Original Article

Corresponding Author

Samuel K. Cho

https://orcid.org/0000-0001-7511-2486

Department of Orthopaedic Surgery,
Icahn School of Medicine at Mount Sinai,
5 East 98th Street, Box 1188, New York,
NY 10029, USA
Tel: +1-212-241-0276
Fax: +1-646-537-8531
E-mail: samuel.cho@mountsinai.org

Predicting Surgical Complications in Adult Patients Undergoing Anterior Cervical Discectomy and Fusion Using Machine Learning

Varun Arvind[1], Jun S. Kim[1], Eric K. Oermann[2], Deepak Kaji[1], Samuel K. Cho[1]

[1]Department of Orthopaedic Surgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[2]Department of Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Objective: Machine learning algorithms excel at leveraging big data to identify complex patterns that can be used to aid in clinical decision-making. The objective of this study is to demonstrate the performance of machine learning models in predicting postoperative complications following anterior cervical discectomy and fusion (ACDF).

Methods: Artificial neural network (ANN), logistic regression (LR), support vector machine (SVM), and random forest decision tree (RF) models were trained on a multicenter data set of patients undergoing ACDF to predict surgical complications based on readily available patient data. Following training, these models were compared to the predictive capability of American Society of Anesthesiologists (ASA) physical status classification.

Results: A total of 20,879 patients were identified as having undergone ACDF. Following exclusion criteria, patients were divided into 14,615 patients for training and 6,264 for testing data sets. ANN and LR consistently outperformed ASA physical status classification in predicting every complication (p < 0.05). The ANN outperformed LR in predicting venous thromboembolism, wound complication, and mortality (p < 0.05). The SVM and RF models were no better than random chance at predicting any of the postoperative complications (p < 0.05).

Conclusion: ANN and LR algorithms outperform ASA physical status classification for predicting individual postoperative complications. Additionally, neural networks have greater sensitivity than LR when predicting mortality and wound complications. With the growing size of medical data, the training of machine learning on these large datasets promises to improve risk prognostication, with the ability of continuously learning making them excellent tools in complex clinical scenarios.

Keywords: Anterior cervical discectomy and fusion, Predicting, Complications, Machine Learning, Spine

INTRODUCTION

With the advent of digital technology, machine learning and deep learning in particular, is increasingly making it possible to utilize big data to more precisely risk stratify and prognosticate how an individual patient will behave based on a given a disease or intervention. Machine learning has already been used in other realms such as retail and search engines. However, healthcare has lagged in the uptake of newer techniques to leverage the rich information contained in electronic health records.

The practice of evidence-based medicine has sustained the progress seen in modern care and diagnosis. Traditional statistical approaches have gleaned much about what is known re-

garding risk factors used for prognostication. Machine learning combines these fundamental statistical insights with modern high-performance computing to learn patterns that can be used for recognition and prediction. Importantly, machine learning often identifies patterns that are not readily apparent to human intuition, thus identifying otherwise unknown connections.[1] Multivariate logistic regression (LR) and artificial neural network (ANN) are the 2 most commonly used machine learning models employed in medicine.[2] ANNs were first developed to model the neural architecture of the brain. Harnessing the structure of biology, ANNs are particularly well suited for modeling complex, nonlinear data when little is known regarding the underlying distribution of the data or colinearity among the variables.[3] Importantly, ANNs can perform these functions without prior assumptions, leading to a highly adaptable system with little bias.[3] Other models such as support vector machines (SVMs) and random forest decision trees (RFs) have been used for classification tasks.

Anterior cervical discectomy and fusion is a commonly performed procedure with excellent and reliable outcomes and a fast recovery.[4-6] The number of ACDF procedures performed in the United States has increased almost 8 fold from 1990 to 2004, accounting for the majority of outpatient cervical spine surgeries.[7] In particular, because of the good outcomes associated with ACDF, complications are difficult to predict. Thus, novel tools that can help predict potential postoperative complications are needed. Furthermore, in the burgeoning era of rising healthcare costs and greater scrutiny over surgical outcomes, there has been increasing emphasis on understanding the risk factors and possible predictors to optimize perioperative planning and management. Data-driven clinical decision support tools have the potential to lead to cost savings by leveraging the information contained in large medical databases. Uptake of machine learning approaches in this realm has lagged due to the sparse data sets associated with ACDF.[8]

This study seeks to develop and validate machine-learning algorithms to precisely predict complications following ACDF using a national database. These algorithms have the capability of continuously "learning" using newly generated information to improve the quality and efficiency of care.

MATERIALS AND METHODS

1. Patient Selection and Preprocessing

The National Surgical Quality Improvement Program (NSQIP) database was used for the purpose of training and validating ANN and LR models. A total of 20,879 patients undergoing anterior cervical discectomy and fusion from the years 2010 through 2014 were reviewed for this study. Patients were excluded from the present analysis due to incomplete data. No other exclusion criteria were employed.

2. Training and Testing Data Sets

For development of our models, 70% of the initial data (training set) was used for training while 30% (training set) was set aside randomly for posttraining evaluation of our models. To overcome the low sample size for positive complication cohorts, the adaptive synthetic sampling (ADASYN) approach for imbalanced learning was utilized to generate positive complications to improve class imbalance. Briefly, ADASYN utilizes a weighted distribution for minority class examples that are difficult to learn, and generates synthetic data based on these examples to improve model learning and generalizability.[9]

This study was reviewed by the Icahn School of Medicine at Mount Sinai Institutional Review Board (IRB) and was deemed appropriate for exemption from IRB oversight as data was supplied from a deidentified national database.

3. Feature Selection

Input features used for training include sex, age, ethnicity (White, Black, Hispanic, or other), history of diabetes, history of smoking, steroid use, history of bleeding disorders, functional status, American Society of Anesthesiologists (ASA) physical status classification ≥III, body mass index (BMI), and presence of pulmonary or cardiac comorbidities. In machine learning applications, the number of training examples required to reach a given accuracy grows exponentially with the number of irrelevant features.[10] To combat this feature selection was performed to prevent overfitting and improve overall generalizability of our models. LR analysis was performed on the training data set, to obtain probability coefficients for each feature. The top 6 features identified as having the greatest regression coefficient magnitudes were chosen as input variables for all machine learning models. Wound complication was defined as superficial or deep surgical site infection, organ space infection or wound dehiscence. Cardiac complication was defined as cardiac arrest requiring cardiopulmonary resuscitation or myocardial infarction. Age and BMI were treated as continuous variables, and all other features were treated categorically.

4. Machine Learning Construction and Testing

Machine learning models were trained to predict occurrence

of mortality, venous thromboembolism (VTE), wound complications, and cardiac complications. ANNs were constructed using the Neural Network toolbox in MatLab 2016b (MathWorks, Inc., Natick, MA, USA). L2 regularization was used to combat ANN overfitting, by augmenting the error function used for training with the squared magnitude of the weights used in the ANN. This prevents overly complex models that are overfitted to a specific dataset, improving predictive generalizability. Multiple ANNs were created by partitioning the majority class (no complication) into subsets in a 1:1 ratio with the minority class (positive complication), generating ANNs trained off of each partition. Subsequently, each ANN was trained in a 5-fold cross validation scheme. Testing data was used for final test of the ANN to provide an unbiased assessment of ANN performance. Final predictions were based off of individual accuracy-weighted predictions surveyed across each ANN. LR, SVM, and RF model was trained and tested on the same data that the ANN was evaluated on. Furthermore, these machine learning models

were compared to the ASA physical status classification system. Classification performance for ANN, LR, SVM, RF, and ASA was evaluated based on area under the receiver operating characteristic curve (AUC) with reported 95% confidence intervals (CIs).

RESULTS

1. Data and Analysis Pipeline

A total of 20,879 patients were identified as having undergone ACDF surgery between 2011 and 2014. Male patients were 48.4%, while female were 51.6%. The mean age was 53.2 years old and the cohort exhibited low rates of complications across all outcomes. Among this cohort, 14,615 patients (70%) were included into the training set and 6,264 patients (30%) were used as a hold out training set for evaluating the trained machine learning models (Fig. 1). Our study uses cardiac complications, VTE, wound complications, and mortality as target outcomes. Spe-
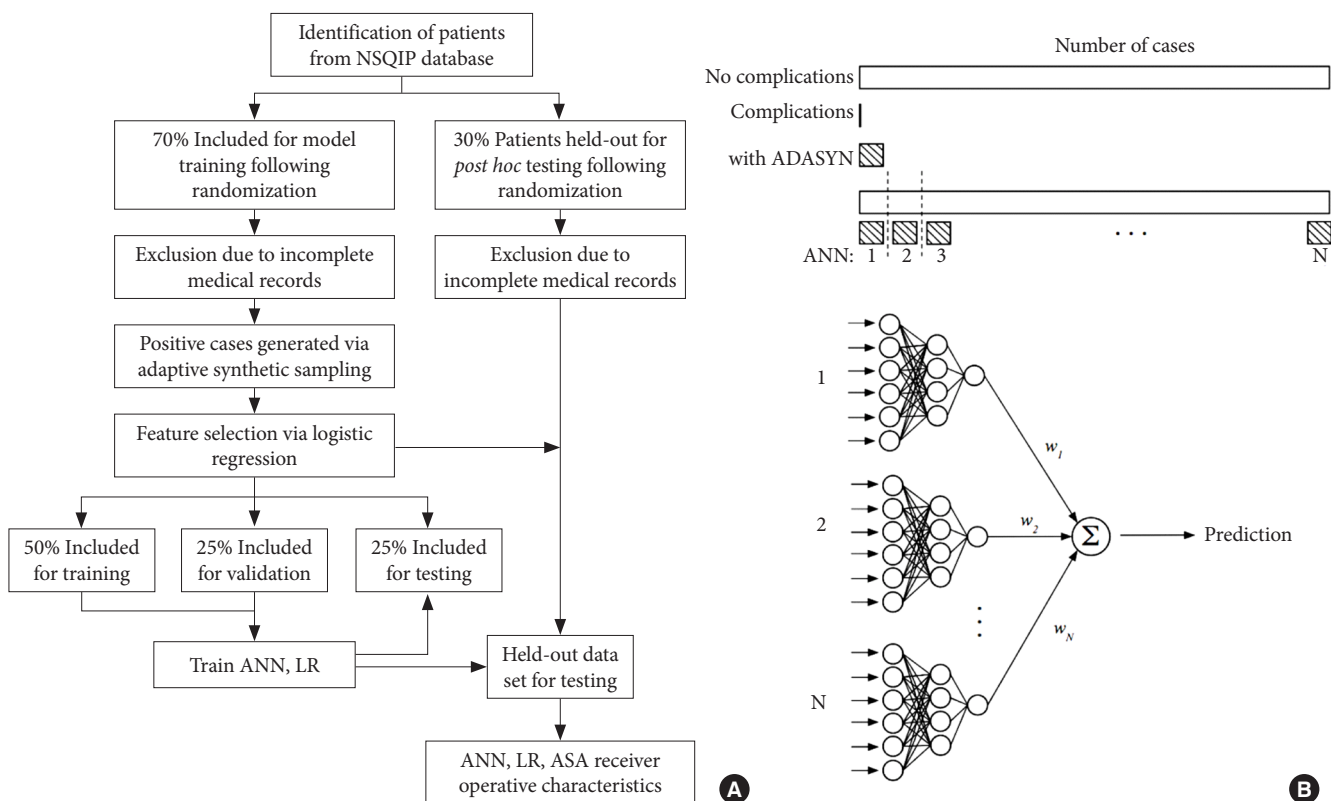
Fig. 1. (A) Schematic of study workflow. (B) Diagram of ANN model. Bar lengths represent number of patient cases. ADASYN increases the number of positive cases to combat class imbalance. Negative cases are then partitioned in a 1:1 ratio with the positive cases to create a class-balanced dataset used for ANN training. Each partition trains an independent neural net. During evaluation, data is fed through each neural net where the responses are surveyed, weighted by the model's accuracy, and the net prediction is used. NSQIP, National Surgical Quality Improvement Program; ANN, artificial neural network; ADASYN, adaptive synthetic sampling; LR, logistic regression; ASA, American Society of Anesthesiologists.

cifically, 0.1% rate of mortality, 0.5% rate of wound complication, 0.3% rate of VTE, 0.2% rate of cardiac complications (Table 1). Additionally, there were low rates of overlap between complications except between mortality and cardiac complications.

Of those patients with cardiac comorbidities, 47.8% did not survive. Unsurprisingly, age was a highly predictive feature across all outcomes. Diabetic status and tobacco usage were also useful features, which is consistent with their known association with

Table 1. Patient characteristics for patients included within the dataset for model construction

Characteristic	Average	Total	Cardiac complication	VTE complication	Wound complication	Mortality
Sex						
Male		9,978 (48.4)	26 (0.3)	39 (0.4)	54 (0.5)	20 (0.2)
Female		10,620 (51.6)	8 (0.1)	15 (0.1)	55 (0.5)	3 (0)
Mean age (yr)	53.2					
Ethnicity						
White		16,900 (82)	29 (0.2)	36 (0.2)	87 (0.5)	19 (0.1)
Black		1,862 (9)	3 (0.2)	15 (0.8)	13 (0.7)	3 (0.2)
Hispanic		175 (0.8)	0 (0)	0 (0)	0 (0)	0 (0)
Other		1,661 (8.1)	2 (0.1)	3 (0.2)	9 (0.5)	1 (0.1)
Diabetes mellitus						
No		17,698 (85.9)	22 (0.1)	48 (0.3)	92 (0.5)	15 (0.1)
Type II		1,913 (9.3)	10 (0.5)	4 (0.2)	12 (0.6)	5 (0.3)
Type I		987 (4.8)	2 (0.2)	2 (0.2)	5 (0.5)	3 (0.3)
Smoking history						
Smoker		6,098 (29.6)	9 (0.1)	9 (0.1)	32 (0.5)	4 (0.1)
Nonsmoker		14,500 (70.4)	25 (0.2)	45 (0.3)	77 (0.5)	19 (0.1)
Steroid use						
Steroid use		613 (3)	1 (0.2)	2 (0.3)	6 (1)	3 (0.5)
No steroid use		19,985 (97)	33 (0.2)	52 (0.3)	103 (0.5)	20 (0.1)
History of bleeding disorder		189 (0.9)	1 (0.5)	0 (0)	1 (0.5)	3 (1.6)
None		20,409 (99.1)	33 (0.2)	54 (0.3)	108 (0.5)	20 (0.1)
Functional status						
Dependent		291 (1.4)	4 (1.4)	3 (1)	2 (0.7)	5 (1.7)
Independent		20,307 (98.6)	30 (0.1)	51 (0.3)	107 (0.5)	18 (0.1)
ASA PS classification						
≥ III		7,700 (37.4)	25 (0.3)	21 (0.3)	50 (0.6)	18 (0.2)
Mean BMI (kg/m²)	30.1					
Comorbidities						
Pulmonary		1,678 (8.1)	9 (0.5)	6 (0.4)	15 (0.9)	8 (0.5)
Cardiac		8,784 (42.6)	26 (0.3)	24 (0.3)	56 (0.6)	18 (0.2)
Complications						
Mortality		23 (0.1)	11 (47.8)	1 (4.3)	0 (0)	23 (100)
Wound complications		109 (0.5)	1 (0.9)	3 (2.8)	109 (100)	0 (0)
VTE		54 (0.3)	1 (1.9)	54 (100)	3 (5.6)	1 (1.9)
Cardiac complications		34 (0.2)	34 (100)	1 (2.9)	1 (2.9)	11 (32.4)

Values are presented as number (%) unless otherwise indicated.
ASA PS, American Society of Anesthesiologists physical status; BMI, body mass index; VTE, venous thromboembolism.

poor clinical outcomes (Fig. 2).[11-13] Due to incomplete data, 923, 927, 894, and 920 patients were excluded from the cardiac complication, VTE, wound complication, and mortality training sets respectively. As previously described, ADASYN was used to generate data from minority class in the training set. To improve learning with class-imbalanced data, 729, 726, 696, and 724 cases were generated by ADASYN in the cardiac complication, VTE, wound complication, and mortality training sets respectively.
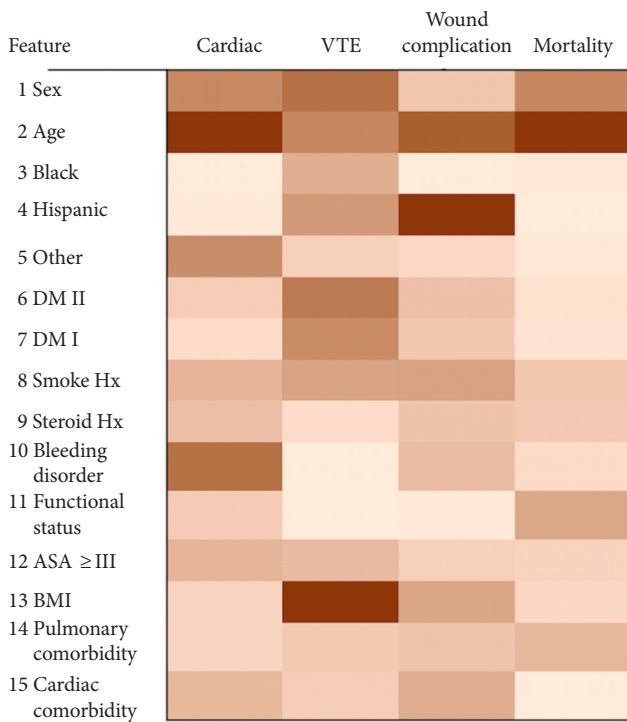
Fig. 2. Coefficient weights obtained from logistic regression analysis used for feature selection. Dark cells indicate highly weighted features indicating a strong predictive value, and lighter cells indicate weakly weighted features. VTE, venous thromboembolism; DM, diabetes mellitus; Hx, history; ASA, American Society of Anesthesiologists; BMI, body mass index.

2. ANN, LR, SVM, RF, and ASA Physical Status Classification Performance

ASA physical status classification was used to benchmark machine learning performance. AUC was used to measure the performance of our classifiers (Fig. 3). The LR and ASA physical status classifiers were outperformed by the ANN for every target (Fig. 4). The ANN performed with an AUC of 0.772 (95% CI, 0.766–0.778) for predicting cardiac complications, 0.656 (95% CI, 0.653–0.658) for predicting VTE, 0.518 (95% CI, 0.510–0.527) for predicting wound complications, and 0.979 (95% CI, 0.978–0.981) for predicting mortality. In contrast, the LR performed consistently better than ASA as a classifier with an AUC of 0.759 (95% CI, 0.738–0.781) for cardiac complications, 0.639 (95% CI, 0.632–0.645) for VTE events, 0.501 (95% CI, 0.500–0.503) for wound complications, 0.974 (95% CI, 0.973–0.976) for mortality. The SVM and RF classifier had the poorest performance. Neither the SVM nor the RF classifiers were able to predict occurrence of postoperative complication better than random chance across all complications (p < 0.05). The ASA physical status classification performed least effectively for all target outcomes with an AUC of 0.566 (95% CI, 0.544–0.587) for cardiac complications, 0.397 (95% CI, 0.388–0.407) for VTE, 0.455 (95% CI, 0.449–0.461), and 0.346 (95% CI, 0.342–0.350) for mortality (Table 2). These findings demonstrate that ANN and LR were consistently the best at predicting postoperative complications. To compare the top 2 performing models, ANN and LR models were asked to predict postoperative mortality and wound complication, the easiest and hardest postoperative complications, respectively, on a blinded dataset (Fig. 5). The ANN had greatly improved sensitivity than LR for predicting postoperative mortality and wound complication.

DISCUSSION

With the advent of large, prospective, multi-institutional clini-

Table 2. Comparison of AUC of machine learning models and ASA evaluated on blinded data

| Variable	LR	ANN	SVM	RF	ASA
Cardiac	0.759 (0.738–0.781)	0.772 (0.766–0.778)	0.559 (0.485–0.633)	0.251 (0.229–0.273)	0.566 (0.544–0.587)
VTE	0.639 (0.632–0.645)	0.656 (0.653–0.658)	0.430 (0.427–0.434)	0.357 (0.299–0.414)	0.397 (0.388–0.407)
Wound	0.501 (0.500–0.503)	0.518 (0.510–0.527)	0.422 (0.413–0.432)	0.489 (0.457–0.522)	0.455 (0.449–0.461)
Mortality	0.974 (0.973–0.976)	0.979 (0.978–0.981)	0.214 (0.193–0.234)	0.393 (0.295–0.491)	0.346 (0.342–0.350)

Values are presented as 95% confidence interval.
AUC, area under the receiver operating characteristic curve; ASA, American Society of Anesthesiologists; LR, logistic regression; ANN, Artificial neural network; SVM, support vector machine; RF, random forest decision tree; VTE, venous thromboembolism.
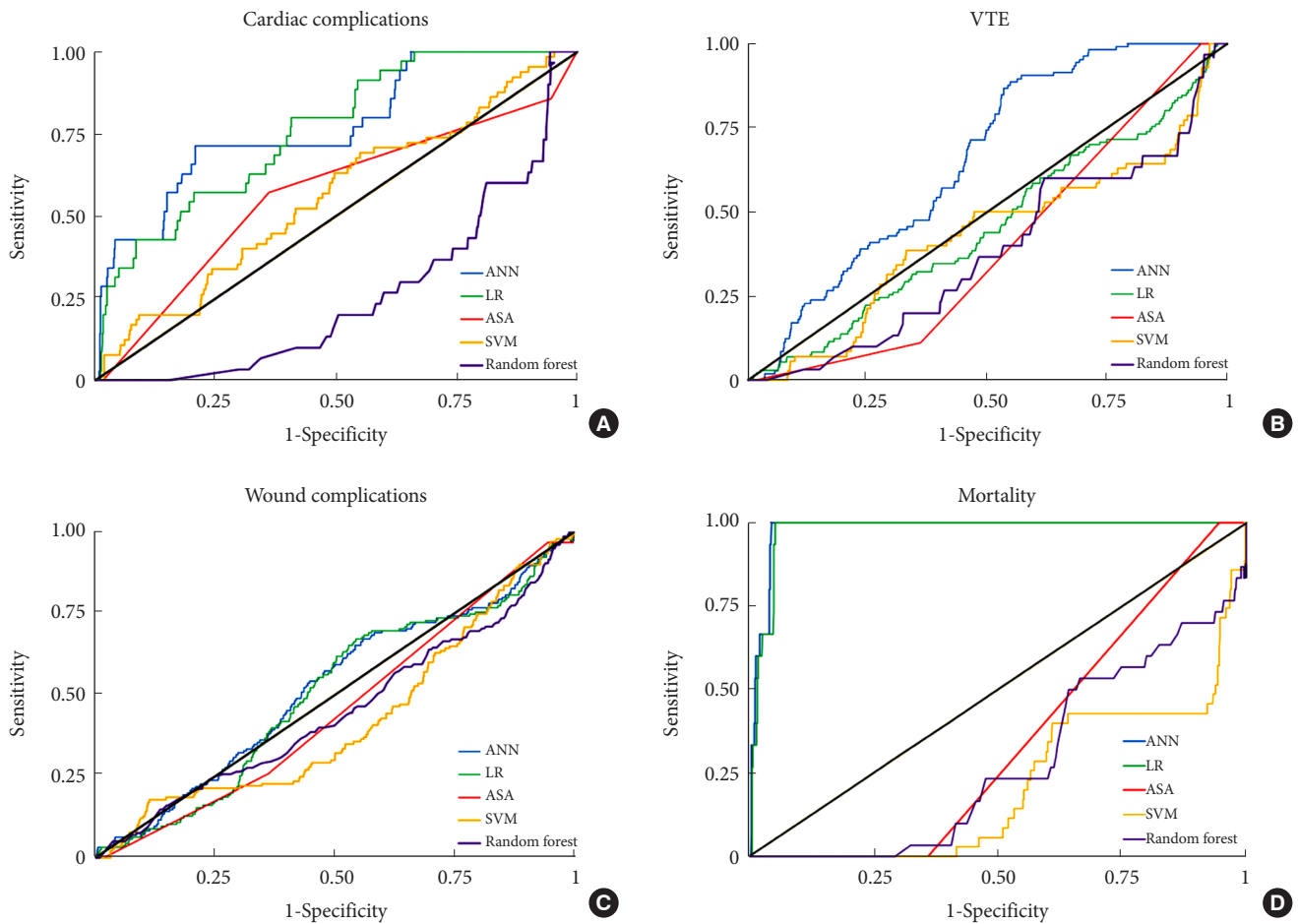
Fig. 3. Receiver operating characteristic curves plotting sensitivity versus 1-specificity for artificial neural network (ANN) (blue), logistic regression (LR) (green), American Society of Anesthesiologists (ASA) physical status classification (red), support vector machine (SVM) (yellow), random forest decision tree (RF) (purple), and random-chance (black). (A) Cardiac complications, (B) venous thromboembolism, (C) wound complications, and (D) mortality.
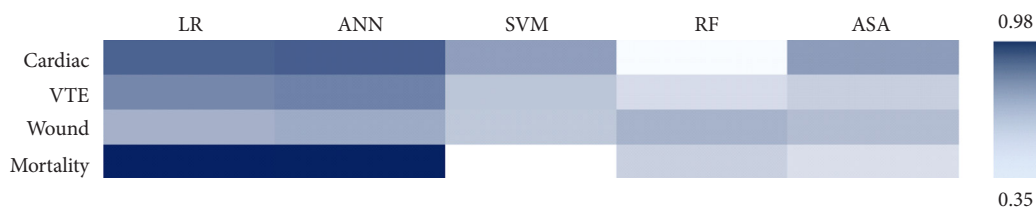
Fig. 4. Heatmap of area under the receiver operating characteristic curve values from LR, ANN, SVM, RF, and ASA when predicting cardiac complications (cardiac), VTE, wound complications (wound), and mortality. LR, logistic regression; ANN, Artificial neural network; SVM, support vector machine; RF, random forest decision tree; ASA, American Society of Anesthesiologists; VTE, venous thromboembolism.

cal registries, physicians have access to large amounts of diverse, high quality clinical data. This has given birth to ideas such as "precision medicine" with the goal of developing quantitative models that can be used to predict health status, prognosticate

disease processes, prevent disease, and reduce complications. Previous groups have employed the use of ANNs and other machine learning models to these data sets.[14-17] However, these studies either trained models on extremely large databases (>1,400,000

Mortality

Wound complication
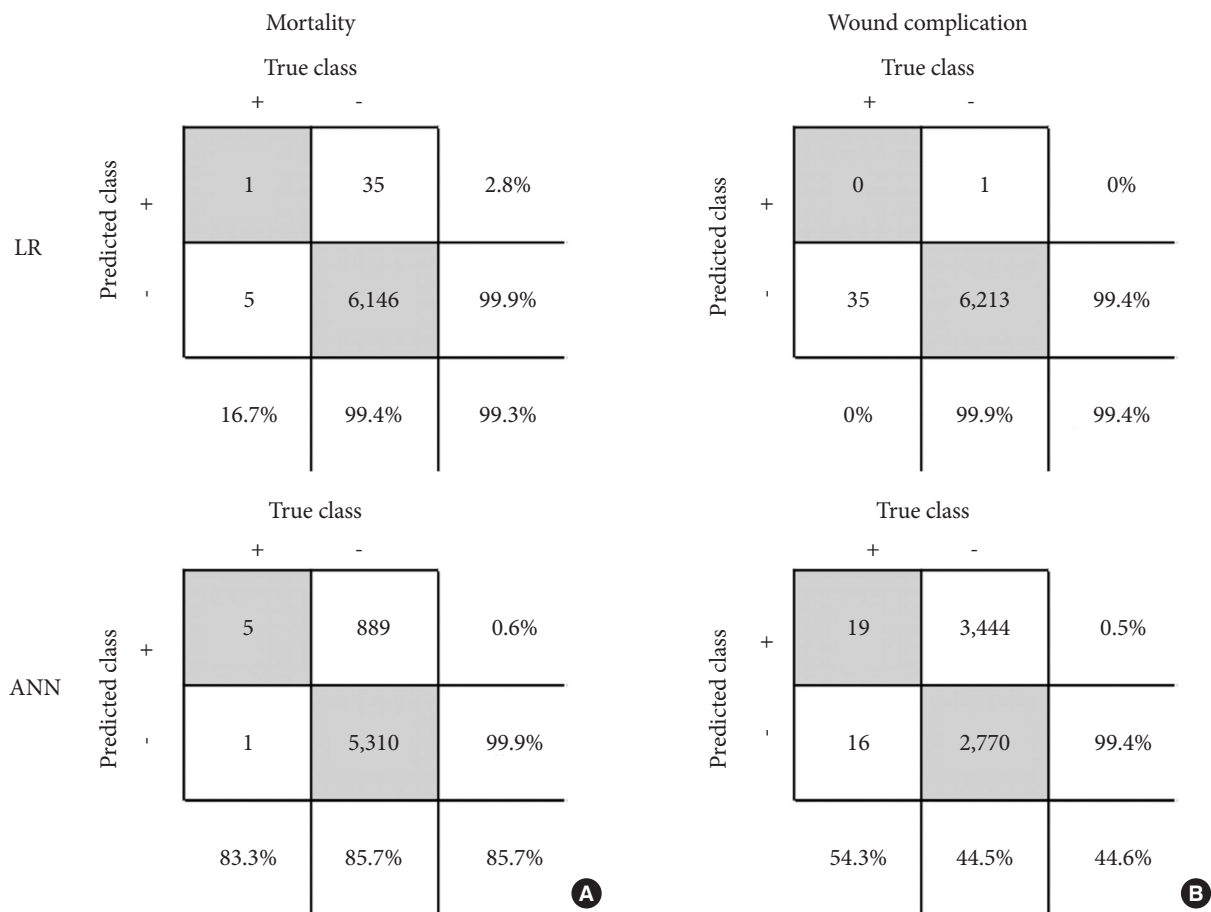
Fig. 5. Confusion matrices of trained ANN and LR machine learners evaluated on testing data set mortality (A) and wound complication (B) data sets to demonstrate real-world performance. LR, logistic regression; ANN, Artificial neural network.

patients) or on complications with high occurrence rates. These examples are impractical for independent institutions or for small scale procedures with rare complications. Low occurrence rates in relatively small datasets lead to large class-imbalances that are a significant challenge in medical machine learning.[18,19] To this end, we have trained several supervised machine learning classifiers to predict the probability of postoperative complications in a relatively small dataset (<15,000 patients) that can accurately learn complications with relatively low occurrence rates (<1%). Furthermore, we have rigorously developed and tested our models by employing the best practices in machine learning in this study by performing automated feature selection, L2 regularization and comparing to a standard risk-scoring system to ensure a high standard that is necessary for implementation of machine learning in clinical settings.

The ANN model was superior to the LR, and both were superior to a clinical benchmark, the ASA physical status classification, with a statistically significantly higher AUC when pre-

dicting VTE, wound complications, and mortality. The sensitivity of the ANN was superior to LR, indicating an ability to identify a greater portion of positive cases, correctly identified as positive. Both the SVM and RF classifiers were unable to perform better than random chance, suggesting that model selection is an important design parameter. This is an important finding that may be worthy of consideration when developing machine learning models for clinical prognostication. Automated feature selection with LR showed that age and male gender were the strongest independent risk factors for mortality, both consistent with current surgical evidence. Wound complications were predicted by age, Hispanic race, BMI, smoking history, diabetes mellitus, and bleeding disorders. VTE was predicted for by BMI, sex, age, diabetes mellitus, and smoking history. Cardiac complications were predicted by age, bleeding disorder, sex, smoking history, cardiac comorbidity, and high ASA physical status classification. These findings are echoed heavily by domain knowledge in prior spine literature.[20,21] While ASA

physical status classification was included as a feature of prediction of cardiac complication, it was not included as a feature used for prediction of VTE, wound complication, and mortality by our models. Thus in these complications ASA physical status classification is a good baseline comparison to benchmark our machine learning algorithms.

A key strength of this study is the adaptability that can be achieved by interrogating medical data with different machine learning models. Indeed, neural network architectures alone are a diverse field of study that seeks to design optimal neural network structures to improve artificial intelligence (AI) predictions.[22] In this study, a grid search was performed to identify optimal hyperparameters. However, this was only carried out over a certain defined domain of hyperparameters not considering all types of network structures and other macro-scale parameters that may be more suited for medical prognostication. This presents a novel opportunity to design machine learners that are adept at learning and prognosticating based on patient data that is highly diverse, class-imbalanced, and often limited in sample size.[23]

The ability of machine learning to identify at-risk-patients and predict potential complications has been clearly demonstrated here, yet the ability to suggest avenues of treatment based on predicted complications has not yet been realized. Future work can take advantage of electronic medical records and medical literature to suggest optimal treatment strategies based on key patient data. Such models can not only guide physicians in the decision making process but can also aide health care systems in low-resource settings, provide personalized care, and improve response times during critical settings. Taken together, the opportunities described here can be used to strengthen medical AI to improve surgical outcomes.

The performance of any classifier, is rooted, in part, in the quality of the training data. Therefore, weaknesses in the NSQIP are represented as weaknesses in the neural network classifier. Larger national in-patient datasets such as the National In-Patient Sample exist. Such data sets sampling patients with a broad demographic spectrum can serve to elucidate patterns in the model that are both more generalizable and predictive of future complications and risk. A major challenge in medicine is the paucity of highly granular and robust large-scale datasets for specific operational cohorts. Large-scale databases remain scattered across institutions and are isolated to protect patient privacy.[24] Furthermore, the NSQIP dataset was not designed with spine surgery outcomes in mind. As a result, many features which may serve as stronger inputs were not available. Addi-

tionally, the large class-imbalance shown in this study serves to skew machine learning models during training. Indeed, Fig. 5 shows that the LR model was skewed to predict the negative outcome, while the ANN was less biased. This highlights the need for future work to further address the issue of class-imbalance to improve machine learning performance in clinical contexts. While these challenges exist, the ability to predict clinical outcomes using NSQIP data is an attractive prospect. The advent of machine learning algorithms and their implementation in a healthcare environment makes the utilization of such machine learning towards increasingly possible. In the past, generalized linear models such as the LR have been the most commonly used classifiers for this purpose. However, the machine learning models described here, particularly the ANN, are similarly powerful and in some circumstances, far exceed LR. As the ability to obtain high quality patient data and computing power increases over time, it is likely that machine learning techniques will find themselves increasingly commonplace in the hospital setting.

CONFLICT OF INTEREST

The authors have nothing to disclose.

REFERENCES

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

2. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35:352-9.

3. Hopfield JJ. Artificial neural networks. IEEE Circuit Device Mag 1988;4:3-10.

4. Baird EO, Egorova NN, McAnany SJ, et al. National trends in outpatient surgical treatment of degenerative cervical spine disease. Global Spine J 2014;4:143-50.

5. Klein GR, Vaccaro AR, Albert TJ. Health outcome assessment before and after anterior cervical discectomy and fusion for radiculopathy: a prospective analysis. Spine (Phila Pa 1976) 2000;25:801-3.

6. Yue WM, Brodner W, Highland TR. Long-term results after anterior cervical discectomy and fusion with allograft and plating: a 5- to 11-year radiologic and clinical follow-up study. Spine (Phila Pa 1976) 2005;30:2138-44.

7. Marawar S, Girardi FP, Sama AA, et al. National trends in

anterior cervical fusion procedures. Spine (Phila Pa 1976) 2010;35:1454-9.

8. Adamson T, Godil SS, Mehrlich M, et al. Anterior cervical discectomy and fusion in the outpatient ambulatory surgery setting compared with the inpatient hospital setting: analysis of 1000 consecutive cases. J Neurosurg Spine 2016;24:878-84.

9. Haibo He, Yang Bai, Garcia EA, et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). IEEE;2008:1322-8.

10. Blum AL, Langley P. Selection of relevant features and examples in machine learning. Artif Intell 1997;97:245-71.

11. Møller AM, Pedersen T, Villebro N, et al. Effect of smoking on early complications after elective orthopaedic surgery. J Bone Joint Surg Br 2003;85:178-81.

12. Iorio R, Williams KM, Marcantonio AJ, et al. Diabetes mellitus, hemoglobin A1C, and the incidence of total joint arthroplasty infection. J Arthroplasty 2012;27:726-9.

13. Chen S, Anderson MV, Cheng WK, et al. Diabetes associated with increased surgical site infections in spinal arthrodesis. Clin Orthop Relat Res 2009;467:1670-3.

14. Van Esbroeck A, Rubinfeld I, Hall B, et al. Quantifying surgical complexity with machine learning: looking beyond patient factors to improve surgical models. Surgery 2014;156: 1097-105.

15. Hu Z, Simon GJ, Arsoniadis EG, et al. Automated detection of postoperative surgical site infections using supervised methods with electronic health record data. Stud Health Technol Inform 2015;216:706-10.

16. Sohn S, Larson DW, Habermann EB, et al. Detection of clinically important colorectal surgical site infection using Bayesian network. J Surg Res 2017;209:168-73.

17. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013;217:833-42.

18. Krell MM, Wilshusen N, Seeland A, et al. Classifier transfer with data selection strategies for online support vector machine classification with class imbalance. J Neural Eng 2017; 14:025003.

19. Wang Q, Luo Z, Huang J, et al. A novel ensemble method for imbalanced data learning: bagging of extrapolation-SMOTE SVM. Comput Intell Neurosci 2017;2017:1827016.

20. Somani S, Di Capua J, Kim JS, et al. Comparing national inpatient sample and national surgical quality improvement program: an independent risk factor analysis for risk stratification in anterior cervical discectomy and fusion. Spine (Phila Pa 1976) 2017;42:565-72.

21. Wang TY, Martin JR, Loriaux DB, et al. Risk assessment and characterization of 30-day perioperative myocardial infarction following spine surgery: a retrospective analysis of 1346 consecutive adult patients. Spine (Phila Pa 1976) 2016;41: 438-44.

22. Tsai JT, Chou JH, Liu TK. Tuning the structure and parameters of a neural network by using hybrid Taguchi-genetic algorithm. IEEE Trans Neural Netw 2006;17:69-80.

23. Cios KJ, Moore GW. Uniqueness of medical data mining. Artif Intell Med 2002;26:1-24.

24. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA 2014;311:2479-80.