scientific reports

Check for updates

OPEN

A novel open-source ultrasound dataset with deep learning benchmarks for spinal cord injury localization and anatomical segmentation

Avisha Kumar[1]✉, Kunal Kotkar[1], Kelly Jiang[2,4], Meghana Bhimreddy[2,4], Daniel Davidar[2,4], Carly Weber-Levine[2,4], Siddharth Krishnan[2], Max J. Kerensky[2], Ruixing Liang[1], Kelley K. Leadingham[2], Denis Routkevitch[2], Andrew M. Hersh[2], Kimberly Ashayeri[2], Betty Tyler[2], Ian Suk[2], Jennifer Son[3], Nicholas Theodore[2], Nitish Thakor[1,2] & Amir Manbachi[1,2]✉

While deep learning has catalyzed breakthroughs across numerous domains, its broader adoption in clinical settings is inhibited by the costly and time-intensive nature of data acquisition and annotation. To further facilitate medical machine learning, we present an ultrasound dataset of 10,223 brightness-mode (B-mode) images consisting of sagittal slices of porcine spinal cords (N = 25) before and after a contusion injury. We additionally benchmark the performance metrics of several state-of-the-art object detection algorithms to localize the site of injury and semantic segmentation models to label the anatomy for comparison and creation of task-specific architectures. Finally, we evaluate the zero-shot generalization capabilities of the segmentation models on human ultrasound spinal cord images to determine whether training on our porcine dataset is sufficient for accurately interpreting human data. Our results show that the YOLOv8 detection model outperforms all evaluated models for injury localization, achieving a mean Average Precision (mAP50-95) score of 0.606. Segmentation metrics indicate that the DeepLabv3 segmentation model achieves the highest accuracy on unseen porcine anatomy, with a Mean Dice score of 0.587, while SAMed achieves the highest mean Dice score generalizing to human anatomy (0.445). To the best of our knowledge, this is the largest annotated dataset of spinal cord ultrasound images made publicly available to researchers and medical professionals, as well as the first public report of object detection and segmentation architectures to assess anatomical markers in the spinal cord for methodology development and clinical applications.

Keywords Object detection, Semantic segmentation, Medical dataset, Ultrasound, Spinal cord injury

Abbreviations

AR	Average recall
ASPP	Atrous spatial pyramid pooling
B-mode	Brightness-mode
CNN	Convolutional neural network
CPU	Central processing unit
CSF	Cerebrospinal fluid
CT	Computed tomography
CVAT	Computer vision annotation tool
DETR	Detection transformer
DICOM	Digital imaging and communications in medicine

[1]Johns Hopkins University, Baltimore, MD, USA. [2]Johns Hopkins University School of Medicine, Baltimore, MD, USA. [3]Cleveland Clinic, Cleveland, OH, USA. [4]Kelly Jiang, Meghana Bhimreddy, Daniel Davidar and Carly Weber-Levine contributed equally to this work. ✉email: akumar80@jh.edu; amir.manbachi@jhu.edu

E-ELAN	Extended efficient layer aggregation network
FPS	Frames per second
GPU	Graphics processing unit
IoU	Intersection over union
mAP	Mean average precision
MIoU	Mean intersection over union
ML	Machine learning
MRI	Magnetic resonance imaging
PNG	Portable networks graphics
PVCSO	Posterior vertebral column subtraction osteotomy
RCNN	Region-based convolutional neural network
SAMed	Segment anything for medical images
SCI	spinal cord injury
SSD	Single shot detector
YOLO	You only look once

Spinal cord injury (SCI) is a devastating condition affecting an estimated 900,000 people worldwide in 2019 alone, predominantly stemming from motor vehicle accidents and violence[1,2]. The aftermath of SCI extends beyond the acute trauma, as patients experience mobility impairments, severe pain, autonomic dysregulation, and increased risk of mortality if not treated effectively and efficiently[3]. SCI can be divided into 2 phases: the primary injury that results immediately from the traumatic event, and the secondary injury, in which the clinical manifestations of SCI are exacerbated by inflammation and edema, compromising vascular perfusion within the cord[4,5]. The current treatment approach involves a surgical decompression procedure, during which the vertebral structures surrounding the injury site are removed to promote blood flow[6].

Interventions such as cerebrospinal fluid (CSF) drainage, augmentation of mean arterial pressure goals, and electrical stimulation show promise to mitigate the adverse physiological sequelae following SCI and may improve functional outcome[7–10]. However, optimally titrating these therapies remains a challenge due to the lack of real-time and automatic monitoring of spinal cord parameters such as hematoma development and tissue inflammation. Ultrasound is a promising diagnostic tool for SCI management, providing clinicians with real-time, remote, and portable imaging capabilities for visualizing anatomical boundaries and identifying pathological abnormalities within the tissue[11]. Brightness-mode (B-mode) ultrasound emits acoustic waves at frequencies $\geq$ 20 kHz to construct a two-dimensional grayscale image representing the anatomy within the imaging plane[12]. Advancements in ultrafast ultrasound imaging and ultrasound localization microscopy further expand on the diagnostic capabilities of ultrasound by enabling visualization of microvascular structures[13]. Without the acoustic barrier of the vertebral bone after surgical decompression (e.g., laminectomy), clinicians can perform intraoperative imaging of the spinal cord with ultrasound. This information allows surgeons to verify sufficient decompression of the cord and evaluate the need for additional interventions[14].

Compared to other widely accepted imaging modalities, like magnetic resonance imaging (MRI), computed tomography (CT), and conventional radiography (X-ray), ultrasound has benefits for rapid detection of tissue irregularities and injury management due to its safety, cost-effectiveness, and real-time imaging[15]. The severity and location of SCI can be understood by studying the inflammation in the tissue and the development of a hematoma (i.e., pooling of blood in a tissue due to broken blood vessels from the injury). Detailed knowledge of the injury and its evolution can provide valuable insights on treatment efficacy and patient prognosis, as spinal cord compression, swelling, and hemorrhage are associated with poor prognosis of neurologic recovery[16,17]. Furthermore, monitoring the swelling of the key anatomy surrounding the injury provides new avenues for researching the underlying physiological mechanisms of secondary injury. While ultrasound's real-time capabilities make it a valuable tool for continuous tissue monitoring, its utility is often compromised by poor image quality, signal reverberation, artifacts, and speckle. These imaging challenges require the expertise of skilled sonographers or radiologists to ensure accurate interpretation[18,19]. Unfortunately, these image quality issues are a significant contributor to inter-observer variability, which further complicates medical image evaluations. As a result, these factors are key drivers in diagnostic errors underscoring the critical need for medical image algorithms for improving diagnostic results[20,21].

To take full advantage of continuous imaging with ultrasound, several research efforts have explored wearable and implantable ultrasound devices for clinical applications[22–24]. These types of devices are particularly well-suited for long-term, continuous diagnostics during the secondary injury phase of SCI, when real-time monitoring is critical but current imaging practices remain infrequent and limited[6] (Fig. 1). However, continuous evaluation of clinical metrics requires integrated computer vision algorithms capable of extracting relevant anatomical and pathological features from ultrasound images. In this study, we explore the efficacy of deep learning models for automatic hematoma tracking and anatomical segmentation using a B-mode ultrasound dataset of 10,223 images of porcine spinal cords (N = 25). To our knowledge, this is the first open-source dataset to capture the anatomy (dura, CSF, pia, spinal cord, and vertebral structures) of both healthy and injured spinal cords.

We benchmark several state-of-the-art object detection models for localizing the injury site (i.e., hematoma) and evaluate semantic segmentation algorithms for delineating spinal cord anatomy. To assess deployment feasibility, we propose a new metric designed to evaluate how well each model can operate on wearable and implantable ultrasound devices to support continuous monitoring of SCI. Finally, we explore the zero-shot generalization capabilities of these models by evaluating porcine-trained segmentation networks on human spinal cord ultrasound images, highlighting the translational potential for clinical use and future efforts for improving performance.

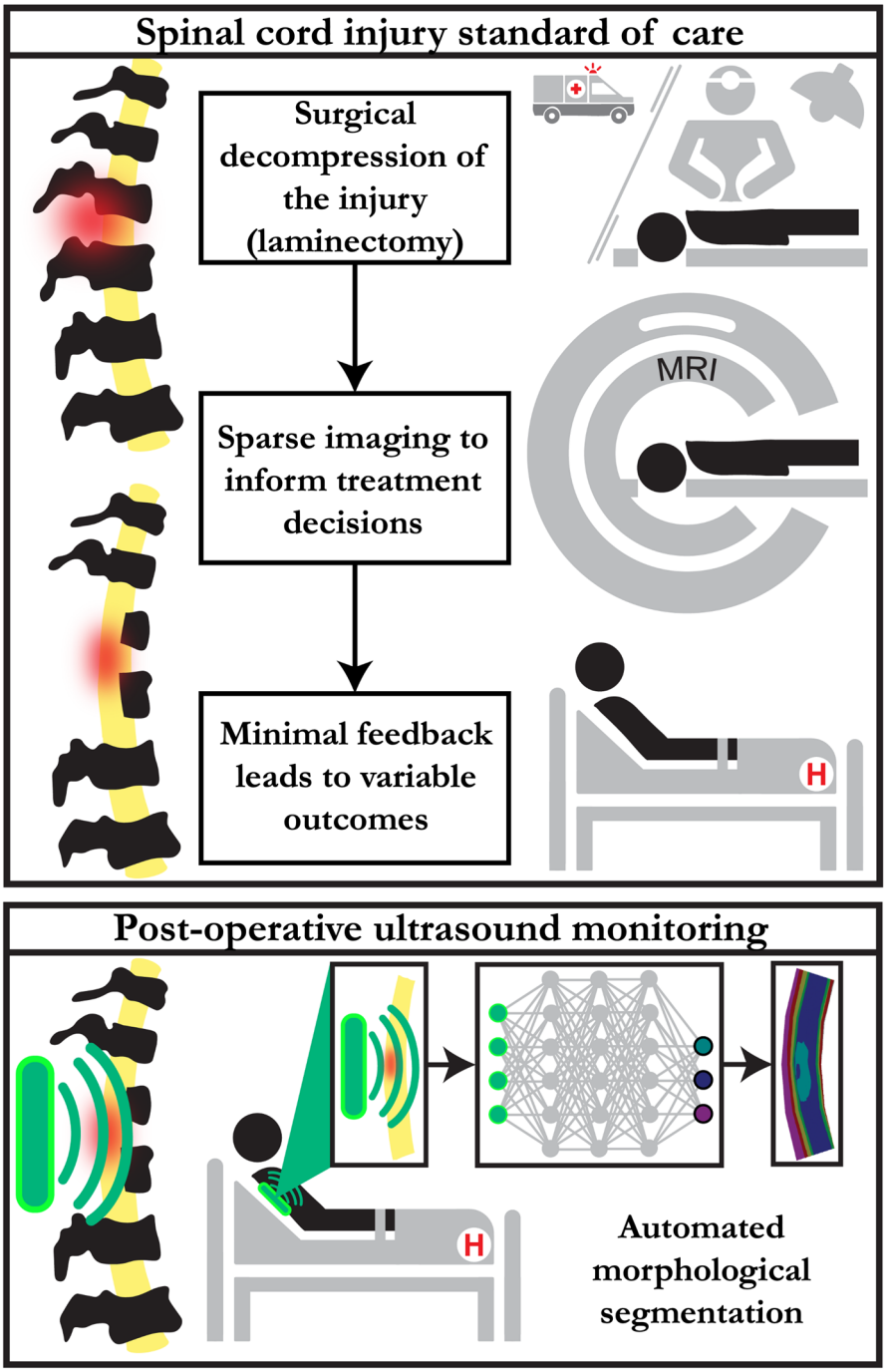

Fig. 1. A comparison of the current standard of care for spinal cord injury with our approach using ultrasound imaging and deep learning for automatic diagnostics. Current treatment approaches do not provide a comprehensive and continuous avenue for monitoring patient spinal cord health after surgical decompression. With automatic injury localization and spinal cord segmentation, clinicians can capture changing biomarkers such as hematoma progression and tissue inflammation to personalize and better understand treatment approach.

Our work represents a pioneering step toward automating continuous diagnostics in SCI, enabling real-time evaluation of morphological features in ultrasound frames to guide diagnosis and uncover clinical insights that were previously inaccessible. With this released dataset, we aim to promote both methodology development for improved computer vision on medical datasets, along with clinical investigations and applications in the domain of SCI.

Previous works

Achieving high accuracy and generalizability in supervised computer vision in spinal cord images typically requires training data that are sufficiently dense and representative of the domain – a costly and time-consuming challenge in the medical field[25]. In contrast to the vast databases of natural images, medical datasets are often scarce and noisy, presenting a substantial challenge for supervised deep learning methods[26–28]. In this section, we highlight some recent efforts for developing medical datasets of spinal cord images.

Prados et al. provided axial MRI images of the cervical spine from 80 healthy subjects across 4 acquisition centers, with annotated white and gray matter segmentation masks[29]. Another published MRI dataset includes spinal cord images from 260 participants to study quantitative metrics (e.g., cross-sectional area of the cord and white/gray matter) of the regions of interest[30]. CTSpine1k provides 1,005 CT volumes with labeled vertebrae for bone segmentation across different spinal cord conditions[31]. However, to our knowledge there are no existing open-source datasets of spinal cord ultrasound images. Unlike CT and MRI, ultrasound has the capability to provide real-time continuous imaging for a unique perspective on spinal cord pulsatility and soft tissue distinction, allowing clinical and surgical decision-making.

There have been several studies which aim to use machine learning (ML) methods to automatically detect injury and segment the spinal cord. Research groups have proposed segment-based classification models[32,33] and hybrid thresholding and convolutional neural network (CNN)-based approaches to identify the severity of SCI and predict disease patterns using the segmented features in an SCI MR image database[34]. Other groups have used deep-learning based techniques for lesion (e.g., cervical disc herniation, traumatic SCI) and compression detection within MR images of the cervical spine[35,36].

Spinal cord segmentation efforts include Paugum et al.'s neural network framework[37] and a spinal cord segmentation challenge to segment white and gray matter in spinal cord structures from axial MRI data[29]. Another CNN-based study proposed a spinal cord segmentation model on 2D axial-view MRI slices from 20 patients (359 images) with cervical spondylotic myelopathy[38].

While there are many computer vision techniques for spinal cord segmentation and disease identification, the published methods at present are limited to MRIs and CTs, which do not provide continuous imaging or detailed soft tissue delineation[39]. As the deep learning field increasingly focuses on video-based understanding, leveraging consecutive frames to enhance network constraints, ultrasound stands out as the optimal modality for ML methods due to its high frame rate for data acquisition. The only other effort for segmentation in ultrasound spinal cord is done by Benjdira et al., where they evaluate models on images (N = 10) post-laminectomy. In their analysis, they tested 3 ML networks on axial slices of the spinal cord to benchmark their performance on outputting a mask of the spinal cord[40]. In our study, we aim to identify all relevant spinal cord anatomy (dura, CSF, pia, spinal cord) along with injury site within high resolution sagittal slices of the cord providing a longitudinal perspective of the spine for detecting conditions that extend over multiple vertebral levels. This dataset is publicly released to further the efforts for medical computer vision and automated injury monitoring.

Methods

Dataset formation

Data acquisition

Over the course of two years (2021–2022), ultrasound data were collected from 25 female Yorkshire pigs weighing approximately 50 pounds. All animal experiments were conducted in compliance with the National Institutes of Health guide for the care and use of laboratory animals (NIH Publications No. 8023, revised 1978). All animal methods were approved by the Johns Hopkins University Animal Care and Use Committee (SW20M221) and were performed following the ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments). To collect the in vivo ultrasound images in the dataset, a laminectomy was performed (Fig. 2a) from the 4th to 6th thoracic vertebrae (T4–T6) to provide an acoustic window (Fig. 2b). Sagittal images were collected at different angles and locations of the spinal cord using a Canon Aplio i800 ultrasound system (Canon Medical Systems, Tustin, CA) connected to either an i22LH8 transducer (operating frequency: 20 MHz) or an i18LX5 transducer (operating frequency: 12 MHz) placed above the spinal cord (Fig. 2c). A controlled injury was subsequently induced with a weight drop (either 20, 40, or 60 grams) from a height of 17 cm, and then images were recollected. Image acquisition parameters, such as spinal cord depth (distance of the spinal cord from the transducer), imaging angle, gain, brightness, and ultrasound signal coupling medium (saline solution, ultrasound gel), were varied to obtain an expressive and diverse dataset. These B-mode ultrasound images were collected by several different people during the data acquisition period and were stored as Digital Imaging and Communications in Medicine (DICOM) files (the international standard for viewing and storing medical imaging information) for further preprocessing to make the final dataset (Fig. 2d).

Additionally, B-mode images of the human spinal cord were captured from 8 patients after a laminectomy was performed from the 11th thoracic level to the 1st lumbar level (T11–L1). Each patient had varying spinal cord curvatures, with 6 undergoing a posterior vertebral column subtraction osteotomy (PVCSO)[41]. Similar to the animal data collection, an i18LX5 transducer connected to a Canon Aplio i800 ultrasound system was placed above the spinal cord and images were collected throughout different stages of the shortening procedure[42]. These images were used to evaluate the semantic segmentation models that were trained on only the porcine spinal cord images. All human data collection methods were approved by the Johns Hopkins Institutional Review Board (IRB00273900). All human data collection methods were performed in accordance with the Declaration of Helsinki and were acquired with informed consent.

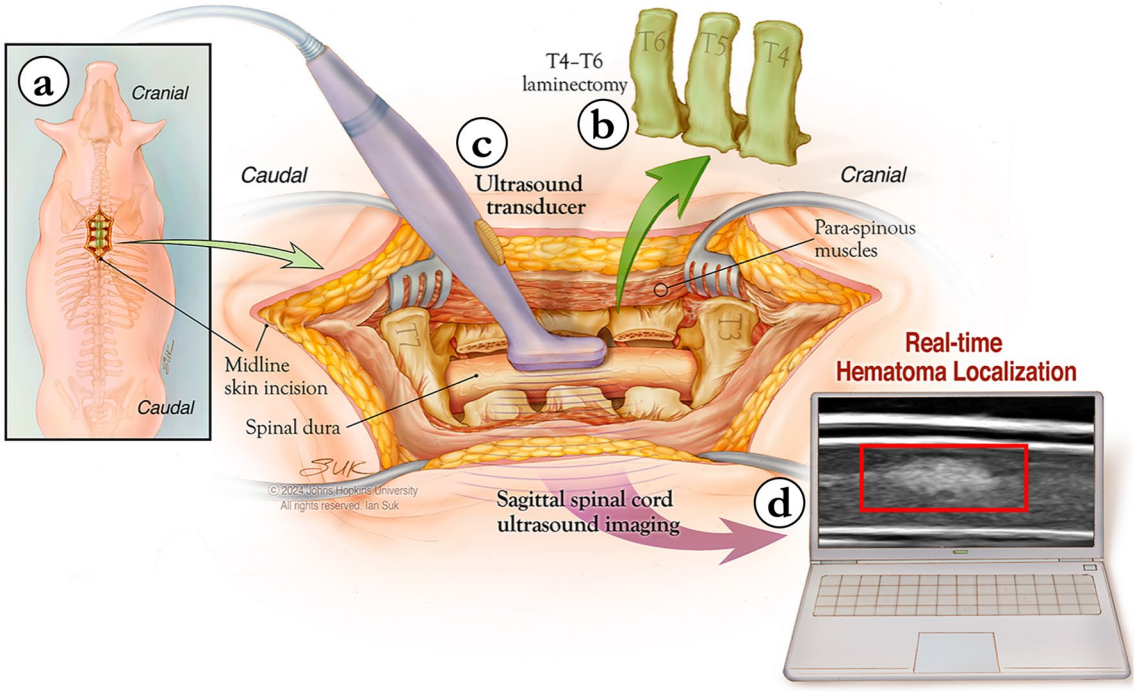

Fig. 2. Data collection of porcine spinal cord. (**a**) An aerial view of the female Yorkshire pig with a laminectomy to expose the spinal cord. (**b**) The spinous processes and lamina of the 4th to 6th thoracic vertebrae (T4–T6) are removed to provide an acoustic window and an injury is induced with a weight drop. (**c**) An i22LH8 transducer connected to Canon Aplio i800 ultrasound system is placed above the spinal cord to capture Brightness-mode (B-mode) images of the region of interest. (**d**) The resulting Digital Imaging and Communications in Medicine (DICOM) image is displayed on a personal computer, showcasing the dura, cerebrospinal fluid (CSF), pia, spinal cord, and the injury location (hematoma). The collected images are included in the final dataset for real-time injury localization and semantic segmentation. This figure was created by Ian Suk.

Data description

There were a total of 3 datasets acquired and created by our group as part of this study for: (1) injury localization in porcine images, (2) semantic segmentation in porcine images, (3) semantic segmentation in human images. After rejecting low quality images in which the spinal cord was occluded or corrupted with noise, the injury localization dataset was developed by randomly selecting 2245 healthy and injured porcine spinal cord images from the collected data (N = 23). The semantic segmentation dataset consists of 10,223 images divided into 4467 healthy spinal cord images (N = 20) and 5756 images of injured spinal cord images (N = 25). To reduce file size, the images (1280 × 960 pixels) were converted to a Portable Network Graphics (PNG) format, scaled to the depth of scan field of the ultrasound image, and cropped to include only the anatomy of interest (690 × 275 pixels) (Fig. 3a–c). After preprocessing, each image shows 25 mm × 8 mm of the sagittal slice of the spinal cord. Finally, a third dataset consisting of 86 human spinal cord ultrasound images (N = 8) was created and preprocessed to develop a unique test set for the semantic segmentation models. While these images did not include a hematoma within the cord, they showed the other anatomy of interest (i.e., dura, CSF, pia, spinal cord), as shown in Fig. 3D. The image statistics of the porcine and human datasets can be found in the Supplementary Materials (Supplementary Table 27, Supplementary Figs. 1 and 2).

Data annotation

The injury localization and anatomical segmentation tasks had different annotation processes. For injury localization, images with a hematoma present were labeled with a bounding box surrounding the injury. The ground truth image masks used for semantic segmentation were created in Computer Vision Annotation Tool (CVAT) (CVAT.ai Corporation, Palo Alto, CA). The anatomical structures of interest were the dura, CSF, pia, spinal cord, and hematoma (Fig. 4A). Additionally, the space above the dorsal dura, where the vertebral bone was removed, was labeled as the dorsal space, and the space below the ventral dura consisting of bony structures and cartilage was labeled as the ventral space. For some injuries, depending on the angle in which the ultrasound images were captured and the impact of the weight drop, the anatomical boundaries between the dura, CSF, and pia were unclear and could not be labeled with sufficient confidence. In order to maintain the integrity of the labels and minimize speculations during the labeling process, those regions were simplified to the dura/pia complex whenever necessary (Fig. 4B). Similarly, the dura/ventral complex label was used when there was ambiguity in the anatomical boundary between the ventral dura and the vertebral structures beneath it (Fig.

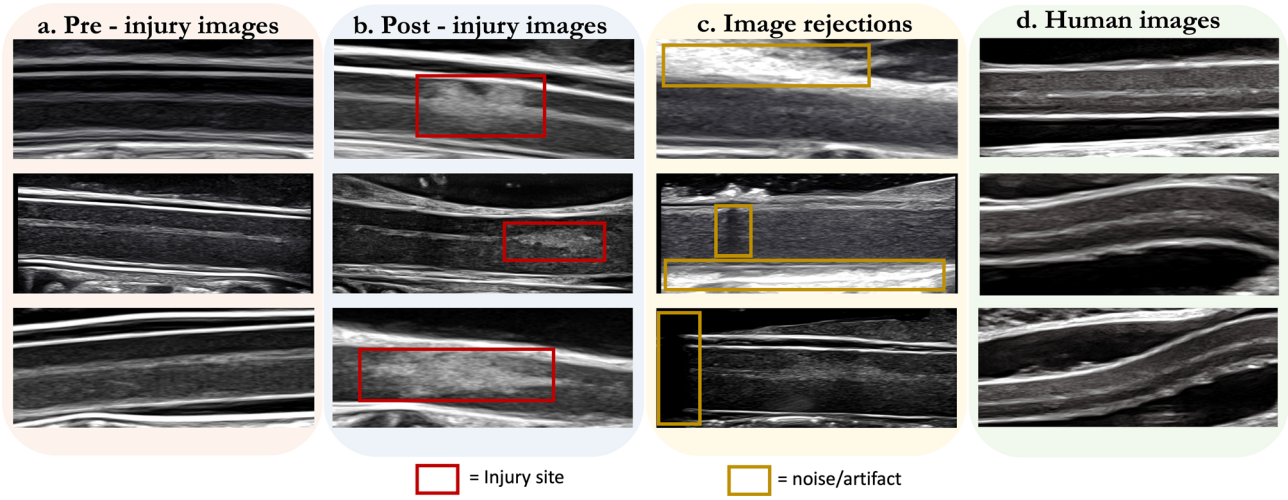

Fig. 3. Curation of the porcine and human ultrasound spinal cord dataset. (**a**) Sample pre-injury images from the porcine spinal cord dataset. These images included the primary anatomy of interest and did not suffer from severe noise or artifacts that would render the image difficult to interpret. (**b**) Sample post-injury images of the dataset that fulfilled the same inclusion criteria as the pre-injury images. The red bounding boxes indicate the hematoma. (**c**) Sample images that were excluded from the final dataset due to shadowing, noise, or artifacts that affect image quality and occlude the anatomy of interest. (**d**) Sample human spinal cord images that were used to test the generalizability of the segmentation models. All images show 25 mm × 8 mm of the spinal cord.

4C). The third dataset of human images was labeled using the same annotation process described above for the porcine spinal cord.

Images were labeled by a team of medical and graduate students trained by a board-certified radiologist. Each image in which there was a lack of clear delineation in anatomical boundaries was verified by the radiologist to ensure high quality and accurate labels. Once all the images were labeled, the masks were then validated by a neurosurgery spine fellow at the Johns Hopkins Hospital to provide a second round of verification and further ensure the robustness of the labels. Table 1 provides an overview of the image and pixel instances for each of these labels. These images and their corresponding annotations are available electronically on our GitHub Repository.

Injury localization

Using the injury localization dataset of 2,245 images, we evaluated the performance of seven state-of-the-art object detection models: Faster RCNN[43], SSD300, SSD512[44], RetinaNet[45], Detection Transformer (DETR)[46], YOLOv7[47], and YOLOv8[48]. Faster RCNN is a two-stage detector where the first stage proposes candidate object bounding boxes with a Region Proposal Network, followed by a second stage that classifies the object and refines its bounding box[43]. Conversely, the SSD model divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell, with SSD300 optimizing for speed with 300 × 300 pixel images, and SSD512 catering to higher-resolution 512 × 512 pixel images[44]. RetinaNet uses a Feature Pyramid Network for building high-level semantic feature maps at multiple scales to efficiently detect objects of various sizes and introduces the Focal Loss function to address the challenge of foreground-background class imbalance[45]. In contrast, DETR uniquely combines a CNN backbone with a transformer encoder-decoder architecture, employing learned positional embeddings and a shared feed-forward network for precise object class and bounding box predictions[46]. YOLOv7 utilizes the Extended Efficient Layer Aggregation Network (E-ELAN) to optimize convergence and network efficiency[47]. Lastly, YOLOv8 utilizes an anchor-free architecture, enhancing detection accuracy across various object sizes and shapes, complemented by a versatile multi-scale prediction method for improved adaptability[48]. More specific details on the models' architectures can be found in their respective references. These models have exhibited high performance when applied to natural image datasets, like COCO[27], PASCAL VOC[28], and ImageNet[26]. The learnable parameters of the model were initialized with pretrained weights and biases using the ImageNet dataset to reduce the training cost.

We also compared the performance of these deep learning models with a traditional image processing approach to gain insight into how heuristic-based methods perform on this computer vision problem. This approach involved preprocessing the ultrasound images using Gaussian blur, Otsu's thresholding, and morphological operations to segment relevant structures, followed by contour detection to generate predicted bounding boxes[49,50]. This algorithm analyzes the distribution of pixel intensities to discriminate regions of interest, which is useful for injury detection in ultrasound images as injuries typically present themselves as hyperechoic regions that appear brighter than surrounding tissues. Effective preprocessing steps, such as Gaussian blur and thresholding, help enhance contrast by suppressing noise, while morphological operations refine the boundaries of the injury region.

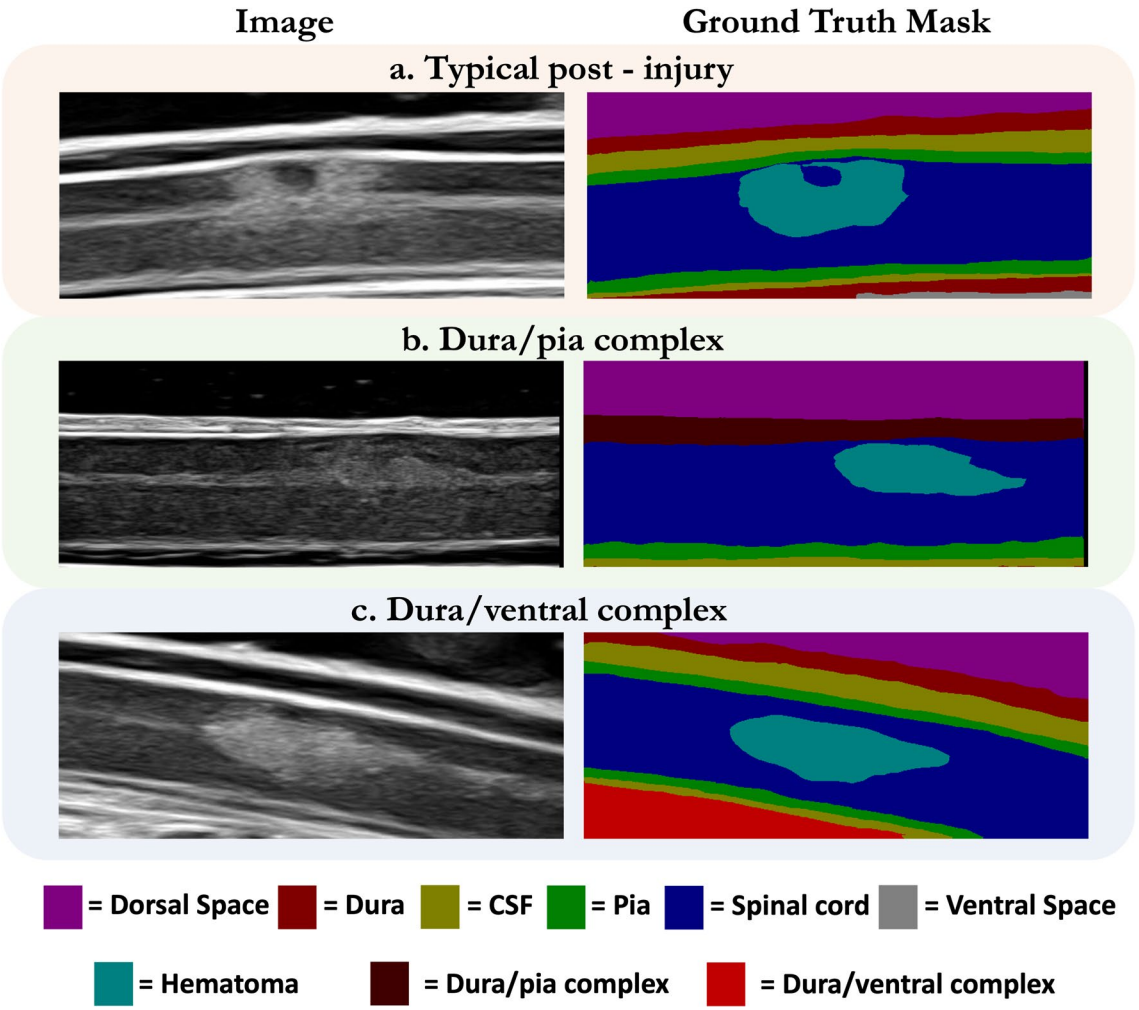

Fig. 4. Example images and their corresponding ground truth masks. (**a**) A typical spinal cord image displaying clear delineation between the anatomical boundaries. The bottom of the image corresponds to the ventral spinal cord. (**b**) A spinal cord image in which the induced injury caused swelling in the tissue, effectively blurring the delineation between the dura, cerebrospinal fluid (CSF), and the pia. To avoid mistakes in interpretation and inconsistencies in labelling, the region is annotated as the dura/pia complex. (**c**) In some images, noise and other intraoperative artifacts resulted in ambiguous delineation between the ventral dura and the ventral region. For these types of images, we label that anatomy as the dura/ventral complex. Each image shows 25 mm $\times$ 8 mm of the spinal cord.

Anatomy	Pixel instances	Instances across images
Dorsal Space	397,350,611	10,223
Dura	142,857,955	9814
Pia	142,721,894	9839
CSF	118,161,757	9613
Dura/Pia complex	57,124,376	2732
Spinal cord	812,894,511	10,223
Hematoma	69,727,644	5756
Dura/Ventral Complex	58,152,053	2671
Ventral Space	89,571,059	6099
Background	18,992,200	6385

Table 1. Pixel- and image-wise instances of each anatomical structure.

The injury localization dataset was split into training, validation, and testing sets by subject, maintaining anatomical independence between splits to prevent data leakage and overfitting (Supplementary Tables 19–23). Roughly 80% of images were used for training, with 10% each for validation and testing, and we balanced the distribution of healthy and injured spinal cord images in each subset. To improve the robustness of the models, we performed a number of data augmentations in the training process. This includes image transformations (e.g., rotating the samples with a degree range of ± 25, horizontal flipping with a probability of 0.5, and randomly zooming in or out of the image by 30%) and image degradation (e.g., blurring the image by a factor of 3 with a probability of 20%). We also adjusted image appearance by randomly manipulating the statistical characteristics of the image intensities such as brightness, saturation, hue, and contrast by a factor of ± 0.2. To optimize computational resources and time, object detection models uniformly resize images to dimensions of 256 × 256 pixels. These models were trained on a Windows 11 Machine (8 GB RAM) with 24 GB NVIDIA GeForce RTX 3090 graphics unit and 14th Gen Intel Core i5-14600 processor (14 cores, 20 threads, 2.7 to 5.2 GHz turbo frequency). The hyperparameters of the model (i.e., batch size, learning rate, training epochs) were tuned to ensure optimal performance using the Neural Network Intelligence software (Supplementary Table 1–7)[51]. The main characteristics of each model are described in Table 2.

Spinal cord segmentation

We evaluated six state-of-the-art semantic segmentation models on our dataset consisting of 10,223 porcine images annotated with masks delineating the anatomical structures. The specific models used were SegFormer[52], U-Net[53], DeepLabv3[54], TransUNet[55], Swin-UNet[56], and Segment Anything for medical images (SAMed)[57]. SegFormer combines a transformer-based encoder to process global information and generate multiscale features, with a straightforward multilayer perceptron (MLP) decoder that integrates these features at various scales[52]. U-Net, renowned for its effectiveness in biomedical imaging, employs a contracting-expanding architecture of the convolutional layers (resulting in a u-shape) that balances spatial and feature information to ensure accurate localization[53]. DeepLabv3 enhances multi-scale context capture through an Atrous Spatial Pyramid Pooling (ASPP) module, which uses dilated convolutions at various scales[54]. TransUNet leverages the global self-attention mechanisms of Transformers to augment the traditional U-Net's convolutional approach, offering improved global information processing and localization[55]. Swin-UNet adopts a Swin Transformer in both its encoder and decoder, utilizing shifted windows to enhance context feature extraction and spatial resolution restoration[56]. Lastly, SAMed builds on the Segment Anything framework, specifically adapting its encoder and introducing a low-rank-based fine-tuning approach for medical image segmentation[57].

We also compared the performance of these segmentation models with traditional image processing methods. This approach involves preprocessing the ultrasound images with Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance contrast followed by Gaussian smoothing to reduce noise, Sobel edge detection to highlight boundaries, Otsu's thresholding to create markers for segmentation, morphological operations like closing and hole filling to refine detected layers, and watershed segmentation to separate anatomical components[49,50,58,59]. These image processing steps are useful for soft tissue segmentation in B-mode sagittal ultrasound images as they enhance contrast, reduce noise/speckle, highlight tissue boundaries, and adaptively threshold regions of interest.

The segmentation dataset was again randomly subdivided into a roughly 80–10–10 split for training, validation, and testing, respectively. All images taken from the same porcine spine were included in a single subset (train, validation, or test) to avoid overlap and prevent artificially inflated results (Supplementary Tables 24–26). Additionally, each subset contained an approximately equal number of healthy and injured spinal cords. The training dataset was augmented to improve the models' generalizability with image transformations (e.g., rotation [± 50-degree limit], horizontal flipping [p = 0.5], image compression [p = 0.2]), and image degradation (random brightness contrast [p = 0.2], random fog [p = 0.2]). For computational and time efficiency, all segmentation models resize the images to 256 × 256 pixels. The hyperparameters (i.e., batch size, learning rate, training epochs) were tuned in the same manner as the object detectors, using the same Windows machine for training (Supplementary Table 8–13)[51]. Table 3 describes the main characteristics of each semantic segmentation model.

Research shows that semantic and structural similarity between source and target domains improves transfer efficiency[60–63]. Because the porcine and human spinal cord have similar anatomical structures and immune

Model	Encoder	# of parameters	Learning rate	Batch size	Training epochs	Optimizer	Loss function
Faster RCNN	ResNet50	41,299,161	0.004935	4	60	SGD	Cross entropy + smooth L1
SSD300	ResNet50	24,641,780	0.002571	32	20	SGD	Cross entropy + smooth L1
SSD512	ResNet50	24,641,780	0.000251	8	20	SGD	Cross entropy + smooth L1
RetinaNet	ResNet50	36,352,630	0.000099	8	60	SGD	Focal loss
DETR	ResNet50	41,524,954	0.000011	8	200	AdamW	Cross entropy + smooth L1
YOLOv7	E-ELAN	37,196,556	0.000123	32	75	Adam	Binary cross entropy + mean square error
YOLOv8	CSPNet	25,856,899	0.000422	48	80	AdamW + SGD	Binary cross entropy + distribution focal loss + complete intersection over union

Table 2. Characteristics of object detection models to detect the site of injury on porcine spinal cord ultrasound images.

Model	Encoder	# of parameters	Learning rate	Batch size	Training epochs	Optimizer	Loss function
SegFormer	MiT-B5	84,601,034	0.000972	4	75	AdamW	Cross entropy
U-Net	ResNet50	31,044,106	0.003011	8	50	SGD	Cross entropy
DeepLabv3	ResNet50	41,998,420	0.000334	4	100	AdamW	Cross entropy
TransUNet	ResNet50 + ViT_B16	105,323,306	0.004468	24	200	SGD	Cross entropy + dice
Swin-UNet	Swin-T	27,153,156	0.061411	16	100	SGD	Cross entropy + dice
SAMed	SAM ViT-B	91,866,903	0.002840	24	200	AdamW	Cross entropy + dice

Table 3. Characteristics of semantic segmentation models for segmenting porcine spinal cord ultrasound images.

Model	mAP50	mAP50-95	AR	CPU			GPU		
				FPS	Load (%)	Implantability score	FPS	Load (%)	Implantability score
Faster RCNN	**0.985**	0.524	0.594	1.53	33	0.676	14.98	62	0.613
SSD300	0.669	0.207	0.249	23.12	36	0.735	**147.62**	27	0.766
SSD512	0.874	0.274	0.324	**23.46**	26	0.866	125.03	18	0.853
RetinaNet	0.912	0.264	0.426	1.89	32	0.646	17.66	48	0.616
DETR	0.787	0.251	0.453	19.78	15	0.812	114.35	26	0.772
YOLOv7	0.923	0.439	0.499	19.54	20	0.865	80.13	28	0.777
YOLOv8	0.979	**0.606**	**0.644**	17.85	22	**0.870**	115.31	27	**0.867**

Table 4. Performance of object detection models on unseen porcine spinal cord ultrasound images to detect the site of injury. Highest performance values are in bold.

response after injury, we evaluated the semantic segmentation models on human spinal cord ultrasound images to understand whether models trained exclusively on porcine images can generalize with sufficient accuracy to human images[64–67]. The annotated dataset containing the 86 human spinal cord images (N = 8 patients) were used as the test set for the trained models to highlight the translatability of our trained models to clinical application.

Results

Injury localization

After hyperparameter tuning, the performance of each model during inference is shown in Table 4, including the mean Average Precision (mAP) value, the Average Recall (AR) value, and speed using frames per second (FPS). mAP is a popular metric ranging from 0 to 1 that measures the accuracy of object detectors as it provides a comprehensive assessment of their performance, considering both precision (how accurate the detected objects are) and recall (how many relevant objects are detected) at multiple levels of confidence (i.e., 50%, and 50–95% at intervals of 5%). The levels of confidence, known as the Intersection over Union (IoU) thresholds, consider a detection accurate if the IoU of the ground truth bounding box and the predicted bounding box is greater than the set threshold (e.g., 0.5 for mAP50). The AR value is the recall metric averaged over a range of IoU thresholds (0.5–1.0). From these results, it is evident that Faster RCNN and YOLOv8 show the strongest performance, achieving a mAP50 score of 0.985 and 0.979, respectively. These models also achieve the highest mAP50-95 score, which is a much more stringent metric for assessing model performance compared to mAP50, at 0.524 for Faster RCNN and 0.606 for YOLOv8. YOLOv8 also attains the highest AR score at 0.644. The traditional image processing approach is evaluated with an average IoU, which is computed as the mean value of IoUs across all predicted and ground truth bounding boxes in the dataset to quantify spatial overlap. On the same test set used to evaluate the deep learning models, this traditional method achieved an average IoU of 0.2034. Given that the IoU threshold for qualifying a prediction as a valid detection in deep learning algorithms is typically 0.5 or higher, this traditional image processing approach exhibits significantly poor performance.

In an effort to better assess the potential of incorporating these models into implantable or wearable devices, we propose a new metric, implantability score, that takes into account the model accuracy, speed, and computational load during inference. We evaluate this implantability score on both the CPU and the GPU with a weighted average of the mAP50 score, normalized average FPS, and average processing load, as shown in Eq. (1).

$$\text{Implantability Score} = \frac{\text{mAP50}}{2} + \frac{\text{FPS}_{\text{norm}}}{4} + \frac{1 - \text{Load}}{4} \tag{1}$$

The mAP50 score is an accuracy metric between 0 and 1. This is weighted twice as much as the other metrics to determine the implantability score as this has a significant impact on the diagnostic capability of the implant. The FPS is normalized to a metric between 0 and 1 by dividing all results with an FPS_{max} value. The FPS_{max} is set to the highest recorded FPS by any model on our CPU (24 FPS). To adapt the FPS normalization process for GPU applications, which generally have much higher speed, the FPS_{max} is set to the highest speed across all

the evaluated models (148 FPS). Finally, because lower computational load is more desirable in this context, we use the inverted average CPU or GPU load (1-Load), which is the percentage increase of computational power during inference, for calculating the score. Our results indicate that YOLOv8 has optimal characteristics for injury localization for automatic monitoring with ultrasound-based implants for both CPU and GPU based applications, with a CPU implantability score of 0.870 and GPU implantability score of 0.867.

Spinal cord segmentation

After hyperparameter tuning, the performance of each segmentation model during inference is shown in Table 5, including the Mean Intersection over Union (MIoU) value, speed, and computational load. MIoU is a measure of the overlap between the predicted segmentation and ground truth segmentation. The IoU for a single class (i.e., each anatomical structure) is calculated as the intersection of the predicted and true positive pixels divided by the union of predicted and true positive pixels and the MIoU is the average IoU across all classes. This provides a more comprehensive understanding of how well the model is performing across different classes. The Mean Dice coefficient, which is another metric between 0 and 1 that indicates the level of similarity between the predicted and ground truth masks, is also evaluated. It is calculated as two times the area of the intersection of the predicted and ground truth masks divided by the sum of the areas of the predicted and ground truth masks.

The class-wise IoU scores and Dice scores are included in the supplementary material (Supplementary Tables 14–15). The metrics are evaluated on both unseen porcine images and human images, which the model had not been exposed to during the training process. To assess the potential for clinical translation, the zero-shot generalizability of these models on human spinal cord images is measured, including the MIoU and Mean Dice scores for the entire anatomy, along with class-wise scores (Supplementary Tables 16–18). The implantable score is again computed using Eq. (1), replacing mAP50 with Mean Dice.

Our results indicate that DeepLabv3 outperforms all other segmentation models in terms of accuracy on porcine anatomy, with a Mean Dice score of 0.587, and SAMed generalizes best to human anatomy, achieving a Mean Dice score of 0.445. TransUNet generalizes best to human spinal cord, with a Dice coefficient of 0.853 for the spinal cord class. Taking accuracy, speed, and computational load into account to determine the model's implantable score (i.e., potential for deployment on an implantable or wearable device), SwinUNet outperforms all other models for CPU-based chips with an implantability score of 0.699 and DeepLabv3 achieves the highest score for GPU-based devices at 0.702. Figure 5 depicts the predicted segmentation masks for each model on an unseen porcine spinal cord image, along with the original ground truth image and mask. We also evaluated the traditional image processing approach for spinal cord segmentation, and found that while this algorithm has reasonable performance on a single image, the hand-crafted rules to create the soft tissue labels do not generalize well across subjects and acquisition angles. We show the qualitative results in Supplementary Figure 3.

Discussion

Deep learning is a promising tool for medical image processing for computer-aided diagnostics in SCI. However, due to the lack of medical image data and their corresponding semantic labels, large-scale computer vision models have been underutilized in this context. With this study, we establish an avenue for automatic monitoring in SCI with our large-scale ultrasound dataset. After benchmarking several object detection and semantic segmentation models, our results show that YOLOv8 and DeepLabv3 are best suited for hematoma tracking and anatomical segmentation, respectively. While it can be challenging to understand the exact reason why some models outperform others, it can be expected that YOLOv8 exhibits the highest performance due to a few distinct characteristics. YOLOv8 incorporates advanced feature aggregation techniques that help in combining features from different scales. This multi-scale approach is particularly useful in medical imaging where the size and shape of the injury can vary significantly. Moreover, YOLOv8 uses mosaic data augmentation, which mixes 4 images together, to provide the model with improved contextual information. Additionally, it is anticipated that YOLOv8's anchor-free detection scheme, in which the model directly predicts an object's mid-point, improves generalization and inference speed for our custom dataset. For semantic segmentation, DeepLabv3 may exhibit the highest accuracy on our dataset because of the atrous convolution approach along with atrous spatial pyramid pooling for capturing multi-scale contextual information without losing resolution. This architecture enables DeepLabv3 to effectively handle boundary information and delineate object boundaries more accurately, which is crucial for our soft tissue segmentation task.

Model	Porcine anatomy		Porcine spinal cord		Human anatomy		Human spinal cord		CPU			GPU		
	MIoU	Dice	IoU	Dice	MIoU	Dice	IoU	Dice	FPS	Load (%)	Implantability score	FPS	Load (%)	Implantability score
SegFormer	0.493	0.570	0.906	0.950	0.232	0.308	0.666	0.773	3.66	23	0.548	23.50	45	0.513
U-Net	0.476	0.553	0.867	0.928	0.253	0.349	0.609	0.722	6.06	32	0.563	62.37	41	0.668
DeepLabv3	**0.515**	**0.587**	0.910	0.952	0.200	0.289	0.506	0.656	4.76	27	0.568	64.11	35	**0.702**
TransUNet	0.500	0.573	**0.921**	**0.958**	0.298	0.388	**0.758**	**0.853**	4.33	35	0.532	40.85	34	0.609
SwinUNet	0.490	0.562	0.913	0.954	0.309	0.401	0.692	0.783	12.75	31	**0.699**	63.36	34	0.690
SAMed	0.497	0.574	0.908	0.951	**0.347**	**0.445**	0.616	0.740	5.40	37	0.535	29.43	35	0.563

Table 5. Performance of semantic segmentation models on unseen porcine and human spinal cord ultrasound images. Highest performance values are in bold.

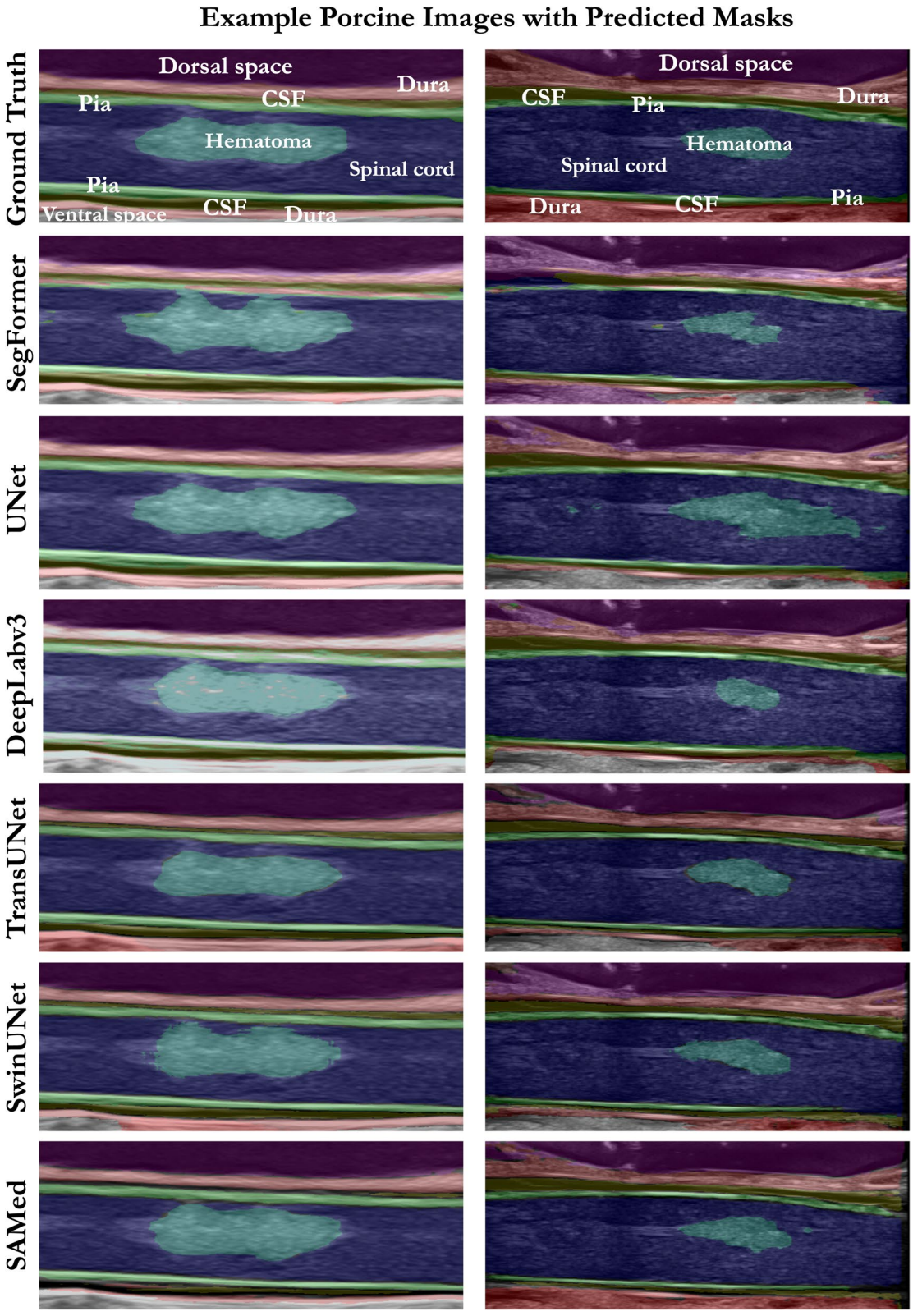

Fig. 5. Visualization of the semantic segmentation models' performance overlaid on example porcine images (25 mm × 8 mm).

Despite this, all evaluated models showed reduced performance on images containing the dura/pia and dura/ventral complexes, likely due to the low frequency of these instances in the dataset. This is quantitatively supported by the class-wise Dice and MIoU scores in Supplementary Tables 14 and 15. However, since spinal cord injury is often characterized by tissue inflammation that leads to CSF disappearance, and because ultrasound image acquisition parameters and angles can obscure anatomical structures farther from the transducer, the inclusion of such images is clinically relevant. Including these challenging cases ensures the dataset reflects the variability and complexity of intraoperative imaging in SCI. To improve classification accuracy on the challenging dura/pia complex and dura/ventral complex regions, future work may benefit from implementing class-specific fine-tuning strategies using hard example mining[68], where model training is progressively focused on underperforming regions. Additionally, some data-level solutions for this class-imbalance can include oversampling rare classes by duplicating or synthesizing images with dura/pia or dura/ventral complexes or performing class-balanced patch sampling[69].

With precise injury localization and anatomical segmentation, deep learning-based computer vision enables tracking of injury progression and tissue inflammation, offering a clearer understanding of patient health trajectories and supporting optimal titration of interventional therapies. Our new dataset and benchmarking results support a new paradigm for computer-aided diagnostics for SCI that address the limitations of the current clinical standard, where anatomical imaging is sparse[6] (Fig. 1). Several academic groups have already demonstrated the feasibility of wearable and implantable ultrasound systems for clinical use[22–24], and there is growing momentum at the intersection of wearable technologies and machine learning, with over 40,000 papers published in 2023 containing the keywords "wearable" and "ML"[70].

ML approaches can analyze long-term patterns from continuous data streams to identify trends that may be difficult for human observers to detect due to the scale and complexity of the information[71,72]. These capabilities also facilitate the integration of ultrasound into robotic and tele-operative systems for remote diagnosis, opening new avenues for enhancing clinical insights and expanding access to care[73]. Additionally, this quantitative approach to image analysis helps mitigate inter-observer variability, a common limitation in clinical ultrasound interpretation. Finally, spinal cord segmentation can support patient-specific computational simulations to explore novel treatment paradigms, such as therapeutic focused ultrasound[74,75].

In our performance comparison with image processing methods, we show that traditional computer vision techniques struggle to localize injury sites and segment soft tissue in ultrasound images. This is primarily due to the inherent noise, speckle artifacts, and low contrast typical of ultrasound imaging, which reduce the effectiveness of thresholding-based methods. Furthermore, the injuries in our dataset exhibit diverse shapes, sizes, and textures across patients, making it difficult for threshold- and contour-based approaches to generalize effectively, as they rely on handcrafted rules that often fail to capture subtle differences between healthy and injured tissues. However, image processing techniques offer advantages in terms of computational efficiency, interpretability, and minimal reliance on extensive training data, making them a valuable option for further exploration, especially in medical applications where human data is often limited and resource-constrained.

It is important to note that there are some limitations to our study which we plan to improve in future works. Firstly, each segmentation model was pretrained on the ImageNet dataset, which contains millions of RGB images, before fine-tuning with our ultrasound dataset. We expect that fine-tuning with these grayscale medical images may cause some performance degradation in the resulting accuracy, which pretraining on a large medical dataset can improve[60,61,63]. With this approach, the underlying learned features can have some foundational similarities with the dataset of interest. One such example of this is the Synapse multi-organ dataset, which contains CT images across several organs[76]. Checkpoints for specific models trained on medical images may also be applicable here[77]. Deep learning algorithms pretrained on ImageNet may learn to make decisions of boundaries between different segmentation regions based on the variance of intensity, which is reasonable in natural images, but less suitable for medical images. This is because medical images are primarily grayscale, which can result in lower intensity variance compared to the rich color information in natural images. The reliability and generalizability of deep learning models are also curtailed by the issue that images were generated by only two different types of transducers at a single collection site (T5), which can be improved with more diverse data collection processes. The accuracy of these models on the test set may also increase with augmentation techniques specific for B-mode ultrasound images that use physics-inspired transformations, including deformation, reverb, and signal-to-noise ratio[78,79]. Future work may also include multiple training runs or bootstrap-based uncertainty estimates, particularly for top-performing models, to better characterize model variability and performance stability, along with model architecture exploration and development tailored to spinal cord-specific tasks using our dataset.

Additionally, the poor zero-shot generalization of these models to human images is influenced by the characteristics and processing of our collected human dataset. While human and porcine spine are morphologically similar, suggesting high generalization accuracy across these species, in the images collected, most patients were undergoing a spinal cord shortening procedure, significantly affecting the curvature of the cord[64,67,80]. Even with these drastic changes to spinal cord geometry, SegFormer, TransUNet, SwinUNet, and SAMed were able to segment the spinal cord with high accuracy (Dice > 0.74). However, because the CSF space is much larger in humans compared to porcine subjects (that were not undergoing PVCSO), after data preprocessing some anatomical classes were not visible in the frame, such as the dorsal space, dura, and ventral space (Fig. 6). This may have a significant impact on the MIoU and Dice scores in Table 5, and should be further explored. Moreover, we had no examples of the hematoma, dura/pia complex, or dura/ventral space classes in this human inference dataset.

Our limited dataset of 86 images is insufficient to fully evaluate generalization performance, especially across a range of spinal cord pathologies or anatomical variations (e.g., age, gender). Future work may involve broader cross-institutional efforts to collect diverse human and porcine spinal cord images across a wider range

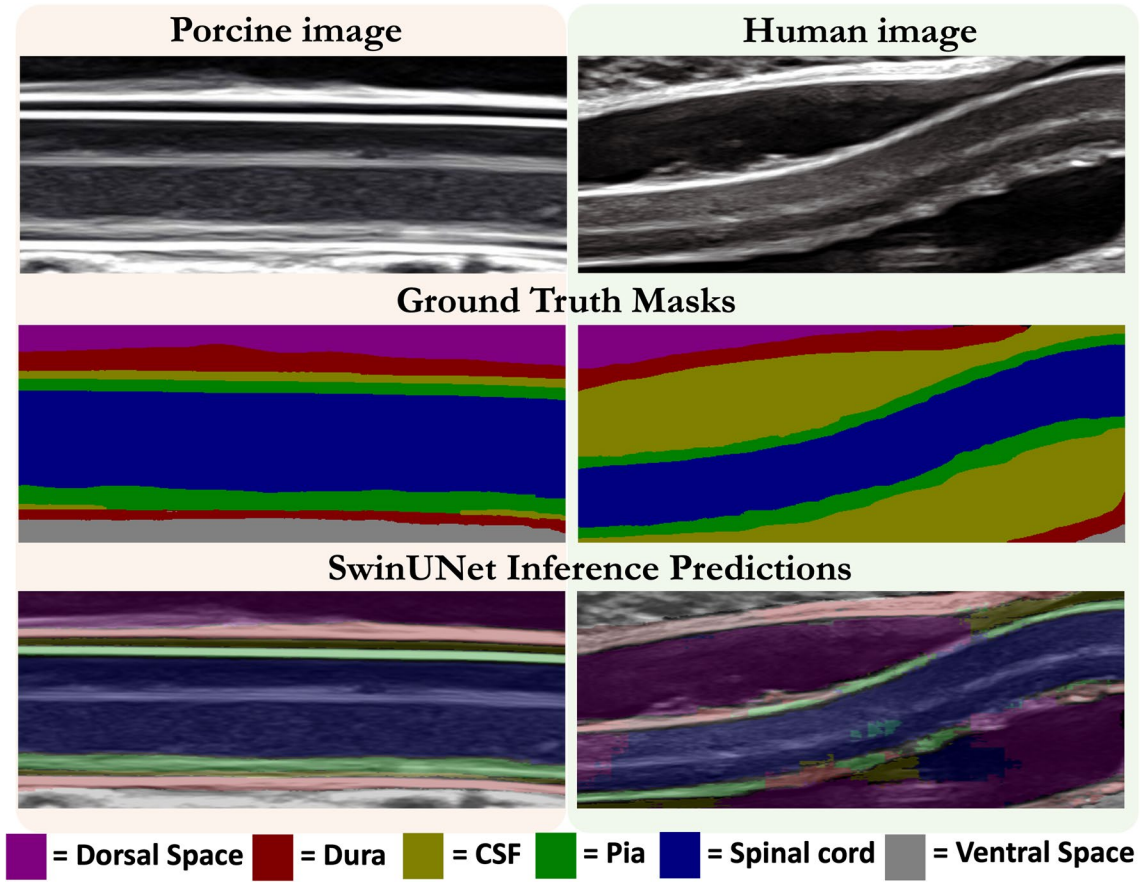

Fig. 6. Visualization of SwinUNET's performance on an example porcine image and example human image (25 mm × 8 mm).

of conditions, allowing for more rigorous evaluation of model generalization. A sufficiently large human dataset could also support fine-tuning of the final layers of deep learning models pretrained on porcine data, improving accuracy on human images while mitigating the need for extensive human data collection to train end-to-end models. Other future efforts to improve human generalization can include transfer learning techniques to adapt the deep learning model trained on porcine data for human applications. Employing domain adaptation techniques, such as matching component analysis, canonical correlation analysis, or optimal transport, can effectively augment porcine data into "synthetic" human data, enhancing the model's ability to generalize across species[81,82]. These approaches would rely on a large set of porcine images and a smaller set of human images that it can map onto, and the morphological similarities in these spinal cord anatomies are favorable as transfer learning typically performs best across domains with related underlying distributions[64,81]. As ultrasound evaluation continues to become the standard-of-care in SCI, there will be increased datasets of human spinal cord images to better inform and train these algorithms[14].

Traditionally, continuous imaging evaluation in SCI has not been feasible. Combining ultrasound's real-time imaging capabilities with deep learning-powered image analysis enables rapid delineation of spinal structures. These models can serve as clinical decision support tools by highlighting subtle abnormalities that may be overlooked in resource-limited settings. Inflammation is a critical biomarker in both acute and chronic SCI. In the acute phase, it contributes to neurotoxicity and expansion of the injury zone and has long been considered an important clinical target to reduce secondary damage and promote maximal recovery[83]. Systemic inflammation also persists chronically after SCI, where it promotes many of the common medical consequences of SCI, including immunosuppression and elevated risk of cardiovascular disease and stroke[80,84,85]. AI-driven tools could support early identification and monitoring of inflammation in spinal cord soft tissue, aiding timely interventions and management.

Conclusion

The deployment of artificial intelligence and computer vision in ultrasound image analysis holds remarkable potential for streamlining diagnostics, with significant enhancements in image quality, diagnostic precision, and accessibility. However, effectively training deep learning models for this purpose presents various challenges, including the need for large, high-quality annotated datasets. We hope that publicly releasing this unique dataset will further facilitate computer vision efforts in medical imaging. By automating diagnostics within ultrasound, clinical workflows for personalized treatment paradigms can be augmented without overburdening clinicians.

The significance of this dataset and research effort lies in ultrasound's unique ability to provide real-time, noninvasive, and continuous imaging at the bedside. With advances in deep learning-based computer vision, this continuous stream of morphological data can be automatically evaluated, enabling a paradigm shift in clinical care by supporting real-time, data-driven insights that were previously unattainable.

Data availability

The authors will maintain this dataset with changes and updates to be described on the following GitHub page: https://github.com/HEPIUSLAB/ultrasound_spinal_cord_dataset.

Received: 29 July 2024; Accepted: 13 August 2025
Published online: 26 September 2025

References

1. Liu, Y. et al. Spinal cord injury: Global burden from 1990 to 2019 and projections up to 2030 using Bayesian age-period-cohort analysis. *Front. Neurol.* **14**, 1304153 (2023).
2. Ding, W. et al. Spinal cord injury: The global incidence, prevalence, and disability from the global burden of disease study 2019. *Spine* **47**, 1532–1540 (2022).
3. Ahuja, C. S. et al. Traumatic spinal cord injury. *Nat. Rev. Dis. Prim.* **3**, 1–21 (2017).
4. Hwang, B. Y. et al. Ultrasound in traumatic spinal cord injury: A wide-open field. *Neurosurgery* **89**, 372–382 (2021).
5. Tsehay, Y. et al. Advances in monitoring for acute spinal cord injury: A narrative review of current literature. *Spine J.* **22**, 1372–1387 (2022).
6. Manbachi, A. et al. Intraoperative ultrasound to monitor spinal cord blood flow after spinal cord injury. In *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Vol. 11317. 58–65 (SPIE, 2020).
7. Martirosyan, N. L. et al. Cerebrospinal fluid drainage and induced hypertension improve spinal cord perfusion after acute spinal cord injury in pigs. *Neurosurgery* **76**, 461–469 (2015).
8. Horn, E. M. et al. The effects of intrathecal hypotension on tissue perfusion and pathophysiological outcome after acute spinal cord injury. *Neurosurg. Focus* **25**, E12 (2008).
9. Schuhmann, M. K., Stoll, G., Bohr, A., Volkmann, J. & Fluri, F. Electrical stimulation of the mesencephalic locomotor region attenuates neuronal loss and cytokine expression in the perifocal region of photothrombotic stroke in rats. *Int. J. Mol. Sci.* **20**, 2341 (2019).
10. Whiting, A. C. et al. Posterior reversible encephalopathic syndrome in the setting of induced elevated mean arterial pressure in patients with spinal cord injury. *Neurosurgery* **83**, 16–21 (2018).
11. Lee, W. & Roh, Y. Ultrasonic transducers for medical diagnostic imaging. *Biomed. Eng. Lett.* **7**, 91–97 (2017).
12. Conlon, T. W. et al. Moving beyond the stethoscope: diagnostic point-of-care ultrasound in pediatric practice. *Pediatrics* **144** (2019).
13. Yu, J., Dong, H., Ta, D., Xie, R. & Xu, K. Super-resolution ultrasound microvascular angiography for spinal cord penumbra imaging. *Ultrasound Med. Biol.* **49**, 2140–2151 (2023).
14. Aarabi, B. et al. Proposal of a management algorithm to predict the need for expansion duraplasty in American Spinal Injury Association impairment scale grades a-c traumatic cervical spinal cord injury patients. *J. Neurotrauma* **39**, 1716–1726 (2022).
15. Kumar, A. et al. Visualizing tactile feedback: An overview of current technologies with a focus on ultrasound elastography. *Front. Med. Technol.* **5** (2023).
16. Miyanji, F., Furlan, J. C., Aarabi, B., Arnold, P. M. & Fehlings, M. G. Acute cervical traumatic spinal cord injury: MR imaging findings correlated with neurologic outcome-prospective study with 100 consecutive patients. *Radiology* **243**, 820–827 (2007).
17. Al-Habib, A. F. et al. Clinical predictors of recovery after blunt spinal cord trauma: Systematic review. *J. Neurotrauma* **28**, 1431–1443 (2011).
18. Ostras, O., Soulioti, D. E. & Pinton, G. Diagnostic ultrasound imaging of the lung: A simulation approach based on propagation and reverberation in the human body. *J. Acoust. Soc. Am.* **150**, 3904–3913 (2021).
19. Long, J., Long, W., Bottenus, N. & Trahey, G. Coherence-based quantification of acoustic clutter sources in medical ultrasound. *J. Acoust. Soc. Am.* **148**, 1051–1062 (2020).
20. Itri, J. N., Tappouni, R. R., McEachern, R. O., Pesch, A. J. & Patel, S. H. Fundamentals of diagnostic error in imaging. *Radiographics* **38**, 1845–1865 (2018).
21. Pinto, A. & Brunese, L. Spectrum of diagnostic errors in radiology. *World J. Radiol.* **2**, 377 (2010).
22. La, T.-G. & Le, L. H. Flexible and wearable ultrasound device for medical applications: A review on materials, structural designs, and current challenges. *Adv. Mater. Technol.* **7**, 2100798 (2022).
23. Hu, H. et al. Stretchable ultrasonic arrays for the three-dimensional mapping of the modulus of deep tissue. *Nat. Biomed. Eng.* 1–14 (2023).
24. Liu, H.-C. et al. Wearable bioadhesive ultrasound shear wave elastography. *Sci. Adv.* **10**, eadk8426 (2024).
25. Tajbakhsh, N. et al. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **63**, 101693 (2020).
26. Deng, J. et al. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255 (IEEE, 2009).
27. Lin, T.-Y. et al. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. 740–755 (Springer, 2014).
28. Everingham, M., Van Gool, L., Williams, C. K., Winn, J. & Zisserman, A. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **88**, 303–338 (2010).
29. Prados, F. et al. Spinal cord grey matter segmentation challenge. *Neuroimage* **152**, 312–329 (2017).
30. Cohen-Adad, J. et al. Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers. *Sci. Data* **8**, 219 (2021).
31. Deng, Y. et al. Ctspine1k: A large-scale dataset for spinal vertebrae segmentation in computed tomography. arXiv preprint arXiv:2105.14711 (2021).
32. Ahammad, S. H., Rajesh, V. & Rahman, M. Z. U. Fast and accurate feature extraction-based segmentation framework for spinal cord injury severity classification. *IEEE Access* **7**, 46092–46103 (2019).
33. Ahammad, S. H., Rahman, M. Z. U., Lay-Ekuakille, A. & Giannoccaro, N. I. An efficient optimal threshold-based segmentation and classification model for multi-level spinal cord injury detection. In *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 1–6 (IEEE, 2020).
34. Ahammad, S. H., Rajesh, V., Rahman, M. Z. U. & Lay-Ekuakille, A. A hybrid cnn-based segmentation and boosting classifier for real time sensor spinal cord injury data. *IEEE Sens. J.* **20**, 10092–10101 (2020).
35. Ma, S., Huang, Y., Che, X. & Gu, R. Faster rcnn-based detection of cervical spinal cord injury and disc degeneration. *J. Appl. Clin. Med. Phys.* **21**, 235–243 (2020).

36. Merali, Z. et al. A deep learning model for detection of cervical spinal cord compression in MRI scans. *Sci. Rep.* **11**, 10473 (2021).
37. Paugam, F. et al. Open-source pipeline for multi-class segmentation of the spinal cord with deep learning. *Magnet. Reson. Imaging* **64**, 21–27 (2019).
38. Zhang, X. et al. Automatic spinal cord segmentation from axial-view MRI slices using cnn with grayscale regularized active contour propagation. *Comput. Biol. Med.* **132**, 104345 (2021).
39. Garg, S. & Bhagyashree, S. Spinal cord MRI segmentation techniques and algorithms: A survey. *SN Comput. Sci.* **2**, 229 (2021).
40. Benjdira, B. et al. Spinal cord segmentation in ultrasound medical imagery. *Appl. Sci.* **10**, 1370 (2020).
41. Theodore, N. et al. Posterior vertebral column subtraction osteotomy for recurrent tethered cord syndrome: A multicenter, retrospective analysis. *Neurosurgery* **88**, 637–647 (2021).
42. Kerensky, M. J. et al. Tethered spinal cord tension assessed via ultrasound elastography in computational and intraoperative human studies. *Commun. Med.* **4**, 4 (2024).
43. Girshick, R. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*. 1440–1448 (2015).
44. Liu, W. et al. SSD: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. 21–37 (Springer, 2016).
45. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*. 2980–2988 (2017).
46. Carion, N. et al. End-to-end object detection with transformers. In *European Conference on Computer Vision*. 213–229 (Springer, 2020).
47. Wang, C.-Y., Bochkovskiy, A. & Liao, H.-Y. M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7464–7475 (2023).
48. Jocher, G., Chaurasia, A. & Qiu, J. Ultralytics YOLO (2023).
49. Bangare, S. L., Dubal, A., Bangare, P. S. & Patil, S. Reviewing Otsu's method for image thresholding. *Int. J. Appl. Eng. Res.* **10**, 21777–21783 (2015).
50. Mukhopadhyay, S. & Chanda, B. Multiscale morphological segmentation of gray-scale images. *IEEE Trans. Image Process.* **12**, 533–549 (2003).
51. Microsoft. Neural Network Intelligence (2021).
52. Xie, E. et al. Segformer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **34**, 12077–12090 (2021).
53. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. 234–241 (Springer, 2015).
54. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587 (2017).
55. Chen, J. et al. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021).
56. Cao, H. et al. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European Conference on Computer Vision*. 205–218 (Springer, 2022).
57. Zhang, K. & Liu, D. Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023).
58. Pisano, E. D. et al. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *J. Digit. Imaging* **11**, 193–200 (1998).
59. Kong, W. et al. Sobel edge detection algorithm with adaptive threshold based on improved genetic algorithm for image processing. *Int. J. Adv. Comput. Sci. Appl.* **14** (2023).
60. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* **27** (2014).
61. Recht, B., Roelofs, R., Schmidt, L. & Shankar, V. Do imagenet classifiers generalize to imagenet? In *International Conference on Machine Learning*. 5389–5400 (PMLR, 2019).
62. Taori, R. et al. Measuring robustness to natural distribution shifts in image classification. *Adv. Neural Inf. Process. Syst.* **33**, 18583–18599 (2020).
63. Miller, J. P. et al. Accuracy on the line: On the strong correlation between out-of-distribution and in-distribution generalization. In *International Conference on Machine Learning*. 7721–7735 (PMLR, 2021).
64. Toossi, A. et al. Comparative neuroanatomy of the lumbosacral spinal cord of the rat, cat, pig, monkey, and human. *Sci. Rep.* **11**, 1955 (2021).
65. Lee, J. H. et al. A novel porcine model of traumatic thoracic spinal cord injury. *J. Neurotrauma* **30**, 142–159 (2013).
66. Kim, K.-T. et al. Review of the ubc porcine model of traumatic spinal cord injury. *J. Korean Neurosurg. Soc.* **61**, 539–547 (2018).
67. Schomberg, D. T. et al. Translational relevance of swine models of spinal cord injury. *J. Neurotrauma* **34**, 541–551 (2017).
68. Shrivastava, A., Gupta, A. & Girshick, R. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 761–769 (2016).
69. Liao, M. et al. Class-balanced sampling and discriminative stylization for domain generalization semantic segmentation. In *IEEE Transactions on Intelligent Transportation Systems* (2024).
70. Olyanasab, A. & Annabestani, M. Leveraging machine learning for personalized wearable biomedical devices: A review. *J. Pers. Med.* **14**, 203 (2024).
71. Benke, K. & Benke, G. Artificial intelligence and big data in public health. *Int. J. Environ. Res. Public Health* **15**, 2796 (2018).
72. Huang, H., Wu, R. S., Lin, M. & Xu, S. *Emerging Wearable Ultrasound Technology* (Ferroelectrics, and Frequency Control, IEEE Transactions on Ultrasonics, 2023).
73. Duan, B., Xiong, L., Guan, X., Fu, Y. & Zhang, Y. Tele-operated robotic ultrasound system for medical diagnosis. *Biomed. Signal Process. Control* **70**, 102900 (2021).
74. Kumar, A. et al. A patient-specific preplanning treatment algorithm for focused ultrasound therapy of spinal cord injury. In *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*. 1–4 (IEEE, 2023).
75. Kumar, A. et al. Computational modeling towards focused ultrasound therapy for spinal cord injury: visualization of beam propagation through patient-specific anatomy. In *Medical Imaging 2023: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 12466. 276–282 (SPIE, 2023).
76. Xu, Z. Multi-atlas labeling beyond the cranial vault-workshop and challenge. In *Synapse Website* (2016).
77. Kang, Q., Gao, J., Li, K. & Lao, Q. Deblurring masked autoencoder is better recipe for ultrasound image recognition. arXiv preprint arXiv:2306.08249 (2023).
78. Tirindelli, M. et al. Rethinking ultrasound augmentation: A physics-inspired approach. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*. 690–700 (Springer, 2021).
79. Al-Dhabyani, W., Gomaa, M., Khaled, H. & Aly, F. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. *Int. J. Adv. Comput. Sci. Appl* **10**, 1–11 (2019).
80. Papatheodorou, A. et al. High-mobility group box 1 (hmgb1) is elevated systemically in persons with acute or chronic traumatic spinal cord injury. *J. Neurotrauma* **34**, 746–754 (2017).

81. Clum, C., Mixon, D. G. & Scarnati, T. Matching component analysis for transfer learning. *SIAM J. Math. Data Sci.* **2**, 309–334 (2020).
82. Andrew, G., Arora, R., Bilmes, J. & Livescu, K. Deep canonical correlation analysis. In *International Conference on Machine Learning*. 1247–1255 (PMLR, 2013).
83. Kwon, B. K. et al. Neurochemical biomarkers in spinal cord injury. *Spinal Cord* **57**, 819–831 (2019).
84. Bloom, O., Herman, P. E. & Spungen, A. M. Systemic inflammation in traumatic spinal cord injury. *Exp. Neurol.* **325**, 113143 (2020).
85. Monahan, R., Stein, A., Gibbs, K., Bank, M. & Bloom, O. Circulating t cell subsets are altered in individuals with chronic spinal cord injury. *Immunol. Res.* **63**, 3–10 (2015).

Acknowledgements

This work received funding support from the National Science Foundation (NSF) STTR Phase 1 Award (#:1938939), Defense Advanced Research Projects Agency (DARPA) Award (#:N660012024075), National Institutes of Health (NIH) awards T32 GM136577 and F30 HL168823, and Johns Hopkins Institute for Clinical and Translational Research (ICTR)'s Clinical Research Scholars Program (KL2), administered by the NIH National Center for Advancing Translational Sciences (NCATS).

Author contributions

AK—project conception, data processing and annotation, deep learning, manuscript preparation, figures. KK—data processing and annotation, deep learning. KJ, MB, DD, CWL, SK, JS, KA—data annotation. MK - data collection, rendered Fig. 6. RL, KKL - manuscript review. AH, DR - data collection. IS—rendered Fig. 1. BT, NT, NT, AM—funding acquisition, supervision.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-16275-z.

Correspondence and requests for materials should be addressed to A.K. or A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025