

XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury

Tomoo Inoue,^{1,*} Daisuke Ichikawa,² Taro Ueno,² Maxwell Cheong,³ Takashi Inoue,¹ William D. Whetstone,⁴ Toshiki Endo,^{1,5} Kuniyasu Nizuma,⁵⁻⁷ and Teiji Tominaga⁵

Abstract

The accurate prediction of neurological outcomes in patients with cervical spinal cord injury (SCI) is difficult because of heterogeneity in patient characteristics, treatment strategies, and radiographic findings. Although machine learning algorithms may increase the accuracy of outcome predictions in various fields, limited information is available on their efficacy in the management of SCI. We analyzed data from 165 patients with cervical SCI, and extracted important factors for predicting prognoses. Extreme gradient boosting (XGBoost) as a machine learning model was applied to assess the reliability of a machine learning algorithm to predict neurological outcomes compared with that of conventional methodology, such as a logistic regression or decision tree. We used regularly obtainable data as predictors, such as demographics, magnetic resonance variables, and treatment strategies. Predictive tools, including XGBoost, a logistic regression, and a decision tree, were applied to predict neurological improvements in the functional motor status (ASIA [American Spinal Injury Association] Impairment Scale [AIS] D and E) 6 months after injury. We evaluated predictive performance, including accuracy and the area under the receiver operating characteristic curve (AUC).

Regarding predictions of neurological improvements in patients with cervical SCI, XGBoost had the highest accuracy (81.1%), followed by the logistic regression (80.6%) and the decision tree (78.8%). Regarding AUC, the logistic regression showed 0.877, followed by XGBoost (0.867) and the decision tree (0.753). XGBoost reliably predicted neurological alterations in patients with cervical SCI. The utilization of predictive machine learning algorithms may enhance personalized management choices through pre-treatment categorization of patients.

Keywords: cervical spinal cord injury; extreme gradient boosting; machine learning; receiver operating curve

Introduction

Cervical spinal cord injury (SCI) leads to poor neurological disabilities that are associated with a deteriorated quality of life and higher rate of unemployment.^{1,2} In the hospital, these patients are assessed

neurologically or radiographically to ascertain their neurological impairments and select adequate treatment strategies. Accurate predictions of neurological outcomes are important for effectively maximizing limited medical resources. Therefore, a dependable

¹Department of Neurosurgery, National Health Organization Sendai Medical Center, Sendai, Miyagi, Japan.

²SUSMED, Inc., Tokyo, Japan.

³Department of Radiology, Stanford University School of Medicine, Palo Alto, California, USA.

⁴Department of Emergency Medicine, University of California, San Francisco, San Francisco, California, USA.

⁵Department of Neurosurgery, ⁷Department of Neurosurgical Engineering and Translational Neuroscience, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan.

⁶Department of Neurosurgical Engineering and Translational Neuroscience, Graduate School of Biomedical Engineering, Tohoku University, Sendai, Miyagi, Japan.

*Address correspondence to Tomoo Inoue, MD, PhD, Department of Neurosurgery, National Health Organization Sendai Medical Center, 2-11-12, Miyaginohara, Miyagino-ku, Sendai, 983-8520, Japan, Email: tomoo49@gmail.com



outcome prediction model is crucial for estimating recovery following SCI and assisting family or informal caregivers in providing personalized care.

Compared with conventional statistical models, such as a logistic regression analysis, machine learning prediction models detect non-linear interactions among prognostic factors.³ Previous studies reported that machine learning models were useful for predicting the outcomes of various clinical entities, such as traumatic brain injury,⁴ congestive heart failure,⁵ sepsis,⁶ asthma,⁷ and chronic obstructive pulmonary disease,⁸ or of intensive care.⁹ Some prognostic models have been validated to optimize management. Also, the requirement for effective outcome prediction in patients with SCI has increased numerous research studies evaluating the efficacy of machine learning algorithms for this cohorts.^{10–14}

Among different machine learning systems, extreme gradient boosting (XGBoost) is widely used to accomplish state-of-the-art analyses in diverse fields with good accuracy or area under the receiver operating characteristic curve (AUC).^{15,16} XGBoost, a decision-tree-based ensemble machine learning algorithm with a gradient boosting framework, was developed by Chen and Guestrin.¹⁷ It has since been used in traffic census and the field of energy consumption.^{18,19} This is the first study to examine the efficacy of XGBoost for predicting neurological outcomes in patients with cervical SCI. The study's purpose was not to improve prognostic models based on a large number of predictor variables, but to innovate machine learning models based on XGBoost using clinical information regularly obtained from patients with SCI on admission.

Methods

Study design, ethical approval, and setting

We retrospectively identified patients with a principal diagnosis of SCI (code S14) according to the *International Classification of Diseases, 10th Revision, Clinical Modifications*, from diagnosis codes on admission. We also included patients with SCI who were initially diagnosed at a local hospital and then transferred to our hospital (National Health Organization Sendai Medical Center) for intensive care. To decrease possible confounding factors due to different surgical strategies, we excluded all patients who underwent surgery in other hospitals. We also excluded patients with a neurological disease or deficits (e.g., Parkinson's disease or stroke) prior to injury. Research procedures were approved by the Institutional Review Board of Sendai Medical Center, which exempted us from the need to obtain consent from individual participants.

Data collection

Basic data were attained from the Sendai Medical Center's Department of Neurosurgery database, and were cross-referenced with trauma records and searchable terms in electronic medical records. Patient demographic data were routinely recorded in our department during the study period, and a data dictionary was utilized to assure consistent data sharing across sites. After collection, all data were evaluated for completeness and accuracy, and then anonymized before investigation. Data were acquired on age, sex, previous medical history, neurological severity, magnetic resonance imaging (MRI) findings, and surgical procedures. Following an intensive review of all variables in database files, we selected 44 basic variables and categorized them into independent categories: demographics and neurological status (8 features), mechanisms of injury (1 feature), treatment strategies (7 features), radiographic information (14 features), and concomitant degenerative spine disease (7 features) (Table 1). To reduce selection bias, the authors responsible for chart reviews were blinded to neurological outcomes.

All patients were evaluated by our multi-disciplinary team immediately after being transferred to our hospital; this team consisted of board-certified neurosurgeons, emergency physicians, and radiologists. All patients were evaluated using the ASIA (American Spinal Injury Association) Impairment Scale (AIS). A complete radiological evaluation, including standard radiographs and computed tomography, was performed for each patient to assess the degree of compression and injury to the spinal cord. Patients also underwent MRI within 24 h of the traumatic event, including T1- and T2-weighted imaging (WI) of the cervical spine in both axial and sagittal views. MRI was performed using the 1.5 Tesla system (Magnetom Avanto, Siemens Medical Solutions, Erlangen, Germany). Previous comorbidities were recorded based on patient self-reports during hospitalization or medical histories in electronic records, using the Charlson Comorbidity Index (CCI) as a measure of clinical importance.^{20,21}

Criteria for surgical decompression

Indications for surgical decompression were the existence of surgically amenable cervical spinal cord compression due to cervical spondylosis, ossification of the posterior longitudinal ligament, or cervical disc herniation, which were considered to be responsible for neurological impairments. The timing of surgical decompression depended on the patient's condition and



Table 1. Patient Characteristics

Forty-four predictors	
Demographics and neurological status (8)	
Age (years), mean (SD)	65.3 (15.4)
Sex (n)	132 males, 33 females
Height, mean (SD)	164.2 (10)
Body weight, mean (SD)	64.1 (14.0)
Body mass index, mean (SD)	24.1 (7.6)
Body surface area, mean (SD)	1.70 (0.24)
American Spinal Injury Association Impairment Scale (AIS) (n)	AIS A = 15, B = 38, C = 66, D = 44, E = 3
Charlson Comorbidity Index, median (IQR)	0 (0 - 1)
Mechanism of injury (8)	
Slip (n)	78
Fall (n)	36
Loss of consciousness (n)	10
Motor vehicle collision (n)	22
Bicycle (n)	11
Sports (n)	8
Spinal cord injury after alcohol consumption (n)	37
Spinal cord injury with traumatic brain injury (n)	137
Therapeutic strategies for spinal cord injury (7)	
Surgical timing, median (days) (IQR)	2 (1 - 7)
Conservative therapy (n)	36
Anterior cervical discectomy and fusion (n)	43
Posterior fixation (n)	6
Laminoplasty (n)	82
Halo-vest stabilization (n)	3
Methylprednisolone use (n)	39
Radiographic information (14)	
Brain and Spinal Cord Injury Center score (BASIC) (n)	BASIC 0 = 21, 1 = 61, 2 = 46, 3 = 24, 4 = 14
Longest measurements of T2 hyperintensity on the sagittal plane (mm), mean (SD)	14.4 (13.7)
Sagittal grading, median (IQR)	2 (2 - 2.25)
Subaxial Injury and Classification system, median (IQR)	6 (5 - 6)
Maximum canal compromise, mean (SD)	75.9 (3.04)
Maximum spinal cord compression, mean (SD)	79.9 (4.56)
Signal intensity at the narrowest level on T1-weighted images, mean (SD)	293.0 (69.3)
Signal intensity at the narrowest level on T2-weighted images, mean (SD)	389.0 (152.2)
Signal intensity at the C7-T1 disc levels on T1-weighted images, mean (SD)	276.7 (66.8)
Signal intensity at the C7-T1 disc levels on T2-weighted images, mean (SD)	268.9 (98.1)
Signal intensity ratio on T1-weighted images, mean (SD)	1.07 (0.15)
Signal intensity ratio on T2-weighted images, mean (SD)	1.40 (0.37)
Cervical alignment (n)	Lordotic (73), reverse S-shape (37), straight (27), kyphosis (15), dislocation (8)
High cervical (n)	36
Concomitant degenerative spine diseases (7)	
Cervical spondylosis	144
Ossification of the posterior longitudinal ligament	52
Cervical disc herniation	64
Osteophyte	115
Ossification of the yellow ligament	2
Ankylosing spondylitis	20
Atlantoaxial dislocation	6

IQR, interquartile range; SD, standard deviation.

any comorbidities, radiological evaluation, and optimum preparation of surgical suites. The surgical approach selected was based on the finding of cord compression and the surgeon's preference. Further, in our institution, we do not regard age as an exclusion criterion for early surgery, and, thus, we often perform this surgery independent of age. Post-operatively, all patients underwent early rehabilitation consisting of physical therapy that was immediately initiated after cardiopulmonary stabilization.

Statistical analysis

Two attending neurosurgeons performed consensus MRI ratings for all metrics while blinded to neurological outcomes. We applied the following axial scoring system, known as the Brain and Spinal Injury Center (BASIC) score²²: grade 0, no intramedullary signal abnormality; grade 1, T2 hyperintensity confined to the gray matter; grade 2, intramedullary T2 hyperintensity extending beyond possible spinal gray matter margins to include the white matter, but not containing the whole transverse extent of the spinal cord; grade 3, intramedullary T2 hyperintensity containing the entire axial level of the spinal cord; and grade 4, a grade 3 injury plus a distinct T2 hypointense area, consistent with macroscopic intramedullary hemorrhage.^{23,24} The longitudinal extent of T2 hyperintensity (in millimeters) was evaluated based on the National Institutes of Health/National Institutes of Neurological Disorders and Stroke SCI common data elements version 1.0.^{22,25,26} Sagittal grading was evaluated based on previous studies^{22,25}: grade 1, no spinal cord abnormal intensity; grade 2, one-level T2 hyperintensity; grade 3, more than a two-level T2 signal hyperintensity; and grade 4, T2 signal hyperintensity with lesions of hypointensity indicating hemorrhage.

We also evaluated the Subaxial Injury and Classification (SLIC) system, which was scored based on the importance of three factors associated with the treatment of cervical injuries: morphology, neurological status, and the integrity of the discoligamentous complex.²⁷ Maximum canal compromise (MCC) and maximum spinal cord compression (MSCC) were evaluated on mid-sagittal images by differentiating the anteroposterior diameter of the canal (on sagittal T1WI for MCC) and that of the spinal cord (on sagittal T2WI for MSCC) by means of the canal or spinal cord above and below as reported previously.^{25,28} The signal intensity ratio (SIR) at the narrowest level of the spinal cord on sagittal views of T1WI and T2WI was measured,



and regions of interest (ROIs) were acquired by 0.05 cm². Normal cord SIs at the C7-T1 disc level were acquired, and ROIs were acquired by 0.3 cm². SIRs between regions of 0.05 and 0.3 cm² were calculated. SIRs on T1WI and T2WI were calculated using the following equation²⁹:

$$\text{SIR} = (\text{SI}[0.05\text{cm}^2]) / (\text{SI of the sagittal normal cord between the C7 and T1 disc levels } [0.3\text{ cm}^2]).$$

Radiographs were also obtained using normal radiographic methods in which the tube was positioned on the C5 disc. The radiographic film cassette was 150 cm from the tube.³⁰ Study participants were categorized into four groups based on differences in alignment in the upright position: lordotic, straight type, kyphotic, S-shape curvature, and dislocation.

Machine learning

By using machine learning algorithms, we built prediction models for neurological improvements evaluated 6 months after injury based on the AIS. We dichotomized the scale as follows: AIS D or E as 1 and AIS A, B, or C as 0.

As predictors for prediction models, we included routinely obtained clinical data on admission, such as age, sex, severity of neurological impairments based on the AIS, and several MRI findings. All predictors are shown in Table 1.

We built multiple prediction models using XGBoost and logistic regressions and evaluated them by 8-fold cross validation. XGBoost is an ensemble learning algorithm and applies decision trees as base learners.¹⁷⁻¹⁹ A logistic regression analysis is a well-known method for building clinical prediction models.^{31,32} It is a type of generalized linear model and features are additively and linearly built into the model.

Evaluation and variable importance

To evaluate prediction models, we drew a receiver operating characteristic curve (ROC curve) and calculated the area under the ROC curve (AUC). We made a confusion matrix and calculated accuracy, the true-positive rate, and the false-positive rate, as follows:

Confusion Matrix

	Actual: Positive (AIS D/E)	Actual: Negative (AIS A/B/C)
Prediction: Positive	TP (true-positive rate)	FP (false-positive rate)
Prediction: Negative	FN (false-negative rate)	TN (true-negative rate)

$$\text{Accuracy}(\%) = \frac{TP + TN}{TP + TN + FP + FN}$$

We acquired the variable importance of each predictor from the XGBoost model. Variable importance indicates the usefulness of each predictor for the prediction model and was calculated for a single decision tree based on the amount that each attribute split-point improved the performance measure, weighted by the number of observations for which the node was responsible.³

Results

Baseline characteristics of participants

During the study period, 165 patients (132 men and 33 women) aged 16 to 93 years (median, 68 years) and diagnosed with cervical SCI were examined. Key demographic, clinical, and outcome parameters in the present study are summarized in Table 1.

Comparison with other prediction models

We performed three algorithms utilizing the training set to obtain better predictors by restoring the parameters of each algorithm and adjusted the predictors based on the validation set. Table 2 shows the prediction capability (accuracy and AUC) of each algorithm using the optimal features subset. As shown in Figure 1, XGBoost and the logistic regression predicted neurological recovery with an AUC greater than 0.800, and XGBoost showed the best performance for outcome predictions. It had the highest accuracy, 81.1%, followed by the logistic regression (80.6%) and decision tree (78.8%). Regarding AUC, the logistic regression showed 0.877, followed by XGBoost (0.867) and the decision tree (0.753).

Variable importance

XGBoost calculates feature importance via the Gini index. To clarify the importance of each predictor, Figure 2 recapitulates the final 15 most significant variables of XGBoost after the exclusion of 30 unimportant predictors. The top 15 predictors, as scored by XGBoost, are as follows:

1. Demographics and neurological status (4): age, AIS B, C, and D
2. Mechanisms of injury (0)
3. Treatment strategies (0)
4. Radiographic information (11): BASIC 1, 3, and 4, longest measurements of T2 hyperintensity on the sagittal plane; MCC, MSCC, SIR at the narrowest level on T1WI and T2WI; SI at C7-T1 on T1WI and T2WI; and the reverse S-shape alignment
5. Concomitant degenerative spine disease (0)



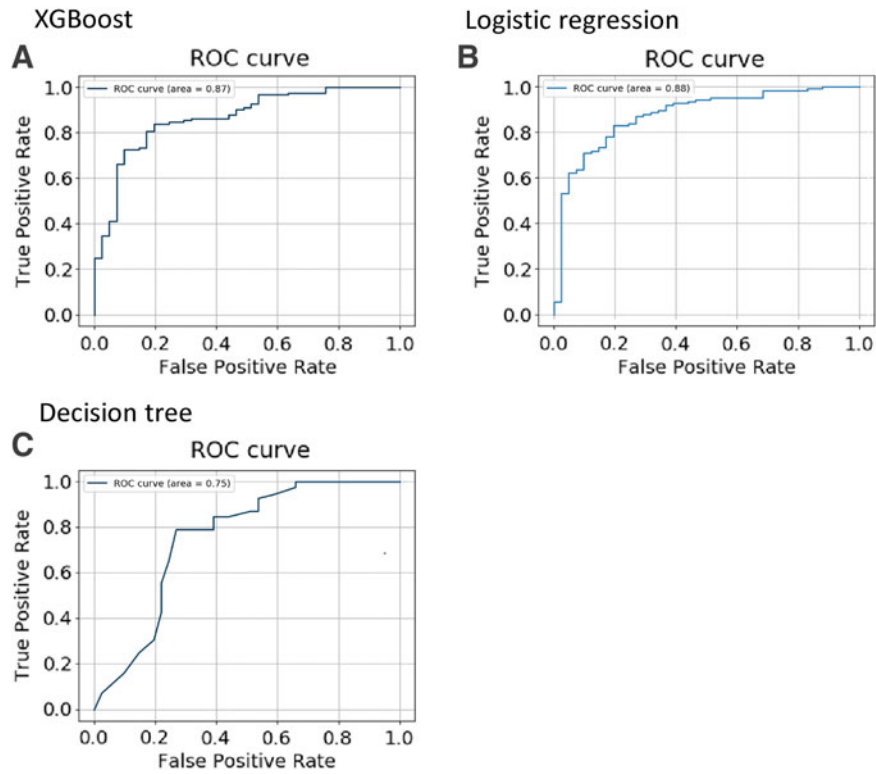


FIG. 1. Receiver operating characteristic curves for models with all algorithms as inputs.

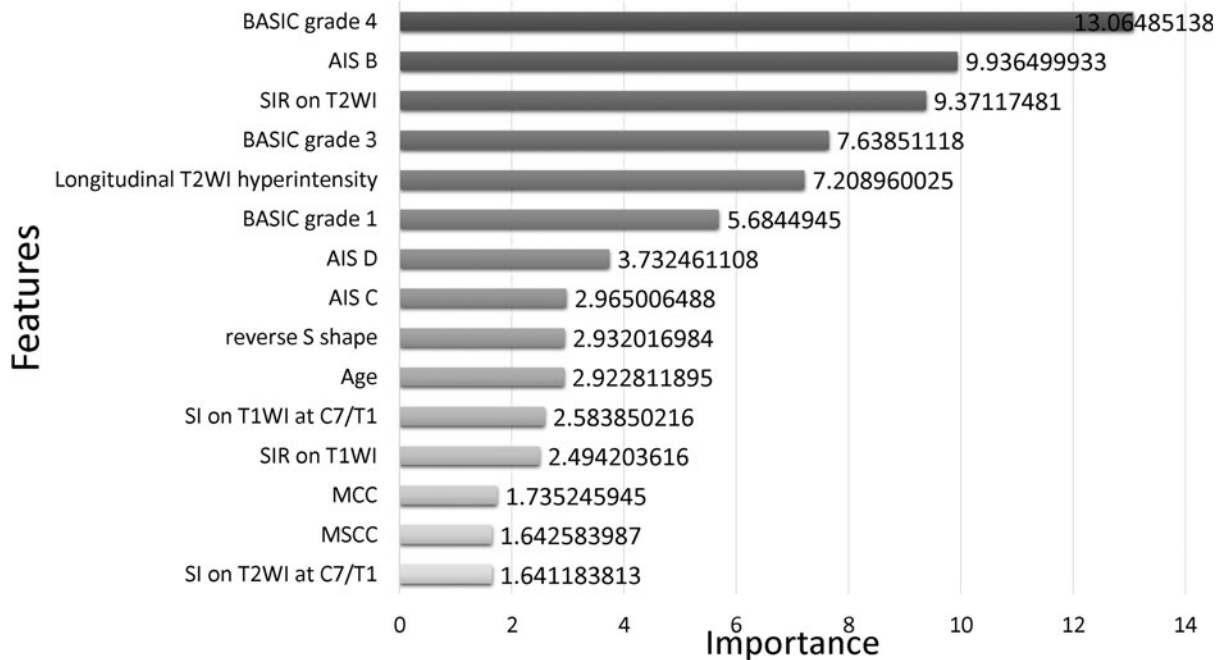


FIG. 2. Feature importance of factors predicting neurological improvements in XGBoost. The top 14 features of importance are shown from high to low.



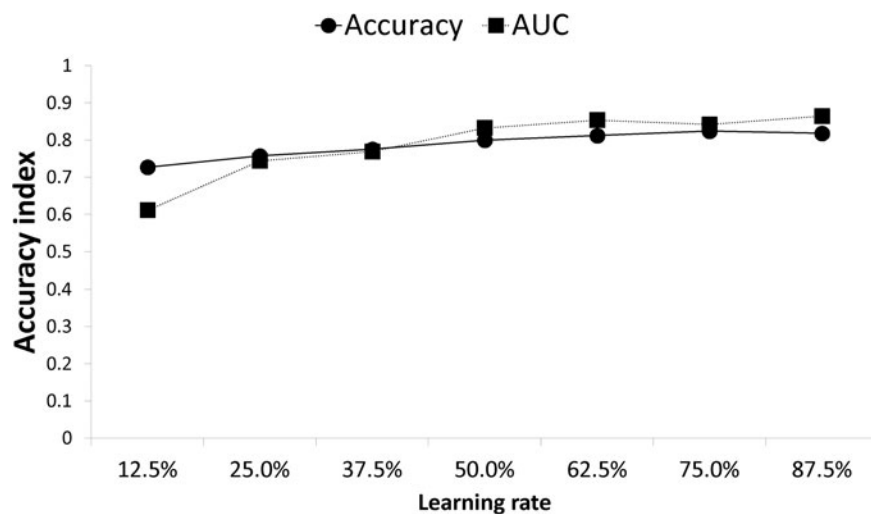


FIG. 3. Relationship between accuracy and the number of evaluators in XGBoost.

In this model, the most important predictive variable was a BASIC score of 4, followed by AIS B, SIR on T2WI, and a BASIC score of 3 as the most significant characteristics for neurological improvements. In comparison with the two other traditional machine learning models, the accuracy rate of XGBoost was satisfactory, and the XGBoost model showed good outcomes on the ROC curve. SIR on T1WI is the most important feature of a logistic regression model, but it contributed only slightly to XGBoost. Figure 3 shows the relationship between accuracy and the number of evaluators in XGBoost. With increases in the number of predictors, computational precision is beyond 0.800 and stable at approximately 0.864, which indicates that the XGBoost algorithm accomplished sustainable predictions. When the estimates surpass 0.875, the calculation exactness of XGBoost declines, suggesting a small number of estimators.

The confusion matrix that resulted from the prediction of neurological improvements based on XGBoost, the logistic regression, and the decision tree corresponded to true-positive rates of 0.833, 0.909, and

0.866, respectively, and false-positive rates of 0.560, 0.243, and 0.415, respectively (Table 2).

Discussion

Results of the present study

Recovery from cervical SCI involves important tasks and significant choices to effectively utilize limited medical resources. The application of clinical information may enhance the accuracy of outcome predictions in patients with SCI. In the present study in 165 patients with cervical SCI, we applied XGBoost to regularly obtained clinical data and achieved greater prediction accuracy than that using two other models: an ordinal logistic regression analysis and a decision tree. Predictive variables analyzed according to statistical calculations for uncomplicatedness utilize simple limited fundamental variables, with measurements being approximately smoothed. Based on this algorithm, clinicians may individualize the management of patients with SCI based on their neurological alterations, which may efficiently reduce medical expenses and establish predictions for personalized neurotherapeutics for these patients.

Table 2. Confusion Matrix for XGBoost, a Logistic Regression, and a Decision Tree

	XGBoost			Logistic regression			Decision tree		
	Actual positive (AIS D/E)	Actual negative (AIS A/B/C)	Total	Actual positive (AIS D/E)	Actual negative (AIS A/B/C)	Total	Actual positive (AIS D/E)	Actual negative (AIS A/B/C)	Total
Prediction: positive	115 (TP)	23 (FP)	138	101 (TP)	10 (FP)	111	110 (TP)	17 (FP)	127
Prediction: negative	9 (FN)	18 (TN)	27	23 (FN)	31 (TN)	54	14 (FN)	24 (TN)	38
Total	124	41	165	124	41	165	124	41	165

AIS, American Spinal Injury Association Impairment Scale; FN, false-negative; FP, false-positive; TN, true-negative; TP, true-positive.



The results of statistical analyses indicate that the majority of the selected features were instructive for predicting neurological improvements. Based on the XGBoost model, severe axial damage with a BASIC score of 4, AIS B, SIR on the T2WI scale, and BASIC grade 3 were strong predictors, in this order of importance. Previous studies reported that the assessment of intramedullary T2 signal abnormalities in the axial plane according to BASIC scores²² or longitudinal lesion lengths³³ may provide important information for predicting neurological outcomes in patients with SCI. In the present study, we demonstrated that among various features, a BASIC score of 4 was the most predictive of the outcome. Talbott and colleagues reported that all patients with a BASIC score of 4 were discharged with an unchanged AIS A.²² In our series, 14 patients received a BASIC score of 4: 9 were AIS A and 5 were ASIA B, and none reached AIS D or E. On the other hand, among 23 patients with a BASIC score of 3, 10 (43.5 %) reached AIS D or E. XGBoost may successfully provide surgeons with selection strategies based on the potential for neurological improvements in patients with SCI.

Various factors contributed to the advancement of XGBoost in prediction modeling. State-of-art machine learning algorithms, such as XGBoost, have the capacity to analyze complex non-linear relationships among various clinical factors.^{34,35} Further, XGBoost may subjectively evaluate a number of clinical prognostic factors that were previously investigated.^{22,36–38} In addition, although overfitting is a common limitation in refined non-linear machine learning algorithms, XGBoost supervises machine learning problems by parallel computing, regularization, cross validation, flexibility, or availability.^{16,39,40}

Comparison with previous research

DeVries and associates previously reported that no clinically significant differences were observed between the use of unsupervised machine learning with complete admission neurological information and established standards.¹⁰ They showed the inherent weakness of applying AUC to imbalanced data sets and outlined a new strategy to evaluate performance.¹⁰ Tay and co-workers proposed a machine learning technique for the diagnosis of SCI using diffusion tensor imaging.¹¹ They developed a classification scheme for identifying healthy individuals and patients, and reported normal case specificity of 0.912 and abnormal case sensitivity of 0.952.¹¹ Khan and colleagues speculated that machine learning-based predictions will become a crucial algorithm in treatment

modalities employed by spinal surgeons.¹² Machine learning has potential and future applicability in multiple clinically significant domains due to its novelty and computational power in the area of SCI.¹² Schwartz and associates reported that machine learning may effectively harness the value of electronic medical records in spine surgery because of developments in algorithms in reading images and in the ability to predict clinical outcomes of patients.¹³ McCoy and co-workers stated that targeted convolutional neural network training in SCI improves algorithm performance for this cohort and provides clinically relevant metrics of cord injury.⁴¹

In future studies, we aim to address the following. First, because XGBoost is a method for optimization, an efficient approach needs to be developed to achieve superior prognostic validity. In addition, we need to confirm whether other developed categorical procedures have a superior prognostic ability to provide clinicians with further state-of-the-art decision-making modalities. More patient information needs to be collected from medical record resources for analyses of the generalization capability of the present algorithm. Significant and precise outcome predictions may be performed when various machine learning systems, including XGBoost, are utilized in diverse clinical areas.

The present study may have been limited by the general validation of prognostic models for other data sets. Our prognostic model was produced using data from a single institution, and this needs to be considered if the model is employed in other hospitals with different treatment procedures or patient backgrounds, because it may lead to invalid prognostic importance. Further, we omitted patients with missing data, therefore logistic regression might be better than XGBoost with respect to AUC. Nusinovi and colleagues have stated that low dimensional settings include low number of events and predictors, so in such settings, logistic regression yields performance as good as machine learning models.⁴²

To overcome this situation, we need to perform multi-center trials to obtain more data sets. Although further examinations on novel acute neurochemical biomarkers, such as S100 β , neurofilaments, and glial fibrillary acidic protein, may increase accuracy,⁴³ this method may not be practical in selection strategies for neurointensive treatments because of low specificity or potential cross contamination by hemolysis. However, analyzed and non-analyzed populations generally have similar backgrounds, clinical findings, or neurological outcomes. In addition, treatment strategies for patients with SCI frequently rely on a surgeon's preference. The present



results showed that surgical timing did not play a major role in predicting neurological alterations assessed by AIS. However, because the importance of surgical intervention has been widely reported in many cases,^{44–49} this result may have been due to the retrospective nature of the present study affecting surgical timing, which was a possible bias. The present study was also restricted by its dependence on neurological outcomes based on AIS; this was due to the effect of selection bias that reduced the statistical power. Further studies that include other outcome measures, such as functional, psychosocial, sexual health, autonomic, bowel and bladder, and pain tools, are needed.^{10,50}

In conclusion, the present study results revealed the potential of XGBoost to predict neurological alterations prior to treatments. By considering neurological recovery in patients with SCI before surgery, we may provide appropriate individualized management strategies for these patients. The present results are promising and represent the primary step for improving prognostic models that may be applied to the management of SCI in patients with a strong possibility of neurological recovery.³³

Acknowledgments

The authors thank Medical English Service for the English language review.

Author Contributions

Conception of the work: all authors; acquisition of data: Inoue, Endo, Nizuma; drafting of the manuscript: Inoue, Ichikawa, Ueno, Cheong; critical revisions: all authors; approval of the final version to submit for publication: all authors; agreement to be accountable for all aspects of the work: all authors.

Funding Information

This article is based on results obtained from a project, P15009, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Author Disclosure Statement

No competing financial interests exist.

References

1. Noonan, V.K., Fingas, M., Farry, A., Baxter, D., Singh, A., Fehlings, M.G., and Dvorak, M.F. (2012). Incidence and prevalence of spinal cord injury in Canada: a national perspective. *Neuroepidemiology* 38, 219–226.
2. Hiremath, S.V., Hogaboom, N.S., Roscher, M.R., Worobey, L.A., Oyster, M.L., and Boninger, M.L. (2017). Longitudinal prediction of quality-of-life scores and locomotion in individuals with traumatic spinal cord injury. *Arch. Phys. Med. Rehabil.* 98, 2385–2392.
3. Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.
4. Matsuo, K., Aihara, H., Nakai, T., Morishita, A., Tohma, Y., and Kohmura, E. (2019). Machine Learning to Predict In-Hospital Morbidity and Mortality after Traumatic Brain Injury. *J. Neurotrauma* 37, 202–210.
5. Mortazavi, B.J., Downing, N.S., Buchholz, E.M., Dharmarajan, K., Manhapra, A., Li, S.X., Negahban, S.N., and Krumholz, H.M. (2016). Analysis of machine learning techniques for heart failure readmissions. *Circ. Cardiovasc. Qual. Outcomes* 9, 629–640.
6. Taylor, R.A., Pare, J.R., Venkatesh, A.K., Mowafi, H., Melnick, E.R., Fleischman, W., and Hall, M.K. (2016). Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. *Acad. Emerg. Med.* 23, 269–278.
7. Finkelstein, J., and Gangopadhyay, A. (2007). Using machine learning to predict asthma exacerbations. *AMIA Annu. Symp. Proc.*, 955.
8. Altan, G., Kutlu, Y., and Allahverdi, N. (2019). Deep Learning on Computerized Analysis of Chronic Obstructive Pulmonary Disease. *IEEE J Biomed Health Inform.* doi: 10.1109/JBHI.2019.2931395 [Epub ahead of print].
9. Wellner, B., Grand, J., Canzone, E., Coarr, M., Brady, P.W., Simmons, J., Kirkendall, E., Dean, N., Kleinman, M., and Sylvester, P. (2017). Predicting unplanned transfers to the intensive care unit: a machine learning approach leveraging diverse clinical elements. *JMIR Med. Inform.* 5, e45.
10. DeVries, Z., Hoda, M., Rivers, C.S., Maher, A., Wai, E., Moravek, D., Stratton, A., Kingwell, S., Fallah, N., Paquet, J., Phan, P., and Network, R. (2019). Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *Spine J.* 20, 213–224.
11. Tay, B., Hyun, J.K., and Oh, S. (2014). A machine learning approach for specification of spinal cord injuries using fractional anisotropy values obtained from diffusion tensor images. *Comput. Math. Methods Med.* 2014, 276589.
12. Khan, O., Badhiwala, J.H., Wilson, J.R.F., Jiang, F., Martin, A.R., and Fehlings, M.G. (2019). Predictive modeling of outcomes after traumatic and nontraumatic spinal cord injury using machine learning: review of current progress and future directions. *Neurospine* 16, 678–685.
13. Schwartz, J.T., Gao, M., Geng, E.A., Mody, K.S., Mikhail, C.M., and Cho, S.K. (2019). Applications of machine learning using electronic medical records in spine surgery. *Neurospine* 16, 643–653.
14. Phan, P., Budhram, B., Zhang, Q., Rivers, C.S., Noonan, V.K., Plashkes, T., Wai, E.K., Paquet, J., Roffey, D.M., Tsai, E., and Fallah, N. (2019). Highlighting discrepancies in walking prediction accuracy for patients with traumatic spinal cord injury: an evaluation of validated prediction models using a Canadian Multicenter Spinal Cord Injury Registry. *Spine J.* 19, 703–710.
15. Mo, X., Chen, X., Li, H., Li, J., Zeng, F., Chen, Y., He, F., Zhang, S., Li, H., Pan, L., Zeng, P., Xie, Y., Li, H., Huang, M., He, Y., Liang, H., and Zeng, H. (2019). Early and accurate prediction of clinical response to methotrexate treatment in juvenile idiopathic arthritis using machine learning. *Front. Pharmacol.* 10, 1155.
16. Liu, L., Yu, Y., Fei, Z., Li, M., Wu, F.X., Li, H.D., Pan, Y., and Wang, J. (2018). An interpretable boosting model to predict side effects of analgesics for osteoarthritis. *BMC Syst. Biol.* 12, 105.
17. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *arXiv* (2016), 1–6. *arXiv preprint arXiv:1603.02754*.
18. Chakraborty, D., and Elzarka, H. (2019). Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold. *Energy and Buildings* 185, 326–344.
19. Alajali, W., Zhou, W., Wen, S., and Wang, Y. (2018). Intersection traffic prediction using decision tree models. *Symmetry* 10, 386.
20. Deyo, R.A., Cherkin, D.C., and Ciol, M.A. (1992). Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J. Clin. Epidemiol.* 45, 613–619.
21. Charlson, M.E., Pompei, P., Ales, K.L., and MacKenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* 40, 373–383.
22. Talbott, J.F., Whetstone, W.D., Readdy, W.J., Ferguson, A.R., Bresnahan, J.C., Saigal, R., Hawryluk, G.W., Beattie, M.S., Mabray, M.C., Pan, J.Z., Manley, G.T., and Dhall, S.S. (2015). The Brain and Spinal Injury Center score: a novel, simple, and reproducible method for assessing the severity of acute cervical spinal cord injury with axial T2-weighted MRI findings. *J. Neurosurg. Spine* 23, 495–504.
23. Bozzo, A., Marcoux, J., Radhakrishna, M., Pelletier, J. and Goulet, B. (2011). The role of magnetic resonance imaging in the management of acute spinal cord injury. *J. Neurotrauma* 28, 1401–1411.



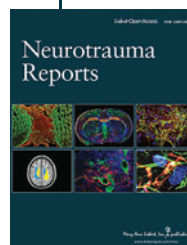
24. Bondurant, F.J., Cotler, H.B., Kulkarni, M.V., McArdle, C.B., and Harris, J.J. (1990). Acute spinal cord injury. A study using physical examination and magnetic resonance imaging. *Spine* 15, 161–168.
25. Haefeli, J., Mabray, M.C., Whetstone, W.D., Dhall, S.S., Pan, J.Z., Upadhyayula, P., Manley, G.T., Bresnahan, J.C., Beattie, M.S., and Ferguson, A.R. (2017). Multivariate analysis of MRI biomarkers for predicting neurologic impairment in cervical spinal cord injury. *Am. J. Neuroradiol.* 38, 648–655.
26. Biering-Sørensen, F., Alai, S., Anderson, K., Charlifue, S., Chen, Y., DeVivo, M., Flanders, A.E., Jones, L., Kleitman, N., and Lans, A. (2015). Common data elements for spinal cord injury clinical research: a National Institute for Neurological Disorders and Stroke project. *Spinal Cord* 53, 265–277.
27. McCormick, P.C. (2002). Guidelines for the management of acute cervical spine and spinal cord injuries. *Neurosurgery* 50, Sii–Siii.
28. Fehlings, M.G., Rao, S.C., Tator, C.H., Skaf, G., Arnold, P., Benzell, E., Dickman, C., Cuddy, B., Green, B., and Hitchon, P. (1999). The optimal radiologic method for assessing spinal canal compromise and cord compression in patients with cervical spinal cord injury: Part II: results of a multicenter study. *Spine* 24, 605–613.
29. Uchida, K., Nakajima, H., Takeura, N., Yayama, T., Guerrero, A.R., Yoshida, A., Sakamoto, T., Honjoh, K., and Baba, H. (2014). Prognostic value of changes in spinal cord signal intensity on magnetic resonance imaging in patients with cervical compressive myelopathy. *Spine J.* 14, 1601–1610.
30. Takeshima, T., Omokawa, S., Takaoka, T., Araki, M., Ueda, Y., and Takakura, Y. (2002). Sagittal alignment of cervical flexion and extension: lateral radiographic analysis. *Spine* 27, E348–E355.
31. Bagley, S.C., White, H., and Golomb, B.A. (2001). Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. *J. Clin. Epidemiol.* 54, 979–985.
32. Deo, R.C. (2015). Machine learning in medicine. *Circulation* 132, 1920–1930.
33. Aarabi, B., Sansur, C.A., Ibrahim, D.M., Simard, J.M., Hersh, D.S., Le, E., Diaz, C., Massetti, J., and Akhtar-Danesh, N. (2017). Intramedullary lesion length on postoperative magnetic resonance imaging is a strong predictor of ASIA Impairment Scale grade conversion following decompressive surgery in cervical spinal cord injury. *Neurosurgery* 80, 610–620.
34. Goto, T., Camargo, C.A., Jr., Faridi, M.K., Yun, B.J., and Hasegawa, K. (2018). Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *Am. J. Emerg. Med.* 36, 1650–1654.
35. Goto, T., Camargo, C.A., Jr., Faridi, M.K., Freishtat, R.J., and Hasegawa, K. (2019). Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw. Open* 2, e186937.
36. van Middendorp, J.J., Goss, B., Urquhart, S., Atresh, S., Williams, R.P., and Schuetz, M. (2011). Diagnosis and prognosis of traumatic spinal cord injury. *Global Spine J.* 1, 1–8.
37. Magu, S., Singh, D., Yadav, R.K., and Bala, M. (2015). Evaluation of traumatic spine by magnetic resonance imaging and correlation with neurological recovery. *Asian Spine J.* 9, 748–756.
38. Roberts, T.T., Leonard, G.R., and Cepela, D.J. (2017). Classifications in brief: American Spinal Injury Association (ASIA) Impairment Scale. *Clin. Orthop. Relat. Res.* 475, 1499–1504.
39. Torlay, L., Perrone-Bertolotti, M., Thomas, E., and Baciú, M. (2017). Machine learning-XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inform.* 4, 159–169.
40. Ogunleye, A.A., and Qing-Guo, W. (2019). XGBoost model for chronic kidney disease diagnosis. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2019.2911071 [Epub ahead of print].
41. McCoy, D., Dupont, S., Gros, C., Cohen-Adad, J., Huie, R., Ferguson, A., Duong-Fernandez, X., Thomas, L., Singh, V., and Narvid, J. (2019). Convolutional neural network-based automated segmentation of the spinal cord and contusion injury: deep learning biomarker correlates of motor impairment in acute spinal cord injury. *Am. J. Neuroradiol.* 40, 737–744.
42. Nusinovič, S., Tham, Y.C., Yan, M.Y.C., Ting, D.S.W., Li, J., Sabanayagam, C., Wong, T.Y., and Cheng, C.-Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *J. Clin. Epidemiol.* 122, 56–69.
43. Yokobori, S., Zhang, Z., Moghieb, A., Mondello, S., Gajavelli, S., Dietrich, W.D., Bramlett, H., Hayes, R.L., Wang, M., Wang, K.K., and Bullock, M.R. (2015). Acute diagnostic biomarkers for spinal cord injury: review of the literature and preliminary research report. *World Neurosurg.* 83, 867–878.
44. Inoue, T., Manley, G.T., Patel, N., and Whetstone, W.D. (2014). Medical and surgical management after spinal cord injury: vasopressor usage, early surgeries, and complications. *J. Neurotrauma* 31, 284–291.
45. Goulet, J., Richard-Denis, A., and Mac-Thiong, J.-M. (2020). The use of classification and regression tree analysis to identify the optimal surgical timing for improving neurological outcomes following motor-complete thoracolumbar traumatic spinal cord injury. *Spinal Cord*, 52, 682–688.
46. Wilson, J.R., Witiw, C.D., Badhiwala, J., Kwon, B.K., Fehlings, M.G., and Harrop, J.S. (2020). Early surgery for traumatic spinal cord injury: where are we now? *Global Spine J.* 10, 845–915.
47. Furlan, J.C., Noonan, V., Cadotte, D.W., and Fehlings, M.G. (2011). Timing of decompressive surgery of spinal cord after traumatic spinal cord injury: an evidence-based examination of pre-clinical and clinical studies. *J. Neurotrauma* 28, 1371–1399.
48. Jug, M., Kežar, N., Vesel, M., Al Mawed, S., Dobravec, M., Herman, S., and Bajrović, F.F. (2015). Neurological recovery after traumatic cervical spinal cord injury is superior if surgical decompression and instrumented fusion are performed within 8 hours versus 8 to 24 hours after injury: a single center experience. *J. Neurotrauma* 32, 1385–1392.
49. Liu, J.-M., Long, X.-H., Zhou, Y., Peng, H.-W., Liu, Z.-L., and Huang, S.-H. (2016). Is urgent decompression superior to delayed surgery for traumatic spinal cord injury? A meta-analysis. *World Neurosurg.* 87, 124–131.
50. Alexander, M.S., Anderson, K.D., Biering-Sorensen, F., Blight, A.R., Brannon, R., Bryce, T., Creasey, G., Catz, A., Curt, A., and Donovan, W. (2009). Outcome measures in spinal cord injury: recent assessments and recommendations for future directions. *Spinal Cord* 47, 582–591.

Cite this article as: Inoue, T, Ichikawa, D, Ueno, T, Cheong, M, Inoue, T, Whetstone, W.D, Endo, T, Nizuma, K, and Tominaga, T (2020) XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury. *Neurotrauma Reports* 1:1, 8–16, DOI:10.1089/neur.2020.0009.

Abbreviations Used

- AIS = ASIA Impairment Scale
- ASIA = American Spinal Injury Association
- AUC = area under the receiver operating characteristic curve
- BASIC = Brain and Spinal Injury Center
- CCI = Charlson Comorbidity Index
- MCC = maximum canal compromise
- MRI = magnetic resonance imaging
- MSCC = maximum spinal cord compression
- ROC curve = receiver operating characteristic curve
- SCI = spinal cord injury
- SIR = signal intensity ratio
- SLIC = Subaxial Injury and Classification
- WI = weighted imaging
- XGBoost = extreme gradient boosting

Publish in *Neurotrauma Reports*



- Immediate, unrestricted online access
- Rigorous peer review
- Compliance with open access mandates
- Authors retain copyright
- Highly indexed
- Targeted email marketing

liebertpub.com/neur

