

Data-driven prediction of spinal cord injury recovery: an exploration of current status and future perspectives

Samuel Håkansson^{1,2,+}, Miklovana Tuci^{1,3,+}, Marc Bolliger³, Armin Curt³, Catherine R. Jutzeler^{1,2,*}, Sarah Brüningk^{1,2,*}

¹ETH Zürich, Department of Health Sciences and Technology (D-HEST), Zürich, Switzerland.

²Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland.

³Spinal Cord Injury Center, University Hospital Balgrist, University of Zürich, Switzerland.

⁺these authors contributed equally to this work.

^{*}shared senior authors.

Abstract

Spinal Cord Injury (SCI) presents a significant challenge in rehabilitation medicine, with recovery outcomes varying widely among individuals. Machine learning (ML) is a promising approach to enhance the prediction of recovery trajectories, but its integration into clinical practice requires a thorough understanding of its efficacy and applicability. We systematically reviewed the current literature on data-driven models of SCI recovery prediction. The included studies were evaluated based on a range of criteria assessing the approach, implementation, input data preferences, and the clinical outcomes aimed to forecast. We observe a tendency to utilize routinely acquired data, such as International Standards for Neurological Classification of SCI (ISNCSCI), imaging, and demographics, for the prediction of functional outcomes derived from the Spinal Cord Independence Measure (SCIM) III and Functional Independence Measure (FIM) scores with a focus on motor ability. Although there has been an increasing interest in data-driven studies over time, traditional machine learning architectures, such as linear regression and tree-based approaches, remained the overwhelmingly popular choices for implementation. This implies ample opportunities for exploring architectures addressing the challenges of predicting SCI recovery, including techniques for learning from limited longitudinal data, improving generalizability, and enhancing reproducibility. We conclude with a perspective, highlighting

possible future directions for data-driven SCI recovery prediction and drawing parallels to other application fields in terms of diverse data types (imaging, tabular, sequential, multimodal), data challenges (limited, missing, longitudinal data), and algorithmic needs (causal inference, robustness).

Introduction

Spinal Cord Injury (SCI) is a complex condition that presents a significant challenge in medical rehabilitation, as it involves a range of physiological response cascades. Beyond the initial physical trauma (e.g. comprising fractures and axonal disruption), the physiological and mechanical damage to spinal tissues triggers neuronal excitotoxicity and increases reactive oxygen concentrations and glutamate levels [1], resulting in neurological dysfunction [2]. As a result, the clinical presentation exhibits extensive heterogeneity in terms of motor abilities (e.g. walking, grasping), sensory function, and various aspects of functionality, including sexual, bowel, and bladder functions. Given the substantial heterogeneity in terms of injury characteristics, clinical presentation, and recovery of bodily functions [3], predicting outcomes at the individual level is challenging. Achieving accurate recovery prediction is, however, essential for establishing realistic expectations among patients and their families regarding long-term prognosis [4]. Moreover, accurate and reliable outcome prediction is crucial for designing and implementing effective clinical trials.

Given the increasing availability of large SCI databases [5–11], it is a logical step to consider data-driven models for recovery prediction. Based on parameterization, these models allow for the establishment of a connection between (multi-modal) input data and outcomes to facilitate predictions. However, it is essential to carefully choose the right model architecture in light of available training data, sample sizes, and clinical questions of interest. For successful clinical deployment, a model must satisfy stringent expectations regarding robustness, generalizability, interpretability, and reproducibility, in addition to meeting performance criteria. Robustness refers to the ability to preserve performance despite conditions such as noisy data and distribution shifts. In general, this requires learning relevant relationships between input data and output labels, rather than overfitting the model to the training data. Generalizability conveys that the method has a high sample size-to-complexity ratio and is externally validated meaning that it performs well on new data, from the same and other origins. In the clinical application context, tracing the prediction’s underlying process is essential to building trust in the model and validating that the decision process is well-founded. Interpretability and explainability analysis facilitate this [12]. Finally, the research and implementation should also be reproducible, meaning that the methods, data processing, and input and output data are transparent and readily usable by others.

This review aims to provide a comprehensive overview of the current state of data-driven applications in the context of SCI recovery prediction. We begin by systematically exploring the diverse landscape within this field, summarizing the various types of input data, outcomes, prediction models, and evaluation methods reported. Transitioning from discussing the status quo and state-of-the-art in the SCI field today,

we provide a perspective regarding promising future directions and inherent challenges. In particular, we stress examples and parallels drawn from other data science applications in healthcare and discuss how concepts and architectural designs from these domains could be adapted for SCI-related tasks.

Methods

Search strategy

We conducted a systematic search across three major bibliographic databases: PubMed, Scopus, and Web of Science. Included in our search were studies published from the inception of these databases up to October 2nd, 2023 (search date). Our search strategy comprised a carefully constructed query for: (“machine learning” OR “statistical model” OR “artificial intelligence” OR “deep learning” OR “decision tree” OR “random forest” OR “prediction rule” OR “prediction model”) AND (“spinal cord injury” OR “spinal cord lesion”) AND (recovery OR prognosis OR prediction OR outcome). Our search was not restricted by language. Additionally, manual searching was also performed, reviewing reference lists of relevant articles and comprehensive reviews.

Inclusion and exclusion criteria

We included peer-reviewed articles focusing on predicting recovery in human adults following acute SCI. Our definition of “recovery” entails predicting motor, sensory, functional, or other clinically important outcomes at a later stage from the acute injury phase. No specific restrictions were imposed regarding the longitudinal timeframe for inclusion. We excluded articles involving pediatric patients or non-human subjects, and those related to non-traumatic or non-ischemic spinal cord injury. Additionally, studies predicting associated or secondary conditions post-injury (e.g. pressure ulcers, surgery-related complications) and those focusing on non-recovery outcomes (like survival, hospital stay length) or investigative treatment effects were excluded. Duplicate studies, abstracts, systematic reviews, meta-analyses, case studies, and validation studies of models already included in this review were also not considered. Reasons for exclusion of studies were defined according to the hierarchy in Table S1.

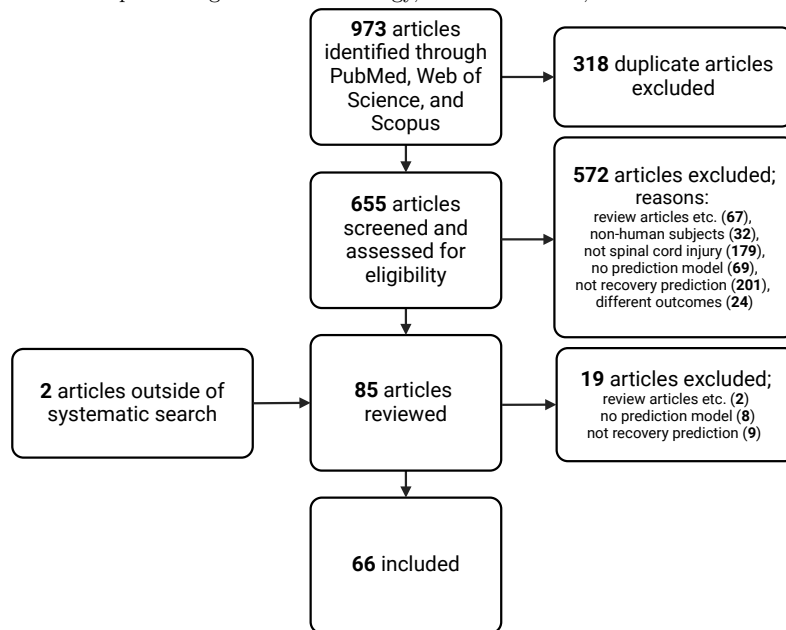
Selection process

We assessed the suitability for inclusion by reviewing titles and abstracts, followed by full-text. This assessment was independently conducted by two reviewers (MT and SH). Conflicting opinions regarding study inclusion or exclusion were settled on a case-by-case basis by mutual discussion with additional investigators (CJ, SB).

Data extraction

Two independent reviewers (MT, SH) performed the data extraction, which was initiated with the systematic retrieval of bibliographic information such as authorship details, publication years, publication titles, and Digital Object Identifiers (DOI).

Fig. 1 Flowchart representing the search strategy, retained studies, and exclusion contributions.



From full texts, we gathered detailed data including: (i) the types of prediction models, (ii) outcomes, (iii) input data, (iv) information about the model architecture(s), implementation, and training, and (v) model performance and metrics used for evaluation.

Qualitative assessment

We evaluated the quality of the included studies based on different criteria spanning five categories: *Clinical Significance*, *Reproducibility*, *Generalizability*, *Machine Learning Quality*, and *Machine Learning Performance*. Each category received a rating from 0 (poor quality) to 5 (high quality) to indicate the quality of the study for the relevant criterion.

To evaluate the *Clinical Significance* of the included studies, we assigned a score of 5 to papers that clearly outlined the clinical problem addressed, employed pertinent outcome measures, and emphasized clinical impact. Conversely, studies lacking a clear motivation for the clinical application of the presented approach, such as purely computational studies, were assigned a score of 0.

Reproducibility refers to the ability to replicate the findings of a study by following the outlined methods as well as using the provided data and code. Ideally, a study provides clear input and output variables, a description of variable coding for input (binary, categorical, or numerical) and output (binary or multi-class classification, or regression), and an overview of the employed model architecture in addition to details regarding its implementation. Both model code (e.g. as a public repository) and

supporting data should be available for external applications. We subtracted scoring points for any missing information.

Understanding the extent to which findings and data-driven models can be generalized across different patient populations, injury severities, or clinical settings is key to examining the potential for clinical deployment. Thus, we assessed the *Generalizability* of the selected studies as follows: a score of 5 was assigned if the risk of overfitting to training data was explicitly addressed. This was considered accomplished through either external validation or by performing a clean data split (with no apparent information leakage) between the training and test set. To achieve an optimal score, we further considered the model complexity with respect to the sample size, and if the utilized data were multi-institutional. We scored 0 when the model was deemed susceptible to overfitting.

Under the umbrella of *Machine Learning Quality*, aspects regarding model architecture, training, and evaluation choices were assessed. We assigned a *Machine Learning Quality* score of 5 to papers that comprehensively reported their entire training pipeline, addressed issues of overfitting and imbalanced data, included strategies for feature selection (where appropriate), and chose relevant evaluation metrics. Ideally, a benchmark of different architectures was performed. We scored a 0 if none of the above criteria were met.

Machine Learning Performance quantifies how well a model accomplishes the task it was designed for. Given the strong effect of the *Machine Learning Quality* on the *Performance*, we fixed the *Machine Learning Quality* score as the upper limit for the performance score. Excellent performance was quantified as an accuracy (or ROCAUC) exceeding 90% or a correlation coefficient of 0.8 while accounting for possible class imbalance. We assigned a score of 0 to articles when the prediction quality was not assessed or did not exceed random prediction. Intermediate scores were given depending on the ratio between excellent and achieved performance.

Statistical analysis

Based on the above scoring metrics, we summarize the included studies by the evaluation criteria using boxplots. Moreover, we identified groups of studies by hierarchical clustering. Clustering was performed such that the cophenetic distance (the height of the dendrogram at the point where the two points are first joined together), between two observations is minimal, with a maximum of 6 flat clusters. The Python library SciPy [13] with the function `fcluster` in the subpackage `cluster.hierarchy` was used to generate the clusters and the hierarchical linkage was generated using `linkage` from the same subpackage.

Results

Study selection

Figure 1 shows a Consort diagram leading to the final set of included studies. Our search identified 973 studies, after the removal of 318 duplicates, 655 article abstracts were screened, resulting in the exclusion of 572 articles. The subsequent review of

83 full-text articles resulted in the exclusion of 19 more articles. Finally, 64 studies met our inclusion criteria. Moreover, one article was added outside of the literature search through the references in one of the included papers (a validation study was exchanged for the study where the method was first suggested), and one article was added manually, resulting in 66 manuscripts [14–79] .

Study characteristics

An overview of the extracted information from all included studies is available in supplementary Table S2.

Input and output data

A range of input and output data types were addressed in the covered studies as shown in Table 1 and Table 2. Patient demographics and injury severity, as measured by the American Spinal Injury Association (ASIA) Impairment Scale (AIS), were identified as primary input factors in 56 (85%) and 47 (71%) papers, respectively.

These data types were followed by the neurological level of injury (NLI, 51.5%); subsets of the ISNCSCI (e.g. pin prick scores, or selected myotomes, 31.8%); and the sum scores for upper extremity (UEMS), lower extremity (LEMS), or total motor scores (TMS) in 17 papers (25.8%).

Imaging data was also a popular input, appearing in 16 papers. “Imaging data” here refers to relevant modalities (e.g. MRI), but also extracted information from images, such as Brain and Spinal Injury Center (BASIC) scores [80]. Only two papers used algorithmic methods to process images rather than pre-extracted image features. Functional input data types included the Functional Independence Measure (FIM, 18.2%), Manual Muscle Tests (MRC, 7.6%), and the Graded Redefined Assessment of Strength, Sensibility, and Prehension (GRASSP, 3.0%).

Eight studies used molecular or hematological information such as blood markers [22] or proteomic analysis of cerebrospinal fluid (CSF) [71].

The specific endpoints addressed covered a variety of scenarios as shown in Table 2. Primarily, functional and neurological outcomes were predicted. The summarized functional endpoints consist of Spinal Cord Independence Measure (SCIM) III [81] (15.2%), 6-minute walk test [82] (16.7%), Functional Independence Measure (FIM) [83] (13.6%), Prehension Performance [84] (1.5%), bladder function (e.g. [81]) (4.6%), and bowel function (e.g. [81]) (3.1%). 34 out of 66 studies predicted neurological outputs, such as AIS grade conversion (31.8%), composite motor scores (LEMS, UEMS or TMS, 7.6%), 2-motor level improvement on the cervical myotomes on either left or right side [46] (1.5%), Percent Deficit Improvement [28] (PDI, 1.5%), which quantifies neurological recovery as the percentage improvement in motor, pin, and touch scores relative to the maximum possible improvement from the baseline to the endpoint score, and pin prick (1.5%) [66].

Moreover, the timing of the data acquisition and prediction endpoints relative to the time of injury were recorded (Table 3). Multiple studies (34) with clearly defined time points since injury considered input data at a maximum of 56 days (i.e. 8 weeks) after injury. One study used input data from between 3 weeks and 3 months [68]. In

Category	Input Data types	N of Papers	%
Neurological	AIS	47	71.2%
	NLI	34	51.5%
	ISNCSCI	21	31.8%
	UEMS/LEMS/TMS	17	25.8%
	VAC/DAP	1	1.5%
	Pin prick	9	13.6%
Functional	FIM	12	18.2%
	MRC	5	7.6%
	GRASSP/MP	2	3.0%
Imaging	MRI features	12	18.2%
	MRI images	4	6.1%
Other	Demographics	56	84.8%
	Comorbidities	14	21.2%
	CSF/Blood	8	12.1%

Table 1 Overview of Input data types. Note that multiple labels can be assigned to one study. NLI: Neurological Level of Injury. AIS: American Spinal Injury Association Impairment Scale. ISNCSCI: International Standards for Neurological Classification of SCI. UEMS: Upper Extremity Motor Score. LEMS: Lower Extremity Motor Score. TMS: Total Motor Score. VAC: Voluntary Anal Contraction. DAP: Deep Anal Pressure. FIM: Functional independence motor score. MRC: Manual Muscle Test. GRASSP: Graded and Redefined Assessment of Strength, Sensibility and Prehension. MP: Motor Point mapping.

the remaining studies, the specific timing of input data acquisition was either unclear (22 studies) or not reported (9 studies). For outcomes, 21 out of 42 studies with reported time points predicted at up to 1 year post-injury. Among these, 12 studies reported 6 months as their prediction time, and 3 studies used an evaluation at 3 months post-injury.

Moreover, six studies reported their prediction time to be > 1 year and two of these studies predicted outcomes at multiple times, ~ 1 st year, ~ 2 nd year, or ~ 5 th year [52, 64]. Similar to input data, several studies ($n=17$) did not specify exact time stamps, but defined categories such as: “at discharge”, or the number of training sessions [54]. Seven studies did not report their prediction time, impeding reproducibility.

Prediction model architecture choices

The included studies encompass various computational architectures, ranging from traditional statistical models to advanced designs incorporating model ensembles and deep learning frameworks. Linear and logistic regression are recurrent choices (50.7%), leveraging their interpretability and user-friendly implementation. Tree-based models and ensembles, including random forest or extreme gradient boosting (XGBoost), were also popular (29%). This showcases the flexibility of tree-based structures to capture complex relationships within the data while being easily implementable given established libraries (e.g. using *scikit-learn* library [85]). Deep learning techniques such as neural networks were reported (9%) but the deliberate motivation for the choice of more complex architectures was rarely reported. Figure 2C summarizes the distribution of the different model architectures over the time of publication. We observed a trend toward alternatives to linear models in recent years.

Category	Output Data types	N of Papers	%
Neurological	AIS	21	31.8%
	ISNCSCI	5	7.6%
	UEMS/LEMS/TMS	5	7.6%
	Contracture*	2	3.0%
	Pain type	1	1.5%
Functional	Walking	11	16.7%
	SCIM III	10	15.2%
	FIM	9	13.6%
	Bowel/Bladder	5	7.6%
	Prehension	1	1.5%
Other	Discharge destination	1	1.5%
	Bone healing**	2	3.0%
	Unspecified***	2	3.0%

Table 2 Overview of output data types. Note that multiple labels can be assigned to one study. SCIM III: Spinal Cord Independence Measure.

*Contracture: Spasticity and Contracture.

**Bone healing: Odontoid fracture nonunion, Bone Mineral Density and Bone Mineral Content.

*** Unspecified: E.g. Soft tissues, Short Form 12-Questionnaire Health Survey.

Prediction Time	N papers
> 1 year	6
~1 year	21
~6 months	12
~3 months	3
Unspecific	17
Not reported	7
Inputs	N papers
< 72 h	6
< 31 days	19
< 56 days	9
3 weeks - 3 months	1
Unspecific	22
Not reported	9

Table 3 Summary of the time points for included input data and evaluation endpoints. Unspecific input time: admission time, acute phase, baseline, and rehab admission. Unspecific output time: at discharge (unknown if this refers to the primary care or rehabilitation facility), after 36 training sessions (unknown absolute time since injury), rehabilitation discharge (unknown absolute time since injury).

In line with the focus on computational architectures comprising limited parameterization, the majority (74.2%) of studies were conducted in small patient cohorts of less than 500 subjects. Despite the large heterogeneity expected in SCI patients, few studies [15, 33, 42, 62] used larger sample sizes beyond 2000 subjects. It is worth noting the study by Kapoor D., et al. [33], comprising 20790 individuals. It benchmarked different machine learning models (ridge classifier, support vector machine,

elastic net, logistic regression, ensemble model, convolutional neural network, random forest, and naive Bayes) on data from the National Spinal Cord Injury Statistical Center (NSCISC) database to predict AIS scores at hospital discharge.

Qualitative assessment of studies

We summarize the qualitative results obtained from scoring different aspects relevant to a data-driven analysis in Figure 3; individual scores are provided in Table S4.

We observe that the majority of papers scored consistently high in clinical significance (median 4.5, minimum 3.5 (excluding three outliers), Figure 3A). The outliers, scoring 0.5, 3 and 3 [19, 37, 40], were computationally focused but lacked a clear definition of the clinical problem they were addressing.

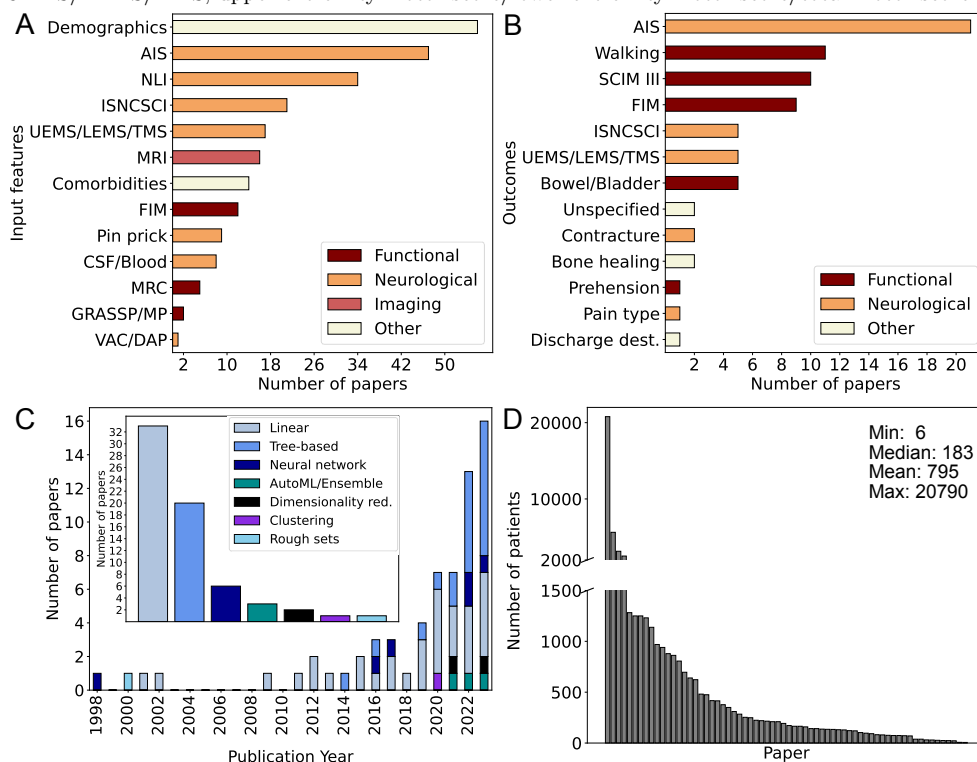
A total of 83% of the papers achieved a *Reproducibility* score of at most 3.5 (Figure 3A). Higher scores (> 3.5) were predominantly related to making the data available (on request) ($n=11$) and/or the code base available ($n=4$). Three percent of the studies did not provide access to code and had unclear definitions of the model inputs and outcomes, resulting in a reproducibility score below 2.5.

For the other evaluation criteria (*Generalizability*, *ML Quality*, and *Performance*), we observed varied results across the included articles, with a median score of 3 in all categories, reflecting a broad range of scores (Figure 3A). Many studies ($n = 34$) addressed the problem of generalization by either external validation or performing a clean train-test data split.

Figure 3B groups all 66 studies using hierarchical clustering. Except for one study [40] that failed to meet both *Clinical Significance* and model *Reproducibility* (cluster 6) and one study that had great *Clinical Significance* and *Reproducibility* but lower grades in the rest of the categories (cluster 3), our clustering revealed four sets of comparable study designs: i) those achieving adequate (score above 3) ratings across all criteria ($n=26$, cluster 5), ii) those yielding intermediate (≈ 3) model *Generalizability*, *ML Quality* and *ML Performance* ($n = 20$, cluster 4), iii) those failing to address the latter three criteria ($n = 16$, *Generalizability*, *ML Quality* and *ML Performance* below 1, cluster 1), and iv) studies of poor generalizability and ML quality but having good *Reproducibility* (cluster 2).

Regarding the longitudinal development across publication years (Figure 3C-E), we did not observe a clear trend of quality improvement despite the noticeable increase in the number of publications focusing on applying data-driven approaches to SCI-related prediction tasks (Figure 3D).

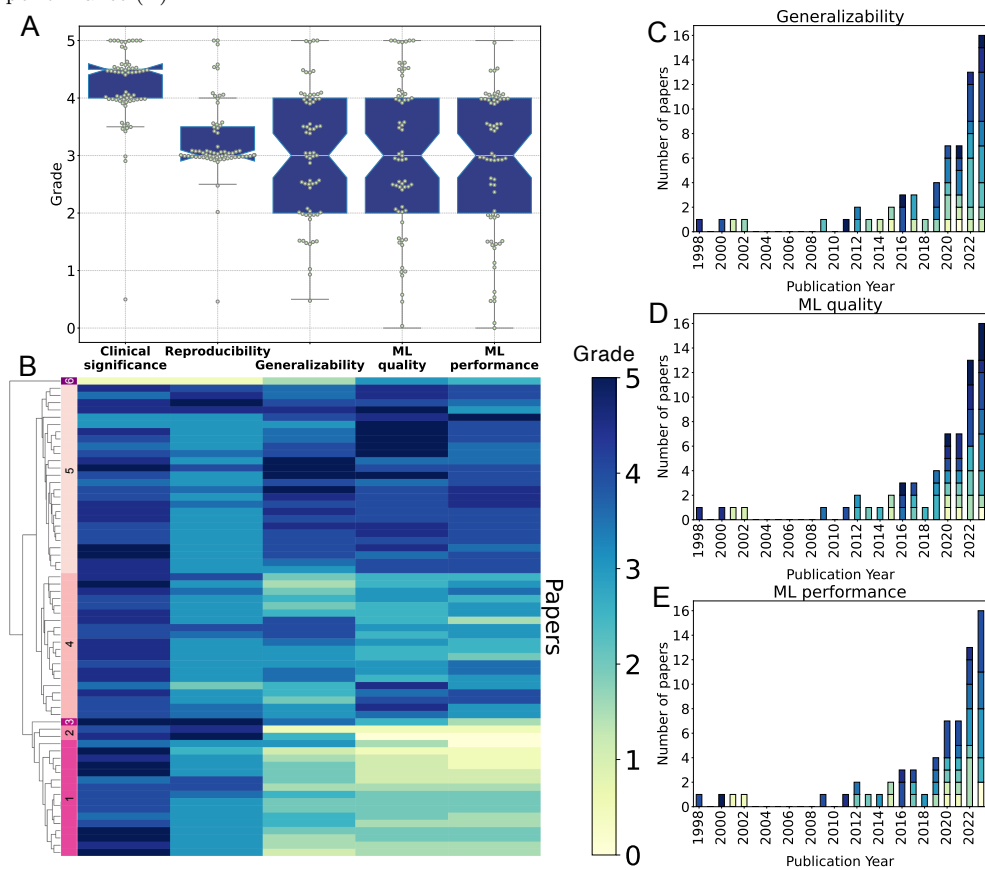
Fig. 2 Overview of the models, inputs, outcomes, and number of patients used in the selected papers. **A) Input features.** Demographic data was used in almost all papers. AIS grade, neurological level of injury (NLI), and neurological assessment data were common. Note that each study could utilize multiple outcomes and input features. **B) Outcomes.** AIS grade was the most common outcome followed by walking ability, FIM, and SCIM III. “Bone healing” refers to bone density and nonunion odontoid fractures. Additionally, “Bowel/Bladder” covers bowel management, bladder recovery, urinary continence, and urodynamic risk factors for upper urinary tract damage. **C) Number of publications per year.** The outer plot shows the increase in publications regarding recovery prediction models in SCI as well as the distribution of model classes. Linear and tree-based models are the most common models. **D) Sample size of analyzed cohorts.** While some papers featured large patient cohorts ranging from 2,000 to 20,000 patients, the majority of papers examined smaller patient populations. Abbreviations: AIS; ASIA Impairment Scale, CSF; Cerebrospinal fluid, FIM; Functional Independence Measure, GRASSP; Graded Redefined Assessment of Strength, Sensation, and Prehension, ISNCSCI; International Standards for Neurological Classification of SCI, MP; motor point, MRC; Medical Research Council manual muscle testing, MRI; magnetic resonance imaging, NLI; neurological level of injury, SCIM III; Spinal Cord Independence Measure III, UEMS/LEMS/TMS; upper extremity motor score/lower extremity motor score/total motor score.



Discussion

Our literature review identified 66 studies addressing the topic of data-driven prediction of recovery after traumatic SCI. These studies employed a range of computational models including linear regression, tree-based models, and neural networks. The choice

Fig. 3 Overview of the grading of clinical relevance and machine learning categories of the included articles. **A) Scoring performance overview** stratified by the five assessment categories. Scores between 0 to 5 with intervals of 0.5 are possible. The boxplot illustrates the median (light blue line) and quartile ranges. The whiskers extend to the farthest data point within 1.5x of the relevant interquartile range. Individual scores are shown with jitter. **B) Heatmap overview of the scores** achieved by each study (row) subject to hierarchical clustering as visualized by the dendrogram. The hierarchy is flattened into 6 clusters indicated by color. **C-E) Visualization of scores as a function of time** for the generalizability (C), machine learning quality (D), and machine learning performance (E)



of model architectures appeared largely influenced by data constraints and the need for reproducible and interpretable implementations inherent in standard machine learning models. Still, a recent trend toward more complex approaches is observed. Data availability also strongly influenced the choice of input features and outcome variables. Predominantly, easily accessible data such as patient demographics and neurological assessments were favored in prediction models. Studies incorporating imaging, functional (e.g. walking ability), molecular, or hematological data as input features were less common. The most commonly used outcomes were AIS grades, walking ability, or

SCIM III. Despite the acknowledged importance of sexual, bowel, and bladder functions for individuals living with SCI [86], these endpoints are understudied, likely due to the lack of available data on these specific aspects. Whereas data on sexual function is particularly limited, bladder and bowel functions are part of the SCIM III assessment. The SCIM III is frequently assessed as part of SCI clinical trials [87–90] and observational cohorts [5–7, 81].

A striking observation was the notable variability and frequently inadequate precision in defining and reporting time points for input features (up to approximately 56 days) and outcome variables (typically around 1 year). The choice of outcome time-point (≈ 1 year) is clinically reasonable, considering the limited remaining recovery potential beyond this timeframe [91].

However, our evaluation also highlights the necessity for improved reporting practices to enhance reproducibility and applicability. In recent decades, the field has established several large-scale observational data registries (e.g. EMSCI, RHSCIR) that address these gaps.

The qualitative assessment revealed consistently high *Clinical significance* scores across all articles, indicating that SCI recovery prediction is driven by clinical application, facilitated by close collaboration between clinical data providers and computational researchers. In contrast, the *Reproducibility* was observed to be relatively low, leaving ample room for improvement. Specifically, while input, output, and chosen models were reported, the provision of code — which is essential for study reproducibility — was rare, yet indispensable upon the publication of data-driven models [92]. We also observed substantial variability in the scoring of *Generalizability*. In general, the importance of *Generalizability* is acknowledged in the field, however, its implementation is often suboptimal. While employing an external validation set would be ideal, adopting a rigorous train-test split can serve as an initial step toward more genuinely externally validated approaches.

In summary, there is a noticeable surge in interest in data-driven SCI recovery predictions, which has inspired a range of promising predictive studies of high clinical relevance. As this field evolves, it is crucial to effectively utilize and report data-driven prediction models to maximize their potential for clinical deployment. Based on our quality assessment, there is room for improvement in terms of reporting and defining outcome variables, as well as ensuring reproducibility and generalizability of the employed models. In the following, we outline possible future directions and key challenges, mention parallels to other clinical applications regarding data types, and outline aspects of promising model architectures and training regimes.

Future directions and key challenges of data-driven recovery prediction

Data availability and diversity remain a limitation for data-driven predictions of SCI recovery. Data-efficient training strategies and optimal use of the available data address this challenge. Missing and incomplete data, particularly in a longitudinal context, is common in biomedical data science. In the SCI field, missing data predominantly arise due to patient transfers between treatment and rehabilitation facilities, comorbidities,

and patient compliance with elaborate assessment protocols such as SCIM and ISNC-SCI [93]. It is widely accepted to rely on complete-case analysis rather than imputed data [93–95]. Carefully applied imputation could, however, increase the pool of training samples and lead to more robust predictions. Bourguignon et al. provide a comprehensive evaluation of diverse imputation strategies in the SCI domain to achieve this [93]. Importantly, depending on the target data type and clinical context, different imputation strategies should be applied (e.g. ordinal imputation [96], last observation carried forward [97], or multiple imputation [98]). In addition to maximizing data usage, strategies from limited data learning, such as auxiliary task learning (training for a subtask that is indirectly linked to the objective of interest) [99], self-supervision (contrasting sample subgroups) [100] and data augmentation can be harnessed.

Traditional data augmentation aims to enhance the performance through various transformations, such as geometric transformations and color space augmentations in imaging [101], aiming to increase the diversity and quantity of the data. Augmenting datasets with limited labels using semi-supervised [102] and active learning [103] could further improve model performance through strategic data labeling and the incorporation of unlabeled datasets [104, 105]. Moreover, data from an external domain might prove useful in settings with limited data. Transfer learning is a compelling strategy that leverages a large dataset from a source domain to (pre-)fit model weights, leading to improved performance upon the target domain upon model refinement. This approach has previously proven successful in related applications such as neurodegenerative disorders [106]. While general data sets are useful, data that are related to the target task (e.g. same data type, same anatomical context) would be preferable [107]. While ISNCSCI data may be unique, parallels for spinal imaging and (multi-) omics data in SCI-related tasks may be readily accessible.

A common limitation to increasing sample size may be data privacy concerns restricting large-scale data centralization. Federated learning represents an alternative way of collaborative and privacy-preserving multisource data digestion [108]. This is especially relevant in scenarios where direct data sharing among institutions is impractical [109]. In the SCI recovery prediction field, techniques for limited data learning combined with a federated learning framework hold the potential to maximize the usage of any data available. This would allow exploration of advanced architectures, beyond the current standard linear and tree-based approaches.

Currently, prediction models are predominantly based on readily available data types (ISNCSCI assessments, image features, demographics). Using underexploited data types like omics and images could improve model predictions. Although the collection and curation of these data types may be challenging, they hold the potential to uncover a deeper biomedical understanding of the physiological processes underpinning SCI recovery in addition to identifying predictive biomarkers. Similarly, combinations of more established data types with imaging and molecular characterization of CSF or blood could be considered. Multi-modal learning that harnesses diverse molecular data alongside clinical information could offer nuanced insight into patient recovery trajectories [110]. Independent of the specific data types, innovative architectures like graph neural networks (GNNs) [111], attention mechanisms [112], and transformers [113]

could pave the way for next-generation modeling of spinal structures and patient recovery patterns. These architectures have previously been shown to effectively interpret complex data structures and longitudinal settings [114–116].

As models become more intricate, interpretability and explainability become more critical. Model transparency is essential for their acceptance as clinical decision support [12, 117, 118]. This is addressed in the topic of causal inference which enables tracing the prediction formulation [119]. The development of a data-driven prediction pipeline involves several critical steps including data collection, preprocessing, hyperparameter tuning, and model evaluation, ideally in collaboration with clinical experts. Automating (part of) this pipeline could accelerate the development and validation of solutions. AutoML (Automated Machine Learning) [120] is one example of such a streamlined pipeline, simplifying model deployment by automating tasks such as model selection and tuning, making it easier and faster to integrate with clinical applications.

In our current review, none of the included studies effectively accounted for the time dependence of SCI recovery beyond the input data and output label inclusion brackets. SCI recovery is a process. The longitudinal character of available SCI data may hold information regarding the specific recovery trajectory, implying that data from repeated assessments should be considered [121]. Going even further in temporal data resolution, sensor data from technologies such as wearables [122] and exoskeletons [123] are available for real-time monitoring of SCI patients. As data-driven models have previously been used to recognize physical activity in patients with SCI [124, 125], these could further enhance recovery prediction.

Finally, to ensure optimal training and objective performance evaluation, the SCI community needs to agree on standards regarding recovery prediction performance evaluation and endpoints of interest. In this review, we saw a variety of tasks associated with SCI recovery. While ISNCSCI composite scores, AIS grade conversion, and SCIM items are widely applied, the clinical relevance, objectivity, and prediction suitability need to be assessed. Ceiling and flooring effects [31], as well as co-linearity of contributing sub-scores, are common limitations in this context [126]. To enhance the understanding and accuracy of SCI recovery evaluations, it is crucial to account for the predominantly non-decreasing nature of patient recovery trajectories (i.e. patients generally exhibit improvement over time). Furthermore, when assessing ordinal scale labels like SCIM or AIS grades, metrics and loss functions suitable for these data types (i.e. ordinal) should be the standard [127]. Stringent data splits are equally important. To demonstrate model generalizability and clinical relevance, even given small sample sizes, (leave-one-out) cross-validation [128] and external test sets are key. In summary, we propose that by embracing state-of-the-art developments in machine learning, beyond standard architectures and training regimes, new insights into spinal cord injury can be discovered. This will pave the way towards robust and clinically relevant predictions.

Conclusion

Data-driven approaches from the realm of machine and deep learning hold promise to improve the prediction of recovery outcomes in SCI. The use of a variety of data types and implementation architectures reflects the growing trend in the SCI research community to harness these approaches in a clinical context. However, accomplishing clinical translation requires careful consideration of model validity, causality, and practical deployability, in addition to growing and maintaining increasingly detailed databases. Currently, data-driven SCI recovery prediction models often rely on standard architectures applied to comparably small sample sizes. This motivates further investigation of modern training paradigms and architectures that embrace the challenges in SCI recovery, such as limited data and model robustness. Data-driven predictions have the potential to inform rehabilitation strategies, enhance the design of clinical trials, and consequently, enhance the overall quality of life for individuals affected by SCI.

Acknowledgements

This study was supported by the Swiss National Science Foundation (Ambizione Grant, #PZ00P3.186101), Wings for Life Research Foundation (#301), Marie Skłodowska-Curie Actions (ReWIRE, #101073374), and the International Foundation for Research in Paraplegia (IRP, P192). SB was supported by the Botnar Research Centre for Child Health Postdoctoral Excellence Programme (#PEP-2021-1008). The funders did not specify the study design, data collection, analysis, or the decision to publish and preparation of the manuscript.

During the preparation of this work, the author(s) used ChatGPT-4 in order to extract information from papers. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

- [1] Anjum, A., Yazid, M.D., Fauzi Daud, M., Idris, J., Ng, A.M.H., Selvi Naicker, A., Ismail, O.H.R., Athi Kumar, R.K., Lokanathan, Y.: Spinal cord injury: Pathophysiology, multimolecular interactions, and underlying recovery mechanisms. *International Journal of Molecular Sciences* **21**(207533) (2020) <https://doi.org/10.3390/ijms21207533>. Citation Key: ijms21207533 tex.pubmedid: 33066029
- [2] Dimitrijevic, M.R., Danner, S.M., Mayr, W.: Neurocontrol of movement in humans with spinal cord injury. *Artificial Organs* **39**(10), 823–833 (2015) <https://doi.org/10.1111/aor.12614>
- [3] Zimmermann, G., Bolter, L.-M., Sluka, R., Höller, Y., Bathke, A., Thom-schewski, A., Leis, S., Lattanzi, S., Brigo, F., Trinka, E.: Sample sizes and statistical methods in interventional studies on individuals with spinal cord

- injury: A systematic review. *Journal of Evidence-Based Medicine* **12**, 200–208 (2019) <https://doi.org/10.1111/jebm.12356>
- [4] Pollard, C., Kennedy, P.: A longitudinal analysis of emotional impact, coping strategies and post-traumatic psychological growth following spinal cord injury: a 10-year review. *British journal of health psychology* **12 Pt 3**, 347–62 (2007) <https://doi.org/10.1348/135910707X197046>
- [5] EMSCI. <https://www.emsci.org/>
- [6] NISCI. <https://nisci-2020.eu/>
- [7] RHSCIR. <https://praxisinstitute.org/research-care/key-initiatives/national-sci-registry/>
- [8] NSCISC. <https://www.nscisc.uab.edu/>
- [9] Cone, D.C., Irvine, K.A., Middleton, P.M.: The methodology of the australian prehospital outcomes study of longitudinal epidemiology (apostle) project. *Prehospital Emergency Care* **16**(4), 505–512 (2012) <https://doi.org/10.3109/10903127.2012.689929>
- [10] Japanese National Spinal Cord Injury Database. https://www.kibirihah.johas.go.jp/200_SCI_DB/top.html
- [11] National Trauma Data Bank. <https://www.facs.org/quality-programs/trauma/quality/national-trauma-data-bank/>
- [12] Tjoa, E., Guan, C.: A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE Transactions on Neural Networks and Learning Systems* **32**, 4793–4813 (2019) <https://doi.org/10.1109/TNNLS.2020.3027314>
- [13] Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, , Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Mulbregt, P.: Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods* **17**(3), 261–272 (2020) <https://doi.org/10.1038/s41592-019-0686-2>
- [14] Ariji, Y., Hayashi, T., Ideta, R., Koga, R., Murai, S., Towatari, F., Terashi, Y., Sakai, H., Kurata, H., Maeda, T.: A prediction model of functional outcome at 6 months using clinical findings of a person with traumatic spinal cord injury at 1 month after injury. *Spinal cord* **58**(11), 1158–1165 (2020) <https://doi.org/10.1038/s41393-020-0488-5>

- [15] Belliveau, T., Jette, A.M., Seetharama, S., Axt, J., Rosenblum, D., Larose, D., Houlihan, B., Slavin, M., Larose, C.: Developing Artificial Neural Network Models to Predict Functioning One Year After Traumatic Spinal Cord Injury. *Archives of Physical Medicine and Rehabilitation* **97**(10), 1663–16683 (2016) <https://doi.org/10.1016/j.apmr.2016.04.014>
- [16] Bersch, I., Krebs, J., Fridén, J.: A Prediction Model for Various Treatment Pathways of Upper Extremity in Tetraplegia. *Frontiers in rehabilitation sciences* **3**, 889577 (2022) <https://doi.org/10.3389/fresc.2022.889577>
- [17] Cathomen, A., Maier, D., Kriz, J., Abel, R., Rohrich, F., Baumberger, M., Scivoletto, G., Weidner, N., Rupp, R., Jutzeler, CR., Steeves, JD., Curt, A., Bolliger, M., EMSCI Study Grp: Walking Outcome After Traumatic Paraplegic Spinal Cord Injury: The Function of Which Myotomes Makes a Difference? *NEUROREHABILITATION AND NEURAL REPAIR* **37**(5), 316–327 (2023) <https://doi.org/10.1177/15459683231166937>
- [18] Chen, S., Li, G., Li, F., Wang, G., Wang, Q.: A dynamic nomogram for predicting the probability of irreversible neurological dysfunction after cervical spinal cord injury: Research based on clinical features and MRI data. *BMC musculoskeletal disorders* **24**(1), 459 (2023) <https://doi.org/10.1186/s12891-023-06570-z>
- [19] Chou, A., Torres-Espin, A., Kyritsis, N., Huie, J.R., Khatry, S., Funk, J., Hay, J., Lofgreen, A., Shah, R., McCann, C., Pascual, L.U., Amorim, E., Weinstein, P.R., Manley, G.T., Dhall, S.S., Pan, J.Z., Bresnahan, J.C., Beattie, M.S., Whetstone, W.D., Ferguson, A.R.: Expert-augmented automated machine learning optimizes hemodynamic predictors of spinal cord injury outcome. *PloS one* **17**(4), 0265254 (2022) <https://doi.org/10.1371/journal.pone.0265254>
- [20] DeVries, Z., Hoda, M., Rivers, C.S., Maher, A., Wai, E., Moravek, D., Stratton, A., Kingwell, S., Fallah, N., Paquet, J., Phan, P.: Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *The spine journal : official journal of the North American Spine Society* **20**(2), 213–224 (2020) <https://doi.org/10.1016/j.spinee.2019.09.007>
- [21] Facchinello, Y., Beauséjour, M., Richard-Denis, A., Thompson, C., Mac-Thiong, J.-M.: Use of Regression Tree Analysis for Predicting the Functional Outcome after Traumatic Spinal Cord Injury. *Journal of neurotrauma* **38**(9), 1285–1291 (2021) <https://doi.org/10.1089/neu.2017.5321>
- [22] Harrington, G.M.B., Cool, P., Hulme, C., Osman, A., Chowdhury, J.R., Kumar, N., Budithi, S., Wright, K.: Routinely Measured Hematological Markers Can Help to Predict American Spinal Injury Association Impairment Scale Scores after Spinal Cord Injury. *Journal of Neurotrauma* **38**(3), 301–308 (2021) <https://doi.org/10.1089/neu.2020.7144>

- [23] Hori, T., Imura, T., Tanaka, R.: Development of a clinical prediction rule for patients with cervical spinal cord injury who have difficulty in obtaining independent living. *The spine journal : official journal of the North American Spine Society* **22**(2), 321–328 (2022) <https://doi.org/10.1016/j.spinee.2021.08.010>
- [24] Hug, A., Schuld, C., Mürle, B., Böttinger, M., Weidner, N., Rupp, R.: Ulnar nerve integrity predicts 1-year outcome in cervical spinal cord injury. *Neurological Research and Practice* (1) (2019) <https://doi.org/10.1186/s42466-019-0017-1>
- [25] Hupp, M., Pavese, C., Bachmann, L.M., Koller, R., Schubert, M.: Electrophysiological Multimodal Assessments Improve Outcome Prediction in Traumatic Cervical Spinal Cord Injury. *Journal of neurotrauma* **35**(24), 2916–2923 (2018) <https://doi.org/10.1089/neu.2017.5576>
- [26] Imura, T., Hori, T., Tanaka, R.: A Simplified Clinical Prediction Rule for Prognosticating Dependent Daily Living in Patients with Thoracic Spinal Cord Injury: A Multicenter Nationwide Japan Registry Study. *European neurology* **86**(3), 178–184 (2023) <https://doi.org/10.1159/000529780>
- [27] Inoue, T., Ichikawa, D., Ueno, T., Cheong, M., Inoue, T., Whetstone, W.D., Endo, T., Nizuma, K., Tominaga, T.: XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury. *Neurotrauma reports* **1**(1), 8–16 (2020) <https://doi.org/10.1089/neur.2020.0009>
- [28] Ishida, Y., Tominaga, T.: Predictors of neurologic recovery in acute central cervical cord injury with only upper extremity impairment. *Spine* **27**(15), 1652–1658 (2002) <https://doi.org/10.1097/00007632-200208010-00011>
- [29] Javeed, S., Greenberg, J.K., Zhang, J.K., Dibble, C.F., Khalifeh, J.M., Liu, Y., Wilson, T.J., Yang, L.J., Park, Y., Ray, W.Z.: Derivation and Validation of a Clinical Prediction Rule for Upper Limb Functional Outcomes After Traumatic Cervical Spinal Cord Injury. *JAMA network open* **5**(12), 2247949 (2022) <https://doi.org/10.1001/jamanetworkopen.2022.47949>
- [30] Jean, S., Mac-Thiong, J.-M., Jean, M.-C., Dionne, A., Bégin, J., Richard-Denis, A.: Early Clinical Prediction of Independent Outdoor Functional Walking Capacity in a Prospective Cohort of Traumatic Spinal Cord Injury Patients. *American journal of physical medicine & rehabilitation* **100**(11), 1034–1041 (2021) <https://doi.org/10.1097/PHM.0000000000001812>
- [31] Kalyani, P., Manasa, Y., Ahammad, S.H., Suman, M., Anwer, T.M.K., Hossain, M.A., Rashed, A.N.Z.: Prediction of patient’s neurological recovery from cervical spinal cord injury through XGBoost learning approach. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* **32**(6), 2140–2148 (2023) <https://doi.org/10.1007/s00586-023-07712-6>

- [32] Kaminski, L., Cordemans, V., Cernat, E., M'Bra, K.I., Mac-Thiong, J.-M.: Functional Outcome Prediction after Traumatic Spinal Cord Injury Based on Acute Clinical Factors. *Journal of neurotrauma* **34**(12), 2027–2033 (2017) <https://doi.org/10.1089/neu.2016.4955>
- [33] Kapoor, D., Xu, C.: Spinal Cord Injury AIS Predictions Using Machine Learning. *eNeuro* **10**(1), 0149–222022 (2023) <https://doi.org/10.1523/ENEURO.0149-22.2022>
- [34] Kato, C., Uemura, O., Sato, Y., Tsuji, T.: Functional outcome prediction after spinal cord injury using ensemble machine learning. *Archives of physical medicine and rehabilitation*, 0003–999323004781 (2023) <https://doi.org/10.1016/j.apmr.2023.08.011>
- [35] Leister, I., Haider, T., Vogel, M., Vastmans, J., Langthaler, P., Mattiassich, G., Christ, A., Etschmaier, M., Eijkenboom, A., Burghuber, J., Kindermann, H., Mach, O., Maier, D., Högel, F.: A Predictive Model to Identify Treatment-related Risk Factors for Odontoid Fracture Nonunion Using Machine Learning. *Spine* **48**(3), 164–171 (2023) <https://doi.org/10.1097/BRS.0000000000004510>
- [36] Mummaneni, N., Burke, J.F., DiGiorgio, A.M., Thomas, L.H., Duong-Fernandez, X., Harris, M., Pascual, L.U., Ferguson, A.R., Russell Huie, J., Pan, J.Z., Hemmerle, D.D., Singh, V., Torres-Espin, A., Omondi, C., Kyritsis, N., Weinstein, P.R., Whetstone, W.D., Manley, G.T., Bresnahan, J.C., Beattie, M.S., Cohen-Adad, J., Dhall, S.S., Talbott, J.F.: Injury volume extracted from MRI predicts neurologic outcome in acute spinal cord injury: A prospective TRACK-SCI pilot study. *Journal of Clinical Neuroscience* **82**, 231–236 (2020) <https://doi.org/10.1016/j.jocn.2020.11.003>
- [37] Øhrn, A., Rowland, T.: Rough sets: A knowledge discovery technique for multifactorial medical outcomes. *American Journal of Physical Medicine and Rehabilitation* **79**(1), 100–108 (2000) <https://doi.org/10.1097/00002060-200001000-00022>
- [38] Okimatsu, S., Maki, S., Furuya, T., Fujiyoshi, T., Kitamura, M., Inada, T., Aramomi, M., Yamauchi, T., Miyamoto, T., Inoue, T., Yunde, A., Miura, M., Shiga, Y., Inage, K., Orita, S., Eguchi, Y., Ohtori, S.: Determining the short-term neurological prognosis for acute cervical spinal cord injury using machine learning. *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia* **96**, 74–79 (2022) <https://doi.org/10.1016/j.jocn.2021.11.037>
- [39] Pfyffer, D., Vallotton, K., Curt, A., Freund, P.: Predictive Value of Midsagittal Tissue Bridges on Functional Recovery After Spinal Cord Injury. *Neurorehabilitation and Neural Repair* **35**(1), 33–43 (2021) <https://doi.org/10.1177/1545968320971787>

- [40] Raju, P.R.S.S.V., Asanambigai, V., Mudunuri, S.B.: A Learning-based Correlated Graph Model for Spinal Cord Injury Prediction from Magnetic Resonance Spinal Images. *International Journal of Advanced Computer Science and Applications* **14**(1), 442–449 (2023) <https://doi.org/10.14569/IJACSA.2023.0140147>
- [41] Rigot, S.K., Boninger, M.L., Ding, D., McKernan, G., Field-Fote, E.C., Hoffman, J., Hibbs, R., Worobey, L.A.: Toward Improving the Prediction of Functional Ambulation After Spinal Cord Injury Through the Inclusion of Limb Accelerations During Sleep and Personal Factors. *Archives of physical medicine and rehabilitation* **103**(4), 676–6876 (2022) <https://doi.org/10.1016/j.apmr.2021.02.029>
- [42] Rowland, T., Ohno-Machado, L., Ohrn, A.: Comparison of multiple prediction models for ambulation following spinal cord injury. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, 528–532 (1998)
- [43] Shimizu, T., Suda, K., Maki, S., Koda, M., Matsumoto Harmon, S., Komatsu, M., Ota, M., Ushirozako, H., Minami, A., Takahata, M., Iwasaki, N., Takahashi, H., Yamazaki, M.: Efficacy of a machine learning-based approach in predicting neurological prognosis of cervical spinal cord injury patients following urgent surgery within 24 h after injury. *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia* **107**, 150–156 (2023) <https://doi.org/10.1016/j.jocn.2022.11.003>
- [44] Sizheng, Z., Boxuan, H., Feng, X., Dianying, Z.: A functional outcome prediction model of acute traumatic spinal cord injury based on extreme gradient boost. *Journal of orthopaedic surgery and research* **17**(1), 451 (2022) <https://doi.org/10.1186/s13018-022-03343-7>
- [45] Smith, A.C., Draganich, C., Thornton, W.A., Berliner, J.C., Lennarson, P.J., Rejc, E., Sevigny, M., Charlifue, S., Tefertiller, C., Weber, K.A.n.: A Single Dermatome Clinical Prediction Rule for Independent Walking 1 Year After Spinal Cord Injury. *Archives of physical medicine and rehabilitation*, 0003–99932300374 (2023) <https://doi.org/10.1016/j.apmr.2023.06.015>
- [46] Tanadini, L.G., Steeves, J.D., Hothorn, T., Abel, R., Maier, D., Schubert, M., Weidner, N., Rupp, R., Curt, A.: Identifying homogeneous subgroups in neurological disorders: Unbiased recursive partitioning in cervical complete spinal cord injury. *Neurorehabilitation and Neural Repair* **28**(6), 507–515 (2014) <https://doi.org/10.1177/1545968313520413>
- [47] Tee, J.W., Rivers, C.S., Fallah, N., Noonan, V.K., Kwon, B.K., Fisher, C.G., Street, J.T., Ailon, T., Dea, N., Paquette, S., Dvorak, M.F.: Decision tree analysis to better control treatment effects in spinal cord injury clinical research. *Journal of neurosurgery. Spine*, 1–9 (2019) <https://doi.org/10.3171/2019.3.SPINE18993>

- [48] Tomioka, Y., Uemura, O., Ishii, R., Liu, M.: Using a logarithmic model to predict functional independence after spinal cord injury: A retrospective study. *Spinal cord* **57**(12), 1048–1056 (2019) <https://doi.org/10.1038/s41393-019-0315-z>
- [49] Torres-Espín, A., Haefeli, J., Ehsanian, R., Torres, D., Almeida, C.A., Huie, J.R., Chou, A., Morozov, D., Sanderson, N., Dirlikov, B., Suen, C.G., Nielson, J.L., Kyritsis, N., Hemmerle, D.D., Talbott, J.F., Manley, G.T., Dhall, S.S., Whetstone, W.D., Bresnahan, J.C., Beattie, M.S., McKenna, S.L., Pan, J.Z., Ferguson, A.R.: Topological network analysis of patient similarity for precision management of acute blood pressure in spinal cord injury. *eLife* **10** (2021) <https://doi.org/10.7554/eLife.68015>
- [50] van Middendorp, J.J., Hosman, A.J.F., Donders, A.R.T., Pouw, M.H., Ditunno, J.F.J., Curt, A., Geurts, A.C.H., Van de Meent, H.: A clinical prediction rule for ambulation outcomes after traumatic spinal cord injury: A longitudinal cohort study. *Lancet (London, England)* **377**(9770), 1004–1010 (2011) [https://doi.org/10.1016/S0140-6736\(10\)62276-3](https://doi.org/10.1016/S0140-6736(10)62276-3)
- [51] Venkatapathiraju, P.R.S.S., Asanambigai, V., Babu Mudunuri, S.: MOD-ELLING A DEEP KERNEL-BASED LEARNING APPROACH FOR SPINAL CORD INJURY PREDICTION. *Journal of Theoretical and Applied Information Technology* **100**(21), 6439–6452 (2022)
- [52] Warschausky, S., Kay, J.B., Kewman, D.G.: Hierarchical linear modeling of FIM instrument growth curve characteristics after spinal cord injury. *Archives of Physical Medicine and Rehabilitation* **82**(3), 329–334 (2001) <https://doi.org/10.1053/apmr.2001.21510>
- [53] Wilson, J.R., Grossman, R.G., Frankowski, R.F., Kiss, A., Davis, A.M., Kulkarni, A.V., Harrop, J.S., Aarabi, B., Vaccaro, A., Tator, C.H., Dvorak, M., Shaffrey, C.I., Harkema, S., Guest, J.D., Fehlings, M.G.: A clinical prediction model for long-term functional outcome after traumatic spinal cord injury based on acute clinical and imaging factors. *Journal of neurotrauma* **29**(13), 2263–2271 (2012) <https://doi.org/10.1089/neu.2012.2417>
- [54] Winchester, P., Smith, P., Foreman, N., Mosby, J.M., Pacheco, F., Querry, R., Tansey, K.: A prediction model for determining over ground walking speed after locomotor training in persons with motor incomplete spinal cord injury. *The journal of spinal cord medicine* **32**(1), 63–71 (2009) <https://doi.org/10.1080/10790268.2009.11760754>
- [55] Yan, X., He, Y., Jia, M., Yang, J., Huang, K., Zhang, P., Lai, J., Chen, M., Fan, S., Li, S., Fan, Z., Teng, H.: Development of a Dynamic Nomogram for Predicting the Probability of Satisfactory Recovery after 6 Months for Cervical Traumatic Spinal Cord Injury. *Orthopaedic surgery* **15**(4), 1008–1020 (2023) <https://doi.org/10.1111/os.13679>

- [56] Yang, F., Guo, X.: Research on Rehabilitation Effect Prediction for Patients with SCI Based on Machine Learning. *World neurosurgery* **158**, 662–674 (2022) <https://doi.org/10.1016/j.wneu.2021.11.040>
- [57] Zariffa, J., Curt, A., Verrier, M.C., Fehlings, M.G., Kalsi-Ryan, S.: Predicting task performance from upper extremity impairment measures after cervical spinal cord injury. *Spinal cord* **54**(12), 1145–1151 (2016) <https://doi.org/10.1038/sc.2016.77>
- [58] Agarwal, N., Aabedi, A.A., Torres-Espin, A., Chou, A., Wozny, T.A., Mummaneni, P.V., Burke, J.F., Ferguson, A.R., Kyritsis, N., Dhall, S.S., Weinstein, P.R., Duong-Fernandez, X., Pan, J., Singh, V., Hemmerle, D.D., Talbott, J.F., Whetstone, W.D., Bresnahan, J.C., Manley, G.T., Beattie, M.S., DiGiorgio, A.M.: Decision tree-based machine learning analysis of intraoperative vasopressor use to optimize neurological improvement in acute spinal cord injury. *Neurosurgical focus* **52**(4), 9 (2022) <https://doi.org/10.3171/2022.1.FOCUS21743>
- [59] El Sammak, S., Michalopoulos, G.D., Arya, N., Bhandarkar, A.R., Moinuddin, F.M., Jarrah, R., Yolcu, Y.U., Shoushtari, A., Bydon, M.: Prediction model for neurogenic bladder recovery one year after traumatic spinal cord injury. *World neurosurgery*, 1878–87502301166 (2023) <https://doi.org/10.1016/j.wneu.2023.08.054>
- [60] HeidarAbadi, N.N., Hakemi, L., Kolivand, P., Safdari, R., Saeidi, M.G.: Comparing performances of intelligent classifier algorithms for predicting type of pain in patients with spinal cord injury. *Electronic physician* **9**(7), 4847–4852 (2017) <https://doi.org/10.19082/4847>
- [61] Kato, C., Uemura, O., Sato, Y., Tsuji, T.: Decision tree analysis accurately predicts discharge destination after spinal cord injury rehabilitation. *Archives of physical medicine and rehabilitation*, 0003–999323004793 (2023) <https://doi.org/10.1016/j.apmr.2023.08.010>
- [62] Pavese, C., Schneider, M.P., Schubert, M., Curt, A., Scivoletto, G., Finazzi-Agrò, E., Mehnert, U., Maier, D., Abel, R., Röhrich, F., Weidner, N., Rupp, R., Kessels, A.G., Bachmann, L.M., Kessler, T.M.: Prediction of bladder outcomes after traumatic spinal cord injury: A longitudinal cohort study. *PLoS Medicine* **13**(6) (2016) <https://doi.org/10.1371/journal.pmed.1002041>
- [63] Pavese, C., Bachmann, L.M., Schubert, M., Curt, A., Mehnert, U., Schneider, M.P., Scivoletto, G., Finazzi Agrò, E., Maier, D., Abel, R., Weidner, N., Rupp, R., Kessels, A.G., Kessler, T.M.: Bowel outcome prediction after traumatic spinal cord injury: Longitudinal cohort study. *Neurorehabilitation and neural repair* **33**(11), 902–910 (2019) <https://doi.org/10.1177/1545968319868722>
- [64] Mills, P., Holtz, K., Szefer, E., Noonan, V., Kwon, B.: Early predictors of developing problematic spasticity following traumatic spinal cord injury: A

- prospective cohort study. *J Spinal Cord Med* **43**(3), 315–330 (2020) <https://doi.org/10.1080/10790268.2018.1527082>
- [65] Ariji, Y., Hayashi, T., Ideta, R., Koga, R., Murai, S., Naka, T., Ifuku, R., Towatari, F., Sakai, H., Kurata, H., Maeda, T.: Identification of a reliable sacral-sparing examination to assess the ASIA Impairment Scale in patients with traumatic Spinal Cord Injury. *J Spinal Cord Med* **47**(2), 286–292 (2024) <https://doi.org/10.1080/10790268.2022.2047548>
- [66] Brown, S.J., Harrington, G.M.B., Hulme, C.H., Morris, R., Bennett, A., Tsang, W.-H., Osman, A., Chowdhury, J., Kumar, N., Wright, K.T.: A preliminary cohort study assessing routine blood analyte levels and neurological outcome after spinal cord injury. *Journal of Neurotrauma* **37**(3), 466–480 (2020) <https://doi.org/10.1089/neu.2019.6495>
- [67] Bykowski, E.A., Petersson, J.N., Dukelow, S., Ho, C., Debert, C.T., Montina, T., Metz, G.A.S.: Urinary biomarkers indicative of recovery from spinal cord injury: A pilot study. *IBRO Neuroscience Reports* **10**, 178–185 (2021) <https://doi.org/10.1016/j.ibneur.2021.02.007>
- [68] Bykowski, E.A., Petersson, J.N., Dukelow, S., Ho, C., Debert, C.T., Montina, T., Metz, G.A.S.: Identification of serum metabolites as prognostic biomarkers following spinal cord injury: A pilot study. *Metabolites* **13**(5) (2023) <https://doi.org/10.3390/metabo13050605>
- [69] Diong, J., Harvey, L.A., Kwah, L.K., Eyles, J., Ling, M.J., Ben, M., Herbert, R.D.: Incidence and predictors of contracture after spinal cord injury: a prospective cohort study. *Spinal Cord* **50**(8), 579–584 (2012) <https://doi.org/10.1038/sc.2012.25>
- [70] Moghaddam, A., Heller, R., Daniel, V., Swing, T., Akbar, M., Gerner, H.-J., Biglari, B.: Exploratory study to suggest the possibility of mmp-8 and mmp-9 serum levels as early markers for remission after traumatic spinal cord injury. *Spinal Cord* **55**(1), 8–15 (2017) <https://doi.org/10.1038/sc.2016.104>
- [71] Skinnider, M.A., Rogalski, J., Tigchelaar, S., Manouchehri, N., Prudova, A., Jackson, A.M., Nielsen, K., Jeong, J., Chaudhary, S., Shortt, K., Gallagher-Kurtzke, Y., So, K., Fong, A., Gupta, R., Okon, E.B., Rizzuto, M.A., Dong, K., Streijger, F., Belanger, L., Ritchie, L., Tsang, A., Christie, S., Mac-Thiong, J.-M., Bailey, C., Ailon, T., Charest-Morin, R., Dea, N., Wilson, J.R., Dhall, S., Paquette, S., Street, J., Fisher, C.G., Dvorak, M.F., Shannon, C., Borchers, C., Balshaw, R., Foster, L.J., Kwon, B.K.: Proteomic portraits reveal evolutionarily conserved and divergent responses to spinal cord injury. *Molecular and Cellular Proteomics* **20** (2021) <https://doi.org/10.1016/j.mcpro.2021.100096>
- [72] Tee, J.W., Chan, P.C.H., Fitzgerald, M.C.B., Liew, S.M., Rosenfeld, J.V.: Early predictors of functional disability after spine trauma: A level 1 trauma

- center study. *Spine* **38**(12), 999–1007 (2013) <https://doi.org/10.1097/BRS.0b013e31828432a3>
- [73] Varzi, D., Coupaud, S.A.F., Purcell, M., Allan, D.B., Gregory, J.S., Barr, R.J.: Bone morphology of the femur and tibia captured by statistical shape modelling predicts rapid bone loss in acute spinal cord injury patients. *Bone* **81**, 495–501 (2015) <https://doi.org/10.1016/j.bone.2015.08.026>
- [74] Zhao, J.-L., Lai, S.-T., Du, Z.-Y., Xu, J., Sun, Y.-R., Yuan, Q., Wu, X., Li, Z.-Q., Hu, J., Xie, R.: Circulating neutrophil-to-lymphocyte ratio at admission predicts the long-term outcome in acute traumatic cervical spinal cord injury patients. *BMC musculoskeletal disorders* **21**(1), 548 (2020) <https://doi.org/10.1186/s12891-020-03556-z>
- [75] Birkhäuser, V., Anderson, C.E., Kozomara, M., Bywater, M., Gross, O., Kiss, S., Knüpfer, S.C., Koschorke, M., Leitner, L., Mehnert, U., Sadri, H., Sammer, U., Stächele, L., Tornic, J., Liechti, M.D., Brinkhof, M.W.G., Kessler, T.M.: Urodynamics are essential to predict the risk for upper urinary tract damage after acute spinal cord injury. *Biomedicines* **11**(6) (2023) <https://doi.org/10.3390/biomedicines11061748>
- [76] Leidinger, A., Zuckerman, S.L., Feng, Y., He, Y., Chen, X., Cheserem, B., Gerber, L.M., Lessing, N.L., Shabani, H.K., Härtl, R., Mangat, H.S.: Predictors of spinal trauma care and outcomes in a resource-constrained environment: a decision tree analysis of spinal trauma surgery and outcomes in tanzania. *Journal of Neurosurgery: Spine* **38**(4), 503–511 (2023) <https://doi.org/10.3171/2022.11.SPINE22763>
- [77] Kwon, S.Y., Shin, J.J., Lee, H.J., Cho, W.H.: Prognostic factors for surgical outcome in spinal cord injury associated with ossification of the posterior longitudinal ligament (opll). *Journal of Orthopaedic Surgery and Research* **10**(1) (2015) <https://doi.org/10.1186/s13018-015-0235-3>
- [78] Pavese, C., Scivoletto, G., Puci, M.V., Abel, R., Curt, A., Maier, D., Rupp, R., Schubert, M., Weidner, N., Montomoli, C., Kessler, T.M.: Prediction of bowel management independence after ischemic spinal cord injury. *European journal of physical and rehabilitation medicine* **58**(5), 709–714 (2022) <https://doi.org/10.23736/S1973-9087.22.07366-X>
- [79] Cathomen, A., Sirucek, L., Killeen, T., Abel, R., Maier, D., Weidner, N., Rupp, R., Hothorn, T., Steeves, J.D., Curt, A., Bolliger, M.: Inclusive trial designs in acute spinal cord injuries: Prediction-based stratification of clinical walking outcome and projected enrolment frequencies. *Neurorehabilitation and Neural Repair* **36**(4–5), 274–285 (2022) <https://doi.org/10.1177/15459683221078302>
- [80] Talbott, J.F., Whetstone, W.D., Readdy, W.J., Ferguson, A.R., Bresnahan, J.C., Saigal, R., Hawryluk, G.W.J., Beattie, M.S., Mabray, M.C., Pan, J.Z., Manley,

- G.T., Dhall, S.S.: The brain and spinal injury center score: a novel, simple, and reproducible method for assessing the severity of acute cervical spinal cord injury with axial t2-weighted mri findings. *Journal of Neurosurgery: Spine* **23**(4), 495–504 (2015) <https://doi.org/10.3171/2015.1.SPINE141033>
- [81] Itzkovich, M., Gelernter, I., Biering-Sorensen, F., Weeks, C., Laramee, M.T., Craven, B.C., Tonack, M., Hitzig, S.L., Glaser, E., Zeilig, G., Aito, S., Scivoletto, G., Mecci, M., Chadwick, R.J., El Masry, W.S., Osman, A., Glass, C.A., Silva, P., Soni, B.M., Gardner, B.P., Savic, G., Bergström, E.M., Bluvshstein, V., Ronen, J., Catz, A.: The spinal cord independence measure (scim) version iii: Reliability and validity in a multi-center international study. *Disability and Rehabilitation* **29**(24), 1926–1933 (2007) <https://doi.org/10.1080/09638280601046302>
- [82] Enright, P.L.: The six-minute walk test. *Respiratory Care* **48**(8), 783–785 (2003)
- [83] Kidd, D., Stewart, G., Baldry, J., Johnson, J., Rossiter, D., Petruckevitch, A., Thompson, A.J.: The functional independence measure: A comparative validity and reliability study. *Disability and Rehabilitation* **17**(1), 10–14 (1995) <https://doi.org/10.3109/09638289509166622>
- [84] Kalsi-Ryan, S., Curt, A., Fehlings, M., Verrier, M.: Assessment of the hand in tetraplegia using the graded redefined assessment of strength, sensibility and prehension (grassp): Impairment versus function. *Topics in Spinal Cord Injury Rehabilitation* **14**(4), 34–46 (2009) <https://doi.org/10.1310/sci1404-34>
- [85] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D.: Scikit-learn: Machine learning in python. *MACHINE LEARNING IN PYTHON*
- [86] Simpson, L.A., Eng, J.J., Hsieh, J.T.C., Wolfe, D.L., Team, S.C.I.R.E.S.R.: The health and life priorities of individuals with spinal cord injury: a systematic review. *Journal of Neurotrauma* **29**(8), 1548–1555 (2012) <https://doi.org/10.1089/neu.2011.2226>
- [87] Jones, M.L., Evans, N., Tefertiller, C., Backus, D., Sweatman, M., Tansey, K., Morrison, S.: Activity-based therapy for recovery of walking in individuals with chronic spinal cord injury: Results from a randomized clinical trial. *Archives of Physical Medicine and Rehabilitation* **95**(12), 2239–22462 (2014) <https://doi.org/10.1016/j.apmr.2014.07.400>
- [88] Derakhshanrad, N., Saberi, H., Yekaninejad, M.S., Joghataei, M.T.: Subcutaneous granulocyte colony-stimulating factor administration for subacute traumatic spinal cord injuries, report of neurological and functional outcomes: a double-blind randomized controlled clinical trial. *Journal of Neurosurgery: Spine* **30**(1), 19–30 (2018) <https://doi.org/10.3171/2018.6.SPINE18209>

- [89] Zamani, H., Soufizomorrod, M., Oraee-Yazdani, S., Naviafar, D., Akhlaghpasand, M., Seddighi, A., Soleimani, M.: Safety and feasibility of autologous olfactory ensheathing cell and bone marrow mesenchymal stem cell co-transplantation in chronic human spinal cord injury: a clinical trial. *Spinal Cord* **60**(1), 63–70 (2022) <https://doi.org/10.1038/s41393-021-00687-5>
- [90] García-Alén, L., Kumru, H., Castillo-Escario, Y., Benito-Penalva, J., Medina-Casanovas, J., Gerasimenko, Y.P., Edgerton, V.R., García-Alías, G., Vidal, J.: Transcutaneous cervical spinal cord stimulation combined with robotic exoskeleton rehabilitation for the upper limbs in subjects with cervical sci: Clinical trial. *Biomedicines* **11**(22), 589 (2023) <https://doi.org/10.3390/biomedicines11020589>
- [91] Kirshblum, S., Millis, S., McKinley, W., Tulskey, D.: Late neurologic recovery after traumatic spinal cord injury. *Archives of Physical Medicine and Rehabilitation* **85**(11), 1811–1817 (2004) <https://doi.org/10.1016/j.apmr.2004.03.015>
- [92] Paton, C., Kobayashi, S.: An open science approach to artificial intelligence in healthcare. *Yearbook of Medical Informatics* **28**, 47–51 (2019) <https://doi.org/10.1055/s-0039-1677898>
- [93] Bourguignon, L., Lukas, L.P., Guest, J.D., Geisler, F.H., Noonan, V., Curt, A., Brüningk, S.C., Jutzeler, C.R.: Studying missingness in spinal cord injury data: challenges and impact of data imputation. *BMC Medical Research Methodology* **24**(1), 5 (2024) <https://doi.org/10.1186/s12874-023-02125-x>
- [94] Kopp, M.A., Watzlawick, R., Martus, P., Failli, V., Finkenstaedt, F.W., Chen, Y., DeVivo, M.J., Dirnagl, U., Schwab, J.M.: Long-term functional outcome in patients with acquired infections after acute spinal cord injury. *Neurology* **88**(9), 892–900 (2017) <https://doi.org/10.1212/WNL.0000000000003652>
- [95] Failli, V., Kopp, M.A., Gericke, C., Martus, P., Klingbeil, S., Brommer, B., Laginha, I., Chen, Y., DeVivo, M.J., Dirnagl, U., Schwab, J.M.: Functional neurological recovery after spinal cord injury is impaired in patients with infections. *Brain* **135**(11), 3238–3250 (2012) <https://doi.org/10.1093/brain/aws267>
- [96] Wu, W., Jia, F., Enders, C.: A comparison of imputation strategies for ordinal missing data on likert scale variables. *Multivariate Behavioral Research* **50**(5), 484–503 (2015) <https://doi.org/10.1080/00273171.2015.1022644>
- [97] Shao, J., Zhong, B.: Last observation carry-forward and last observation analysis. *Statistics in Medicine* **22**(15), 2429–2441 (2003) <https://doi.org/10.1002/sim.1519>
- [98] Rubin, D.B.: Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91**(434), 473–489 (1996) <https://doi.org/10.1080/01621459.1996.10473171>

1996.10476908

- [99] Liebel, L., Körner, M.: Auxiliary tasks in multi-task learning. ArXiv **abs/1805.06334** (2018)
- [100] Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F.: A survey on contrastive self-supervised learning. *Technologies* **9**(11), 2 (2021) <https://doi.org/10.3390/technologies9010002>
- [101] Shorten, C., Khoshgoftaar, T.: A survey on image data augmentation for deep learning. *Journal of Big Data* **6**, 1–48 (2019) <https://doi.org/10.1186/s40537-019-0197-0>
- [102] Zhu, X.J.: Semi-Supervised Learning Literature Survey, (2005). Accepted: 2012-03-15T17:19:12Z. <https://minds.wisconsin.edu/handle/1793/60444>
- [103] Settles, B.: Active Learning Literature Survey, (2009). Accepted: 2012-03-15T17:23:56Z. <https://minds.wisconsin.edu/handle/1793/60660>
- [104] Cheplygina, V., Bruijne, M., Pluim, J.P.W.: Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis* **54**, 280–296 (2019) <https://doi.org/10.1016/j.media.2019.03.009>
- [105] Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis* **71**, 102062 (2021) <https://doi.org/10.1016/j.media.2021.102062>
- [106] Hon, M., Khan, N.M.: Towards alzheimer’s disease classification through transfer learning
- [107] Kim, H.E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M.E., Ganslandt, T.: Transfer learning for medical image classification: a literature review. *BMC Medical Imaging* **22**(1), 69 (2022) <https://doi.org/10.1186/s12880-022-00793-7>
- [108] Li, W., Milletari, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M., Feng, A.: Privacy-preserving federated brain tumour segmentation. ArXiv **abs/1910.00962** (2019) https://doi.org/10.1007/978-3-030-32692-0_16
- [109] Noonan, V., Thorogood, N., Joshi, P., Fehlings, M., Craven, B.C., Linassi, G., Fourney, D., Kwon, B., Bailey, C., Tsai, E., Drew, B., Ahn, H., Tsui, D., Dvorak, M.: Meeting the privacy requirements for the development of a multi-centre patient registry in canada: the rick hansen spinal cord injury registry. *Health-care policy = Politiques de sante* **8** 4, 87–99 (2013) <https://doi.org/10.12927/HCPOL.2013.23397>

- [110] Li, Y., Wu, F.-X., Ngom, A.: A review on machine learning principles for multi-view biological data integration. *Briefings in Bioinformatics* **19**, 325–340 (2016) <https://doi.org/10.1093/bib/bbw113>
- [111] Paul, S.G., Saha, A., Hasan, M.Z., Noori, S.R.H., Moustafa, A.: A systematic review of graph neural network in healthcare-based applications: Recent advances, trends, and future directions. *IEEE Access* **12**, 15145–15170 (2024) <https://doi.org/10.1109/ACCESS.2024.3354809>
- [112] Lu, S., Liu, M., Yin, L., Yin, Z., Liu, X., Zheng, W.: The multi-modal fusion in visual question answering: a review of attention mechanisms. *PeerJ Computer Science* **9**, 1400 (2023) <https://doi.org/10.7717/peerj-cs.1400>
- [113] Nerella, S., Bandyopadhyay, S., Zhang, J., Contreras, M., Siegel, S., Bumin, A., Silva, B., Sena, J., Shickel, B., Bihorac, A., Khezeli, K., Rashidi, P.: *Transformers in Healthcare: A Survey* (2023)
- [114] Bürgin, V., Prevost, R., Stollenga, M.F.: Robust vertebra identification using simultaneous node and edge predicting graph neural networks. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*. Lecture Notes in Computer Science, pp. 483–493. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-43996-4_46
- [115] Tao, R., Liu, W., Zheng, G.: Spine-transformers: Vertebra labeling and segmentation in arbitrary field-of-view spine cts via 3d transformers. *Medical Image Analysis* **75**, 102258 (2022) <https://doi.org/10.1016/j.media.2021.102258>
- [116] Tipirneni, S., Reddy, C.: Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **16**, 1–17 (2021) <https://doi.org/10.1145/3516367>
- [117] Vellido, A.: The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*, 1–15 (2020) <https://doi.org/10.1007/s00521-019-04051-w>
- [118] Tonekaboni, S., Joshi, S., McCradden, M.D., Goldenberg, A.: What clinicians want: Contextualizing explainable machine learning for clinical end use. *arXiv* (2019) <https://doi.org/10.48550/arxiv.1905.05134>
- [119] Sanchez, P., Voisey, J.P., Xia, T., Watson, H.I., O’Neil, A.Q., Tsiftaris, S.A.: Causal machine learning for healthcare and precision medicine. *Royal Society Open Science* **9**(8), 220638 (2022) <https://doi.org/10.1098/rsos.220638>
- [120] Waring, J., Lindvall, C., Umeton, R.: Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artificial intelligence in medicine* **104**, 101822 (2020) <https://doi.org/10.1016/j.artmed.2020.101822>

- [121] Venkatraman, A., Hebert, M., Bagnell, J.: Improving multi-step prediction of learned time series models, 3024–3030 (2015) <https://doi.org/10.1609/aaai.v29i1.9590>
- [122] Popp, W.L., Schneider, S., Bär, J., Bösch, P., Spengler, C.M., Gassert, R., Curt, A.: Wearable sensors in ambulatory individuals with a spinal cord injury: From energy expenditure estimation to activity recommendations. *Frontiers in Neurology* **10** (2019) <https://doi.org/10.3389/fneur.2019.01092>
- [123] Esquenazi, A., Talaty, M., Packel, A., Saulino, M.: The rewalk powered exoskeleton to restore ambulatory function to individuals with thoracic-level motor-complete spinal cord injury. *American Journal of Physical Medicine & Rehabilitation* **91**(11), 911 (2012) <https://doi.org/10.1097/PHM.0b013e318269d9a3>
- [124] Alhammad, N., Al-Dossari, H.: Recognizing physical activities for spinal cord injury rehabilitation using wearable sensors. *Sensors* **21**(1616), 5479 (2021) <https://doi.org/10.3390/s21165479>
- [125] Sok, P., Xiao, T., Azeze, Y., Jayaraman, A., Albert, M.V.: Activity recognition for incomplete spinal cord injury subjects using hidden markov models. *IEEE Sensors Journal* **18**(15), 6369–6374 (2018) <https://doi.org/10.1109/JSEN.2018.2845749>
- [126] Graves, D.E., Frankiewicz, R.G., Donovan, W.H.: Construct validity and dimensional structure of the asia motor scale. *Journal of Spinal Cord Medicine* **29**(1), 39–45 (2006) <https://doi.org/10.1080/10790268.2006.11753855>
- [127] Gutierrez, P.A., Perez-Ortiz, M., Sanchez-Monedero, J., Fernandez-Navarro, F., Hervas-Martinez, C.: Ordinal regression methods: Survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering* **28**(1), 127–146 (2016) <https://doi.org/10.1109/TKDE.2015.2457911>
- [128] Wong, T.-T.: Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition* **48**(9), 2839–2846 (2015) <https://doi.org/10.1016/j.patcog.2015.03.009>

Supplementary

Order of labeling for study exclusion:

Item	Description
1	Review, collection or commentary
2	Non-human
3	Not traumatic (or ischemic) spinal cord injury, such as stroke, degenerative (e.g. myelopathy), pediatric
4	No prediction model or validation study of a previously described model (including linear regression models fitted to new data but excluding linear regression models with parameters from a previous paper.)
5	No recovery prediction
6	Different outcome measure

Table S1 Description of hierarchical exclusion criteria.

	Author	Dataset name	Number of patients	Models	Code publicly available	Outcome	Input	Evaluation	Prediction
1	Raju, P.R.S.S.V., et al.	Cancer Imaging Archive for MRI data	72	Correlated Graph model, KNN, SVM, RNN, Convolutional neural network *	-	Other	Imaging	Statistical Significance Metrics, Other Metrics	Accuracy 99%
2	Ariji, Y., et al.	Japan Single Center Study for Spinal Cord Injury Database (JSSCI-DB)	137	Linear regression	-	Functional (SCIM III)	Functional, Demographics, Neurological	Predictive Accuracy Metrics	R-Squared 0.83
3	Torres-Espín, A., et al.	-	118	Linear Regression	Yes	Neurological (AIS)	Neurological, Functional, Patient Demographics	Predictive Accuracy Metrics, Statistical Significance Metrics	Accuracy 73% , 0.82 %, 0.81%
4	Sizheng, Z., et al.	Different Japanese hospitals	249	XGBoost, Linear regression*	Yes	Functional (SCIM III)	Imaging, Neurological	Predictive Accuracy metrics	MAE 3.01
5	Kapoor, D., et al.	National Spinal Cord Injury Statistical Center (NSCISC)	20790	Random forest, SVM, Naive Bayes, Convolutional neural network, Ridge classifier, Elastic net, Ensemble model *	Yes	Neurological (AIS)	Neurological	Predictive Accuracy Metrics, Other Metrics	Accuracy 74%
6	Øhrn, A., et al.	Spinal Cord Injury Model System Database (SCIMS)	1138	Rough Sets, Logistic regression, Neural Network*	-	Functional (Walking)	Demographics, Neurological	Discrimination Metrics, Statistical Significance Metrics	ROC 91.4
7	Chou, A., et al.	Transforming Research and Clinical Knowledge for Spinal Cord Injury (TRACK-SCI) program	74	Regularized Logistic Regression, Decision Trees*	-	Neurological (AIS)	Neurological, Patient Demographics, Other	Discrimination Metrics, Other Metrics	AUC 0.68, AUC 0.67
8	Shimizu, T., et al.	-	135	CatBoost, LightGBM, XGBoost*	-	Neurological (AIS)	Neurological, Imaging	Predictive Accuracy Metrics, -	Accuracy 84%
9	Inoue, T., et al.	Sendai Medical Center's Department of Neurosurgery database,	165	XGBoost, Logistic regression, Decision Trees*	-	Neurological (AIS)	Neurological, Imaging	-	Accuracy 81%
10	Leister, I., et al.	Different German and Austrian Hospitals	415	Binary Logistic regression, XGBoost*	-	Bone healing (Odontoid fracture nonunion)	Discrimination Metrics, Statistical Significance Metrics	AUC 0.71, AUC 0.68	
11	Yang, F., et al.	Beijing hospital	1231	Random Forest, HH0 - Random Forest, XGBoost, SVM, Bayesian Ridge*	-	Functional (FIM)	Functional, Neurological, Patient Demographics	Predictive Accuracy Metrics	RMSE 0.13, RMSE 0.11
12	Smith, A.C., et al.	National Spinal Cord Injury Model Systems (SCIMS)	623	Regression analysis	-	Functional (Walking)	Neurological	Predictive Accuracy Metrics	Accuracy 81%
13	Skinnider, M.A. et al.	-	131	Ensemble model	-	Neurological (AIS, UEMS/LEM-S/TMS)	Neurological, Patient demographics, Other	Predictive Accuracy metrics	AUC 0.91
14	Belliveau, T., et al.	National Spinal Cord Injury Model Systems (SCIMS)	3142	Logistic regression, Neural Network*	-	Functional (Walking and FIM)	Neurological, Patient Demographics, Functional	Discrimination Metrics, Predictive Accuracy Metrics	AUC 0.86 - 0.90
15	Hug, A., et al.	EMSCI	444	Logistic regression	-	Neurological (AIS)	Neurological, Patient Demographics, Other	Predictive Accuracy Metrics, Statistical Significance Metrics	AUC 0.90
16	van Midden-dorp, Joost J., et al.	EMSCI	640	Logistic regression	-	Functional	Functional, Demographics	Predictive Accuracy Metrics	AUC 0.97
17	Javeed, S., et al.	Spinal Cord Injury Model System Database (SCIMS)	940	Logistic regression	-	Functional (FIM)	Patient demographics, Neurological, Functional	Predictive Accuracy Metrics	AUC 0.90
18	Pavese, C., et al.	EMSCI	1250	Logistic Regression	-	Functional (Bowel/Bladder)	Demographics, Neurological	Discrimination Metrics, Discrimination Metrics, Predictive Accuracy Metrics, Statistical Significance Metrics	ROC 0.81
19	Chen, Si, et al.	Hospital of Southwest Medical University, Chengdu 363 Hospital of Southwest Medical University	193	Logistic regression, Nomogram	-	Neurological (AIS)	Neurological, Imaging, Patient Demographics	Discrimination Metrics	C-index 0.89
20	Yan, X., et al.	-	475	Logistic regression, Nomogram*	-	Neurological (JOA score)	Patient Demographics, Neurological, JOA score, Imaging	Predictive Accuracy Metrics, Statistical Significance Metrics, Other Metrics	AUC 0.82
21	Rowland, T., et al.	Spinal Cord Injury Model System Database (SCIMS)	5626	Logistic regression, Multilayer perceptron, Rough sets*	-	Functional (Walking)	Patient Demographics, Neurological, Other	Discrimination Metrics	ROC 0.91, ROC 0.89
22	Kalyani, P., et al.	-	165	XGBoost, Logistic Regression, Decision trees*	-	Neurological (AIS)	Patient Demographics, Neurological, Imaging	Predictive Accuracy Metrics, Discrimination Metrics	AUC 0.81
23	Mills, P.B., et al.	Vancouver site of the national Rick Hansen SCI Registry (RHSCIR)	350	Logistic Regression	-	Neurological (Contracture/S-pasticity)	Patient demographics, Neurological	Predictive Accuracy Metrics	AUC 0.85
24	Harrington, G.M.B., et al.	Internal hospital dataset	417	Linear Regression, Generalized Linear models with elastic net penalization*	-	Functional (SCIM III), Neurological (AIS)	Neurological, Patient Demographics, Other (blood markers)	Predictive Accuracy Metrics, Statistical Significance Metrics	RMSE 11-18
25	Zariffa, J., et al.	-	129	Random forest	-	Functional (Prehension Performance)	Neurological, Functional	Statistical Significance Metrics	Spearman's ρ 0.84
26	Kato, C., et al.	-	210	Linear Regression, Decision Trees, Support Vector Machines, Neural Network*	-	Functional (SCIM)	Patient Demographics, Functional, Neurological	Predictive Accuracy Metrics, Statistical Significance Metrics, Other Metrics	RMSE 10-12/100
27	HeidarAbadi, et al.,	-	253	Neural Network	-	Neurological (Type of pain)	McGill Pain Questionnaire	Predictive Accuracy Metrics	Accuracy 91%
28	Bykowski, E.A. et al.,	-	7	Dimensionality reduction	-	Functional (SCIM III)	Other (Serum metabolites)	Predictive Accuracy Metrics	Accuracy 99%
29	Tanadini, G.L., et al.	EMSCI	122	Decision trees	-	Neurological (ISNCSCI)	Demographics, Neurological	Predictive Accuracy Metrics	RMSE 10.36
30	Tomioka, Y., et al.	-	31	Logarithmic model	-	Functional (SCIM III)	Demographics	Statistical Significance Metrics, Other Metrics	-
31	DeVries, Z., et al.	RHSCIR database	862	Logistic regression, K-means clustering*	-	Functional (FIM)	Demographics	Statistical Significance Metrics, Predictive Accuracy Metrics	Accuracy 0.87
32	Hupp, M., et al.	EMSCI	224	Logistic regression	-	Functional (SCIM III)	Neurological, Demographics	Discrimination Metrics, Predictive Accuracy Metrics	AUC 0.94

33	El Sammak, et al.	NSCIMS	2515	Chi-square Automatic Interaction Detector (CHAID)	-	Neurological (UEMS/LEM-S/TMS)	Demographics, Functional	Predictive Accuracy Metrics	AUC
34	Tee, J.W., et al.	Vancouver Rick Hansen Spinal Cord Injury Registry (RHSCIR)	806	Chi-square Automatic Interaction Detector (CHAID)	-	Neurological (UEMS/LEM-S/TMS)	Neurological	Statistical Significance Metrics	p-value <= 0.001
35	Imura, T., et al.	Japan Rehabilitation Database	310	Decision Trees	-	Functional (FIM)	Functional, Demographics	Predictive Accuracy Metrics	AUC 0.72
36	Hori, T., et al.	Japan Rehabilitation Database (JRD)	1282	Decision Trees	-	Neurological (AIS), Functional (FIM)	Neurological, Functional, Demographics	Statistical Significance Metrics, Other Metrics	AUC 0.72
37	Facchinello, Y., et al.	-	172	Linear regression, Decision Trees*	-	Functional (SCIM III)	Demographics	Predictive Accuracy Metrics	R ² 0.5
38	Kaminski, L., et al.	-	76	Linear regression	-	Functional (SCIM)	Neurological, Patient Demographics	Predictive Accuracy Metrics	R ² 0.37
39	Diong, J., et al.	-	92	Linear Regression	-	Neurological (Contracture)	Demographics, Neurological	Predictive Accuracy Metrics	R ² 0.31
40	Kato, et al.	-	210	Decision Trees	-	Other (Discharge destination)	Neurological, Demographics, Functional	Predictive Accuracy Metrics, Statistical Significance Metrics	AUC 0.832
41	Wilson, J.R., et al.	NACTN and STASCIS	696	Linear regression, Logistic regression*	-	Functional (FIM)	Imaging, Demographics	Discrimination Metrics	AUC 0.92
42	Pavese, C., et al.	EMSCI	1250	Logistic Regression	-	Functional (Bowel/Bladder)	Demographics, Neurological, Functional	Discrimination Metrics, Predictive Accuracy Metrics, Statistical Significance Metrics	ROC 0.837
43	Venkatapathiraju, P.R.S.S., et al.	-	-	Deep kernel based model	-	Other tissues) (Soft)	MRI	Predictive Accuracy Metrics, Statistical Significance Metrics	Accuracy 75%
44	Winchester, P., et al.	-	38	Linear regression	-	Functional (Walking)	Demographics	Other Metrics	-
45	Rigot, S.K., et al.	Hospital internal dataset	27	Random Forest	-	Functional (Walking)	Patient Demographics, Other, LA	Predictive Accuracy Metrics	Accuracy 70%
46	Okimatsu, S., et al.	-	215	Convolutional neural network and Random forest	-	Neurological (AIS)	Neurological, Imaging, Demographics	Discrimination Metrics, Predictive Accuracy metrics, Statistical Significance Metrics	Accuracy 71%
47	Birkhäuser, V. et al.	EMSCI	97	Logistic Regression	-	Functional (Bowel/Bladder/Urodynamic risk factors)	Neurological, Demographics, Functional	Predictive Accuracy metrics, Statistical Significance Metrics	AUC 0.79
48	Cathomen, A. et al.	EMSCI	481	URP-CTREE	Yes	Functional (Walking)	Neurological, Demographics	Statistical Significance Metrics	-
49	Pfyffer, D., et al.	EMSCI	70	URP-CTREE	-	Neurological (AIS), UEMS/LEMS/TMS), Functional(SCIM)	Neurological, Imaging, Patient Demographics	Statistical Significance Metrics	p-value 0.001
50	Cathomen, A. et al.	EMSCI	879	URP-CTREE	Yes	Functional (Walking)	Neurological, Demographics	Statistical Significance Metrics	p-value 0.01
51	Leidinger, A., et al.	-	284	Decision Trees	-	Neurological (AIS)	Imaging, Neurological, Demographics	Predictive Accuracy metrics	Accuracy 34%
52	Mummaneni, N., et al.	TRACK-SCI (Transforming Research and Clinical Knowledge in Spinal Cord Injury)	24	Linear Regression	-	Neurological (UEMS/LEM-S/TMS)	Imaging, Patient Demographics	Statistical Significance Metrics	p-value 0.001
53	Warschausky, S., et al.	Spinal Cord Injury Model Systems database	142	Linear regression	-	Functional (FIM)	Functional, Demographics, Neurological	Only feature importance	-
54	Ishida, Y., et al.	Tsushimi Hospital and Yamaguchi Rosai Hospital	22	Logistic regression	-	Neurological (PDI)	Neurological, Imaging	Statistical Significance Metrics	p-value 0.006
55	Kwon, S.Y., et al.	-	38	Logistic Regression	-	Neurological (AIS)	Neurological, Demographics, Other	Statistical Significance Metrics	-
56	Zhao, J.L., et al.	-	377	Logistic Regression	-	Neurological (AIS)	Neurological, Demographics, Other	Predictive Accuracy metrics, Statistical Significance Metrics	AUC 0.94
57	Ariji, Y., et al.	Japan Single Center study for Spinal Cord Injury Database (JSSCI-DB)	80	Linear Regression	-	Neurological (AIS)	Neurological, VAC/DAP	Predictive Accuracy metrics	AUC 0.87
58	Pavese, C., et al.	EMSCI	142	Logistic Regression	-	Functional (Bowel/Bladder)	Neurological	Predictive Accuracy metrics, Statistical Significance Metrics	AUC 0.780
59	Jean, S., et al.	-	159	Linear regression	-	Neurological, Functional (Walking)	Neurological, Patient Demographics	Predictive Accuracy Metrics	Accuracy 85%
60	Tee, J.W., et al.	Institution's spine trauma registry dataset	969	Logistic Regression	-	Functional (FIM/-Glasgow Outcome Scale)	Neurological, Demographics, Other	Predictive Accuracy Metrics, Statistical Significance Metrics	Accuracy 61%
61	Bersch, I., et al.	-	220	Ordinal regression	-	Neurological (ISNCSCI)	MP, Demographics	Other Metrics	-
62	Agarwal et al.	-	74	Decision trees	-	Neurological (AIS)	Demographics, Neurological, Other	Statistical Significance Metrics	-
63	Brown, S.J., et al.	Midlands Centre for Spinal Injuries (MCSI)	82	Logistic Regression	-	Neurological (AIS, Pin Prick)	Neurological, Demographics, Other	Predictive Accuracy Metrics, Statistical Significance Metrics	R ² 0.81
64	Moghaddam, A. et al.	BG Trauma Centre	30	Logistic Regression	-	Neurological (AIS)	Neurological, Demographics, Other	Predictive Accuracy Metrics, Statistical Significance Metrics	Accuracy 97%
65	Bykowski, E.A., et al.	-	6	Dimensionality Reduction	-	Functional (SCIM III)	Functional, Other (Urinary metabolites)	Predictive Accuracy Metrics, Statistical Significance Metrics	-
66	Varzi, D., et al.	Queen Elizabeth National Spinal Injury Unit (QENSIU)	25	Linear Regression	-	Functional (Bone healing/BMC/BMD)	Neurological, Demographics, Imaging	Predictive Accuracy Metrics, Statistical Significance Metrics	R ² 0.61

Table S2: Overview of the information retrieval from the systematic review. The order of articles is the same as in the heatmap in Figure 3B.

Metric Category	Metrics
Discrimination Metrics	ROCAUC (Receiver Operating Characteristic Area Under the Curve) AUPRC (Area Under the Precision-Recall Curve) Sensitivity/Specificity Positive/Negative Predictive Value Positive/Negative Likelihood Ratio
Predictive Accuracy Metrics	AUC (Area Under the Curve) RMSE (Root Mean Squared Error) or MSE (Mean Squared Error) MAE (Mean Absolute Error) Accuracy (Percent Error) F1 Score R-squared
Statistical Significance Metrics	P-value Statistical Tests (depends on the context and the statistical test used)
Other Metrics	Absolute/Relative Error Correlation Coefficient Odds Ratio Concordance Index Interpretability Analysis/Feature Importance Linear Mixed Model Pearson Correlation

Table S3 Evaluation metrics by category.

		<i>Clinical significance</i>	<i>Reproducibility</i>	<i>Generalizability</i>	<i>ML quality</i>	<i>ML performance</i>
1	Raju, P.R.S.S.V., et al.	0.5	0.5	1.5	3	2.5
2	Ariji, Y., et al.	4.5	4	3.5	4.5	4
3	Torres-Espín, A., et al.	3.5	4.5	3.5	4.5	4
4	Sizheng, Z., et al.	4.5	5	4	4	3.5
5	Kapoor, D., et al.	4.5	4.5	4.5	5	3
6	Øhrn, A., et al.	3	3	4	4.5	5
7	Chou, A., et al.	3	3	3	5	4
8	Shimizu, T., et al.	4.5	3	3.5	5	4
9	Inoue, T., et al.	4	3	3.5	5	4
10	Leister, I., et al.	3.5	3	4	5	3.5
11	Yang, F., et al.	4	3.5	4	5	3.5
12	Smith, A.C., et al.	4.5	3	5	3.5	3.5
13	Skinnider, M.A. et al.	5	4	5	4	4
14	Belliveau, T., et al.	4	3	5	5	4
15	Hug, A., et al.	3.5	3	4	3.5	4
16	van Middendorp, Joost J., et al.	4	3.5	5	4	4.5
17	Javeed, S., et al.	4	3	4.5	4	4.5
18	Pavese, C., et al.	4.5	3.5	4	4	4.5
19	Chen, Si, et al.	4.5	3	4.5	4	4
20	Yan, X., et al.	4.5	3	4	4	4
21	Rowland, T., et al.	4	3	4.5	4.5	4
22	Kalyani, P., et al.	4	3	4	4.5	4
23	Mills, P.B., et al.	4	3	4	4	4
24	Harrington, G.M.B., et al.	5	3	4	4.5	3.5
25	Zariffa, J., et al.	5	3	4	3.5	4
26	Kato, C., et al.	4.5	3	3.5	4	4
27	HeidarAbadi, et al.,	4.5	3	3	4	4
28	Bykowski, E.A. et al.	4.5	4	2	2.5	2.5
29	Tanadini, G.L., et al.	5	3	1.5	2.5	3
30	Tomioka, Y., et al.	4.5	3.5	2	3	3.5
31	DeVries, Z., et al.	4	3	2.5	3	2.5
32	Hupp, M., et al.	4	3	2	2.5	3
33	El Sammak, et al.	4.5	3	2.5	2.5	3
34	Tee, J.W., et al.	4.5	3	4	2.5	1.5
35	Imura, T., et al.	4	4	4	3	3
36	Hori, T., et al.	4	3.5	4	2.5	3
37	Facchinello, Y., et al.	4.5	3	3.5	3	2.5
38	Kaminski, L., et al.	4.5	3	3	2.5	2.5
39	Diong, J., et al.	4.5	3	3	2.5	2
40	Kato, et al.	4	3	3	3	3.5
41	Wilson, J.R., et al.	4.5	3	3.5	3	3.5
42	Pavese, C., et al.	4.5	3	3.5	2.5	3
43	Venkatapathiraju, P.R.S.S., et al.	3.5	2	2.5	4.5	3
44	Winchester, P., et al.	4.5	3	2.5	3.5	4
45	Rigot, S.K., et al.	4	3	2	4	4
46	Okimatsu, S., et al.	4	3.5	3	4.5	3
47	Birkhäuser, V. et al.	4	3.5	3	3.5	3
48	Cathomen, A. et al.	5	5	3.5	2.5	1.5
49	Pfyffer, D., et al.	4	4.5	0.5	0.5	0.5
50	Cathomen, A. et al.	4.5	5	2.5	0	0
51	Leidinger, A., et al.	3.5	3	3	1.5	0
52	Mummaneni, N., et al.	5	2.5	1	0.5	0.5
53	Warschausky, S., et al.	4.5	3	1.5	1	0.5
54	Ishida, Y., et al.	5	3	2	1	0.5
55	Kwon, S.Y., et al.	5	3	2	1	1
56	Zhao, J.L., et al.	3.5	4	2	1	1
57	Ariji, Y., et al.	4	4	1.5	1.5	1.5
58	Pavese, C., et al.	4	3.5	2.5	2	2
59	Jean, S., et al.	4	3	2	2	2
60	Tee, J.W., et al.	4	3	2	2	2
61	Bersch, I., et al.	3.5	3	2.5	2	1.5
62	Agarwal et al.	4	3	2.5	1.5	1.5
63	Brown, S.J., et al.	5	3	2	2	2
64	Moghaddam, A. et al.	5	3	1.5	2	2
65	Bykowski, E.A., et al.	4.5	3	1.5	1.5	1.5
66	Varzi, D., et al.	5	3	1	1.5	1.5

Table S4: Grades of the studies in the systematic search on the scale 0-5 using the mean of two scores (two raters). The order of articles is the same as in the heatmap in Figure 3B.