20	60	45	Single center
Shamim, 2009	Lumbar disc surgery	Cluster analysis	Post-op poor outcomes	501	41	31	Single center
Azimi, 2014	Lumbar spinal stenosis	Deep learning/ANN and regression analysis	Post-op patient satisfaction	168	60	65	Single center
Azimi, 2015	Lumbar disk herniation	Deep learning/ANN and regression analysis	Post-op recurrent lumbar disc herniation	402	50	54	Single center
Azimi, 2017	Lumbar disk herniation	Deep learning/ANN	Post-op successful outcomes	203	48	53	Single center
Buchlak, 2017	ASD	Regression analysis	Post-op complications	136	63	74	Single center
Karhade, 2018	Spinopelvic chordoma surgery	Deep Learning/ANN, decision tree, SVM, and Bayesian networks	5-year mortality	265	64	39	NCI SEER registry
Khor, 2018	Lumbar fusion	Regression analysis	Clinical improvement	1965	61	60	Multi-center database
Bekelis, 2014	ACDF	Regression analysis	Cardiac, VTE, wound infection, and 30-day mortality	2732	56	46	ACS-NSQIP database
Siccoli, 2019	Lumbar decompression	Deep learning/ANN, decision tree, and Bayesian networks	Clinical improvement (6wk, 12wk)	635	62	48	Single center
Ames, 2019	Adult spinal deformity	Regression analysis, decision tree, and boosting/ensemble learning	PRO/functional outcomes (SRS-22R)	561	54.4	75.9	Single center

Abbreviations: ANN, artificial neural network; SVM, support vector machine; LOS, length of stay; ASD, adult spinal deformity; ACDF, anterior cervical discectomy and fusion; CHF, congestive heart failure; VTE, venous thromboembolism; UTI, urinary tract infection; PRO, patient-reported outcomes; SF-36, short-form 36 questionnaire; MCS, mental health composite score; PCS, physical health composite score; ODI, Oswestry disability index; ACS-NSQIP, American College of Surgery-National Surgical Quality Improvement Program; CMS, centers for Medicare and Medicaid services; NCI SEER, National Cancer Institute Surveillance, Epidemiology, and End Results database; AOSpine CSM, AOSpine North America cervical spondylotic myelopathy study.

machines (SVMs) had the lowest average AUC (.63, SD .18, 17 models). Tukey post hoc testing found significant differences in AUC between SVM and several AL/ML algorithms, including Bayesian network ($P = .009$), decision tree ($P = .018$), and deep learning/ANN ($P = .019$) (Table 5). There was significant variation in reported average sensitivity of the AL/ML algorithms ($P = .006$). Deep learning/ANN had the highest reported average sensitivity of 81.5% (SD 12.1%, 8 models), while Bayesian network and boosting/ensemble learning had the lowest reported sensitivity, with 63.7%

(SD 11.0%, 4 models) and 55.7% (SD 21.7%, 7 models), respectively. ANOVA testing did not detect significant differences in accuracy or specificity between the AI/ML algorithms ($P = .083$ and $P = .554$, respectively) (Table 5).

Discussion

This systematic review is the first to evaluate and summarize AI/ML applications in optimizing patient selection and predicting surgical outcomes and complications in spine surgery.

Table 4. Statistical Comparisons of Reported Model Performance Metrics, by Postoperative Prediction/Management Application.

Postoperative prediction/management applications	Performance metrics: Mean (SD, N)			
	AUC	Accuracy	Sensitivity	Specificity
Postoperative cardiovascular complications	.69 (.12, 21)	–	81.0 (4.2, 2)	52.0 (1.4, 2)
Other postoperative complications	.68 (.12, 31)	85.8 (7.9, 4)	77.6 (4.4, 5)	51.6 (.5, 5)
Postoperative mortality	.82 (.08, 30)	–	–	–
Postoperative functional or clinical outcomes	.75 (.09, 30)	72.2 (11.2, 28)	73.8 (15.5, 24)	60.9 (17.5, 24)
ANOVA	$P < .001$	$P = .027$	$P = .487$	$P = .278$
Tukey post hoc tests	1 vs 3 ($P < .001$)	–	–	–
	2 vs 3 ($P < .001$)	–	–	–
	3 vs 4 ($P = .035$)	–	–	–

Abbreviations: AUC, area under the curve; SD, standard deviation; N, number of models; LOS, length of stay.

Table 5. Statistical Comparisons of Reported Model Performance Metrics, by AI/ML Algorithm.

AI/ML algorithm	Performance metrics: Mean (SD, N)			
	AUC	Accuracy	Sensitivity	Specificity
Bayesian network (BN)	.80 (.09, 13)	76.9 (11.9, 8)	63.7 (11.0, 4)	67.4 (17.7, 4)
Boosting/ensemble learning (BEL)	.76 (.10, 13)	74.1 (9.6, 8)	55.7 (21.7, 7)	71.7 (11.4, 7)
Decision tree (DT)	.77 (.11, 29)	74.0 (8.7, 13)	75.4 (13.7, 12)	62.5 (21.7, 12)
Deep learning/artificial neural network (ANN)	.77 (.11, 34)	83.0 (10.7, 10)	81.5 (12.1, 8)	71.8 (10.1, 8)
Logistic regression (LR)	.74 (.11, 56)	70.4 (10.6, 13)	70.6 (12.4, 19)	61.0 (12.4, 19)
Support vector machines (SVM)	.63 (.18, 17)	67.5 (12.9, 3)	72.3 (18.3, 3)	56.0 (42.9, 3)
ANOVA	$P = .007$	$P = .083$	$P = .006$	$P = .554$
Tukey post hoc tests	BN vs SVM ($P = .009$)	–	BEL vs ANN ($P = .002$)	–
	DT vs SVM ($P = .018$)	–	–	–
	ANN vs SVM ($P = .019$)	–	–	–

Abbreviations: AUC, area under the curve; SD, standard deviation; N, number of models; AI/ML, artificial intelligence and machine learning.

Our review included 41 studies from the literature which tested AI/ML-based prediction and optimization models that may help guide clinical decision-making and surgical planning. Among all the reviewed studies, AI/ML models were fairly accurate, averaging 74.9% overall accuracy and AUC of .75, across all AI/ML methods. In particular, AI/ML models performed best in optimizing preoperative patient selection and planning and predicting costs, hospital discharge, and length of stay. Model performance was also good or fair (AUC between .70 and .89) in predicting postoperative mortality and functional and clinical outcomes. However, model performance was considered poor (AUC between .50 and .69) in predicting postoperative complications (including cardiovascular complications), adverse events, and readmissions and reoperations, which may be due to the difficulty in predicting random events which are out of the surgeon's control in the postoperative period. In addition, model performance metrics such as AUC must also be carefully interpreted, especially because AUC balances a model's precision and recall (and resulting false positives and false negatives), and in certain clinical applications such as cancer screening or prediction of

potentially fatal complications after spine surgery, providers may prefer a model with a lower AUC that minimizes false negatives.^{4,53}

Although AI/ML models did not perform well in predicting postoperative complications, they offer a potentially beneficial tool for providers to optimize preoperative planning and improve cost-efficiency. For example, a practicing surgeon may use an electronic medical record system with an integrated AI/ML application that accurately predicts which patients will almost certainly require inpatient vs outpatient surgery to ensure that these high-risk patients have access to specialized care and supervision post-operatively. As a result, surgeons can have an incredibly accurate aid in patient selection, thus ensuring that patients are treated in the appropriate setting. In a systematic review of AI/ML applications in neurosurgery, Buchlak et al.⁵⁴ reported similar model performance results for deep learning/ANN and logistic regression models as our study. However, their findings reported SVM performance to have an average AUC of .80 and accuracy of 81.8%, which is significantly higher than the results from our study for SVM. It appears that SVM may be shown

to be more accurate in certain non-spine neurosurgical applications, such as image classification,⁵⁵⁻⁶⁰ but is perhaps less accurate for guided decision-making in spine surgery.

AI/ML-based predictive modeling may be especially beneficial in spine surgery, which usually involves complex procedures with potentially high complication rates in often highly comorbid patient population. Ames et al.²⁶ showed that an AI/ML-based classification system for ASD surgical candidates optimizes personalized treatment plans based on patient-specific risk factors. This application may aid surgeons with pre-operative decision-making by informing them about which treatment options may offer optimal clinical improvement and value with the lowest risk of adverse events. In our review, several of the AI/ML prediction and optimization models that were used to improve patient care and postoperative outcomes also showed the potential to reduce unnecessary healthcare expenditures and even provide risk-adjusted reimbursement models for providers and hospitals.^{13,14,22,26} The predictive capabilities of AI/ML models enable decision makers to forecast costs related to postoperative outcomes and complications, pain medication use, patient discharges and discharge placements, length of stay, unplanned readmissions, and other postoperative interventions. The authors highlighted the potential of AI/ML to improve clinical decision making and patient care by predicting likely postoperative outcomes, which enables providers to optimize resource allocation for post-surgical monitoring and focused care of high-risk patients.^{16,17,20,22} Of particular relevance to curtailing rising inpatient costs, accurate forecasting of hospital length of stay has important implications for management of bed utilization and other hospital resources. Kalagara et al.¹³ analyzed hospital readmissions after laminectomy and used patient-specific variables to develop predictive models for identifying readmitted patients with over 95% accuracy.

AI/ML may also aid surgeons and clinical decision makers to more efficiently plan for surgery and select patients for the optimal surgical setting (for instance, outpatient vs inpatient) that will produce the best care outcomes while improving cost-efficiency. Several studies highlighted the potential value of predictive modeling during the pre-operative period in helping surgeons with optimizing patient selection for surgery and surgical planning, which also allows providers to efficiently allocate needed hospital resources and plan for possible postoperative interventions to ensure the best possible outcomes.^{22,25-27} The recent shift toward value-based health care has likely also spurred the recent spike in research of AI/ML applications in optimizing cost-efficiency and resource allocation, especially because post-surgical inpatient care and other associated hospital costs are major drivers of US healthcare expenditures⁶¹ and spine surgery costs.⁶²⁻⁶⁷ In contrast, outpatient surgical procedures have been shown to be comparatively less costly than inpatient treatment, and treating suitable surgical candidates in the outpatient setting may offer significant cost savings.⁶⁸⁻⁷³ Development of well-defined and

accurate patient selection criteria for outpatient surgery, along with optimized anesthesia and postoperative pain management protocols, are associated with reduced patient readmission risk and surgical costs.⁷⁴⁻⁷⁶ Predictive modeling of patient length of stay, based on their medical comorbidities, demographic profile, and other variables, may aid surgeons in the selection of outpatient surgical candidates, and has been shown to be effective in selecting patients for outpatient posterior spinal fusions.²⁷ Through the use of patient-specific risk factors, AI/ML applications may also enable development of risk-adjusted insurance reimbursement models which compensate providers and hospitals commensurate with the case complexity and patient complication risk and comorbidities, providing a potential solution for unwillingness to treat medically complex patients. However, issues of data privacy and security when using AI/ML remain a major challenge which must be addressed, as patients may feel uncomfortable with their personal health information being used on such a large scale.

Although there has been a recent significant increase in the number of AI/ML publications in spine surgery, there remains a general lack of large, powered, and externally validated studies which would elucidate more information on their efficacy in spine surgery practice. In addition, it is important to note that although that our review included studies through February 2020, it still provides a detailed overview of the recent trends in the literature and the potential early applications of AI/ML. Many of the reviewed studies involved different spine procedures that vary in complexity and risk and included studies with models which varied in the quality and quantity of their training and validation data. As such, any conclusions about the efficacy of AI/ML applications in spine surgery require further investigation. This study does not make conclusive relationships between AI/ML and clinical efficacy, but instead presents statistical findings and trends from recent studies. Future directions in research of AI/ML applications in spine surgery, and in health care, must focus on developing externally validated and commercially viable systems that can be easily implemented and incorporated with already-existing hospital systems in a cost-efficient manner. In addition, future studies should evaluate optimal methods that aid in determining surgical candidates and which can use a wide range of preoperative data. An understanding of AI/ML-based applications is becoming increasingly important, particularly in spine surgery, as the volume of reported literature, technology accessibility, and clinical applications continues to rapidly expand.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

IRB statement

This study utilized national, de-identified data and is exempt from IRB review.

ORCID iDs

Venkat Boddapati  <https://orcid.org/0000-0002-3333-2234>

Justin Mathew  <https://orcid.org/0000-0002-7699-780X>

Nicholas C. Danford  <https://orcid.org/0000-0002-9620-2023>

References

- Deo RC. Machine Learning in Medicine. *Circulation*. 2015; 132(20):1920–1930.
- Awan SE, Soheli F, Sanfilippo FM, Bennamoun M, Dwivedi G. Machine learning in heart failure: ready for prime time. *Curr Opin Cardiol*. 2018;33(2):190–195.
- Smektala R, Endres HG, Dasch B, et al. The effect of time-to-surgery on outcome in elderly patients with proximal femoral fractures. *BMC Musculoskelet Disord*. 2008;9:171.
- Wong D, Yip S. Machine learning classifies cancer. *Nature*. 2018;555(7697):446–447.
- Lin CC, Ou YK, Chen SH, Liu YC, Lin J. Comparison of artificial neural network and logistic regression models for predicting mortality in elderly patients with hip fracture. *Injury*. 2010;41(8):869–873.
- Chung SW, Han SS, Lee JW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. 2018;89(4):468–473.
- Ramkumar PN, Navarro SM, Haeberle HS, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. *J Arthroplasty*. 2019;34(4):632–637.
- OCEBM. OCEBM levels of evidence. Oxford, UK: University of Oxford; 2020. <https://www.cebm.net/2016/05/ocebml-levels-of-evidence/>
- Biron DR, Sinha I, Kleiner JE, et al. A novel machine learning model developed to assist in patient selection for outpatient total shoulder arthroplasty. *J Am Acad Orthop Surg*. 2019;28(13):e580–e585.
- Navarro SM, Wang EY, Haeberle HS, et al. Machine learning and primary total knee arthroplasty: patient forecasting for a patient-specific payment model. *J Arthroplasty*. 2018;33(12):3617–3623.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143(1):29–36.
- Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. United Kingdom: Wiley; 2020. <http://handbook-5-1.cochrane.org/>.
- Kalagara S, Eltorai AEM, Durand WM, DePasse JM, Daniels AH. Machine learning modeling for predicting hospital readmission following lumbar laminectomy. *J Neurosurg Spine*. 2018;30(3):344–352.
- Stopa BM, Robertson FC, Karhade AV, et al. Predicting non-routine discharge after elective spine surgery: external validation of machine learning algorithms [published ahead of print July 26, 2019]. *J Neurosurg Spine*. doi:10.3171/2019.5.SPINE1987.
- Goyal A, Ngufo C, Kerezoudis P, McCutcheon B, Storlie C, Bydon M. Can machine learning algorithms accurately predict discharge to nonhome facility and early unplanned readmissions following spinal fusion? Analysis of a national surgical registry. *J Neurosurg Spine*. 2019:1–11.
- Ogink PT, Karhade AV, Thio Q, et al. Development of a machine learning algorithm predicting discharge placement after surgery for spondylolisthesis. *Eur Spine J*. 2019;28(8):1775–1782.
- Ogink PT, Karhade AV, Thio Q, et al. Predicting discharge placement after elective surgery for lumbar spinal stenosis using machine learning methods. *Eur Spine J*. 2019;28(6):1433–1440.
- Kuo CY, Yu LC, Chen HC, Chan CL. Comparison of models for the prediction of medical costs of spinal fusion in Taiwan diagnosis-related groups by machine learning algorithms. *Health Inform Res*. 2018;24(1):29–37.
- Siccoli A, de Wispelaere MP, Schroder ML, Staartjes VE. Machine learning-based preoperative predictive analytics for lumbar spinal stenosis. *Neurosurg Focus*. 2019;46(5):E5.
- Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. *Neurosurg Focus*. 2018;45(5):E6.
- Hopkins BS, Yamaguchi JT, Garcia R, et al. Using machine learning to predict 30-day readmissions after posterior lumbar fusion: an NSQIP study involving 23,264 patients [published ahead of print November 29, 2019]. *J Neurosurg Spine*. doi:10.3171/2019.9.SPINE19860.
- Karnuta JM, Golubovsky JL, Haeberle HS, et al. Can a machine learning model accurately predict patient resource utilization following lumbar spinal fusion? *Spine J*. 2019;20(3):329–336.
- Bekelis K, Desai A, Bakhom SF, Missios S. A predictive model of complications after spine surgery: the National surgical quality improvement program (NSQIP) 2005–2010. *Spine J*. 2014;14(7):1247–1255.
- Varghese V, Krishnan V, Kumar GS. Evaluating pedicle-screw instrumentation using decision-tree analysis based on pullout strength. *Asian Spine J*. 2018;12(4):611–621.
- Chia K, Fischer I, Thomason P, Graham K, Sangeux M. Is it feasible to use an automated system to recommend orthopaedic surgeries? *Gait & Posture*. 2017;57:89.
- Ames CP, Smith JS, Pellise F, et al. Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. *Spine (Phila Pa 1976)*. 2019;44(13):915–926.
- Lerner J, Ruppenkamp J, Etter K, et al. Preoperative behavioral health, opioid, and antidepressant utilization and 2-year costs after spinal fusion-revelations from cluster analysis. *Spine (Phila Pa 1976)*. 2020;45(2):E90–E98.
- Huang KT, Silva MA, See AP, et al. A computer vision approach to identifying the manufacturer and model of anterior cervical spinal hardware [published ahead of print September 6, 2019]. *J Neurosurg Spine*. doi:10.3171/2019.6.SPINE19463.

29. Khatri R, Varghese V, Sharma S, Kumar GS, Chhabra HS. Pullout strength predictor: a machine learning approach. *Asian Spine J.* 2019;13(5):842–848.
30. Assi KC, Labelle H, Cheriet F. Statistical model based 3D shape prediction of postoperative trunks for non-invasive scoliosis surgery planning. *Comput Biol Med.* 2014;48:85–93.
31. Arvind V, Kim JS, Oermann EK, Kaji D, Cho SK. Predicting surgical complications in adult patients undergoing anterior cervical discectomy and fusion using machine learning. *Neurospine.* 2018;15(4):329–337.
32. Kim JS, Merrill RK, Arvind V, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine (Phila Pa 1976).* 2018;43(12):853–860.
33. Kim JS, Arvind V, Oermann EK, et al. Predicting surgical complications in patients undergoing elective adult spinal deformity procedures using machine learning. *Spine Deform.* 2018;6(6):762–770.
34. Karhade AV, Ogink PT, Thio Q, et al. Machine learning for prediction of sustained opioid prescription after anterior cervical discectomy and fusion. *Spine J.* 2019;19(6):976–983.
35. Karhade AV, Schwab JH, Bedair HS. Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty. *J Arthroplasty.* 2019;34(10):2272–2277.e2271.
36. Karhade AV, Thio Q, Ogink PT, et al. Predicting 90-day and 1-year mortality in spinal metastatic disease: development and internal validation. *Neurosurgery.* 2019;85(4):E671–E681.
37. Karhade AV, Shah AA, Bono CM, et al. Development of machine learning algorithms for prediction of mortality in spinal epidural abscess. *Spine J.* 2019;19(12):1950–1959.
38. Karhade AV, Ahmed AK, Pennington Z, et al. External validation of the SORG 90-day and 1-year machine learning algorithms for survival in spinal metastatic disease. *Spine J.* 2020;20(1):14–21.
39. Han SS, Azad TD, Suarez PA, Ratliff JK. A machine learning approach for predictive models of adverse events following spine surgery. *Spine J.* 2019;19(11):1772–1781.
40. Durand WM, DePasse JM, Daniels AH. Predictive modeling for blood transfusion after adult spinal deformity surgery: a tree-based machine learning approach. *Spine (Phila Pa 1976).* 2018;43(15):1058–1066.
41. Scheer JK, Smith JS, Schwab F, et al. Development of a preoperative predictive model for major complications following adult spinal deformity surgery. *J Neurosurg Spine.* 2017;26(6):736–743.
42. Janssen DMC, van Kuijk SMJ, d'Aumerie BB, Willems PC. External validation of a prediction model for surgical site infection after thoracolumbar spine surgery in a Western European cohort. *J Orthop Surg Res.* 2018;13(1):114.
43. Ryu SM, Lee SH, Kim ES, Eoh W. Predicting survival of patients with spinal ependymoma using machine learning algorithms with the SEER database [published ahead of print December 28, 2018]. *World Neurosurg.* doi:10.1016/j.wneu.2018.12.091.
44. Khan O, Badhiwala JH, Witiw CD, Wilson JR, Fehlings MG. Machine learning algorithms for prediction of health-related quality-of-life after surgery for mild degenerative cervical myelopathy [published ahead of print February 8, 2020]. *Spine J.* doi:10.1016/j.spinee.2020.02.003.
45. Staartjes VE, de Wispelaere MP, Vandertop WP, Schroder ML. Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: feasibility of center-specific modeling. *Spine J.* 2019;19(5):853–861.
46. Hoffman H, Lee SI, Garst JH, 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(9):1444–1449.
47. 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(6):565–572; discussion 572.
48. 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(3):300–305.
49. 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(3):E161–165.
50. 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(2):173–177.
51. Buchlak QD, Yanamadala V, Leveque JC, Edwards A, Nold K, Sethi R. The Seattle spine score: predicting 30-day complication risk in adult spinal deformity surgery. *J Clin Neurosci.* 2017;43:247–255.
52. Khor S, Lavalley D, Cizik AM, et al. Development and validation of a prediction model for pain and functional outcomes after lumbar spine surgery. *JAMA Surg.* 2018;153(7):634–642.
53. Lynch CM, Abdollahi B, Fuqua JD, et al. Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int J Med Inform.* 2017;108:1–8.
54. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev.* 2019;43(5):1235–1253.
55. Yu S, Tan KK, Sng BL, Li S, Sia AT. Lumbar Ultrasound Image Feature Extraction and Classification with Support Vector Machine. *Ultrasound Med Biol.* 2015;41(10):2677–2689.
56. Ramirez L, Durdle NG, Raso VJ. A machine learning approach to assess changes in scoliosis. *Stud Health Technol Inform.* 2008;140:254–259.
57. Adankon MM, Dansereau J, Labelle H, Cheriet F. Non invasive classification system of scoliosis curve types using least-squares support vector machines. *Artif Intell Med.* 2012;56(2):99–107.
58. Valsky D, Marmor-Levin O, Deffains M, et al. Stop! border ahead: automatic detection of subthalamic exit during deep brain stimulation surgery. *Mov Disord.* 2017;32(1):70–79.

59. Akbari H, Macyszyn L, Da X, et al. Pattern analysis of dynamic susceptibility contrast-enhanced MR imaging demonstrates peritumoral tissue heterogeneity. *Radiology*. 2014;273(2):502–510.
60. Akbari H, Macyszyn L, Da X, et al. Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma. *Neurosurgery*. 2016;78(4):572–580.
61. Papanicolas I, Woskie LR, Jha AK. Health care spending in the United States and other high-income countries. *JAMA*. 2018;319(10):1024–1039.
62. Missios S, Bekelis K. Hospitalization cost after spine surgery in the United States of America. *J Clin Neurosci*. 2015;22(10):1632–1637.
63. Singh K, Nandyala SV, Marquez-Lara A, et al. A perioperative cost analysis comparing single-level minimally invasive and open transforaminal lumbar interbody fusion. *Spine J*. 2014;14(8):1694–1701.
64. Liu CY, Zygourakis CC, Yoon S, et al. Trends in utilization and cost of cervical spine surgery using the National inpatient sample database, 2001 to 2013. *Spine (Phila Pa 1976)*. 2017;42(15):E906–E913.
65. Zygourakis CC, Liu CY, Keefe M, et al. Analysis of national rates, cost, and sources of cost variation in adult spinal deformity. *Neurosurgery*. 2018;82(3):378–387.
66. Shields LB, Clark L, Glassman SD, Shields CB. Decreasing hospital length of stay following lumbar fusion utilizing multidisciplinary committee meetings involving surgeons and other caretakers. *Surg Neurol Int*. 2017;8:5.
67. Epstein NE, Schwall G, Reilly T, Insinna T, Bahnken A, Hood DC. Surgeon choices, and the choice of surgeons, affect total hospital charges for single-level anterior cervical surgery. *Spine (Phila Pa 1976)*. 2011;36(11):905–909.
68. Ford MC, Walters JD, Mulligan RP, et al. Safety and cost-effectiveness of outpatient unicompartmental knee arthroplasty in the ambulatory surgery center: a matched cohort study. *Orthop Clin North Am*. 2020;51(1):1–5.
69. Walters JD, Walsh RN, Smith RA, Brolin TJ, Azar FM, Throckmorton TW. Bundled payment plans are associated with notable cost savings for ambulatory outpatient total shoulder arthroplasty. *J Am Acad Orthop Surg*. 2019;28(19):795–801.
70. Brolin TJ, Throckmorton TW. Outpatient shoulder arthroplasty. *Orthop Clin North Am*. 2018;49(1):73–79.
71. Brolin TJ, Mulligan RP, Azar FM, Throckmorton TW. Neer award 2016: outpatient total shoulder arthroplasty in an ambulatory surgery center is a safe alternative to inpatient total shoulder arthroplasty in a hospital: a matched cohort study. *J Shoulder Elbow Surg*. 2017;26(2):204–208.
72. Fournier MN, Brolin TJ, Azar FM, Stephens R, Throckmorton TW. Identifying appropriate candidates for ambulatory outpatient shoulder arthroplasty: validation of a patient selection algorithm. *J Shoulder Elbow Surg*. 2019;28(1):65–70.
73. Oh J, Perlas A, Lau J, Gandhi R, Chan VW. Functional outcome and cost-effectiveness of outpatient vs inpatient care for complex hind-foot and ankle surgery. A retrospective cohort study. *J Clin Anesth*. 2016;35:20–25.
74. Berger RA, Sanders S, Gerlinger T, Della Valle C, Jacobs JJ, Rosenberg AG. Outpatient total knee arthroplasty with a minimally invasive technique. *J Arthroplasty*. 2005;20(7 suppl 3):33–38.
75. Hanson NA, Lee PH, Yuan SC, Choi DS, Allen CJ, Auyong DB. Continuous ambulatory adductor canal catheters for patients undergoing knee arthroplasty surgery. *J Clin Anesth*. 2016;35:190–194.
76. Cullom C, Weed JT. Anesthetic and analgesic management for outpatient knee arthroplasty. *Curr Pain Headache Rep*. 2017;21(5):23.