Published in final edited form as:

World Neurosurg. 2019 July ; 127: e436–e442. doi:10.1016/j.wneu.2019.03.165.

Machine Learning For The Prediction Of Cervical Spondylotic Myelopathy: A Post Hoc Pilot Study Of 28 Participants.

Benjamin Hopkins, BS¹, Kenneth A. Weber II, DC, PhD², Kartik Kesavabhotla, MD¹, Monica Paliwal, PhD¹, Donald Cantrell, MD, PhD³, and Zachary A. Smith, MD¹

¹Northwestern University Feinberg School of Medicine, Department of Neurological Surgery

²Stanford University School of Medicine, Department of Radiology

³Northwestern University Feinberg School of Medicine, Department of Radiology

Introduction

Cervical Spondylotic Myelopathy (CSM) is a debilitating degenerative disease of the upper spine affecting primarily elderly patients.^{1–3} While compression of the spinal cord is thought to be the main driver of symptomology, pathophysiology is not entirely understood.^{1–3} It has been well characterized that imaging can be suggestive of severe compression in asymptomatic patients or on the contrary normal appearing in patients with debilitating disease.^{3,4} As such, imaging remains of limited benefit to clinicians without the appropriate associated clinical information.^{1,3,4} As such, CSM remains a clinical diagnosis with imaging playing a role of supporting or refuting evidence at most.^{1,3,4}

Similarly, as technology advances, MRI imaging of the cervical spine continues to provide promising progress towards diagnostic capabilities of such patients.^{4–7} Numerous studies have attempted to better categorize symptomatic compression through methods such as new MR techniques or sequences.^{4–7} While many continue to show promise, none have been overwhelmingly successful, further propelling the strong need for further innovation and new, more granular diagnostic techniques.^{8–13}

With the advent and recent explosion of machine learning, statistical predictive capabilities are growing faster than ever before.^{14–17} Relatively new to the medical imaging space, machine learning leverages innovation in computing power with ever improving modeling and statistical approaches.^{14–16} Many medically trained machine learning algorithms are further capable of discriminating between discrete patterns of data often overlooked by even the best human experts.¹⁵ While these techniques remain promising, much of the field of artificial intelligence remains focused primarily on diagnostics and automation, with only a small minority focusing on development of tools capable directly predicting clinical

Corresponding Author: Zachary A. Smith, M.D., Associate Professor Neurological Surgery, Northwestern University School of Medicine, 676 N. St. Clair Street, Chicago, IL 60611, zsmith1@nm.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

outcomes.¹⁵ As such, the purpose of this pilot study is to explore the potential use of machine learning algorithms in predicting CSM and correlated clinical scores based on imaging characteristics alone.

Methods

13 CSM and 15 controls underwent imaging of the cervical spine. All CSM patients included were diagnosed at a single large academic institution by a board certified practicing neurosurgeon based upon a combination of both clinical and radiographic findings. (Table 1) Inclusion criteria for entry included the following in all patients diagnosed with CSM: classic CSM symptoms, including exam findings of weakness, hyperreflexia, or change in coordination; radiographic signs of spinal compression; Nurick grade I-IV¹⁸; and modified Japanese Orthopedic Association (mJOA) scores of <18¹⁹. Exclusion criteria included the following: age <21 or >80, comorbid neural disease (e.g., multiple sclerosis), pregnant or nursing, active systemic rheumatological disease, active peripheral or vascular neuropathy, urgent need for surgery. The study was conducted with the approval of the university's Institutional Review Board (IRB).

Image Acquisition and Analysis

All imaging data were collected with a 3.0 Tesla Siemens Prisma magnetic resonance scanner (Siemens, Erlangen, Germany) equipped with a 64-channel head/neck coil. Participants were placed supine on the scanner bed, and a localizer scan was obtained to identify the location of the intervertebral discs of the cervical spine (C2-3, C3-4, C4-5, C5-6, C6-7, and C7-T1). Six high-resolution transverse slices with high white matter to gray matter contrast were acquired within the plane of each cervical intervertebral disc using a multi-echo gradient-echo sequence (TR=300 ms, TE=18 ms, Flip angle=30°, FOV=180×180, Matrix size=384×384, In-plane resolution=0.47×0.7 mm², Slice thickness=4 mm, number of averages=2).

Model 1- Predicting CSM diagnosis

Images were reviewed in a post-hoc fashion and the degree of cord compression was graded using 3 common literature scales (Kang, Nagata and Chang) alongside 3 MRI measurements (sagittal canal width, vertebral body height to vertebral disk height ratio and the C5 vertebral body sagittal width) all at the point of greatest compression on MRI. These six features were used to train a deep neural network (DNN) classification model (Figure 1) using the Keras open source Python package. The model was trained and tested using cross-validation, in which the data were randomly partitioned into training (n=18) and testing (n=10) datasets. In training, the 18 training images were fed through a series of 7 layers each with varying degrees of forward and backward communicating nodes (neurons). Dropout layers were introduced sporadically, preventing a certain percentage of neurons from communicating forward or backwards at different time points during the training to prevent overfitting and keep the model generalizable. The model was then trained and tested across a total of 200 random partitions using a batch size of 4 and 25 iterations for training the model. Mean and median cross-validated accuracies, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were used to assess the model performance.

Model 2- Predicting CSM Severity

Images for each transverse slice were further analyzed using the Spinal Cord Toolbox (version 3.0.7) and the PAM50 spinal cord template as described previously.^{20,21} Regions of the spinal cord were segmented and volumetric and cross-sectional measurements were collected for each region of interest. Regions of interest included anterior/posterior diameter, eccentricity of the spinal cord, ventral corticospinal tract, ventral reticulospinal tract, medial reticulospinal tract, lateral corticospinal tract, rubrospinal tract, lateral reticulospinal tract, ventrolateral reticulospinal tract and medial longitudinal fasciculus (each measured volumetrically using voxels and metrically using millimeters). The following gender, age, height, weight, level and the above noted parameters were input into our deep neural network, comprising a total of 23 input variables with the only output variable being mJOA score. Model 2 specifications are outlined in Figure 2. The model was trained with data partitioned into two datasets: training (n=78) and testing (n=26). Similarly to above, the 78 training data points were input into 9 layers, each with varying degrees of nodes (neurons). Sporadic dropout layers were added just as in Model 1 to prevent overfitting. Upon completion, the model was further trained and tested using a batch size of 3 and 1,250 iterations. This process was, similarly, repeated a total of 150 times in order to better characterize population data. Outputs were defined as a numeric prediction of mJOA score. Model performance was evaluated based upon mean squared error and subsequent average error in predictions were calculated. Error was defined as the total difference between predicted mJOA scale value and actual mJOA scale value.

Results

Model 1- Predicting CSM

The mean cross-validated accuracy of the trained model was 86.50% (95% CI 85.16%-87.83%) with a median accuracy of 90.00%. A distribution of accuracies across the 200 partitions is shown in Figure 3. Out of 200 partitions, the machine learning model was able to predict CSM (versus controls) with 100% accuracy 32 times (16.0%), 90% accuracy 91 times (45.5%), and 80% accuracy 59 times (29.5%). The program failed to predict CSM at an accuracy of 80% or greater in only 18 times (9.0%). Area under the curve (AUC) was calculated for each repetition. Average AUC for each repetition was 0.947 with a median AUC of 1.0. Average sensitivity, specificity, PPV and NPV were 90.25%, 85.05%, 81.58% and 91.94% respectively. A representative receiver operator curve from one iteration is shown in Figure 4.

Model 2- Modeling mJOA score

The mJOA score predictive model tended to slightly underpredict mJOA scores, with a mean error of -0.29 mJOA points and a median of -0.08 mJOA points. Average mean squared error (MSE) was 0.47 ± 1.13 . The standard deviation of mean error per batch was 0.714 mJOA points. A representative curve of mean errors per batch is represented in Figure 5. Average error of within 1 mJOA point of the actual mJOA score occurred roughly 81.0% of the time. Similarly, an average error of .5 mJOA points or less was noted approximately 48.5% of the time.

Discussion

Machine learning within medical imaging has grown immensely over the past 5-10 years. With the invention of Graphical Processing Units (GPUs), increased computing power has propelled technology to the possibility of raw unprocessed images into computer algorithms, each with the ability to produce designated unaided, unaltered, outputs without supervision. Numerous real-life applications have been documented in the literature with applications ranging as wide as from pulmonary embolism segmentation on emergency CT, to cognitive state of impairment in Alzheimer disease using fMRI.²²⁻²⁸

Literature on machine learning techniques specific to spine remains scarce, with the majority of papers focusing on identifying regions of interest in the lumbar spine or streamlining radiographic analysis.²⁹⁻³³ The current study is among the first to attempt to bridge the gap between strictly imaging diagnoses and clinical symptoms in the cervical spine. To date, few papers exist using machine learning techniques to look at clinical outcomes. Kim et. al attempted to predict clinical outcomes from pre-operative data points and complications after single level lumbar posterior fusion, noting that deep neural networks performed better than alternative statistical methods with AUCs ranging from 0.606-0.710 from 20,000 patients.³⁴ Similarly, Durand et. al notes the use machine learning approaches to predict the need for blood transfusion after lumbar spinal surgery to be promising reporting an AUC of 0.79.³⁵ The current paper produced a mean classification accuracy of 88% and an AUC of 0.947 with very few subjects available for training. Similarly, our paper further was able to predict mJOA score to within an average of 0.04 points based purely on computer automated spinal cord volumes.

The current paper improves upon literature suggesting imaging alone may eventually prove adequate in diagnosis and prognosis of patients with CSM. Currently, much of the literature involving CSM and imaging studies has been done on the use of DTI imaging and MT imaging.³⁶⁻³⁹ While each show elements of promise individually, no studies have demonstrated overwhelmingly conclusive results. As such, current investigational methods remain research based only, as currently they add little value to the physician in practice. Our paper demonstrates a much needed first step at the implementation of machine learning as a tool with potential to augment decision making of clinicians.

The goal of machine learning in medicine should not be misinterpreted however. With the recent boom in technological advancement, media and news too often tend to exaggerate as to the true abilities and impact machine learning can have on medicine.⁴⁰ Primarily, the authors believe machine learning should not be intended to replace physician intuition and judgement. The authors rather believe machine learning has the ability to be used clinically simply as yet one more data point in the already complex decision making process; in the case of CSM these decisions may be as straightforward as on whom and when to operate. Such scoring decision tree methods are already in place in medical decision making in numerous specialties, noteworthy examples being HAS-BLED⁴¹ or CHA₂DS₂-VASc⁴² scores among many others.

Our study is not without limitations. First, due to the prospective nature of our data we were limited in sample size and as such were perhaps not able to realistically approximate true predictive power. Similarly, the study uses close to equal numbers of CSM and control subjects in development and implementation of the model. This ratio is likely much smaller in the real world, with the population of CSM patients compared to controls being much lower than was in the current study. Our study is further limited by the lack of consistent industry standards for optimization of machine learning algorithms.⁴³ Currently, gold standards for optimization of DNN models involve simple guess and check methods, with comparative partition analysis commonly used as the main decision factor in determining the optimum model.⁴³ In doing so, single variables are altered manually and arbitrarily and compared to the current best known model.⁴³ These changes are either accepted or declined depending on the relative performance of the model over many simulations.⁴³ As such, due to lack of standardized method, bias in model construction is a definite confounder in our above study as no current theory exists on best practices in varying model parameters. Lastly, as our study was a controlled prospective study, only patients with clearly diagnosed, classic CSM were enrolled. While ideal for a pilot study, the lack of patients on the fringes of diagnosis may introduce artificially inflated predictive statistics. In real life, patients are more heterogeneous, often presenting without an obvious diagnosis. These types of patients were not included in our model unfortunately due to the nature of study design. Despite these limitations, the authors believe the above findings to be important and useful for the future of improving outcomes amongst patients with CSM.

Potential next steps in exploring machine learning related to CSM imaging include larger, more complex model development. Currently, methods are available to feed entire cervical MR images into computer models without human preprocessing.⁴⁴ While requiring a large amount of computing power, such methods would eliminate the problematic human error or bias introduced by much of current diagnostic methods. Similarly, by feeding entire images into the model without alteration, it gives training the needed freedom to find new predictors not otherwise previously noted by human investigators.

Conclusions

Machine learning provides a promising method for prediction and diagnosis for patients with CSM. In this pilot study, after reviewing features from only 18 images, our classification model was able to predict CSM from controls with a median accuracy of 90% as well as predict mJOA scores within 0.4 points using imaging characteristics alone. While still only preliminary, the current study demonstrates promise and feasibility for the use of machine learning to better improve diagnostic and predictive methods for CSM as well as other cervical spine disorders.

Acknowledgments

** Funding contributing to this manuscript has been received by the National Institute of Drug Abuse [Grant Number T32DA035165], the National Institute on Neurological Disorders and Stroke [Grant Number K23NS104211], and the Neurosurgical Research Education Fund Summer Student Research Fellowship. The authors otherwise have nothing to disclose and no further conflicts of interest. The below manuscript has been submitted and was rejected from *Spine*. The below manuscript has not been submitted, published or presented elsewhere.

Abbreviations-

CSM	Cervical Spondylotic Myelopathy
mJOA	modified Japanese Orthopedic Association
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
AUC	Area Under Curve
IRB	Institutional Review Board
DNN	Deep Neural Network
PPV	Positive Predictive Value
NPV	Negative Predictive Value
MSE	Mean Squared Error
GPU	Graphical Processing Unit

References

1. Ames CP, Blondel B, Scheer JK, et al. Cervical radiographical alignment: comprehensive assessment techniques and potential importance in cervical myelopathy. *Spine (Phila Pa 1976)*. 2013;38(22 Suppl 1):S149–160. [PubMed: 24113358]
2. Ichihara K, Taguchi T, Sakuramoto I, Kawano S, Kawai S. Mechanism of the spinal cord injury and the cervical spondylotic myelopathy: new approach based on the mechanical features of the spinal cord white and gray matter. *J Neurosurg*. 2003;99(3 Suppl):278–285. [PubMed: 14563145]
3. Klineberg E Cervical spondylotic myelopathy: a review of the evidence. *Orthop Clin North Am*. 2010;41(2):193–202. [PubMed: 20399358]
4. Hopkins BS, Weber KA 2nd, Cloney MB, Paliwal M, Parrish TB, Smith ZA. Tract-Specific Volume Loss on 3T MRI in Patients With Cervical Spondylotic Myelopathy. *Spine (Phila Pa 1976)*. 2018;43(20):E1204–E1209. [PubMed: 29649085]
5. Cloney M, Smith AC, Coffey T, et al. Fatty infiltration of the cervical multifidus musculature and their clinical correlates in spondylotic myelopathy. *J Clin Neurosci*. 2018.
6. Cui JL, Li X, Chan TY, Mak KC, Luk KD, Hu Y. Quantitative assessment of column-specific degeneration in cervical spondylotic myelopathy based on diffusion tensor tractography. *Eur Spine J*. 2015;24(1):41–47. [PubMed: 25150714]
7. Gao SJ, Yuan X, Jiang XY, et al. Correlation study of 3T-MR-DTI measurements and clinical symptoms of cervical spondylotic myelopathy. *Eur J Radiol*. 2013;82(11):1940–1945. [PubMed: 23932097]
8. You JY, Lee JW, Lee E, Lee GY, Yeom JS, Kang HS. MR Classification System Based on Axial Images for Cervical Compressive Myelopathy. *Radiology*. 2015;276(2):553–561. [PubMed: 25906184]
9. Uda T, Takami T, Tsuyuguchi N, et al. Assessment of cervical spondylotic myelopathy using diffusion tensor magnetic resonance imaging parameter at 3.0 tesla. *Spine (Phila Pa 1976)*. 2013;38(5):407–414. [PubMed: 22914703]
10. Chang V, Ellingson BM, Salamon N, Holly LT. The Risk of Acute Spinal Cord Injury After Minor Trauma in Patients With Preexisting Cervical Stenosis. *Neurosurgery*. 2015;77(4):561–565; discussion 565. [PubMed: 26191975]

11. Kang Y, Lee JW, Koh YH, et al. New MRI grading system for the cervical canal stenosis. *AJR Am J Roentgenol*. 2011;197(1):W134–140. [PubMed: 21700974]
12. Nagata K, Kiyonaga K, Ohashi T, Sagara M, Miyazaki S, Inoue A. Clinical value of magnetic resonance imaging for cervical myelopathy. *Spine (Phila Pa 1976)*. 1990;15(11):1088–1096. [PubMed: 2267602]
13. Pavlov H, Torg JS, Robie B, Jahre C. Cervical spinal stenosis: determination with vertebral body ratio method. *Radiology*. 1987;164(3):771–775. [PubMed: 3615879]
14. Giger ML. Machine Learning in Medical Imaging. *J Am Coll Radiol*. 2018;15(3 Pt B):512–520. [PubMed: 29398494]
15. Erickson BJ. Machine Learning: Discovering the Future of Medical Imaging. *J Digit Imaging*. 2017;30(4):391. [PubMed: 28653122]
16. Wong KKL, Wang L, Wang D. Recent developments in machine learning for medical imaging applications. *Comput Med Imaging Graph*. 2017;57:1–3. [PubMed: 28456276]
17. Marr B How Is AI Used In Healthcare - 5 Powerful Real-World Examples That Show The Latest Advances. In: *Forbes* 2018.
18. Nurick S The pathogenesis of the spinal cord disorder associated with cervical spondylosis. *Brain*. 1972;95(1):87–100. [PubMed: 5023093]
19. Benzel EC, Lancon J, Kesterson L, Hadden T. Cervical laminectomy and dentate ligament section for cervical spondylotic myelopathy. *J Spinal Disord*. 1991;4(3):286–295. [PubMed: 1802159]
20. De Leener B, Levy S, Dupont SM, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage*. 2017;145(Pt A):24–43. [PubMed: 27720818]
21. De Leener B, Fonov VS, Collins DL, Callot V, Stikov N, Cohen-Adad J. PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. *Neuroimage*. 2017;165:170–179. [PubMed: 29061527]
22. Schoepf UJ, Schneider AC, Das M, Wood SA, Cheema JI, Costello P. Pulmonary embolism: computer-aided detection at multidetector row spiral computed tomography. *J Thorac Imaging*. 2007;22(4):319–323. [PubMed: 18043385]
23. Dundar MM, Fung G, Krishnapuram B, Rao RB. Multiple-instance learning algorithms for computer-aided detection. *IEEE Trans Biomed Eng*. 2008;55(3):1015–1021. [PubMed: 18334393]
24. Summers RM. Improving the accuracy of CTC interpretation: computer-aided detection. *Gastrointest Endosc Clin N Am*. 2010;20(2):245–257. [PubMed: 20451814]
25. Wei J, Hadjiiski LM, Sahiner B, et al. Computer-aided detection systems for breast masses: comparison of performances on full-field digital mammograms and digitized screen-film mammograms. *Acad Radiol*. 2007;14(6):659–669. [PubMed: 17502255]
26. Bauer C, Adam R, Stoltz DA, Beichel RR. Computer-aided analysis of airway trees in micro-CT scans of ex vivo porcine lung tissue. *Comput Med Imaging Graph*. 2012;36(8):601–609. [PubMed: 22959430]
27. Mitchell TM, Shinkareva SV, Carlson A, et al. Predicting human brain activity associated with the meanings of nouns. *Science*. 2008;320(5880):1191–1195. [PubMed: 18511683]
28. Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging*. 2008;29(4):514–523. [PubMed: 17174012]
29. Wang SQ, Li X, Cui JL, Li HX, Luk KD, Hu Y. Prediction of myelopathic level in cervical spondylotic myelopathy using diffusion tensor imaging. *J Magn Reson Imaging*. 2015;41(6):1682–1688. [PubMed: 25044870]
30. Zhu X, He X, Wang P, et al. A method of localization and segmentation of intervertebral discs in spine MRI based on Gabor filter bank. *Biomed Eng Online*. 2016;15:32. [PubMed: 27000749]
31. Oktay AB, Akgul YS. Localization of the lumbar discs using machine learning and exact probabilistic inference. *Med Image Comput Comput Assist Interv*. 2011;14(Pt 3):158–165. [PubMed: 22003695]
32. Oktay AB, Albayrak NB, Akgul YS. Computer aided diagnosis of degenerative intervertebral disc diseases from lumbar MR images. *Comput Med Imaging Graph*. 2014;38(7):613–619. [PubMed: 24972858]

33. Koh J, Chaudhary V, Dhillon G. Disc herniation diagnosis in MRI using a CAD framework and a two-level classifier. *Int J Comput Assist Radiol Surg.* 2012;7(6):861–869. [PubMed: 22392057]
34. Kim JS, Merrill RK, Arvind V, et al. Examining the Ability of Artificial Neural Networks Machine Learning Models to Accurately Predict Complications Following Posterior Lumbar Spine Fusion. *Spine (Phila Pa 1976).* 2018;43(12):853–860. [PubMed: 29016439]
35. Durand WM, DePasse JM, Daniels AH. Predictive Modeling for Blood Transfusion After Adult Spinal Deformity Surgery: A Tree-Based Machine Learning Approach. *Spine (Phila Pa 1976).* 2018;43(15):1058–1066. [PubMed: 29215501]
36. Guan X, Fan G, Wu X, et al. Diffusion tensor imaging studies of cervical spondylotic myelopathy: a systemic review and meta-analysis. *PLoS One.* 2015;10(2):e0117707. [PubMed: 25671624]
37. Ellingson BM, Salamon N, Holly LT. Advances in MR imaging for cervical spondylotic myelopathy. *Eur Spine J.* 2015;24 Suppl 2:197–208. [PubMed: 23917647]
38. Suleiman LI, Weber KA 2nd, Rosenthal BD, et al. High-resolution magnetization transfer MRI in patients with cervical spondylotic myelopathy. *J Clin Neurosci.* 2018;51:57–61. [PubMed: 29530383]
39. Cloney MB, Smith ZA, Weber KA 2nd, Parrish TB. Quantitative Magnetization Transfer MRI Measurements of the Anterior Spinal Cord Region are Associated With Clinical Outcomes in Cervical Spondylotic Myelopathy. *Spine (Phila Pa 1976).* 2018;43(10):675–680. [PubMed: 29068880]
40. A realistic look at the hype of machine learning and AI. In: News HI, ed2018.
41. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest.* 2010;138(5):1093–1100. [PubMed: 20299623]
42. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest.* 2010;137(2):263–272. [PubMed: 19762550]
43. Ben-Nun TH T. Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis. 2018.
44. Niftynet.io. 2018 <http://www.niftynet.io/>.

Figure 1-

Algorithm specifications for Deep Neural Network (DNN) used in Model 1. All variables underwent processing through a 7 layer artificial neural network with above specified neurons in each layer. Images were analyzed in groups of 4 at a time until all 18 were completed, at which point the algorithm would adjust predictive variable weights and repeat for 25 epochs/repeats. Upon completion, the final algorithm would be used to verify accuracy on 10 unseen images from the original dataset. Images were redistributed randomly into test/training groups, model weights were reinitialized and randomized and the entire process was repeated 200 times for population data collection.

Figure 2-

Algorithm specifications for Deep Neural Network (DNN) used in Model 2. All variables underwent processing through a 9 layer artificial neural network with above specified neurons in each layer. Images were analyzed in groups of 3 at a time until all 78 were completed, at which point the algorithm would adjust predictive variable weights and repeat for 1250 epochs/repeats. Upon completion, the final algorithm would be used to verify error on 26 unseen data points from the original dataset. Images were redistributed randomly into test/training groups, model weights were reinitialized and randomized and the entire process was repeated 150 times for population data collection.

Figure 3-
Distribution of cross validation accuracies of CSM predictions on 10 previously unseen patients after first training and model adjustment on random sub-sets of 18 patients.

Figure 4-

Representative receiver operator curve for Model 1. Area under the curve for this particular repetition was 0.880. Average AUC for all models was 0.947 with a median AUC being 1.0.

Figure 5-

Outcome Distribution of Model 2. (A) Distribution of mean error per batch in units of points of mJOA score. (B) Distribution of median error per batch in units of points of mJOA score.

Table 1.

Baseline Demographic and Clinical Characteristics

	Controls (Mean $\pm$ SD)	CSM (Mean $\pm$ SD)
Age	50.31 $\pm$ 11.56	59.38 $\pm$ 11.84
Gender (% MALE)	54%	62%
Height (in)	68.58 $\pm$ 3.37	67.08 $\pm$ 4.23
Weight (lbs)	171.69 $\pm$ 23.35	190.54 $\pm$ 41.45
Nurick	0 $\pm$ 0	1.85 $\pm$ 0.9
mJOA	18 $\pm$ 0	14.31 $\pm$ 2.14
Neck NRS	0.15 $\pm$ 0.55	5 $\pm$ 2.45
Arm NRS	0.23 $\pm$ 0.44	4.46 $\pm$ 3.26
NDI	1.23 $\pm$ 2.35	19 $\pm$ 9.5
Pain SF 6a	42.52 $\pm$ 5.1	61.73 $\pm$ 7.69
SF-36 PCS	32.61 $\pm$ 28.81	18.17 $\pm$ 23.82
SF-36 MCS	63.16 $\pm$ 6.29	60.23 $\pm$ 13.11

Abbreviations: mJOA- modified Japanese Orthopedic Association; NRS- Numeric Rating Scale; NDI- Neck Disability Index; Pain SF 6a- Pain Short Form Health Survey 6a; SF-36 PCS- Short Form (36) Health Survey Physical Component Summary; SF-36 MCS- Short Form (36) Health Survey Mental Component Summary.