Title

Development and External Validation of a Risk Calculator for Prediction of Major Complications and Readmission After Anterior Cervical Discectomy and Fusion

Journal
Spine, 48(7)

ISSN
0362-2436

Authors
Shah, Akash A
Devana, Sai K
Lee, Changhee
et al.

Publication Date
2023-04-01

DOI
10.1097/brs.0000000000004531

Peer reviewed

HHS Public Access
Author manuscript
Spine (Phila Pa 1976). Author manuscript; available in PMC 2024 April 01.

Published in final edited form as:
Spine (Phila Pa 1976). 2023 April 01; 48(7): 460–467. doi:10.1097/BRS.0000000000004531.

Development and external validation of a risk calculator for prediction of major complications and readmission after anterior cervical discectomy and fusion

Akash A. Shah[1,†], **Sai K. Devana**[1], **Changhee Lee**[2], **Thomas E. Olson**[1], **Alexander Upfill-Brown**[1], **William L. Sheppard**[1], **Elizabeth L. Lord**[1], **Arya N. Shamie**[1], **Mihaela van der Schaar**[3,4], **Nelson F. SooHoo**[1], **Don Y. Park**[1]

[1]Department of Orthopaedic Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA

[2]Department of Artificial Intelligence, Chung-Ang University, Seoul, South Korea

[3]Department of Electrical and Computer Engineering, University of California, Los Angeles, CA

[4]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom

Abstract

Study Design: Retrospective, case-control study

Objective: We aim to build a risk calculator predicting major perioperative complications after anterior cervical fusion. Additionally, we aim to externally validate this calculator with an institutional cohort of patients who underwent anterior cervical discectomy and fusion (ACDF).

Summary of Background Data: The average age and proportion of patients with at least one comorbidity undergoing ACDF have increased in recent years. Given the increased morbidity and cost associated with perioperative complications and unplanned readmission, accurate risk stratification of patients undergoing ACDF is of great clinical utility.

Methods: This is a retrospective cohort study of adults who underwent anterior cervical fusion at any non-federal California hospital between 2015–2017. The primary outcome was major perioperative complication or 30-day readmission. We built standard and ensemble machine learning models for risk prediction, assessing discrimination and calibration. The best-performing model was validated on an external cohort comprised of consecutive adult patients who underwent ACDF at our institution between 2013–2020.

Results: A total of 23,184 patients were included in this study; there were 1,886 cases of major complication or readmissions. The ensemble model was well-calibrated and demonstrated an area under the receiver operating characteristic curve (AUROC) of 0.728. The variables most important for the ensemble model include male sex, medical comorbidities, history of complications, and

[†] Corresponding author: Akash A. Shah, Department of Orthopaedic Surgery, David Geffen School of Medicine at UCLA, 10833 Le Conte Avenue 76-116 CHS, Los Angeles, CA 90095, AAShah@mednet.ucla.edu.

teaching hospital status. The ensemble model was evaluated on the validation cohort (n=260) with an AUROC of 0.802. The ensemble algorithm was used to build a web-based risk calculator.

Conclusion: We report derivation and external validation of an ensemble algorithm for prediction of major perioperative complications and 30-day readmission after anterior cervical fusion. This model has excellent discrimination and is well-calibrated when tested on a contemporaneous external cohort of ACDF cases.

Abstract

Accurate risk stratification of patients undergoing ACDF is of great clinical utility. With a cohort of 23,184 patients, we derive an ensemble machine learning algorithm for prediction of major perioperative complications and unplanned readmission after ACDF. When externally validated on an institutional cohort, the model displays excellent discrimination (AUROC 0.802).

Introduction

Anterior cervical fusion procedures such as anterior cervical discectomy and fusion (ACDF) have grown in popularity due to excellent clinical outcomes, low risk of perioperative complications, and reduced length of hospitalization.[1–4] The number of anterior cervical fusion procedures is expected to grow rapidly for all indications.[5,6] The average age and percentage of patients with at least one medical comorbidity undergoing ACDF have similarly increased.[1] Elderly patients with an increased comorbidity burden are at elevated risk for perioperative complications.

Given the increased morbidity and cost associated with perioperative complications and unplanned readmissions, it would be of utility to predict which patients are likely to suffer complications. Accurate prediction of a patient’s complication risk would allow for appropriate pre-operative counseling and risk stratification. Unfortunately, prediction of complication risk after ACDF with conventional methods is challenging given the low incidence of perioperative complications. Due to their ability to detect complex non-linear relationships, machine learning (ML) methods have been increasingly applied to spinal surgery in recent years.[7] ML models have been built for prediction of outcomes after treatment for spinal pathology including degenerative disease, infection, and malignancy.[8–14] ML studies for prediction of outcomes after ACDF remain scarce.[15,16]

We primarily aim to build an ensemble ML model for prediction of major perioperative complications or unplanned readmission after anterior cervical fusion. We hypothesize that the ensemble model will identify risk factors for complications after anterior cervical fusion. To determine the generalizability of the developed model, we aim to externally validate the model on an institutional cohort of patients who underwent ACDF. Finally, we aim to use the validated ensemble model to build a risk calculator to predict patient-specific risk for adverse outcomes after ACDF.

Methods

Study design and subjects

This study is a retrospective review of patients using the California Office of Statewide Health and Planning and Development (OSHPD) Patient Discharge Database (PDD), a statewide discharge database containing admissions data for all non-federal hospital admissions in California. Complications and future admissions for patients in this database can be tracked regardless of whether the complication or readmission occurred at a different hospital from where the index procedure was performed – so long as the hospital is a non-federal facility in California. To build the derivation cohort, we included patients ≥18 years who underwent anterior cervical fusion between 2015 and 2017 using International Classification of Diseases, Tenth Revision (ICD-10) procedure codes for this procedure.

Our institution is a non-federal hospital within California. In order to avoid any overlap between the derivation cohort and the institutional validation cohort, we only included a case in the validation cohort if it did not meet criteria for inclusion in the OSHPD PDD. We included all adult patients who underwent ACDF at our academic medical center comprised of two tertiary care hospitals between 2013–2015 and 2017–2020. Additionally, we included patients who underwent ACDF between 2015–2017 only if they were not admitted (i.e. same-day surgery). The cohort was identified by querying surgeon schedules and operative notes. Our institutional review board approved a waiver of consent for this retrospective study.

Outcome and other variables

The primary outcome measure was any major complication or 30-day readmission after index anterior cervical fusion. Complications were identified by adapting the total joint arthroplasty ICD-10 coding algorithm developed by the Centers for Medicare and Medicaid (CMS).[17,18] These complications include acute myocardial infarction, pneumonia, sepsis, pulmonary embolism, surgical site bleeding, and wound infection. Myocardial infarction, pneumonia, and sepsis must occur during the index admission or within seven days of start of index admission. Pulmonary embolism must occur during the index admission or within 30 days of admission. Surgical site bleeding and wound infection must occur during the index admission or within 90 days. Readmissions must occur within 30 days of discharge to be included.[18]

Explanatory features collected for the cohort include patient demographic characteristics and patient medical comorbidities using the CMS Condition Categories as defined by the CMS Hierarchical Condition Category (HCC) risk adjustment model.

Model development

The derivation cohort was divided into a training cohort (comprising of 80% of the study population) and a hold-out testing cohort (comprising of 20% of the study population). An ML-based risk prediction model for major complications after ACDF was developed using AutoPrognosis, which employs a Bayesian optimization algorithm to generate a model as a weighted ensemble of ML pipelines. Each pipeline comprises design choices for classification methods.[19] To train the ensemble model, we conduct 100 iterations of the Bayesian optimization. During each iteration, the algorithm explores a new pipeline of classification methods and their corresponding hyperparameters. Five-fold stratified cross-validation was performed within the training set to evaluate performance of the ensemble pipeline.

We built five standard ML models spanning different classes of ML modeling approaches: logistic regression, gradient boosting, XGBoost, AdaBoost, and random forest. Logistic regression is a linear classifier and random forest is a tree-based ensemble classifier; AdaBoost, gradient boosting, and XGBoost are all boosting ensemble classifiers.[20–23] We implemented logistic regression, random forest, AdaBoost, and gradient boosting using the *scikit-learn* Python library.[24] XGBoost was built using the *xgboost* Python library.[23] The hyperparameters of each benchmark model were selected via grid search. For random forest, AdaBoost, and gradient boosting models, the number and trees are chosen from the set {50, 100, 200, 300}. For XGBoost, the number of trees and maximum depth of each tree were selected from sets {50, 100, 200, 300} and {2, 3, 4, 5}, respectively.

Performance metrics

Within the derivation cohort, we employed five-fold stratified cross-validation – where the study population is split into a training cohort and hold-out testing cohort – to evaluate discrimination and calibration for prognostic models.

Discrimination determines how well a model distinguishes patients who developed a complication or readmission from those who did not. We assessed discrimination with the area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC). AUROC represents the probability that a model can distinguish between patients who developed a complication and those who did not; a value of 0.5 indicates random prediction and a value of 1 indicates perfect discrimination.[25,26] AUPRC may be assessed when analyzing an imbalanced dataset – one in which negative cases far outnumber positive cases. Constructed by plotting positive predictive value (precision) versus the sensitivity (recall), the precision-recall curve depicts a model's ability to correctly identify positive cases.[27,28] Random prediction will result in the baseline AUPRC value, which is the proportion of true positive cases in the cohort. The higher the AUPRC is compared to the random prediction value, the better the model identifies positive cases.

Calibration is a measure of the agreement between predicted risk and the observed outcome proportion in model's predictions and the observed outcomes in the study population. The calibration slope is a measure of prediction spread by the model, where a slope of 1 is consistent with perfect spread. A calibration intercept close to 0 indicates minimal overestimation or underestimation of an outcome by the model.[29,30] The Brier score – the mean squared error between the observed values and the predicted probabilities – is a measure of both discrimination and calibration. A Brier score close to zero indicates low deviation of the model's predicted values from the observed probability and therefore a more accurate model.[31] The Brier score for the null model was calculated; the null model assigns a probability for all patients equivalent to the prevalence of major complication in the cohort.[12]

Feature importance and validation

We utilize a partial dependence function to measure the importance of an individual feature to model performance by assessing the average effect in predicted risks when its value is altered.[22] The features found to be most important for model performance were collected for each patient in the institutional validation cohort. The ensemble algorithm was then tested on the validation cohort. Discrimination and calibration of the model on the validation cohort were determined. Discrimination was assessed with AUROC and calibration was assessed with the Brier score.

Risk calculator

The best-performing model on the testing cohort was deployed as a web-based application with the open-access platform Heroku (San Francisco, CA).

Results

Baseline cohort demographics

A total of 23,184 patients met inclusion criteria for the derivation cohort. The median age was 58 years; just over half of the patients (50.9%) were female. A plurality of patients (39.8%) was privately insured and 34.4% were insured through Medicare. The most common medical comorbidity present in the cohort was diabetes mellitus (4.5%), followed by coronary atherosclerosis (3.2%) and chronic obstructive pulmonary disease (3.1%). Four-hundred and ninety-five patients (2.1%) had prior medical complications during hospitalizations, 491 (2.1%) had a history of cardiorespiratory complications, and 458 (2.0%) suffered prior implant complication. Four hundred and sixty-nine patients (2.0%) had protein-calorie malnutrition and 422 (1.8%) had dementia. Four hundred and thirty-three patients (1.0%) had prior musculoskeletal infection and 395 (1.7%) had metastatic cancer or leukemia. A complete description of the cohort demographics is provided in Table 1. There were 1,886 patients (8.1%) who suffered a major complication or readmission. The most common complications observed were pneumonia, sepsis, pulmonary embolism, and acute myocardial infarction (Table 2).

We performed a univariate analysis to determine whether patients who were treated at a teaching institution had significantly different baseline characteristics compared to those treated at a non-teaching institution. Patients treated at teaching hospitals were significantly more likely to have the following comorbidities: diabetes mellitus, malignancy, severe chronic kidney disease coronary atherosclerosis, angina pectoris, protein-calorie malnutrition, and major depression and/or bipolar disorder (Supplementary Table 1).

The validation cohort was comprised of 260 patients who underwent ACDF. There were 10 patients (3.8%) with a major complication or 30-day readmission.

Model performances

The ensemble model was built with a weighted ensemble of eight ML pipelines (Supplementary Table 2). This ensemble model demonstrates an AUROC of 0.728 ± 0.011. It is well-calibrated with a calibration slope of 1.399 and calibration intercept of −0.007. The Brier score of the ensemble model is 0.071, compared to 0.074 for the null model. The receiver-operating characteristic curve for the ensemble model is depicted in Figure 1. The AUPRC of the ensemble model is 0.273 ± 0.021; a random classifier would return an AUPRC of 0.081 (Table 3). The precision-recall curve for the ensemble model is shown in Figure 2.

Relative feature importance

The importance of each explanatory feature to model performance for the ensemble model is displayed in Table 4. The features important for risk prediction in the ensemble model include: male sex, musculoskeletal infection, implant complication, malnutrition, history of medical complications, cardio-respiratory failure, teaching hospital status, metastatic cancer or leukemia, chronic obstructive pulmonary disease (COPD).

Validation performance

The most important features for ensemble model performance were collected for the validation cohort (Table 5). The ensemble model was tested on this cohort, resulting in an AUROC of 0.802 and AUPRC of 0.259. The Brier score was 0.054 (Table 6). The receiver-operating characteristic curve for the validation cohort is depicted in Figure 3.

Risk calculator

The ensemble model was used to build a web-based risk calculator that is available for clinicians. The web application can be accessed at: https://risk-calculator-anterior-cervc.herokuapp.com/. Users may input values for each explanatory feature and observe the updated risk for major complication or 30-day readmission.

Discussion

The age and comorbidity burden of patients undergoing ACDF has increased with the significant uptick in utilization of ACDF; many of these patients are at elevated risk of perioperative complications with significant associated cost and morbidity.[1,32,33] Accurate pre-operative assessment of a patient's risk of developing major perioperative complications with unplanned readmission is thus of great utility. ML methods have been increasingly employed in spinal surgery for prediction of outcomes such as surgical site infection, discharge disposition, treatment failure for spinal infections, and mortality in metastatic disease.[9,12,13,34] Most risk prediction tools have been developed with multivariable logistic regression; ML has been employed sparingly for evaluation of outcomes after anterior cervical fusion. Arvind and colleagues employed an artificial neural network to predict risk of specific complications and mortality after ACDF – identifying increased age, diabetes, and tobacco use as predictive features for cardiac complications. Their model performed well for prediction of cardiac complications and mortality; however, it performed poorly for prediction of venous thromboembolism and wound complications.[15] ML has also been used to stratify patients for inpatient versus outpatient ACDF based on outpatient surgery risk prediction.[16] To our knowledge, no externally validated ML models for outcomes after ACDF have been reported. The lack of external validation remains a pervasive problem in the ML literature, limiting the generalizability of these models.[35]

With a cohort of 23,184 patients, we report the development of a well-calibrated ML-based ensemble model that predicts major perioperative complications and unplanned readmission after anterior cervical fusion. With an AUROC of 0.728 and an AURPC of 0.273, this model demonstrates moderate discrimination. Additionally, we externally validate this model with a contemporaneous institutional cohort of patient who underwent ACDF. The model performs excellently on the validation cohort with an AUROC of 0.802.

The most important feature for model performance is patient sex. Patient sex has been previously shown to be predictive of major complications and readmission in those undergoing anterior cervical fusion.[36–38] We also show that comorbidities such as COPD, metastatic cancer, and protein-calorie malnutrition are important contributors to the accuracy of our ensemble model. In patients undergoing anterior cervical fusion, COPD has been implicated as a predictor of unplanned readmission and post-operative complications such as pneumonia.[39–41] This may be due to the dissection of the anterior cervical approach and the necessary retraction of the airway and throat structures during anterior cervical surgery, which may contribute to airway swelling, dysphagia, and possible aspiration events that may trigger COPD exacerbations and pneumonia. A regression analysis on elective cervical and lumbar spinal surgery showed that disseminated cancer is associated with post-operative complication.[42] Optimization of pre-operative nutritional status may reduce risk of surgical site infection after posterior cervical fusion.[43] Neither metastatic cancer nor malnutrition have been previously linked to poor outcomes in anterior cervical fusion specifically, however. Age is an important feature to model performance, consistent with multiple studies that have shown increasing age as predictive of major complications and unplanned readmission after anterior cervical fusion.[15,36,39,40,44,45]

We find that a history of implant complication is important to model performance. Implant complications may be associated with pseudarthrosis leading to screw breakage, screws backing out of the plate, and failure of the construct leading to recurrent or worsened symptoms. Revision cervical spinal surgery after implant failure is technically challenging due to dissection through scar tissue and limited bone stock; it is associated with a significantly higher risk of perioperative complications compared to the index surgery.[46] Additionally, history of complications in past hospital admissions (e.g. excessive transfusion requirement, wound infection) is important to model performance. This is an intuitive finding suggesting that past hospitalization-associated complications are likely correlated with future complications.

Teaching hospital status is an important feature for ensemble model performance. Patients undergoing cervical spine surgery at a teaching hospital are more likely to have increased length of stay and mortality than those at non-teaching hospitals. In addition, resident involvement in cervical fusion has been shown to be associated with increased blood transfusion, pulmonary complication, and increased operative time.[47,48] The medical comorbidities of patients and complexity of cases at academic medical centers are often greater than those at non-teaching hospitals, potentially increasing the likelihood of complication. Indeed, we found that patients treated at a teaching hospital in our cohort were more likely to have medical comorbidities than those not treated at a teaching facility.

This study has limitations, first of which is its retrospective design. The use of a de-identified state database without access to underlying patient records limits the features and outcomes that can be extracted. The OSHPD PDD does not provide access to important risk factors such as pre-operative functional status, surgical technique, number of levels treated, staged surgery, and frailty index. Reliance on ICD-10 diagnosis codes to assign complications may underestimate complication rates compared to chart review. Specific laboratory values such as serum albumin are not available for manual search, requiring reliance on CMS-HCC Condition Categories that are comprised of groups of ICD-9/10 codes to determine presence of comorbidities such as malnutrition. This database does not contain data on patient-reported functional outcomes, neurologic complications, or specific implant-related complications. Additionally, the PDD does not allow for collection of data on specific major complications that are associated with ACDF such as dysphagia and C5 nerve root palsy. Modeling outcomes based on procedure type alone is a limitation given the heterogeneous indications for ACDF procedures. Larger multi-institutional cohort studies would allow for more granular analysis that can predict outcomes for specific surgical indications and further evaluate the generalizability of this model. Finally, advanced machine learning models sacrifice interpretability for improved predictive performance; while we can determine the relative importance of a feature for model performance, this analysis is unable to quantify the effect size of a given feature. The importance of a feature is not necessarily reflective of the feature’s effect size. The importance of a given feature to model performance may be due interactions with other features or how it separates data.

With a cohort of 23,184 patients, we have developed an ensemble ML algorithm that predicts major perioperative complications and readmission after anterior cervical fusion. This model is well-calibrated and identifies features important for prediction. Additionally, this model is externally valid when on tested on a contemporaneous institutional cohort. By providing accurate prognostic information, this tool may facilitate improved pre-operative shared decision-making and appropriate patient selection. We also identify potentially modifiable risk factors for complications such as malnutrition and COPD; pre-operative optimization of these patient features may minimize the risk of developing complications after ACDF. To encourage direct use of this algorithm by healthcare providers, we incorporated this model into a web-based risk calculator. This tool may help define the perioperative risk profile of an individual patient, which may be beneficial when determining whether an ACDF should be performed at an ambulatory surgery center versus an inpatient facility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

The research reported in this publication was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under the Ruth L. Kirschstein National Research Service Award Number T32AR059033 (AAS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. CL is supported through the IITP grant (MSIT 2021–0-01341, AI Graduate School Program, CAU)

References

1. Memtsoudis S, Hughes A, Ma Y, Chiu Y, Sama A, Girardi F. Increased in-hospital complications after primary posterior versus primary anterior cervical fusion. Clin Orthop Relat Res. 2011;469:649–657. [PubMed: 20838946]
2. Badhiwala J, Ellenbogen Y, Khan O, et al. Comparison of the inpatient complications and health care costs of anterior versus posterior cervical decompression and fusion in patients with multilevel degenerative cervical myelopathy: a retrospective propensity score-matched analysis. World Neurosurg. 2020;134:e112–e119. [PubMed: 31574327]
3. Passias P, Marascalchi B, Boniello A, et al. Cervical spondylotic myelopathy: national trends in the treatment and peri-operative outcomes over 10 years. J Clin Neurosci. 2017;42:75–80. [PubMed: 28476459]
4. Marawar S, Girardi F, Sama A, et al. National trends in anterior cervical fusion procedures. Spine (Phila Pa 1976). 2010;35(15):1454–1459. [PubMed: 20216341]
5. Wang M, Kreuter W, Wolfla C, Maiman D, Deyo R. Trends and variations in cervical spine surgery in the United States. Spine (Phila Pa 1976). 2009;34(9):955–961. [PubMed: 19352223]
6. Goz V, Weinreb J, McCarthy I, Schwab F, Lafage V, Errico T. Perioperative complications and mortality after spinal fusions: analysis of trends and risk factors. Spine (Phila Pa 1976). 2013;38(22):1970–1976. [PubMed: 23928714]
7. Martin B, Bono C. Artificial intelligence and spine: rise of the machines. Spine J. 2021;21(10):1604–1605. [PubMed: 34146706]
8. Hopkins B, Yamaguchi J, 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. 2020;32:399–406.
9. Hopkins B, Mazmudar A, Driscoll C, Svet M, Goergene J, Kelsten M. Using artificial intelligence (AI) to predict postoperative surgical site infection: a retrospective cohort of 4046 posterior spinal fusions. Clin Neurol Neurosurg. 2020;192:105718. [PubMed: 32065943]
10. Thio QCBS, Karhade AV, Ogink PT, et al. Can machine-learning techniques be used for 5-year survival prediction of patients with chondrosarcoma? Clin Orthop Relat Res. 2018;476(10):2040–2048. doi:10.1097/CORR.0000000000000433 [PubMed: 30179954]
11. Kim J, Merrill R, 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. [PubMed: 29016439]
12. Karhade A, Thio Q, Ogink P, et al. Development of machine learning algorithms for prediction of 30-day mortality after surgery for spinal metastasis. Neurosurgery. 2019;85(1):E83–E91. doi:10.1016/j.wneu.2018.07.276 [PubMed: 30476188]
13. Shah A, Karhade A, Bono C, Harris M, Nelson S, Schwab J. Development of a machine learning algorithm for prediction of failure of nonoperative management in spinal epidural abscess. Spine J. 2019;19(10):1657–1665. [PubMed: 31059819]
14. Shah A, Devana S, Lee C, et al. Machine learning-driven identification of novel patient factors for prediction of major complications after posterior cervical spinal fusion. Eur Spine J. 2021:Online ahead of print.
15. Arvind V, Kim J, Oermann E, Kaji D, Cho S. Predicting surgical complications in adult patients undergoing anterior cervical discectomy and fusion using machine learning. Neurospine. 2018;15(4):329–337. [PubMed: 30554505]

16. Wang K, Suresh K, Puvanesarajah V, Read M, Margalit A, Jain A. Using predictive modeling and machine learning to identify patients appropriate for outpatient ACDF. Spine (Phila Pa 1976). 2021;46(10):665–670. [PubMed: 33306613]

17. Yale New Haven Health Services Corporation / Center for Outcomes Research & Evaluation. 2017 Procedure-specific measure updates and specifications report hospital-level risk-standardized complication measure: elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) - version 6.0. 2017.

18. Shah A, Devana S, Lee C, Kianian R, van der Schaar M, SooHoo N. Development of a novel, potentially universal machine learning algorithm for prediction of complications after total hip arthroplasty. J Arthroplasty. 2021;36(5):1655–1662. [PubMed: 33478891]

19. Alaa A, van der Schaar M. AutoPrognosis: automated clinical prognostic modeling via Bayesian optimization with structured kernel learning. Proc 35th Int Conf Mach Learn PMLR. 2018;80:139–148.

20. Breiman L Random forests. Mach Learn. 2001;45:5–32.

21. Ratsch G, Onoda T, Muller K. Soft margins for AdaBoost. Mach Learn. 2001;42:287–320.

22. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189–1232. doi:10.1214/aos/1013203451

23. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. KDD ‘16 Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min. 2016:785–794.

24. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12(85):2825–2830.

25. Harris AHS, Kuo AC, Bozic KJ, et al. American joint replacement registry risk calculator does not predict 90-day mortality in veterans undergoing total joint replacement. Clin Orthop Relat Res. 2018;476(9):1869–1875. doi:10.1097/CORR.0000000000000377 [PubMed: 30113939]

26. Fischer J, Bachmann L, Jaeschke R. A readers’ guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Med. 2003;29:1043–1051. [PubMed: 12734652]

27. Ozenne B, Subtil F, Maucort-Boulch D. The precision-recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. J Clin Epidemiol. 2015;68:855–859. [PubMed: 25881487]

28. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10(3):e0118432. [PubMed: 25738806]

29. Steyerberg E, van Calster B, Pencina M. Performance measures for prediction models and markers: evaluation of predictions and classifications. Rev Esp Cariol. 2011;64(9):788–794.

30. Steyerberg E, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925–1931. [PubMed: 24898551]

31. Manning DW, Edelstein AI, Alvi HM. Risk prediction tools for hip and knee arthroplasty. J Am Acad Orthop Surg. 2016;24(1):19–27. doi:10.5435/JAAOS-D-15-00072 [PubMed: 26604220]

32. Zileli M, Dursun E. How to improve outcomes of spine surgery in geriatric patients. World Neurosurg. 2020;140:519–526. [PubMed: 32797984]

33. Goyal A, Bhandarkar A, Alvi M, et al. Cost of readmissions following anterior cervical discectomy and fusion: insights from the Nationwide Readmissions Database. Neurosurgery. 2020;87(4):679–688. [PubMed: 31642499]

34. Ogink P, Karhade A, 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. [PubMed: 30941521]

35. Alsoof D, McDonald C, Kuris E, Daniels A. Machine learning for the orthopaedic surgeon: uses and limitations. J Bone Jt Surg Am. 2022:Online ahead of print.

36. Gruskay J, Fu M, Basques B, et al. Factors affecting length of stay and complications after elective anterior cervical discectomy and fusion. Clin Spine Surg. 2016;29(1):E34–E42. [PubMed: 24525748]

37. Shah N, Jain I, Moattari C, et al. Comparing predictors of complications after anterior cervical diskectomy and fusion, total disk arthroplasty, and combined anterior cervical diskectomy and

fusion-total disk arthroplasty with a minimum 2-year follow-up. J Am Acad Orthop Surg. 2020;28(17):e759–e765. [PubMed: 31860582]

38. Schafer E, Bazydlo M, Schultz L, et al. Rates and risk factors associated with 90-day readmission following cervical spine fusion surgery: analysis of the Michigan Spine Surgery Improvement Collaborative (MSSIC) registry. Spine J. 2020;20(5):708–716. [PubMed: 31958576]
39. Bohl D, Ahn J, Rossi V, Tabaraee E, Grauer J, Singh K. Incidence and risk factors for pneumonia following anterior cervical decompression and fusion procedures: an ACS-NSQIP study. Spine J. 2016;16(3):335–342. [PubMed: 26616171]
40. Katz A, Mancini N, Karukonda T, Cote M, Moss I. Comparative and predictor analysis of 30-day readmission, reoperation, and morbidity in patients undergoing multilevel ACDF versus single and multilevel ACCF using the ACS-NSQIP dataset. Spine (Phila Pa 1976). 2019;44(23):E1379–E1387. [PubMed: 31348176]
41. Malik A, Jain N, Kim J, Khan S, Yu E. Chronic obstructive pulmonary disease is an independent predictor for 30-day complications and readmissions following 1- to 2- level anterior cervical discectomy and fusion. Glob Spine J. 2019;9(3):298–302.
42. Broda A, Sanford Z, Turcotte J, Patton C. Development of a risk prediction model with improved clinical utility in elective cervical and lumbar spine surgery. Spine (Phila Pa 1976). 2020;45(9):E542–E551. [PubMed: 31770338]
43. Zreik J, Goyal A, Alvi M, Freedman B, Bydon M. Utility of preoperative laboratory testing in assessing risk of adverse outcomes after anterior cervical discectomy and fusion: insights from National Surgical Registry. World Neurosurg. 2020;136:e398–e406. [PubMed: 31931235]
44. Veeravagu A, Connolly I, Lamsam L, et al. Surgical outcomes of cervical spondylotic myelopathy: an analysis of a national, administrative, longitudinal database. Neurosurg Focus. 2016;40(6):E11.
45. Kurian S, Wahood W, Yolcu Y, et al. Risk factors for unplanned readmissions following anterior cervical discectomy and fusion and posterior lumbar fusion procedures: comparison of two national databases. World Neurosurg. 2020;143:e613–e630. [PubMed: 32781151]
46. Shimokawa N, Sato H, Matsumoto H, Takami T. Complex revision surgery for cervical deformity or implant failure. Neurospine. 2020;17(3):543–553. [PubMed: 33022159]
47. Lee N, Kothari P, Kim C, et al. The impact of resident involvement in elective posterior cervical fusion. Spine (Phila Pa 1976). 2018;43(5):316–323. [PubMed: 26839988]
48. Kim R, Garcia R, Smith Z, Dahdaleh N. Impact of resident participation on outcomes after single-level anterior cervical diskectomy and fusion: an analysis of 3265 patients from the American College of Surgeons National Surgical Quality Improvement Program Database. Spine (Phila Pa 1976). 2016;41(5):E289–96. [PubMed: 26555830]

- With a cohort of 23,184 patients, we develop a well-calibrated novel ensemble machine learning algorithm for prediction of major perioperative complications and 30-day readmission after anterior cervical fusion.
- When externally validated on a contemporaneous institutional cohort who underwent ACDF, the ensemble algorithm displays excellent discrimination with an AUROC of 0.802.
- We built a web-based risk calculator with the ensemble model: https://risk-calculator-anterior-cervc.herokuapp.com/. Users input values for each explanatory feature and observe the updated risk for major complication or 30-day readmission.
- This externally validated tool may facilitate improved pre-operative risk stratification, determining the risk profile of an individual patient as well as guiding whether a planned ACDF should be performed in an ambulatory surgery center or the hospital setting.

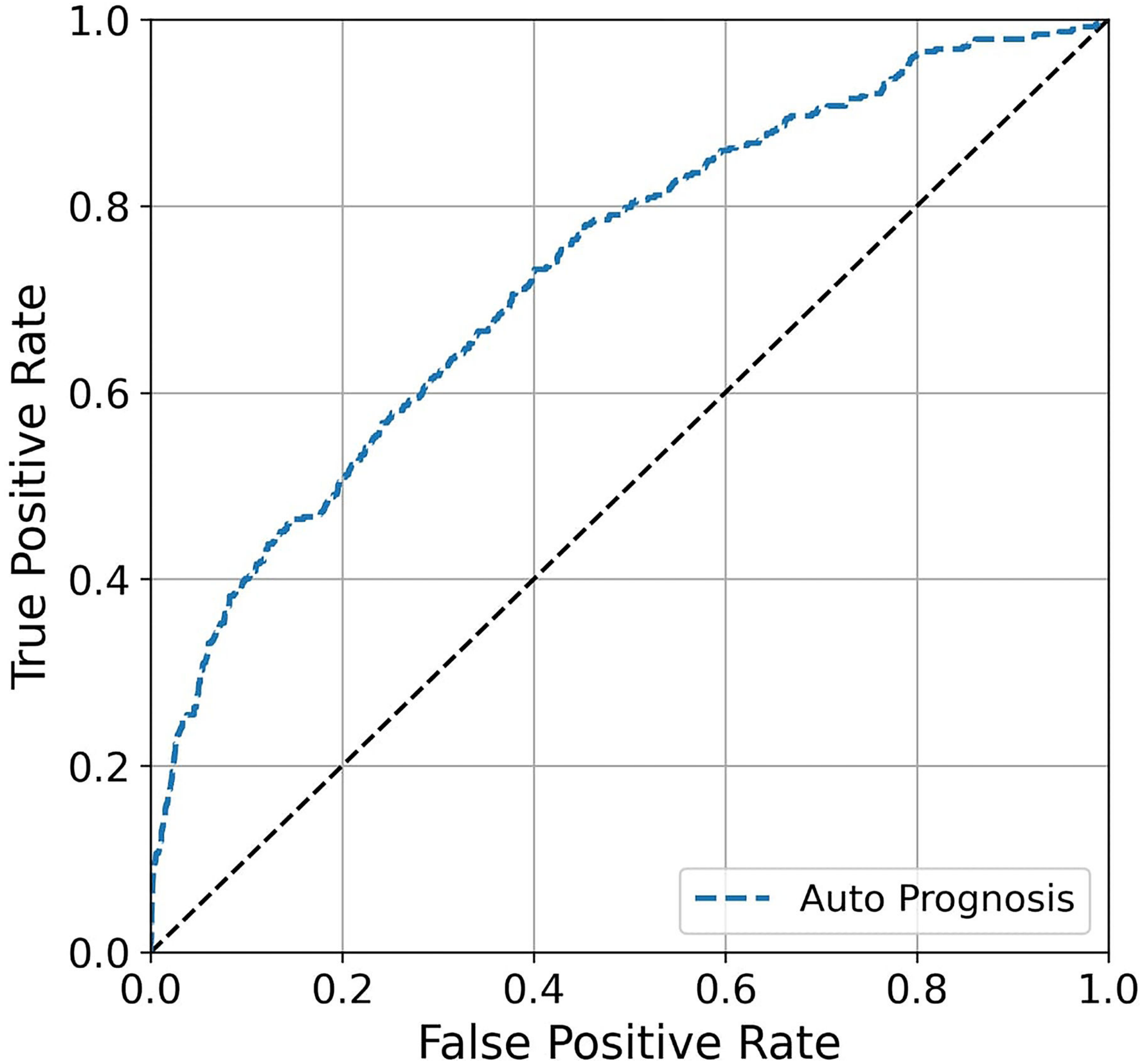

Figure 1:
Receiver-operating characteristic curve for the ensemble algorithm (training set)

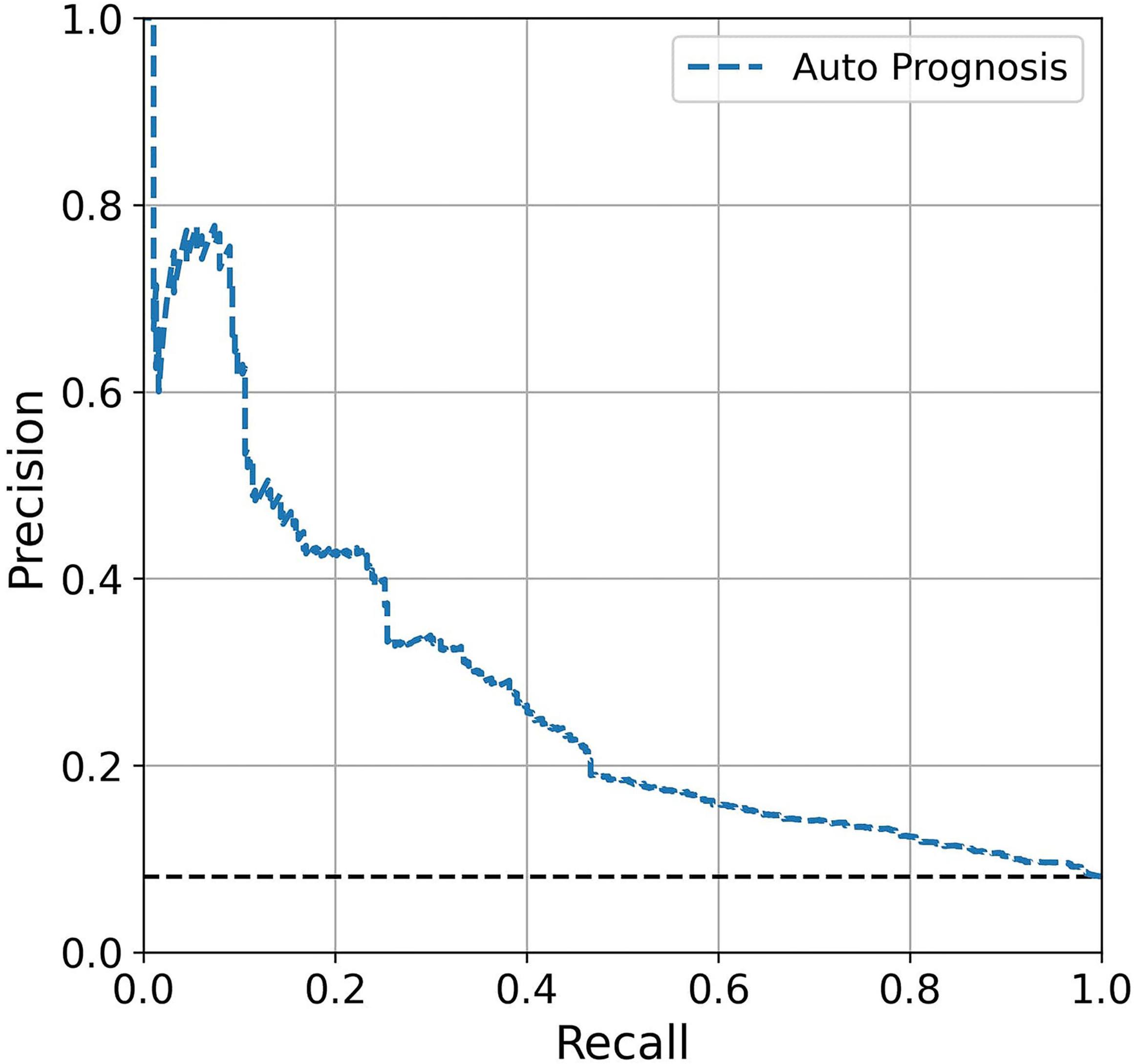

Figure 2:
Precision-recall curve for the ensemble algorithm (training set)

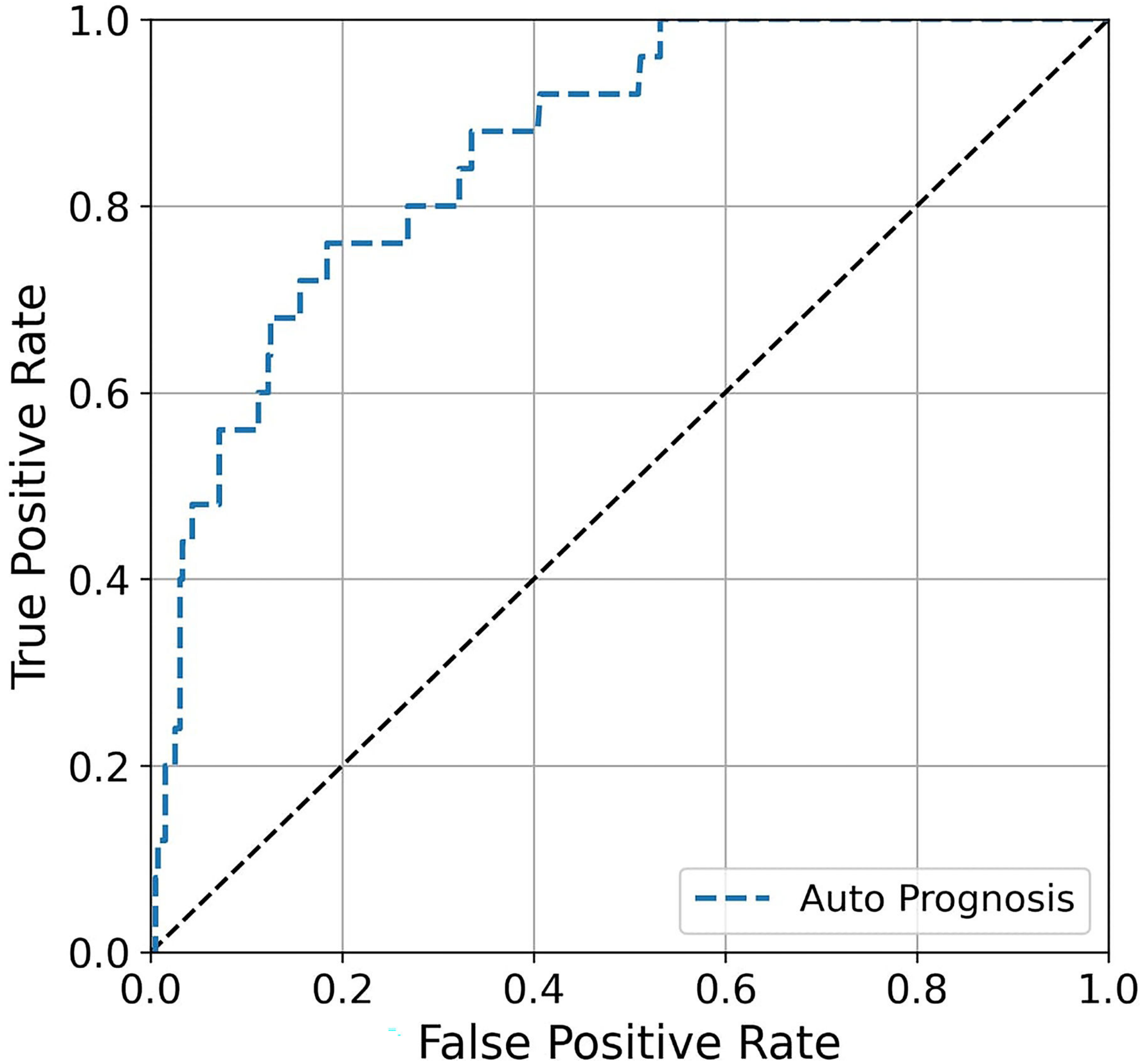

Figure 3:
Receiver-operating characteristic curve for the ensemble algorithm (validation set)

Table 1.

Derivation cohort characteristics

Variable	All Patients (n = 23,184)
	Median (IQR)
Age (years)	58 (51 – 67)
Hospital volume[†]	276 (143 – 406)
	Number (%)
Male	11,383 (49.1)
Race	
White	17,824 (76.9)
Black	1,545 (6.7)
Asian / Pacific Islander	1,356 (5.9)
Native American	113 (0.5)
Other	2,130 (9.2)
Unknown	216 (0.9)
Insurance	
Private	9,226 (39.8)
Medicare	7,967 (34.4)
Medi-Cal	3,258 (14.1)
Workers’ compensation	2,037 (8.8)
Other	696 (3.0)
Procedure performed at teaching hospital	4,283 (18.5)
Medical comorbidities	
Diabetes mellitus	1,045 (4.5)
Coronary atherosclerosis	743 (3.2)
Chronic kidney disease	658 (2.8)
COPD	722 (3.1)
Metastatic cancer or acute leukemia	395 (1.7)
Morbid obesity	532 (2.3)
Protein-calorie malnutrition	469 (2.0)
Paraplegia	373 (1.6)
Quadriplegia	412 (1.8)
Dementia	422 (1.8)
Schizophrenia	407 (1.8)
Major depressive or bipolar disorder	589 (2.5)
Stroke	383 (1.7)
Vertebral fractures without spinal cord injury	501 (2.2)
Spinal cord injury	666 (2.9)
Cardiorespiratory failure or shock	491 (2.1)
Implant complication	458 (2.0)
Complications of medical care	495 (2.1)

IQR = Interquartile range; COPD = chronic obstructive pulmonary disease

†Cases of cervical fusions performed between 2015 and 2017

Table 2.

Major complications and 30-day readmission for derivation cohort

Major complications	All Patients (n = 23,184)
	Number (%)
At least one complication or readmission	1,886 (8.1)
Pneumonia	581 (2.5)
Sepsis	348 (1.5)
Pulmonary embolism	95 (0.4)
Acute myocardial infarction	67 (0.3)
Surgical site bleeding or wound infection	39 (0.2)

Table 3.

Model performance for ML and logistic regression models

Model	AUROC	AUPRC	Brier score
Ensemble	**0.728 ± 0.011**	**0.273 ± 0.021**	0.071 ± 0.006
XGBoost	0.723 ± 0.019	0.261 ± 0.024	0.068 ± 0.001
Gradient boosting	0.722 ± 0.015	0.258 ± 0.021	0.068 ± 0.001
AdaBoost	0.719 ± 0.013	0.256 ± 0.018	0.246 ± 0.002
Logistic regression	0.717 ± 0.012	0.248 ± 0.021	0.069 ± 0.001
Random forest	0.668 ± 0.016	0.162 ± 0.012	0.080 ± 0.001

Table 4.

Relative feature importance for the ensemble model predicting complications and 30-day readmission after anterior cervical fusion

Feature	Rank
Male sex	1
Musculoskeletal infection	2
Implant complication	3
Protein-calorie malnutrition	4
Complication of medical care	5
Cardio-respiratory failure	6
Teaching hospital	7
Metastatic cancer of leukemia	8
COPD	9
Dementia	10

Table 5.

Validation ACDF cohort characteristics

Feature	All Patients (n = 260)
	Median (IQR)
Age (years)	54 (46 – 65)
	Number (%)
Male	148 (56.9)
Implant complication	14 (5.3)
Quadriplegia	4 (1.5)
Spinal cord injury	75 (28.8)
Dialysis	1 (0.4)
Chronic kidney disease	16 (6.2)
Prior complication	4 (1.5)
Metastatic cancer or leukemia	2 (0.8)
Cardio-respiratory failure	8 (3.1)
COPD	6 (2.3)
Bacterial pneumonia	4 (1.5)
Stroke	1 (0.4)
Vascular disease	12 (4.6)
Coronary atherosclerosis	7 (2.7)
Other circulatory disease	8 (3.1)
Major depression or bipolar disorder	77 (29.6)
Procedure performed at teaching hospital	260 (100)
Insurance	
Private	204 (78.5)
Medicare	68 (26.2)
Medi-Cal	17 (6.5)
Other	12 (4.6)
Major perioperative complication or readmission	10 (3.8)

IQR = Interquartile range

Table 6.

Derived model performance on validation ACDF cohort

Model	AUROC	AUPRC	Brier score
Ensemble	**0.802**	**0.259**	0.054
Gradient boosting	0.718	0.128	0.094
AdaBoost	0.715	0.123	0.214
Random forest	0.638	0.128	0.080
Logistic regression	0.557	0.115	0.038
XGBoost	0.520	0.109	0.049