

Application of machine learning algorithms to predict 30-day hospital readmission following cement augmentation for osteoporotic vertebral compression fractures

Andrew Cabrera^a, Alexander Bouterse^a, Michael Nelson^a, Luke Thomas^a, Omar Ramos^b, Wayne Cheng^c, Olumide Danisa^{d,*}

^a School of Medicine, Loma Linda University, Loma Linda, CA, 92354, USA

^b Twin Cities Spine Center, Minneapolis, MN 55404, USA

^c Jerry L Pettis Memorial Veterans Hospital, Loma Linda, CA, 92354, USA

^d Department of Orthopedics, Loma Linda University, Loma Linda, CA, 92354, USA

ARTICLE INFO

Keywords:

Cement augmentation
Kyphoplasty
Machine learning
Readmission
Vertebroplasty

ABSTRACT

Objective: Osteoporosis is a common skeletal disease that greatly increases the risk of pathologic fractures and accounts for approximately 700,000 vertebral compression fractures (VCFs) annually in the United States. Cement augmentation procedures such as balloon kyphoplasty (KP) and percutaneous vertebroplasty (VP) have demonstrated efficacy in the treatment of VCFs, however, some studies report rates of readmission as high as 10.8% following such procedures. The purpose of this study was to employ Machine Learning (ML) algorithms to predict 30-day hospital readmission following cement augmentation procedures for the treatment of VCFs using the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database.

Methods: ACS-NSQIP was queried to identify patients undergoing either KP or VP from 2011 to 2014. Three ML algorithms were constructed and tasked with predicting post-operative readmissions within this cohort of patients. Results: Postoperative pneumonia, ASA Class 2 designation, age, partially-dependent functional status, and a history of smoking were independently identified as highly predictive of readmission by all ML algorithms. Among these variables postoperative pneumonia ($p < 0.01$), ASA Class 2 designation ($p < 0.01$), age ($p = 0.002$), and partially-dependent functional status ($p < 0.01$) were found to be statistically significant. Predictions were generated with an average AUC value of 0.757 and an average accuracy of 80.5%.

Conclusions: Postoperative pneumonia, ASA Class 2 designation, partially-dependent functional status, and age are perioperative variables associated with 30-day readmission following cement augmentation procedures. The use of ML allows for quantification of the relative contributions of these variables toward producing readmission.

1. Introduction

Osteoporosis is a common skeletal disease affecting more than 50 million people in the United States, with half of individuals older than 50 years demonstrating osteopenia or osteoporosis.¹ The progressive microarchitectural deterioration and loss of bone density seen in osteoporosis greatly increases the risk of pathologic fractures, particularly of the spine.² By age 50, the risk of osteoporotic vertebral compression fractures (VCFs) in women is as high as 50% and, in total, osteoporosis accounts for approximately 700,000 vertebral fractures annually in the United States.³ As such, VCFs account for a significant portion of the

morbidity associated with osteoporosis and pose a considerable challenge to both patients and providers.

Minimally invasive cement augmentation procedures such as balloon kyphoplasty (KP) and percutaneous vertebroplasty (VP) are viable options for the treatment of VCFs and are frequently utilized in patients that have failed non-operative treatment.⁴ While these procedures may provide improved quality of life, reduced back pain, and increased fracture stabilization in patients with VCFs, they are not without inherent risks.^{1,4,5} Of note, prior studies have recorded 30-day hospital readmission rates as high as 10.8% in patients undergoing KP or VP.⁶ While a number of variables including age, chronic obstructive

* Corresponding author. 25805 Barton Road Suite A106, Loma Linda, CA, 92354, USA.

E-mail address: odanisa@llu.edu (O. Danisa).

<https://doi.org/10.1016/j.wnsx.2024.100338>

Received 23 March 2023; Received in revised form 13 November 2023; Accepted 21 February 2024

Available online 2 March 2024

2590-1397/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

pulmonary disease (COPD), and chronic steroid use have previously been found to contribute to this outcome, the application of machine learning (ML) algorithms may provide knowledge toward the clinical variables that produce readmission.^{7,8}

ML, a subfield of artificial intelligence, allows for accurate and efficient predictions of clinically relevant outcomes through the analysis of large datasets. The capabilities of ML have been previously illustrated in studies predicting a number of pertinent patient outcomes including readmission, reoperation, and increased lengths of hospital stay.^{9–11} The extensive analysis provided by ML, as well as its ability to quantify relative variable importance, has led to its vast utilization throughout medical research as a means of outcome prediction.¹² The purpose of this study was to employ three ML algorithms to predict 30-day hospital readmission following cement augmentation procedures for the treatment of VCFs using the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database.

2. Methods

After receiving a waiver of the study protocol from the institutional review board, the ACS-NSQIP database was queried using Rstudio (RStudio, PBC, Boston, MA) to identify adult patients undergoing elective KP and VP from 2011 to 2014. Patients who underwent KP and VP in the ACS-NSQIP database were identified using Current Procedural Terminology (CPT) codes, including 22510, 22511, and 22512 for KP and 22520, 22521, and 22522 for VP. Patients undergoing non-elective or emergency procedures and those designated as having disseminated cancer or malignancy were excluded from our analysis to ensure non-osteoporotic VCF were not present among our population. Additionally, those aged less than 50 years, those with compression fractures greater than three levels, and those with International Classification of Diseases Ninth Revision (ICD9) codes for indications other than osteoporotic fractures were excluded from our population.

Patients matching our specified criteria were analyzed by three independent supervised ML classification algorithms, namely Support Vector Machine Classifier (SVM), Gaussian Naive Bayes Classifier (GNB), and Random Forest Classifier (RF). Support Vector Machine (SVM), Random Forest (RF), and Gaussian Naive Bayes (GNB) classifiers were chosen as they are commonly used and well-established methods of predictive binary classification within scientific literature. SVM functions by creating a hyperplane, effectively separating classes in feature space. This facilitates maximization of the margin between classes, thus allowing for effective performance in higher-dimensional datasets.¹³ RF algorithms employ an ensemble or group of decision trees, with each tree generating a prediction and the majority vote of these trees producing a final predicted outcome.¹⁴ GNB applies Bayes' theorem with the assumption of feature independence to estimate class probabilities.¹⁵ Thus, these algorithms represent a diverse approach to variable classification, which ultimately provides a more rigorous evaluation of a given dataset than would be produced by relying on a single technique or multiple algorithms with similar design. Each algorithm was constructed using the SciKit-Learn library in the python programming language and tasked with predicting 30-day readmission based on a given set of patient variables.^{16,17} Variable selection was influenced by presence of greater than 20% missing data within each variable and review of variable selection from previously published literature utilizing the ACS-NSQIP database to generate predictive modeling.^{6,7,9} These variables included demographic information, preoperative lab values, comorbidities and peri-operative complications (see Table 1).

Our patient population was systematically evaluated for missing data within each variable. To address these inconsistencies, missForest through utilization of the missForest package (Stekhoven, 2022) in the R statistical programming language (R Core Team, 2022) was applied for imputation of the missing data.^{18,19} Categorical variables (eg, Race, Sex, ASA Class) were preprocessed using SciKit-Learn's one-hot-encoder to transform each unique variable within a categorical feature into a

Table 1
Patient characteristics (n = 1870).

Characteristic	Mean (±SD) or Percentage (n)
Hospital Status	
Inpatient	25.19% (658)
Outpatient	64.81% (1212)
Age (yrs.)	75.417 ± 9.369
Sex	
Male	28.40% (531)
Female	71.60% (1339)
BMI	26.57 ± 5.74
Underweight (< 18.5 kg/m ²)	4.22% (79)
Normal weight (18.5 - 24.9 kg/m ²)	39.68% (742)
Overweight (25.0 - 29.9 kg/m ²)	31.98% (598)
Obese (≥30 kg/m ²)	24.12% (451)
Ethnicity	
Hispanic	5.78% (108)
Race	
American Indian or Alaska Native	0.16% (3)
Asian	3.96% (74)
Black or African American	1.55% (29)
Native Hawaiian or Pacific Islander	0.16% (3)
White	89.52% (1674)
Unknown/Not Reported	4.65% (87)
ASA Class	
Class I - No disturb	0.70% (13)
Class II - Mild disturb	26.15% (489)
Class III - Severe Disturb	64.33% (1203)
Class IV - Life Threat	8.72% (163)
Class V - Moribund	0.11% (2)
Preoperative Functional status	
Independent	88.98% (1664)
Partially Dependent	10.11% (189)
Totally Dependent	0.91% (17)
Transferred From	
Acute Care	1.44% (27)
Home	94.60% (1769)
Nursing Home/Chronic Care	2.99% (56)
Transferred From Other	0.96% (18)
Diabetes	
Requiring Insulin	8.02% (150)
Non-Insulin Dependent	9.57% (179)
Steroid Use	11.93% (223)
History of Smoking	12.46% (233)
History of COPD	15.56% (291)
History of CHF	2.03% (38)
Hypertension Requiring Medication	66.90% (1251)
Preoperative Renal Failure	0.16% (3)
Dialysis	0.75% (14)
Bleeding Disorder	7.59% (142)
Dyspnea	
At rest	1.39% (26)
Moderate exertion	11.44% (214)
Preoperative Transfusion	0.53% (10)
>10lbs weight loss in the last 3 months	1.18% (22)
Lab Values	
Preoperative Sodium (mmol/L)	138.096 ± 3.497
Preoperative BUN (mg/dl)	18.437 ± 8.831
Preoperative Creatinine (mg/dl)	0.943 ± 0.587
Preoperative WBC (x10 ³ /ul)	7.640 ± 2.638
Preoperative HCT (%)	37.803 ± 4.988
Preoperative Platelet Count (x10 ³ /ul)	245.349 ± 81.667
Operative Time (min.)	33.102 ± 8.761
Vertebroplasty	
Thoracic	3.90% (73)
Lumbar	5.35% (100)
Kyphoplasty	
Thoracic	90.75% (1697)
Lumbar	43.48% (813)
Lumbar	47.27% (884)
Levels	
1 Level	82.03% (1534)
2 Level	15.61% (292)
3 Level	2.35% (44)
Intraoperative/postoperative Transfusion	0.59% (11)
Total Length of Hospital Stay (days)	2.415 ± 5.729
Discharge Destination	
Home	84.06% (1572)

(continued on next page)

Table 1 (continued)

Characteristic	Mean (±SD) or Percentage (n)
Non-Home	15.94% (298)
Minor complication	
Urinary Tract Infection	2.19% (41)
Pneumonia	1.55% (29)
Progressive Renal Insufficiency	0.21% (4)
Superficial SSI	0.05% (1)
Wound Dehiscence	0.11% (2)
Major Complications	
Deep Vein Thrombosis	0.43% (8)
Pulmonary Embolism	0.43% (8)
Sepsis	0.43% (8)
Septic Shock	0.43% (8)
Renal Failure Requiring Dialysis	0.16% (3)
Cardiac Arrest	0.27% (5)
Myocardial Infarction	0.27% (5)
Deep Wound Infection	0.05% (1)
Organ Space Surgical Space Infection	0
CVA/Stroke with Neurologic Deficit	0.05% (1)
Readmission	8.88% (166)

unique binary column. Additionally, SciKit-Learn’s StandardScaler was applied to our dataset, bringing all features to a mean of 0 and standard deviation of 1, thus standardizing our population’s data.¹⁶ A train test split was performed using Scikit-Learn’s `train_test_split` method in which 80% of our populations data was used for training and the remaining 20% used for later testing of the model’s performance.^{16,20} Skit-Learn’s `RandomizedSearchCV` and `StratifiedKFoldCV` methods were applied in the construction of each model to determine optimal hyperparameters through utilization of a stratified tenfold cross-validation process.^{16,21} This stratified tenfold cross-validation method ensured the training data was randomly divided into ten subsets with equal class distribution. Nine of the ten subsets were used in the training process, with the remaining subset used for model validation. This process was repeated ten times, with each of the ten subsets being used for validation to ensure model generalizability. Once the appropriate hyperparameters were chosen, the final model was evaluated using the testing data obtained from the train test split to determine the model’s performance. Average cross validation Area Under Receiver Operating Characteristics Curve (AUROC) scores from the stratified tenfold cross validation process were obtained from each model and compared to the AUCs generated on the testing data. This process was used to ensure models were not over-fitting, a process where models perform well on training data, but are unable to generalize and perform appropriately on unseen testing data. Through utilization of the ELI5 library (version 0.11.0), importance of each variable was quantified based on permutation feature importance (PFI).^{22,23} PFI is generated by randomly omitting or shuffling a single variable and assessing the variation in model performance. Thus, this process removes the relation between the variable and the predicted outcome. As a result, a decline in model performance correlates to the extent to which the model relied upon the variable for its prediction.

Classification accuracy, sensitivity, specificity, and AUROC, commonly used metrics for ML performance evaluation, were used to evaluate each of our ML models.^{24–26} Graphic visualization of the ROCs produced by each model was performed using the Matplotlib library.²⁷ AUROC is widely regarded as a valuable metric for the evaluation of a classification model’s performance, as reliance on accuracy to evaluate performance can be a deceptive metric in imbalanced classification predictions.^{25, 26} The PFIs from each model were then compared to identify variables that were deemed predictive across the three ML algorithms. Commonly identified variables were then ranked based on the number of algorithms for which each variable was considered to be within the top 20 most predictive variables. Those that were featured in the top 20 most predictive variables by each algorithm independently can be seen in the supplementary file. Python files of the code used to generate the results in this manuscript can be made available from the corresponding author upon reasonable request.

Statistical analysis utilized SPSS version 28 (IBM Corporation, 2021, Armonk, NY, USA) with statistical significance defined as $p < 0.05$. Descriptive statistics utilized percentages, mean, and standard deviations (SD). Categorical differences between groups were computed using Pearson’s Chi-Square test or Fischer’s exact test when conditions for Chi-Square test were not met. For univariate analysis, independent sample *t*-tests with Levene’s test for equality of variance were used to compare numerical differences between groups.

3. Results

Initial filtering of the ACS-NSQIP dataset using relevant CPT codes identified 2560 patients undergoing KP or VP within our selected study period. Additional selection using ICD9 diagnosis codes and application of our exclusion criteria refined this cohort to 1870 patients for final review. Among those identified through our selection process, 1697 (90.7%) and 173 (9.3%) underwent KP and VP, respectively. The demographic and clinical characteristics of this population are included in Table 1.

The cohort included in our review was independently analyzed by the RF, GNB, and SVM algorithms, with each algorithm tasked with identifying patient variables that were predictive of postoperative readmission. The aggregate of these algorithms generated predictions with an average AUC of 0.757, with the RF algorithm demonstrating the highest efficacy in outcome prediction. The combined accuracy of these algorithms was 80.5% and predictions were made with a sensitivity of 56.3% and a specificity of 83.0%. For a summary of performance metrics generated by each algorithm and their plotted AUCs, please see Table 2 and Fig. 1. For details pertaining to model training using the stratified tenfold cross-validation with `RandomizedSearchCV` along with additional performance metrics generated by each algorithm upon testing, please refer to both [supplementary file Table 1](#) and [supplementary file Table 2](#) respectively.

The RF algorithm identified preoperative hematocrit, a history of COPD, postoperative pneumonia, total length of stay, and inpatient status as the perioperative variables most predictive of readmission within 30 days of undergoing KP or VP (Supplementary File: Table 3). The predictions provided by this algorithm were generated with an accuracy of 79.7%, a sensitivity of 59.4%, and specificity of 82.5%. Among the algorithms utilized in our analysis, the RF algorithm performed at the highest level, producing an AUROC of 0.775 (Table 2).

The analysis generated by the GNB algorithm identified a history of COPD, a diagnosed bleeding disorder, an American Society of Anesthesiologists (ASA) class 4 designation, postoperative pneumonia, and a history of smoking as variables predictive of a 30-day readmission (Supplementary File: Table 4). The GNB algorithm generated predictions with an AUROC of 0.758 as well as an accuracy of 83.2%, a sensitivity of 53.1%, and a specificity of 86.0% (Table 2).

The SVM algorithm identified postoperative pneumonia, partially dependent functional status, use of MAC anesthesia, preoperative transfer from a nursing home, and discharge destination as predictive variables for postoperative readmission (Supplementary File: Table 5). Predictions generated by this algorithm were made with an accuracy of 78.6%, a sensitivity of 56.3%, and a specificity of 80.7%, however the algorithm’s overall performance was comparatively poor, producing an AUROC of 0.737 (Table 2).

Table 2
Performance of algorithms in the prediction of 30-day readmission following cement augmentation.

Algorithm	Accuracy	Sensitivity	Specificity	AUC
Random Forest	79.68%	0.5938	0.8246	0.7753
Gaussian Naive Bayes	83.16%	0.5313	0.8596	0.7578
Support Vector Machine	78.61%	0.5625	0.8070	0.7372
Average	80.48%	0.5625	0.8304	0.7567

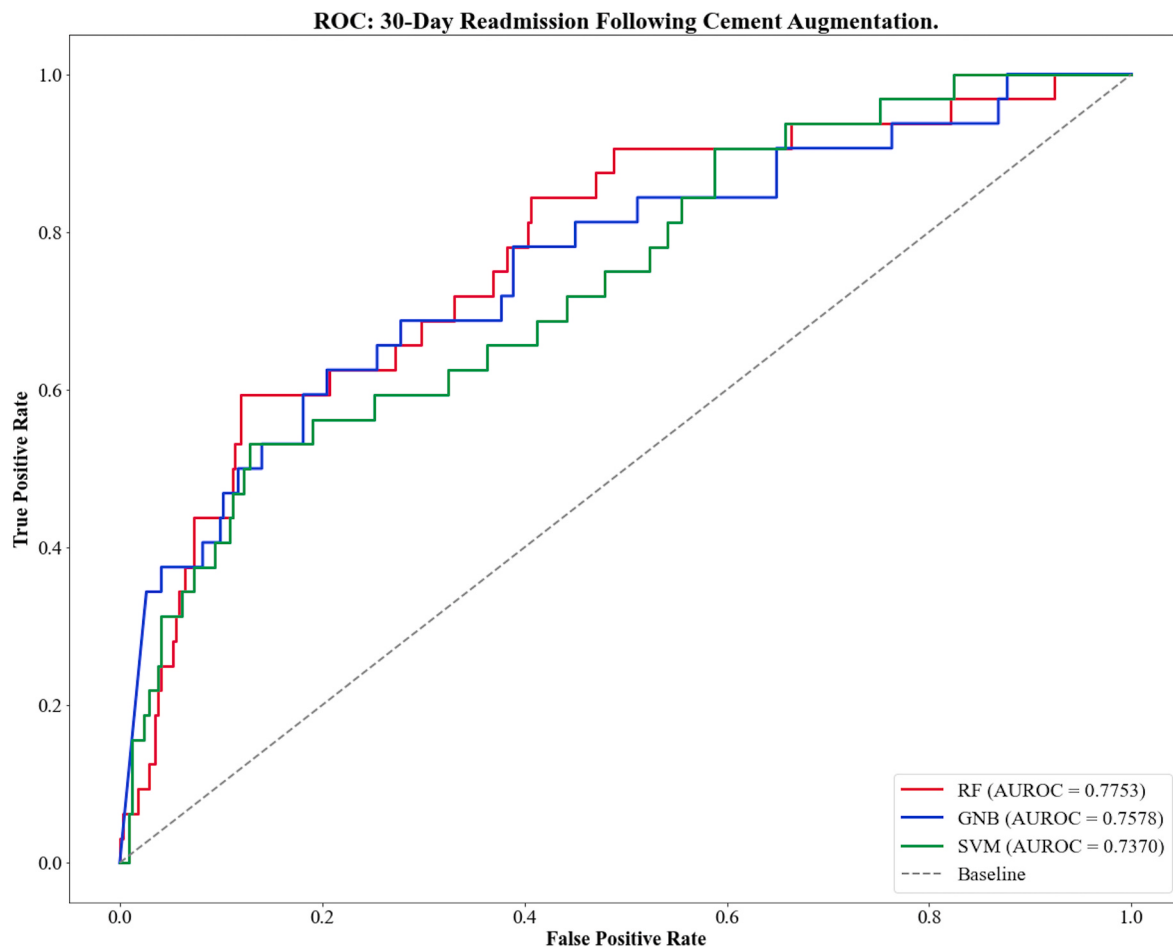


Fig. 1. Auroc of machine learning algorithm generated prediction of 30-day readmission following cement augmentation.

Upon comparison and statistical analysis of the predictive variables identified by each algorithm, several variables were identified as being highly predictive of 30-day readmission across all three algorithms (Fig. 2). These variables include postoperative pneumonia, ASA Class 2 designation, age, partially-dependent functional status, and a history of smoking. Among these variables postoperative pneumonia ($p < 0.001$), ASA Class 2 designation ($p < 0.001$), age ($p = 0.002$), and partially-dependent functional status ($p < 0.01$) were found to be statistically significant (Table 3).

4. Discussion

This study's application of ML-based data analysis to the ACS-NSQIP database identified a number of perioperative variables that were predictive of 30-day readmission for patients undergoing cement augmentation procedures. Among the variables identified by our methodology, postoperative pneumonia, ASA Class 2 designation, age, partially dependent functional status, and a history of smoking were the most frequently cited predictors of readmission across three independent ML algorithms. This study is the first to utilize multiple ML algorithms to document these risk factors and their relative contributions to predicting 30-day readmission following KP and VP.

Knowledge of the clinical variables highlighted by our analysis may aid providers in accurately identifying and mitigating risk factors for readmission within this surgical population. Due to the progressive nature of osteoporosis, patients requiring cement augmentation for VCFs are often elderly and may feature an increased number of medical comorbidities – both of which entail increased risk for perioperative complications, such as readmission.⁴ In conjunction with the fact that

hospital readmission is associated with increased levels of morbidity and mortality, it is critical that surgeons be able to identify risk factors for this outcome in this frail and high-risk population.^{28–30} This study, as well as other investigations into the outcomes of cement augmentation, report rates of readmission approaching 10%, which further calls attention to the need for accurate and efficient identification of patient risk factors. As such, the results of this study may prove useful to providers seeking to minimize this complication, as our analysis provides a series of highly-predictive risk factors for readmission related to this procedure. Concurrently, application of study findings to clinical practice may allow for the anticipation of postoperative readmission and, subsequently, preemptive implementation of strategies designed to reduce the risk of morbidity and mortality upon readmission.

The results of our analysis concur with those of previous studies that have reported on factors contributing to readmission in cement augmentation. Using the ACS-NSQIP database, Choo et al investigated risk factors for 30-day readmission following KP or VP.⁷ Their analysis identified factors such as age, ASA classification ≥ 2 , and a history of COPD as being individually associated with increased rates of readmission. Unlike Choo et al's use of ASA Class greater than 2 as a threshold, this study utilized each ASA class as separate binary variables allowing for examination of their individual contributions to the risk of readmission. Despite differing methodologies for handling of ASA Classes, ASA Class 2–4 were found to be statistically significant and were identified as potential risk factors by at least one of the algorithms (see Table 3). Additionally, in their analysis of risk factors for 30-day complications, functional dependence, inpatient admission status, and a history of COPD were each found to incur increased rates of cardiovascular, infectious, respiratory, and wound-related complications –

Table 3

Permutation feature importance for prediction of 30-day readmission following cement augmentation.

Outcome and Features	Times ranked in top 20 features of importance	Average Importance	Mean $\pm$ SD or % Sample (n) – Readmission	Mean $\pm$ SD or % Sample (n) – No Readmission	p
<i>Pneumonia</i>	3	0.0430	0.86% (16)	0.70% (13)	<0.001
<i>Partially Dependent Functional Status</i>	3	0.0271	1.66% (31)	8.45% (158)	<0.001
<i>Age (yrs.)</i>	3	0.0177	77.45 $\pm$ 8.201	75.22 $\pm$ 9.454	0.003
<i>Asa Class 2</i>	3	0.0174	0.96% (18)	25.19% (471)	<0.001
<i>History of Smoking</i>	3	0.0089	1.39% (26)	11.07% (207)	0.191
<i>History of COPD</i>	2	0.0315	2.73% (51)	12.83% (240)	<0.001
<i>Preoperative HCT</i>	2	0.0207	36.58 $\pm$ 5.57	37.92 $\pm$ 4.91	<0.001
<i>Asa Class 4</i>	2	0.0196	1.44% (27)	7.27% (136)	<0.001
<i>Bleeding Disorder</i>	2	0.0175	1.22% (21)	6.47% (121)	0.01
<i>MAC/IV Sedation</i>	2	0.0172	0.91% (17)	10.43% (195)	0.641
<i>Preoperative WBC Count ($\times 10^3$/ul)</i>	2	0.0123	8.43 $\pm$ 3.29	7.56 $\pm$ 2.55	<0.001
<i>Total Hospital Length of Stay (days)</i>	2	0.0116	3.60 $\pm$ 4.80	2.30 $\pm$ 5.80	0.005
<i>Independent Functional Status</i>	2	0.0084	7.22% (135)	81.76% (1529)	<0.001
<i>Dyspnea on Moderate Exertion</i>	2	0.0060	1.71% (32)	9.73% (182)	<0.001
<i>Asian</i>	2	0.0050	0.16% (3)	3.80% (71)	0.137
<i>Black or African American</i>	2	0.0022	0.05% (1)	1.50% (28)	0.509
<i>BMI (kg/m^2)</i>	2	0.0019	26.53 $\pm$ 6.68	26.57 $\pm$ 5.64	0.922
<i>Discharge Destination</i>	1	0.0097	7.11% (133)	76.95% (1439)	0.146
<i>Inpatient Hospital Status</i>	1	0.0085	4.44% (83)	30.75% (575)	<0.001
<i>Transferred from Outside Emergency Department</i>	1	0.0066	0.16% (3)	0.64% (12)	0.142
<i>Myocardial Infarction</i>	1	0.0061	0.16% (3)	0.11% (2)	0.006
<i>Preoperative Sodium (mmol/L)</i>	1	0.0058	137.76 $\pm$ 4.04	138.13 $\pm$ 3.44	0.197
<i>Preoperative Platelet Count ($\times 10^3$/ul)</i>	1	0.0056	245.34 $\pm$ 96.55	245.35 $\pm$ 80.10	0.999
<i>Sex (male)</i>	1	0.0051	2.41% (45)	25.99% (486)	0.7
<i>Preoperative BUN (mg/dl)</i>	1	0.0045	19.57 $\pm$ 10.13	18.33 $\pm$ 8.69	0.083
<i>Transferred from Home</i>	1	0.0039	8.24% (154)	86.36% (1615)	0.275
<i>Non-Insulin Dependent Diabetes</i>	1	0.0024	0.64% (12)	8.93% (167)	0.282
<i>Hispanic Ethnicity</i>	1	0.0021	0.32% (6)	5.45% (102)	0.211
<i>Asa Class 3</i>	1	0.0020	6.42% (120)	57.91% (1083)	0.025
<i>Preoperative Creatinine (mg/dl)</i>	1	0.0019	1.05 $\pm$ 0.83	0.93 $\pm$ 0.56	0.015
<i>Totally Dependent Functional Status</i>	1	0.0015	(0)	0.91% (17)	0.391
<i>Insulin Dependent Diabetes</i>	1	0.0014	0.70% (13)	7.33% (137)	0.925
<i>Operative Time (mins.)</i>	1	0.0014	29.88 $\pm$ 7.31	33.42 $\pm$ 9.30	0.13
<i>Transferred from Chronic Care Facility</i>	1	0.0135	0.43% (8)	2.57% (48)	0.151
<i>Caucasian or White</i>	1	0.0002	8.34% (156)	81.18% (1518)	0.5
<i>General Anesthesia</i>	1	0.0001	7.97% (149)	80.27% (1501)	0.523
<i>Dyspnea at Rest</i>	1	0.0001	0.37% (7)	1.02% (19)	0.006
<i>Hypertension Requiring Medication</i>	1	0.0001	5.88% (110)	61.02% (1141)	0.856

factors which may inherently contribute to the likelihood of readmission.^{7,31} Similarly, Segal et al applied the 5-item modified Frailty Index (5i-mFI) to the ACS-NSQIP database in order to identify risk factors for 30-day postoperative complications following KP.³² Their analysis found that increasing frailty on the 5i-mFI, an index that accounts for patient-specific factors such as COPD and functional dependence, was predictive of 30-day readmission. Lastly, Toy et al utilized ACS-NSQIP to identify a history of pulmonary disease and inpatient status as independent risk factors for readmission in KP or VP procedures.⁶ In light of the commonalities between these findings and those produced by our analysis, the results of our study are able to lend validity to these established risk factors and further describe their contributions toward generating adverse postoperative outcomes.

While prior studies investigating this topic have previously identified risk factors for 30-day readmission, the application of ML algorithms provides a further level of insight by quantifying the relative importance of each variable in predicting our selected outcome. By utilizing PFI to provide a measure of each variable's contribution to an algorithm's predictive analysis, clinicians may be better equipped to recognize the clinical factors that carry comparatively greater significance in producing postoperative readmissions. Prior application of this methodology has been used to predict a number of perioperative outcomes across spine surgery and has become increasingly incorporated into clinical risk stratification and decision-making algorithms.^{33–36} Additionally, this study provides further validation of the variables contributing to

readmission by utilizing three ML algorithms to isolate factors that were deemed predictive by multiple independent analyses. Through this method, variables that consistently hold significance in the prediction of readmission may be identified and incorporated into perioperative decision-making. Furthermore, as the variables highlighted by our analysis concur with the findings of previous authors, our study serves to reinforce the efficacy of ML models in accurately and efficiently analyzing large volumes of data. The ML algorithms that were utilized in this study produced performance metrics similar to those of studies that have previously applied ML to cement augmentation procedures, thus demonstrating a comparable level of predictive efficacy within our analysis.^{37,38}

Several limitations must be recognized when considering the results of this study. Namely, the use of the ACS-NSQIP database to perform predictive analysis inherently limits the scope to which our findings may be applied. Although the data provided by this resource is extensive, it evidently does not represent the entirety of clinical practice and the complexity of our population of interest. Further contributing to this limitation, our study draws from the data provided by the 2011–2014 ACS-NSQIP datasets, thus potentially limiting its applicability to current practice. Additionally, as the outcomes reported by ACS-NSQIP are confined to within 30 days of our chosen procedure, the conclusions that are able to be drawn from our study are limited and may not necessarily hold true beyond this time frame. Furthermore, as with all database research, the data utilized for our analysis is subject to potential

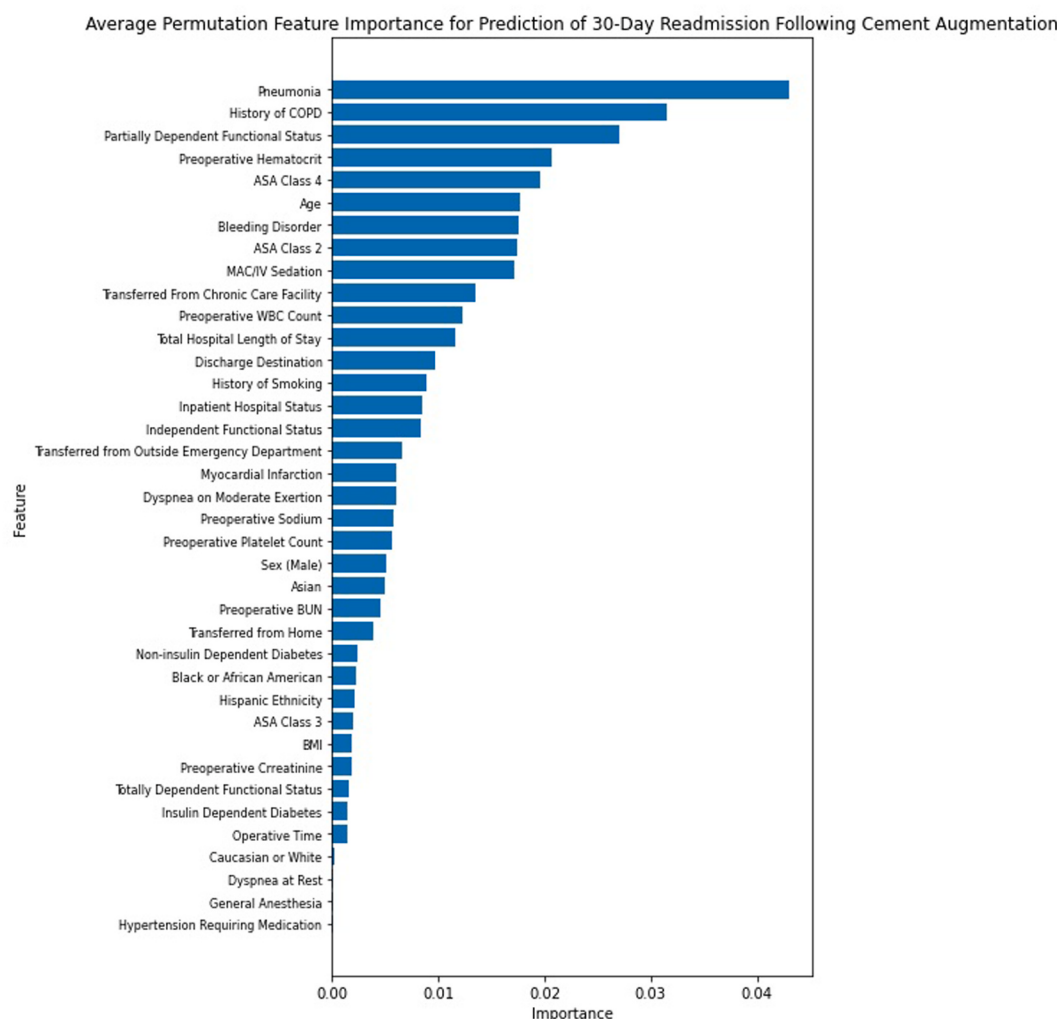


Fig. 2. Permutation feature importance in prediction of readmission following cement augmentation.

confounders as errors may occur during the process of encoding clinical encounters. It must also be qualified that ACS-NSQIP and other large repositories of patient data and outcomes frequently contain missing data points within a given sample.³⁹ For instance, the cohort examined in this study featured certain variables that were missing data in approximately 20% of patients. This missing data was, in part, accounted for through the utilization of missForest, a mixed variable imputation method that has demonstrated a high level of performance in datasets containing complex interaction and non-linear relationships.¹⁸ Additionally, despite differences in both prevalence and procedural technique, VP and KP are treated as equivalent within this study in order to provide a comprehensive description of the outcomes pertaining to cement augmentation procedures. While this approach has been employed in previous studies, it may inherently lack the same detail that would be provided by individually analyzing these two procedures.⁴⁰ It is also important to qualify that while PFI may be utilized as means of demonstrating a variable's clinical importance, it does not directly represent the actual importance of any one risk factor. Rather, PFI records the variation in algorithmic model performance when a given variable is omitted or shuffled from the available data, thus providing an indirect measure of that features importance to generating accurate predictions of an outcome.⁴¹ Finally, it must be acknowledged that, though ML and other AI-based methods of analysis can provide notable insight into clinically-relevant outcomes, its utility must not supersede a provider's clinical judgment and the consideration of the psychosocial, spiritual, and cultural factors that inherently influence patient care.

5. Conclusions

This study presents a number of risk factors for 30-day readmission following cement augmentation procedures as identified and quantified by three independent ML algorithms. Commonly documented risk factors include postoperative pneumonia, ASA Class 2 designation, age, partially-dependent functional status, and a history of smoking, as well as several other highly-predictive variables identified by individual algorithms. The variables recorded by this study align with those of previous studies and serve to add further insight into these factors through the use of ML analysis and PFI. Application of study findings may assist providers in identifying and mitigating risk factors for readmission in high-risk patients undergoing KP or VP.

Sources of support

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Andrew Cabrera: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Alexander Bouterse:** Investigation, Methodology, Writing – original draft, Writing – review & editing. **Michael Nelson:** Investigation, Writing – original draft, Writing – review & editing. **Luke Thomas:**

Investigation, Methodology, Writing – original draft, Writing – review & editing. **Omar Ramos:** Conceptualization, Project administration, Supervision, Writing – review & editing. **Wayne Cheng:** Conceptualization, Data curation, Investigation, Methodology, Supervision, Writing – review & editing. **Olumide Danisa:** Conceptualization, Data curation, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviations

ACS-NSQIP	American College of Surgeons, National Surgical Quality Improvement Program
ANOVA	Analysis of Variance
ASA	American Society of Anesthesiologists
AUC	Area Under the Curve;
COPD	Chronic Obstructive Pulmonary Disease
CPT	Current Procedural Terminology
GNB	Gaussian Naive Bayes Classifier
KP	Kyphoplasty
ML	Machine Learning
PFI	Permutation Feature Importance
RF	Random Forest Classifier
SD	Standard Deviation
SVM	Support Vector Machine Classifier
VCF	Vertebral Compression Fracture
VP	Vertebroplasty
5i-mFI	5-item modified Frailty Index

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.wnsx.2024.100338>.

References

- Chandra RV, Maingard J, Asadi H, et al. Vertebroplasty and kyphoplasty for osteoporotic vertebral fractures: what are the latest data? *Am J Neuroradiol*. 2018;39(5):798–806. <https://doi.org/10.3174/ajnr.A5458>.
- Dennison E, Cooper C. Epidemiology of osteoporotic fractures. *Horm Res Paediatr*. 2000;54(Suppl. 1):58–63. <https://doi.org/10.1159/000063449>.
- Cauley JA, Hochberg MC, Lui L-Y, et al. Long-term risk of incident vertebral fractures. *JAMA*. 2007;298(23):2761. <https://doi.org/10.1001/jama.298.23.2761>.
- McCarthy J, Davis A. Diagnosis and management of vertebral compression fractures. *Am Fam Physician*. 2016;94(1):44–50.
- Boss S, Srivastava V, Anitescu M. Vertebroplasty and kyphoplasty. *Phys Med Rehabil Clin*. 2022;33(2):425–453. <https://doi.org/10.1016/j.pmr.2022.01.008>.
- Toy JO, Basques BA, Grauer JN. Morbidity, mortality, and readmission after vertebral augmentation: analysis of 850 patients from the American College of surgeons national surgical quality improvement Program database. *Spine*. 2014;39(23):1943–1949. <https://doi.org/10.1097/BRS.0000000000000563>.
- Choo S, Malik AT, Jain N, Yu E, Kim J, Khan SN. 30-day adverse outcomes, readmissions and mortality following vertebroplasty/kyphoplasty. *Clin Neurol Neurosurg*. 2018;174:129–133. <https://doi.org/10.1016/j.clineuro.2018.08.014>.
- Bernatz JT, Anderson PA. Thirty-day readmission rates in spine surgery: systematic review and meta-analysis. *Neurosurg Focus*. 2015;39(4):E7. <https://doi.org/10.3171/2015.7.FOCUS1534>.
- 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. <https://doi.org/10.1016/j.jocn.2022.10.029>.
- Le Lay J, Alfonso-Lizarazo E, Augusto V, et al. Prediction of hospital readmission of multimorbid patients using machine learning models. *PLoS One*. 2022;17(12), e0279433. <https://doi.org/10.1371/journal.pone.0279433>.
- Yagi M, Michikawa T, Yamamoto T, et al. Development and validation of machine learning-based predictive model for clinical outcome of decompression surgery for lumbar spinal canal stenosis. *Spine J*. 2022;22(11):1768–1777. <https://doi.org/10.1016/j.spinee.2022.06.008>.
- Shamout F, Zhu T, Clifton DA. Machine learning for clinical outcome prediction. *IEEE Reviews in Biomedical Engineering*. 2021;14:116–126. <https://doi.org/10.1109/RBME.2020.3007816>.
- Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20:273–297. <https://doi.org/10.1007/BF00994018>.
- Breiman L. Random forests. *Mach Learn*. 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
- Zhang H. *The Optimality of Naive Bayes*. The Florida AI Research Society; 2004.
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12(null):2825–2830.
- Van Rossum G, Drake FL. *Python 3 reference manual*. CreateSpace; 2009.
- Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112–118. <https://doi.org/10.1093/bioinformatics/btr597>.
- Ambler G, Omar RZ, Royston P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. *Stat Methods Med Res*. 2007;16(3):277–298. <https://doi.org/10.1177/0962280206074466>.
- Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. *BMC Med Genom*. 2011;4(1):31. <https://doi.org/10.1186/1755-8794-4-31>.
- Agrawal T. Hyperparameter optimization using scikit-learn. In: Agrawal T, ed. *Hyperparameter Optimization in Machine Learning*. 2021:31–51. https://doi.org/10.1007/978-1-4842-6579-6_2. Apress.
- Kaneko H. Cross-validated permutation feature importance considering correlation between features. *Analytical Sci Adv*. 2022;3(9–10):278–287. <https://doi.org/10.1002/ansa.202200018>.
- 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 Bioinf*. 2009;10(1):213. <https://doi.org/10.1186/1471-2105-10-213>.
- Alam Md Z, Rahman MS, Rahman MS. A Random Forest based predictor for medical data classification using feature ranking. *Inform Med Unlocked*. 2019;15, 100180. <https://doi.org/10.1016/j.imu.2019.100180>.
- Erickson BJ, Kitamura F. Magician's corner: 9. Performance metrics for machine learning models. *Radiology: Artif Intell*. 2021;3(3), e200126. <https://doi.org/10.1148/ryai.2021200126>.
- 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. 2671. Springer Berlin Heidelberg; 2003:329–341. https://doi.org/10.1007/3-540-44886-1_25.
- Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3): 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- Visade F, Babykina G, Puisieux F, et al. Risk factors for hospital readmission and death after discharge of older adults from acute geriatric units: taking the rank of admission into account. *Clin Interv Aging*. 2021;16:1931–1941. <https://doi.org/10.2147/CIA.S327486>.
- Shaw JA, Stiliannoudakis S, Kaiser R, Layman E, Sima A, Ali A. Thirty-day hospital readmissions: a predictor of higher all-cause mortality for up to two years. *Cureus*. 2020. <https://doi.org/10.7759/cureus.9308>.
- Fluck D, Murray P, Robin J, Fry CH, Han TS. Early emergency readmission frequency as an indicator of short-, medium- and long-term mortality post-discharge from hospital. *Internal and Emergency Medicine*. 2021;16(6):1497–1505. <https://doi.org/10.1007/s11739-020-02599-3>.
- Piper K, DeAndrea-Lazarus I, Algattas H, et al. Risk factors associated with readmission and reoperation in patients undergoing spine surgery. *World Neurosurg*. 2018;110:e627–e635. <https://doi.org/10.1016/j.wneu.2017.11.057>.
- Segal DN, Wilson JM, Staley C, Michael KW. The 5-item modified frailty index is predictive of 30-day postoperative complications in patients undergoing kyphoplasty vertebral augmentation. *World Neurosurg*. 2018;116:e225–e231. <https://doi.org/10.1016/j.wneu.2018.04.172>.
- Elsamadicy AA, Koo AB, Reeves BC, et al. Utilization of machine learning to model important features of 30-day readmissions following surgery for metastatic spinal column tumors: the influence of frailty. *Global Spine J*. 2022;219256822211380. <https://doi.org/10.1177/21925682221138053>.
- Martini ML, Neifert SN, Gal JS, Oermann EK, Gilligan JT, Caridi JM. Drivers of prolonged hospitalization following spine surgery: a game-theory-based approach to explaining machine learning models. *J Bone Joint Surg*. 2021;103(1):64–73. <https://doi.org/10.2106/JBJS.20.00875>.
- Shah AA, Devana SK, Lee C, et al. Machine learning-driven identification of novel patient factors for prediction of major complications after posterior cervical spinal fusion. *Eur Spine J*. 2022;31(8):1952–1959. <https://doi.org/10.1007/s00586-021-06961-7>.
- Giordano C, Brennan M, Mohamed B, Rashidi P, Modave F, Tighe P. Accessing artificial intelligence for clinical decision-making. *Front Digital Health*. 2021;3, 645232. <https://doi.org/10.3389/fgdh.2021.645232>.
- Dong S, Zhu J, Yang H, Huang G, Zhao C, Yuan B. Development and internal validation of supervised machine learning algorithm for predicting the risk of recollapse following minimally invasive kyphoplasty in osteoporotic vertebral compression fractures. *Front Public Health*. 2022;10, 874672. <https://doi.org/10.3389/fpubh.2022.874672>.
- Liao P-H, Tsuei Y-C, Chu W. Application of machine learning in developing decision-making Support models for decompressed vertebroplasty. *Healthcare*. 2022;10(2): 214. <https://doi.org/10.3390/healthcare10020214>.
- Hamilton BH, Ko CY, Richards K, Hall BL. Missing data in the American College of surgeons national surgical quality improvement Program are not missing at random:

- implications and potential impact on quality assessments. *J Am Coll Surg*. 2010;210(2):125. <https://doi.org/10.1016/j.jamcollsurg.2009.10.021>.
40. Kim HJ, Zuckerman SL, Cerpa M, Yeom JS, Lehman Jr RA, Lenke LG. Incidence and risk factors for complications and mortality after vertebroplasty or kyphoplasty in the osteoporotic vertebral compression fracture-analysis of 1,932 cases from the American College of surgeons national surgical quality improvement. *Global Spine J*. 2022;12(6):1125–1134. <https://doi.org/10.1177/2192568220976355>.
41. Cava WL, Bauer C, Moore JH, Pendergrass SA. Interpretation of machine learning predictions for patient outcomes in electronic health records. *AMIA Annual Symposium Proceedings. AMIA Symposium*. 2019;2019:572–581.