Prediction of In-Hospital Mortality Following Vertebral Fracture Fixation in Patients With Ankylosing Spondylitis or Diffuse Idiopathic Skeletal Hyperostosis: Machine Learning Analysis

ANDREW CABRERA, BS[1]; ALEXANDER BOUTERSE, BS[1]; MICHAEL NELSON, BS[1]; COLEMAN DIETRICH, BA[1]; JACOB RAZZOUK, BS[1]; UDOCHUKWU OYOYO, MPH[2]; CHRISTOPHER M. BONO, MD[3]; AND OLUMIDE DANISA, MD[4]

[1]School of Medicine, Loma Linda University, Loma Linda, CA, USA; [2]Office of Dental Education Services, Loma Linda University School of Dentistry, Loma Linda, CA, USA; [3]Department of Orthopedics, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA; [4]Department of Orthopedics, Loma Linda University, Loma Linda, CA, USA

ABSTRACT

Background: Ankylosing spondylitis (AS) and diffuse idiopathic skeletal hyperostosis (DISH) are distinct pathological entities that similarly increase the risk of vertebral fractures. Such fractures can be clinically devastating and frequently portend significant neurological injury, thus making their prevention a critical focus. Of particular significance, spinal fractures in patients with AS or DISH carry a considerable risk of mortality, with reports on 1-year injury-related deaths ranging from 24% to 33%. As such, the purpose of this study was to conduct machine learning (ML) analysis to predict postoperative mortality in patients with AS or DISH using the Nationwide Inpatient Sample Healthcare Cost and Utilization Project (HCUP-NIS) database.

Methods: HCUP-NIS was queried to identify adult patients carrying a diagnosis of AS or DISH who were admitted for spinal fractures and underwent subsequent fusion or corpectomy between 2016 and 2018. Predictions of in-hospital mortality in this cohort were then generated by three independent ML algorithms.

Results: An in-hospital mortality rate of 5.40% was observed in our selected population, including a rate of 6.35% in patients with AS, 2.81% in patients with DISH, and 8.33% in patients with both diagnoses. Increasing age, hypertension with end-organ complications, spinal cord injury, and cervical spinal fractures each carried considerable predictive importance across the algorithms utilized in our analysis. Predictions were generated with an average area under the curve of 0.758.

Conclusions: This study's application of ML algorithms to predict in-hospital mortality among patients with AS or DISH identified a number of clinical risk factors relevant to this outcome.

Clinical Relevance: These findings may serve to provide physicians with an awareness of risk factors for in-hospital mortality and, subsequently, guide management and shared decision-making among patients with AS or DISH.

Level of Evidence: 4.

Complications

Keywords: machine learning, ankylosing spondylitis, diffuse idiopathic skeletal hyperostosis, HCUP-NIS, in-hospital mortality

INTRODUCTION

Ankylosing spondylitis (AS) and diffuse idiopathic skeletal hyperostosis (DISH) are distinct pathological entities that similarly contribute to fractures of the spine through alterations of its biomechanical properties.[1,2] Specifically, both AS and DISH entail ankylosis of contiguous vertebral segments, which ultimately renders the spine rigid, brittle, and susceptible to fracture with even minor trauma.[3–6] Such fractures are clinically devastating and frequently portend significant neurological injury, thus making their prevention a critical focus in the treatment of patients with AS or DISH.[2]

Of particular significance, spinal fractures in patients with AS or DISH carry a considerable risk of mortality, with reports on 1-year injury-related deaths ranging from 24% to 33%.[7] While the causes of mortality in these patients are multifactorial and often stem from comorbid conditions, risk factors such as age, female sex, and spinal cord injuries have been found to predict increased mortality in these populations.[7] However, with the increasing capabilities of machine learning (ML) analysis and its demonstrated efficacy in the prediction of patient outcomes, further analysis using this methodology is warranted to identify and mitigate the

variables associated with mortality in patients with AS or DISH.

ML is a widely utilized means of predictive analysis that can employ artificial intelligence to classify and quantify risk factors for a chosen clinical outcome. This methodology has been broadly incorporated into medical research to provide insight into a number of perioperative outcomes, including reoperation, discharge destination, and mortality.[8–11] Thus, similar utilization of this approach to examine mortality in patients with AS or DISH may be useful for expanding knowledge toward the variables that incur an increased risk of mortality in this unique subset of patients. As such, the purpose of this study was to conduct ML analysis in order to predict postoperative mortality in patients with AS or DISH using the Nationwide Inpatient Sample Healthcare Cost and Utilization Project (HCUP-NIS) database.

METHODS

After receiving a notice of exemption from our institution's review board, the HCUP-NIS database was queried using SAS 9.4 (Cary, NC) to identify adult patients carrying a diagnosis of AS or DISH who were admitted for spinal fractures and underwent subsequent fusion or corpectomy between 2016 and 2018. A full list of the International Classification of Diseases 10 revision (ICD-10) diagnostic and procedural codes used to identify this population can be found in the . Patients for which spinal fracture was not the primary cause of admission and those with malignancy were excluded from our analysis. Additionally, the use of the Agency for Healthcare Research and Quality Elixhauser Comorbidity Software Refined for the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), v2023.1, was used to identify comorbid conditions within our refined population.[12]

Three supervised ML classification algorithms were constructed in the Python programming language using the scikit-learn library.[13,14] These algorithms included Random Forest Classifier, Gradient Boosting Classifier, and Adaptive Boosting Classifier (ADAboost). These algorithms were tasked with predicting in-hospital mortality based on a given set of patient variables, including age, race, socioeconomic characteristics, comorbidities, and surgical factors.

Prior to analysis, preprocessing was performed using scikit-learn's StandardScaler to standardize variables across our cohort in order to bring all features to the same magnitude.[14] A train test split was performed using scikit-learn's train_test_split method in which 70% of our population's data was randomly chosen and used for training, and the remaining 30% was used for later testing of the model's performance.[14] Training of each algorithm involved application of sckit-learn's RandomizedSearchCV method along with a stratified 5-fold crossvalidation to determine optimal hyperparameters on the training data and ensure model generalizability.[14,15] Once the appropriate hyperparameters were determined, the final models were subsequently evaluated using the testing data to determine the model's performance.

The performance of the 3 ML models was then evaluated by a series of commonly used metrics, including classification accuracy, sensitivity, specificity, and area under receiver operating characteristics curve.[16–18] The Matplotlib library in Python was used for graphical visualization of the receiver operating characteristics produced by the three models.[19] Importance of each variable was quantified using two commonly used feature importance methods, namely, permutation feature importance (PFI) and Gini feature importance (GFI) through utilization of ELI5 library (version 0.11.0) and scikit-learn's feature_importance method, respectively.[14,20,21] PFI is generated by measuring the change in model performance when a single variable or a feature is removed, breaking the relationship between the feature and the predicted variable. Thus, a variable given a greater PFI value indicates a decrease in the predictive performance of the model when that feature is removed.[20,21] GFI quantifies the decrease in the Gini impurity index that is seen after a node split in an anglorithm's decision tree and utilizes this value as a measure of feature importance. Thus, features with larger decreases in impurity after a certain node split are deemed more important in predicting the outcome of interest.[14,21,22] As our study utilized multiple ML algorithms, the PFIs and GFIs generated by each algorithm for a given variable were averaged to yield a composite measure of their relative importance. Thus, variables with higher average PFI (aPFI) and average GFI (aGFI) values represented features that contributed significantly to the performance of multiple algorithms and thus carried greater overall value in producing effective predictions across these algorithms.

Statistical analysis was performed using SAS 9.4 (Cary, NC) with statistical significance defined as $P < 0.05$. Discharge weights were applied to calculate the nationally presentable frequency. Differences between categorical variables were assessed using the Pearson's χ^2 test, while numerical differences were assessed using independent sample t test. Categorical variables are

presented as frequencies in percentages, and continuous variables are presented as means and SDs.

RESULTS

Following a query of the HCUP-NIS database and application of our selected inclusion and exclusion criteria, 2960 patients were identified for further analysis within our study (Table 1). Of this cohort, 1890 and 890 patients featured a diagnosis of AS or DISH, respectively, while 180 patients carried both diagnoses. With regard to our primary outcome, an in-hospital mortality rate of 5.40% was observed in our selected population, including a rate of 6.35% in patients with AS, 2.81% in patients with DISH, and 8.33% in patients with both diagnoses. A full description of the clinical and demographic characteristics of this cohort is included in Table 1.

Each algorithm constructed for our analysis was then tasked with independently identifying predictors of in-hospital mortality among this cohort. In total, these algorithms yielded an average accuracy of 87.52% and generated predictions with an average sensitivity and specificity of 30.0% and 90.9%, respectively. With regard to predictive performance, the aggregate of our constructed algorithms produced an area under the curve (AUC) of 0.758, with the Adaptive Boosting classifier performing most effectively (AUC = 0.766), followed by the Random Forest (AUC = 0.756) and Gradient Boosting (AUC = 0.751) classifiers. The ROCs and information regarding individual performance of each algorithm is provided in Figure 1 and Table 2.

Upon comparison of the PFI values generated by these algorithms, a number of variables were distinguished as highly predictive of in-hospital mortality within our sample. Specifically, fractures of the thoracic spine (aPFI = 0.0843), complicated hypertension (aPFI = 0.0541), spinal cord injuries (aPFI = 0.0541), fractures of the cervical spine (aPFI = 0.0402), DISH affecting the thoracic spine (aPFI = 0.038), and age (aPFI = 0.034) each carried considerable importance while also demonstrating statistical significance among those experiencing in-hospital mortality. A record of the aPFI values and sample means for these variables can be found in Table 3.

Similarly, each algorithm also reported a number of statistically significant variables deemed important through the use of GFI. These included age (aGFI = 0.172), fractures of the cervical spine (aGFI = 0.065), spinal cord injuries (aGFI = 0.057), AS of the lumbar spine (aGFI = 0.044), complicated hypertension (aGFI = 0.037), perivascular disease (aGFI = 0.036), spinal

Table 1. Patient characteristics (n = 2960).

Characteristic	Mean ± SD or % (n)
Age, y, mean ± SD	73.13 ± 10.10
Sex	
Male	84.5% (2500)
Female	15.5% (460)
Race	
White	76.0% (2250)
African American	4.6% (135)
Hispanic	8.3% (245)
Asian/Pacific Islander	2.0% (60)
Native Americans	0.5% (15)
Others	2.4% (70)
Unknown	6.2% (185)
Patient Location	
"Central" counties of metro areas of ≥1 million population	29.5% (875)
"Fringe" counties of metro areas of ≥1 million population	24.2% (720)
Counties in metro areas of 250,000–999,999 population	20.0% (590)
Counties in metro areas of 50,000–249,999 population	7.1% (205)
Micropolitan counties	11.3% (335)
Not metropolitan or micropolitan counties	7.4% (220)
Unknown	0.5% (15)
Median Household Income	
0–25th percentile	21.8% (645)
26th–50th percentile	26.9% (795)
51st–75th percentile	26.2% (775)
76th–100th percentile	22.8% (675)
Unknown	2.4% (70)
Days From Admission to Procedure	2.36 ± 2.68
Primary Payer	
Medicare	71.7% (2125)
Medicaid	2.3% (65)
Primary insurance	20.6% (610)
Self-pay	0.3% (10)
No charge	0.0% (0)
Other	4.7% (140)
Unknown	0.3% (10)
Comorbidities	
Alcoholism	7.8% (230)
Diabetes complicated	23.0% (680)
Drug abuse	1.4% (40)
Hypertension complicated	25.0% (740)
Chronic lung condition	16.9% (500)
Obese	26.2% (775)
Perivascular disease	8.3% (245)
Fracture Level	
Cervical	31.9% (945)
Thoracic	57.2% (1695)
Lumbar	10.8% (320)
Surgical Procedure	
Fusion	
Cervical fusion	21.5% (635)
Cervicothoracic fusion	3.2% (95)
Thoracic fusion	44.6% (1320)
Thoracolumbar fusion	9.0% (265)
Lumbar fusion	3.7% (110)
Lumbosacral fusion	0.8% (25)
>2 levels	54.7% (1620)
Corpectomy	
Cervical corpectomy	3.7% (110)
Thoracic corpectomy	10.8% (320)
Lumbar corpectomy	2.2% (65)
>2 levels	0.2% (5)
Spinal cord injury	13.3% (395)
Ankylosing Spondylitis	
Cervical	13.3% (395)
Cervicothoracic	1.2% (35)

Table 1. Continued.

Characteristic	Mean ± SD or % (n)
Thoracic	24.8% (735)
Thoracolumbar	2.9% (85)
Lumbar	5.2% (155)
Multiple sites	4.2% (125)
Unspecified region	19.9% (590)
Ankylosing Hyperostosis (DISH)	
Cervical	7.8% (230)
Cervicothoracic	0.8% (25)
Thoracic	13.2% (390)
Thoracolumbar	1.7% (50)
Lumbar	2.0% (60)
Multiple sites	2.9% (85)
Unspecified region	6.1% (180)
Mixed (AS + DISH)	6.1% (180)
In-hospital deaths	5.4% (160)

Abbreviations: AS, ankylosing spondylitis; DISH, diffuse idiopathic skeletal hyperostosis.

fusion >2 levels (aGFI = 0.027), and Hispanic ethnicity (aGFI = 0.022). The GFI rankings produced by each algorithm as well as the sample means associated with each variable are available in Table 4.

Further comparison of the variables identified by both PFI and GFI allowed for isolation of clinical characteristics that were deemed important to predicting in-hospital mortality across both methodologies. Of the variables that were previously independently identified, cervical spine fractures, spinal cord injuries, and complicated hypertension carried statistical significance and predictive importance as quantified by both GFI and PFI.
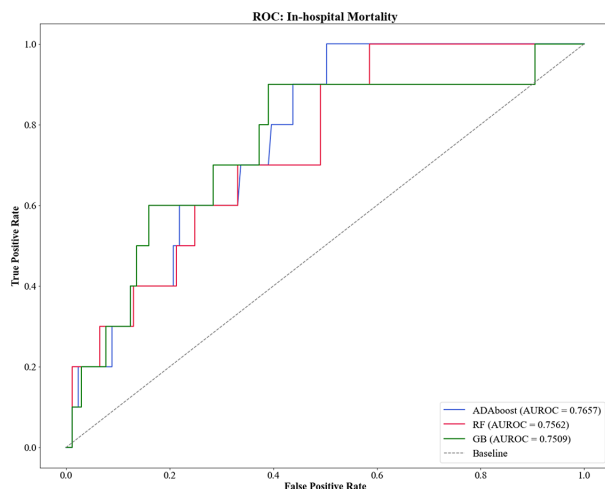

Figure. Graphical representation of each algorithm's area under the receiver operating characteristics curve (AUROC) in prediction of in-hospital mortality. Abbreviations: ADAboost, Adaptive Boosting Classifier; GB, Gradient Boosting Classifier; RF, Random Forest Classifier.

DISCUSSION

Application of ML analysis to the HCUP-NIS database identified a number of predictive variables for in-hospital mortality following spinal fracture fixation in patients with AS or DISH. Notably, age, thoracic spinal fractures, complicated hypertension, spinal cord injury, and cervical spinal fractures were among the greatest predictive importance across the ML algorithms utilized in our analysis. Additionally, fractures of the cervical spine, spinal cord injuries, and complicated hypertension were found to carry importance across both methods of feature importance quantification utilized in our analysis. While previous studies have documented several risk factors that contribute to in-hospital mortality in this population, the use of ML to validate these risk factors and their relative contributions to this outcome provides an additional level of understanding to this issue.

The results of this study serve to corroborate many of the variables that have been previously documented as risk factors for in-hospital mortality in patients with AS or DISH. In a review of in-hospital complications experienced by this population, Bernstein et al[23] identified age and cervical spine fractures as independent predictors of mortality, while spinal cord injuries were correlated with increased rate of complications throughout admission. Concurrently, studies by Ull et al and Robinson et al also reported spinal cord injuries, age, and increasing medical comorbidities including hypertension as significant predictors of mortality within the acute postoperative period.[24,25] In similarly identifying these variables through the use of ML-based predictive analysis, our study serves to substantiate the importance of these risk factors while also providing a measure of their relative contributions to producing in-hospital mortality. Furthermore, alignment of our results with those of prior studies demonstrates the utility of ML as an effective means of outcome prediction as previously reported variables were also deemed important by our study despite differences in methodology and data sources.

Knowledge of the variables identified by our analysis provides clinicians with validated and quantified predictors of mortality that may serve to guide perioperative decision-making in this high-risk patient population. For instance, the risk of mortality in patients with AS or DISH that feature complicated hypertension may be diminished if providers are able to promptly identify this risk factor and coordinate medical optimization throughout the perioperative period. Additionally, by providing a measure of each variable's

Table 2. Algorithm performance in predicting in-hospital mortality.

Algorithm	Accuracy	Sensitivity	Specificity	Area Under the Curve
Adaptive Boosting Classifier	79.88%	0.400	0.8225	0.7657
Random Forest	92.74%	0.300	0.9645	0.7562
Gradient Boosting	89.94%	0.200	0.9408	0.7509
Average	87.52%	0.300	0.9093	0.7576

relative importance, physicians may effectively triage the requirements of these at-risk and medically complex patients through identification of the characteristics most associated with short-term mortality. Furthermore, as nonoperative management of vertebral fractures in these populations has been found to carry a comparable risk of mortality to that of surgical intervention, the analysis provided by this study may be utilized to identify patients who are better suited for conservative treatment according to their unique profile of perioperative risk factors.[2] In this high-risk and medically complex cohort of patients, the decision for surgical vs nonoperative management is undoubtedly challenging and thus must be guided by a thorough consideration of relevant risk factors for mortality, including those identified by our analysis.

The findings produced by our analysis also serve to provide comparison of 2 distinct methods of generating feature importance measures within ML analysis. Interestingly, despite being applied to the same data source, the features identified by GFI and PFI differed considerably, with only 3 variables sharing common importance across these 2 methods. While this highlights the variation that may be present between GFI and PFI, it also serves to both identify a broader sample of predictive variables and more thoroughly validate those that were deemed predictive by both metrics. Similarly, this study's use of multiple independent ML algorithms adds to the value of our findings by ensuring that each variable reported was deemed important by several independent analyses rather than the predictions

of a single algorithm. Both aspects of our methodology serve to further support the importance of the variables identified in our analysis, in turn, supplying providers with the most pertinent predictors of mortality in our population of interest.

There are several limitations that must be recognized when considering the results of this study. Namely, the use of the HCUP-NIS database limits the conclusions that may be drawn from our analysis. For instance, although this database provides an extensive and widely utilized source of procedural outcomes, it is inherently incapable of capturing the entirety of our cohort of interest, thus limiting the generalizability of our findings. Similarly, the outcomes reported by this study are reflective solely of the complications occurring throughout the duration of hospital admission, thus limiting the span of our observation and analysis. As such, the risk factors exhibited by patients experiencing short-term postoperative mortality at home, in rehabilitation centers, or at nursing facilities may not be fully characterized by our study. Additionally, it is important to note that both PFI and GFI, while widely being the utilized methods of quantifying variable importance, are not necessarily reflective of the reality of clinical practice. Rather, each serves as an indirect measure of the importance of features within the context of a trained model.[26,27] Furthermore, both feature importance methods are subject to bias. Bias in PFI arises when features are highly correlated, leading to the permutation of one variable inadvertently impacting the predictive power of other highly correlated features, potentially leading to inaccurate importance measures.[26,27] GFI is also subject to bias

Table 3. Permutation feature importance for prediction of in-hospital death following surgical treatment of vertebral fractures in patients with AS and DISH.

Outcome and Features	Average Permutation Feature Importance	Mean ± SD or % (n)		P
		Survived to Discharge	In-Hospital Death	
Days from admission to surgery	0.0851	2.36 ± 2.72	2.14 ± 1.58	0.828
Thoracic fracture	0.0843	55.2% (1635)	2.0% (60)	<0.001
Hypertension with end-organ complications	0.0541	22.5% (665)	2.5% (75)	<0.001
Spinal cord injury	0.0541	11.5% (340)	1.9% (55)	<0.001
Cervical fracture	0.0402	29.4% (870)	2.5% (75)	<0.001
Age	0.0338	72.89 ± 10.17	77.34 ± 7.69	<0.001
Primary payer - Medicare	0.0221	67.1% (1985)	4.7% (140)	<0.001
Thoracic fusion	0.0212	42.3% (1250)	2.4% (70)	0.825
Median household income: 76th–100th percentile	0.0161	22.3% (660)	0.5% (15)	<0.001
Cervical fusion	0.0156	19.8% (585)	1.69% (50)	0.002

Table 4. Gini feature importance for prediction of in-hospital death following surgical treatment of vertebral fractures in patients with AS and DISH.

Outcome and Features	Average Gini Feature Importance	Mean ± SD or % (n)		P
		Survived to Discharge	In-Hospital Death	
Age	0.172	72.89 ± 10.17	77.34 ± 7.69	<0.001
Days from admission to surgery	0.101	2.36 ± 2.72	2.14 ± 1.58	0.828
Cervical fracture	0.0646	29.4% (870)	2.5% (75)	<0.001
Spinal cord injury	0.0567	11.5% (340)	1.9% (55)	<0.001
AS of lumbar spine	0.0443	4.4% (130)	0.8% (25)	<0.001
Hypertension with end-organ complications	0.0368	22.5% (665)	2.5% (75)	<0.001
Perivascular disease	0.0363	7.6% (225)	0.7% (20)	0.046
Obesity	0.0344	25.0% (740)	1.2% (35)	0.203
Fusion >2 evels	0.0272	52.2% (1545)	2.5% (75)	0.04
Hispanic ethnicity	0.0225	7.6% (225)	0.7% (20)	0.045

as it has a tendency to favor features with high cardinality or a large number of unique values, potentially leading to overestimation of their importance compared to other features.[26,27] Clinical decisions often involve complex interactions among multiple variables and considerations beyond predictive ability alone. These important measures can help to guide the understanding of feature relevance, but they should be considered as complementary tools rather than absolute determinants of clinical significance.

CONCLUSION

This study's application of ML algorithms to predict in-hospital mortality among patients with AS or DISH identified a number of clinical risk factors relevant to this outcome through analysis of the NIS-HCUP database. Namely, age, thoracic spinal fractures, complicated hypertension, spinal cord injury, and cervical spinal fractures each carried notable importance and statistical significance as independent predictors of mortality with this cohort of high-risk patients. These findings may serve to provide physicians with an awareness of risk factors for in-hospital mortality and, subsequently, strategies through which such variables may be mitigated.

ACKNOWLEDGMENTS

REFERENCES

1. Angelopoulou F, Kraniotis P, Daoussis D. DISH vs spondyloarthritides. *Mediterr J Rheumatol*. 2020;31(1):81–83. doi:10.31138/mjr.31.1.81

2. Westerveld LA, Verlaan JJ, Oner FC. Spinal fractures in patients with ankylosing spinal disorders: a systematic review of the literature on treatment, neurological status and complications. *Eur Spine J*. 2009;18(2):145–156. doi:10.1007/s00586-008-0764-0

3. Furukawa M, Okuyama K, Ninomiya K, et al. Association of continuous vertebral bone bridges and bone mineral density with the fracture risk in patients with diffuse idiopathic skeletal hyperostosis. *Asian Spine J*. 2022;16(1):75–81. doi:10.31616/asj.2020.0352

4. Pray C, Feroz NI, Nigil Haroon N. Bone mineral density and fracture risk in ankylosing spondylitis: a meta-analysis. *Calcif Tissue Int*. 2017;101(2):182–192. doi:10.1007/s00223-017-0274-3

5. Rustagi T, Drazin D, Oner C, et al. Fractures in spinal ankylosing disorders: a narrative review of disease and injury types, treatment techniques, and outcomes. *J Orthop Trauma*. 2017;31:S57–S74. doi:10.1097/BOT.0000000000000953

6. Badve SA, Bhojraj SY, Nene AM, et al. Spinal instability in ankylosing spondylitis. *Indian J Orthop*. 2010;44(3):270–276. doi:10.4103/0019-5413.65151

7. Robinson Y, Robinson A-L, Olerud C. Complications and survival after long posterior instrumentation of cervical and cervicothoracic fractures related to ankylosing spondylitis or diffuse idiopathic skeletal hyperostosis. *Spine (Phila Pa 1976)*. 2015;40(4):E227–E233. doi:10.1097/BRS.0000000000000726

8. Cabrera A, Bouterse A, Nelson M, et al. Accounting for age in prediction of discharge destination following elective lumbar fusion: a supervised machine learning approach. *Spine J*. 2023;23(7):997–1006. doi:10.1016/j.spinee.2023.03.015

9. Cabrera A, Bouterse A, Nelson M, et al. Use of random forest machine learning algorithm to predict short term outcomes following posterior cervical decompression with instrumented fusion. *J Clin Neurosci*. 2023;107:167–171. doi:10.1016/j.jocn.2022.10.029

10. Karhade AV, Shah AA, Bono CM, et al. Development of machine learning algorithms for prediction of mortality in spinal epidural abscess. *Spine J*. 2019;19(12):1950–1959. doi:10.1016/j.spinee.2019.06.024

11. Shan Z-M, Ren X-S, Shi H, et al. Machine learning prediction model and risk factor analysis of reoperation in recurrent lumbar disc herniation patients after percutaneous endoscopic lumbar discectomy. *Global Spine J*. 2023:21925682231173353. doi:10.1177/21925682231173353

12. Agency for Healthcare Research and Quality. Elixhauser Comorbidity Software Refined for ICD-10-CM. Healthcare Cost and Utilization Project (HCUP). October 2021. www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp.

13. RossumG, Drake FL. Python 3 reference manual. CreateSpace. 2009.

14. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–2830.

15. Agrawal T. Hyperparameter optimization using Scikit-learn. In: *Hyperparameter Optimization in Machine Learning*. Apress; 2021:31–51. doi:10.1007/978-1-4842-6579-6

16. Alam M, Rahman MS, Rahman MS. A random forest based predictor for medical data classification using feature ranking. *Informatics in Medicine Unlocked*. 2019;15:100180. doi:10.1016/j.imu.2019.100180

17. Erickson BJ, Kitamura F. Magician's corner: 9. Performance metrics for machine learning models. *Radiol Artif Intell*. 2021;3(3):e200126. doi:10.1148/ryai.2021200126

18. Ling CX, Huang J, Zhang H. AUC: A better measure than accuracy in comparing learning algorithms. In: Xiang Y, Chaib-draa B, eds. *Advances in Artificial Intelligence*. Vol Berlin Heidelberg: Springer; 2003:2671. 329–341. doi:10.1007/3-540-44886-1

19. Hunter JD. Matplotlib: A 2d graphics environment. *Comput Sci Eng*. 2007;9(3):90–95. doi:10.1109/MCSE.2007.55

20. Kaneko H. Cross-validated permutation feature importance considering correlation between features. *Analytical Science Advances*. 2022;3(9–10):278–287. doi:10.1002/ansa.202200018

21. Menze BH, Kelm BM, Masuch R, et al. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*. 2009;10:213. doi:10.1186/1471-2105-10-213

22. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. doi:10.1023/A:1010933404324

23. Bernstein DN, McCalla DJ, Molinari RW, Rubery PT, Menga EN, Mesfin A. An analysis of patient and fracture characteristics and clinical outcomes in patients with hyperostotic spine fractures. *Global Spine J*. 2020;10(8):964–972. doi:10.1177/2192568219887157

24. Robinson Y, Robinson A-L, Olerud C. Complications and survival after long posterior instrumentation of cervical and cervicothoracic fractures related to ankylosing spondylitis or diffuse idiopathic skeletal hyperostosis. *Spine*. 2015;40(4):E227. doi:10.1097/BRS.0000000000000726

25. Ull C, Yilmaz E, Hoffmann MF, et al. Factors associated with major complications and mortality during hospitalization in patients with ankylosing spondylitis undergoing surgical management for a spine fracture. *Global Spine J*. 2022;12(7):1380–1387. doi:10.1177/2192568220980702

26. Cava WL, Bauer C, Moore JH, Pendergrass SA. Interpretation of machine learning predictions for patient outcomes in electronic health records. *AMIA Annu Symp Proc*. 2019;2019:572–581.

27. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*. 2007;8:25. doi:10.1186/1471-2105-8-25

Declaration of Conflicting Interests: Dr. Danisa serves on the board of directors for the Musculoskeletal Transplantation Foundation. He performs consulting for Stryker and receives royalties from Globus Medical. Dr. Bono receives a stipend for being the editor in chief of The Spine Journal. He also consults for United Health Care and receives royalties from Wolters Kluwer and Elsevier. The remaining authors have no disclosures.

IRB Approval: IRB exemption was granted by our institutional review board. IRB #5230174

Corresponding Author: Olumide Danisa, Department of Orthopedics, Loma Linda University, Loma Linda, CA 92354, USA; odanisa@llu.edu