Received: 10 December 2024 | Revised: 18 February 2025 | Accepted: 1 March 2025
DOI: 10.1111/vsu.14250

CLINICAL RESEARCH

WILEY

Machine-learning-based prediction of functional recovery in deep-pain-negative dogs after decompressive thoracolumbar hemilaminectomy for acute intervertebral disc extrusion

Daniel Low BVetMed, CertAVP(GSAS), PGCertVPS, MRCVS[1,2] | **Sophie Stables BVSc, MRCVS**[3] | **Laura Kondrotaite DVM, MRCVS**[4] | **Ben Garland BVetMed, CertAVP(GSAS), MRCVS**[5] | **Scott Rutherford BVMS, CertSAS, DipECVS, MRCVS**[1]

[1]Frank. Pet Surgeons, Leeds, UK
[2]Swift Referrals, Wetherby, UK
[3]Anderson Moores Veterinary Specialists, Winchester, UK
[4]Northwest Veterinary Specialists, Sutton Weaver, UK
[5]Debenham Vets Surgical Referrals, Stowmarket, UK

Correspondence
Daniel Low, Frank. Pet Surgeons, Leeds, UK.
Email: daniel.kuan.chong.low@gmail.com

Funding information
IVC Evidensia

Abstract

Objective: To develop and compare machine-learning algorithms to predict recovery of ambulation after decompressive surgery for acute intervertebral disc extrusion (IVDE).

Study design: Multicenter retrospective cohort study.

Sample population: Deep-pain-negative dogs with acute IVDE ($n = 162$).

Methods: Clinical variables were preprocessed for machine learning and split into independent training and test sets in an 80:20 ratio. Each model was trained and internally validated on the full test set ($Test_{full}$) and the XGBoost algorithm validated on the same test set with preoperative variables withheld ($Test_{wh}$).

Results: Recovery of ambulation was recorded in 86/162 dogs (53.1%) in this sample population after decompressive surgery. The XGBoost algorithm achieved the best performance with an area under the receiver operating characteristic curve (AUC) of .9502 (95% CI: .8919–.9901), an accuracy of .8906 (95% CI: .8125–.9531), a sensitivity of .8750, and a specificity of .9063 on $Test_{full}$. XGBoost performance on $Test_{wh}$ was decreased, with an AUC of .8271 (95% CI: .7186–.9209), an accuracy of .7187 (95% CI: .6093–.8281), a sensitivity of .5625, and a specificity of .8750.

Conclusion: Machine-learning algorithms may predict outcomes accurately in deep-pain-negative dogs with IVDE after decompressive surgery. The

Abbreviations: AdaBoost, Adaptive Boosting; AUC, area under the receiver operating characteristic curve; IQR, interquartile range; IVDE, intervertebral disc extrusion; NPV, negative predictive value; PPV, positive predictive value; SHAP, Shapley additive explanations; T2W:L2, T2-weighted hyperintensity to L2 ratio; $Test_{full}$, full test set; $Test_{wh}$, test set with postoperative variables withheld; XGBoost, Extreme Gradient Boosting.

A subset of these data was presented as an abstract at the British Veterinary Neurology Symposium, 2024.

XGBoost algorithm performed best on tabular data from this veterinary population undergoing spinal surgery.

Clinical significance: Machine-learning algorithms outperform current methods of prognostication. Pending external validation, machine-learning algorithms may be useful as assistive tools for surgical decision making.

1 | INTRODUCTION

Thoracolumbar intervertebral disc extrusion (IVDE) is a common cause of acute spinal cord injury and pelvic limb dysfunction in dogs,[1,2] with an estimated lifetime risk of up to 15.7% in high-risk breeds.[3] The severity of spinal cord injury is variable and the degree of neurological deficits is associated strongly with functional outcome.[4] The loss of deep pain perception represents the most severe grade of spinal cord injury. Even with surgical decompression, it is associated with a guarded prognosis, with 30% to 75% of dogs making a functional recovery.[1,4] Prognosticating deep-pain-negative dogs with IVDE is of great importance for pet owners, who may not be willing or able to commit resources, financial or otherwise, to a treatment with an uncertain outcome. Surgery is not benign and is associated with a reasonable rate of perioperative morbidity and euthanasia.[5,6] Prognostic factors such as patient signalment, clinical presentation, anatomical location of extrusion, delay to surgery, and magnetic resonance imaging (MRI) biomarkers have been investigated extensively but remain imperfect markers compared with the loss of deep pain perception.[1] It is unclear if clinical variables such as speed of onset,[7,8] nonambulatory duration,[9,10] and T2-weighted spinal cord hyperintensities[11–13] are associated with prognosis for recovery. Improved prognostication methods are important to guide veterinarians' treatment recommendations, and to advise pet owners appropriately.

Traditional statistical methods, such as multivariable logistic regression, rely on specific assumptions within datasets, which limit the ability to capture real-world complex interactions between variables and outcomes.[14–16] This limitation reduces their performance on external patient populations, which reduces their clinical applicability.[17] In contrast, machine learning algorithms trained on electronic health records have demonstrated superior prognostic capabilities compared to traditional statistical models[18] and are increasingly explored as predictive tools for risk assessment in both medical and surgical contexts in human populations.[19,20]

Accurate prediction of postoperative outcomes would allow for selection of individual dogs likely to benefit from surgery, and possibly for the avoidance of surgery in a dog unlikely to benefit from this treatment. This information is also useful to counsel pet owners and to manage their expectations. This study primarily aimed to develop, internally validate, and compare various machine-learning algorithms to predict postoperative recovery of ambulation after decompressive thoracolumbar hemilaminectomy for acute IVDE. Secondarily, we aimed to derive model explanations to identify the most influential clinical variables associated with recovery of ambulation.

2 | METHODS

This study was reported in accordance with recommendations from the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View.[21] All data in this study were obtained retrospectively with written client consent granting permission for the anonymized use of patient data for research purposes, and with additional regulatory ethical approval for multicenter data sharing and model development (Royal College of Veterinary Surgeons Ethics Review Panel, 2024-043-Low).

2.1 | Data extraction and outcome definition

Retrospective data extraction was performed via manual search of each institution's electronic health records, with coauthors working independently. Records from 2018 to 2024 were searched for "deep-pain negative" or "grade 5," and "thoracolumbar hemilaminectomy." Inclusion criteria were dogs with acute thoracolumbar IVDE less than 14 days in duration, undergoing decompressive surgery. Patient records were excluded if adjunctive surgical procedures such as partial corpectomy or vertebral stabilization were performed, if loss of deep pain was equivocal, or if 28-day follow up was unavailable for outcome data collection. Clinical variables collected included patient age at the time of surgery, breed, body weight, sex, and neuter status, previous spinal surgeries, duration of clinical signs, nonambulatory time, deep-pain-negative time, imaging modality, T2-weighted hyperintensity to L2 ratio (T2W:L2), blood test results and abnormalities, attending surgeon, time delay to surgery, surgical technique, use of fenestration and/or durotomy, intraoperative complications, surgery time, anesthesia time, duration of hospitalization, method of follow up, and time to follow up.

Duration of clinical signs, nonambulatory time, and deep-pain-negative time were collected retrospectively based on clinical history from the pet owner and from the referring veterinarian.

Recovery of ambulation was the outcome of interest and these data were collected either via neurological examination or remotely, depending on the institution.

2.2 | Predictive modeling

Twenty-eight preoperative variables were selected based on retrospective data availability and for potential clinical utility. They were readily available from electronic health records. Continuous data included patient age at the time of surgery, body weight, duration of clinical signs, nonambulatory time, deep-pain-negative time, T2W:L2, time delay to surgery, surgery time, anesthesia time, duration of hospitalization, and time to follow up. Categorical data included the institution; patient's breed, sex, and neuter status; previous spinal surgeries; imaging modality; blood test results and abnormalities; attending surgeon; surgical technique; use of fenestration and/or durotomy; intraoperative complications, and method of follow up. Continuous data were preprocessed using a robust scaler to handle outliers, and categorical data were preprocessed with one-hot encoding. Missing data were imputed with a k-nearest neighbors ($k = 5$) imputer and 2:1 interpolation oversampling applied to the dataset. The dataset was randomly split into training and testing sets in an 80:20 ratio using a pseudorandom split function with a consistent seed. Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), Ridge regression, and Gaussian Naive Bayes were selected and were independently trained and tuned on the data with automated Bayesian hyperparameter tuning, guided by area under the receiver operating characteristic curve (AUC) as the target maximizing metric. Each model was then validated internally on the full test set ($\text{Test}_{\text{full}}$), with predictive performance reported as AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV). Sensitivity and specificity were defined as model ability to predict recovery and nonrecovery of ambulation respectively. The 95% confidence intervals (CI) for AUC and accuracy were calculated with bootstrap resampling. The XGBoost model was tested on the same independent test set with postoperative variables withheld (Test_{wh}). For the best performing model, Shapley additive explanations (SHAP) values were calculated to identify the clinical variables with the greatest contributions to model performance.[22] Data normality was assessed with the Shapiro–Wilk test. Nonparametric data were expressed as median (interquartile range [IGR]) and parametric data were expressed as means and standard deviations. Univariable statistical analysis was performed with the χ^2 test for categorical data and the Student's *t*-test and Mann–Whitney *U*-test for parametric and nonparametric data respectively. Statistical significance was set at $p < .05$. All statistical analysis and predictive modeling were performed with *pandas*, *scipy*, *numpy*, *scikit-learn*, *imbalanced-learn*, *xgboost*, *AdaBoost-Classifier*, *GaussianNB*, *Ridge*, *optuna*, and *matplotlib* in Python, version 3.10.12.[23–33]

3 | RESULTS

Cases were recruited from five referral hospitals in the United Kingdom. Records of 206 deep-pain-negative dogs that underwent decompressive surgery were retrieved by searching electronic health records. Nineteen dogs were excluded, having been lost to follow up, leaving 187 dogs that met the inclusion criteria. Twenty-five dogs died or were euthanized postoperatively before outcome data collection. Eighteen (8.7%) of these deaths were due to progressive myelomalacia. Of dogs surviving to follow up, the rate of recovery of ambulation after decompressive surgery in this sample population was 86/162 dogs (53.1%). Summary statistics for each cohort are shown in Table 1. There was one intraoperative complication in each cohort, due to iatrogenic peripheral nerve root damage. The T2W:L2 in the cohort that recovered ambulation (median: 0 [IQR: 0–2]) was different from the cohort that did not recover ambulation (median: 2 [IQR: 1–4.9]; $p < .0001$). The category of attending surgeon was different between both cohorts ($p = .0266$). All other variables were not different with regard to univariable analysis.

Each of the four algorithms was trained and tuned independently with Bayesian hyperparameter optimization. Final model parameters achieved after optimization are reported in Data S1. XGBoost achieved the highest AUC of .9502 (95% CI: .8919–.9901), and accuracy of .8906 (95% CI: .8125–.9531), in comparison with the other machine-learning algorithms (Table 2). XGBoost also achieved balanced sensitivity and specificity while maintaining discriminatory performance. It showed a measurable decrease in predictive performance when tested on Test_{wh}. The other machine-learning algorithms investigated showed good to fair discriminatory performance, relative to XGBoost (Figure 1).

TABLE 1 Summary statistics of each cohort.

	Recovered ambulation (n = 86)	Did not recover ambulation (n = 76)	p
Institution	NWVS (n = 13)	NWVS (n = 5)	.1389
	fPS (n = 40)	fPS (n = 41)	
	AMVS (n = 3)	AMVS (n = 7)	
	SR (n = 30)	SR (n = 23)	
Age (years)	4.5 (3.7–5.7)	4.4 (3.5–5.8)	.6123
Breed	Dachshund (n = 29)	Dachshund (n = 23)	.4271
	French bulldog (n = 23)	French bulldog (n = 22)	
	Crossbreed (n = 23)	Crossbreed (n = 16)	
	Eight other breeds (n = 11)	Seven other breeds (n = 15)	
Bodyweight (kg)	10.1 (7.9–13.1)	11.0 (7.6–14.2)	.6016
Sex	Male neutered (n = 28)	Male neutered (n = 26)	.4547
	Male entire (n = 16)	Male entire (n = 21)	
	Female neutered (n = 26)	Female neutered (n = 19)	
	Female entire (n = 16)	Female entire (n = 10)	
Previous spinal surgery	Previous surgery (n = 3)	Previous surgery (n = 6)	.2243
	No previous surgery (n = 83)	No previous surgery (n = 70)	
Duration of clinical signs (days)	2 (1–3)	2 (1–3)	.8101
Nonambulatory time (h)	12 (6–24)	18 (6–24)	.3990
Deep-pain-negative time (h)	12 (6–24)	12 (6–24)	.1910
Imaging modality	MRI (n = 43)	MRI (n = 33)	.7043
	Plain CT (n = 30)	Plain CT (n = 30)	
	CT myelogram (n = 13)	CT myelogram (n = 13)	
T2W:L2	0 (0–2)	2 (1–4.9)	**<.0001**
Blood tests and abnormalities (where performed)	Normal (n = 14)	Normal (n = 9)	.6657
	Any abnormality (n = 3)	Any abnormality (n = 4)	
Attending surgeon	Board-certified surgeon (n = 39)	Board-certified surgeon (n = 36)	**.0266**
	Board-certified neurologist (n = 16)	Board-certified neurologist (n = 4)	
	Resident (n = 5)	Resident (n = 12)	
	Nonboarded clinician (n = 26)	Nonboarded clinician (n = 24)	
Time delay to surgery (h)	3 (2–6)	3 (2–6)	.2744
Surgical technique	Hemilaminectomy (n = 74)	Hemilaminectomy (n = 71)	.1981
	Pediculectomy (n = 12)	Pediculectomy (n = 5)	
Fenestration	None (n = 53)	None (n = 60)	.0699
	Same level (n = 21)	Same level (n = 11)	
	Multilevel (n = 11)	Multilevel (n = 5)	
Durotomy	3/86 (3.5%)	4/76 (5.3%)	.5820
Intraoperative complications	1/86 (1.2%)	1/76 (1.3%)	.9304

TABLE 1 (Continued)

	Recovered ambulation ($n = 86$)	Did not recover ambulation ($n = 76$)	p
Surgery time (min)	75 (56.5–107)	75.5 (60–107)	.8467
Anesthesia time (min)	176.5 (135–240)	185 (150–240)	.1730
Duration of hospitalization (h)	96 (65–168)	90.5 (69.5–192)	.7433
Method of follow up	Neurological examination ($n = 66$)	Neurological examination ($n = 65$)	.1678
	Remote assessment ($n = 20$)	Remote assessment ($n = 11$)	
Time to follow up	42.5 (32–80)	42 (31–66.5)	.0698

Note: Parametric numerical data are represented as means (standard deviations). Nonparametric numerical data are represented as medians (IQRs). Bold indicates significant value, $p < .05$.
Abbreviations: AMVS, Anderson moores veterinary specialists; fPS, frank pet surgeons; NWVS, Northwest veterinary specialists; SR, swift referrals; T2W:L2, T2-weighted hyperintensity to L2 ratio.

TABLE 2 Independent test set predictive performance of four machine-learning algorithms for outcome prediction after decompressive thoracolumbar hemilaminectomy for acute intervertebral disc extrusion.

Algorithm		AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
XGBoost	$Test_{full}$	.9502 (.8919–.9901)	.8906 (.8125–.9531)	.8750	.9063	.9032	.8788
	$Test_{wh}$	.8271 (.7186–.9209)	.7187 (.6093–.8281)	.5625	.8750	.8182	.6667
AdaBoost		.8262 (.6987–.9159)	.7500 (.6563–.8594)	.6563	.8438	.8077	.7105
Ridge regression		.9004 (.8095–.9705)	.8125 (.7031–.9063)	.8750	.7500	.7778	.8571
Gaussian Naive Bayes		.7568 (.6196–.8750)	.6406 (.5156–.7656)	.4375	.8438	.7368	.600

Note: 95% confidence intervals in brackets. Sensitivity and specificity are defined as model ability to predict recovery and nonrecovery of ambulation respectively.
Abbreviations: AdaBoost, adaptive boosting; AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value; $Test_{full}$, full independent test set; $Test_{wh}$, test set with preoperative variables withheld; XGBoost, extreme gradient boosting.

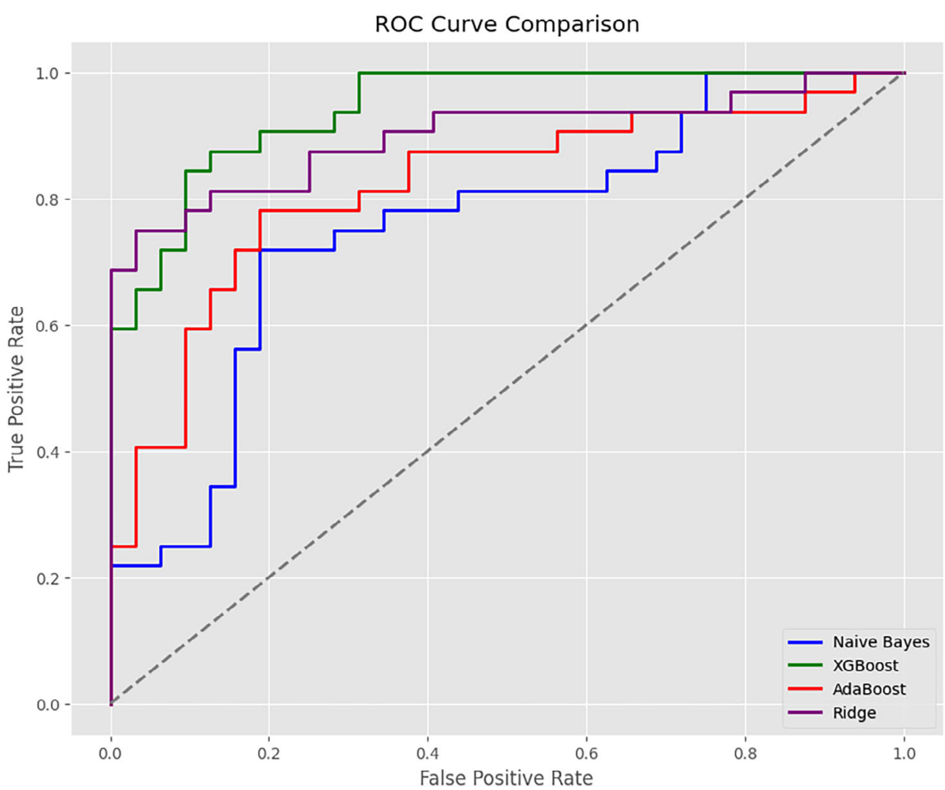

FIGURE 1 Receiver operating characteristic (ROC) curves of four machine-learning algorithms for outcome prediction after decompressive thoracolumbar hemilaminectomy for acute intervertebral disc extrusion. AdaBoost, adaptive boosting; Ridge, ridge regression; ROC, receiver operating characteristic; XGBoost, extreme gradient boosting.

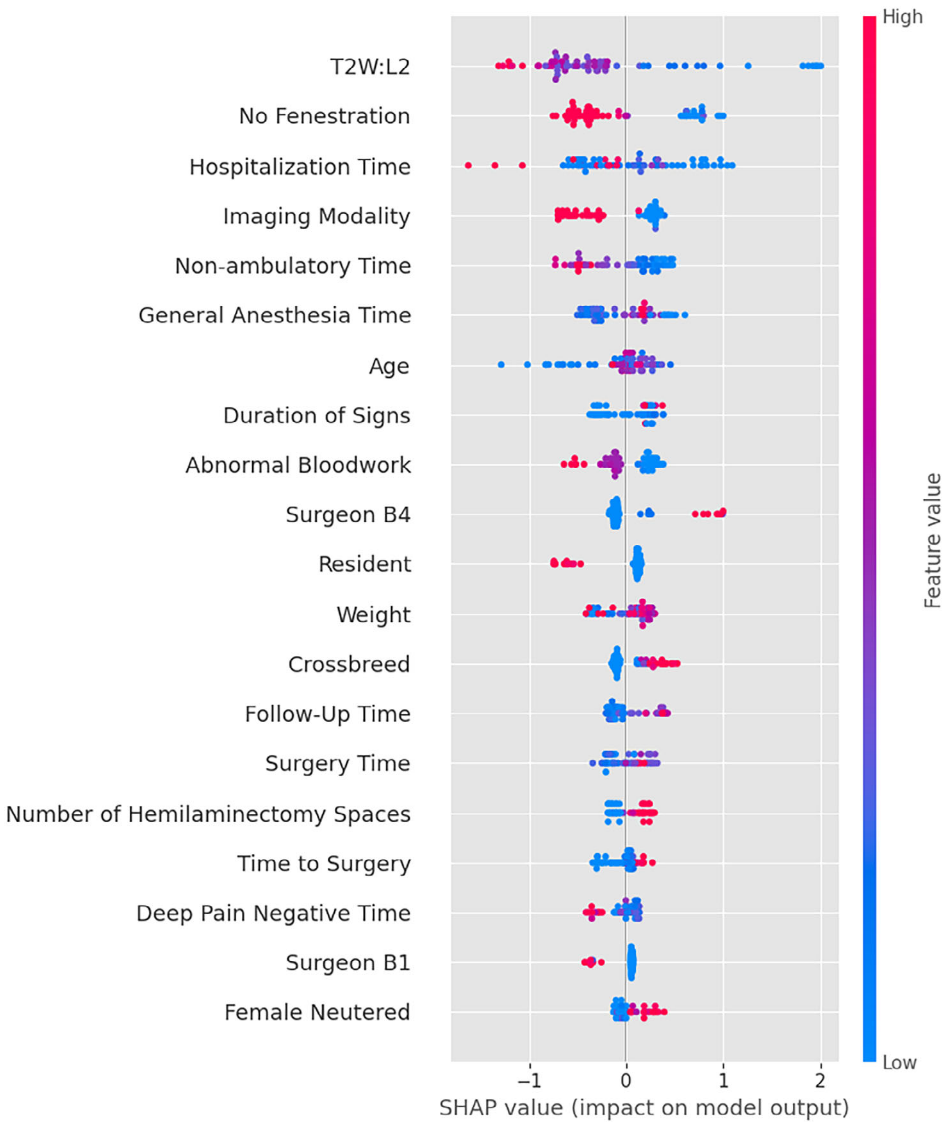

FIGURE 2 Shapley additive explanations (SHAP) summary plot showing the 20 most important variables for extreme gradient boosting (XGBoost) model prediction after decompressive thoracolumbar hemilaminectomy for acute intervertebral disc extrusion. Variables are ordered in descending order of cumulative importance. Points to the right and left of the *y*-axis indicate increased and decreased likelihood, respectively, of postoperative recovery of ambulation. Variables not shown had minimal impact on the outcome.

The five variables with the greatest impact on XGBoost model predictions were T2W:L2, the use of fenestration, hospitalization time, imaging modality used, and nonambulatory time, as calculated using SHAP. Figure 2 shows the 20 most important variables and their directionality of impact. Variables not shown were considered to be minimally important.

4 | DISCUSSION

This study demonstrated that veterinary electronic health records can be used to train machine-learning algorithms. It also showed that machine-learning algorithms trained on retrospectively collected data can predict accurately postoperative recovery of ambulation after decompressive surgery for deep-pain-negative dogs with acute IVDE, and specifically, that the XGBoost algorithm achieved the best all-round performance. Finally, it demonstrated a workflow using the XGBoost algorithm as a preoperative predictive tool without input of any postoperative variables.

Machine learning has been investigated in veterinary internal medicine, public health, radiology, and pathology,[34–38] but remains relatively poorly studied in veterinary surgery.[39] Machine-learning algorithms have been investigated in human medicine and surgery and have been shown to be accurate risk prediction models after general and reconstructive surgery.[40–42] In dogs, the loss of deep pain perception has been shown repeatedly to be the most reliable prognostic indicator; however, this still only identifies up to 60% of individuals who may benefit from surgery.[1,10,43] Previous cohort studies have failed to identify clinical risk factors which may be prognostically important in deep-pain negative dogs.[8,10] Magnetic resonance imaging features as biomarkers have uncertain clinical utility[1] as there are conflicting reports about the association between spinal cord hyperintensities and functional outcome.[11–13] Thus, even today, deep-pain-negative dogs with IVDE represent a population that is clinically challenging to prognosticate.

Machine-learning algorithms identify patterns in tabular data such as electronic health records and their association with a target outcome. Traditional multivariable statistics assume simple and linear relationships between variables and outcomes, whereas machine learning more adequately captures complex nonlinear relationships, better reflecting real-world interactions.[44] Machine-learning models incorporate regularization techniques to limit overfitting and thus improve generalizability to independent datasets,[45] whereas results of studies using traditional statistical methods do not always withstand external validation.[17]

In this study, even with a modestly sized dataset, machine-learning techniques were applied and good predictive performance was achieved on the independent test set. The XGBoost algorithm outperformed other algorithms, concurring with previous studies showing the algorithm to be well suited to machine learning with tabular data,[46] outperforming other gradient-boosting methods.[47,48] The XGBoost algorithm can accept datasets with missing data which is a common problem with retrospective electronic health record searching.[29] To simulate a real-world scenario for surgical decision making, the XGBoost algorithm was tested with postoperative variables withheld. As expected, a decrease in predictive performance was observed, indicating that larger sample sizes and external validation are needed to support this study's findings.

In analogous studies of humans with spinal cord metastasis or spinal cord injury, with sample populations of similar sizes to that in this study, the XGBoost algorithm has also been shown to be a good predictor of functional outcome after surgery.[49,50] Traditional statistics relies on null hypothesis testing and arbitrary probability value cutoffs, which have been criticized in both the veterinary and statistical community.[51,52] This approach risks discarding data with borderline significance, whereas machine-learning models rely on the additive effects of both strong and weak predictors,[29] which may additionally explain why these models perform better. The XGBoost model in this study achieved a sensitivity of 88% and specificity of 91% in predicting recovery of ambulation on the independent test set, which outperforms all current prognostication methods.

A criticism of machine-learning algorithms is their lack of transparency, which can make their decision-making processes difficult to interpret.[53] The SHAP summary plot visually illustrates the contributions of variables to the XGBoost algorithm's predictions and therefore enhances model interpretability. The T2W:L2, which was highly significant in univariable analysis, emerged as the most influential variable in the XGBoost model's decision making. However, this pattern no longer held for the other clinical variables. On univariable analysis, fenestration approached statistical significance as a predictor of functional outcome. However, because of p-value thresholds, fenestration was considered nonsignificant. However, fenestration was identified as the second most important variable in the XGBoost model. Conversely, the attending surgeon, although significant in univariable analysis, appeared to have less predictive value in the XGBoost model. Our results suggest that the T2W:L2 and use of fenestration are important predictors of recovery, with lower T2W:L2 values and the use of fenestration associated with more favorable outcomes.

The model was not significantly influenced by onset of clinical signs, nonambulatory time and time delay to surgery, which concurs with previous studies.[8,54,55] Quantitative MRI biomarkers and their influence on recovery from loss of deep pain remains a topic of debate.[1,6] and the results of our study may differ from other studies because our sample population comprised only deep-pain-negative dogs. Therapeutic fenestration has been proposed as a treatment option for acute IVDE, as an alternative to decompressive surgery.[55–57] The findings in this study suggest that the T2W:L2 significantly influenced outcome in deep-pain-negative dogs, and that fenestration may have a therapeutic effect as well, on top of the decompressive effect of surgery. Further research into both these areas is warranted. In general, although this study was not designed primarily to identify risk factors for poor outcomes, further investigation using machine-learning techniques may provide novel information about potential prognostic factors.

The participating centers in this study are generally representative of the practice of canine spinal surgery in the United Kingdom, with board-certified surgeons, board-certified neurologists, and nonboarded clinicians represented. The sample size also represents one of the largest studies of deep-pain-negative dogs with IVDE. The sample population is also consistent with previous reports where young, chondrodystrophic dogs are overrepresented.[58] The rate of recovery of ambulation[1] and rate of progressive myelomalacia[10,59] in this study was also within previously reported limits. However, this study is limited by the nonrandomized application of surgery and study case inclusion. Board-certified surgeons were overrepresented and practice pattern differences between neurologists and surgeons have previously been reported.[60] It is also likely that preoperative surgeon selection and institutional imaging availability was an important factor in explaining the difference in outcomes between surgeon types, with only selected patients undergoing cross-sectional imaging and, subsequently, only selected patients undergoing surgery. The relatively low rate (8.7%) of progressive myelomalacia in this sample population supports the fact that attending clinicians were likely selecting candidates for surgery. The heterogeneity of care across different institutions would have been another source of uncontrolled bias, with different imaging modalities possibly influencing case selection for surgery and the extent of decompression performed.[61]

Retrospective electronic health record review is also limited by missing data. Although missing data may be mitigated by data imputation, the single imputation technique employed in this study may have underestimated variance within the missing data points. Certain clinical data such as duration of onset of signs and nonambulatory time are also reliant on owner reporting, which may be subject to recall bias and inaccurate reporting by laymen.[54]

Deep-pain-negative time may have also been misassessed by general practitioner veterinarians and would have been a source of uncontrolled bias.[62] The assessment of deep pain can also be subjective and false-positive enrolment or false-negative exclusion may have been possible. Death or euthanasia within 28 days, or loss to follow up would have also been a source of uncontrolled bias. The threshold for outcome data collection was defined as 28 days to maximize case recruitment; however, this may have misclassified dogs with delayed recovery of ambulation, which can take up to 6 months.[1] Remote assessment of dogs may have also risked misclassification of spinal walking as recovery of independent ambulation.[10]

Given the challenges, mentioned above, of prognosticating deep-pain-negative dogs with IVDE, and the uncertainty around risk factors for nonrecovery, accurate machine-learning predictive models will provide much-needed information for both veterinarians and pet owners to guide decision making. However, this study's results are based on a limited cohort and are only preliminary in nature. Prospective, and external, validation with larger sample sizes is required before clinical application of these machine-learning algorithms. Machine-learning techniques also offer additional opportunities to identify patient cohorts that may benefit from decompressive surgery, potentially revealing novel insights into canine IVDE.

In conclusion, veterinary electronic health records can be used to train accurate machine-learning algorithms that withstand internal validation. The XGBoost algorithm demonstrated the best balanced predictive performance on tabular electronic health record data from this veterinary population undergoing spinal surgery. Further investigation and development may allow the XGBoost algorithm to be used as a preoperative assistive tool in surgical decision making.

ACKNOWLEDGMENTS

Author Contributions: Low D, BVetMed, CertAVP(GSAS), PGCertVPS, MRCVS: Study conception, study design, data collection, model development and data analysis, interpretation of results, and writing the manuscript. Stables S, BVSc, MRCVS: Data collection, interpretation of results, and manuscript revision. Kondrotaite L, DVM, MRCVS: Data collection, interpretation of results, and manuscript revision. Garland B, BVetMed, CertAVP(GSAS), MRCVS: Data collection, interpretation of results, and manuscript revision. Rutherford S, BVMS, CertSAS, DipECVS, MRCVS: Study conception, study design, interpretation of results, and manuscript revision. All authors approved the final manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest related to this report.

DATA AVAILABILITY STATEMENT

The analytic code, model parameters, and trained model weights are available at the following public repository https://github.com/daniel-low-vet/XGBoost-DeepPainNegative.

ORCID

Daniel Low https://orcid.org/0000-0003-0215-4997
Scott Rutherford https://orcid.org/0000-0002-2044-3656

REFERENCES

1. Olby NJ, da Costa RC, Levine JM, Stein VM. Prognostic factors in canine acute intervertebral disc disease. *Front Vet Sci*. 2020; 7:596059. doi:10.3389/fvets.2020.596059
2. Rossi G, Stachel A, Lynch AM, Olby NJ. Intervertebral disc disease and aortic thromboembolism are the most common causes of acute paralysis in dogs and cats presenting to an emergency clinic. *Vet Rec*. 2020;187(10):e81. doi:10.1136/vr.105844
3. Packer RMA, Seath IJ, O'Neill DG, De cker S, Volk HA. DachsLife 2015: an investigation of lifestyle associations with the risk of intervertebral disc disease in dachshunds. *Canine Genet Epidemiol*. 2016;3(1):8. doi:10.1186/s40575-016-0039-8
4. Langerhuus L, Miles J. Proportion recovery and times to ambulation for non-ambulatory dogs with thoracolumbar disc extrusions treated with hemilaminectomy or conservative treatment: a systematic review and meta-analysis of case-series studies. *Vet J*. 2017;220:7-16. doi:10.1016/j.tvjl.2016.12.008
5. Low D, Rutherford S. Investigating the weekend effect in decompressive thoracolumbar hemilaminectomy for acute intervertebral disc extrusion: an observational cohort study of 460 cases (2018–2023). *Vet Surg*. 2024;53(8):vsu.14089. doi:10.1111/vsu.14089
6. Jones GMC, Cherubini GB, Llabres-Diaz F, Caine A, De Stefani A. A case series of 37 surgically managed, paraplegic, deep pain negative French bulldogs, with thoracolumbar intervertebral disc extrusion, from two English referral centres. *Vet Rec Open*. 2023;10(1):e61. doi:10.1002/vro2.61
7. Scott HW, McKee WM. Laminectomy for 34 dogs with thoracolumbar intervertebral disc disease and loss of deep pain perception. *J Small Anim Pract*. 1999;40(9):417-422. doi:10.1111/j.1748-5827.1999.tb03114.x
8. Jeffery ND, Barker AK, Hu HZ, et al. Factors associated with recovery from paraplegia in dogs with loss of pain perception in the pelvic limbs following intervertebral disk herniation.

J Am Vet Med Assoc. 2016;248(4):386-394. doi:10.2460/javma.248.4.386
9. Davis GJ, Brown DC. Prognostic indicators for time to ambulation after surgical decompression in nonambulatory dogs with acute thoracolumbar disk extrusions: 112 cases. *Vet Surg*. 2002;31(6):513-518. doi:10.1053/jvet.2002.36015
10. Olby N, Levine J, Harris T, Muñana K, Skeen T, Sharp N. Long-term functional outcome of dogs with severe injuries of the thoracolumbar spinal cord: 87 cases (1996–2001). *J Am Vet Med Assoc*. 2003;222(6):762-769. doi:10.2460/javma.2003.222.762
11. Levine JM, Fosgate GT, Chen AV, et al. Magnetic resonance imaging in dogs with neurologic impairment due to acute thoracic and lumbar intervertebral disk herniation. *J Vet Intern Med*. 2009;23(6):1220-1226. doi:10.1111/j.1939-1676.2009.0393.x
12. Boekhoff TM, Flieshardt C, Ensinger EM, Fork M, Kramer S, Tipold A. Quantitative magnetic resonance imaging characteristics: evaluation of prognostic value in the dog as a translational model for spinal cord injury. *J Spinal Disord Tech*. 2012;25(3):E81-E87. doi:10.1097/BSD.0b013e31823f2f55
13. Wang-Leandro A, Siedenburg JS, Hobert MK, et al. Comparison of preoperative quantitative magnetic resonance imaging and clinical assessment of deep pain perception as prognostic tools for Early recovery of motor function in paraplegic dogs with intervertebral disk herniations. *J Vet Intern Med*. 2017;31(3):842-848. doi:10.1111/jvim.14715
14. Ranganathan P, Pramesh C, Aggarwal R. Common pitfalls in statistical analysis: logistic regression. *Perspect Clin Res*. 2017;8(3):148-151. doi:10.4103/picr.PICR_87_17
15. Moss M, Wellman DA, Cotsonis GA. An appraisal of multivariable logistic models in the pulmonary and critical care Literaturea. *Chest*. 2003;123(3):923-928. doi:10.1378/chest.123.3.923
16. Ottenbacher KJ, Ottenbacher HR, Tooth L, Ostir GV. A review of two journals found that articles using multivariable logistic regression frequently did not report commonly recommended assumptions. *J Clin Epidemiol*. 2004;57(11):1147-1152. doi:10.1016/j.jclinepi.2003.05.003
17. Bishop RC, Gutierrez-Nibeyro SD, Stewart MC, McCoy AM. Performance of predictive models of survival in horses undergoing emergency exploratory laparotomy for colic. *Vet Surg*. 2022;51(6):891-902. doi:10.1111/vsu.13839
18. Bonde A, Varadarajan KM, Bonde N, et al. Assessing the utility of deep neural networks in predicting postoperative surgical complications: a retrospective study. *Lancet Digit Health*. 2021;3(8):e471-e485. doi:10.1016/S2589-7500(21)00084-4
19. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. *Stat Methods Med Res*. 2018;27(11):3271-3285. doi:10.1177/0962280217696115
20. Hofer IS, Lee C, Gabel E, Baldi P, Cannesson M. Development and validation of a deep neural network model to predict postoperative mortality, acute kidney injury, and reintubation using a single feature set. *Npj Digit Med*. 2020;3(1):58. doi:10.1038/s41746-020-0248-0
21. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res*. 2016;18(12):e323. doi:10.2196/jmir.5870
22. Lundberg S, Lee SI. A unified approach to interpreting model predictions. Published online November 25, 2017. Accessed November 12, 2024. http://arxiv.org/abs/1705.07874
23. McCallum A, Nigam K. A comparison of event models for naive Bayes text classification. *AAAI-98 Workshop on Learning for Text Categorization*. Vol 752. Association for the Advancement of Artificial Intelligence; 1998:41-48.
24. McKinney W. Data structures for statistical computing in python. *SciPy*. 2010;445:56-61. doi:10.25080/Majora-92bf1922-00a
25. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*. 2020;17(3):261-272. doi:10.1038/s41592-019-0686-2
26. Harris CR, Millman KJ, Van Der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357-362. doi:10.1038/s41586-020-2649-2
27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2012;12:2825-2830. doi:10.48550/ARXIV.1201.0490
28. Lemaitre G, Nogueira F, Aridas CK. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*. 2016;18:1-5. doi:10.48550/ARXIV.1609.06570
29. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM; 2016:785-794.
30. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. Published online July 25, 2019. Accessed November 12, 2024. http://arxiv.org/abs/1907.10902
31. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95. doi:10.1109/MCSE.2007.55
32. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. 1997;55(1):119-139. doi:10.1006/jcss.1997.1504
33. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Dent Tech*. 1970;12(1):55-67. doi:10.1080/00401706.1970.10488634
34. Barge P, Oevermann A, Maiolini A, Durand A. Machine learning predicts histologic type and grade of canine gliomas based on MRI texture analysis. *Vet Radiol Ultrasound*. 2023;64(4):724-732. doi:10.1111/vru.13242
35. Bollig N, Clarke L, Elsmo E, Craven M. Machine learning for syndromic surveillance using veterinary necropsy reports Pławiak P, ed. *PLoS One*. 2020;15(2):e0228105. doi:10.1371/journal.pone.0228105
36. Guitiãn J, Snary EL, Arnold M, Chang Y. Applications of machine learning in animal and veterinary public health surveillance: -EN- applications of machine learning in animal and veterinary public health surveillance -FR- applications de l'apprentissage automatique dans la surveillance de la santé animale et la santé publique vétérinaire -ES- Aplicaciones del aprendizaje automático en la vigilancia zoosanitaria y de salud pÃºblica veterinaria. *Rev Sci Tech OIE*. 2023;42:230-241. doi:10.20506/rst.42.3366
37. Schofield I, Brodbelt DC, Kennedy N, et al. Machine-learning based prediction of Cushing's syndrome in dogs attending UK primary-care veterinary practice. *Sci Rep*. 2021;11(1):9035. doi:10.1038/s41598-021-88440-z
38. Tahghighi P, Appleby RB, Norena N, Ukwatta E, Komeili A. Classification of the quality of canine and feline ventrodorsal and dorsoventral thoracic radiographs through machine learning. *Vet Radiol Ultrasound*. 2024;65(4):417-428. doi:10.1111/vru.13373

39. Low D, Rutherford S. Development and assessment of a deep neural network to predict functional outcome and perioperative adverse events after decompressive thoracolumbar hemilaminectomy for canine acute intervertebral disc extrusion: a retrospective cohort study of 460 cases. *Proceedings of the 9th British Veterinary Neurology Society Symposium*. British Veterinary Neurology Society; 2024.
40. Zhang H, Zhao J, Farzan R, Alizadeh Otaghvar H. RETRACTED: risk predictions of surgical wound complications based on a machine learning algorithm: a systematic review. *Int Wound J*. 2024;21(1):e14665. doi:10.1111/iwj.14665
41. Xue B, Li D, Lu C, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open*. 2021;4(3):e212240. doi:10.1001/jamanetworkopen.2021.2240
42. Asaad M, Lu SC, Hassan AM, et al. The use of machine learning for predicting complications of free-flap head and neck reconstruction. *Ann Surg Oncol*. 2023;30(4):2343-2352. doi:10.1245/s10434-022-13053-3
43. Aikawa T, Fujita H, Kanazono S, Shibata M, Yoshigae Y. Long-term neurologic outcome of hemilaminectomy and disk fenestration for treatment of dogs with thoracolumbar intervertebral disk herniation: 831 cases (2000–2007). *J Am Vet Med Assoc*. 2012;241(12):1617-1626. doi:10.2460/javma.241.12.1617
44. Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216-1219. doi:10.1056/NEJMp1606181
45. Beam AL, Kohane IS. Big data and machine learning in health care. *Jama*. 2018;319(13):1317. doi:10.1001/jama.2017.18391
46. Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. *Inf Fusion*. 2022;81:84-90. doi:10.1016/j.inffus.2021.11.011
47. Nielsen D. *Tree boosting with XGBoost - why does XGBoost win "every" machine learning competition?*. Norwegian University of Science and Technology; 2016.
48. Ke G, Meng Q, Finley T. A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Curran Associates; 2017.
49. Sizheng Z, Boxuan H, Feng X, Dianying Z. A functional outcome prediction model of acute traumatic spinal cord injury based on extreme gradient boost. *J Orthop Surg*. 2022;17(1):451. doi:10.1186/s13018-022-03343-7
50. Chavalparit P, Wilartratsami S, Santipas B, Ittichaiwong P, Veerakanjana K, Luksanapruksa P. Development of machine-learning models to predict ambulation outcomes following spinal metastasis surgery. *Asian Spine J*. 2023;17(6):1013-1023. doi:10.31616/asj.2023.0051
51. Lakens D. The practical alternative to the *p* value is the correctly used *p* value. *Perspect Psychol Sci*. 2021;16(3):639-648. doi:10.1177/1745691620958012
52. Jeffery ND, Budke CM, Chanoit GP. What is the value of statistical testing of observational data? *Vet Surg*. 2022;51(7):1043-1051. doi:10.1111/vsu.13845
53. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. Published online March 2, 2017. Accessed November 12, 2024. http://arxiv.org/abs/1702.08608
54. Upchurch DA, Renberg WC, Turner HS, McLellan JG. Effect of duration and onset of clinical signs on short-term outcome of dogs with Hansen type I thoracolumbar intervertebral disc extrusion. *Vet Comp Orthop Traumatol*. 2020;33(3):161-166. doi:10.1055/s-0039-1700988
55. Compagnone K, Upchurch DA, Pompermaier E, Motta L. Thoracolumbar intervertebral disk extrusion in dogs: do onset of clinical signs, time of surgery, and neurological grade matter? *Vet Comp Orthop Traumatol*. 2023;36(6):287-293. doi:10.1055/s-0043-1770355
56. Freeman P, Jeffery ND. Re-opening the window on fenestration as a treatment for acute thoracolumbar intervertebral disc herniation in dogs. *J Small Anim Pract*. 2017;58(4):199-204. doi:10.1111/jsap.12653
57. Jeffery ND, Freeman PM. The role of fenestration in Management of Type I Thoracolumbar Disk Degeneration. *Vet Clin North Am Small Anim Pract*. 2018;48(1):187-200. doi:10.1016/j.cvsm.2017.08.012
58. Bergknut N, Auriemma E, Wijsman S, et al. Evaluation of intervertebral disk degeneration in chondrodystrophic and nonchondrodystrophic dogs by use of Pfirrmann grading of images obtained with low-field magnetic resonance imaging. *Am J Vet Res*. 2011;72(7):893-898. doi:10.2460/ajvr.72.7.893
59. Balducci F, Canal S, Contiero B, Bernardini M. Prevalence and risk factors for presumptive ascending/descending myelomalacia in dogs after thoracolumbar intervertebral disk herniation. *J Vet Intern Med*. 2017;31(2):498-504. doi:10.1111/jvim.14656
60. Moore SA, Early PJ, Hettlich BF. Practice patterns in the management of acute intervertebral disc herniation in dogs. *J Small Anim Pract*. 2016;57(8):409-415. doi:10.1111/jsap.12496
61. Noyes JA, Thomovsky SA, Chen AV, et al. Magnetic resonance imaging versus computed tomography to plan hemilaminectomies in chondrodystrophic dogs with intervertebral disc extrusion. *Vet Surg*. 2017;46(7):1025-1031. doi:10.1111/vsu.12700
62. Bailie C, Sousa F, Madden M, Nagendran A, De Fria J. Reliability of deep pain perception test prior to referral for dogs diagnosed with thoracolumbar disc extrusion. *Proceedings of the 36th Symposium of the European Society of Veterinary Neurology-European College of Veterinary Neurology (ESVN-ECVN)*. European Society of Veterinary Neurology-European College of Veterinary Neurology; 2024.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Low D, Stables S, Kondrotaite L, Garland B, Rutherford S. Machine-learning-based prediction of functional recovery in deep-pain-negative dogs after decompressive thoracolumbar hemilaminectomy for acute intervertebral disc extrusion. *Veterinary Surgery*. 2025;54(4):665-674. doi:10.1111/vsu.14250