

Using machine learning and big data for the prediction of venous thromboembolic events after spine surgery: A single-center retrospective analysis of multiple models on a cohort of 6869 patients

ABSTRACT

Objective: Venous thromboembolic event (VTE) after spine surgery is a rare but potentially devastating complication. With the advent of machine learning, an opportunity exists for more accurate prediction of such events to aid in prevention and treatment.

Methods: Seven models were screened using 108 database variables and 62 preoperative variables. These models included deep neural network (DNN), DNN with synthetic minority oversampling technique (SMOTE), logistic regression, ridge regression, lasso regression, simple linear regression, and gradient boosting classifier. Relevant metrics were compared between each model. The top four models were selected based on area under the receiver operator curve; these models included DNN with SMOTE, linear regression, lasso regression, and ridge regression. Separate random sampling of each model was performed 1000 additional independent times using a randomly generated training/testing distribution. Variable weights and magnitudes were analyzed after sampling.

Results: Using all patient-related variables, DNN using SMOTE was the top-performing model in predicting postoperative VTE after spinal surgery (area under the curve [AUC] = 0.904), followed by lasso regression (AUC = 0.894), ridge regression (AUC = 0.873), and linear regression (AUC = 0.864). When analyzing a subset of only preoperative variables, the top-performing models were lasso regression (AUC = 0.865) and DNN with SMOTE (AUC = 0.864), both of which outperform any currently published models. Main model contributions relied heavily on variables associated with history of thromboembolic events, length of surgical/anesthetic time, and use of postoperative chemoprophylaxis.

Conclusions: The current study provides promise toward machine learning methods geared toward predicting postoperative complications after spine surgery. Further study is needed in order to best quantify and model real-world risk for such events.

Keywords: Big data, machine learning, venous thromboembolism, venous thromboembolic events

BENJAMIN S. HOPKINS, MICHAEL B. CLONEY¹, EKAMJEET S. DHILLON², PAVLOS TEXAKALIDIS¹, JONATHAN DALLAS, VINCENT N. NGUYEN, MATTHEW ORDON¹, NAJIB EL TECLÉ¹, THOMAS C. CHEN, PATRICK C. HSIEH, JOHN C. LIU, TYLER R. KOSKI¹, NADER S. DAHDALEH¹

Department of Neurological Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, ¹Department of Neurological Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, ²Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, WA, USA

Address for correspondence: Dr. Jonathan Dallas, Department of Neurological Surgery, Keck School of Medicine, University of Southern California, 1200 North State Street, Suite 3300, Los Angeles, CA 90033, USA.
E-mail: jonathan.dallas@med.usc.edu

Submitted: 29-Jun-23

Accepted: 12-Jul-23

Published: 18-Sep-23

INTRODUCTION

Venous thromboembolic events (VTEs) are feared complications of spinal surgery.^[1-6] They are associated with longer hospital stays, decreased satisfaction, and worse outcomes.^[2-12] While undergoing spine surgery is, itself, a risk factor for developing VTE, many factors are patient

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Hopkins BS, Cloney MB, Dhillon ES, Texakalidis P, Dallas J, Nguyen VN, *et al.* Using machine learning and big data for the prediction of venous thromboembolic events after spine surgery: A single-center retrospective analysis of multiple models on a cohort of 6869 patients. *J Craniocerv Junction Spine* 2023;14:221-9.

Access this article online	
Website: www.jcvjs.com	Quick Response Code 
DOI: 10.4103/jcvjs.jcvjs_69_23	

specific.^[2-12] Despite prevalent literature on the issue, predictive capabilities remain limited.^[13] In the age of big data and machine learning, it is now possible to extract previously unnoticed trends from large datasets.^[13-18] For spine surgery, numerous large studies involving machine learning exist; however, few to date have focused specifically on the risk of postoperative VTE after spine surgery.^[13,19] Those that have are confined to low-power exploratory studies or used nongranular national databases. Others have been limited in analysis of single procedural types, or have not been designed to focus specifically on postoperative VTE.^[13,20-22] To date, this study is the first to analyze deep learning as a method to predict postoperative VTE after spinal surgery. We present a single-center retrospective cohort of 6869 consecutive spine surgeries analyzing trends found through deep learning as well as more traditional machine learning methods.

METHODS

All patients who underwent spine surgery at a single academic institution between January 1, 2009, and May 31, 2015, were included. Spine surgeries were detected using current procedural terminology codes, and all identified primary spine surgeries were included in the analysis. If patients had multiple procedures that required different admissions, each operation was analyzed separately. We excluded any patient who underwent reoperation within 30 days, minor spine surgery (including electrode placement or hardware removal), or a secondary procedure (e.g., surgery for wound dehiscence and hematoma evacuation). For each surgery included in the study, data about the patient, the procedure, and the postoperative management and recovery were collected. The study was approved by the institutional review board. Database demographics and reference outcomes are shown in Tables 1 and 2, respectively. Further information regarding demographics and outcomes is further noted in Dhillon *et al.*^[23]

Database maintenance and statistical analysis were performed using Python 3.9. Key packages used in the analysis included TensorFlow, Keras, Matplotlib, Sklearn, and NumPy. For phase 1, seven different models were screened using 108 database variables [Table 3]. A separate variable set was further selected excluding variables that would have only been known postoperatively. In this set, only variables and data available at preoperative planning stage were included, limiting our dataset to 62 input variables. For each set of variables, training and testing data were split (75/25) and each model was trained 100 independent times, using a randomly generated training/testing set for mean/median performance comparison as well as mean and median model weights.

Table 1: Preoperative characteristics of patients

Variables	Nonchemoprophylaxis group	Chemoprophylaxis group	P
Number of patients	4965	1904	
Age (years), mean	51.86	59.7	<0.001
Sex (male) (%)	56	47	<0.001
BMI (kg/m ²), mean	28.11	29.31	<0.001
Smoking (% current smoker)	15	12	<0.001
VTE history (%)	2.10	5.80	<0.001
Epidural hematoma history	0.07	0.13	0.4
Bleeding disorder history	2.20	3.60	<0.001
Neurological deficit (%)	4	6	0.004
Number of comorbidities (mean)	1.89	2.69	<0.001

BMI - Body mass index; VTE - Venous thromboembolism

Table 2: Procedure characteristics

Characteristics	Nonchemoprophylaxis group	Chemoprophylaxis group	P
Timing of anticoagulant use (days)	-	1.46	
IVC filter placed	1.93	8.56	<0.001
Surgical site (%)			
Cervical	26.5	24.2	<0.001
Thoracic	3.7	11.2	
Lumbar	69.3	62.8	
Other	0.5	1.79	
Fusion (%)	24.7	46.1	<0.001
Decompression (%)	33.5	37.3	<0.001
Staged (%)	44.3	62.5	0.013
Surgical time (min), mean	145.66	278.59	<0.001
Anesthesia time (min), mean	205	363	<0.001
EBL (cc), mean	447.98	994.77	<0.001
ICU admission (%)	10.4	39.4	<0.001

IVC - Inferior vena cava; EBL - Estimated blood loss; ICU - Intensive care unit

A deep neural network (DNN) consisting of two dense layers of five neurons and one neuron, respectively, was trained, using Adam optimizer and binary cross-entropy as a loss function. Sigmoidal activation was used in the final layer to simulate binary outcomes. The model was trained for 10 epochs and all training was performed using a batch size of 300. The above architecture was selected through a trial-and-error method of bootstrapping as is the current standard of practice using such DNNs. Given the relatively low probability of any given procedure having a VTE, a second DNN was trained using a similar architecture and synthetic minority oversampling technique (SMOTE) to synthetically balance class sizes during the training process and not bias the model toward negative

predictions due to frequency. Likewise, further models were trained using the same variables including simple logistic regression, linear regression, and penalized linear regression using both L1 (lasso) and L2 (ridge) regressions, as well as a gradient boosting classifier ($\alpha = 0.01$). Relevant metrics were compared between each model including area under the receiver operator curve, sensitivity, specificity, as well as mean squared error for each model.

For phase 2, the top four models were selected based on a metric of area under the receiver operator curve. These models included a DNN with SMOTE, a linear regression model, a lasso regression model, and a ridge regression model as noted above. Separate random sampling of each model was performed 1000 additional independent times, again using a randomly generated training/testing distribution for each run. For the DNN model, to best understand the relationship between variables, a mean relative weight assignment was created for the DNN. For each of the linear models, mean and median coefficients were retrieved for each variable for comparison. For linear models, all variables were ranked using the absolute value of each mean and median coefficient. For each of the three linear models, each variable magnitude was placed in ascending order and the median rank for each variable was obtained in order to best analyze variable weight across models.

RESULTS

Phase 1 model results are shown in Figure 1. On initial evaluation of model performance over 100 random samples, we found that a DNN using SMOTE technique performed the best of all models with a mean area under the curve (AUC)

of 0.902. The linear regression model (AUC 0.866), ridge regression model (AUC 0.875), and lasso regression model (AUC 0.897) each performed within a reasonable distance of the aforementioned DNN with SMOTE model. Notably, logistic regression, DNN without SMOTE, and the gradient-boosted classifier significantly underperformed the others, with mean AUCs of 0.680, 0.625, and 0.683, respectively. Using only preoperative values, similar trends were noted with the highest four models being the DNN with SMOTE, lasso regression, ridge regression, and linear regression with AUCs of 0.865, 0.867, 0.852, and 0.852, respectively. Similar to prior samples, DNN without SMOTE, logistic regression, and the gradient-boosted classifier significantly underperformed relative to the prior models with AUCs of 0.588, 0.628, and 0.561, respectively.

Phase 2 model results are shown in Figure 2. After 1000 random samples, the DNN with SMOTE remained the best-performing model with a mean AUC of 0.904. Lasso regression, ridge regression, and linear regression performed only slightly worse with mean AUCs of 0.894, 0.873, and 0.864, respectively. When considering preoperative variables only, lasso regression was the top-performing model, however only a margin compared to the DNN with SMOTE (with a mean AUC of 0.865 vs. 0.864, respectively). Ridge regression and linear regression performed only slightly worse with mean AUCs of 0.852 and 0.852, respectively.

Table 4 shows the top 10 variables of the highest magnitude based on the median and mean coefficients over the 1000 random samples contributing to the three linear models as described above. The top three coefficients in terms of magnitude for the three best-performing linear models using

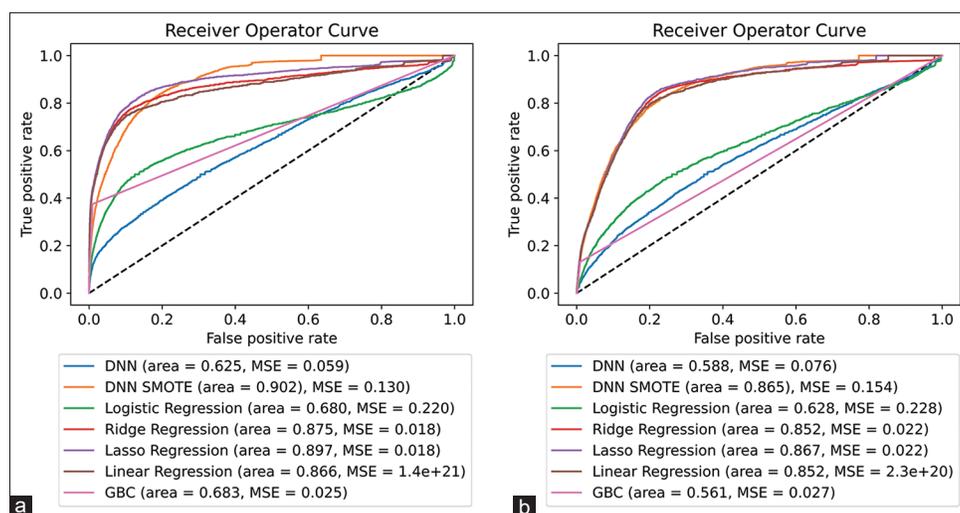


Figure 1: (a) 100 independent random samples for model selection with associated area under the curve (AUC) for all variables ($n = 108$) and (b) 100 independent random samples for model selection with associated AUC for preoperative variables only ($n = 62$). DNN - Deep neural network, SMOTE - Synthetic minority oversampling technique, MSE - Mean squared error, GBC - Gradient boosted classifier

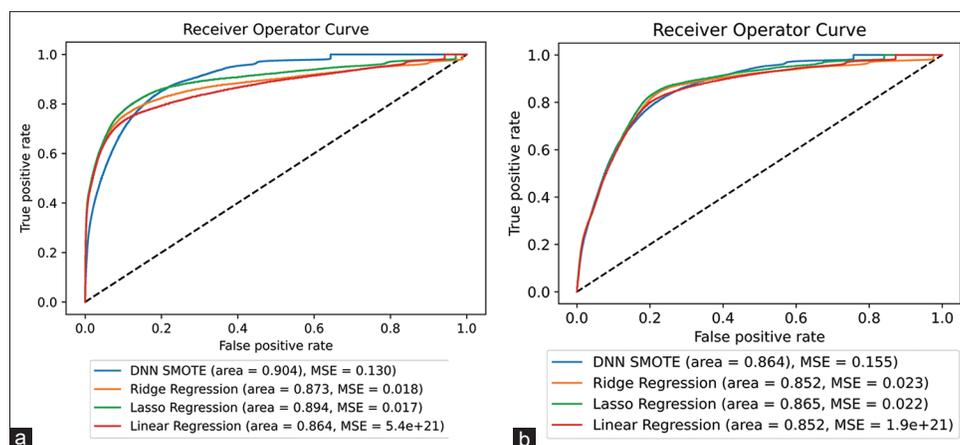


Figure 2: (a) 1000 independent random samples for the top four selected models using all variables ($n = 108$) and (b) 1000 independent random samples for the top four selected models using preoperative variables only ($n = 62$). DNN - Deep neural network, SMOTE - Synthetic minority oversampling technique, MSE - Mean squared error

both the mean and median methods were presence of any VTE prior to surgery, surgery time in minutes, and utilization of any form of chemoprophylaxis after surgery within 30 days. The top three variables when considering only preoperative factors were not unanimous. Both presence of a deep venous thrombosis (DVT) and presence of a pulmonary embolism (PE) were consistently in the top three. Using a median method, the top contributor was age, whereas using a mean method, the top contributor was involvement of surgery in the thoracic spine.

The top ten weights in order of magnitude for the DNN with SMOTE are shown in Table 5. While many of the top ten highest magnitude variables were consistent with the linear models, the top three differed in all, except for the presence of any VTE prior to surgery (with the DNN model holding the presence of cardiac disease and surgical use of decompression/laminectomy at the top two spots, respectively). Using only preoperative variables, the model focused heavily on history of events with number of VTEs over the past 12 months, presence of a DVT in the past 12 months, and presence of any VTE in the past 12 months taking the top three spots, respectively. A full schematic of all relative model weights for DNN SMOTE across the 1000 random samples is further shown in Figure 3.

DISCUSSION

Machine learning has become increasingly relevant with the advent of the electronic medical record.^[14,15] This study presents an example where advanced statistical modeling may add value in a clinically relevant scenario. To the authors' knowledge, this is the first study that uses machine learning strategies to predict post-operative VTE using a large, single-center database.^[13,20,21] While other national database studies

have looked into postoperative outcomes after spine surgery, and even VTE specifically, no studies have been performed using machine learning to evaluate VTE as a primary outcome.^[13,20,21] The results of this study provide exciting evidence that future preoperative risk stratification could be undertaken using more tailored machine learning models.

Model selection

Significant research remains ongoing regarding how to most accurately select model type. While some advocates argue for a "kitchen sink" model, more data are not always "better" for a model.^[24-26] Current research in advanced statistics instead focuses on tradeoffs between marginal added predictive value and data size.^[24-26] There are pros and cons to each approach, and as such, many different models and applications need to be explored.^[24-26] For the current article, we elected to use a "kitchen sink" approach as a starting point. Furthermore, current tendencies are to explore and attempt to use "new or novel" models whenever possible, such as convoluted neural networks, DNNs, and even generative adversarial networks. While these models are interesting, there remains a tradeoff between complexity and increasing model data requirement.^[24-26] As the use of machine learning models becomes increasingly prevalent, it becomes crucial to critically evaluate whether or not a complex machine learning model adds value over traditional statistical approaches.^[27,28] Current literature, likewise, employs traditional linear regression and/or logistic regression as a commonly used gold standard.^[29,30] As such, we elected to use traditional linear regression as a baseline for added model value.

We found that DNN using SMOTE, ridge regression, and lasso regression all outperformed traditional linear regression models for both phase 1 and phase 2. For phase 1, using all variables within the database, our top-performing model

Table 3: List of database variables used in model

Preoperative variables (n=62)	Additional variables for all variable models (n=108)
Age	Date/time of warfarin first dose postoperatively
Presence of DVT within previous 12-month period	Warfarin first dose (mg)
Presence of PE within previous 12-month period	Utilization of chemoprophylactic agents within 30 days (yes/no)
Presence of VTE within previous 12-month period	Utilization of chemoprophylactic agents within 30 days (coded)
BMI	Utilization of chemoprophylactic agents within 3 days
Staged procedure (yes/no)	Transfusion (yes/no)
Cyst/abscess (yes/no)	Total surgery time (included all stages)
Anterior involvement (yes/no)	Total anesthesia time (included all stages)
Smoking status (current, former, never)	Timing of warfarin (days)
Fusion (yes/no)	Timing of onset of any anticoagulant
Decompressions or laminectomies (yes/no)	Timing of heparin (days)
Foraminotomy (yes/no)	Timing of fondaparinux (days)
Cancer (yes/no)	Timing of enoxaparin (days)
Gender (female or male)	Timing of dalteparin (days)
Ever smoker (yes/no)	Time in ICU (h)
Department (neurosurgery/orthopedic surgery)	Surgery time (min)
Race	Stage III surgery time
Number of comorbidities (based on CCI)	Stage III surgery stop
Cervical involvement (yes/no)	Stage III surgery start
Endoscopic (yes/no)	Stage III anesthesia time
Number of stages	Stage III anesthesia stop time
Presence of hematoma within previous 12-month period	Stage III anesthesia start time
Diabetes (yes/no)	Stage II surgery time
Thoracic involvement (yes/no)	Stage II surgery stop
Autograft use (yes/no)	Stage II surgery start
Allograft use (yes/no)	Stage II anesthesia time
Osteotomy (yes/no)	Stage II anesthesia stop time
Presence of hypertension (yes/no)	Stage II anesthesia start time
Intraoperative neuromonitoring use (yes/no)	Use of red blood cell transfusion (yes/no)
Presence of neurological deficit preoperatively (yes/no)	Number of stages (range)
Endocrinology disease (yes/no)	Date/time of heparin first dose postoperatively
Surgical duration > 6 h (yes/no)	Heparin first dose (mg)
Presence of diagnosis of HTN with previous 12-month period	Date/time of fondaparinux first dose postoperatively
Corpectomy (yes/no)	Fondaparinux first dose (mg)
Levels unspecified	Date/time of enoxaparin first dose postoperatively
Worker's comp (yes/no)	Enoxaparin first dose (mg)
Facetectomy (yes/no)	EBL > 750
Lumbar involvement (yes/no)	EBL > 500 (coded)
Surgical duration > 4 h (yes/no)	EBL > 500
Fracture (yes/no)	EBL > 1000
Tumor (yes/no)	EBL (numeric)

Contd...

Table 3: Contd...

Preoperative variables (n=62)	Additional variables for all variable models (n=108)
ICU admission (yes/no)	Date/time of dalteparin first dose postoperatively
MIS/percutaneous (yes/no)	Dalteparin first dose (mg)
Presence of renal disease	Date/time of aspirin first dose postoperatively
Presence of diagnosis of bleeding disorders with previous 12-month period	Aspirin first dose (mg)
Preoperative scoliotic deformity (yes/no)	Anesthesia > 6 h
Medicare (yes/no)	Anesthesia > 4 h
Involvement of kyphoplasty and/or cement (yes/no)	Anesthesia stop time
Private insurance (yes/no)	Anesthesia start time
Lateral approach (yes/no)	Anesthesia length (min)
Posterior approach (yes/no)	Amount transfused
Number of levels	
Medicaid/public aid (yes/no)	
Number of surgical stages if > 1	
Presence of diagnosis of bleeding disorders with previous 12-month period	
Cardiac disease	
Pulmonary disease	
Number of VTEs in previous 12 months	
Stenosis present (yes/no)	
Discectomy (yes/no)	

DVT - Deep venous thrombosis; PE - Pulmonary embolism; VTE - Venous thromboembolic event; BMI - Body mass index; ICU - Intensive care unit; HTN - Hypertension; EBL - Estimated blood loss; MIS - Minimally invasive spine; CCI - Charlson Comorbidity Index

boasted an AUC of 0.904 predictive of 30-day VTE after spinal surgery when averaged over 1000 independent random samples. For phase 2, using only variables known prior to surgery, our top model had an AUC of 0.865 predictive of 30-day VTE after spinal surgery when averaged over 1000 independent random samples. Both inclusion and exclusion of postoperative variables drastically outperform prior literature models on assessment of VTE (with the next highest reported AUC in the literature reporting 0.80).^[13,19-22]

Inferences from model performance

One of the challenges associated with newer machine learning methods is understanding how to interpret results. While excellent at predicting outcomes, newer models have notoriously difficult-to-understand variable-to-variable interactions.^[31,32] Despite this, there are a number of surrogate techniques that provide at least some interpretation.^[32] For our DNN using SMOTE, we took the average weights of each of the first layers and analyzed them, as shown in Table 5 and Figure 3. While this is a commonly used surrogate for variable importance, it is not entirely reflective.^[33] While understanding the input layer is helpful, it in no way gives insight into how nodes affect each other further down the network as a near infinite number of possible modifications could be made as

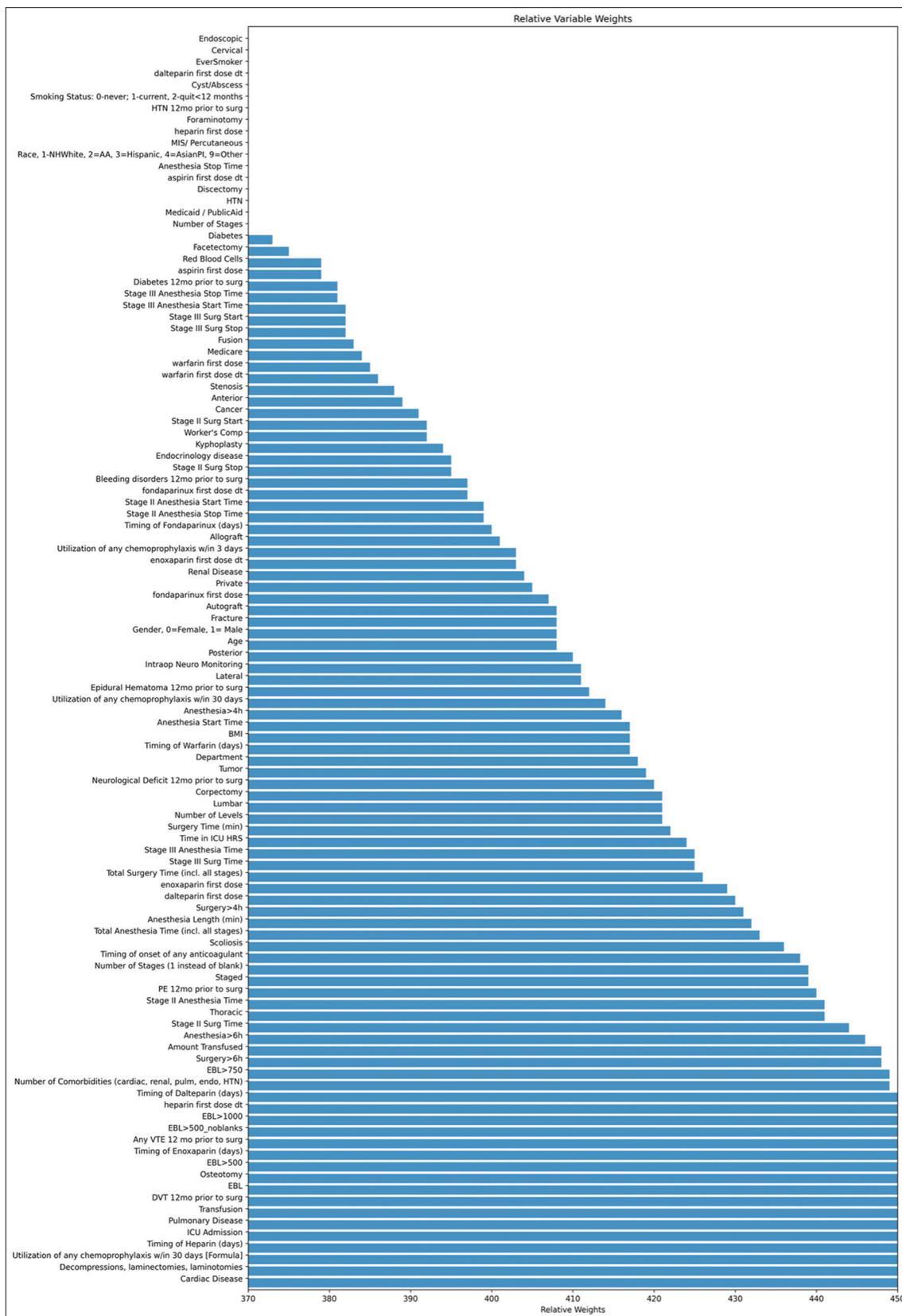


Figure 3: Mean relative weights across 1000 random samples of deep neural network using synthetic minority oversampling technique using all variables. BMI - Body mass index, EBL - Estimated blood loss, PE - Pulmonary embolism, VTE - Venous thromboembolism, DVT - Deep venous thrombosis, ICU - Intensive care unit, HTN - Hypertension

Table 4: Linear models' coefficient rank

All variables	
Top 10 (median)	Top 10 (mean)
Any VTE 12 months prior to surgery	Any VTE 12 months prior to surgery
Surgery time (min)	Surgery time (min)
Utilization of any chemoprophylaxis weeks/in 30 days (coded)	Utilization of any chemoprophylaxis weeks/in 30 days (coded)
Timing of heparin (days)	Timing of heparin (days)
Stage II surgery time	Timing of onset of any anticoagulant
EBL	Total anesthesia time (included all stages)
Enoxaparin first dose	EBL
Total anesthesia time (included all stages)	Stage III surgery start
DVT 12 months prior to surgery	Anesthesia length (min)
Diabetes	Enoxaparin first dose
Preoperative variables only	
Top 10 (median)	Top 10 (mean)
Age	Thoracic
DVT 12 months prior to surgery	DVT 12 months prior to surgery
PE 12 months prior to surgery	PE 12 months prior to surgery
Any VTE 12 months prior to surgery	Department
BMI	Any VTE 12 months prior to surgery
Staged	Foraminotomy
Cyst/abscess	BMI
Anterior	Fusion
Smoking status	Gender
Fusion	Endoscopic

VTE - Venous thromboembolic event; EBL - Estimated blood loss; DVT - Deep venous thrombosis; PE - Pulmonary embolism; BMI - Body mass index

Table 5: Deep neural network with synthetic minority oversampling technique performance variables

All variables - Top 10 (DNN)	Preoperative variables only - Top 10 (DNN)
Cardiac disease	Number of VTEs 12 months prior to surgery
Decompression/laminectomy	DVT 12 months prior to surgery
Utilization of any chemoprophylaxis weeks/in 30 days (formula)	Presence of VTE 12 months prior to surgery
Timing of heparin (days)	PE 12 months prior to surgery
ICU admission	Staged
Pulmonary disease	Osteotomy
Transfusion	Number of stages
DVT 12 months prior to surgery	Thoracic
EBL	Epidural hematoma 12 months prior to surgery
Osteotomy	Scoliosis

DNN - Deep neural network; EBL - Estimated blood loss; VTEs - Venous thromboembolic events; DVT - Deep venous thrombosis; PE - Pulmonary embolism; ICU - Intensive care unit

soon as one layer deeper within the model.^[33] Despite these limitations, such evaluations are commonly considered to be one of the better available to determine variable importance.^[33] In our model specifically, for phase 1, we found cardiac disease, presence of a decompression/laminectomy, and utilization of chemoprophylactic agents to be the top three contributors to our model performance. It is not unreasonable that we see

the presence of cardiac disease as the top predictor, as many patients with cardiac disease have predisposing ischemic changes in coronary vessels which could easily be extrapolated to other vasculature.^[4,6,34] While the authors do not entirely understand laminectomy/decompression being representative, we believe that the model could have evaluated the variable as a surrogate for simpler/quicker surgeries, therefore allowing the model to extrapolate lower risk.^[4] Finally, the use of chemoprophylaxis agents is no surprise, with its importance repeatedly shown in prior literature. For phase 2 variables, we found, not unsurprisingly, prior history to be the top three most relevant variables with the number of VTEs over the past 12 months, the presence of DVT in the past 12 months, and the presence of any VTE in 12 months prior to surgery, respectively.^[35] Other variables of high significance are noted to be timing of heparin initiation, intensive care unit admission, pulmonary disease, the presence or absence of a transfusion, estimated blood loss, the presence of osteotomy during surgery, history of epidural hematoma within 12 months prior to surgery, extension of surgery to the thoracic spine, and the number of stages performed.^[35] As many of these have been shown over numerous studies to pose a higher risk of VTEs, it is not surprising to the authors that the model focused on these variables.^[35]

When further analyzing trends within the linear models, we found similar variables to be of importance with a few exceptions. The top three variables after ranking the three linear models in phase 1 as described above were consistently history of any VTE in the 12 months prior to surgery, length of surgery, and utilization of any chemoprophylactic agent in the 30 days postoperatively, respectively [Table 4]. For phase 2, however, we found differences within the above-noted trends. When accessing median variable weights, we found the top three most significant variables to be age, history of DVT 12 months prior to surgery, and history of PE 12 months prior to surgery. When accessing mean variable weights, however, we found extension to the thoracic spine, history of DVT in 12 months prior to surgery, and history of PE 12 months prior to surgery to be the largest contributors. This suggests that, perhaps, more variability in model variable selection existed in phase 2 as the variance in variable weights between random samples was higher. As age increases, both the likelihood of immobility and the associated VTE rate both increase.^[35,36] Similarly, as extension of surgery to the thoracic spine likely suggests larger surgeries, it is not surprising that the model decided to focus on this variable as a relevant predictor.^[35,36] Furthermore, VTEs after traumatic injuries are more common than elective procedures.^[36-38] Likewise, as the thoracic spine is a common location necessitating surgical spinal evaluation and fixation, it is further possible

that thoracic surgeries could further be a marker for sicker overall patients.^[36-38]

Limitations of machine learning approaches

While powerful tools, ultimately there are numerous downfalls to employing advanced machine learning techniques without critical thought and evaluation. Just as with basic statistical models, machine learning relies on clean data that adequately approximate a global population. As such, machine learning models are extremely susceptible to data bias, perhaps even more so than traditional statistical models. Furthermore, the field of machine learning is relatively new, providing a rudimentary foundation and understanding as to how such models handle complex statistical problems. In many traditional statistical methods, excluding covariant variables and heteroskedastic data is required for basic assumptions of modeling. With the advent of big data and large models, it becomes significantly trickier to tease out such relationships. When implementing models with hundreds of variables, inevitably covariance is likely to be introduced. Furthermore, in models such as DNNs with hundreds to thousands of nodes and connections, it is difficult to ascertain whether such data principles are problematic or mitigable. In theory, such models should be able to train and minimize problematic covariance, with certain models even attempting to capitalize on known covariance in order to augment predictions. While experimental methods for correcting and mitigating these issues within newer models are being suggested, the literature is still young on such methods and, as such, caution should be employed when making conclusions from such newer techniques and models.

Limitations of our models

As described above, class imbalance and covariance are definitive limitations to the above noted models. While attempts to correct and augment model training to account for such problems (i.e., SMOTE, etc.) were performed, the nature of the event of interest (VTE) is naturally rare and therefore susceptible to skews in training data and bias.^[39] Furthermore, our data are collected at a single center retrospectively which further adds to question the overall generalizability of our data. Despite these limitations, the authors believe that the current model serves as a useful proof of concept that with extensive data, complex relationships and predictions be discovered and elucidated through newer machine learning techniques. As such, the authors believe that further, more powered, multicentered studies are needed to improve clinical practice and find the optimal role for machine learning in clinical practice.

CONCLUSIONS

VTE after spine surgery is a rare yet devastating complication,

leading to considerable comorbidity for patients. As such, better predictive methods and surveillance are needed to best prevent events during more treatable timepoints within the natural history of VTE. With the advent of big data and machine learning approaches, an opportunity exists for more robust modeling and use of patient data to prevent and understand these relatively rare events. Using all patient-related variables over 1000 independent random samples, DNN using SMOTE was the top-performing model boasting an AUC of 0.904 in predicting postoperative VTE after spinal surgery, with linear regression, lasso regression, and ridge regression performing only slightly worse with AUCs of 0.864, 0.894, and 0.873, respectively. When analyzing a subset of variables known at the preoperative timepoint only over 1000 independent random samples, the top-performing model was lasso regression with DNN using SMOTE as a close second with AUCs of 0.865 and 0.864, respectively, both of which outperform any currently published models within the literature to date. Main model contributions relied heavily on variables associated with history of thromboembolic events and length of surgical and anesthetic time alongside the use of postoperative chemoprophylaxis. The current study provides promise toward big data and machine learning methods geared toward predicting, preventing, and early surveillance of postoperative complications after spine surgery. Further study is needed to best quantify and model real-world risk for such events.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Bui A, Lashkari N, Formanek B, Wang JC, Buser Z, Liu JC. Incidence and risk factors for postoperative venous thromboembolic events in patients undergoing cervical spine surgery. *Clin Spine Surg* 2021;34:E458-65.
2. Cloney MB, Driscoll CB, Yamaguchi JT, Hopkins B, Dahdaleh NS. Comparison of inpatient versus post-discharge venous thromboembolic events after spinal surgery: A single institution series of 6869 consecutive patients. *Clin Neurol Neurosurg* 2020;196:105982.
3. Cloney MB, Goergen J, Hopkins BS, Dhillon ES, Dahdaleh NS. Factors associated with venous thromboembolic events following ICU admission in patients undergoing spinal surgery: An analysis of 1269 consecutive patients. *J Neurosurg Spine* 2018;30:99-105.
4. Cloney MB, Hopkins B, Dhillon E, El Tecle N, Koski TR, Dahdaleh NS. Chemoprophylactic anticoagulation following lumbar surgery significantly reduces thromboembolic events after instrumented fusions, not decompressions. *Spine (Phila Pa 1976)* 2023;48:172-9.
5. Cloney MB, Hopkins B, Dhillon E, El Tecle N, Swong K, Koski TR, *et al.* Anterior approach lumbar fusions cause a marked increase in thromboembolic events: Causal inferences from a propensity-matched analysis of 1147 patients. *Clin Neurol Neurosurg* 2022;223:107506.
6. Cloney MB, Hopkins B, Dhillon ES, Dahdaleh NS. The timing of venous

- thromboembolic events after spine surgery: A single-center experience with 6869 consecutive patients. *J Neurosurg Spine* 2018;28:88-95.
7. Lo BD, Qayum O, Penberthy KK, Gyi R, Lester LC, Hensley NB, *et al.* Dose-dependent effects of red blood cell transfusion and case mix index on venous thromboembolic events in spine surgery. *Vox Sang* 2023;118:76-83.
 8. Love D, Bruckner J, Ye I, Thomson AE, Pu A, Cavanaugh D, *et al.* Dural tear does not increase the rate of venous thromboembolic disease in patients undergoing elective lumbar decompression with instrumented fusion. *World Neurosurg* 2021;154:e649-55.
 9. Rudic TN, Moran TE, Kamalopathy PN, Werner BC, Bachmann KR. Venous thromboembolic events are exceedingly rare in spinal fusion for adolescent idiopathic scoliosis. *Clin Spine Surg* 2023;36:E35-9.
 10. Taghlabi K, Carlson BB, Bunch J, Jackson RS, Winfield R, Burton DC. Chemoprophylactic anticoagulation 72 hours after spinal fracture surgical treatment decreases venous thromboembolic events without increasing surgical complications. *N Am Spine Soc J* 2022;11:100141.
 11. Thota DR, Bagley CA, Tamimi MA, Nakonezny PA, Van Hal M. Anticoagulation in elective spine cases: Rates of hematomas versus thromboembolic disease. *Spine (Phila Pa 1976)* 2021;46:901-6.
 12. Tran KS, Issa TZ, Lee Y, Lambrechts MJ, Nahi S, Hiranaka C, *et al.* Impact of prolonged operative duration on postoperative symptomatic venous thromboembolic events after thoracolumbar spine surgery. *World Neurosurg* 2023;169:e214-20.
 13. Kim JS, Merrill RK, Arvind V, Kaji D, Pasik SD, Nwachukwu CC, *et al.* Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine (Phila Pa 1976)* 2018;43:853-60.
 14. Celtikci E. A systematic review on machine learning in neurosurgery: The future of decision-making in patient care. *Turk Neurosurg* 2018;28:167-73.
 15. Mofatteh M. Neurosurgery and artificial intelligence. *AIMS Neurosci* 2021;8:477-95.
 16. Hopkins BS, Yamaguchi JT, Garcia R, Kesavabhotla K, Weiss H, Hsu WK, *et al.* Using machine learning to predict 30-day readmissions after posterior lumbar fusion: An NSQIP study involving 23,264 patients. *J Neurosurg Spine* 2019;32:399-406.
 17. Hopkins BS, Weber KA 2nd, Kesavabhotla K, Paliwal M, Cantrell DR, Smith ZA. Machine learning for the prediction of cervical spondylotic myelopathy: A post hoc pilot study of 28 participants. *World Neurosurg* 2019;127:e436-42.
 18. Hopkins BS, Murthy NK, Texakalidis P, Karras CL, Mansell M, Jahromi BS, *et al.* Mass Deployment of deep neural network: Real-time proof of concept with screening of intracranial hemorrhage using an open data set. *Neurosurgery* 2022;90:383-9.
 19. Lopez CD, Boddapati V, Lombardi JM, Lee NJ, Mathew J, Danford NC, *et al.* Artificial learning and machine learning applications in spine surgery: A systematic review. *Global Spine J* 2022;12:1561-72.
 20. Wang KY, Ikwuezunma I, Puvanesarajah V, Babu J, Margalit A, Raad M, *et al.* Using predictive modeling and supervised machine learning to identify patients at risk for venous thromboembolism following posterior lumbar fusion. *Global Spine J* 2023;13:1097-103.
 21. Arvind V, Kim JS, Oermann EK, Kaji D, Cho SK. Predicting surgical complications in adult patients undergoing anterior cervical discectomy and fusion using machine learning. *Neurospine* 2018;15:329-37.
 22. Kim JS, Arvind V, Oermann EK, Kaji D, Ranson W, Ukogu C, *et al.* Predicting surgical complications in patients undergoing elective adult spinal deformity procedures using machine learning. *Spine Deform* 2018;6:762-70.
 23. Dhillon ES, Khanna R, Cloney M, Roberts H, Cybulski GR, Koski TR, *et al.* Timing and risks of chemoprophylaxis after spinal surgery: a single-center experience with 6869 consecutive patients. *J Neurosurg Spine* 2017;27:681-93.
 24. Blum AL, Langley P. Selection of relevant features and examples in machine learning. *Artificial Intelligence* 1997;97:245-71.
 25. Cooper A, Levy K, De Sa C. Accuracy-Efficiency Trade-Offs and Accountability in Distributed ML Systems. *Equity and Access in Algorithms, Mechanisms, and Optimization* 2021;4:1-11.
 26. Koch O. The Trade-Offs of Large-Scale Machine Learning: The Price of Time. Medium; 2020. Available from: <https://medium.com/criteo-engineering/the-trade-offs-of-large-scale-machine-learning-71ad0cf7469f>. [Last accessed on 2023 Jan 23]
 27. Stewart M. The Actual Difference Between Statistics and Machine Learning. Medium; 2019. Available from: <https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3>. [Last accessed on 2023 Jan 23].
 28. Stewart M. The Limitations of Machine Learning. Medium; 2019. Available from: <https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6>. [Last accessed on 2023 Jan 23].
 29. Lazar A, Jin L, Brown C, Spurlock CA, Sim A, Wu K. Performance of the gold standard and machine learning in predicting vehicle transactions. *Lawrence Berkeley National Laboratory* 2021:1-5.
 30. Lee KY, Kim KH, Kang JJ, Choi SJ, Im YS, Lee YD, *et al.* Comparison and Analysis of Linear Regression & Artificial Neural Network. *Int J Appl Eng Res* 2017;12:9820-5.
 31. Sheu YH. Illuminating the black box: Interpreting deep neural network models for psychiatric research. *Front Psychiatry* 2020;11:551299.
 32. Blazek P. Why We Will Never Pen Depp Learning's Black Box. Medium; 2022. Available from: <https://towardsdatascience.com/why-we-will-never-open-deep-learning-black-box-4c27cd335118>. [Last accessed on 2023 Jan 23].
 33. Montavon G, Samek W, Muller KR. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing* 2018;73:1-15.
 34. Gregson J, Kaptoge S, Bolton T, Pennells L, Willeit P, Burgess S, *et al.* Cardiovascular risk factors associated with venous thromboembolism. *JAMA Cardiol* 2019;4:163-73.
 35. Zhang L, Cao H, Chen Y, Jiao G. Risk factors for venous thromboembolism following spinal surgery: A meta-analysis. *Medicine (Baltimore)* 2020;99:e20954.
 36. Gephart MG, Zygorakis CC, Arrigo RT, Kalanithi PS, Lad SP, Boakye M. Venous thromboembolism after thoracic/thoracolumbar spinal fusion. *World Neurosurg* 2012;78:545-52.
 37. Platzer P, Thalhammer G, Jaindl M, Obradovic A, Benesch T, Vecsei V, *et al.* Thromboembolic complications after spinal surgery in trauma patients. *Acta Orthop* 2006;77:755-60.
 38. Cloney MB, Yamaguchi JT, Dhillon ES, Hopkins B, Smith ZA, Koski TR, *et al.* Venous thromboembolism events following spinal fractures: A single center experience. *Clin Neurol Neurosurg* 2018;174:7-12.
 39. He J, Cheng MX. Weighting methods for rare event identification from imbalanced datasets. *Front Big Data* 2021;4:715320.