



Review Article

Corresponding Author

Michael G. Fehlings

<https://orcid.org/0000-0002-5722-6364>

Department of Surgery, University of Toronto, Division of Neurosurgery and Spinal Program, Toronto Western Hospital, University Health Network, 399 Bathurst St., Suite 4W-449, Toronto, ON M5T 2S8, Canada

Tel: +1-416-603-5627

Fax: +1-416-603-5298

E-mail: michael.fehlings@uhn.ca

Received: November 18, 2019

Revised: December 9, 2019

Accepted: December 15, 2019



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2019 by the Korean Spinal Neurosurgery Society

Predictive Modeling of Outcomes After Traumatic and Nontraumatic Spinal Cord Injury Using Machine Learning: Review of Current Progress and Future Directions

Omar Khan¹, Jetan H. Badhiwala^{1,2}, Jamie R.F. Wilson^{1,2}, Fan Jiang^{1,2}, Allan R. Martin², Michael G. Fehlings^{1,2}

¹Department of Surgery, University of Toronto, Toronto, ON, Canada

²Spinal Program, Toronto Western Hospital, University Health Network, Toronto, ON, Canada

Machine learning represents a promising frontier in epidemiological research on spine surgery. It consists of a series of algorithms that determines relationships between data. Machine learning maintains numerous advantages over conventional regression techniques, such as a reduced requirement for a priori knowledge on predictors and better ability to manage large datasets. Current studies have made extensive strides in employing machine learning to a greater capacity in spinal cord injury (SCI). Analyses using machine learning algorithms have been done on both traumatic SCI and nontraumatic SCI, the latter of which typically represents degenerative spine disease resulting in spinal cord compression, such as degenerative cervical myelopathy. This article is a literature review of current studies published in traumatic and nontraumatic SCI that employ machine learning for the prediction of a host of outcomes. The studies described utilize machine learning in a variety of capacities, including imaging analysis and prediction in large epidemiological data sets. We discuss the performance of these machine learning-based clinical prognostic models relative to conventional statistical prediction models. Finally, we detail the future steps needed for machine learning to become a more common modality for statistical analysis in SCI.

Keywords: Machine learning, Spinal cord injury, Outcomes, Degenerative cervical myelopathy, Magnetic resonance imaging

INTRODUCTION

The prediction of outcomes in spinal cord injury (SCI) is a challenging task requiring robust statistical techniques and the development of powerful clinical prognostic models (CPMs).¹ CPMs are statistical rules relating a desired outcome to one or more predictor variables. Clinicians frequently employ CPMs to make treatment decisions, manage patient expectations, and predict the course of an illness. In both the traumatic and nontraumatic forms of SCI, the underlying pathophysiological changes result in secondary glial scarring and cystic cavity formation,

which impair the regeneration and healing of neurons.² Clinically, this manifests as significant and occasionally irreversible functional deterioration, which results in a considerable burden on patients, families, and society at large.³

Because of the substantial impact of SCI, accurate CPMs are immensely useful for both the clinician and patient as tools to navigate the challenging landscape and sequelae of SCI. The development of these CPMs is now increasingly being done with machine learning (ML), a series of computational algorithms that can determine relationships within datasets.⁴ Compared to traditional prognostic models, which employ some

variant of logistic regression, ML has numerous advantages. First, ML requires little a priori knowledge of important predictors, as it often automatically determines the important predictors based on the dataset.^{5,6} Second, ML is generally less restrictive than logistic regression about the number of predictors that can be used for a given dataset.⁷ This makes ML useful for large datasets (e.g., oncology and pharmacogenomics) where large numbers of predictors may be present and relationships between predictors may not be immediately obvious. Third, ML can find complex, nonlinear relationships within datasets which are less amenable to being developed and analysed via logistic regression.⁸

Given these advantages, ML is often found to be more accurate and powerful than logistic regression techniques on the same dataset.^{9,10} This relative benefit has caused ML to become increasingly used in predictive modeling for SCI, to the extent

that many research groups have recently begun to develop ML-based CPMs for SCI. This review discusses recent works in which ML-based prognostic models were created for the prediction of outcomes in SCI. It is organized as follows: Section 2 presents the methods used to arrive at the body of literature reviewed in this article, while section 3 presents an overview of the methodology behind ML. Sections 4 and 5 describe the current progress with ML in developing prediction models for traumatic and nontraumatic SCI, respectively. Section 6 describes potential future directions with the use of ML in SCI with a case illustration, while section 7 concludes the paper.

METHODS

This article is a narrative review with the goal of providing the reader with an overview of ML applied to SCI. The body of

Table 1. Summary of literature review of machine learning in outcome prediction after SCI

Study	Description
A ML approach for specification of spinal cord injuries using fractional anisotropy values obtained from diffusion tensor images. ²⁷	Developed KNN and SVM models to predict the presence of spinal cord injury in individual axial slices of the spinal cord collected from DTI, specifically the fractional anisotropy parameter.
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. ²⁹	Developed a convolutional neural network to perform segmentation of the spinal cord in tSCI. Segmentation helped authors conclude that contusion injury volume was significantly correlated with motor scores at admission and discharge.
Development of an unsupervised ML algorithm for the prognostication of walking ability in spinal cord injury patients. ³¹	Constructed unsupervised ML algorithm predicting independent ambulation ability post-SCI at discharge or at the 1-year follow-up. Compared ML algorithm to logistic regression model – no significant difference found in performance.
Use of multivariate linear regression and support vector regression to predict functional outcome after surgery for cervical spondylotic myelopathy.	Compared a support vector regression model with a multivariate logistic regression model in the prediction of functional outcome after surgery for DCM. Support vector regression model was found to be superior.
Using a ML approach to predict outcome after surgery for degenerative cervical myelopathy. ¹⁵	Formulated random forest predicting quality-of-life and functional outcomes after decompression surgery for DCM (AUC = 0.70).
ML for prediction of sustained opioid prescription after anterior cervical discectomy and fusion (ACDF).	Developed stochastic gradient boosting model (AUC = 0.81) to predict sustained opioid prescription after ACDF. Major predictors of lengthened opioid prescription included preoperative opioid prescription, antidepressant use, tobacco use, and Medicaid insurance status.
Prognosis of cervical myelopathy based on diffusion tensor imaging with artificial intelligence methods.	Utilized multiple supervised learning models (e.g., SVM) that used DTI features to predict the mJOA recovery rate at the 1-year postsurgery follow-up.
Development of ML algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation	Created an elastic-net penalized logistic regression model (AUC = 0.81) to predict sustained opioid prescription after lumbar disc herniation surgery. Major predictors of lengthened opioid prescription included instrumentation, preoperative opioid duration, and comorbid depression.
Development of ML algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders	Created a neural network (AUC = 0.82) to predict nonroutine (i.e., not home) discharge for patients undergoing surgery for lumbar degenerative disc disease based on age, comorbid status, etc.

SCI, spinal cord injury; KNN, k-nearest neighbor; SVM, support vector machine; DTI, diffusion tensor imaging; tSCI, traumatic SCI; ML, machine learning; DCM, degenerative cervical myelopathy; AUC, area under the curve; mJOA, modified Japanese Orthopaedic Association.

literature for this review was collected by searching the PubMed database using the keywords ‘machine learning’ and ‘spinal cord injury’ or ‘machine learning’ and ‘cervical myelopathy’ or ‘machine learning’ and ‘lumbar spine.’ The combinations of these keywords produced 9 articles which used ML as a tool to predict either functional, health-related quality-of-life, or other outcomes (e.g., opioid use after surgery) pertinent to spine surgical practice. Three of these articles apply ML to the prediction of outcomes after traumatic SCI, while 6 of the articles use ML in the context of outcome prediction after nontraumatic SCI. Table 1 summarizes the articles found, including the ML algorithms used and the specific outcomes predicted.

OVERVIEW OF MACHINE LEARNING

As datasets continue to burgeon in size and complexity, the need for powerful software tools for data analysis has dramatically increased over the last few decades. At the same time, as computational capacity and technological sophistication continue to rise with time, the implementation of these tools has become increasingly feasible. One such software tool is ML, a series of complex mathematical algorithms frequently used for the development of mathematical models describing relationships between data.¹¹ Broadly, ML is divided into 3 categories: supervised ML, unsupervised ML,¹² and reinforcement learning.¹³

In supervised ML, the model is built using data which has its inputs and outputs labeled by the user.¹² That is, the user ‘supervises’ the algorithm by labelling the data beforehand, after which the ML algorithm creates the model relating the inputs and outputs. In unsupervised ML, the data is not given input and output labels.¹² Instead, the algorithm determines features within the data that allow it to group different data points. In reinforcement learning, the algorithm works within an environment to determine a policy that maximizes reward.¹³ For instance, a robot playing tennis works in an environment (i.e., tennis court) and is trained to determine a policy (i.e., make moves) that maximizes reward (i.e., number of points) while minimizing penalty (i.e., the opponent’s points). Reinforcement learning is not as amenable to epidemiological datasets in SCI as unsupervised or supervised learning. In fact, the majority of ML models developed in SCI are supervised learning models, which include classification and regression algorithms.¹⁴

Supervised ML algorithms are developed using a set of training data, on which the ML model is developed and optimized, and a separate set of testing data. Fig. 1 illustrates a schematic of

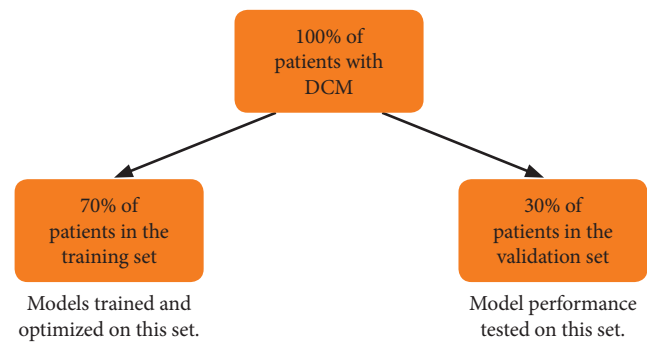


Fig. 1. Schematic of the train-test split performed in Merali et al.¹⁵ DCM, degenerative cervical myelopathy.

such a train-test split carried out by Merali et al.¹⁵ Maintaining a training-testing split is crucial for the successful design of a ML model, as it prevents overfitting and provides a good preliminary check of the external validity of the model.¹⁶ It is important to note that ML does not comprise a singular, monolithic entity; rather, there is a wide variety of algorithms that can be used to create ML models. These supervised algorithms include support vector machines (SVMs),¹⁷ classification trees,¹⁸ k-nearest neighbor (KNN),¹⁹ and naïve Bayes.²⁰

In addition, ML models can be further modified by using ensemble techniques, such as bagging and boosting.^{21,22} For example, multiple classification trees can be trained using different subsets of the training data. Then, the overall outcome of bagging the classification trees can be set equal to the majority outcome of the individual classification trees, resulting in a random forest model.²³ Stacking can also be employed by combining the outputs of entirely different ML models in a similar manner.²⁴ The variety and computational power provided by these techniques make ML the tool of choice when developing high-performance predictive models. Further, the nature of these models allows them to determine nonlinear relationships, which would otherwise be very difficult to incorporate in simple logistic regression. Table 2 compares machine learning to conventional logistic regression, while Table 3 compares the 3 modalities of machine learning with respect to multiple characteristics.

MACHINE LEARNING ALGORITHMS IN TRAUMATIC SPINAL CORD INJURY

Traumatic SCI (tSCI) has a significant impact on both patients and the healthcare system. In Western countries, it affects between 15 and 53 people per million every year.²⁵ What makes tSCI especially troublesome is that affected patients frequently

Table 2. Key characteristics of machine learning, organized by feature

Feature	Machine learning characteristics	Logistic regression characteristics
Knowledge of predictors	Little a priori knowledge of predictors needed	Requires knowledge of predictors for elimination of unimportant variables from model
No. of predictors	Fewer restrictions on number of predictors in machine learning	Number of predictors is restricted based on number of data points available
Nonlinear relationships	Capable of capturing complex, nonlinear relationships	Has difficulty with modelling nonlinear relationships
Algorithm variety	A host of machine learning algorithms exist, each with its own separate advantages and disadvantages. In addition, additional variations to enhance performance (e.g., bagging, boosting, stacking) may also be used in machine learning.	Multiple types of logistic regression models exist, but models generally have a similar foundation

Table 3. Attributes of the 3 major categories of machine learning

Attribute	Supervised learning	Unsupervised learning	Reinforcement learning
Labelling	Data outcomes are labeled beforehand	Data outcomes are not labeled	Data outcomes are not labeled
Description	Algorithm makes prediction of outcomes based on predictors using labeled data as a reference	Algorithm is used to separate data into clusters	Algorithm is used to build a policy that maximizes a cumulative reward
Evaluation Metrics	Algorithms are evaluated based on area under the curve and accuracy relative to the 'ground truth' (i.e. the true values of the outcomes)	Difficult to evaluate algorithm performance in the absence of 'ground truth' data	Algorithms are evaluated based on cumulative reward
Examples	Examples include classification algorithms (e.g., support vector machine) and regression algorithms (e.g., regression tree)	Examples include k-means clustering and principal component analysis	Examples include Q-learning

experience sequelae of persistent neurological dysfunction in multiple functional domains.^{3,26} These complications result in lifelong disability and a significant burden on the healthcare system. The devastating consequences of tSCI, as well as many unanswered issues concerning its management and prognosis, have created many questions for ML to answer. However, the current literature on the application of ML to prediction of outcomes in tSCI is fairly limited, possibly due to the relative novelty of ML and the tendency to pursue a more familiar regression modality instead.

Despite this paucity of literature, a number of important articles applying ML to tSCI have been published in multiple areas of SCI. For instance, Tay et al.²⁷ developed ML tools to predict the presence of SCI in individual axial slices of the spinal cord collected from diffusion tensor imaging (DTI). The authors used a specific DTI metric known as the fractional anisotropy,²⁸ a descriptor of the degree of directional water diffusion and tissue orientation in an image, which is used as a surrogate measure for the structural integrity of the imaged tissue. To perform their analysis, the authors first built tools to isolate the spinal cord in the image. Next, using the fractional anisotropy of different sections of the isolated spinal cord, the authors trained

and tested ML algorithms (specifically KNN and SVM with a radial kernel) to predict whether the spinal cord in a particular image slice was 'injured' or 'normal.' The ML algorithms performed well in predicting the presence of SCI from DTI images, yielding a specificity of 0.952 and a sensitivity of 0.912.

Like Tay et al.,²⁷ McCoy et al.²⁹ recently applied a ML approach to spinal cord imaging in tSCI. They developed a convolutional neural network (CNN) to perform segmentation of the spinal cord in tSCI. The model was trained using a training set of axial spinal magnetic resonance (MR) images annotated by board-certified radiologists, and it was then validated and tested using separate sets of axial MR images. The Dice coefficient,³⁰ a metric used to quantify the performance of an image segmentation algorithm, was 0.93 for the developed CNN, indicating excellent predictive performance above that of previous models. Based on this segmentation, the authors determined that the contusion injury volume (i.e., the volume of the spinal cord which was deemed injured by the CNN) was significantly correlated with motor scores at admission and discharge.

In addition to prediction of tSCI from imaging, the latest studies applying ML to tSCI have made important strides in the prognostication of functional outcomes after cord injury. DeVries et

al.³¹ recently used tSCI data from the RHSCIR (Rick Hansen Spinal Cord Injury Registry) database to construct an unsupervised ML algorithm predicting independent ambulation ability post-tSCI at discharge or at 1-year follow-up. This algorithm was then compared to previously established logistic regression models predicting the same outcome, and the authors found no significant difference between the performance of the logistic regression models and the unsupervised ML algorithm. While this equivalence is not immediately encouraging, DeVries et al.³¹ set a valuable foundation for future comparisons involving ML and established statistical models.

MACHINE LEARNING ALGORITHMS IN NONTRAUMATIC SPINAL CORD INJURY

Nontraumatic SCI is defined as a series of pathological changes resulting in damage to the spinal cord that do not arise directly from trauma. In the cervical region, nontraumatic SCI is often termed degenerative cervical myelopathy (DCM), the most common cause of spinal cord dysfunction worldwide.³² DCM results from age-related degenerative changes to the cervical spine that result in spinal cord compression, such as ossification of the posterior longitudinal ligament, hypertrophy of the ligamentum flavum, and osteophyte formation in the cervical vertebrae.³³ In the lumbar region, another area commonly affected by nontraumatic SCI, degenerative disc disease and lumbar stenosis can cause damage to the spinal cord.³⁴ Taken together, both DCM and nontraumatic lumbar SCI are caused by age-related degenerative changes, whose incidence is expected to rise with the growth in the elderly population.

As these population trends have begun, the need for ML has risen in order to meet the increasing requirement for powerful predictive tools guiding clinical management of degenerative cervical and lumbar cord compression. To this end, multiple groups have recently published articles on the use of ML in predictive modelling related to both DCM and nontraumatic lumbar SCI. One of the early articles on applying ML to DCM was published by Hoffman and colleagues,³⁵ who compared a support vector regression model (SVR - a regression variant of SVM) with a multivariate logistic regression model in the prediction of functional outcome (specifically, the Oswestry Disability Index or ODI) after surgery for DCM. The authors found that the SVR model outperformed its logistic regression counterpart. This was possibly due to the SVR model's relatively improved capability of capturing complex nonlinear relationships between the prognostic variables and the outcome. Additionally, the au-

thors identified a number of important variables affecting the final ODI score, such as preoperative ODI and symptom duration.

Merali et al.¹⁵ built upon this work and formulated ML models predicting quality-of-life and functional outcomes after decompression surgery for DCM. The authors trained and tested multiple classification models to predict improvement in SF-6D (a quality-of-life measure) and modified Japanese Orthopaedic Association (mJOA; a functional outcome measure) at 3 follow-up time points: 6-month postsurgery, 1-year postsurgery, and 2-year postsurgery. Overall, the random forest model was found to have the best performance, with an average area under the curve (AUC) of 0.70, a sensitivity of 78%, and an accuracy of 77% on the testing cohort. From a clinical relevance standpoint, the models developed identified many important features contributing to poor surgical outcomes in DCM. These included longer duration of myelopathy symptoms, worse preoperative disease severity, increased age, increased weight, and current smoking status. These findings have the potential to guide future surgical practice by potentially serving as a foundation on which to counsel patients.

Beyond prediction of functional outcomes, ML has also been used as a predictor of opioid use after surgery for DCM and lumbar disc disease. Anterior cervical discectomy and fusion (ACDF) is a surgical procedure for DCM that involves removal of a diseased intervertebral disc and fusion of 2 adjacent vertebral bodies.³⁶ Karhade et al.³⁷ developed 5 ML models to predict sustained opioid use after ACDF and found stochastic gradient boosting to be the highest-performing algorithm (AUC = 0.81 with good calibration). The authors determined that prolonged opioid use after ACDF was driven by preoperative opioid prescription, antidepressant use, tobacco use, and Medicaid insurance status. To apply the model created in the article, Karhade et al.³⁸ developed a webpage, making their ML algorithms more accessible to clinical practice. Karhade et al.³⁸ also extended this work to the lumbar spine domain, by developing ML algorithms predicting extended opioid use after surgery for lumbar disc herniation. Here, the authors found that an elastic-net penalized logistic regression model had the best performance (AUC = 0.81 with good calibration) and that instrumentation, preoperative opioid duration, and comorbid depression were the major predictors of prolonged opioid use.

From a health economics perspective, discharge disposition is one of the most important considerations for physicians working in an inpatient setting. Patients discharged to a setting other than home (i.e., nonroutine discharges) often experience a great-

er length-of-stay and are associated with a greater economic burden than those who are discharged directly home. In the context of lumbar spine disease, Karhade et al.³⁹ created a neural network predicting nonroutine discharge for patients undergoing surgery for lumbar degenerative disc disease. The neural network used a sample of over 26,000 patients from the National Surgical Quality Improvement Program database and extracted 8 key variables (e.g., age, body mass, comorbid disease status) to classify patients as either routine discharge or nonroutine discharge. The algorithm achieved an AUC of 0.82 with good calibration on the testing set, potentially guiding future practice by providing spine surgeons with tools to prepare in advance for nonroutine discharge and decrease hospital stays.

With respect to imaging interpretation, Jin et al.⁴⁰ have used DTI to determine prognosis based on the mJOA recovery ratio, defined as the ratio of postoperative improvement in mJOA to the ideal improvement in mJOA (i.e., recovery from preoperative baseline to maximum mJOA score). Patients were dichotomized at the 1-year follow-up according to whether their recovery ratio was 'good' or 'poor,' which facilitated the development of classification algorithms. In their analysis, the authors employed both supervised ML approaches (SVM, KNN) and multivariate logistic regression. These approaches used DTI features such as fractional anisotropy, axial diffusivity, radial diffusivity, and mean diffusivity (where diffusivity is defined as the rate of molecular diffusion in different directions) to make predictions about the recovery rate. The deep learning model showed poor accuracy in the prediction of recovery rate (59.2%); however, the SVM model showed excellent accuracy in its prediction (89.7%), exceeding the performance of the logistic regression model.

FUTURE DIRECTIONS AND RECOMMENDATIONS

With the increasing complexity of modern epidemiological data, ML is a prime candidate as the tool of choice for analysis. Because of its novelty and computational power in the area of SCI, ML possesses tremendous future potential and applicability in multiple clinically important domains (e.g., diagnosis, image processing, and prognosis). However, as discussed in sections 3 and 4, much of the work done so far in applying ML to SCI has been foundational and in a limited set of areas. To make ML a more mainstream tool in spine practice, we believe that 3 conditions need to be met. First, ML models need to be made more accessible to potential end-users in publications. Owing to their architecture, ML models can necessarily be more com-

plex and difficult to decipher on initial examination. Specifically, the models are not as intuitive and the relationships between outcome and predictors is often difficult to ascertain in a ML algorithm. In addition, they may involve more features than their logistic regression counterparts, making hand calculations of prognostic probabilities more difficult. To overcome this issue, applications such as Shiny can be used to create web-based tools that incorporate ML models. The development of these web-based tools (e.g., as seen in the study of Karhade et al.³⁷⁻³⁹) is a necessary step in allowing ML models to be more easily applicable in the clinical setting.

Second, ML models need to have a robust evidence base supporting them. While current studies frequently show superiority of ML over conventional techniques and generally employ a sound methodology, the mainstream appeal of ML may benefit from a consistent, evidence-based approach to developing algorithms. Conventional prediction models benefit from the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) checklist,⁴¹ which guides model development. For ML, guidelines to aid in ML algorithm creation were published in 2016 by Luo et al.⁷ These guidelines provide a step-by-step overview of the key tasks to be accomplished when creating a robust ML model for an epidemiological study. Future works employing ML models would benefit greatly from having a similar systematic approach, which would undoubtedly increase the appeal of ML especially as studies continue to demonstrate its superiority over conventional tools.

Finally, for ML to become a more mainstream tool in SCI, it needs to be used more widely to develop CPMs. Significant research needs to be performed to expand the scope of ML by applying it to other areas. For instance, ML may be used to predict functional outcomes after SCI in larger datasets, such as the European Multicenter SCI dataset (ClinicalTrials.gov Identifier: NCT01571531) and the North American Clinical Trials Network registry (ClinicalTrials.gov Identifier: NCT00178724). For cervical myelopathy, ML can be expanded to include imaging data as explanatory variables or to evaluate different outcome measures. In the basic science world, the close relationship of ML with so-called 'big data' can be leveraged to propose biomarkers of SCI or to analyse and develop advanced imaging techniques. The foundation currently established with ML and clinical imaging in SCI can also be further expanded, with ML eventually serving a crucial role in determining imaging characteristics that could be potential prognostic tools in SCI. The abundance of possible routes for ML makes these future areas of study ideal for future exploration.

CASE ILLUSTRATION

To better define the potential role of ML in SCI, we present a case of SCI in which ML can serve a substantial clinical role beneficial to both the spine clinician and the patient. Consider a 26-year-old male who presents to the Emergency Department with a SCI after a motor vehicle accident. After initial stabilization, the patient undergoes a complete neurological exam to assess the degree of his SCI. He is found to have decreased lower extremity neurological function (i.e., decreased motor scores and sensation) but near normal upper extremity neurological function. Based on a ML algorithm predicting independence in various activities of daily living, it is determined that the patient has a low probability of independently performing activities requiring the lower extremities (e.g., walking, climbing stairs, and bladder function) at 1 year. The patient is given early counseling to prepare him for this eventuality, and early intervention (e.g., urological studies, targeted physiotherapy) is started to maximize the patient's lower extremity function. With the early prediction from the ML algorithm, the patient's psychosocial status is improved thanks to early counseling, and the patient's lower extremity function is improved with the early intervention. The potential role for ML in improving lives with early prediction is immense, which is why it is a promising tool for spine specialists going forward.

CONCLUSIONS

Current clinical practice and everyday decision-making still rely on established CPMs based on logistic regression, since regression models are easier to utilize, have a greater body of research applying them, and are better understood compared to ML. Nevertheless, as ML begins to occupy a greater role in the research setting, and as studies begin to show the relative superiority of ML over conventional tools, we expect the reliance on the latter to decrease in favor of ML. Achieving these goals is no simple task; deeper, more impactful work needs to be done in order to build on the current foundation and expand the scope of ML in SCI. However, with the strides made by current literature in the realm of ML, we anticipate that ML-based prediction will soon become an essential tool in the armament of the spine physician.

CONFLICT OF INTEREST

The authors have nothing to disclose.

ACKNOWLEDGMENTS

MGF wishes to acknowledge the Gerry and Tootsie Halbert Chair in Neural Repair and Regeneration and the DeZwirek Family Foundation.

REFERENCES

1. Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009; 338:b604.
2. Ahuja CS, Nori S, Tetreault L, et al. Traumatic spinal cord injury-repair and regeneration. *Neurosurgery* 2017;80(3S): S9-22.
3. Pavese C, Schneider MP, Schubert M, et al. Prediction of bladder outcomes after traumatic spinal cord injury: a longitudinal cohort study. *PLoS Med* 2016;13:e1002041.
4. Karhade AV, Thio QCBS, Ogink PT, et al. Development of machine learning algorithms for prediction of 30-day mortality after surgery for spinal metastasis. *Neurosurgery* 2019; 85:E83-91.
5. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2007;2:59-77.
6. Panesar SS, D'Souza RN, Yeh FC, et al. Machine learning versus logistic regression methods for 2-year mortality prognostication in a small, heterogeneous glioma database. *World Neurosurg* X 2019;2:100012.
7. 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:e323.
8. Kuhle S, Maguire B, Zhang H, et al. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. *BMC Pregnancy Childbirth* 2018;18:333.
9. Oermann EK, Rubinsteyn A, Ding D, et al. Using a machine learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations. *Sci Rep* 2016;6: 21161.
10. Lee SI, Celik S, Logsdon BA, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat Commun* 2018;9:42.
11. Ghahramani Z. Probabilistic machine learning and artificial intelligence. *Nature* 2015;521:452-9.
12. Jordan MI, Mitchell TM. Machine learning: Trends, perspec-

- tives, and prospects. *Science* 2015;349:255-60.
13. Wiering M, van Otterlo M. Reinforcement learning: state-of-the-art. New York: Springer; 2012.
 14. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009.
 15. Merali ZG, Witiw CD, Badhiwala JH, et al. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PLoS One* 2019;14:e0215133.
 16. Koul A, Becchio C, Cavallo A. Cross-validation approaches for replicability in psychology. *Front Psychol* 2018;9:1117.
 17. Ben-Hur A, Weston J. A user's guide to support vector machines. *Methods Mol Biol* 2010;609:223-39.
 18. Kingsford C, Salzberg SL. What are decision trees? *Nat Biotechnol* 2008;26:1011-3.
 19. Mitchell TM. Machine learning. New York: McGraw-Hill; 1997.
 20. Shen Y, Li Y, Zheng HT, et al. Enhancing ontology-driven diagnostic reasoning with a symptom-dependency-aware Naïve Bayes classifier. *BMC Bioinformatics* 2019;20:330.
 21. Leo Breiman. Bagging Predictors. *Machine Learning* 1996; 24:123-40.
 22. Schapire RE. A brief introduction to boosting. In: *IJCAI-99 Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*; 1999 Jul 31-Aug 6; Stockholm, Sweden. San Francisco (CA); Morgan Kaufmann Publishers Inc.; 1999:1401-6.
 23. Breiman L. Random forests. *Mach Learn* 2001;45:5-32.
 24. Wang SQ, Yang J, Chou KC. Using stacked generalization to predict membrane protein types based on pseudo-amino acid composition. *J Theor Biol* 2006;242:941-6.
 25. Lee BB, Cripps RA, Fitzharris M, et al. The global map for traumatic spinal cord injury epidemiology: update 2011, global incidence rate. *Spinal Cord* 2014;52:110-6.
 26. van Middendorp JJ, Hosman AJ, Donders AR, et al. A clinical prediction rule for ambulation outcomes after traumatic spinal cord injury: a longitudinal cohort study. *Lancet* 2011; 377:1004-10.
 27. Tay B, Hyun JK, Oh S. A machine learning approach for specification of spinal cord injuries using fractional anisotropy values obtained from diffusion tensor images. *Comput Math Methods Med* 2014;2014:276589.
 28. Soares JM, Marques P, Alves V, et al. A hitchhiker's guide to diffusion tensor imaging. *Front Neurosci* 2013;7:31.
 29. McCoy DB, Dupont SM, Gros C, et al. 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. *AJNR Am J Neuroradiol* 2019;40:737-44.
 30. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol* 2004;11:178-89.
 31. DeVries Z, Hoda M, Rivers CS, et al. Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *Spine J* 2019 Sep 13 [Epub]. pii: S1529-9430(19)30971-4. <https://doi.org/10.1016/j.spinee.2019.09.007>.
 32. Nouri A, Tetreault L, Singh A, et al. Degenerative cervical myelopathy: epidemiology, genetics, and pathogenesis. *Spine (Phila Pa 1976)* 2015;40:E675-93.
 33. Davies BM, Mowforth OD, Smith EK, et al. Degenerative cervical myelopathy. *BMJ* 2018;360:k186.
 34. Ammendolia C. Degenerative lumbar spinal stenosis and its imposters: three case studies. *J Can Chiropr Assoc* 2014;58: 312-9.
 35. Hoffman H, Lee SI, Garst JH, et al. Use of multivariate linear regression and support vector regression to predict functional outcome after surgery for cervical spondylotic myelopathy. *J Clin Neurosci* 2015;22:1444-9.
 36. Maxey L, Magnusson J. Rehabilitation for the postsurgical orthopedic patient. 2nd ed. St. Louis (MO): Mosby; 2007.
 37. Karhade AV, Ogink PT, Thio QCBS, et al. Machine learning for prediction of sustained opioid prescription after anterior cervical discectomy and fusion. *Spine J* 2019;19:976-83.
 38. Karhade AV, Ogink PT, Thio QCBS, et al. Development of machine learning algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J* 2019;19:1764-71.
 39. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. *Neurosurg Focus* 2018;45:E6.
 40. Jin R, Luk KD, Cheung JPY, et al. Prognosis of cervical myelopathy based on diffusion tensor imaging with artificial intelligence methods. *NMR Biomed* 2019;32:e4114.
 41. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Ann Intern Med* 2015; 162:735-6.