Artificial Intelligence and Machine Learning Applications in Spine Surgery

Nathan J. Lee, Joseph M. Lombardi and Ronald A. Lehman

Int J Spine Surg 2023, 17 (S1) S18-S25
doi: https://doi.org/10.14444/8503
https://www.ijssurgery.com/content/17/S1/S18

The International Journal of Spine Surgery
2397 Waterbury Circle, Suite 1,
Aurora, IL 60504, Phone: +1-630-375-1432

International Journal of Spine Surgery, Vol. 17, No. S1, 2023, pp. S18–S25
https://doi.org/10.14444/8503

Artificial Intelligence and Machine Learning Applications in Spine Surgery

NATHAN J. LEE, MD[1]; JOSEPH M. LOMBARDI, MD[1]; AND RONALD A. LEHMAN, MD[1]
[1]*Department of Orthopaedics, Columbia University Medical Center, The Och Spine Hospital at New York-Presbyterian, New York, NY, USA*

ABSTRACT

The complexity of patients with spine pathology and high rates of complications has driven extensive research directed toward optimizing outcomes and reducing complications. Traditional statistical analysis has been limited both in validity and in the number of predictor variables considered. Over the past decade, artificial intelligence and machine learning have taken center stage as the possible solution to creating more accurate and applicable patient-centered predictive models in spine surgery. This review discusses the current published machine learning applications on preoperative optimization, risk stratification, and predictive modeling for the cervical, lumbar, and adult spinal deformity populations.

Focus Issue Article

Keywords: machine learning, artificial intelligence, cervical spine, lumbar spine, adult spinal deformity, predictive model

INTRODUCTION

The development of predictive models is not a novel concept in spine surgery. For decades, surgeons have relied on various statistical analyses to identify risk factors for complications with the hope of creating a valid model. A popular technique is the use of multivariate logistic regression (LR), which produces odds ratios for independent variables on the outcomes of interest. The advantage of such analysis includes the relative ease of interpretation and application. However, an important limitation of predictive models is the limited number of predictive variables included. Furthermore, these traditional analyses are static in nature, assume a "linear" relationship between the input and output variables, and may have minimal applicability for addressing the intricacies of patient-specific needs as new data are introduced.

Over the past decade, health care providers have gained access to an immense amount of patient information through the digitation of electronic medical records. As a result, artificial intelligence (AI) and machine learning (ML) have taken center stage as the potential solution for implementing more accurate and generalizable predictive models. The major reasons for the increasing attraction toward AI and ML include the potential to process large amounts of data quickly, create models that adapt to new data, and understand complex, nonlinear relationships that conventional regression models might fail to comprehend. Spine studies are already showing promise in the ability of ML methods to provide improved preoperative risk stratification and diagnostics as well as leverage imaging data for better clinical prognostication.[1–3] The purpose of this review is to highlight the current applications of ML in spine surgery, compare the performance of common ML models, and explore the potential of ML in future studies.

WHAT IS ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING?

AI is the broader concept of applying systems to simulate human learning and thinking. One of the main applications of AI is ML, which utilizes various computational techniques to continuously learn and self-adjust from past data in order to determine mathematical relationships inherent in the data. The majority of prior outcomes research in spine have involved statistical analyses to characterize relationships between independent and dependent variables. By doing so, the focus of these statistical analyses has been to identify the parameters of a model and understand how each impacts the prediction. Although these analyses are valuable and often offer the researcher an ease of interpretability, they are static in nature and are often subject to selection bias and limited external validity. In contrast, the focus of ML is less about the parameters of a model and more about the prediction. This often leads to "black box" algorithms, which may be difficult to conceptualize. However, ML has the potential to process an inordinate amount of data, learn and adapt to varying

patient populations, and provide more powerful predictive capabilities.

In general, there are 3 main paradigms within ML: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is a more "hands-on" approach than other forms of ML. This approach relies on labeled datasets to "supervise" algorithms to classify and predict outcomes accurately. Tasks suited for this style of learning involve classification and regression. One potential drawback is that it may be time-consuming to train the data as it requires labels for each input and output, which may be extensive for large datasets. Common models used in supervised learning include decision trees (DTs), support vector machines (SVMs), and LR.

The main difference between supervised learning and unsupervised learning is the use of labeled datasets. Unsupervised learning trains algorithms based on unlabeled data without any guidance. This style of learning attempts to discover the inherent patterns of unlabeled data on its own. Tasks suited for this approach involve associations and clustering. Unsupervised learning may be used to cluster large unlabeled datasets based on their similarities and/or differences as well as process anomalies in visual or medical imaging data.

Similar to unsupervised learning, reinforcement learning does not involve external supervision. However, the goal of this approach involves a "trial-and-error" method of exploring new, unlabeled data to minimize cost or maximize reward parameters without any guidance. Over time, these algorithms will continue to explore new data and develop their own rules to maximize an outcome. Artificial neural networks (ANNs) often serve as the learning component in this approach.

As with any statistical analysis, it is important to make the distinction between correlation and causation in ML. A correlation is an association between variables. For instance, variable A may be associated with variable B. It is possible that variable A caused variable B or vice versa. However, there could be a third factor, variable C, which changes or confounds both variables A and B independently. Only after controlling for all confounders and random chance can it be assumed that a causal relationship exists between variables A and B. Although ML enables users to identify nuanced, complex relationships, these relationships are still correlations, not causations. For instance, ML may be trained on hundreds of x-ray images to detect a spinal deformity pattern; however, changes in pixel gradients or variances in patient positioning in x-ray images may introduce new variables that can influence the model. The assumption made with ML is that with a large enough training dataset, the ML will be able to encode all possible relationships into the model. One of the key factors behind the incremental gains in the accuracy of ML is the growing availability of data and stronger computer processing power. However, there may be practical limits to this. Currently, AI researchers argue for integrating a causal understanding into ML, which may require fewer training samples; however, the implementation of "causal ML" remains at the conceptual level.[4,5]

COMMON MACHINE LEARNING MODELS USED IN SPINE RESEARCH

DT and random forests (RFs) are examples of supervised learning, which uses a set of inputs and outputs to predict classification and regression problems such as readmission rates, patient-reported outcomes, and surgical complications. In short, a DT iteratively asks questions to partition data to reach the eventual outcome. An RF is an ensemble of many DTs. In comparison with a single DT, RF reduces overfitting and ultimately may improve accuracy.

SVMs are supervised learning algorithms that use decision boundaries or hyperplanes to help categorize 2 classes of data points and maximize the "distance" between the data of those 2 classes. SVMs may be useful in classifying images and conducting image segmentation analyses.

ANN is a commonly used ML tool that emulates the framework of the nervous system. ANN involves multiple layers with at least 1 hidden layer between the input and output. These hidden layers are interconnected by weighted linkages.

Convolutional neural networks (CNNs) are a subset of ANN that leverage a mathematical operation called convolution and are best suited for image, speech, and audio signal inputs. In short, CNNs take an input picture, assign relevance (eg, weights) to different aspects in the image through multiple layers, and distinguish them.

APPLICATIONS OF MACHINE LEARNING IN THE CERVICAL SPINE

A number of studies have sought to develop ML algorithms to predict outcomes and complications after cervical spine surgery.[6–8] One of the earliest applications of ML in cervical spine outcomes was introduced by Arvind et al in 2018, who examined more than 20,000 patients who underwent anterior cervical discectomy and fusion (ACDF) from the National Surgical Quality

Improvement Program (NSQIP) database and compared the performance of various ML models (ANN, LR, SVM, and RF) to predict surgical complications.[9] These authors found that ANN and LR outperformed the American Society of Anesthesiologists (ASA) physical status classification for every complication (area under curve [AUC] for cardiac, venous thromboembolism, wound complication, and mortality: ANN 0.772, 0.656, 0.518, 0.979; LR 0.759, 0.639, 0.501, 0.974; ASA 0.566, 0.397, 0.455, 0.346). Furthermore, ANN outperformed LR in predicting venous thromboembolism, wound complication, and mortality rate. In comparison, both the SVM and RF models were unable to perform better than random chance, which suggests that selecting the appropriate ML algorithm is an important factor to consider. As with any ML model, the performance hinges on the quality of the training data. Although NSQIP includes a large national sample of surgical patients, it lacks highly granular data pertinent to spine surgery. As a result, many surgery-specific variables, which may serve as stronger inputs, are not available to include in these models.

In 2019, one of the first studies to use ML to predict patient-reported outcomes after surgery for degenerative cervical myelopathy was performed by Merali et al.[10] These authors used a prospective, multicentered database to review 757 patients and compared multiple ML models (DT, RF, SVM, LR, and ANN) to predict patient-reported outcomes (Short Form-6D and modified Japanese Orthopaedic Association [mJOA]) after surgery. As measured by the AUC, the RF and SVM models outperformed LR and DT in up to 2-year follow-ups, which was attributed to the ability of these models to process more complex nonlinear and conditional relationships than either LR or DT models. The RF model found that the longer duration of preoperative symptoms, worse preoperative disease severity, older age, greater body weight, and current smoking status were associated with worse postoperative outcomes. Of note, this study did not include radiographic parameters in their models, which may have improved the performance. Furthermore, ANN performed poorly compared with the other models. This was likely due to the relatively limited number of training samples since ANN models are known to generally require larger training samples than SVM or RF for adequate training.

Wong et al used SVM to identify demographic, radiographic, and paraspinal muscle parameters that would predict proximal/distal adjacent segment disease after 2-level ACDF.[11] The SVM model achieved high accuracy (96.7%) and an AUC (0.97) for predicting adjacent segment disease. This study found that preoperative total cross-sectional area of cervical paraspinal muscle, relative fat composition, and asymmetry at C5 to C7 were predictive for early onset of adjacent segment disease.

Another application of ML in cervical spine has involved medical image analysis.[12] Computer vision techniques have shown the potential for a wide spectrum of medical applications and recently shown adoption within spine surgery. Huang et al reviewed over 300 images of 9 different ACDF systems from 5 different companies to identify the manufacturer and model of anterior cervical spinal hardware.[13] Knowing the manufacturer and model of prior cervical implants can facilitate faster and safer revision surgeries. These authors used a computer vision algorithm "bag of visual worlds" instead of other well-known techniques such as CNN because the primary task was the classification of a relatively sparse dataset. Images were subsequently classified by an SVM classifier. These authors achieved an accuracy of greater than 90% with a relatively small sample size, and this performance persisted for 1-level, 2-level, and 3-level plates.

Recently, ML applications in predicting the diagnosis and severity of cervical spondylotic myelopathy have shown promise. Hopkins et al used ANN on clinical and radiographic factors on 18 images to predict cervical spondylotic myelopathy as well as mJOA scores.[14] These authors achieved a median accuracy of 90% as well as mJOA scores within 0.4 points. This study was limited by the small sample size and lack of comparative partition analyses for training, testing, and validating, which are commonly used to optimize ML models.

APPLICATIONS OF MACHINE LEARNING IN THE LUMBAR SPINE

Several studies have used ML algorithms to predict complications after lumbar surgery. Kim et al used the NSQIP database to study more than 22,000 patients who underwent lumbar fusion surgery.[15] They trained and validated both ANN and LR models to predict postoperative complications (cardiac, wound, venous thromboembolism, and mortality). On the basis of AUC values, ANN and LR had comparable AUC values for predicting all complication types; however, ANN had greater sensitivity for detecting wound complications and mortality. The main difference between ANN and LR models is that ANN allows the characterization of nonlinear relationships in the data and has the potential

for slightly better accuracy for relatively rare complications, which is important for medical prognostication.[16,17]

Hopkins et al examined the efficacy of deep neural network to predict 30-day hospital readmission after posterior lumbar fusion based on more than 20,000 patients in the NSQIP database.[18] This model took advantage of the 177 unique input variables of the NSQIP database and fed data through a series of 7 layers, each with varying degrees of forward and backward communicating neurons and sporadic dropout layers to avoid overfitting. Ultimately, this model outperformed LR by achieving an AUC of 0.81 (vs 0.72).

Ogink et al used the NSQIP database to study patients with degenerative spondylolisthesis and compared 4 ML algorithms (DT, SVM, Bayes point machine, and ANN) in their ability to predict discharge placement after surgery.[19] The AUC was fairly similar among all 4 models (DT AUC = 0.733, SVM AUC = 0.742, ANN AUC = 0.755, and Bayes point machine AUC = 0.753). This study demonstrated the predictive capabilities of ML on discharge placement with good discrimination and performance. These authors have subsequently developed a web application for further clinical utility.

Kuris et al examined the records of more than 60,000 patients who underwent lumbar fusion from the NSQIP database and ANN to predict 30-day readmission.[20] The ANN algorithm was able to achieve an accuracy of 94.6% for anterior lumbar interbody fusion, 94% for posterior lumbar interbody fusion, and 92.6% for posterior spinal fusion with AUC values 0.64 to 0.65.

Harada et al performed a retrospective, single-center review of 2630 patients who underwent lumbar microdiscectomy to examine the utility of ML to predict the risk of recurrent herniated nucleus pulposus.[21] Input variables included patient demographics/comorbidities, clinical factors (eg, number of herniated levels, type of herniation, duration of symptoms, motor examination, surgical approach), patient-reported outcomes, and radiographic factors. These authors used an ML algorithm called "extreme gradient boosting classification (XGBoost classification)," which involves both linear model and tree learning algorithms. This model demonstrated excellent model discrimination with an accuracy 0.7, an AUC 0.72, and a Brier score of 0.21.

To the authors' knowledge, Roller et al were the first to use ML to predict the level of lumbar spinal decompression. These authors used CNN to predict the lumbar decompression level based on magnetic resonance images of 141 patients.[22] This algorithm was able to predict with an accuracy of 65%, with higher mean scores for L3-S1 surgical levels.

MACHINE LEARNING APPLICATIONS IN ADULT SPINAL DEFORMITY

Postoperative Complications

Adult spinal deformity (ASD) is known to be associated with a high complication rate, and numerous studies have sought to characterize the risk profile for this population. Until recently, a common limitation has been a lack of patient-specific, predictive models to account for this diverse patient population and wide-ranging potential risk factors. A commonly used statistical analysis has been LR because it is simple, easy to interpret, and readily transparent. Over the past few years, numerous studies have explored more robust ML algorithms in ASD surgery. Using a prospective multicenter database of 557 patients, Scheer et al used an ensemble of DTs to predict which patients sustained at least 1 major intraoperative or perioperative (6 weeks) complication after ASD surgery.[23] In comparison with other multicenter registries, this database included a comprehensive set of patient factors, surgical variables, implant characteristics, radiographic measures, and patient-reported outcomes. The overall model accuracy was 87.6% with an AUC of 0.89. According to the authors, a DT framework was used due to its ability to process a large number of input variables (both categorical and continuous), ease of construction, feasibility with missing data, and deal with potential nonlinear relationships in the data. Furthermore, an ensemble of 5 DTs was used to increase accuracy but at the cost of decreasing interpretability. A limitation of this study was combining both intra- and perioperative complications as a single outcome, which likely was due to the sample size (148 patients with at least 1 complication). Nevertheless, this study was one of the first to demonstrate the potential role of predictive analytics in ASD outcomes research.

Yagi et al similarly created a predictive model (ensemble of 5 DTs) for complications of 195 ASD patients using extensive demographic, surgical, and radiographic data.[24] The overall accuracy of their model was 92.5% with an AUC of 0.963, with an 84% accuracy in the external validation. Compared with Scheer et al, these authors focused on predicting postoperative complications occurring within 2 years after surgery in older patients (age ≥50 years) as well as the inclusion of patient frailty, which is a known risk factor for major complications.[25]

Jain et al leveraged the State Inpatient Database to study a larger population of 37,852 patients who underwent long-segment lumbar posterior spine fusion.[26] RF, LR, and elastic net regression models were used to predict discharge disposition, 90-day readmission, and 90-day major medical complications after surgery. Interestingly, the more traditionally used LR model appeared to outperform the other 2 ML algorithms for discharge to facility (AUC: LR 0.77 vs RF 0.75 vs elastic net 0.76), 90-day readmission (AUC: LR 0.65 vs RF/elastic net 0.63), and 90-day medical complications (AUC: LR 0.7 vs RF/elastic net 0.68). However, these differences are likely not clinically relevant. Therefore, this study's findings should not be taken to imply that LR is a superior ML algorithm. Readers should be aware that given the linear relationship between the independent variables and the log-odds of the outcome, the results are often more easily interpretable with LR. In comparison, RF is a nonparametric ML algorithm that is capable of accounting for nonlinear relationships, yet the results have limited inferential capabilities.

Similar to Jain et al study, Kim et al used a national database to study the 30-day postoperative complications (cardiac, wound, venous thromboembolism, and mortality) of patients undergoing elective ASD surgery.[27] Both ANN and LR models outperformed the ASA scoring system in predicting every complication. Furthermore, these authors found that ANN outperformed LR for all complications except venous thromboembolism ($P < 0.05$). Durand et al used the NSQIP database on 1029 patients with ASD and compared both DT and RF models to predict intra-/postoperative blood transfusion.[28] The RF achieved a higher AUC (0.85 vs 0.79), but this difference was not statistically significant ($P = 0.155$). These findings highlight that none of the models performs the best on all datasets and that it is advisable to include multiple ML techniques for comparative purposes. Although national registries (the State Inpatient Database and NSQIP) provide large patient samples, these databases are deidentified and limited by the lack of radiographic data and surgery-specific variables, which may influence model performance.

Other Surgery-Specific Complications

Recent studies have used ML to predict other surgery-specific complications, including proximal junctional kyphosis (PJK), proximal junctional failure (PJF), and pseudarthrosis after ASD surgery. In 2016, Scheer et al used an ensemble of DT on 510 patients to predict PJK and PJF.[29] Their model accuracy was 86.3%, with an AUC of 0.89. They found that the strongest predictors were age, lower instrumented level, preoperative sagittal vertical axis, upper instrumented vertebra implant type, upper instrumented vertebra, preoperative pelvic tilt, and preoperative pelvic incidence–lumbar lordosis (PI-LL) mismatch. However, a limitation of this study was combining both PJK and PJF as 1 outcome rather than as 2 complications, which may be managed differently. In contrast, Yagi et al performed an ensemble of DT models to predict PJF alone on 145 patients as well as included bone mineral density as an input feature among several other patient, radiographic, and surgical factors.[30] They reported an accuracy of 98% with an AUC of 1.0. In another study, Scheer et al used DT to predict pseudarthrosis with 2-year follow-up based on an array of demographic, radiographic, and surgical factors.[31] Their model achieved a 91.3% accuracy with an AUC of 0.94. To our knowledge, this was the first study to use ML to predict pseudarthrosis after ASD surgery.

Patient-Reported Outcomes

Another popular application of ML has been predicting the minimum clinically important difference patient-reported outcome following ASD surgery. Scheer et al applied DT models on 198 patients and achieved an accuracy of 86% with an AUC of 0.94.[32] Interestingly, the top predictors were not surgical parameters but clinical and radiographic factors (gender, sagittal vertical axis, PI-LL mismatch, T1 spinopelvic inclination angle, ASA, T1 pelvic angle, Scoliosis Research Society [SRS] pain, and SRS total). The following year, Ames et al augmented this analysis by performing a retrospective analysis of prospectively collected, multicenter data on 570 patients to predict the likelihood of reaching a minimum clinically important difference in patient-reported outcomes at 1- and 2-year postoperative follow-up.[33] Multiple ML algorithms were used (including ordinary least squares, ordinary least squares with partitions, elastic net, gradient boosting machines, XGBoost tree, XGBoost linear, RF, and generalized linear modeling), and the composite of these models created a prediction tool for each patient. Model performance was evaluated using the mean absolute error (MAE), and the

final model was selected based on the minimization of MAE and goodness of fit using R^2. The MAE ranged from 8% to 15%, which demonstrated successful model performance.

Clinical Decision-Making

Other applications of ML involve models to support clinical decision-making. Durand et al used various ML models to predict whether patients with ASD were managed operatively vs nonoperatively.[34] More than 1500 patients were followed up to 1 year after their baseline visit. RF, SVM, and elastic net regression models were used and compared against LR. The SVM (AUC 0.914) and elastic net (AUC 0.913) models had excellent discrimination compared with LR (AUC 0.896) and RF (AUC 0.830). The SVM model showed 86% accuracy. Ultimately, this study showed that the shared decision-making process between operative and nonoperative management could be computationally predicted with excellent performance given the complexity of factors involved in such decisions. However, it is important to note that these models predicted patients who underwent surgery and not necessarily who "should" undergo surgery.

Another approach to augment preoperative clinical decision-making is leveraging unsupervised ML algorithms. One of the challenges in providing patient-specific management is examining the hundreds of possible data points (eg, demographics, comorbidities, and radiographic and patient-reported outcomes) for any given patient in the preoperative setting. Although classification systems exist, such as the SRS-Schwab classification, these often focus on radiographic parameters and fail to include other potentially important variables.[35] To address this, Ames et al developed a phylogenic dendrogram based on hierarchical clustering of patient parameters.[36] This resulted in 3 clusters: (1) "young coronal," which included the youngest patients (mean age 47.6 years) with a deformity predominantly with scoliosis (mean Cobb angle 50.4°), (2) "old revision," which involved relatively older patients (mean age 62.3 years) with a high incidence of prior surgery (48%), and (3) "old primary," which included older patients (mean age 61 years) with a low incidence of prior spine surgery (7%). The same was performed for surgical parameters, which had 4 distinct groups: (1) 3-column osteotomy patients, (2) no osteotomy and no interbody fusion, (3) interbody fusion, and (4) Smith-Peterson osteotomy. In comparison with prior ML applications in predictive models for complications, this study illustrated the potential for broader applications of ML. Through this novel application of an unsupervised hierarchical clustering model, Ames et al demonstrated the potential of ML to sift through the complex features of ASD and define relevant patient clusters in a purely data-driven approach.

FUTURE DIRECTIONS OF ARTIFICIAL INTELLIGENCE IN SPINE SURGERY

The current literature on ML in spine surgery is promising as it has already demonstrated wide-ranging applications for the cervical spine, the lumbar spine, and ASD. Other possible ML applications include integration in the preoperative planning phase and with other existing technologies, such as robot-assisted surgery and/or augmented reality systems. A variety of surgical options exist for ASD (eg, 3-column osteotomies, level selection, and interbody fusion), each with their own potential risks and benefits. The determination of the optimal surgery for each individual patient is often at the discretion of the surgeon. Furthermore, predicting postoperative alignment and specifically reducing the risk of PJK/PJF remain a significant challenge for ASD surgery. ML has the potential to augment surgeon decision-making and improve surgical outcomes. This will hinge on continuing to build an ensemble of models based on high-quality and robust databases as well as further validation and ultimate consolidation of published models to better integrate in clinical practice. It is important to keep in mind, however, that the goal of AI and ML is not to replace but to complement the surgeon in providing safer and more efficient patient-centered care.

REFERENCES

1. Hornung AL, Hornung CM, Mallow GM, et al. Artificial intelligence in spine care: current applications and future utility. *Eur Spine J*. 2022;31(8):2057–2081. doi:10.1007/s00586-022-07176-0

2. Stephens ME, O'Neal CM, Westrup AM, et al. Utility of machine learning algorithms in degenerative cervical and lumbar spine disease: a systematic review. *Neurosurg Rev*. 2022;45(2):965–978. doi:10.1007/s10143-021-01624-z

3. Joshi RS, Lau D, Scheer JK, et al. State-of-the-art reviews predictive modeling in adult spinal deformity: applications of advanced analytics. *Spine Deform*. 2021;9(5):1223–1239. doi:10.1007/s43390-021-00360-0

4. Schölkopf B, Locatello F, Bauer S, et al. *Towards Causal Representation Learning*. 2021.

5. Sanchez P, Voisey JP, Xia T, Watson HI, O'Neil AQ, Tsaftaris SA. Causal machine learning for healthcare and precision medicine. *R Soc Open Sci*. 2022;9(8):220638. doi:10.1098/rsos.220638

6. 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. *Spine J*. 2021;21(10):1659–1669. doi:10.1016/j.spinee.2020.02.003

7. Khan O, Badhiwala JH, Akbar MA, Fehlings MG. Prediction of worse functional status after surgery for degenerative cervical myelopathy: a machine learning approach. *Neurosurgery*. 2021;88(3):584–591. doi:10.1093/neuros/nyaa477

8. Liew BXW, Peolsson A, Rugamer D, et al. Clinical predictive modelling of post-surgical recovery in individuals with cervical radiculopathy: a machine learning approach. *Sci Rep*. 2020;10(1):16782. doi:10.1038/s41598-020-73740-7

9. 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. doi:10.14245/ns.1836248.124

10. Merali ZG, Witiw CD, Badhiwala JH, Wilson JR, Fehlings MG. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PLoS One*. 2019;14(4):e0215133. doi:10.1371/journal.pone.0215133

11. Wong AYL, Harada G, Lee R, et al. Preoperative paraspinal neck muscle characteristics predict early onset adjacent segment degeneration in anterior cervical fusion patients: a machine-learning modeling analysis. *J Orthop Res*. 2021;39(8):1732–1744. doi:10.1002/jor.24829

12. Schwartz JT, Valliani AA, Arvind V, et al. Identification of anterior cervical spinal instrumentation using a smartphone application powered by machine learning. *Spine (Phila Pa 1976)*. 2022;47(9):E407–E414. doi:10.1097/BRS.0000000000004172

13. Huang KT, Silva MA, See AP, et al. A computer vision approach to identifying the manufacturer and model of anterior cervical spinal hardware. *J Neurosurg Spine*. 2019:1–7. doi:10.3171/2019.6.SPINE19463

14. Hopkins BS, Weber KA, Kesavabhotla K, Paliwal M, Cantrell DR, Smith ZA. Machine learning for the prediction of cervical spondylotic myelopathy: a post hoc pilot study of 28 participants. *World Neurosurg*. 2019;127:e436–e442. doi:10.1016/j.wneu.2019.03.165

15. 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. doi:10.1097/BRS.0000000000002442

16. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol*. 1996;49(11):1225–1231. doi:10.1016/s0895-4356(96)00002-9

17. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform*. 2002;35(5–6):352–359. doi:10.1016/s1532-0464(03)00034-0

18. 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. *J Neurosurg Spine*. 2019:1–8. doi:10.3171/2019.9.SPINE19860

19. Ogink PT, Karhade AV, Thio QCBS, et al. Development of a machine learning algorithm predicting discharge placement after surgery for spondylolisthesis. *Eur Spine J*. 2019;28(8):1775–1782. doi:10.1007/s00586-019-05936-z

20. Kuris EO, Veeramani A, McDonald CL, et al. Predicting readmission after anterior, posterior, and posterior interbody lumbar spinal fusion: a neural network machine learning approach. *World Neurosurg*. 2021;151:e19–e27. doi:10.1016/j.wneu.2021.02.114

21. Harada GK, Siyaji ZK, Mallow GM, et al. Artificial intelligence predicts disk re-herniation following lumbar microdiscectomy: development of the "RAD" risk profile. *Eur Spine J*. 2021;30(8):2167–2175. doi:10.1007/s00586-021-06866-5

22. Roller BL, Boutin RD, O'Gara TJ, et al. Accurate prediction of lumbar microdecompression level with an automated MRI grading system. *Skeletal Radiol*. 2021;50(1):69–78. doi:10.1007/s00256-020-03505-w

23. 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. doi:10.3171/2016.10.SPINE16197

24. Yagi M, Hosogane N, Fujita N, et al. Predictive model for major complications 2 years after corrective spine surgery for adult spinal deformity. *Eur Spine J*. 2019;28(1):180–187. doi:10.1007/s00586-018-5816-5

25. Leven DM, Lee NJ, Kothari P, et al. Frailty index is a significant predictor of complications and mortality after surgery for adult spinal deformity. *Spine (Phila Pa 1976)*. 2016;41(23):E1394–E1401. doi:10.1097/BRS.0000000000001886

26. Jain D, Durand W, Burch S, Daniels A, Berven S. Machine learning for predictive modeling of 90-day readmission, major medical complication, and discharge to a facility in patients undergoing long segment posterior lumbar spine fusion. *Spine (Phila Pa 1976)*. 2020;45(16):1151–1160. doi:10.1097/BRS.0000000000003475

27. 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. doi:10.1016/j.jspd.2018.03.003

28. 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. doi:10.1097/BRS.0000000000002515

29. Scheer JK, Osorio JA, Smith JS, et al. Development of validated computer-based preoperative predictive model for proximal junction failure (pjf) or clinically significant pjk with 86 % accuracy based on 510 ASD patients with 2-year follow-up. *Spine (Phila Pa 1976)*. 2016;41(22):E1328–E1335. doi:10.1097/BRS.0000000000001598

30. Yagi M, Fujita N, Okada E, et al. Fine-Tuning the predictive model for proximal junctional failure in surgically treated patients with adult spinal deformity. *Spine (Phila Pa 1976)*. 2018;43(11):767–773. doi:10.1097/BRS.0000000000002415

31. Scheer JK, Oh T, Smith JS, et al. Development of a validated computer-based preoperative predictive model for pseudarthrosis with 91 % accuracy in 336 adult spinal deformity patients. *Neurosurg Focus*. 2018;45(5):2018.8.FOCUS18246. doi:10.3171/2018.8.FOCUS18246

32. Scheer JK, Osorio JA, Smith JS, et al. Development of a preoperative predictive model for reaching the oswestry disability index minimal clinically important difference for adult spinal deformity patients. *Spine Deform*. 2018;6(5):593–599. doi:10.1016/j.jspd.2018.02.010

33. Ames CP, Smith JS, Pellisé F, et al. Development of deployable predictive models for minimal clinically important difference achievement across the commonly used health-related quality of life instruments in adult spinal deformity surgery. *Spine (Phila Pa 1976)*. 2019;44(16):1144–1153. doi:10.1097/BRS.0000000000003031

34. Durand WM, Daniels AH, Hamilton DK, et al. Artificial intelligence models predict operative versus nonoperative management of patients with adult spinal deformity with 86 % accuracy. *World Neurosurgery*. 2020;141:e239–e253. doi:10.1016/j.wneu.2020.05.099

35. Schwab F, Ungar B, Blondel B, et al. Scoliosis research society-schwab adult spinal deformity classification: a validation study. *Spine (Phila Pa 1976)*. 2012;37(12):1077–1082. doi:10.1097/BRS.0b013e31823e15e2

36. Ames CP, Smith JS, Pellisé 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. doi:10.1097/BRS.0000000000002974

Funding: The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests: The authors report no conflicts of interest in this work.

Disclosures: Nathan John Lee has nothing to report. Joseph M. Lombardi reports consulting fees from Medtronic and Stryker and stock/stock options from OnPoint Surgical. Ronald A. Lehman Jr reports grants/contracts from the Department of Defense and royalties/licenses and patents from Medtronic and Stryker.

Corresponding Author: Nathan J. Lee, Columbia University Medical Center, 161 Fort Washington Ave, New York, NY 10032, USA; njl2116@cumc.columbia.edu

Published 12 May 2023