JNS SPINE

CLINICAL ARTICLE

J Neurosurg Spine 41:639–647, 2024

Spinal cord metrics derived from diffusion MRI: improvement in prognostication in cervical spondylotic myelopathy compared with conventional MRI

Presented at the 2024 AANS/CNS Joint Section on Disorders of the Spine and Peripheral Nerves

Justin K. Zhang, MD, MSCI,[1,2] Salim Yakdan, MD, MSCI,[2] Muhammad I. Kaleem, MBBS,[2] Saad Javeed, MD, MSCI,[2] Jacob K. Greenberg, MD, MSCI,[2] Kathleen S. Botterbush, BS,[2] Braeden Benedict, MS,[2] Martin Reis, MD,[3] Natasha Hongsermeier-Graves, MD, MPH,[1] Spencer Twitchell, MD,[1] Brandon Sherrod, MD,[1] Marcus S. Mazur, MD,[1] Mark A. Mahan, MD,[1] Andrew T. Dailey, MD,[1] Erica F. Bisson, MD, MPH,[1] Sheng-Kwei Song, PhD,[3] and Wilson Z. Ray, MD, MBA[2]

[1]Department of Neurosurgery, Clinical Neurosciences Center, University of Utah, Salt Lake City, Utah; and Departments of [2]Neurosurgery and [3]Radiology, Washington University School of Medicine, St. Louis, Missouri

OBJECTIVE A major shortcoming in optimizing care for patients with cervical spondylotic myelopathy (CSM) is the lack of robust quantitative imaging tools offered by conventional MRI. Advanced MRI modalities, such as diffusion MRI (dMRI), including diffusion tensor imaging (DTI) and diffusion basis spectrum imaging (DBSI), may help address this limitation by providing granular evaluations of spinal cord microstructure.

METHODS Forty-seven patients with CSM underwent comprehensive clinical assessments and dMRI, followed by DTI and DBSI modeling. Conventional MRI metrics included 10 total qualitative and quantitative assessments of spinal cord compression in both the sagittal and axial planes. The dMRI metrics included 12 unique measures including anisotropic tensors, reflecting axonal diffusion, and isotropic tensors, describing extraaxonal diffusion. The primary outcome was the modified Japanese Orthopaedic Association (mJOA) score measured at 2 years postoperatively. Extreme gradient boosting–supervised classification algorithms were used to classify patients into disease groups and to prognosticate surgical outcomes at 2-year follow-up.

RESULTS Forty-seven patients with CSM, including 24 (51%) with a mild mJOA score, 12 (26%) with a moderate mJOA score, and 11 (23%) with a severe mJOA score, as well as 21 control subjects were included. In the classification task, the traditional MRI metrics correctly assigned patients to healthy control versus mild CSM versus moderate/severe CSM cohorts, with an accuracy of 0.647 (95% CI 0.64–0.65). In comparison, the DTI model performed with an accuracy of 0.52 (95% CI 0.51–0.52) and the DBSI model's accuracy was 0.81 (95% CI 0.808–0.814). In the prognostication task, the traditional MRI metrics correctly predicted patients with CSM who improved at 2-year follow-up on the basis of change in mJOA, with an accuracy of 0.58 (95% CI 0.57–0.58). In comparison, the DTI model performed with an accuracy of 0.62 (95% CI 0.61–0.62) and the DBSI model had an accuracy of 0.72 (95% CI 0.718–0.73).

CONCLUSIONS Conventional MRI is a powerful tool to assess structural abnormality in CSM but is inherently limited in its ability to characterize spinal cord tissue injury. The results of this study demonstrate that advanced imaging techniques, namely DBSI-derived metrics from dMRI, provide granular assessments of spinal cord microstructure that can offer better diagnostic and prognostic utility.

https://thejns.org/doi/abs/10.3171/2024.4.SPINE24107

KEYWORDS cervical spondylotic myelopathy; magnetic resonance imaging; diffusion MRI; machine learning

ABBREVIATIONS ADC = apparent diffusion coefficient; AUC = area under the curve; CSM = cervical spondylotic myelopathy; DBSI = diffusion basis spectrum imaging; dMRI = diffusion MRI; DTI = diffusion tensor imaging; FA = fractional anisotropy; FOV = field of view; ICC = intraclass correlation coefficient; LMC = level of maximum cord compression; LOOCV = leave-one-out cross-validation; MCID = minimum clinically important difference; MEDIC = multiecho data image combination; mJOA = modified Japanese Orthopaedic Association; RFE = recursive feature elimination; XGBoost = extreme gradient boosting.

SUBMITTED March 20, 2024. **ACCEPTED** April 23, 2024.
INCLUDE WHEN CITING Published online July 26, 2024; DOI: 10.3171/2024.4.SPINE24107.

Cervical spondylotic myelopathy (CSM) is the leading cause of progressive disability in patients older than 65 years.[1] A major shortcoming in optimizing care for patients with CSM is the lack of quantifiable imaging metrics to guide surgical decision-making. Conventional imaging techniques such as MRI provide valuable data on structural abnormalities in CSM but cannot quantitatively assess white matter tracts. Biomarkers derived from quantitative diffusion MRI (dMRI), such as diffusion tensor imaging (DTI), have shown promise in predicting disease severity and neurological outcomes,[2] but can be confounded by the complex pathophysiology of CNS disorders.[2] To address these shortcomings, we previously developed diffusion basis spectrum imaging (DBSI), a data-driven multiple-tensor model capable of resolving dMRI signals to accurately reflect both complex intraaxonal and confounding extraaxonal structures and pathologies.[3–8]

Despite the increasing popularity of advanced MRI modalities, no study has assessed their application in CSM in relation to traditional MRI metrics. As such, we compared conventional MRI measures with those attained from DTI and DBSI in both characterizing CSM pathophysiology and prognosticating long-term outcomes. We hypothesized that traditional MRI measures would be useful to characterize disease pathology but lack detailed evaluations of white matter tract integrity that can help inform long-term outcomes. The results of this study may help identify clinically meaningful imaging biomarkers to deliver more targeted therapeutics and provide more personalized surgical decision-making when treating patients with CSM.

Methods

Study Description

This is a prospective cohort study comprising 50 patients with CSM who were enrolled after presenting to the neurosurgery outpatient clinic between 2018 and 2020. Patients 18–75 years of age with a history of ongoing spinal cord compression and symptomatic CSM with clinical signs of myelopathy were included. Patients were excluded if they were pregnant or unable to tolerate MRI sequencing or if they had concomitant thoracic and/or lumbar stenosis, rheumatoid arthritis, demyelinating diseases, spine tumor, or human immunodeficiency virus–related myelopathy. Age-matched and sex-matched healthy control subjects were also enrolled in the same time period. This study adhered to STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines,[9] and institutional review board approval and informed consent were obtained.

Clinical Measures

At initial evaluation, comprehensive demographic and clinical characteristics were assessed. All patients with CSM underwent decompressive cervical surgery with or without fusion, with the approach (i.e., anterior vs posterior) based on the surgeon's discretion. All patients were monitored for 2 years after surgery. The primary outcome was the modified Japanese Orthopaedic Association (mJOA) score, with patients categorized as having mild (mJOA 15–17), moderate (12–14), or severe (0–11) CSM.[1] A conservative minimum clinically important difference (MCID) of 2 points was used to determine improvement at 2 years after surgery. Patients with moderate and severe CSM were combined into one group for analyses because of small sample size.

Imaging Data

All patients underwent preoperative imaging evaluation, which included both traditional MRI sequencing and specific dMRI sequences. The following images were obtained. 1) Sagittal T2-weighted images—a 3D turbo spin echo sequence was used with the following parameters: acquisition time 4 minutes 33 seconds; TR 1500 msec; TE 15 msec; in-plane resolution 0.8 × 0.8 mm^2; 64 × 0.8–mm–thick slices; field of view (FOV) 256 × 256 cm^2; and data matrix 320 × 320. 2) Axial T2*-weighted images from C3 to C6—a multiecho data image combination (MEDIC) sequence was used with the following parameters: acquisition time 4 minutes 43 seconds; TR 766 msec; TE 13 msec; in-plane resolution 0.5 × 0.5 mm^2; four 7.5-mm-thick slices; FOV 160 × 160 cm^2; data matrix 320 × 320; 3 combined echoes; and 2 excitations. 3) dMRI data—a vendor-supplied ZOOMIt sequence in axial view was used with the same slice planning as MEDIC: acquisition time approximately 2 minutes per scan; TR 620 msec; TE 70 msec; in-plane resolution 0.75 × 0.75 mm^2; four 7.5-mm-thick slices covering C3 to C6; cardiac gated; FOV 76 × 38 cm^2; and data matrix 102 × 51. dMRI data were obtained using a 26-direction diffusion weighting with a maximum b-value of 1000 sec/mm^2 (0–1000 at step of 40 distributed randomly assigned to each diffusion-weighting direction).

Traditional MRI

All imaging measurements were made using routine imaging programs built into electronic health record systems, and were performed under the guidance of a board-certified neuroradiologist (Table 1, Fig. 1). Metrics included qualitative and quantitative assessments in both the sagittal and axial planes. All metrics were measured in reference to the level of maximum cord compression (LMC). In the sagittal plane, qualitative measures included the presence of T2 hyperintensity within the cord, the LMC of the cervical spinal cord, the grade of spinal canal stenosis,[10] and the degree of spinal cord compression.[11] Quantitative measures included the total number of compressed levels in the cervical spinal cord (defined as at least a grade of ≥ 1 for spinal stenosis), maximum spinal cord compression (given by lines C/[{B+D}/2] in Fig. 1),[12] and sagittal compression ratio (given by lines C/A in Fig. 1).[13] In the axial plane, qualitative assessments included the type of spinal cord compression (i.e., partial or circumferential),[14] the axial compression ratio (given by lines C/E in Fig. 1),[15] and the transverse area of the cord at the LMC.

Diffusion MRI

The apparent disconnect between MRI findings and clinical manifestations may stem from the inability of

TABLE 1. Spinal cord measurements on traditional MRI

Metric	Plane	Type	Comment
T2 hyperintensity	Sagittal	Qualitative	Presence of T2 signal w/in spinal cord
LMC	Sagittal	Qualitative	Most compressed cervical spinal cord level
Degree of spinal cord compression	Sagittal	Qualitative	Qualitative assessment of obliteration of subarachnoid space
Grade of spinal cord compression	Sagittal	Qualitative	Qualitative assessment of obliteration of subarachnoid space, taking into account T2 signal change
Levels of spinal cord compression	Sagittal	Quantitative	Total no. of levels of spinal cord compression*
Compression ratio—sagittal	Sagittal	Quantitative	Ratio of midsagittal diam of spinal cord at the LMC divided by avg diam of spinal cord at the closest noncompressed regions above & below the LMC
Max spinal cord compression	Sagittal	Quantitative	Ratio of spinal cord diam at the LMC to diam at the C1 level
Type of spinal cord compression	Axial	Qualitative	Partial vs circumferential compression at the LMC
Compression ratio—axial	Axial	Qualitative	Ratio of sagittal diam divided by transverse diam of spinal cord at the LMC
Transverse area	Axial	Quantitative	Cross-sectional transverse area at the LMC

Avg = average; diam = diameter; max = maximum.
* Defined as at least a grade of ≥ 1 for spinal stenosis.

conventional MRI to quantify microstructural spinal cord injuries present in CSM.[2] To address this limitation, dMRI has been increasingly used. In dMRI the thermal motion of water molecules is used as the dominant contrast mechanism to study the microstructure of biological tissues.[2] In this study, the above-mentioned dMRI parameters were used to generate dMRI data, which then underwent post-processing with DTI and DBSI modeling.

DTI. In conventional DTI modeling, the dMRI signals within an image voxel are averaged to model as a single tensor. DTI-derived tensors included fractional anisotropy (FA), which reflects overall white matter tract integrity,[16] and apparent diffusion coefficient (ADC), which measures water motion without reference to any one direction (e.g., vasogenic edema). DTI modeling also generates axial and radial diffusivity, which quantifies axonal injury and demyelination, respectively (Table 2).

DBSI. As opposed to DTI modeling, DBSI models the diffusion properties of water as a linear combination of anisotropic tensors (reflecting axonal diffusion) and iso-

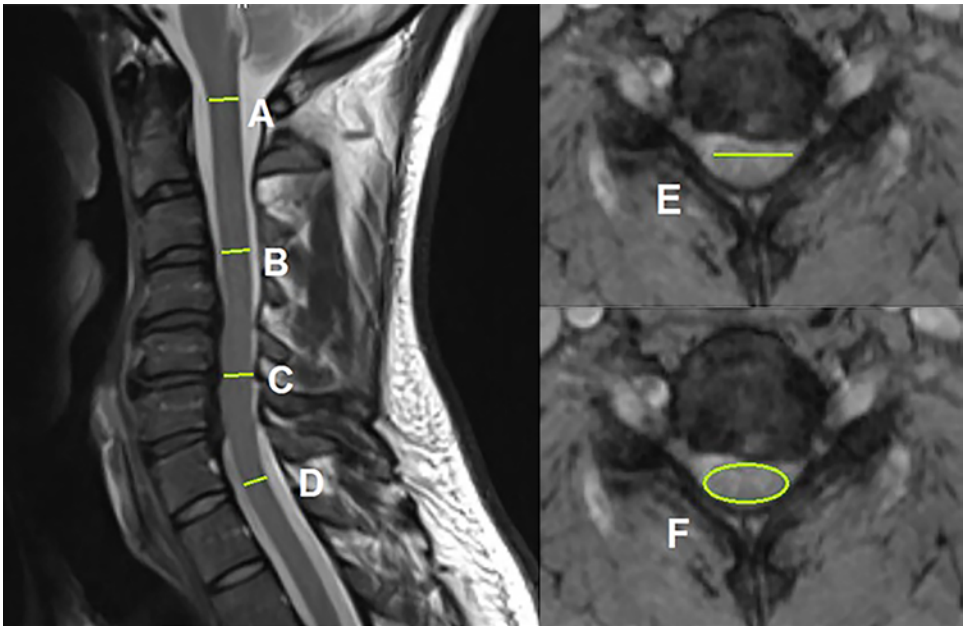

FIG. 1. Measurements from sagittal (**left**) and transverse (**right**) views of T2-weighted MRI of the cervical spinal cord. These measurements include lines marked *A*: diameter of cord at the level of C1 vertebra; *B*: diameter of cord at the first level superior to the LMC that is normal; *C*: diameter of cord at the LMC; *D*: diameter of cord at the first level inferior to the LMC that is normal; *E*: transverse diameter of cord at the LMC; and *F*: transverse area of cord at the LMC (*circle*).

TABLE 2. Spinal cord measurements on dMRI

dMRI Modality	Clinical Correlate
DTI, single tensor	
ADC	Edema
FA	White matter tract integrity
Axial diffusivity	Axonal integrity
Radial diffusivity	Myelin integrity
DBSI, multiple tensors	
FA	White matter tract integrity
Fiber fraction	Axonal density
Axial diffusivity	Axonal integrity
Radial diffusivity	Myelin integrity
Restricted fraction	Cellularity
Nonrestricted fraction	Vasogenic edema
Intraaxonal axial diffusivity	Axonal integrity
Extraaxonal fiber fraction	Vasogenic edema

tropic tensors (describing extraaxonal diffusion).[17] DBSI-derived anisotropic tensors also include FA^2 and axial and radial diffusivity.[18] DBSI modeling also generates fiber fraction, a measure of axonal density.[18] DBSI intraaxonal axial diffusivity minimizes water signals from extraaxonal compartments, and therefore is expected to possess greater sensitivity for axonal injury (Table 2).

DBSI-derived isotropic tensors rely on ADC, a measure of the magnitude of free diffusion of water.[17] Unlike DTI modeling, however, ADC cutoffs are used to generate more granular isotropic tensors. Small ADC values reflect water diffusion within cells, whereas large ADC values signal the presence of extracellular water.[7] These isotropic tensors include restricted fraction (ADC ≤ 0.3 μm^2/msec), which reflects levels of cellularity,[19] and nonrestricted fraction (ADC > 0.3 μm^2/msec), which measures edematous water.[20] DBSI extraaxonal fraction measures water signals from extraaxonal compartments and provides a robust assessment of vasogenic edema. By measuring both anisotropic and isotropic diffusion tensors, DBSI can describe coinciding pathologies, such as axonal injury (anisotropic tensors) and edema (isotropic tensors). Refer to Supplementary Methods for detailed information on DTI and DBSI modeling of dMRI signals.

Derivation of MRI Metrics

dMRI of the spinal cord was processed using a Python-implemented pipeline developed in our laboratory to process all data. Image data were converted from native (Digital Imaging and Communications in Medicine) to the Neuroimaging Informatics Technology Initiative format for all scans. Using a 2D registration procedure with slicewise translations in the axial plane, the image data were preprocessed to correct for motion artifacts due to subject motion or physiological noise. Each dMRI slice was then nonlinearly registered using Advanced Neuroimaging Tools to the corresponding image slice within the PAM50 template (available for T1-, T2-, and T2*-weighting contrasts). Hand-drawn regions of interest were applied to the warped maps to extract diffusion metrics in white matter tracts.

Statistical Analyses

Our analysis flow sheet is summarized in Fig. 2.

Univariate Analysis

Imaging metrics were compared across patient groups (i.e., healthy control, mild CSM, and moderate/severe CSM) by using independent samples t-tests, ANOVA, Fisher exact tests, and chi-square tests. Pearson correlations or ANOVA were used to explore the relationship between imaging metrics and mJOA scores (i.e., preoperative, postoperative, and change). A correlation matrix was built comparing all imaging metrics with each other by using Pearson correlations. Given the exploratory nature of our analyses, multiple comparisons testing was not performed. Probability values < 0.05 were considered statistically significant. The intraclass correlation coefficient (ICC) was calculated for a subset of 25% of the MR images to assess the interrater reliability.

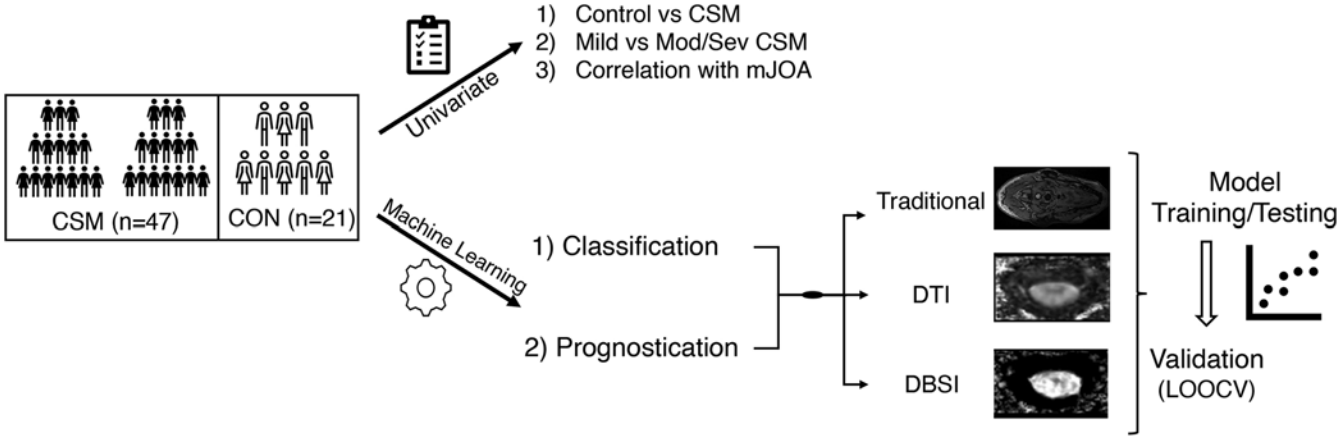

FIG. 2. A schematic representation illustrating the analysis strategy used in the study. CON = control; Mod/Sev = moderate/severe.

Model Building

A supervised classification machine learning model that uses the extreme gradient boosting (XGBoost) algorithm[21] was chosen because it is well known for its high predictive accuracy and computational efficiency. Gradient boosting is an ensemble algorithm that accumulates weak models such as decision trees to the ensemble to minimize previous model errors.[22,23]

Three sets of imaging predictors based on MRI were examined: traditional, DTI, and DBSI features (Fig. 2). First, in the preoperative setting, the XGBoost model was used to assess how accurately MRI measures could correctly classify patients into healthy control, mild CSM, and moderate/severe CSM groups. Next, at 2-year follow-up, including only patients with CSM, the model was used to predict which patients experienced a clinically meaningful improvement after surgery, defined by the MCID of the mJOA. For each objective (i.e., classification and prognostication), 3 models were built using traditional MRI metrics only, DTI metrics only, or DBSI metrics only, for a total of 6 models (Fig. 2). Performance measures included accuracy, area under the curve (AUC), precision, and F1 score. Nonoverlapping 95% confidence intervals between models for a given performance metric (e.g., accuracy) were considered statistically significantly different.[24]

Numerical variables were centered and scaled, and categorical variables were preprocessed using one-hot encoding.[25] Hyperparameter tuning of the XGBoost models was performed to identify the combination of parameters of the XGBoost model that yielded the highest accuracy.[26] This was performed using grid search, which involves defining a grid of hyperparameters to be tested and using 5-fold cross-validation for model evaluation. Recursive feature elimination (RFE), a feature selection algorithm, was used to identify variables with the greatest predictive power. Based on the RFE, select variables were eliminated and the remaining predictors were used to train the final model.[27,28]

Model Validation

Model validation was carried out using leave-one-out cross-validation (LOOCV) to internally validate feature sets. LOOCV is more suitable and provides less biased performance estimates for small sample size.[29] Different models were evaluated using accuracy, AUC, precision, and F1 score. The 95% confidence intervals were calculated by generating 1000 bootstrapped samples, and statistically significant differences between feature sets were determined by nonoverlapping confidence intervals. All analyses were conducted using Python version 3.12.1 and R version 4.3.2.

Results

Among the 50 patients with CSM who were initially included, 3 were excluded for clinical reasons. One patient underwent preoperative evaluation but ultimately did not undergo surgery, another was found to have severe lumbar stenosis at a later visit, and a third was found to have multiple comorbidities (which violated our inclusion criteria). Twenty-one age- and sex-matched healthy control subjects from the same period were included. Complete demographic, clinical, and surgical characteristics are described in Table 3. Of the 47 included patients with CSM, 24 (51%) had mild mJOA scores and 23 (49%) had a moderate or severe mJOA classification. The mean age of patients with CSM and control subjects was 57 ± 9 and 57 ± 8 years ($p > 0.05$), respectively, and the mean mJOA scores were 14 ± 3 and 17 ± 1 ($p < 0.001$), respectively.

TABLE 3. Demographic, clinical, and surgical characteristics of study participants

Characteristic	Control Pts, n = 21	Pts w/ CSM, n = 47
Age in yrs	57 ± 8	57 ± 9
BMI, kg/m^2	28 ± 7	29 ± 6
Sex, M:F	11:10	28:19
Tobacco use	8 (38%)	26 (55%)
ECI		
<0	NA	0
0	NA	12 (26%)
1–4	NA	32 (68%)
5–13	NA	3 (6%)
mJOA score	17 ± 1	14 ± 3
Mild	NA	24 (51%)
Moderate	NA	12 (26%)
Severe	NA	11 (23%)
Median Sx duration, mos	NA	13.5, range 7–28
Surgery type		
Anterior	NA	29 (62%)
Posterior	NA	17 (36%)
Combined	NA	2 (4%)
Multilevel surgery	NA	38 (81%)
Postop mJOA score	NA	15 ± 3
Change in mJOA	NA	0.9 ± 2.4
% improved*	NA	14 (30%)

BMI = body mass index; ECI = Elixhauser Comorbidity Index; NA = not applicable; Pts = patients; Sx = symptom.
Unless otherwise indicated, values are expressed as the number of patients (%) or the mean ± SD.
* Based on an MCID of 2 points.

The ICC revealed excellent agreement in measuring imaging parameters (average ICC 0.77). All MRI metrics (traditional, DTI, DBSI) were significantly different when comparing control subjects and patients with CSM ($p < 0.05$), except for the LMC, DBSI FA, and DBSI radial diffusivity ($p > 0.05$) (Supplementary Table 1). However, when comparing imaging measures between the patients with mild and moderate/severe CSM, for all traditional MRI metrics, except for the transverse area, there were no statistically significant differences (Supplementary Table 2, $p = 0.045$). Most dMRI metrics also were not statistically significantly different between mJOA cohorts, except for DBSI intraaxonal axial diffusivity and DBSI extraaxonal fiber fraction ($p < 0.001$). When correlating MRI metrics with preoperative, postoperative, and change in mJOA, there were few statistically significant associations (Sup-

TABLE 4. Model performance measures of XGBoost classification algorithms on 2-year mJOA outcomes

Model	MRI Modality	Accuracy [95% CI]	Precision [95% CI]	F1 Score [95% CI]	AUC × 100 [95% CI]
Classification	Traditional	0.647 [0.64–0.65]	0.655 [0.65–0.66]	0.646 [0.64–0.65]	0.8 [0.797–0.802]
	DTI	0.515 [0.51–0.52]	0.523 [0.52–0.53]	0.514 [0.51–0.52]	0.687 [0.68–0.69]
	DBSI	0.81 [0.808–0.814]	0.82 [0.817–0.823]	0.81 [0.808–0.814]	0.904 [0.902–0.906]
Prognostication	Traditional	0.575 [0.57–0.58]	0.543 [0.54–0.55]	0.552 [0.55–0.56]	0.418 [0.41–0.42]
	DTI	0.618 [0.61–0.62]	0.592 [0.59–0.6]	0.596 [0.59–0.6]	0.52 [0.514–0.524]
	DBSI	0.72 [0.718–0.73]	0.71 [0.7–0.714]	0.7 [0.696–0.706]	0.645 [0.639–0.651]

Accuracy is calculated as the number of correct predictions (TP+TN)/total number of predictions (TP+TN+FP+FN), where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. Precision is calculated as the correct positive predictions (TP)/all positive predictions (TP+FP). Recall is calculated as the TP divided by all actual positive predictions (TP+FN). F1 score is a combination of precision and recall used to compare the performance between models.

plementary Table 3). Importantly, DBSI intraaxonal axial diffusivity and DBSI extraaxonal fiber fraction were significantly associated with preoperative, postoperative, and change in mJOA. When correlating traditional MRI measures with dMRI (i.e., DTI and DBSI) metrics, 30/56 (54%) relationships were significant (Supplementary Table 4).

MRI metrics were then incorporated into XGBoost classification algorithms. Specifically, for each objective (i.e., classification and prognostication), 3 models were built using 1) traditional MRI metrics only, 2) DTI metrics only, or 3) DBSI metrics only (Fig. 2). In both the classification and prognostication tasks, the XGBoost models built using DBSI metrics performed the best (Table 4).

In the classification task (Fig. 3A), the traditional MRI metrics correctly assigned patients to healthy control versus mild CSM versus moderate/severe CSM cohorts with an accuracy of 0.647 (95% CI 0.64–0.65) and an AUC of 0.8 (95% CI 0.797–0.802). In comparison, the DTI model performed with an accuracy of 0.515 (95% CI 0.51–0.52) and an AUC of 0.687 (95% CI 0.68–0.69), and the DBSI model with an accuracy of 0.81 (95% CI 0.808–0.814) and an AUC of 0.904 (95% CI 0.902–0.906). In the prognostication task (Fig. 3B), the traditional MRI metrics correctly predicted patients with CSM who improved at 2-year follow-up with an accuracy of 0.575 (95% CI 0.57–0.58) and an AUC of 0.418 (95% CI 0.41–0.42). In comparison, the DTI model performed with an accuracy of 0.618 (95% CI 0.61–0.62) and an AUC of 0.52 (95% CI 0.514–0.524), and the DBSI model with an accuracy of 0.72 (95% CI 0.718–0.73) and an AUC of 0.645 (95% CI 0.639–0.651) (Table 4).

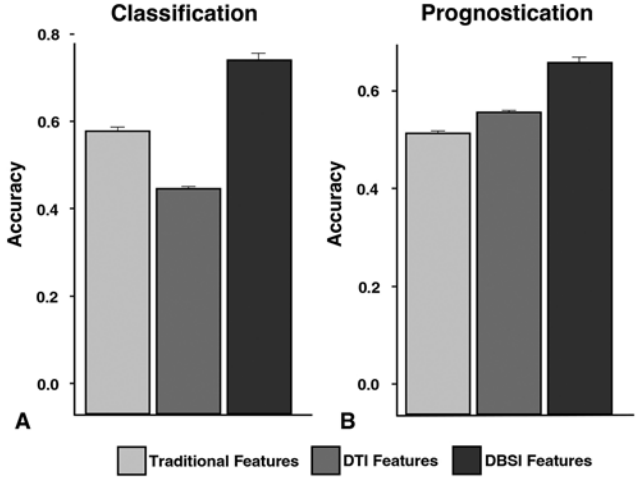

FIG. 3. XGBoost model accuracies for each outcome measure stratified by radiological feature sets (traditional, DTI, and DBSI).

Discussion

In recent decades, there has been a growing interest in applying advanced imaging modalities in CSM. In the present study, we first investigated the performance of individual metrics derived from 3 MRI modalities—traditional, DTI, and DBSI—to differentiate control subjects from those with CSM, as well as patients with mild CSM from those with moderate/severe CSM (Fig. 2). We then investigated the performance of combined imaging measures in classifying patient group preoperatively (i.e., control vs mild CSM vs moderate/severe CSM) and prognosticating long-term outcomes postoperatively (i.e., predicting change in mJOA at 2 years). Overall, our results suggest that in our cohort, metrics of spinal cord health derived from advanced dMRI are generally superior to traditionally derived measures in their application in CSM on both univariate analyses and machine learning approaches.

Generally speaking, metrics derived from all imaging modalities were capable of separating control subjects and patients with CSM. However, it is important to note that a nonnegligible percentage of healthy patients exhibited imaging evidence of spinal cord compression, supporting the well-known notion that there is a significant percentage of the population with asymptomatic spinal cord compression.[30] It is also important to note that there was not a significant difference in DBSI FA between control subjects and those with CSM. However, this was an expected finding, because DBSI FA is a fiber-specific metric, meaning that it reflects the integrity of residual axons.[4] This is in opposition to the previously described DTI-derived FA,[16] which averages fibers with surrounding structure properties in a voxel. Comparable DBSI FA between control subjects and patients with CSM suggests that residual, intact fibers in patients with CSM were not different from those of control subjects.

When comparing imaging metrics between the mild and moderate/severe CSM cohorts, there were few significant differences between groups (Supplementary Table 2). Corresponding with this finding, few imaging measures were significantly associated with mJOA in the preoperative or postoperative period (Supplementary Table 3). These results are not entirely unexpected, however, because individual MRI metrics likely lack the power to adequately parse out nuances in CSM pathology. Each measure assesses a specific facet of radiographic pathology, and our group discovered in further analyses that it is necessary to use the collective, symbiotic relationship among imaging measures to provide the most predictive power. Indeed, previous groups have tried to identify certain imaging biomarkers that could be used in the evaluation of CSM, but existing evidence remains limited.[31,32]

Despite the limited significant findings of imaging metrics on univariate analyses, transverse area and DBSI markers of vasogenic edema (intraaxonal axial diffusivity and extraaxonal fraction) were significantly different in all univariate analyses. The cross-sectional transverse area of the spinal cord at the LMC has been frequently reported in the literature as a useful measure to assess CSM disease burden. The spinal cord is reported to lose function when the transverse area is < 55%–75%, and previous groups have defined a critical threshold of transverse area in patients with CSM.[33] Although measuring the transverse area of the spinal cord is informative in characterizing baseline disease severity, available evidence from two high-quality studies from the prospective, multicenter AOSpine North America group demonstrated that it was not significantly associated with recovery.[31,32]

DBSI intraaxonal diffusivity and DBSI extraaxonal fraction are both markers of vasogenic edema.[7] Although these metrics are scarcely mentioned in the literature because of their novelty in recent years, their significance is intuitive when taking the pathophysiology of CSM into account. In CSM, chronic spinal cord compression causes local ischemia, leading to endothelial cell dysfunction and compromise of the blood–spinal cord barrier.[34] This in turn promotes increased cellularity (e.g., microglial infiltration) as well as release of inflammatory cytokines, promoting vasogenic edema. These results suggest that vasogenic edema may play a significant role in CSM pathophysiology and may be a target for future therapeutics.

Surprisingly, we did not find significant differences between groups with other traditional MRI measures such as T1-weighted hypointensity and T2-weighted hyperintensities. These imaging findings are thought to be evidence for demyelinization, gliosis, edema, or myelomalacia,[34] and have been associated with greater clinical impairment.[35–37] In fact, certain older studies have suggested that T2-weighted signal hyperintensity is necessary to diagnose CSM.[38] However, more recent large studies have found that these imaging findings are present in approximately 60% of all patients with CSM and in less than half of patients with mild CSM.[39]

Although a correlation exists between the sagittal diameter of the spinal canal and CSM severity, spinal cord compression is not uncommon in the asymptomatic population. Indeed, it is well known that spinal cord function can be preserved even under severe compression. As such, it was not entirely surprising that measures assessing compression in the sagittal plane (Table 1) were not significantly associated with worse disease. The degree and the grade of spinal canal stenosis are the most common mechanisms that radiologists use to characterize spinal cord compression in the clinical setting. Although they are very useful to provide a general sense of compression severity, these measures are not only subjective but also lack the power behind quantitative measurements.

Surprisingly, DTI-derived measures were globally not significant. One reason behind this finding may be the underlying postprocessing details behind DTI modeling. In contrast to DBSI modeling, DTI assumes a single diffusion tensor averaging the diffusion profile of multiple microstructural compartments within an image voxel.[40] Intuitively, this may cause issues, because an image voxel may contain multiple spinal microstructural elements, including axons, inflammatory cells, and the extracellular matrix. This effect is worsened in the context of significant vasogenic edema, as is expected in CSM. This, in turn, can confound DTI measurements intended to assess white matter tracts. In addition, DTI metrics can have a wide range of normal values, modest test–retest reliability, and limited sensitivity to pathology.[41]

Our initial results on univariate analyses demonstrated that there is no singular imaging metric derived from MRI that is particularly powerful, and that multiple measures assessing different aspects of spinal cord pathology are needed. As such, we leveraged XGBoost, a dynamic machine learning algorithm able to capture complex nonlinear relationships and variable interactions. In addition, we used RFE to identify top predictors and limit overfitting, combined with LOOCV to provide the most accurate performance. We specifically tested the ability of each imaging modality (e.g., traditional, DTI, DBSI) to 1) classify patients into control, mild CSM, or moderate/severe CSM groups and 2) prognosticate 2-year long-term outcomes in patients with CSM.

Overall, models built on DBSI metrics had good discrimination in correctly assigning patients into their corresponding groups, with an accuracy of 0.81 (95% CI 0.808–0.814) and AUC of 0.904 (95% CI 0.902–0.906) (Fig. 3A). Understandably, accuracies were worse when predicting 2-year outcomes, with an accuracy of 0.72 (95% CI 0.718–0.73) and AUC of 0.645 (95% CI 0.639–0.651) (Fig. 3B). These accuracies were statistically significantly better than those from the DTI and traditional models, as reflected by the nonoverlapping 95% confidence intervals. Worse accuracies on prognostication are expected, because we intentionally did not include any clinical covariates in the model, which are indispensable when predicting long-term outcome. The purpose of our analysis was to assess the individual performance of imaging data and directly compare with the performance of these different imaging modalities.

Our contention behind the improved accuracy of DBSI is that the intricate processing of its modeling allows for the generation of granular imaging modalities that can capture subtle facets of disease pathophysiology that cannot be appreciated on standard MRI. For example, T1 hy-

pointensities and T2 hyperintensities are nonspecific and do not measure specific pathophysiological processes that occur at the microstructural level. Imaging techniques that can characterize microstructural changes (i.e., dMRI) may have greater potential for prognostication because they can provide insight between reversible and irreversible changes in neurons. As reflected by Supplementary Table 4, indeed we found that only approximately 50% of DBSI metrics were significantly correlated with traditional measures, and most of these were weak correlations. This suggests that, although DBSI metrics are generally related to traditional MRI metrics, they are capturing a different facet of CSM pathophysiology.

Limitations

Our study has several limitations, as follows. Our study was constrained by a relatively small sample size at a single center with an institutional-specific imaging protocol that is still undergoing refinements.[42] Along similar lines, another limitation of our study is that we binned patients with moderate and severe CSM into one group to increase our statistical power. These limitations may compromise the generalizability of our findings. Future studies across multiple centers with larger cohorts and standardized imaging protocols are essential to validate our results and enhance statistical power. Furthermore, the internal validation of our models using LOOCV may have limitations, such as potential overfitting, particularly in small sample sizes. Finally, because of the exploratory nature of our analysis, we did not perform corrections for multiple comparisons, increasing the risk of type I errors.

Conclusions

In CSM, a disproportion exists between the degree of spinal cord compression and the clinical presentation. Conventional MRI is a powerful tool to assess structural abnormality but is inherently limited in its ability to characterize spinal cord tissue injury. Our results demonstrate that advanced imaging techniques, namely DBSI-derived metrics from dMRI, provide granular assessments of spinal cord microstructure than can offer better diagnostic and prognostic utility.

References

1. Badhiwala JH, Ahuja CS, Akbar MA, et al. Degenerative cervical myelopathy—update and future directions. *Nat Rev Neurol*. 2020;16(2):108-124.
2. Shabani S, Kaushal M, Budde MD, Wang MC, Kurpad SN. Diffusion tensor imaging in cervical spondylotic myelopathy: a review. *J Neurosurg Spine*. 2020;33(1):65-72.
3. Jayasekera D, Zhang JK, Blum J, et al. Analysis of combined clinical and diffusion basis spectrum imaging metrics to predict the outcome of chronic cervical spondylotic myelopathy following cervical decompression surgery. *J Neurosurg Spine*. 2022;37(4):588-598.
4. Zhang JK, Jayasekera D, Song C, et al. Diffusion basis spectrum imaging provides insights into cervical spondylotic myelopathy pathology. *Neurosurgery*. 2023;92(1):102-109.
5. Zhang JK, Javeed S, Greenberg JK, et al. Feasibility of postoperative diffusion-weighted imaging to assess representations of spinal cord microstructure in cervical spondylotic myelopathy. *Neurosurg Focus*. 2023;55(3):E7.
6. Zhang JK, Javeed S, Greenberg JK,Dibble CF, Song SK, Ray WZ. Diffusion basis spectrum imaging identifies clinically relevant disease phenotypes of cervical spondylotic myelopathy. *Clin Spine Surg*. 2023;36(3):134-142.
7. Zhang JK, Jayasekera D, Javeed S, et al. Diffusion basis spectrum imaging predicts long-term clinical outcomes following surgery in cervical spondylotic myelopathy. *Spine J*. 2023;23(4):504-512.
8. Zhang JK, Sun P, Jayasekera D, et al. Utility of diffusion basis spectrum imaging in quantifying baseline disease severity and prognosis of cervical spondylotic myelopathy. *Spine (Phila Pa 1976)*. 2022;47(24):1687-1693.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344-349.
10. 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-W140.
11. Morio Y, Yamamoto K, Kuranobu K, Murata M, Tuda K. Does increased signal intensity of the spinal cord on MR images due to cervical myelopathy predict prognosis? *Arch Orthop Trauma Surg*. 1994;113(5):254-259.
12. Fehlings MG, Rao SC, Tator CH, et al. 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 (Phila Pa 1976)*. 1999;24(6):605-613.
13. Oshima Y, Seichi A, Takeshita K, et al. Natural course and prognostic factors in patients with mild cervical spondylotic myelopathy with increased signal intensity on T2-weighted magnetic resonance imaging. *Spine (Phila Pa 1976)*. 2012;37(22):1909-1913.
14. Shimomura T, Sumi M, Nishida K, et al. Prognostic factors for deterioration of patients with cervical spondylotic myelopathy after nonsurgical treatment. *Spine (Phila Pa 1976)*. 2007;32(22):2474-2479.
15. Chen CJ, Lyu RK, Lee ST, Wong YC, Wang LJ. Intramedullary high signal intensity on T2-weighted MR images in cervical spondylotic myelopathy: prediction of prognosis with type of intensity. *Radiology*. 2001;221(3):789-794.
16. Rao A, Soliman H, Kaushal M, et al. Diffusion tensor imaging in a large longitudinal series of patients with cervical spondylotic myelopathy correlated with long-term functional outcome. *Neurosurgery*. 2018;83(4):753-760.
17. Cross AH, Song SK. "A new imaging modality to non-invasively assess multiple sclerosis pathology". *J Neuroimmunol*. 2017;304:81-85.
18. Song SK, Sun SW, Ju WK, Lin SJ, Cross AH, Neufeld AH. Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *Neuroimage*. 2003;20(3):1714-1722.
19. Wang Y, Wang Q, Haldar JP, et al. Quantification of increased cellularity during inflammatory demyelination. *Brain*. 2011;134(Pt 12):3590-3601.
20. Chiang CW, Wang Y, Sun P, et al. Quantifying white matter tract diffusion parameters in the presence of increased extra-fiber cellularity and vasogenic edema. *Neuroimage*. 2014;101:310-319.
21. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13–17, 2016; San Francisco, CA. Accessed May 10, 2024. https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf
22. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189-1232.
23. Yakdan S, Botterbush K, Xu Z, Lu C, Ray WZ, Greenberg

JK. Machine learning and lumbar spondylolisthesis. *Semin Spine Surg*. 2023;35(3):101048.
24. Cumming G. Inference by eye: reading the overlap of independent confidence intervals. *Stat Med*. 2009;28(2):205-220.
25. Kosaraju N, Sankepally SR, Mallikharjuna Rao K. Categorical data: need, encoding, selection of encoding method and its emergence in machine learning models—a practical review study on heart disease prediction dataset using Pearson correlation. In: Saraswat M, Chowdhury C, Mandal CK, Gandomi AH, eds. *Proceedings of International Conference on Data Science and Applications – ICDSA 2022*. Vol 1. Springer; 2023:369-382. Accessed May 10, 2024. https://doi.org/10.1007/978-981-19-6631-6_26
26. Ahsan MM, Mahmud MAP, Saha PK, et al. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies (Basel)*. 2021;9(3):52.
27. Kuhn M. Variable selection using the caret package. 2012. Accessed May 10, 2024. http://cran.cermin.lipi.go.id/web/packages/caret/vignettes/caretSelection.pdf
28. Granitto PM, Furlanello C, Biasioli F, Gasperi F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometr Intell Lab Syst*. 2006;83(2):83-90.
29. Cawley GC, Talbot NLC. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. *Neural Netw*. 2004;17(10):1467-1475.
30. Smith SS, Stewart ME, Davies BM, Kotter MRN. The prevalence of asymptomatic and symptomatic spinal cord compression on magnetic resonance imaging: a systematic review and meta-analysis. *Global Spine J*. 2021;11(4):597-607.
31. Tetreault LA, Côté P, Kopjar B, Arnold P, Fehlings MG. A clinical prediction model to assess surgical outcome in patients with cervical spondylotic myelopathy: internal and external validations using the prospective multicenter AOSpine North American and international datasets of 743 patients. *Spine J*. 2015;15(3):388-397.
32. Tetreault LA, Kopjar B, Vaccaro A, et al. A clinical prediction model to determine outcomes in patients with cervical spondylotic myelopathy undergoing surgical treatment: data from the prospective, multi-center AOSpine North America study. *J Bone Joint Surg Am*. 2013;95(18):1659-1666.
33. Kadanka Z, Kerkovsky M, Bednarik J, Jarkovsky J. Cross-sectional transverse area and hyperintensities on magnetic resonance imaging in relation to the clinical picture in cervical spondylotic myelopathy. *Spine (Phila Pa 1976)*. 2007;32(23):2573-2577.
34. Karadimas SK, Erwin WM, Ely CG, Dettori JR, Fehlings MG. Pathophysiology and natural history of cervical spondylotic myelopathy. *Spine (Phila Pa 1976)*. 2013;38(22 Suppl 1):S21-S36.
35. Yukawa Y, Kato F, Yoshihara H, Yanase M, Ito K. MR T2 image classification in cervical compression myelopathy: predictor of surgical outcomes. *Spine (Phila Pa 1976)*. 2007;32(15):1675-1679.
36. Yagi M, Ninomiya K, Kihara M, Horiuchi Y. Long-term surgical outcome and risk factors in patients with cervical myelopathy and a change in signal intensity of intramedullary spinal cord on magnetic resonance imaging. *J Neurosurg Spine*. 2010;12(1):59-65.
37. Mastronardi L, Elsawaf A, Roperto R, et al. Prognostic relevance of the postoperative evolution of intramedullary spinal cord changes in signal intensity on magnetic resonance imaging after anterior decompression for cervical spondylotic myelopathy. *J Neurosurg Spine*. 2007;7(6):615-622.
38. Al-Mefty O, Harkey LH, Middleton TH, Smith RR, Fox JL. Myelopathic cervical spondylotic lesions demonstrated by magnetic resonance imaging. *J Neurosurg*. 1988;68(2):217-222.
39. Nouri A, Martin AR, Kato S, Reihani-Kermani H, Riehm LE, Fehlings MG. The relationship between MRI signal intensity changes, clinical presentation, and surgical outcome in degenerative cervical myelopathy: analysis of a global cohort. *Spine (Phila Pa 1976)*. 2017;42(24):1851-1858.
40. Le Bihan D, Mangin JF, Poupon C, et al. Diffusion tensor imaging: concepts and applications. *J Magn Reson Imaging*. 2001;13(4):534-546.
41. Martin AR, De Leener B, Cohen-Adad J, et al. Clinically feasible microstructural MRI to quantify cervical spinal cord tissue injury using DTI, MT, and T2*-weighted imaging: assessment of normative data and reliability. *AJNR Am J Neuroradiol*. 2017;38(6):1257-1265.
42. Xu J, Shimony JS, Klawiter EC, et al. Improved in vivo diffusion tensor imaging of human cervical spinal cord. *Neuroimage*. 2013;67:64-76.

Disclosures

Dr. Sherrod reported grants from the Cervical Spine Research Society (CSRS) outside the submitted work. In addition, Dr. Sherrod had a patent (no. 18/114,157) pending. Dr. Mazur reported fellowship support from Medtronic and personal fees from Cerapedics outside the submitted work. In addition, Dr. Mazur had a patent for motorized skeletal traction (no. 18/114,157) issued. Dr. Bisson reported personal fees from Stryker, Medtronic, and minus; and stock from nView, Proprio, and SeeAll outside the submitted work. Dr. Song reported a patent (no. 61/345,367) issued.

Author Contributions

Conception and design: Zhang, Yakdan, Kaleem, Javeed, Greenberg, Song. Acquisition of data: Zhang, Yakdan, Kaleem, Javeed, Botterbush, Reis. Analysis and interpretation of data: Zhang, Yakdan, Kaleem, Javeed, Greenberg, Botterbush, Reis, Sherrod, Song. Drafting the article: Zhang, Yakdan, Kaleem, Javeed, Twitchell. Critically revising the article: Ray, Zhang, Kaleem, Javeed, Greenberg, Botterbush, Benedict, Twitchell, Sherrod, Mazur, Mahan, Dailey, Bisson, Song. Reviewed submitted version of manuscript: Zhang, Kaleem, Javeed, Greenberg, Botterbush, Benedict, Hongsermeier-Graves, Twitchell, Sherrod, Mazur, Mahan, Dailey, Bisson, Song. Approved the final version of the manuscript on behalf of all authors: Ray. Statistical analysis: Zhang, Yakdan, Kaleem. Administrative/technical/material support: Botterbush, Song. Study supervision: Ray, Bisson, Song.

Supplemental Information

Online-Only Content

Supplemental material is available with the online version of the article.

Supplementary Methods and Tables. https://thejns.org/doi/suppl/10.3171/2024.4.SPINE24107.

Previous Presentations

This work was presented at the Spine Summit 2024 annual meeting in Las Vegas, NV, on February 22–24, 2024.

Correspondence

Wilson Z. Ray: Washington University in St. Louis, MO. rayz@wustl.edu.