BMJ Neurology Open

Spinal cord demyelination predicts neurological deterioration in patients with mild degenerative cervical myelopathy

Abdul Al-Shawwa,[1] Michael Craig,[2] Kalum Ost,[1] David Anderson,[3] Steve Casha,[2,4] W Bradley Jacobs,[2,4] Nathan Evaniew,[2,5] Saswati Tripathy,[2] Jacques Bouchard,[2,5] Peter Lewkonia,[2,5] Fred Nicholls,[2,5] Alex Soroceanu,[2,5] Ganesh Swamy,[2,5] Kenneth C Thomas,[2,5] Stephan duPlessis,[2,4] Michael MH Yang,[2,4] Julien Cohen-Adad ,[6,7,8] Nicholas Dea,[9] Jefferson R Wilson,[10] David W Cadotte [1,2,3,4]

To cite: Al-Shawwa A, Craig M, Ost K, *et al*. Spinal cord demyelination predicts neurological deterioration in patients with mild degenerative cervical myelopathy. *BMJ Neurology Open* 2025;**7**:e000940. doi:10.1136/bmjno-2024-000940

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjno-2024-000940).

This work has been presented at the 25th Annual Scientific Conference of the Canadian Spine Society.

Received 07 October 2024
Accepted 18 January 2025

Check for updates

© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Dr David W Cadotte;
david.cadotte@ucalgary.ca

ABSTRACT

Background Degenerative cervical myelopathy (DCM) is the most common form of atraumatic spinal cord injury globally. Clinical guidelines regarding surgery for patients with mild DCM and minimal symptoms remain uncertain. This study aims to identify imaging and clinical predictors of neurological deterioration in mild DCM and explore pathophysiological correlates to guide clinical decision-making.

Methods Patients with mild DCM underwent advanced MRI scans that included T2-weighted, diffusion tensor imaging and magnetisation transfer (MT) sequences, along with clinical outcome measures at baseline and 6-month intervals after enrolment. Quantitative MRI (qMRI) metrics were derived above and below maximally compressed cervical levels (MCCLs). Various machine learning (ML) models were trained to predict 6 month neurological deterioration, followed by global and local model interpretation to assess feature importance.

Results A total of 49 patients were followed for a maximum of 2 years, contributing 110 6-month data entries. Neurological deterioration occurred in 38% of cases. The best-performing ML model, combining clinical and qMRI metrics, achieved a balanced accuracy of 83%, and an area under curve-receiver operating characteristic of 0.87. Key predictors included MT ratio (demyelination) above the MCCL in the dorsal and ventral funiculi and moderate tingling in the arm, shoulder or hand. qMRI metrics significantly improved predictive performance compared to models using only clinical (bal. acc=68.1%) or imaging data (bal. acc=57.4%).

Conclusions Reduced myelin content in the dorsal and ventral funiculi above the site of compression, combined with sensory deficits in the hands and gait/balance disturbances, predicts 6-month neurological deterioration in mild DCM and may warrant early surgical intervention.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Mild degenerative cervical myelopathy (DCM) presents a challenge in clinical management due to its unpredictable natural history. This is caused by a lack of reliable prognostic indicators of neurological deterioration, leading to a lack of consensus on when to recommend surgical intervention for patients with minimal symptoms but who may deteriorate.

WHAT THIS STUDY ADDS

⇒ This study demonstrates that combining quantitative MRI (qMRI) metrics, particularly magnetisation transfer (MT) ratios, with clinical outcome measures significantly enhances the ability to predict early neurological deterioration in mild DCM.

⇒ Our single-centre study comprises one of few longitudinal, multiyear cohorts of patients with non-operative mild DCM.

⇒ Demyelination in the dorsal and ventral spinal cord regions above the level of compression was found to be a key marker of deterioration within patients.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Incorporating qMRI methods, such as tract-specific MT ratio, may allow for better-informed surgical decision-making and early intervention in individual patients with mild DCM.

INTRODUCTION

Degenerative cervical myelopathy (DCM) is the most common form of atraumatic cervical spinal cord injury.[1 2] DCM is an umbrella term for several degenerative changes in the spinal column that cause progressive compression of the cervical spinal cord. These underlying changes may include cervical spondylosis, ossification of the posterior longitudinal ligament, ossification of the ligamentum flavum and degenerative disc disease.[1] Typical symptoms include gait imbalance, dexterity impairment in the hands, upper

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

and lower extremity numbness and sphincter dysfunction.[3] Currently, the most widely accepted method of classifying DCM severity is through the clinician-reported modified Japanese Orthopaedic Association (mJOA) scale. The mJOA scale assesses upper and lower extremity motor dysfunction, upper extremity sensory dysfunction and sphincter dysfunction.[4] Scores are stratified into mild (mJOA 15–18), moderate (mJOA 12–14) and severe (mJOA<12) DCM.[5] The minimum clinically important difference, for patients with mild, moderate and severe DCM with a change in mJOA of 1, 2 or 3 points across time, respectively.[6]

Surgery is currently the only effective treatment for DCM, yet clinical guidelines remain unclear on its effectiveness for patients with mild disease.[7] Likewise, an international survey of 699 surgeons, physicians and academics found no consensus on the management of mild DCM.[8] A cost-utility analysis study suggested that early surgery for patients with mild DCM was cost-effective for the Canadian healthcare system, being associated with lifetime gains in health-related quality of life (HRQOL).[9] However, if an individual patient is minimally affected by symptoms and their risk of neurological deterioration could be reliably estimated as low, then surgical treatment would be unlikely to be superior to the natural history.

The widespread adoption of MRI has allowed for the confirmation of DCM diagnosis and the development of sophisticated treatment plans.[10] Conventional MRI acquisitions typically include T_2-weighted series in the sagittal and axial plane which provide sharp contrast between the spinal cord, cerebrospinal fluid and surrounding structures.[11] Support in the literature has grown with respect to quantitative and advanced MRI techniques capable of capturing metrics of axon integrity and spinal cord demyelination. A systematic review of clinical studies suggested the importance of diffusion tensor imaging (DTI) and magnetisation transfer (MT) sequences in improving the diagnosis and treatment of DCM.[12] DTI measures the directional diffusivity of water per voxel; thus, axon integrity metrics could be derived and could predict functional impairment in mild-to-moderate DCM.[13] MT scans provide a measure of myelin content through the excitation of protons in water molecules bound to macromolecules and can ultimately be represented as a ratio (MT ratio (MTR)).[14] However, the inclusion of quantitative MRI (qMRI) protocols requires advanced analytical techniques that are not implemented routinely.[15]

The use of semi-automated machine learning (ML) based systems is one means of integrating qMRI metrics into a clinical decision-making framework. Currently, predictive modelling of neurological deterioration in patients with mild DCM is inaccurate, unreliable and rarely seen in the literature. One study was able to predict changes in the Short Form-36 HRQOL measure with an area under curve (AUC) of 0.77–0.78 in 193 patients with mild DCM.[16] The study did not include qMRI metrics within its modelling pipeline. While showing promise, no ML models have been published that predict neurological deterioration in non-operative patients with mild DCM. Furthermore, existing mild DCM ML models lack the inclusion of comprehensive qMRI-derived metrics, such as fractional anisotropy (FA) and MTR.

In this study, we developed supervised ML models aimed at predicting neurological deterioration among patients with non-operative mild DCM who were enrolled in our observational study. We observed patients from the time of diagnosis and at 6-month intervals after enrolment with clinical outcome measures and qMRI metrics focussing on axonal integrity (DTI-derived) and demyelination (MT-derived). We also explored pathophysiological correlates that may offer clinicians insight to help guide decision-making.

METHODS

Study subjects

As of January 2023, 508 patients with a diagnosis of DCM were enrolled in the Canadian Spine Outcomes Research Network Registry through the University of Calgary Spine Program. A subset of these patients were enrolled in concurrent longitudinal studies titled 'Personalized decision making after a diagnosis of cervical myelopathy: quantitative MRI within an artificial intelligence framework' and 'Natural History of Mild Degenerative Cervical Myelopathy: an observational study'. These longitudinal studies included the collection of an advanced MRI and clinical metric protocol. Exclusion criteria for this study included patients with other confounding neurological conditions (ie, Parkinson's disease), patients with other compressive spinal cord pathologies (ie, spinal cord tumours) and patients who did not fall within the mild mJOA severity class.

For this study, we identified 49 consecutive patients with non-operative mild DCM whose data was collected for up to 2 years using 6-month intervals. All collected data was de-identified and stored in a Brain Imaging Data Structure format before analysis.[17]

Patient and public involvement

No patient and public involvement.

Clinical metrics data collection

The mJOA was collected at baseline and the changes in mJOA after 6 months were used as the primary outcome in this study.[4] Deterioration was considered as a loss of 1 or more points on mJOA over the study period. Other clinical metrics collected include gait and balance function (30-metre walk test and the Berg Balance Scale),[18] hand function (grip strength test and quick disabilities of the arm, shoulder and hand (quickDASH) test)[19] and global function (Myelopathy Disability Index).[20] We also collected included age, sex, body mass index (BMI), smoking status and the presence or absence of a Hoffman sign.

MRI data collection

Advanced imaging protocol

All MRI data was acquired on the same 3T GE scanner. A T_2-weighted acquisition was acquired as follows:

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

FIESTA-C sequence (T_2-weighted); 512×512, NEX 1.0, FOV 200 mm, slice thickness 0.4 mm; resulting in a voxel size of 0.4×0.4 × 0.4 mm^3, collected in the coronal plane. MRI records also contained sequences suggested in the spine-generic protocol, namely: DTI and MT image acquisitions, each with 13 axial slices across the C1–C7 vertebrae, using a variable gap to alternate between bone and intervertebral discs. DTI used spin-echo single-shot echo planar imaging with three acquisitions averaged offline, b=800 s/mm^2 in 25 directions, 5 images with b=0 s/mm^2 and resolution of 1×1 × 5 mm^3 (7 min total). MT used two-dimensional spoiled gradient echo with/without MT pre-pulse (offset=1.2 kHz), with 1×1 × 5 mm^3 voxels, for 8 min total. Total imaging time for all acquisitions together came to approximately 30 min after accounting for patient positioning, slice prescription and high-order shimming.

Atlas-based spinal cord analysis

DTI and MT-derived metrics were used to quantitatively assess spinal cord axon integrity and demyelination respectively. Quantitative metrics were derived using the Spinal Cord Toolbox (SCT), an all-in-one open-source spinal cord specific Python package designed to analyse and derive metrics of the spinal cord.[21] Initially, the spinal cord was segmented on T_2-weighted images,[22] followed by semi-automated vertebral level labelling.[23] The vertebral levels were then resampled in the DTI/MT space. This allowed for a more accurate estimation of the vertebral levels using the anatomical contrast advantages of T_2-weighted images. Next, the PAM50 template[24] was registered to the DTI/MT space and the white matter atlas[25] was warped to fit the image appropriately. Quality control inspection of the final warped atlas and vertebral levels was done by a graduate student with 5 years of MRI analysis experience.

DTI metrics, including FA, axial diffusivity, medial diffusivity, radial diffusivity, along with MT metrics such as MTR were derived from one vertebral level above and below the maximally compressed cervical level (MCCL). DTI/MT metrics were collected in this way due to poor quality control of SCT metrics at the levels of severe compression and consistent analysis failure. The MCCL was determined individually by a graduate student focused on DCM research with 5 years of experience in MRI analysis and a neurosurgery-trained spine fellow on all collected MRI scans. MCCL was established as a consensus between both raters. Using the SCT atlas-based analysis, DTI and MT metrics were derived as described for the entire spinal cord, white matter tracts collectively, grey matter tracts collectively, the ventral funiculi, dorsal funiculi and lateral funiculi (figure 1). A complete list of derived metrics can be found in online supplemental appendix A.

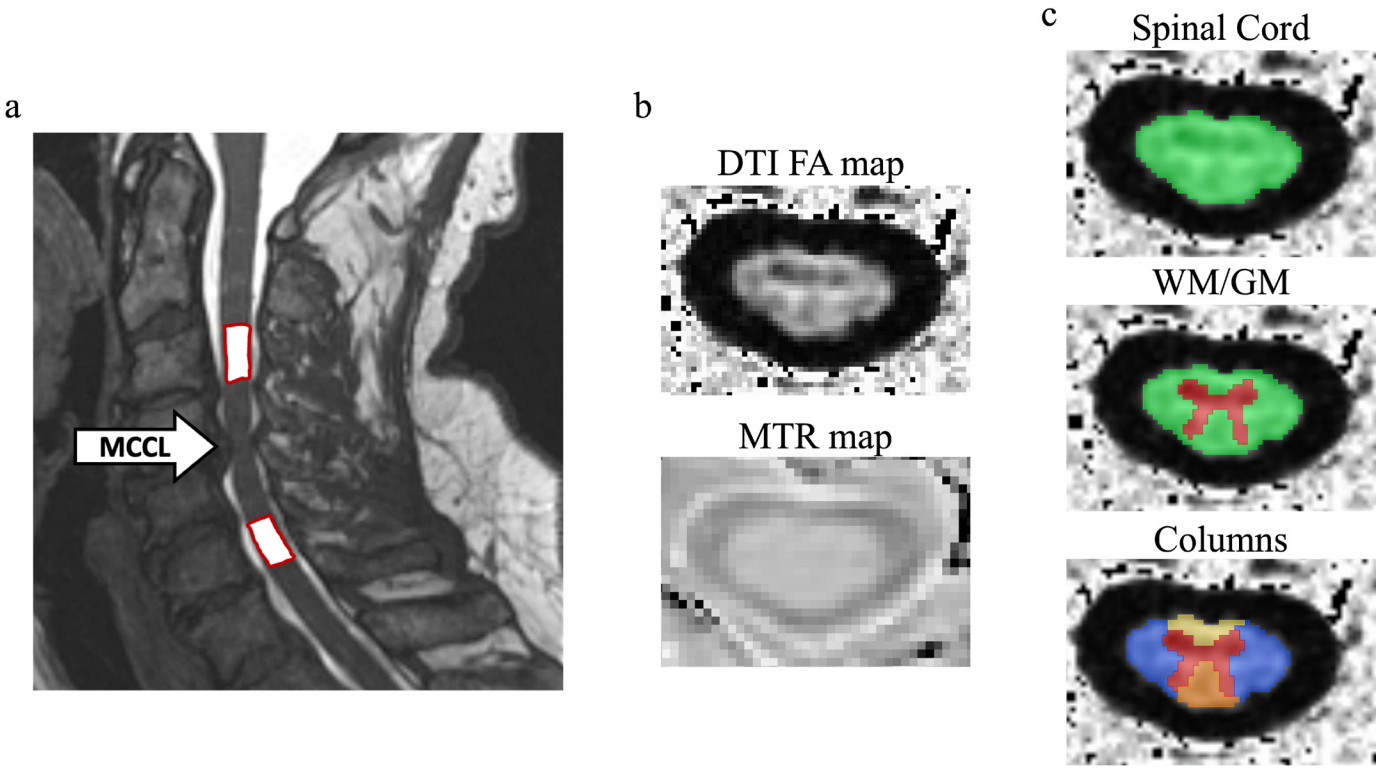

Figure 1 Representation of collected metrics above and below MCCL. Subplot A shows the T2w sagittal image of a patient with mild DCM with focal cord compression at the C4/C5 disc level. Subplot B contains axial images of the DTI FA map and MTR map above the level of MCCL. Subplot C contains segmentation and atlas examples of the regions collected for data extraction with the study. DCM, degenerative cervical myelopathy; DTI, diffusion tensor imaging; FA, fractional anisotropy; GM, grey matter; MCCL, maximally compressed cervical level; MTR, magnetisation transfer ratio; WM, white matter.

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Manually derived MRI metrics

Each T2-weighted scan underwent manual analysis of cord compression and spinal canal diameter measurement in both the axial and sagittal plane. This consisted of evaluating the C2-T1 disc levels as either having cerebrospinal fluid present, minimal cord compression or severe cord compression. Furthermore, at each vertebral level spinal canal diameter was measured between the middle of the dorsal surface of the vertebral body to the middle of the ventral surface of the lamina. This methodology was followed based on previous work aimed at identifying neurological deterioration in patients with non-operative DCM.[26]

A subset of 50 randomly selected scans underwent repeat evaluations by a neurosurgery fellow with a subspecialty focus in spine surgery to validate assessment reliability. This was done through the use of the Cohen kappa statistic and via two-way random effect, single random rater intraclass correlation testing between the two raters.

ML approach

Data preprocessing

All data preprocessing was done using the SciKit-Learn python package. Data preprocessing included splitting the data into 80% training data and 20% testing data. This was followed by data imputation of missing data using a linear regression iterative imputer for continuous variables and a mode-based imputer for categorical variables. Data scaling included minimum–maximum scaling for continuous variables and one hot encoding for categorical variables. Recursive feature elimination with cross-validation (RFECV) was used to eliminate redundant features, followed by Principal Component Analysis (PCA) to reduce the dimensionality of the feature set. This dimensionality reduction is intended to combat model overfitting and improve model generalisability. Following this reduction, Synthetic Minority Over-Sampling was implemented to combat the data set imbalance.

Hyperparameter tuning and model evaluation

Logistic regression (LR), random forest classifier (RFC) and support vector classifier (SVC) models were all tuned and evaluated to determine the best-fit model. These were chosen based on previous evidence suggesting their efficacy within a DCM population.[27] Hyperparameter tuning was done through a fivefold, 10-repetition stratified cross-validated grid search aimed at maximising the model's balanced accuracy score. A complete list of tested hyperparameters is included in online supplemental appendix B. The tuned model was then evaluated on the testing set and evaluated for balanced accuracy, F1 score (harmonic mean of precision and recall), receiver operating characteristic-AUC (ROC-AUC), sensitivity and specificity. Shapley Additive Explanations (SHAP), a game theoretic approach to explain ML model predictions, was employed on the best-performing model to evaluate global model feature importance.[28] Local Interpretable Model-agnostic Explanations (LIME) were employed on the most reliably predicted cases in each of the deteriorate and non-deteriorate cohorts to evaluate local model feature importance.[29]

RESULTS

Patient characteristics

Initially, 74 patients with a diagnosis of mild DCM were enrolled in our longitudinal studies. 25 patients were excluded from this analysis as the patient and surgeon elected for operative treatment rather than an observational treatment plan or the patient was missing their 6-month mJOA data. Ultimately, there were 110 included scans from 49 unique patients. 42 scans correspond to patients who experienced neurological deterioration within 6 months and 68 scans correspond to patients for whom the mJOA remained stable or improved across 6 months (figure 2). The longitudinal nature of the study meant that each unique patient underwent multiple assessments and each assessment was used as an independent data point within the model (figure 3).

The mean baseline mJOA of the patient group was 16.7±1.0. The mean change in mJOA over 6 months was 1.4±0.8 and 0.4±0.7 for patients who deteriorated and did not deteriorate, respectively. The mean overall age of the participants was 58.7±12.9 and the average BMI was 26.5±4.8. A minority of patients (14.5%) smoked cigarettes. The C5/C6 disc level was the most common MCCL for both patients who deteriorate (43%) and those who do not deteriorate (41%). A summary of patient demographics between the deterioration and non-deterioration cohorts can be seen in table 1.

Model performance

The best-performing model, as measured by average balanced accuracy, contained the data set with a combination of qMRI-derived metrics and clinical metrics (table 2). The ML models containing only imaging-based metrics had the greatest range of accuracy with a~30% difference between the tuned LR model and the SVC model. The best imaging-only model (0.643) and clinical-only model (0.732) performed worse than the worst-performing combination model (0.795). Generally, the imaging-only models had lower sensitivity than specificity, averaging 0.458 and 0.691, respectively. The combined models showed an opposite trend with a mean sensitivity of 0.917 and a mean specificity of 0.714. On the other hand, the clinical-only model had an average sensitivity and specificity of 0.667 and 0.691. When compared across data sets, the SVC was the best-performing model type with an average balanced accuracy of 0.735 followed by RFCs (0.702) and LRs (0.631). The area under the ROC curve follows a similar trend with the combined support vector classifier model (0.87), outperforming the clinical-only (0.71) and imaging-only (0.72) models (figure 4). The F1-scores in model performance followed the same trend with SVC outperforming both LR and RFC models. The SVC F1 score for the combined model (0.778)

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

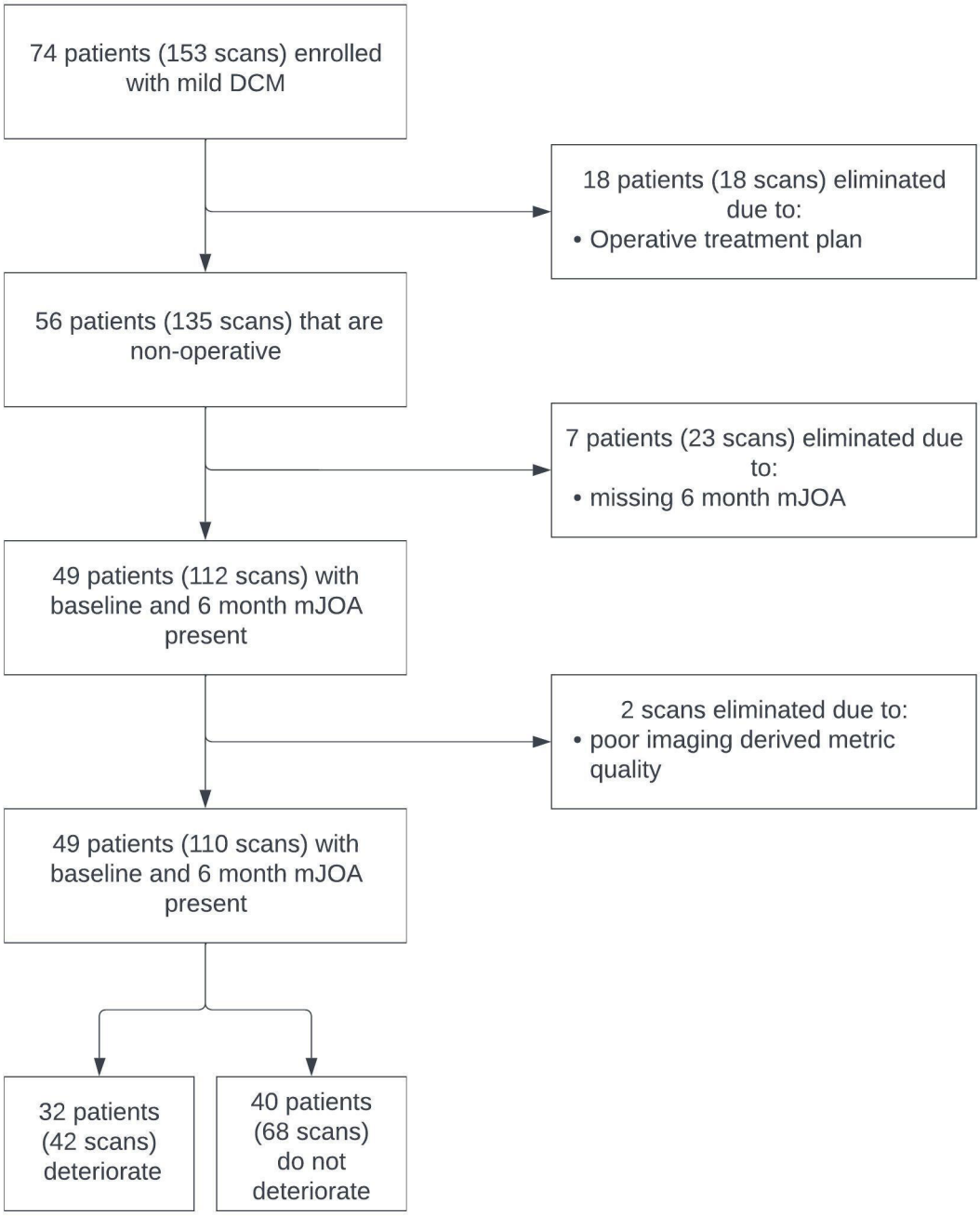

Figure 2 Patient enrolment overview. The final included data set contained 49 unique patients totalling in 110 6-month scan entries. With data collected every 6 months, many patients were categorised as both patients who deteriorate (n=42) and those who do not (n=40). Patients who have been enrolled in the study for a longer duration and contain more data points are more likely to have experienced deterioration at some point in their enrolment. Each patient scan was entered within the model as an independent data point. Overall, there are more patients who do not deteriorate than those who do, aligning with the expected natural history of mild DCM. 18 patients were eliminated due to enrolment within an operative treatment plan. A small number of scans (n=2) were eliminated due to poor imaging-derived metrics as a result of poor template registration that could not be manually corrected. DCM, degenerative cervical myelopathy; mJOA, modified Japanese Orthopaedic Association.

outperformed the clinical-only (0.667) and imaging-only (0.533) models.

Model explainability

Following RFECV, the data set was reduced from 171 unique features to 102, which was further refined to 32 principal components, which captured 90% of the variance in the data. From the original feature set, SHAP analysis was conducted on the best-performing model and highlighted the 20 most important features for model performance (figure 5) which includes 13 imaging-derived metrics and 7 clinical metrics. Of the 13 imaging-derived, 10 came from DTI/MT images and 3 were T_2-weighted based metrics. Five of the most important qMRI-derived metrics were collected above the MCCL

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

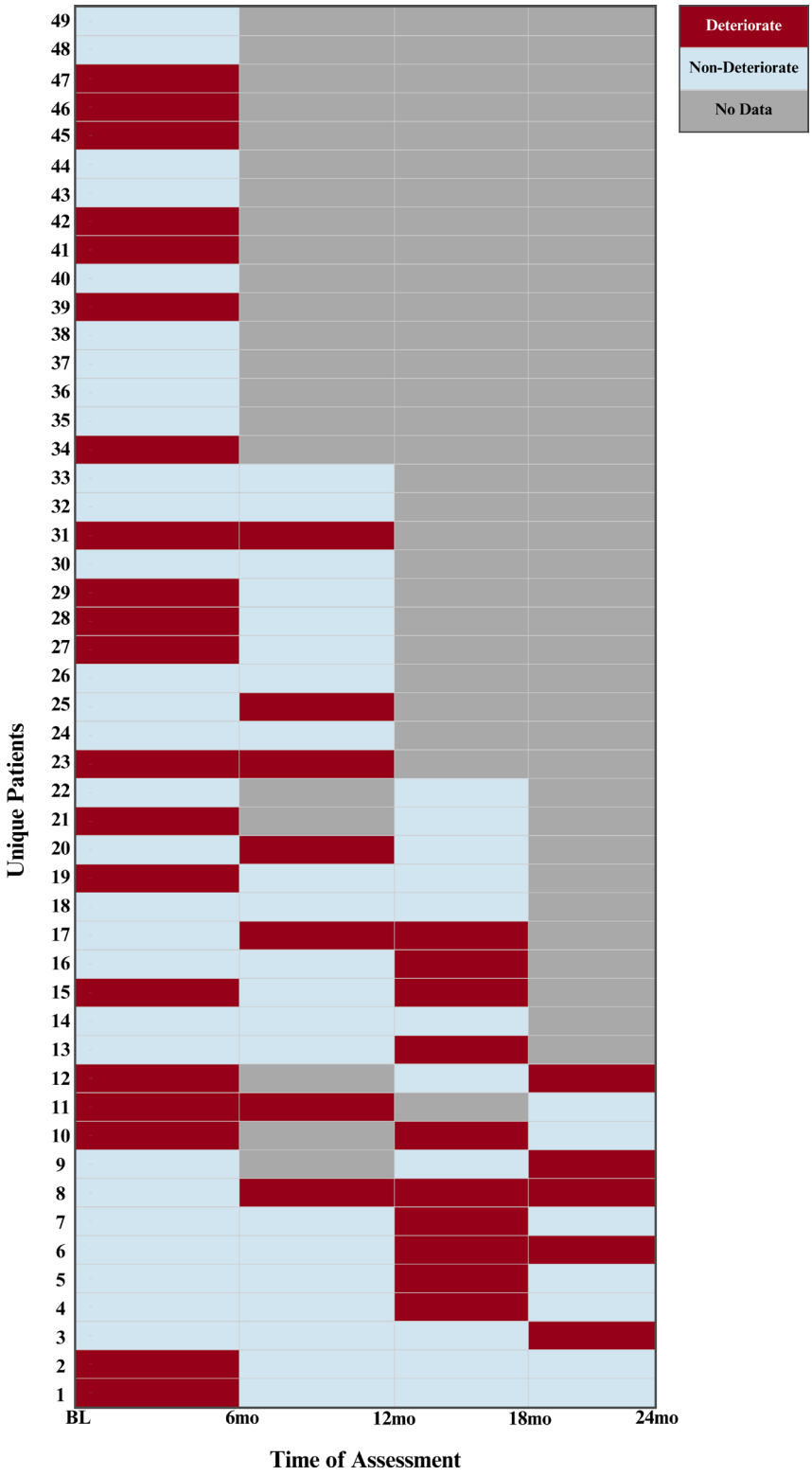

Figure 3 Representation of the change in the neurological condition of the uniquely enrolled patients. Each box represents an independent 6-month time period, with consecutive boxes representing the consecutive periods of enrolment and data collection for a single patient. A blue box means that the patient did not experience a decrease in mJOA, particularly in that 6-month duration. A red box means that the patient did experience a decrease in mJOA by 1 or more points within the past 6 months. Consecutive red boxes indicate a repeated decrease in mJOA at multiple data collection sessions. Grey boxes indicate missing or incomplete data collection for that 6-month time period. Patients who have been enrolled for longer in the study are likely to experience deterioration at some point in their enrolment, with only patients who have been enrolled for 1 year not experiencing deterioration. While deterioration is described as a decrease in mJOA by 1 or more points, the functional and quality of life decline is not captured in this figure. BL, baseline; mJOA, modified Japanese Orthopaedic Association; 6mo, 6 months; 12mo, 12 months; 18mo, 18 months; 24mo, 24 months.

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest.
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 Characteristics of patients included in the ML model data set. There is a balance between patients who are male (47%) and female (53%) and the mean age of all subjects is 58.7 years. A majority of patients contain a negative Hoffman sign. 14% of enrolled patients smoke. The most common cervical level of maximal compression is the C5/C6 region (41%) followed by the C4/C5 level (31%). A majority of patients contain a combination of pathologies leading to their DCM, with degenerative disc disease and spondylosis being the most common independently and most likely to be seen within the same patient

Characteristics	Study subjects (n=49)
Age in years	58.7 (12.9)
Sex (male/female)	23/26 (47/53)
BMI (kg/m2)	26.5 (4.8)
Smoking status (yes/no)	7/42 (14/86)
Hoffman sign present (positive/negative/NA)	19/29/1 (38/60/2)
Six-month change in mJOA score	−0.3 (1.2)
Maximally compressed cervical level	
C3/C4	6 (12)
C4/C5	15 (31)
C5/C6	20 (41)
C6/C7	8 (16)

Data is given as absolute numbers (percentage) or mean (SD).
BMI, body mass index; DCM, degenerative cervical myelopathy; mJOA, modified Japanese Orthopaedic Association Scale; ML, machine learning; NA, missing data.

and five were collected below the MCCL. The five most influential metrics to model performance include MTR of the dorsal/ventral funiculi above the MCCL, moderate tingling in the arm, shoulder and hand (quickDASH item 10), the total quickDASH score and MTR of the lateral funiculus below the MCCL, respectively. MTR above the MCCL of the dorsal funiculus had the most significant model impact with a mean model output magnitude of 0.12 (SHAP value), with the next most important feature (MTR above the MCCL of the ventral funiculus) having a magnitude of 0.07. These feature importance trends are reflected at both the overall and individual patient prediction levels (figure 6).

DISCUSSION

For patients with mild DCM, there is uncertainty about whether to pursue surgical intervention because symptoms are relatively mild and it has been difficult to reliably predict if neurological deterioration will occur. This study demonstrated that a combination of clinical and qMRI metrics that capture information relevant to the pathophysiology of spinal cord injury offer improved predictive ability with respect to clinical deterioration after a diagnosis of mild DCM. In doing so, this study has demonstrated the utility of semi-automated qMRI-derived metrics and supports the need for advanced imaging in the monitoring and clinical decision-making process for patients with mild DCM.

Retrograde degeneration as a marker of early neurological deterioration

Our models indicate that MTR may be more sensitive for detecting early neurological deterioration than DTI-based metrics in mild DCM. MTR metrics account for three of the five most important metrics in this ML model's prediction. Furthermore, MTR metrics above and below the MCCL were identified as important, suggesting the potential usefulness in identifying retrograde and anterograde (Wallerian) axonal degeneration. In a novel cervical spondylotic myelopathy rat model, evidence of

Table 2 Tuned ML model testing set performance on differing data set variations. The overall best-performing model based on balanced accuracy was the SVC when clinical and imaging metrics were combined (0.830). That model's sensitivity and specificity were 0.875 and 0.786, respectively. The worst-performing model based on balanced accuracy was the logistic regression model of imaging-only metrics. A majority of models performed within the 0.6–0.7 balanced accuracy range. The RFC of the combined clinical and imaging metrics contained a sensitivity score of 1.000, indicating that it was capable of correctly predicting which patients would deteriorate in every case

Data set variation	Model type	Balanced accuracy	F1 score	Sensitivity	Specificity
Imaging metrics only	LR	0.464	0.400	0.500	0.429
	RFC	0.616	0.462	0.375	0.857
	SVC	0.643	0.533	0.500	0.786
Clinical metrics only	LR	0.634	0.556	0.625	0.643
	RFC	0.670	0.588	0.625	0.714
	SVC	0.732	0.667	0.750	0.714
Combined imaging and clinical metrics	LR	0.795	0.737	0.875	0.714
	RFC	0.821	0.762	1.000	0.643
	SVC	0.830	0.778	0.875	0.786

LR, logistic regression; ML, machine learning; RFC, random forest classifier; SVC, support vector classifier.

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

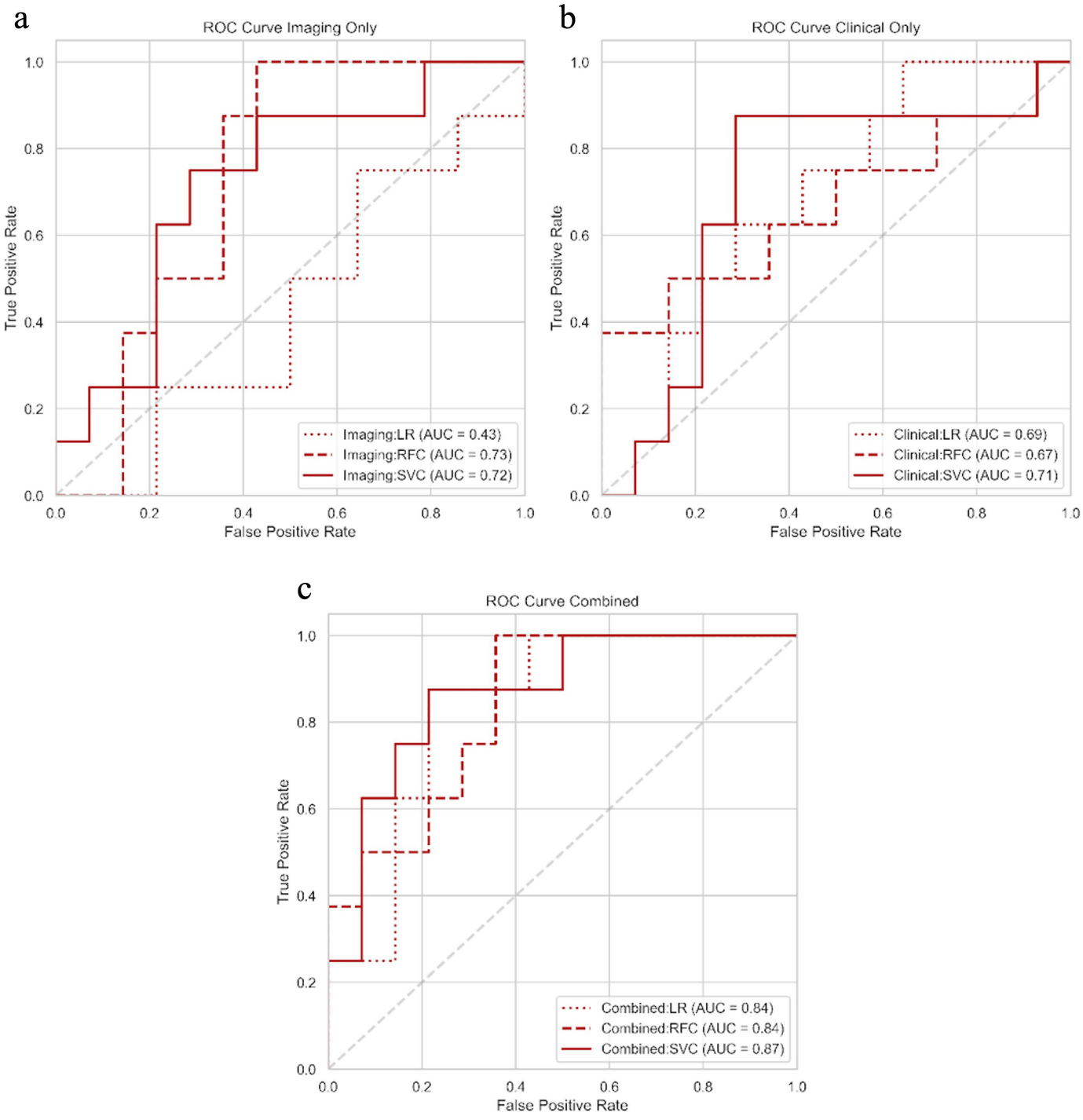

Figure 4 ROC curves for the imaging-only, clinical-only and combined data sets across varying model types. Subplot A contains the imaging-only metric models, subplot B contains the clinical metric-only models and subplot C contains the models using a combination of clinical and imaging metrics. The AUC of each model is reported in the figure caption. When imaging and clinical metrics were combined, the most performed better with AUC scores above 0.8 than when imaging and clinical metrics were used independently. The logistic regression model for imaging-only metrics performed worse with an AUC lower than random guessing (0.43). AUC, area under curve; LR, logistic regression; RFC, random forest classifier; ROC, receiver operating characteristic; SVC, support vector classifier.

retrograde and Wallerian degeneration, particularly in spinal cords facing prolonged compression has been shown.[30] Furthermore, our ML model highlighted retrograde degeneration in the dorsal funiculi as an important indicator of neurological deterioration, aligning with electrophysiological findings from Kanchiku *et al* rabbit model study.[31]

Ellingson *et al* have indicated that a reduction in FA values at the site of compression is a strong biomarker for differentiating between patients with moderate DCM

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

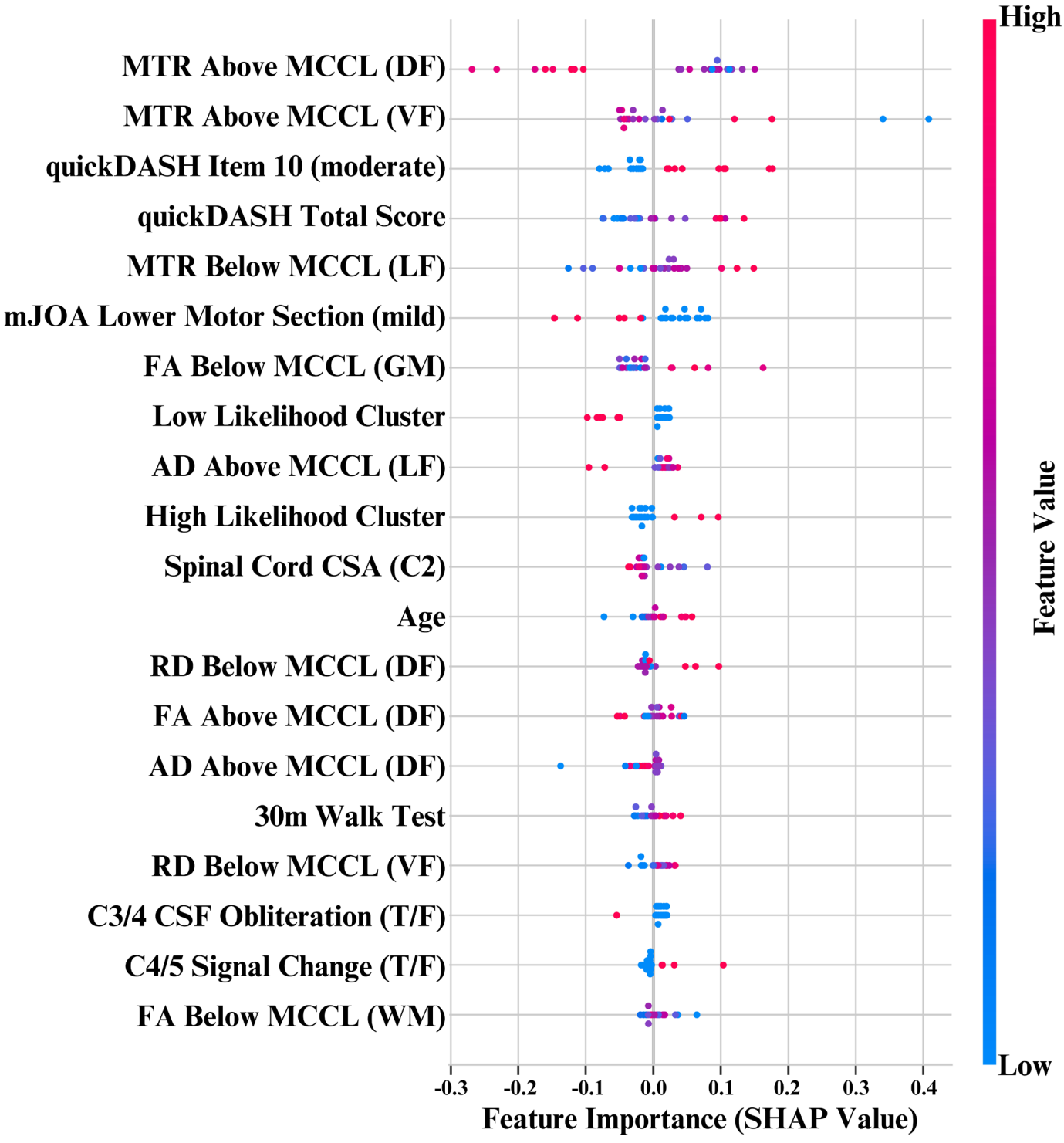

Figure 5 The 20 most important features for model performance are represented through SHAPely values. The represented plot comes from the combined clinical and imaging SVC model following tuning, validation and testing. The plot contains the 20 most important metrics for model performance with 10/20 coming from qMRI metrics. 3/5 most influential metrics for the model's performance were metrics of demyelination as represented through the MTR. The most important clinical metrics were Item 10 of the quickDASH scale (tingling and numbness in the arm, shoulder or hand) and the overall quickDASH score of the patient. AD, axial diffusivity; CSA, cross-sectional area; CSF: cerebrospinal fluid; DF, dorsal funiculus; FA, fractional anisotropy; GM, grey matter; LF, lateral funiculus; MCCL, maximally compressed cervical level; mJOA, modified Japanese Orthopaedic Association; MTR, magnetisation transfer ratio; quickDASH, quick disabilities of the arm, shoulder and hand questionnaire; qMRI, quantitative MRI; RD, radial diffusivity; SHAP, Shapley Additive Explanations; SVC, support vector classifier; T/F, True/False; VF, ventral funiciulus; WM, white matter.

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

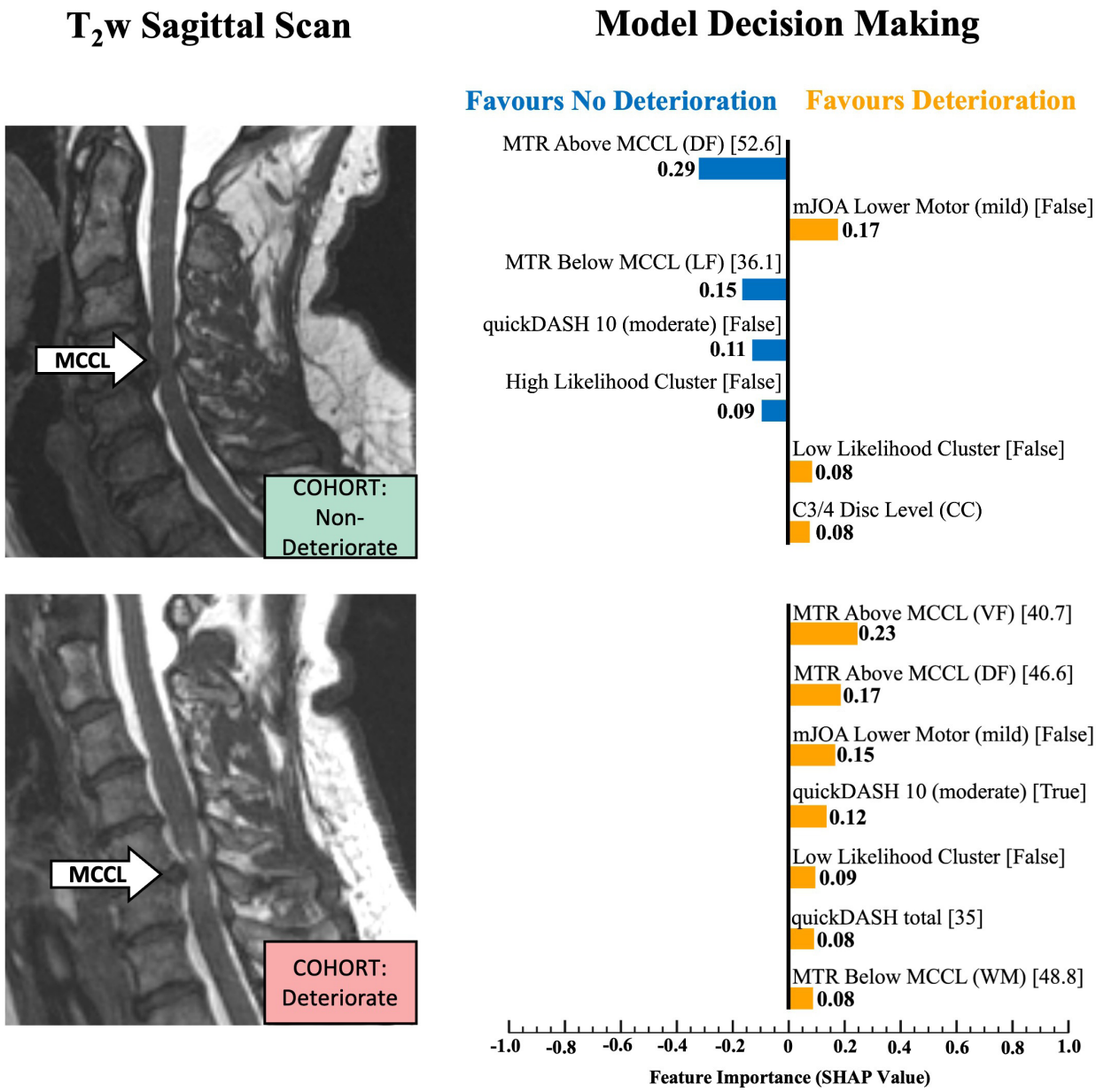

Figure 6 Model prediction feature importance for the most reliably predicted case in each cohort. The prediction probability of the non-deteriorate patient was 0.98 and the prediction probability of the deteriorate patient was 0.92. The original value of the metric for the patient is represented through (value). CC, cord compression; DF, dorsal funiculus; LF, lateral funiculus; MCCL, maximally compressed cervical level; mJOA, modified Japanese Orthopaedic Association; MTR, magnetisation transfer ratio; quickDASH, quick disabilities of the arm, shoulder and hand questionnaire; SHAP, Shapley Additive Explanations; VF, ventral funiciulus; WM, white matter.

and patients with mild/asymptomatic DCM.[13] However, the study indicated that FA values could not significantly differentiate between patients with no neurological impairment and those with mild neurological impairment.[13] This is demonstrated in this study in which FA was the most important of the DTI metrics with an overall feature importance of seventh place.

Clinical and qMRI metrics play complementary roles in determining the risk of neurological deterioration. When trained on clinical or qMRI metrics independently, the ML model's performance accuracy decreased on average by 9%. This indicates that while qMRI metrics are important to model performance, the inclusion of clinical metrics that symptomatically express spinal cord deterioration can act to improve the model's predictive capacity. Interestingly, the most important clinical metric of the model's performance was the quickDASH score, specifically a 'moderate' score in quickDASH item 10 (moderate tingling in the arm, shoulder and hand). This aligns with qMRI MTR metrics in the dorsal funiculus which contain sensory tracts and were highlighted as the most important feature in the model. Ultimately, it appears that early myelin breakdown in the spinal cord at regions affecting fine sensory function (such as upper extremity sensory

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

function) may be indicative of early neurological deterioration in patients with mild DCM and may warrant early surgical intervention. Although, it must be highlighted that our target outcome, the mJOA scale, is heavily influenced by hand sensory function and therefore may cause sensory clinical metrics to be weighed more significantly within the model. Further validation and investigation are needed.

DCM pathobiology

There are many pathobiological consequences of prolonged and chronic compression of the cervical spinal cord.[1] The chronic pathobiological consequences for patients with mild DCM remain unclear, particularly for patients who may face mild compression over many years and who are more likely to induce compensatory mechanisms within the cord.[32] Recent studies have indicated that chronic intraparenchymal ischaemia may be the cause of neural degeneration within the spinal cord,[30 32] inducing a unique immune response.[33] Based on the feature importance of our ML model, which is most strongly affected by demyelination surrogates such as MTR and less strongly by axonal markers, we suggest that this immune response may impact cord myelination more than axon integrity within patients with mild DCM, although further investigation is required.

Another study has suggested that neuronal loss and demyelination can be explained through spinal cord tethering which increases intramedullary pressure.[34] Interestingly, the same study found demyelination to begin in the ventral funiculus and progress to the lateral and dorsal funiculus.[34] This contrasts with our finding suggesting a greater emphasis on the dorsal funiculus in patients with mild DCM followed by the ventral and lateral funiculus. It has been suggested that compensatory mechanisms, particularly supraspinal and 'corticospinal reserve capacity' may help to preserve neurological function while degeneration takes place.[35 36] Similarly, our ML model shows low importance to many gross motor skills, such as tasks in the Berg Balance Scale, as they are likely to remain unchanged in patients with mild DCM. Our model did highlight the lower extremity motor component of the mJOA score as an important metric, even if only mild compromise to stability is seen. We speculate that if compromise to gross motor function such as walking stability is seen in the early stages of DCM progression and is not masked by compensatory mechanisms, spinal tract compromise by way of demyelination may already be occurring.

Limitations

The primary limitation of this study is its single-centre design which restricts the model's generalisability. The use of 6-month intervals led to multiple scans from the same patients, who are monitored more closely than typical clinical practice, possibly enhancing the detection of minor deterioration compared with broader cohorts. Furthermore, the use of only non-operative patients may limit the model's generalisability in the context of a broader patient population. Another limitation of this study is the lack of inter-rater testing of the collected clinical metrics. The reliance on SHAP and LIME, through surrogate-model-based explanations, is a limitation as some models lack inherent interpretability making complete verification of the identified important features challenging. Despite this, the overall trends in feature importance are likely robust. The inclusion of numerous qMRI and clinical metrics on a limited number of patient scans increases model complexity which may lead to overfitting and reduced statistical power. We mitigated this through PCA, RFECV and cross-validation, but a larger data set of unique patients would enhance reliability. This study is also limited in SCT's segmentation algorithm capability which currently does not segment at regions of severe compression. Future studies with updated SCT segmentation algorithms will help to overcome this limitation, allowing for metric collection at the site of maximal compression.

CONCLUSION

This study presents a new ML model using qMRI-derived and clinical metrics to predict 6-month neurological deterioration in patients with non-operative mild DCM, achieving 83% peak balanced accuracy. Key imaging and clinical metrics identified as potential deterioration markers include retrograde MTR in the dorsal and ventral funiculus, mild gait imbalance and moderate tingling in the arm, shoulder and hand. While these factors reliably contribute to the model's predictions, further investigation and validation are necessary. Future studies plan to expand the data set to multicentre patient groups, including a wider range of disease severity and work toward validating these models for clinical application. Furthermore, future studies should work to include a more comprehensive set of imaging and clinical metrics, including morphological spinal cord features, and test advanced ML approaches such as deep learning methods.

Author affiliations
[1]Hotchkiss Brain Institute, University of Calgary Cumming School of Medicine, Calgary, Alberta, Canada
[2]Combined Orthopaedic and Neurosurgery Spine Program, University of Calgary Department of Surgery, Calgary, Alberta, Canada
[3]Department of Biochemistry and Molecular Biology, University of Calgary Cumming School of Medicine, Calgary, Alberta, Canada
[4]University of Calgary Department of Clinical Neurosciences, Calgary, Alberta, Canada
[5]Section of Orthopaedic Surgery, University of Calgary Department of Surgery, Calgary, Alberta, Canada
[6]Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, Quebec, Canada
[7]Mila - Quebec Artificial Intelligence Institute, Montreal, Quebec, Canada
[8]Functional Neuroimaging Unit, Université de Montréal, Montreal, Quebec, Canada
[9]Combined Neurosurgical and Orthopaedic Spine Program, The University of British Columbia Department of Surgery, Vancouver, British Columbia, Canada
[10]Division of Neurosurgery, University of Toronto Department of Surgery, Toronto, Ontario, Canada

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Contributors AA-S: data analysis/collection, model creation, study design, statistical analysis, manuscript writing and editing, figure creation, final approval; MC: data analysis/collection, study design, manuscript writing and editing, figure creation, final approval; KO: data analysis, study design, manuscript revision/editing, final approval; DA: study design, model feedback/revision, manuscript editing, final approval; SC: study design, study revision, manuscript editing, final approval; WBJ: study design, data collection, manuscript revision/editing, final approval; NE: study design, data collection, manuscript revision/editing, final approval; ST: study design, data analysis/collection, manuscript revision/editing, final approval; JB: data collection, manuscript revision/editing, final approval; PL: data collection, manuscript revision/editing, final approval; FN: data collection, manuscript revision/editing, final approval; AS: data collection, manuscript revision/editing, final approval; GS: data collection, manuscript revision/editing, final approval; KCT: data collection, manuscript revision/editing, final approval; SdP: data collection, manuscript revision/editing, final approval; MMHY: data collection, manuscript revision/editing, final approval; JC-A: study design, model revision/design, manuscript revision/editing, final approval; ND: study design, data collection, manuscript revision/editing, final approval; JW: study design, data collection, manuscript revision/editing, final approval; DWC: study design/supervision, data collection, manuscript revision/editing, final approval, guarantor.

Funding This work was supported by the Department of Clinical Neurosciences and the Hotchkiss Brain Institute; University of Calgary, Alberta Spine Foundation and the Canadian Institutes for Health Research.

Competing interests No, there are no competing interests.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by University of Calgary, ethics certificate number: REB-15-1332, REB-18-1614 and REB-20-0186. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Julien Cohen-Adad http://orcid.org/0000-0003-3662-9532
David W Cadotte http://orcid.org/0000-0003-2581-5903

REFERENCES

1 Nouri A, Tetreault L, Singh A, *et al*. Degenerative Cervical Myelopathy: Epidemiology, Genetics, and Pathogenesis. *Spine (Phila Pa 1976)* 2015;40:E675–93.
2 Fehlings MG, Ibrahim A, Tetreault L, *et al*. A global perspective on the outcomes of surgical decompression in patients with cervical spondylotic myelopathy: results from the prospective multicenter AOSpine international study on 479 patients. *Spine (Phila Pa 1976)* 2015;40:1322–8.
3 Davies BM, Mowforth OD, Smith EK, *et al*. Degenerative cervical myelopathy. *BMJ* 2018;360:k186.
4 Benzel EC, Lancon J, Kesterson L, *et al*. Cervical laminectomy and dentate ligament section for cervical spondylotic myelopathy. *J Spinal Disord* 1991;4:286–95.
5 Tetreault L, Kopjar B, Nouri A, *et al*. The modified Japanese Orthopaedic Association scale: establishing criteria for mild, moderate and severe impairment in patients with degenerative cervical myelopathy. *Eur Spine J* 2017;26:78–84.
6 Tetreault L, Nouri A, Kopjar B, *et al*. The Minimum Clinically Important Difference of the Modified Japanese Orthopaedic Association Scale in Patients with Degenerative Cervical Myelopathy. *Spine (Phila Pa 1986)* 2015;40:1653–9.
7 Fehlings MG, Tetreault LA, Riew KD, *et al*. A Clinical Practice Guideline for the Management of Patients With Degenerative Cervical Myelopathy: Recommendations for Patients With Mild, Moderate, and Severe Disease and Nonmyelopathic Patients With Evidence of Cord Compression. *Global Spine J* 2017;7:70S–83S.
8 Brannigan JFM, Davies BM, Mowforth OD, *et al*. Management of mild degenerative cervical myelopathy and asymptomatic spinal cord compression: an international survey. *Spinal Cord* 2024;62:51–8.
9 Malhotra AK, Shakil H, Harrington EM, *et al*. Early surgery compared to nonoperative management for mild degenerative cervical myelopathy: a cost-utility analysis. *Spine J* 2024;24:21–31.
10 Al-Mefty O, Harkey LH, Middleton TH, *et al*. Myelopathic cervical spondylotic lesions demonstrated by magnetic resonance imaging. *J Neurosurg* 1988;68:217–22.
11 Nouri A, Martin AR, Mikulis D, *et al*. Magnetic resonance imaging assessment of degenerative cervical myelopathy: a review of structural changes and measurement techniques. *Neurosurg Focus* 2016;40:E5.
12 Martin AR, Aleksanderek I, Cohen-Adad J, *et al*. Translating state-of-the-art spinal cord MRI techniques to clinical use: A systematic review of clinical studies utilizing DTI, MT, MWF, MRS, and fMRI. *Neuroimage Clin* 2016;10:192–238.
13 Ellingson BM, Salamon N, Grinstead JW, *et al*. Diffusion tensor imaging predicts functional impairment in mild-to-moderate cervical spondylotic myelopathy. *Spine J* 2014;14:2589–97.
14 Grossman RI, Gomori JM, Ramer KN, *et al*. Magnetization transfer: theory and clinical applications in neuroradiology. *Radiographics* 1994;14:279–90.
15 Martin AR, Tadokoro N, Tetreault L, *et al*. Imaging Evaluation of Degenerative Cervical Myelopathy: Current State of the Art and Future Directions. *Neurosurg Clin N Am* 2018;29:33–45.
16 Khan O, Badhiwala JH, Witiw CD, *et al*. Machine learning algorithms for prediction of health-related quality-of-life after surgery for mild degenerative cervical myelopathy. *Spine J* 2021;21:1659–69.
17 Gorgolewski KJ, Alfaro-Almagro F, Auer T, *et al*. BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput Biol* 2017;13:e1005209.
18 Berg KO, Wood-Dauphinee SL, Williams JI, *et al*. Measuring balance in the elderly: validation of an instrument. *Can J Public Health* 1992;83 Suppl 2:S7–11.
19 Beaton DE, Wright JG, Katz JN, *et al*. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am* 2005;87:1038–46.
20 Casey AT, Bland JM, Crockard HA. Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy. *Ann Rheum Dis* 1996;55:901–6.
21 De Leener B, Lévy S, Dupont SM, *et al*. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage* 2017;145:24–43.
22 Gros C, De Leener B, Badji A, *et al*. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage* 2019;184:901–15.
23 Ullmann E, Pelletier Paquette JF, Thong WE, *et al*. Automatic labeling of vertebral levels using a robust template-based approach. *Int J Biomed Imaging* 2014;2014:719520.
24 De Leener B, Fonov VS, Collins DL, *et al*. PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. *Neuroimage* 2018;165:170–9.
25 Lévy S, Benhamou M, Naaman C, *et al*. White matter atlas of the human spinal cord with estimation of partial volume effect. *Neuroimage* 2015;119:262–71.
26 Al-Shawwa A, Craig M, Ost K, *et al*. Focal compression of the cervical spinal cord alone does not indicate high risk of neurological deterioration in patients with a diagnosis of mild degenerative cervical myelopathy. *J Neurol Sci* 2024;461:123042.
27 Stephens ME, O'Neal CM, Westrup AM, *et al*. Utility of machine learning algorithms in degenerative cervical and lumbar spine disease: a systematic review. *Neurosurg Rev* 2022;45:965–78.
28 Lundberg SM, Lee S-I, *et al*. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, eds. *Advances in neural information processing systems*. Curran Associates, Inc, 2017.
29 Ribeiro MT, Singh S, Guestrin C. 'Why should i trust you?': explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016:1135–44.

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

30 Karadimas SK, Moon ES, Yu W-R, *et al*. A novel experimental model of cervical spondylotic myelopathy (CSM) to facilitate translational research. *Neurobiol Dis* 2013;54:43–58.
31 Kanchiku T, Taguchi T, Kaneko K, *et al*. A new rabbit model for the study on cervical compressive myelopathy. *J Orthop Res* 2001;19:605–13.
32 Karadimas SK, Gatzounis G, Fehlings MG. Pathobiology of cervical spondylotic myelopathy. *Eur Spine J* 2015;24 Suppl 2:132–8.
33 Karadimas SK, Klironomos G, Papachristou DJ, *et al*. Immunohistochemical profile of NF-κB/p50, NF-κB/p65, MMP-9, MMP-2, and u-PA in experimental cervical spondylotic myelopathy. *Spine (Phila Pa 1976)* 2013;38:4–10.
34 Shimizu K, Nakamura M, Nishikawa Y, *et al*. Spinal kyphosis causes demyelination and neuronal loss in the spinal cord: a new model of kyphotic deformity using juvenile Japanese small game fowls. *Spine (Phila Pa 1976)* 2005;30:2388–92.
35 Wang C, Laiwalla A, Salamon N, *et al*. Compensatory brainstem functional and structural connectivity in patients with degenerative cervical myelopathy by probabilistic tractography and functional MRI. *Brain Res* 2020;1749:147129.
36 Zdunczyk A, Schwarzer V, Mikhailov M, *et al*. The Corticospinal Reserve Capacity: Reorganization of Motor Area and Excitability As a Novel Pathophysiological Concept in Cervical Myelopathy. *Neurosurgery* 2018;83:810–8.

BMJ Neurology Open: first published as 10.1136/bmjno-2024-000940 on 31 January 2025. Downloaded from https://neurologyopen.bmj.com on 1 October 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.