International Journal of Research in Orthopaedics
Kricfalusi M et al. Int J Res Orthop. 2024 Mar;10(2):306-313
http://www.ijoro.org

DOI: https://dx.doi.org/10.18203/issn.2455-4510.IntJResOrthop20240402

Original Research Article

Supervised machine learning algorithms used to predict post-surgical outcomes following anterior surgical fixation of odontoid fractures

Mikayla Kricfalusi[1]*, Richard C. Rice[2], Andrew Cabrera[3], Prathyusha Dasari[4], David Chung[2], Trevor Case[1], David E. Ruckle[2], Joseph N. Liu[5], Wayne K. Cheng[2], Olumide Danisa[2]

[1]California University of Science and Medicine School of Medicine, 1501 Violet St, Colton, CA, USA
[2]Loma Linda University Health, Department of Orthopedic Surgery, Loma Linda, CA, USA
[3]Loma Linda University School of Medicine, 11021 Campus Street, Loma Linda, CA, USA
[4]University of California Riverside School of Medicine, 92521 Botanic Gardens Dr, Riverside, CA, USA
[5]USC Epstein Family Center for Sports Medicine, Keck Medicine of USC, 1520 San Pablo St. Los Angeles, CA, USA

Received: 14 November 2023
Accepted: 04 January 2024

***Correspondence:**
Dr. Mikayla Kricfalusi,
E-mail: mikayla.kricfalusi@student.cusm.edu

ABSTRACT

Background: Odontoid fractures have a high mortality rate, and numerous classification systems have previously predicted surgical outcomes with mixed consensus. We generated a machine learning (ML) construct to predict post-operative adverse events following anterior (ORIF) of odontoid fractures.

Methods: 266 patients from the American college of surgeons-national surgical quality improvement program (ACS-NSQIP) with anterior ORIF (CPT 22318) of odontoid fractures from 2008-2018 were analyzed using ML algorithms random forest classifier (RF), gradient boosting classifier (GB), support vector machine classifier (SVM), Gaussian Naive Bayes classifier (GNB), and multi-layer perceptron classifier (MLP), and were compared to logistic regression classifier (LR). Algorithms predicted increased length of stay (LOS), need for transfusion (Transf), non-home discharge (NHD), and any adverse event (AAE). Permutation feature importance (PFI) identified risk factors.

Results: ML algorithms outperformed LR. The average AUC for predicting Transf was 0.635 (accuracy=77.4%), extended LOS=0.652 (accuracy 59.6%), NHD 0.788 (accuracy=71.9%) and AAE 0.649 (accuracy 68.1%). GB performed highest for Transf (AUC=0.861), identifying operative time (PFI 0.253, p=0.016). GB and RF performed equally for NHD (AUC=0.819), highlighting preoperative hematocrit (PFI=0.157, p<0.001). GB predicted AAE (AUC=0.720) also identifying preoperative hematocrit (PFI=0.112, p<0.001). RF predicted extended LOS (AUC=0.669) highlighting preoperative hematocrit (PFI=0.049, p<0.001).

Conclusions: ML outperformed LR, successfully predicting Transf, extended LOS, NHD, and AAE for anterior ORIF of odontoid fractures. Our construct may complement conventional risk stratification to reduce adverse outcomes and excess cost.

Keywords: ML, Odontoid fracture, Anterior odontoid screw fixation, Hematocrit

INTRODUCTION

Odontoid fractures are common fractures of the cervical spine that portend significant instability requiring operative intervention. Upper cervical spine fractures account for 69% of all cervical spine injuries in the elderly population, the vast majority of which are Anderson-D'Alonzo type II odontoid fractures.[1] Interestingly, the incidence of odontoid fractures appears to be surging in a bimodal age distribution hypothesized to be from a multitude of etiologies, including an increasing prevalence of motor vehicle accidents (in younger population) and longevity (in elderly).[2] Dens fractures have previously been identified as etiologic mediators of numerous, concomitant adverse events including mortality rates between 25-37.5% (comparable to hip fractures).[3] Management of type II odontoid fractures is controversial with significant dissonance amongst spine experts.[1,2] Anterior fixation of odontoid fractures is associated with good clinical outcomes, including 88% union by 6 months, high biomechanical stability, and theoretical maintenance of cervical spine rotational range of motion (C1/2 fusion mediates up to 50% reduction in cervical rotation).[4-6] In spite of this, many authors diametrically advocate posterior fusion over anterior dens screw in management odontoid fractures in elderly population due to ADS-specific adverse events, including lower bony fusion when using single screw, higher rates of re-operation, and increased incidence of acute postop genitourinary, pulmonary, renal, and gastrointestinal complications.[7] While many previous studies have focused on risks of osseous nonunion following fixation, few studies have successfully predicted major 30-day postoperative complications that portend immediate deleterious outcomes and injury-related financial burden.[8,9]

Technological advancements in artificial intelligence and ML have revolutionized predictive modeling in many fields, including medicine.[10] Application of ML to healthcare has vastly expanded our predictive capabilities, especially when for outcome-driven analysis of retrospective cohort studies.[11] These models are effective and efficient tools capable of analyzing vast databases for variables of interest to ultimately produce data-driven forecasts.[12]

Numerous other studies have similarly attempted to evaluate co-morbidities, admission variables, or treatment modalities to establish correlations with patient outcomes following anterior ORIF. Age, frailty score, neurologic deficit, ASA class, preoperative Rankin score, Charlson comorbidity index, and hemoglobin on admission, among others, have been implicated as independent predictors of nonunion, revision surgeries, morbidity, and mortality.[13,14] However, to our knowledge, no team has previously employed ML in this endeavor. Therefore, this study utilizes ML algorithms to predict deleterious outcomes following anterior fixation of odontoid fractures and identify predictive perioperative variables to inform optimal management, mitigate morbidity, facilitate preoperative risk stratification, while reducing surgical cost.

METHODS

This observational study was conducted using deidentified data from the ACS-NSQIP database. ACS-NSQIP was queried using RStudio (RStudio, PBC, Boston, MA) to identify adult patients undergoing anterior open treatment of dens fractures from 2008 to 2018 using current procedural terminology (CPT) codes 22318 and 22319. Patients with fractures due to malignancy or those with missing data were excluded from analyses within our study.

Patients matching our specified criteria were analyzed by five supervised ML classification algorithms, namely RF, GB, SVM, GNB, and MLP. These algorithms are adept at identifying complex interactions between features, especially when training data is abundant. SVM is considered a robust algorithm for datasets in which there are relatively large numbers of features in comparison to a smaller number of training cases, which was expected to be beneficial considering the smaller population of odontoid fracture patients.[15-17] The decision tree algorithms such as RF are considered some of the best general models for ML in many applications, since they generate multiple trees based on layering of different features in the training phase, and then take the majority conclusion of all decision trees during the testing phase.[19] Such algorithms have strong predictive capabilities while maintaining sufficient flexibility to incorporate the highly variable characteristics of individual patients.[16] Each algorithm was constructed using the SciKit-Learn library in the python programming language and tasked with predicting extended LOS, non-home discharge (NHD), transfusion, and any adverse event based on a given set of patient variables.[18-20] Extended LOS was defined as greater than 7 days, based on previous studies reporting 5 days to 9 days as the average LOS.[12] NDH was defined as discharge to “skilled care”, “rehabilitation facility”, “separate acute care”, “unskilled facility not home”, or “multi-level senior community”, as coded in ACS-NSQIP. Home discharge was defined as discharge locations encoded as “home”, “facility which was home”, or “against medical advice.” Transfusion was defined as any red blood cell transfusion in the perioperative period. Any adverse event was defined as those having any one or multiple of the following: surgical site infection, renal complications, sepsis, intubation, transfusion, pneumonia, deep vein thrombosis (DVT), urinary tract infection (UTI), cerebrovascular accidents, cardiac arrest, myocardial infarction (MI), return to operating room, or death. All postoperative outcomes are within 30-days of surgery, which is the maximum follow-up time recorded by ACS-NSQIP.[22]

Patient variables included demographic information, preoperative lab values, comorbidities, operative-time, ASA score, and the 5-factor modified frailty index, which evaluates functional status, diabetes, heart failure, chronic obstructive pulmonary disease, and hypertension to determine the level of decreased physiologic reserve in the elderly (Table 1).[23] Preprocessing was performed by removing the mean and scaling to unit variance to bring all features to the same magnitude, thereby standardizing patient variables using SciKit-Learn’s standard scaler.[18] A train test split was performed using Scikit-Learn’s train_test_split method in which a subset of 70% of our population was used for training and the remaining 30% of our population was held out for later testing of the model's performance.[24] For each model, GridSearchCV along with a stratified five-fold cross validation was used to ensure generalizability through optimization of hyperparameters.[25,26] The final models were then evaluated using the 30% held out testing subset to determine each model's performance.

The performance of the five ML models was then evaluated by a series of standard metrics, primarily area under receiver operating characteristics curve (AUROC), as well as classification accuracy, sensitivity, and specificity.[21,27] The negative and positive likelihood ratios were also calculated from the sensitivity and specificity for each algorithm in predicting a certain outcome of interest. The negative likelihood ratio (NLR) is the probability that an algorithm identified a negative result in a patient without a specific adverse outcome, calculated by the equation NLR=(1-sensitivity)/specificity. Positive likelihood ratio (PLR) describes the probability that a positive result would be expected in a patient with an outcome of interest, calculated by the equation: PLR=sensitivity/(1-specificity).[28] Each model was subsequently categorized based on AUC as either acceptable (0.7-0.79), excellent (0.8-8.9), or outstanding (0.9-1.0).[28] The graphical visualization of the ROCs produced by each of the models was accomplished through utilization of the Matplotlib library in python.[29] Permutation feature importance (PFI) was derived from the highest-performing models to determine the predictive value of each variable, via utilization of the ELI5 library (version 0.11.0). PFI is generated by measuring variations in model performance after removing individual features at random. This allows us to determine the relationship between a variable and the predicted outcome, as a decrease in a model's performance corresponds to the extent that the model depends on a particular variable for prediction.[28,30]

Statistical analysis utilized SPSS version 29 (IBM corporation, 2021, Armonk, NY, USA) with statistical significance defined as $p<0.05$. Categorical differences between groups were assessed using Pearson's chi-square test and presented as frequencies in percentages.

RESULTS

The patient dataset initially contained 344 patients with CPT codes 22318 or 22319. After exclusion of malignancy and missing data, the final population consisted of 266 patients (122 male, 181 female) with a mean age of 68.25 years (Table 1). The majority of patients, 92%, were functionally independent upon admission and 58% had an ASA of 3. A frailty index of 1 was most common, followed by 0, accounting for 41% and 38% of patients respectively. It is important to note that none of the patients had a smoking history, a risk factor well known to increase the likelihood of non-union following the orthopedic surgery.[31]

Additionally, 35% of patients had an extended LOS, with an average LOS of 6.81 days, while 51% were discharged to a non-home location. Transfusion was required for 12% of patients and 24% experienced a postoperative adverse event, as defined previously.

We evaluated the performance of five different algorithms alongside LR in predicting several postoperative events of interest, including transfusion, extended LOS, NHD, and any adverse event. In predicting transfusion, GB achieved an AUC of 0.861, specificity of 1.0, sensitivity 0.2, and overall accuracy of 90.7%, indicating very high predictive capability. However, the low sensitivity indicates that GB often misses patients who eventually needed transfusion. The SVM algorithm also performed well, with an AUC of 0.715, specificity, 0.7, sensitivity, 0.2, and accuracy of 64.8%.

The RF algorithm was adequately predictive with an AUC of 0.628, specificity of 0.854, sensitivity 0.5, and accuracy 79.6%, while the remaining algorithms (MLP and GNB) only resulted in AUCs of 0.486 and 0.483 respectively. In comparison, all ML algorithms outperformed LR, which yielded an AUC of 0.458, specificity 0.6, sensitivity 0.3, and accuracy of 55.6% for predicting transfusion.

Several algorithms were adequately predictive of extended LOS, with the RF algorithm performing highest with an AUC of 0.669, specificity 0.7, sensitivity 0.5, and accuracy of 66.7%. The MLP, GB, SVM, GNB algorithms performed similarly, achieving AUCs of 0.663, 0.654, 0.650, and 0.623, respectively. Of note, GB and MLP both showed a specificity and sensitivity of 1.0, and accuracy of 64.8%. All algorithms outperformed LR (AUC=0.616, specificity 0.6, sensitivity 0.6, and accuracy 61.1%) for predicting extended LOS.

The RF and GB algorithms were highly predictive of NHD following anterior odontoid fixation. RF achieved an AUC of 0.839, specificity 0.8, sensitivity 0.8, and accuracy of 81.5%. GB had an AUC of 0.819, specificity of 0.8, sensitivity of 0.7, and accuracy of 76.0%. MLP and GNB algorithms also performed well, with AUCs of 0.749 and 0.758, and specificities of 1.0 and sensitivities of 0.2 for both algorithms.

The accuracy of MLP was 70.4% and 59.3% for GNB. In predicting NHD, all ML algorithms (including the SVM, AUC 0.775, specificity 0.9, sensitivity 0.6, and accuracy 72.2%) were greater than the AUC threshold of 0.7, and outperformed LR with an AUC of 0.659, specificity 0.6, sensitivity 0.6, and accuracy of 59.3%.

When looking at any adverse events, GB performed the highest at an AUC of 0.720, specificity of 0.9, sensitivity 0.2, and accuracy of 72.2%. The RF algorithm achieved an AUC of 0.672, specificity of 0.9, sensitivity 0.3, and accuracy of 75.9%. SVM, MLP, and GNB reported slightly lower AUCs of 0.65, 0.62, and 0.59, respectively, and accuracies of 64.8%, 64.8% and 63.0% respectively, but still outperformed LR with an AUC of 0.485, specificity=0.6, sensitivity=0.4, and accuracy of the 53.7%.

The comparative AUCs of all algorithms for the outcomes of interest are shown in Figure 1.

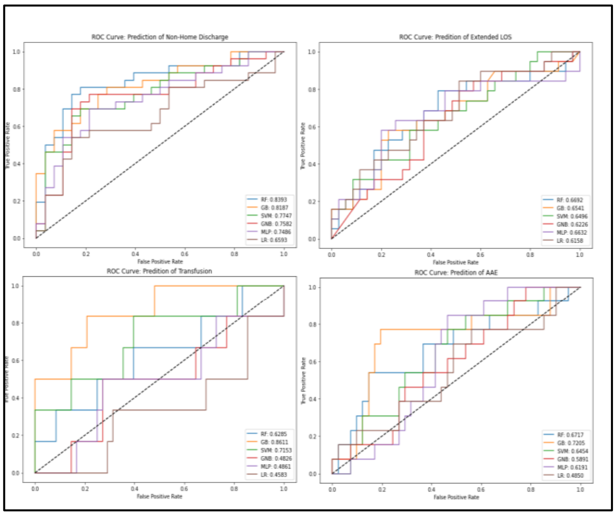

Figure 1: ROC of various machine learning algorithms applied to the outcome variables of interest: extended LOS, NHD, transfusion, and AAE.

RF-Randomfo-rest, GB-Gradient boosting classifier, SVM-Support vector machine classifier, GNB- Gaussian Naive Bayes classifier, MLP-Multi-layer perceptron, LR-Logistic regression.

Also examined PFI for highest performing algorithms in order to determine the most influential variables for prediction of outcomes. GB algorithm most accurately predicted need for transfusion and identified operative time as the best predictor within GB algorithm structure, with mean PFI of 0.253 (p=0.016). Preo hematocrit (PFI=0.133, p<0.001) was also shown to be an important variable for GB prediction of transfusion and is similarly implicated in predicting extended LOS within RF algorithm (PFI=0.049, p<0.001), NHD via RF (PFI=0.0157, p<0.001) and any adverse event for GB (PFI=0.112, p<0.001). For extended LOS, ASA class 4 was highly predictive, with mean PFI of 0.020 (p<0.001), although it was ranked similarly to presence of a bleeding disorder (PFI=0.016, p=0.239). PFI for RF prediction of NHD showed preop hematocrit as most important variable (PFI=0.157, p<0.001), followed by age (PFI=0.051, p<0.001) and operative time (PFI=0.253, p=0.016).

Finally, the GB algorithm also identified preoperative hematocrit as highly predictive of any adverse event (PFI=0.112, p<0.001). Age was also highly predictive, with a PFI of 0.081 (p=0.007).

Table 1: Preoperative characteristics of study population with anterior odontoid fixation, (n=266).

Variables	N (%)
Demographics	
Age (In years)	68.25±17.04
Body mass index (Mean ± SD)	25.80±5.95
Gender	
Male	114 (43)
Female	152 (57)
Race	
Asian	6 (2)
Black or African American	19 (7)
Native Hawaiian or Pacific Islander	1 (0)
White	225 (85)
Unknown/ not reported	15 (6)
Ethnicity Hispanic	13 (5)
Functional status	
Independent	245 (92)
Partially dependent	17 (6)
Totally dependent	4 (2)
Comorbidities	
Smoking	0 (0)
Diabetes	
Non-insulin dependent	29 (11)
Insulin dependent	13 (5)
Congestive heart failure	3 (1)
COPD	22 (8)
Dialysis	2 (1)
Hypertension requiring medication	150 (56)
Ascites	0 (0)
Cancer	11 (4)
Dyspnea	
Moderate exertion	14 (5)
At rest	7 (3)
History of oral steroid use	21 (8)
Bleeding disorder	19 (7)
Weight loss	10 (4)

Continued.

Variables	N (%)
Laboratory values	
Hematocrit (%)	37.65±5.82
Creatinine	0.90±0.54
White blood cell count	8.49±3.28
Platelet count	226.32±79.75
Sodium (nmol/L)	138.57±0.49
ASA classification	
1	5 (2)
2	54 (20)
3	155 (58)
4	52 (20)
5	1 (0)
Fragility index	
0	101 (38)
1	110 (41)
2	42 (16)
3	10 (4)
4	2 (1)
5	1 (0)
Operative time	133.62±97.86

ASA-American society of anesthesiologists, SD-Standard deviation.

DISCUSSION

As the longevity of Americans continues to increase, characterizing diseases and injuries with a relatively higher incidence within the elderly population will be of paramount importance for reducing morbidity and practicing cost-effective medicine. Upper cervical spine fractures are common in elderly patients, especially Anderson-D'Alonzo type II odontoid fractures.[1] However, these injuries are becoming increasingly prevalent not only in this subpopulation, but also in younger patients for a variety of reasons.[3] Therefore, characterizing the etiology, pathogenesis, and management of odontoid fractures is of paramount public health importance as the generalizable application of the findings will not only improve outcomes, but also imbue large, financial ramifications within healthcare. Numerous studies have attempted to determine the most important risk factors for adverse outcomes following odontoid fractures, as well as propose the best preoperative scoring system to guide clinical decision making. Carlstorm et al proposed the frailty score as a predictor of mortality in elderly odontoid fractures, while other studies examined treatment modality, age, preoperative living arrangement, the Grauer classification, preoperative Rankin score, and the Charlson comorbidity index.[12-14] Bajada et al similarly used preoperative hemoglobin and neurological status to predict 30-day mortality following odontoid fracture.[5] These previous studies focused primarily on mortality, however etiologic agents mediating deleterious outcomes following dens fracture outcomes remain poorly understood.[3] In contrast, this study employed artificial intelligence to predict mortality using many of the same independent variables as Carlstrom et al and Bajada et al while concomitantly identifying variables mediating morbidity and undesired outcomes, including extended LOS and the adverse events.[3,14]

Our construct utilized several different ML algorithms to predict postoperative outcomes of interest, and compared their predictive abilities to a LR, a traditional statistical model. Notably, our construct exhibited outstanding (AUC>0.8) success for predicting 'need for transfusion' with GB. Specifically, GB accurately predicted the correct outcome with 90.7% accuracy and an AUC of 0.861 (compared to LR's 55.6% and 0.458, respectively) and identified 'operative time' and 'preoperative hematocrit' as statistically significant predictive variables. Moreover, it demonstrated a specificity of 1.0, suggesting the clinical implementation of this ML index may exhibit high affinity for ruling out patients unlikely to require transfusion. Our construct also demonstrated outstanding predictive ability for NHD, with RF resulting in accuracy of 81.5% and AUC of 0.839, compared to LR results of 59.3% and 0.659. Any adverse event was predicted using GB, with accuracy of 72.2% and AUC of 0.72, compared to LR's 53.7% and 0.485, respectively. Preoperative hematocrit and age were identified as two statistically significant variables in predicting this outcome. As ML algorithms are reliably shown to outperform LR, we anticipate these models will eventually replace classic risk stratification tools to evaluate preoperative risk.

Conventional statistical analyses were previously employed to evaluate outcomes of odontoid fracture surgeries in ACS-NSQIP, and found that, compared to posterior cervical fusion, anterior fixation is associated with greater relative risk of need for revision surgery and 30-day hospital readmission.[44] While these findings are clinically important, this statistical methodology is unable to identify or predict patient and surgical variables associated with these outcomes. Our findings corroborate those of previous studies which utilized traditional statistical modeling to characterize preoperative risk and surgical outcomes.[14] Namely, the predictive value of preoperative hematocrit for extended LOS, NHD, and any adverse event echoed the findings of Carlstrom et al which identified preoperative hemoglobin as a strong predictor of mortality.[14] While anterior ORIF may result in less blood loss (79 mL) than posterior fusion (379 mL), even small amounts of blood loss may predispose frail patients to poor outcomes.[32] One study demonstrated that for every 50 mL of blood drawn, the risk of anemia increased by 18%.[33] Additionally, low preoperative hematocrit values have been shown to warrant prophylactic transfusion due to a significantly higher 30-day mortality in anemic patients following non-cardiac surgery.[33] This is an effect that was compounded by increased age, and thus could be applicable to the elderly dens fracture population. Our study corroborates the importance of preoperative hematocrit and its predictive significance, even in comparatively bloodless procedures. This opens opportunities for further research into strategies for mitigating adverse outcomes.

In addition to risk stratification, our construct may provide valuable information to help guide clinical decision-making and postoperative support. As Hill et al highlighted, one of the clearest benefits of ML is that it is both automated and able to take full advantage of the information contained in electronic medical records.[34] Several studies argue for conservative treatment due to the frailty of odontoid fracture patients, and there is conflicting research into rates of nonunion among different treatment modalities.[3] Therefore, it is imperative to appropriately balance the risks and benefits of surgery, considering the increased rates of comorbidities in the elderly. This will also improve clarity when setting preoperative expectations, including providing objective, numerical values for outcomes and allowing ample time for social planning, such as obtaining supervision for children or pets if high risk of extended LOS and/or non-home discharge. Finally, our algorithm can also be employed to predict and prophylactically attenuate the rate of preventable deleterious outcomes. Identifying at-risk patients preoperatively may not only prevent morbidity/mortality, but also ameliorate associated healthcare-related expenditure.

The individual cost associated with surgical management of dens fractures is tremendous, estimated to be $131,855 per patient and approximately 1.5 billion dollars annually for all dens fractures.[35,36] Several studies attribute adverse events leading to increased length of inpatient stay as the critical mediator of elevated perioperative cost in odontoid fractures. Therefore, mitigating preventable complications not only improves outcomes, but imparts benefits from the perspective of financial stewardship at both the individual and national levels.

While this study provides general insight into the care of patients undergoing anterior dens ORIF, it also builds upon a greater body of work which seeks to validate ML as a tool for clinical decision-making. Other studies have demonstrated the use of ML in predicting outcomes for procedures such as shoulder arthroplasty, lumbar fusion, cardiac surgery, and neurosurgery.[37-39] A study by Hill et al examining all patients who underwent surgical procedures at university of California Los Angeles Health found that ML outperformed all other predictors in anticipating postoperative outcomes.[34] Our results show that similar levels of predictive capability can be obtained for specific procedures, providing tuned results for providers. Future studies should consider utilizing machine-learning models to predict procedure-specific, and even somewhat esoteric complications following individual surgeries.

The key limitations of this study hinge on the modest sample size. A smaller sample size limits training of ML algorithms, and fewer patients for training can impair algorithm performance.[24] Additionally, the anterior approach for odontoid fixation is only one surgical technique for dens fixation, so cannot be generalized to all patients undergoing dens ORIF (including posterior fusion). The lack of smokers in the dataset also limits the generalizability of this study to a non-smoker population, due to the well-known risk that smoking poses to postoperative adverse events.[30,40] Our study was also retrospective in nature, so inherent limitations exist within the study design. We used ACS-NSQIP to source our data and are therefore confined by the variables recorded in the dataset. For example, one of the most important adverse outcomes of odontoid fracture is nonunion, which is determined more than 6 months postoperatively.[8] However, ACS-NSQIP only records data up to 30-days after a procedure, so nonunion was not able to be considered as an outcome of interest evaluated in this study. Finally, while machine-learning identifies important variables for the prediction of outcomes within certain algorithms, the PFIs do not directly identify risk factors. Rather, by measuring how the absence of a variable changes algorithm performance, PFIs provide insight into potential risk factors. However, confounding variables could lead to misinterpretation of results, therefore the findings of this study should be weighed in the context of existing literature.

CONCLUSION

To our knowledge, however, this is the first study employing ML to implicate key patient characteristics and perioperative variables to predict adverse events of interest. ML algorithms successfully outperformed LR and predicted NHD, need for transfusion, extended LOS, and any adverse event with high accuracy. Preoperative hematocrit stood out as a consistently high predictor of all outcomes studied. These findings provide a procedure-specific, weighted index of individual variables and their association with predicting adverse outcomes of interest following anterior ORIF of odontoid fractures. This indicates that ML can facilitate the traditional preoperative risk stratification process, provide concrete risk estimates when discussing patients' preferences on operative versus conservative management, help attenuate perioperative morbidity and cost, and assist with postoperative expectations and social planning.

ACKNOWLEDGEMENTS

Authors would like to thanks to Loma Linda university Orthopedics Department.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Delcourt T, Begue T, Saintyves G, Mebtouche N, Cottin P. Management of upper cervical spine fractures in elderly patients: current trends and outcomes. Injury. 2015;46(1):S24-7.
2. Carvalho AD, Figueiredo J, Schroeder GD, Vaccaro AR, Rodrigues-Pinto R. Odontoid fractures- a critical review of current management and future directions. Clin Spine Surg. 2019;32(8):1-11.
3. Bajada S, Ved A, Dudhniwala AG, Ahuja S. Predictors of mortality following conservatively managed fractures of the odontoid in elderly patients. Bone Joint J. 2017;99-B(1):116-21.
4. Heggeness MH, Doherty BJ. The trabecular anatomy of the axis. Spine (Phila Pa 1976). 1993;18(14):1945-9.
5. Watanabe M, Saai D, Yamamoto Y, Sato M, Mochida J. Upper cervical spine injuries: age-specific clinical features. J Orthop Sci. 2010;15(4):485-92.
6. Apfelbaum RI, Lonser RR, Veres R, Casey A. Direct anterior screw fixation for recent and remote odontoid fractures. J Neurosurg. 2000;93(2):227-36.
7. Texakalidis P, Matsoukas S, Karras CL, Hyman GF, Kevin S, Geoffrey PS et al. Outcomes following anterior odontoid screw versus posterior arthrodesis for odontoid fractures: a systematic review and meta-analysis. J Neurosurg Spine. 2023;39(2):196-205.
8. Wittauer M, Burch MA, McNally M, Thomas V, Martin C, Gregory JDR et al. Definition of long-bone nonunion: A scoping review of prospective clinical trials to evaluate current practice. Injury. 2021;52(11):3200-5.
9. Dhall SS, Yue JK, Winkler EA, Mummaneni PV, Manley GT, Tarapore PE. Morbidity and Mortality Associated with Surgery of Traumatic C2 Fractures in Octogenarians. Neurosurgery. 2017;80(6):854-62.
10. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov Today. 2017;22(11):1680-85.
11. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. Radiographics. 2017;37(2):505-15.
12. El Naqa I, Ruan D, Valdes G, Andre D, Todd McN, Yaorong G et al. Machine learning and modeling: Data, validation, communication challenges. Med Phys. 2018;45(10):e834-40.
13. De Bonis P, Iaccarino C, Musio A, Antonio M, Federico DI, Roberto D et al. Functional Outcome of Elderly Patients Treated for Odontoid Fracture: A Multicenter Study. Spine (Phila Pa 1976). 2019;44(13):951-8.
14. Carlstrom LP, Helal A, Perry A, Lakomkin N, Graffeo CS, Clarke MJ. Too frail is to fail: Frailty portends poor outcomes in the elderly with type II odontoid fractures independent of management strategy. J Clin Neurosci. 2021;93:48-53.
15. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920-30.
16. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics. 2023;24(1):48.
17. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825-30.
18. Van Rossum G, Drake FL. Python 3 Reference Manual. CreateSpace. 2009.
19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825-30.
20. Smith JS, Kepler CK, Kopjar B, James SH, Paul A, Jens RC et al. Effect of type II odontoid fracture nonunion on outcome among elderly patients treated without surgery: based on the AOSpine North America geriatric odontoid fracture study. Spine (Phila Pa 1976). 2013;38(26):2240-46.
21. User Guide for the 2017 ACS NSQIP. 2018.
22. Shi SM, McCarthy EP, Mitchell SL, Kim DH. Predicting Mortality and Adverse Outcomes: Comparing the Frailty Index to General Prognostic Indices. J Gen Intern Med. 2020;35(5):1516-22.
23. Feghali J, Gami A, Rapaport S, Jaimin P, Adham MK, Sakibul H et al. Adapting the 5-factor modified frailty index for prediction of postprocedural outcome in patients with unruptured aneurysms. J Neurosurg. 2021;136(2):456-63.
24. Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. University of Texas at El Paso. ScholarWorks@UTEP. 2018;1-6.
25. Yu T, Zhu H. Hyper-parameter optimization: a review of algorithms and applications. arXiv. 2020.
26. Zhou J, Gandomi AH, Chen F, Holzinger A. Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. Electronics. 2021;10(5):593.
27. Hosmer DW, Lemeshow S. Applied logistic regression, 2nd ed. Wiley. 2000;156-64.
28. Shreffler J, Huecker MR. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. In: StatPearls. Treasure Island (FL): StatPearls Publishing. 2023.
29. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing Sci Engin. 2007;9(3):90-95.
30. Hak DJ, Fitzpatrick D, Bishop JA, Lawrence JM, Susanne T, Reinhard S et al. Delayed union and nonunions: epidemiology, clinical issues, and financial aspects. Injury. 2014;45(2):S3-7.

31. Patterson JT, Theologis AA, Sing D, Tay B. Anterior Versus Posterior Approaches for Odontoid Fracture Stabilization in Patients Older Than 65 Years: 30-day Morbidity and Mortality in a National Database. Clin Spine Surg. 2017;30(8):E1033-8.
32. Salisbury AC, Reid KJ, Alexander KP, Frederick AM, Sue-Min L, Paul SC et al. Diagnostic blood loss from phlebotomy and hospital-acquired anemia during acute myocardial infarction. Arch Intern Med. 2011;171(18):1646-53.
33. Wu WC, Schifftner TL, Henderson WG, Charles BE, Roy MP, Georgette U et al. Preoperative hematocrit levels and postoperative outcomes in older patients undergoing noncardiac surgery. JAMA. 2007;297(22):2481-8.
34. Hill BL, Brown R, Gabel E, Nadav R, Christine L, Maxime C et al. An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. Br J Anaesth. 2019;123(6):877-86.
35. Alluri R, Bouz G, Solaru S, Kang H, Wang J, Hah RJ. A Nationwide Analysis of Geriatric Odontoid Fracture Incidence, Complications, Mortality, and Cost. Spine (Phila Pa 1976). 2021;46(2):131-7.
36. Daniels AH, Arthur M, Esmende SM, Vigneswaran H, Palumbo MA. Incidence and cost of treating axis fractures in the United States from 2000 to 2010. Spine (Phila Pa 1976). 2014;39(18):1498-505.
37. Lopez CD, Constant M, Anderson MJJ, Confino JE, Heffernan JT, Jobin CM. Using machine learning methods to predict nonhome discharge after elective total shoulder arthroplasty. JSES Int. 2021;5(4):692-8.
38. Penny-Dimri JC, Bergmeir C, Perry L, Hayes L, Bellomo R, Smith JA. Machine learning to predict adverse outcomes after cardiac surgery: A systematic review and meta-analysis. J Card Surg. 2022;37(11):3838-45.
39. Wang H, Fan T, Yang B, Lin Q, Li W, Yang M. Development and Internal Validation of Supervised Machine Learning Algorithms for Predicting the Risk of Surgical Site Infection Following Minimally Invasive Transforaminal Lumbar Interbody Fusion. Front Med (Lausanne). 2021;8:771608.
40. Wilcox T, Smilowitz NR, Xia Y, Beckman JA, Berger JS. Cardiovascular Risk Factors and Perioperative Myocardial Infarction After Noncardiac Surgery. Can J Cardiol. 2021;37(2):224-31.

Cite this article as: Kricfalusi M, Rice RC, Cabrera A, Dasari P, Chung D, Case T et al. Supervised machine learning algorithms used to predict post-surgical outcomes following anterior surgical fixation of odontoid fractures. Int J Res Orthop 2024;10:306-13.