

Titre: Outcomes Prediction in Acute Cervical Traumatic Spinal Cord Injury from Clinical and Preoperative Magnetic Resonance Imaging-Based Models: The Role of Imaging Biomarkers
Title:

Auteur: Maxime Bouthillier
Author:

Date: 2025

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Bouthillier, M. (2025). Outcomes Prediction in Acute Cervical Traumatic Spinal Cord Injury from Clinical and Preoperative Magnetic Resonance Imaging-Based Models: The Role of Imaging Biomarkers [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. <https://publications.polymtl.ca/66529/>
Citation:

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/66529/>
PolyPublie URL:

Directeurs de recherche: Julien Cohen-Adad, & Laurent Létourneau-Guillon
Advisors:

Programme: Génie biomédical
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Outcomes Prediction in Acute Cervical Traumatic Spinal Cord Injury
from Clinical and Preoperative Magnetic Resonance Imaging-Based Models:
The Role of Imaging Biomarkers**

MAXIME BOUTHILLIER

Institut de génie biomédical

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie biomédical

Juin 2025

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Outcomes Prediction in Acute Cervical Traumatic Spinal Cord Injury
from Clinical and Preoperative Magnetic Resonance Imaging-Based Models:
The Role of Imaging Biomarkers**

présenté par **Maxime BOUTHILLIER**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Benjamin DE LEENER, président

Julien COHEN-ADAD, membre et directeur de recherche

Laurent LÉTOURNEAU-GUILLON, membre et codirecteur de recherche

Jean-Marc MAC-THIONG, membre

DEDICATION

*To my loved ones,
thank you for everything..*

*À mes proches,
merci pour tout. . .*

ACKNOWLEDGEMENTS

Many individuals and organizations contributed directly or indirectly to the realization of this work. I am deeply grateful for their support, which allowed me to engage meaningfully in research—an opportunity that has significantly enriched my development and broadened my competencies as a future radiologist. Pursuing a Master’s degree in biomedical engineering research alongside a diagnostic radiology residency has been a significant challenge—one that would not have been possible without meaningful support across multiple levels. The planning of this work began in the summer of 2021 (now spring 2025), just weeks after I began my residency, due to the substantial organizational effort it required. I sincerely appreciate all the collaborations and opportunities that made this journey possible.

First, I would like to thank the University of Montreal Department of Radiology, Radiation Oncology and Nuclear Medicine for their institutional support in curriculum scheduling, academic insight, and for allowing me protected time for research. Time is the most critical factor in the life of a resident physician, and this protected time was vital to make this implication a success. More specifically, I extend my gratitude to:

- Dre. Isabelle Trop, Director of the University of Montreal Radiology Residency Program, for her continuous support and guidance;
- Dr. Carl Chartrand-Lefebvre, Director of the Department of Radiology, Radiation Oncology and Nuclear Medicine, for his support and leadership;
- Dr. Gilles Soulez, for his dedication in promoting research within the field of radiology;
- The Radiology Program Competency Committee, for proposing an on-call policy tailored to residents in the Clinical Investigator Program;
- The administrative staff, for their invaluable assistance in management.

Second, I acknowledge the University of Montreal Clinical Investigator Program (PCC), which promotes high-level research training during and beyond residency. I recognize:

- Dr. Dang Khoa Nguyen, previous PCC Program Director, for his leadership in research;
- Dre. Catherine Larochelle, current PCC Program Director, for her insightful feedback;
- Sylvie Labelle, PCC Secretary, whose outstanding administrative skills and responsiveness deserve special recognition.

Third, I express my gratitude to the organizations that provided the financial support necessary for completing this research:

- The Radiological Society of North America Research & Education Foundation Resident Research Grant (#RR2313, 2023-2025), which made this project feasible;
- “Bourse de la Régie de l’Assurance-maladie du Québec (RAMQ) pour le programme de cliniciens-chercheurs de l’Université de Montréal”, providing salary complement during research periods (2023-2024, renewed for 2024-2025);
- “Bourse Fondation Marie-Robert NeuroTrauma pour la recherche sur les traumatismes crâniens” from the “Fondation Marie-Robert de l’Hôpital du Sacré-Coeur-de-Montréal” (2023-2024, renewed for 2024-2025);
- “Subvention *ad hoc* – Programme de recherche”, from the University of Montreal Department of Radiology, Radiation Oncology and Nuclear Medicine, which I intend to use to present this thesis work at academic conferences;
- “Bourses de la Financière des Professionnels pour l’excellence en recherche du Programme de cliniciens-chercheurs”, for which I was awarded in recognition of my research contributions during my residency.

Fourth, I thank the Praxis Spinal Cord Institute imaging substudy team, whose contributions were indispensable to the success and completion of this work (details in Appendix A). I would like to especially express my gratitude to Naama Rotem-Kohavi, Suzanne Humphreys, Vanessa Noonan, and Dr. David W. Cadotte for their invaluable insights and thoughtful feedback, as well as for making this collaboration such an enjoyable and rewarding experience.

Fifth, I thank my collaborators at the Neuropoly Lab, and more specifically Jan Valošek (for his complementary work and collaboration), Alexia Mahlig (shared coursework) and Kalum Ost (for his teaching and guidance in modeling). I also express my recognition to my colleagues in radiology residency for working with them in a positive way and their gratitude by awarding me the 2024 “Research Implication, voted by radiology residents” departmental honor. I also acknowledge Delphine Pilon and Kuan Yi Wang for their contributions to advancing the development of SlicerCART during my clinical rotations.

Finally, I would like to express my deepest gratitude to my research director and co-director Prof. Julien Cohen-Adad and Dr. Laurent Létourneau-Guillon: their supervision, thoughtful reflection, and constructive feedback have greatly shaped my research journey. It has been an honor to work under their guidance. I will carry forward many lessons from their mentorship, and it would be a true privilege if I could one day reach their level of scientific excellence.

RÉSUMÉ

Le présent mémoire se dédie aux lésions médullaires traumatiques (LTM), qui figurent parmi les conditions les plus dévastatrices d'un point de vue neurologique fonctionnel, avec des répercussions inestimables sur la qualité de vie des personnes atteintes. En l'absence de traitement actuellement reconnu pour renverser la paralysie des LTM, la sévérité initiale de l'atteinte constitue un déterminant central de l'orientation thérapeutique et du pronostic neurologique, dont la portée peut être limitée en phase aiguë.

L'imagerie par résonance magnétique (IRM) offre des perspectives intéressantes par le biais de biomarqueurs d'imagerie pouvant compléter l'évaluation clinique: leur utilité réelle demeure toutefois controversée. Des approches multimodales de prédiction du pronostic neurologique incorporant données cliniques et d'imagerie, notamment par apprentissage machine, demeurent encore peu exploitées dans les LTM cervicales aiguës préopératoires. Dans ce contexte, ce mémoire vise à évaluer la valeur pronostique des biomarqueurs IRM en lien avec la sévérité neurologique initiale, finale (au congé de la réadaptation) et son évolution.

La première partie se consacre au développement de SlicerCART, un module d'annotation d'images permettant la création d'étiquettes et de masques de segmentation, soulignant l'importance d'outils adaptés. À partir de ces masques, des biomarqueurs IRM sont extraits.

La deuxième partie traite de l'évaluation de la reproductibilité des biomarqueurs IRM acquis manuellement. Les valeurs obtenues sur les séquences sagittales sont comparées aux séquences axiales pour chaque méthode d'acquisition. Une analyse de la variabilité intra-observateur est également réalisée pour un sous-groupe de participants. Les résultats sont mitigés pour la concordance et reproductibilité des mesures manuelles.

La dernière partie intègre les biomarqueurs IRM aux données cliniques pour prédire le grade de l'*American Spinal Injury Association Impairment Scale* (AIS), la performance motrice totale, ainsi que celle des membres supérieurs et inférieurs, au congé de réadaptation (finale) et leur évolution depuis l'évaluation initiale. Les résultats montrent un rôle des biomarqueurs dans la prédiction des issues cliniques en LTM à partir de données préopératoires.

En somme, ce mémoire met en lumière : 1) la nécessité d'outils applicables à de grands jeux de données d'imagerie pour les tâches de segmentation, classification et de contrôle-qualité; 2) une reproductibilité modérée des biomarqueurs IRM dérivés de segmentations manuelles; 3) la performance des modèles de prédiction d'issues cliniques en LTM.

ABSTRACT

This thesis focuses on traumatic spinal cord injuries (tSCI), among the most devastating neurological conditions with profound impact on quality of life. Without any proven therapy capable of significantly reversing paralysis, the initial severity of the injury remains central for clinical decision-making and prognosis, although it may be limited in the acute phase.

Magnetic resonance imaging (MRI) presents promising avenues through imaging biomarkers that could complement the clinical assessment. However, the real-world prognostic value of these biomarkers remains controversial. Advanced multimodal prognostic models integrating clinical and imaging data—particularly those leveraging machine learning—are still under-used in the context of acute preoperative cervical tSCI. Thus, this thesis aims to evaluate the prognostic relevance and associations of MRI biomarkers in relation to initial injury severity, final severity at discharge from rehabilitation, and recovery trajectory.

The first part details the development of SlicerCART, a dedicated image annotation module enabling the creation of segmentation masks and classification labels, highlighting the importance of tailored tools for large-scale imaging studies. Ground-Truths MRI biomarkers are extracted from these manual segmentations.

The second part assesses the variability of MRI biomarkers derived from manual segmentations. Quantitative comparisons are conducted between sagittal and axial sequences values for each acquisition, reflecting moderate correlation. For a subset of subjects, intra-observer variability is analyzed: results indicate good agreement but moderate reproducibility.

The final section explores the integration of MRI biomarkers with clinical data to predict the *American Spinal Injury Association Impairment Scale* (AIS) grade, total motor score, as well as upper and lower extremity motor score at discharge, and their changes from baseline. Findings suggest that MRI biomarkers play a complementary role in outcomes prediction at rehabilitation discharge, using acute preoperative clinical and MRI data.

In summary, this thesis highlights: 1) the critical need for scalable, task-specific tools for segmentation, classification, and efficient quality-control in research studies using large imaging datasets; 2) the moderate reproducibility of MRI biomarkers derived from manual segmentation; and 3) the overall performance of prediction modeling in acute cervical tSCI.

TABLE OF CONTENTS

| | |
|--|------|
| DEDICATION | iii |
| ACKNOWLEDGEMENTS | iv |
| RÉSUMÉ | vi |
| ABSTRACT | vii |
| LIST OF TABLES | xi |
| LIST OF FIGURES | xiii |
| LIST OF SYMBOLS AND ABBREVIATIONS | xvi |
| LIST OF APPENDICES | xvii |
| CHAPTER 1 INTRODUCTION | 1 |
| 1.1 Problem Statement | 1 |
| 1.2 Objectives and Hypotheses | 1 |
| 1.3 Thesis Outline | 2 |
| CHAPTER 2 LITERATURE REVIEW | 3 |
| 2.1 Acute Traumatic Cervical Spinal Cord Injury | 3 |
| 2.1.1 Pathophysiology | 4 |
| 2.1.2 Epidemiology | 4 |
| 2.1.3 Injury Severity Classification Systems | 5 |
| 2.1.4 Role of Magnetic Resonance Imaging | 7 |
| 2.2 Ground-Truths in Medical Imaging Analysis | 8 |
| 2.2.1 Terminology | 9 |
| 2.2.2 Variability is Problematic | 10 |
| 2.2.3 Solutions for Efficient Manual Annotation: Lacking | 11 |
| 2.3 Approach in Prediction Models | 12 |
| 2.3.1 Regression Models | 13 |
| 2.3.2 Machine Learning Models | 13 |
| 2.3.3 Validation and the Role of Data Sampling Strategies | 15 |
| 2.3.4 Performance Assessment | 16 |
| 2.3.5 Previous Work in Acute Cervical Traumatic Spinal Cord Injury | 16 |

| | | |
|-----------|--|----|
| CHAPTER 3 | METHODOLOGY | 18 |
| 3.1 | Study Design | 18 |
| 3.1.1 | Inclusion and Exclusion Criteria | 18 |
| 3.1.2 | Sample Size | 19 |
| 3.2 | Objective 1: To Develop a Software Ecosystem Enhancing Manual Annotation Tasks in Imaging Datasets | 20 |
| 3.2.1 | Software Selection for Efficient Segmentation and Classification | 20 |
| 3.2.2 | Manual Annotation: Standard Operating Procedure | 26 |
| 3.3 | Objective 2: To Measure Variability of MRI Biomarkers Extracted from Manual Segmentation | 29 |
| 3.3.1 | Quantitative Metrics | 29 |
| 3.3.2 | Qualitative Labels | 29 |
| 3.3.3 | Hemorrhage in Acute Traumatic Spinal Cord Injury | 33 |
| 3.3.4 | Reliability Assessment | 34 |
| 3.4 | Objective 3: To Develop Outcome Prediction and Classification Models From Clinical and MRI Metrics | 35 |
| 3.4.1 | Descriptive Statistics: Contextualization | 35 |
| 3.4.2 | Predictive Modeling and Classification | 37 |
| CHAPTER 4 | RESULTS | 42 |
| 4.1 | SlicerCART | 42 |
| 4.2 | Imaging Biomarkers Variability | 50 |
| 4.3 | Predictive Modeling and Outcomes Classification | 52 |
| 4.3.1 | RHSCIR Imaging Substudy Cohort Profile | 52 |
| 4.3.2 | Univariate Analyses | 56 |
| 4.3.3 | Model Performance Evaluation | 61 |
| CHAPTER 5 | DISCUSSION | 71 |
| 5.1 | SlicerCART | 71 |
| 5.1.1 | Scope of realizations | 72 |
| 5.1.2 | Future Developments | 74 |
| 5.2 | MRI Biomarkers | 74 |
| 5.2.1 | Axial Versus Sagittal | 74 |
| 5.2.2 | Automatic Versus Manual | 75 |
| 5.2.3 | Intra-Rater Variability | 76 |
| 5.2.4 | Intramedullary Hemorrhage and Outcomes | 77 |
| 5.2.5 | Future Analyses | 77 |

| | | |
|------------|--|-----|
| 5.3 | Predictive Modeling | 77 |
| 5.3.1 | Cohort Profile: Selection Bias? | 77 |
| 5.3.2 | Associations are not Predictions | 79 |
| 5.3.3 | Performance Assessment | 79 |
| 5.3.4 | The Role of Imaging Biomarkers | 84 |
| CHAPTER 6 | CONCLUSION | 85 |
| REFERENCES | | 87 |
| APPENDICES | | 110 |

LIST OF TABLES

| | | |
|-----------|---|-----|
| Table 2.1 | Comparison between the ASIA Impairment Scale and the Frankel Classification | 6 |
| Table 3.1 | Description of Central Files and Folders in the SlicerCART Repository | 21 |
| Table 3.2 | Description of SlicerCART Module Core Classes | 23 |
| Table 3.3 | Complementary Python Files that Support SlicerCART | 24 |
| Table 3.4 | Listing of MRI Labels for Sagittal T2-Weighted Manual Classification Tasks | 32 |
| Table 3.5 | Listing of Demographic, Clinical, and MRI Features | 36 |
| Table 3.6 | Listing of Outcomes Used for Model Development | 40 |
| Table 4.1 | Intra-Rater Reliability in Hemorrhage Binary Classification 6-months Apart on Sagittal T2-weighted MRI and 3-months Apart on Axial T2* MRI | 51 |
| Table 4.2 | Frequency of Significant Association Grouped Per Variable Category Using $p = 0.05$ | 56 |
| Table 4.3 | Summary of Sample Size and Selected Predictive Features by Outcome and Subpopulation Based on <i>American Spinal Injury Association Impairment Scale</i> Grade at Initial Admission | 70 |
| Table B.1 | Listing of Existing Open-Source 3D Slicer Module or Software for Manual Segmentation and/or Classification Tasks (Page 1) | 112 |
| Table B.2 | Listing of Existing Open-Source 3D Slicer Module or Software for Manual Segmentation and/or Classification Tasks (Page 2) | 113 |
| Table B.3 | Listing of Existing Open-Source 3D Slicer Module or Software for Manual Segmentation and/or Classification Tasks (Page 3) | 114 |
| Table B.4 | Summary of Literature Review for Open-Source Solutions in Manual Segmentation and/or Classification Tasks (Page 1) | 115 |
| Table B.5 | Summary of Literature Review for Open-Source Solutions in Manual Segmentation and/or Classification Tasks (Page 2) | 116 |
| Table B.6 | Summary of Literature Review for Open-Source Solutions in Manual Segmentation and/or Classification Tasks (Page 3) | 117 |
| Table B.7 | Summary of Literature Review for Open-Source Solutions in Manual Segmentation and/or Classification Tasks (Page 4) | 118 |
| Table G.1 | Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 1) | 140 |

| | | |
|-----------|---|-----|
| Table G.1 | Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 2) | 141 |
| Table G.1 | Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 3) | 142 |
| Table G.1 | Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 4) | 143 |
| Table G.1 | Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 5) | 144 |
| Table G.1 | Univariate Analyses Results for ≤ 0 vs 0 Change in Severity (Page 1) | 145 |
| Table G.1 | Univariate Analyses Results for ≤ 0 vs 0 Change in Severity (Page 2) | 146 |
| Table G.1 | Univariate Analyses Results for ≤ 0 vs 0 Change in Severity (Page 3) | 147 |

LIST OF FIGURES

| | | |
|-------------|---|----|
| Figure 2.1 | Thesis Context Overview | 3 |
| Figure 2.2 | What are Segmentation and Classification? | 9 |
| Figure 2.3 | Contour Delineation: Right or Wrong? | 10 |
| Figure 3.1 | <i>Rick Hansen Spinal Cord Injury Registry</i> Imaging Substudy and Thesis Flowchart | 19 |
| Figure 3.2 | Overview of SlicerCART Module Code Structure | 22 |
| Figure 3.3 | Example of a Pop-Up Window Notifying Incompatibility Work Envi- ronment | 25 |
| Figure 3.4 | Illustrated Examples of Quantitative MRI Metrics Measured | 30 |
| Figure 3.5 | Illustrated Examples of Qualitative MRI Labels Classified | 31 |
| Figure 3.6 | Suspicion and Confirmation of Intramedullary Hemorrhage on MRI . | 33 |
| Figure 3.7 | Outcome-Dependent Prediction Models Development Summary (Ob- jective 3) | 41 |
| Figure 4.1 | Overview | 43 |
| Figure 4.2 | Configuration Set-Up | 44 |
| Figure 4.3 | Label Selection | 45 |
| Figure 4.4 | Label Standardization | 46 |
| Figure 4.5 | User Interface | 47 |
| Figure 4.6 | Use Case Example: Manual Correction of Automated Segmentation Masks | 48 |
| Figure 4.7 | Work Environment | 49 |
| Figure 4.8 | MRI Metrics Values Calculated on Axial Versus Sagittal View for Man- ual and Automatic Segmentation Methods | 50 |
| Figure 4.9 | Distribution of Dice Scores for Non-Empty Masks at 6-months Interval | 51 |
| Figure 4.10 | Imaging Substudy Cohort Characteristics and Visual Summary . . . | 53 |
| Figure 4.11 | Subgroup Analysis: MRI versus No MRI | 54 |
| Figure 4.12 | Initial and Final <i>American Spinal Injury Association Impairment Scale</i> (AIS) Grades at Initial Evaluation and Final Discharge from Rehabil- itation | 54 |
| Figure 4.13 | Rick Hansen Imaging Substudy Cohort Profile: Severity Assessment . | 55 |
| Figure 4.14 | Top 15 Most Significant Associations for AIS Grade | 57 |
| Figure 4.15 | Top 15 Most Significant Associations for Total Motor Score | 58 |
| Figure 4.16 | Top 15 Most Significant Associations for Upper Extremity Motor Score | 59 |

| | | |
|-------------|---|-----|
| Figure 4.17 | Top 15 Most Significant Associations for Lower Extremity Motor Score | 60 |
| Figure 4.18 | Outcomes Baseline Modeling Results Using Full Available Data at Final Discharge, Change from Baseline, and Change ≤ 0 Versus > 0 in Scores | 62 |
| Figure 4.19 | Top 10 Feature Coefficients Magnitude in Predictive Modeling Using Full Available Data for Outcomes Prediction at Final Discharge, Change from Baseline, and Change (≤ 0 versus > 0) in Scores | 63 |
| Figure 4.20 | Model Performance for Outcomes Prediction and Classification at Final Discharge, Change from Baseline, and Change (≤ 0 Versus > 0) . . . | 64 |
| Figure 4.21 | Assessment of <i>American Spinal Injury Association Impairment Scale</i> Grade Prediction and Classification at Final Discharge, Change from Baseline, and Change from Baseline ≤ 0 Versus > 0 | 65 |
| Figure 4.22 | Assessment of Total Motor Score Prediction at Final Discharge, and Change from Baseline | 66 |
| Figure 4.23 | Assessment of Upper Extremity Motor Score Prediction at Final Discharge, and Change from Baseline | 67 |
| Figure 4.24 | Assessment of Lower Extremity Motor Score Prediction at Final Discharge, and Change from Baseline | 68 |
| Figure 4.25 | The Role of Magnetic Resonance Imaging Biomarkers in Acute Preoperative Cervical Traumatic Spinal Cord Injury | 69 |
| Figure 5.1 | Comparison of Manual and Automatic Segmentation on Sagittal T2-weighted MRI | 75 |
| Figure B.1 | Results for Available Code Searching Related to Manual Segmentation and/or Classification Tools (Page 1) | 119 |
| Figure B.2 | Results for Available Code Searching Related to Manual Segmentation and/or Classification Tools (Page 2) | 120 |
| Figure C.1 | STROBE Statement Filled for the <i>Rick Hansen Spinal Cord Injury Registry</i> Imaging Substudy and Thesis (Page 1) | 121 |
| Figure C.2 | STROBE Statement Filled for the <i>Rick Hansen Spinal Cord Injury Registry</i> Imaging Substudy and Thesis (Page 2) | 122 |
| Figure C.3 | STROBE Statement Filled for the <i>Rick Hansen Spinal Cord Injury Registry</i> Imaging Substudy and Thesis (Page 3) | 123 |
| Figure E.1 | TRIPOD-AI 2024 Abstract Checklist Filled for the <i>Rick Hansen Spinal Cord Injury Registry</i> Imaging Substudy Thesis Predictive Modeling . | 132 |
| Figure E.2 | TRIPOD-AI 2024 Checklist Filled for the <i>Rick Hansen Spinal Cord Injury Registry</i> Imaging Substudy Thesis Predictive Modeling (Page 1) | 133 |

| | | |
|------------|---|-----|
| Figure E.3 | TRIPOD-AI 2024 Checklist Filled for the <i>Rick Hansen Spinal Cord Injury Registry</i> Imaging Substudy Thesis Predictive Modeling (Page 2) | 134 |
| Figure F.1 | PROBAST Guidelines – Pages 1 and 2 | 135 |
| Figure F.2 | PROBAST Guidelines – Pages 3 and 4 | 136 |
| Figure F.3 | PROBAST Guidelines – Pages 5 and 6 | 137 |
| Figure F.4 | PROBAST Guidelines – Pages 7 and 8 | 138 |

LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|----------|---|
| AIS | American Spinal Injury Association Impairment Scale |
| AUC | Area Under the Receiver Operating Characteristic Curve |
| BASIC | Brain and Spinal Injury Center |
| BIDS | Brain Imaging Data Structure |
| CI | Confidence Intervals |
| FIM | Functional Independence Measure |
| GUI | Graphical User Interface |
| ISNCSCI | International Standards for Neurological Classification of Spinal Cord Injury |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| LEMS | Lower Extremity Motor Score |
| MCC | Maximum Cord Compression |
| MRI | Magnetic Resonance Imaging |
| MSCC | Maximum Spinal Canal Compromise |
| PRESS | Predicted Residual Error Sum of Squares |
| PROBAST | Prediction Model Risk-of-Bias Assessment Tool |
| PROGRESS | Prognosis Research Strategy |
| RHSCIR | Rick Hansen Spinal Cord Injury Registry |
| R^2 | Coefficient of Determination |
| SCIM | Spinal Cord Independence Measure |
| STROBE | Strengthening the Reporting of Observational Studies in Epidemiology |
| tSCI | Traumatic Spinal Cord Injury |
| TRIPOD | Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis |
| UEMS | Upper Extremity Motor Score |

LIST OF APPENDICES

| | | |
|------------|---|-----|
| Appendix A | Detailed Acknowledgments about the Rick Hansen Spinal Cord Injury Registry (RHSCIR) | |
| | Sites and Participants | 110 |
| Appendix B | Existing Open-Source Solutions for Imaging Analysis | 111 |
| Appendix C | STROBE Statement | 121 |
| Appendix D | Model Hyperparameters | 124 |
| Appendix E | TRIPOD-AI, Checklists | 132 |
| Appendix F | PROBAST | 135 |
| Appendix G | Univariate Analyses for Demographics, Clinical and MRI Data | 139 |

CHAPTER 1 INTRODUCTION

1.1 Problem Statement

Traumatic cervical spinal cord injury (tSCI) [1, 2] is among the most disabling neurological conditions per individual [3, 4], with profound impact on quality of life [5–7]. In the absence of any proven therapy capable of significantly reversing the paralysis resulting from severe motor dysfunction [8], the initial severity of the injury remains central for clinical decision-making and early prognostication [9]. The expected recovery from early prognostication may also guide in prioritizing patients for investigational therapies. While the *American Spinal Injury Association Impairment Scale* [10, 11] (AIS) grade remains a gold standard for assessment of the severity, it has limitations and may be suboptimal in the early acute period [12, 13].

Magnetic resonance imaging (MRI) presents promising avenues through imaging biomarkers that could complement traditional clinical assessment in tSCI [14, 15]. While clinical indications define the diagnostic use of MRI in tSCI [1, 16–20], some systematic reviews increasingly highlight the prognostic value of MRI-based biomarkers [21–23]. Furthermore, recent consensus guidelines advocate for the integration of MRI biomarkers into clinical trials targeting spinal cord injury [16, 24]. However, the real-world prognostic value of these biomarkers remains controversial [25, 26]. Advanced multimodal prognostic models integrating clinical and imaging data—particularly machine learning—are still underused in acute preoperative cervical tSCI (most studies focus on postoperative imaging). Thus, this thesis aims to evaluate the prognostic relevance and associations of MRI biomarkers with initial (acute setting) and final (discharge from rehabilitation) injury severity, along with recovery trajectory. The developed prediction models are intended to support clinicians and researchers in tSCI.

1.2 Objectives and Hypotheses

The overall objective is to improve neurological outcomes prediction at rehabilitation discharge using clinical and preoperative MRI data. The specific aims (**O**) are:

- O1.** To develop a software ecosystem enhancing manual annotation tasks in imaging datasets.
- O2.** To measure variability of MRI biomarkers extracted from manual segmentation.
- O3.** To develop outcome prediction and classification models from clinical and MRI metrics.

The related hypotheses (**H**) and sub-hypotheses (**H**) are:

H1. An optimized platform for manual segmentation, classification, and quality-control tasks in large imaging datasets will enable annotation times less than 5 minutes per subject in acute cervical preoperative spinal cord injury lesion on MRI.

H2. MRI biomarkers extracted from manual segmentation will have good agreement.

H2.1 The correlation coefficient between quantitative MRI biomarkers extracted from segmentations will be ≥ 0.60 when comparing intra-subject sagittal and axial values;

H2.2 The mean Dice coefficient will be ≥ 0.70 for intra-rater manual segmentation masks across two independent annotation sessions;

H2.3 The kappa coefficient (κ) for manual qualitative MRI biomarkers will be ≥ 0.60 .

H3. Incorporating MRI biomarkers in prediction and classification modeling will increase performance in comparison to clinical variables-only based models.

H3.1 The acute cervical tSCI population with preoperative MRI and without have different clinical characteristics, which define the context's applicability of the findings;

H3.2 MRI biomarkers extracted from both manual and automatic segmentation masks are significantly associated with initial, final, and change from baseline severity;

H3.3 Statistical and machine-learning models will accurately predict ($R^2 \geq 0.70$) and classify (AUC or $Accuracy \geq 0.80$) discharge and change in neurological severity;

H3.3.1 Feature importance will highlight MRI biomarkers.

The original research contribution of this work relates to the large multicentric sample size that incorporate clinical and preoperative MRI data in outcomes prediction modeling.

1.3 Thesis Outline

This thesis is structured as follows. Chapter 2 presents the concepts that support the rationale for this work through a review of the current literature. Chapter 3 describes the methodology employed to address the research objectives and hypotheses. The results and principal findings are presented in Chapter 4, followed by a detailed interpretation and discussion in Chapter 5. Chapter 6 concludes by providing key messages and recommendations for future work. Supplementary materials are provided in the Appendices.

CHAPTER 2 LITERATURE REVIEW

This chapter provides the theoretical foundations that contextualize and support this work: Figure 2.1 illustrates its rationale. The chapter begins with an overview of acute traumatic cervical spinal cord injury, linking its pathophysiology and epidemiology to current severity classification systems. The complementary role of MRI-based biomarkers is then introduced, emphasizing the importance and challenges of establishing reliable ground truths as a foundation for further analysis. The chapter concludes with a review of predictive modeling approaches, since early prognostication may help overcome limitations of initial assessment.

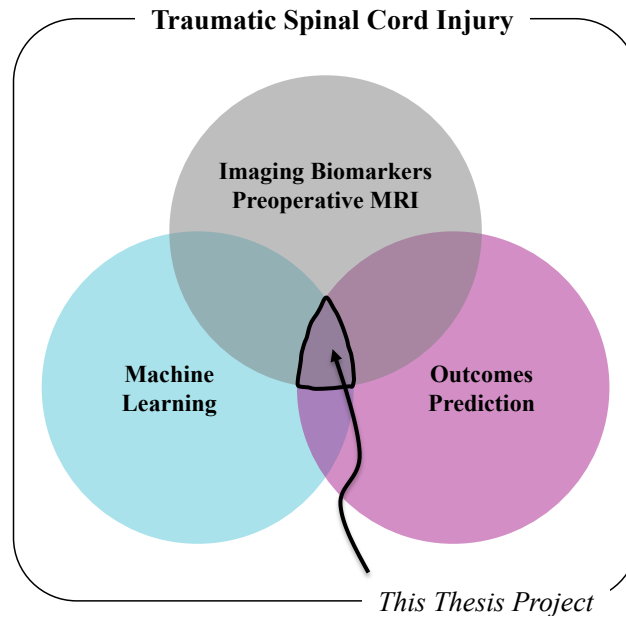


Figure 2.1 Thesis Context Overview

2.1 Acute Traumatic Cervical Spinal Cord Injury

According to Ahuja et al. [1], spinal cord injury “is defined as damage to the spinal cord that temporarily or permanently causes changes in its function” and “is divided into traumatic and non-traumatic aetiologies”. Traumatic spinal cord injury (tSCI) “occurs when an external physical impact (for example, a motor vehicle injury, fall, sports-related injury or violence) acutely damages the spinal cord, whereas non-traumatic [spinal cord injury] occurs when an acute or chronic disease process, such as a tumour, infection or degenerative disc disease, generates the primary injury.” Thereby, the scope of this work is limited to individuals with spinal cord injury acquired through a sudden external event.

2.1.1 Pathophysiology

The spinal cord is part of the central nervous system, and its function resumes to transmit afferently and efferently volunteer and unvolunteerly sympathetic, parasympathetic, sensory and locomotion signal information between the brain and other organs [27]. It contains central grey matter surrounded by peripheral white matter and acts as an intermediary between the peripheral nervous system and the brain [28]. Motricity enables locomotion [29] which yeilds functional living capabilities. Rostro-caudal organization of the spinal cord makes that more neural pathways are present in rostral (e.g. cervical) spinal cord than caudal (e.g. thoracic), which implies that cervical spinal cord are likely more disabling than thoracic lesions. From a pathophysiological standpoint, a cervical spinal cord lesion is likely to significantly impact the expected functions of the human body, and that is reflected in its epidemiology. Moreover, Bennett et al. [2] delineate the temporal progression of spinal cord injury into distinct phases—early acute, acute, subacute, intermediate, and chronic—to its pathophysiological mechanism. In their framework, early acute is defined as the first two to 48 hours post-injury, and acute phase is defined as the first two hours to two weeks, respectively. Thereby, this work focuses on acute tSCI (see Subsection 3.1.1 in Methodology).

2.1.2 Epidemiology

In Canada, the prevalence of traumatic spinal cord injury (tSCI) was estimated in 2019 to be 30,239 cases, corresponding to approximately 804 per 1,000,000 individuals, with 53% presenting paraplegia and 48% tetraplegia [30]. That same year, the incidence of hospital discharge for tSCI was estimated at 1,076 cases, or roughly 29 per 1,000,000 person-years [30]. In the United States, a higher incidence of 54 per 1,000,000 person-years has been reported, with prevalence ranging from 721 to 906 per 1,000,000 individuals as of 2016 [31]. Globally, variations in incidence have been observed between developed and developing countries [32], alongside an increasing trend in overall incidence from 1986 to 2016 [31]. The estimated lifetime cost per individual with tSCI ranges from \$1.5 million for incomplete paraplegia to \$3 million for complete tetraplegia [3]. In Canada, this translates to an estimated annual societal burden of approximately \$2.67 billion. Quality of life is frequently diminished following tSCI [5–7]. Although tSCI is a rare condition, it imposes a significant socioeconomic burden due to the lasting and severe disabilities it causes.

2.1.3 Injury Severity Classification Systems

The injury severity in tSCI reflects the expected chances of recovery, and complementary severity classification systems have been proposed accordingly. Around World War II, Stokes Manville Hospital in England [33] introduced a five-category classification system for spinal injuries, which was later formalized as the Frankel Classification in 1969 [34]. This classification largely inspired the currently widely used *American Spinal Injury Association Impairment Scale* (AIS) [10]. The *International Standards for Neurological Classification of Spinal Cord Injury* (ISNCSCI) [11] have also been proposed and widely adopted for detailed clinical assessment. In that context, one of the most foundational classification element is the presence or absence of perianal sensory function [35, 36], which primarily informs the chances of recovery (referring to complete [no function] or incomplete [function preserved]). The AIS grade and Frankel classification systems have been compared regarding this completeness aspect and their diagnostic stability [37], with the first showing superiority, thereby contributing to its wider use (AIS grade relies on the ISNCSCI classification). ~20% of patients with initial AIS grade A will spontaneously convert to AIS grade B or C [38]. For patients classified as AIS grade B or C, 60–80% may show spontaneous neurological improvement [38–40], while AIS grade D patients “usually improve substantially” [36].

A strength of the ISNCSCI classification system is its interpretability and ease of communication. However, it does not incorporate advances in the understanding of tSCI [15]. For large-scale clinical trials, it is essential to stratify patients based on their individual risk of recovery [15], but current clinical classification systems are not designed to do so. Although widely used, the ISNCSCI classification remains somewhat subjective due to the nature of the examination. Furthermore, the ISNCSCI exam was designed for clinical use, not specifically for research stratification [36, 41]. The reliability of the classification also depends on the timing of the examination. For instance, Brown et al. [42] showed that an examination performed at 24 hours is less reliable than one performed at 72 hours in predicting 3-month recovery. Additionally, long-term walking ability is poorly predicted in AIS grade B or C patients [43]. It is worth noting that the ISNCSCI classification has been revised several times, most recently in 2019 [11, 44], and some subjects still remain difficult to classify [13].

Another important consideration when referring to spinal cord injury is the level of injury, which is clinically defined “as the most caudal segment with motor function rated at greater than or equal to 3/5, with pain and temperature preserved” [35]. This informs the expected functional and motor disabilities related to the injury. Functional assessment tools such as the *Functional Independence Measure* (FIM) [45] and the *Spinal Cord Independence Measure* (SCIM) [46] are primarily used during the rehabilitation process rather than for severity

classification. In brief, when referring to the severity of spinal cord injury, both the injury level and whether the lesion is complete or incomplete (e.g. via AIS) are usually communicated to patients, as this informs about disability and prognosis, although outcomes can vary. In the pursuit of improved clinical trial design, it has also been proposed that advanced imaging techniques could support better patient stratification in tSCI [15]. Many have noted that MRI has introduced imaging relevance in the care of patients with tSCI [47], but as discussed in this section, the most commonly clinically-used classification systems do not currently incorporate imaging data in their evaluation. There is a perpetual need for improved severity classification systems in tSCI, as reflected by the 2025 prognostication challenge launched by the *American Spinal Injury Association* [48]. Accordingly, exploring classification approaches that integrate imaging data—and potentially other multimodal inputs such as biochemical biomarkers [49–51] is worthwhile! The Table 2.1 cites both classification from the current ASIA and original Frankel systems:

| ASIA Impairment Scale (AIS) [10] | Frankel Classification [34] |
|---|---|
| A=Complete. No sensory or motor function is preserved in the sacral segments S4-5. | A=Complete. This means that the lesion was found to be complete both motor and sensory below the segmental level marked. If there was an alteration of level but the lesion remained complete below the new level, then the arrow would point up or down the 'complete' column. |
| B=Sensory Incomplete. Sensory but not motor function is preserved below the neurological level, including S4–S5 AND no motor function is preserved more than three levels below the motor level on either side of the body. | B=Sensory only. This implies that there was some sensation present below the level of the lesion but that the motor paralysis was complete below that level. This column does not apply when there is a slight discrepancy between the motor and sensory level but does apply to sacral sparing. |
| C=Motor Incomplete. Motor function is preserved at the most caudal sacral segments for voluntary anal contraction OR the patient meets the criteria for sensory incomplete status, and has some sparing of motor function more than three levels below the ipsilateral motor level on either side of the body [...] less than half of key muscle functions below the single NLI have a muscle grade ≥ 3 . | C=Motor Useless. This implies that there was some motor power present below the lesion but it was of no practical use to the patient. |
| D=Motor Incomplete. Motor incomplete status as defined above, with at least half [...] of key muscle functions below the single NLI having a muscle grade ≥ 3 . | D=Motor Useful. This implies that there was useful motor power below the level of the lesion. Patients in this group could move the lower limbs and many could walk, with or without aids. |
| E=Normal. If sensation and motor function as tested with the ISNCSCI are graded as normal in all segments, and the patient had prior deficits, then the AIS grade is E. Someone without an initial SCI does not received an AIS grade. | E=Recovery. This implies that the patient was free of neurological symptoms, i.e. no weakness, no sensory loss, no sphincter disturbance. Abnormal reflexes may have been present. |

Table 2.1 Comparison between the ASIA Impairment Scale and the Frankel Classification

2.1.4 Role of Magnetic Resonance Imaging

MRI findings are not included in widely used clinical classification systems, but there are recognized clinical indications to perform cervical MRI in acute tSCI. Mainly, the role of MRI is to assess the spinal cord, identify the presence of intramedullary lesions (signal abnormalities), determine the lesion’s etiology, identify the neurological level(s) involved, and evaluate the extent of spinal cord compression. Other complementary and specific diagnostic indications include: suspected ligamentous injury on radiograph/computed tomography, identification of epidural hematoma or disc herniation (“before attempting a closed reduction of cervical facet dislocations” [17]), assessment of spinal cord pathology in patients with neurological deficits (e.g. edema, hemorrhage [for prognosis], cord transection), evaluation of spine stability, exclusion of occult bony injuries, and detection of vascular injuries such as spinal cord infarction [1, 17–20]. Except for identifying the presence of intramedullary hemorrhage, these indications remain primarily for **diagnostic** purposes rather than for **prognostication**. Consequently, while MRI is often routinely performed in acute tSCI, the level of evidence supporting its use as a baseline [26] and prognostic value [52, 53] remain low.

Imaging Biomarkers

Since its initial use for tSCI in the early 1990s [54], MRI has been investigated for its potential role in outcome prediction. Numerous studies have reported associations between specific MRI features and neurological outcomes [55–58]. Imaging-based severity classification systems, such as the *Brain and Spinal Injury Center* (BASIC) score [59], have been developed to provide imaging classification-based assessment. The concept of imaging biomarkers—“defined as anatomic, physiologic, biochemical, or molecular parameters detectable with imaging methods used to establish the presence or severity of disease” [60]—has emerged over the years and increasingly gained traction for applications in tSCI [14], including patient selection in clinical trials [24] and longitudinal tracking of neurodegeneration [61, 62].

Examples of spinal cord imaging biomarkers include (not exhaustively): maximum canal compromise (MSCC) [58, 63], maximum cord compression (MCC) [58, 63], lesion volume [64], lesion length [65], and axial damage ratio [66] (illustrated examples are shown in Subsection 3.3.1 in Chapter 3). While many of these biomarkers are used interchangeably in different spinal cord pathologies, certain like tissue bridges [67, 68] are more specific to tSCI.

Most recently, the concept of Radiomics [69–71]—quantitative/mathematical features extraction from medical images (initially studied for tumors)—has emerged as an additional tool to enrich imaging analysis and facilitate integration with multimodal data—e.g. clinical or

biological. This has introduced the concept of shape, first-order, second-order and higher-order [72] imaging features (where lesion volume is a typical example of a shape-related feature). Despite its promising role, the application of Radiomics remains exploratory and limited [73,74], especially in tSCI. When such first-order quantitative biomarkers are derived from manually acquired segmentations, for example MCC and MSCC, intra and inter-rater variability are reported [75]. Such variability can directly impact reproducibility, and potentially limit or complicate their clinical utility.

Importantly, while many studies describe associations between MRI findings and outcomes, methodological concepts are often confused. Association and prediction serve distinct purposes: the former seeks to understand relationships between variables, while the latter aims to forecast future outcomes based on data patterns [76,77]. The two are frequently conflated. For instance, Tarawneh et al. [78] use the terms interchangeably; in Bozzo et al. [79], MRI “prognostication” is discussed without quantifying predictive performance. When imaging biomarkers are evaluated within true predictive frameworks [25,80], their added value over clinical examination often becomes minimal or nonexistent. It is to note that prediction models that combined MRI and clinical data have been attempted [81], yet these studies typically relied on basic MRI descriptors (e.g. presence of cord edema) and lacked advanced feature extraction or modern predictive modeling approaches such as machine learning. Furthermore, heterogeneity in timing (e.g. preoperative versus postoperative imaging), and disease phase affect imaging appearance and interpretation (see Figure 1 in [24] for a visual example): this seems to not be consistently reported in tSCI imaging studies [14]

Overall, imaging biomarkers encompass a broad range of metrics with potential to improve diagnosis and prognosis, though their clinical utility remains limited by various challenges. While imaging biomarkers present theoretical value, their practical utility remains controversial [82] and yet minimally explored. Indeed, the role of imaging biomarkers in acute preoperative tSCI is still worth to investigate!

2.2 Ground-Truths in Medical Imaging Analysis

Entire PhD theses—and even full careers—can be devoted to the subject of ground-truths and variability in medical imaging analysis. Nevertheless, understanding practical concepts is essential to grasp how imaging biomarkers are derived and generated. This section first defines common terminology used in the field, then highlights the problematic surrounding ground-truth variability, and finally outlines current limitations and gaps in manual annotation workflows. These issues form the basis for research contributions of this thesis.

2.2.1 Terminology

The extraction of imaging biomarkers usually requires image labeling, whether via segmentation or classification. In medical image analysis, segmentation is “typically defined as identifying the set of voxels which make up either the contour or the interior of the object(s) of interest” [83]. In other words, it refers to the annotation of a region of interest (ROI), which may involve anatomical structures or pathologies (e.g. spinal cord injury lesions). Concepts of semantic and instance segmentation [84] go beyond the scope of the current intended analyses. Figure 2.2 resumes segmentation and classification definition for this work:

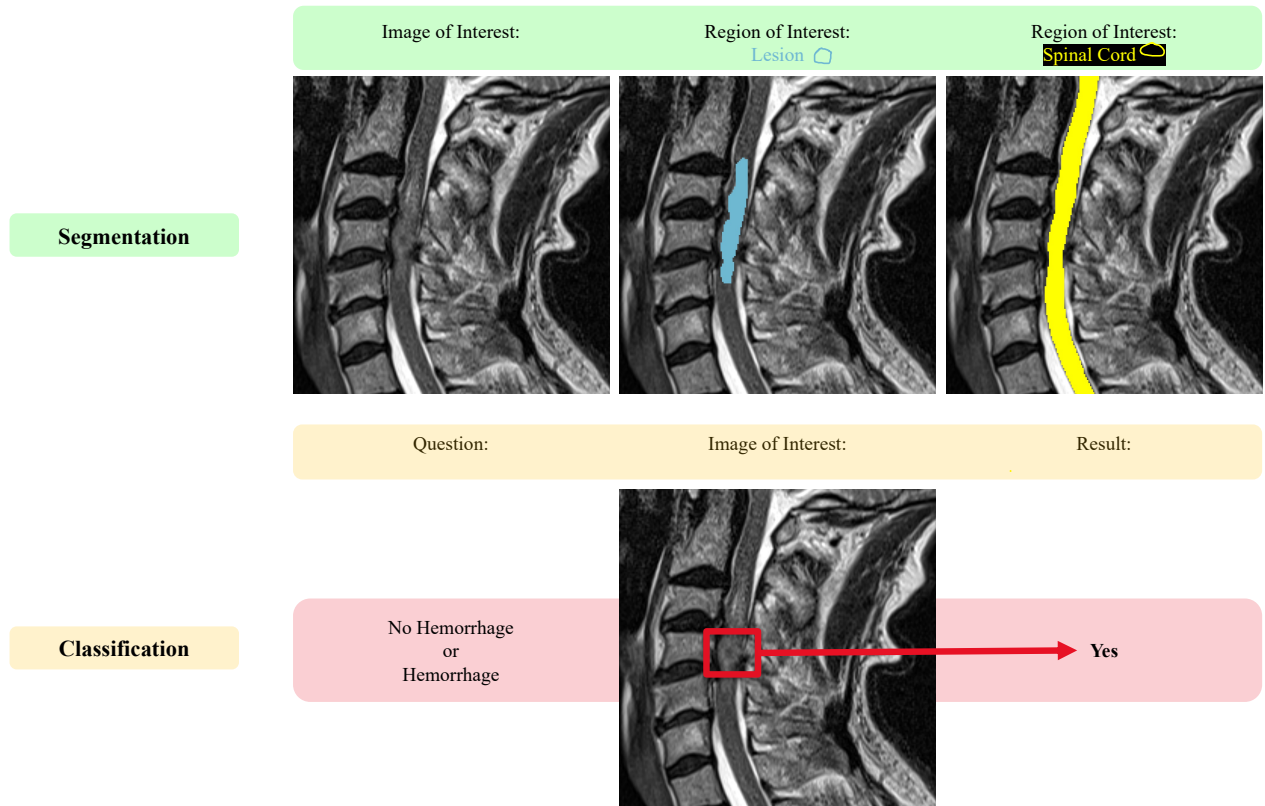


Figure 2.2 What are Segmentation and Classification?

As described by Litjens et al. [83], lesion segmentation presents unique challenges, combining aspects of object detection and anatomical delineation. In contrast, image classification tasks involve assigning diagnostic labels to entire exams or images (e.g. presence or absence of intramedullary hemorrhage) [83]. In essence, lesion segmentation can be considered a pixel-wise classification problem—i.e., a binary decision at each pixel or voxel location as to whether it belongs to the lesion or not. However, for the purpose of this thesis, segmentation

and classification are treated as distinct annotation tasks. Imaging biomarkers may be derived either manually—where a clinician provides segmentation masks of the ROI (e.g. traumatic spinal cord injury lesions, as in this work)—or automatically, using automatic segmentation models. Manual annotation remains a fundamental component, as it serves as the reference standard against which the performance of automatically generated measures is evaluated.

2.2.2 Variability is Problematic

While different learning paradigms—supervised, semi-supervised, or unsupervised—can be applied to medical imaging analysis model training [85–87], lesion segmentation often begins with manually delineated masks provided by domain experts. These masks are commonly referred to as “ground-truths” (i.e., reference standards). In diagnostic imaging, radiologists or clinicians are often tasked with this process due to their ability to characterize lesions on medical images. However, the act of segmentation differs significantly from the diagnostic interpretation typically performed (confirmed by this thesis author background) in radiology. Manual segmentation is inherently subjective, and has been shown to exhibit inter- and intra-rater variability [88, 89], though this has been less thoroughly investigated in the context of tSCI. Many automatic segmentation studies lack methodological rigor. For instance, Wiguna et al. [90] omit details on the manual segmentation procedure and conditions under which ground-truths were created. Stating that two senior residents performed the task offers little insight: Was it their first time? Were consistent screen sizes or lighting conditions used—which affect visual perception? In contrast, the OPERA trials in multiple sclerosis define their procedure [91]. Techniques like smoothing exist [92], but are beyond the scope of this work. Figure 2.3 illustrates a typical challenge faced during manual segmentation:

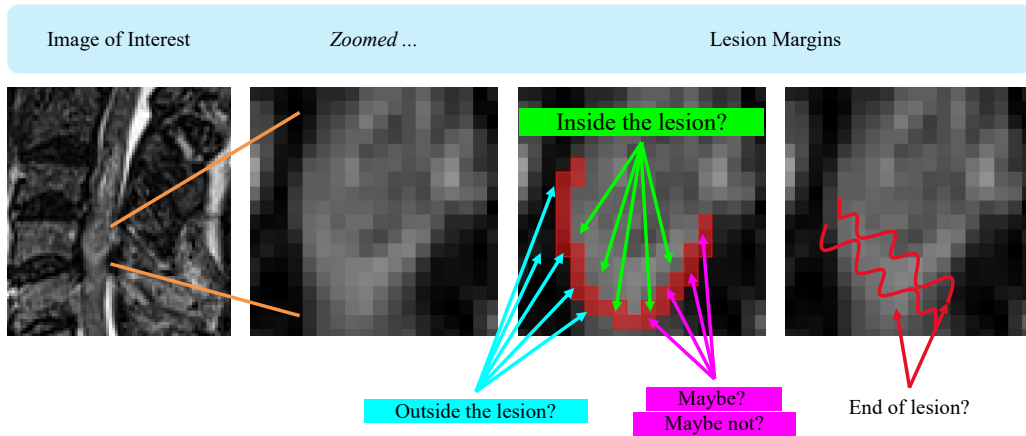


Figure 2.3 Contour Delineation: Right or Wrong?

Contour delineation is an example that is subject to intra and inter-rater variability in manual tasks.

This lack of standardization is problematic given that the performance of predictive models is highly dependent on the quality of training data. While segmentation variability is acknowledged in the literature, its impact is seldom quantified or discussed in depth [93–95]. Variability has also been reported in other domains [96, 97], and methods have been proposed to address it [98]. A key goal is to evaluate reproducibility—defined by the *National Academies of Sciences* as “obtaining consistent results using the same input data” [99] (2019 version). Reproducibility is often assessed using metrics like the Dice Similarity Coefficient and other overlap-based measures [100], though these can be unreliable for small volumes.

2.2.3 Solutions for Efficient Manual Annotation: Lacking

Before assessing reproducibility, the context in which manual segmentations are performed must be considered. Clinicians are frequently involved in curating imaging datasets, yet the process is resource-intensive and often poorly supported. The quality and quantity of labeled data are both critical for the development of high-performance imaging analysis models.

A non-systematic literature review (see [Appendix B](#)) revealed a scarcity of software tools designed to facilitate manual segmentation, classification, and quality control for large imaging datasets. Existing open-source platforms often lack efficient navigation, version control, and customizable annotation workflows. Their unintuitive interfaces or installation process further hinder adoption by clinicians, who have limited time for research and require practical, collaborative solutions. The scarcity of suitable solutions may stem from a disconnect between clinicians, who prioritize data quality, and developers, who emphasize data volume. For example, an observation comes from *Brainhack School* at Polytechnique Montreal [101], which emphasizes the efficiency of command-line tools for computational tasks—a technical advantage often inaccessible to clinicians with limited programming experience. A critical first step to address this challenge is the development of an ecosystem supporting efficient segmentation, classification, and quality control of imaging data.

In response, the first objective of this thesis focuses on developing a custom module using 3D Slicer, specifically designed to address the unmet needs in manual segmentation, classification, and quality control workflows. The second objective investigates the variability inherent in manual segmentations. By tackling the challenge of ground-truth variability and promoting greater transparency, this thesis aims to address barriers in the assessment and application of imaging biomarkers in acute tSCI. Before going into methodological details, the third section of the literature review outlines practical approaches for prediction modeling.

2.3 Approach in Prediction Models

The purpose of this section is not to provide an in-depth exposition of the theoretical mathematics underlying statistical and predictive modeling techniques. Rather, it aims to offer the reader a practical framework for understanding the application of these methods, including the contexts in which they are most suitable, along with their strengths and limitations.

To begin with, there is an important distinction between *explaining* and *predicting* models. For example, Shmueli et al. [76] defines predictive modeling as “the process of applying a statistical model or data mining algorithm to data for the purpose of predicting new or future observations [...] where the goal is to predict the output value (Y) for new observations given their input values (X)” and explanatory models as “a set of underlying factors that are measured by variables X [that] are assumed to cause an underlying effect, measured by variable Y .” The scope of this work is thereby towards developing *predicting* models. In addition, an essential distinction lies between prediction and classification. Section 1.3 in Harrell *Regression modeling strategie...* [102] details prediction as “the act of using a model—or device, algorithm—to produce values of new, existing, or future observations” [103]. Classification “consists of a hierarchical structure of group and subgroup names with corresponding identifiers (numeric or alphanumeric codes)” [104]. Dichotomization of outcomes to simplify model development is not recommended [102] due to loss of information, but authors also recognize that “[i]n the vast majority of cases, classification is the task of the user of the predictive model, at the point in which utilities (costs) and preferences are known.” [102] The distinction between prediction and classification is essential for making informed methodological decisions in model development and interpretation, based on strengths and limitations.

Moreover, establishing a reliable prognosis is essential in the clinical practice, which explains prognostic research in medicine and related guidelines publications [105–111]. In fact, from the manyfolds prediction models being published each year, few are used in real practice. Those guidelines aim to overcome frequent limitations, and include the PROGRESS framework (prognosis research strategy) [107–110], TRIPOD statement (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) [106] (also TRIPOD-AI [112]), and PROBAST (prediction model risk-of-bias assessment tool) [107].

In that context, Efthimiou et al. [105] proposes 13 steps designed to optimize prediction models development, from team creation to clinical implementation. Emphasis is put on strategies for model development, including imputation of missing data, model and variable selection, and validation. The following subsections outline such key foundational concepts in predictive modeling.

2.3.1 Regression Models

Regression remains a central statistical method for modeling the relationship between input variables (predictors) and an outcome of interest [113,114]. Regression techniques provide a flexible and interpretable framework for both explanatory and predictive modeling [76], making it widely applicable in clinical sciences. Various types of regression models exist, and the choice of model depends primarily on the nature of the outcome variable. Each model relies on specific assumptions, which must be met to ensure valid and generalizable results, otherwise introducing bias or overfitting that limits external applicability. Among these, generalizable linear models such as linear regression, logistic regression, [115], ridge [116], elastic net [117] or linear discriminant analysis [118] are commonly used due to their advantages, including straightforward inference, computer efficiency and interpretability [119]. Regression also serves as a baseline for comparing more complex machine learning approaches.

2.3.2 Machine Learning Models

The PubMed National Library of Medicine’s [120] defines machine learning as “[a] type of *artificial intelligence* that enable *computers* to independently initiate and execute *learning* when exposed to new data.” [121] This definition is widely supported in the literature [85,122–124]. Lo Vercio et al. [125] explains its framework in medicine. A wide array of machine learning algorithms—e.g. decision trees [126], random forests [127], neural networks [128], gradient boosting (XGBoost, LightGBM) [129], support vector machines [130], and k-nearest neighbors [131]—can be leveraged thanks to advances in computational power [132–135]. Machine learning workflows typically require additional steps beyond those in classical statistical modeling, including systematic feature selection, model training, and hyperparameter tuning—all essential to develop robust and generalizable predictive solutions.

Feature Selection and Regularization

Biases [136] such as overfitting and underfitting compromise generalization to unseen data. [137] describe “over-fitting [...] to mean minimisation of the model selection criterion beyond the point at which generalisation performance ceases to improve and subsequently begins to decline”. Underfitting occurs when a model is too simplistic to generalize. Mitigation strategies include input feature selection and the use of regularization techniques. Feature selection refers to the “study of algorithms for reducing dimensionality of data to improve [...] performance” [138], while regularization involves the use of a regularizer to quantify and penalize model complexity [139]. This is achieved by removing “irrelevant and/or redundant

features and retain[ing] only relevant features” [138]. Such algorithms are categorized as supervised, unsupervised, or semi-supervised [138]. Three classic methodological frameworks apply: filter, wrapper, and embedded models. LASSO *Least Absolute Shrinkage and Selection Operator* is one of the most prominent embedded methods (also a generalizable linear model) which penalizes the absolute sum of coefficients to induce sparsity [115, 140]: it denotes a small subset of non-zero coefficients after regularization [132]. LASSO is particularly useful as it simultaneously optimizes prediction, perform variable selection [141] and helps address multicollinearity. Regularization techniques [142] address overfitting by introducing a penalty term to the loss function, discouraging overly complex models. They are essential in small-sample settings, and include penalization, early stopping and ensembling [142].

Hyperparameter Tuning

Hyperparameters are “settings that we can use to control the behavior of the learning algorithm” [143]. Hyperparameter tuning is the process of searching for the best combination of hyperparameter values to optimize a model’s performance. From Goodfellow et al. [143]:

There are two basic approaches to choosing these hyperparameters: choosing them manually and choosing them automatically. Choosing the hyperparameters manually requires understanding what the hyperparameters do and how machine learning models achieve good generalization. Automatic hyperparameter selection algorithms greatly reduce the need to understand these ideas, but they are often much more computationally costly. (Section 11.4 in [143])

Using automatic hyperparameter selection, grid and random search are common strategies. While grid search exhaustively evaluates all combinations (common practice when there are three or fewer hyperparameters [143]), “[t]he obvious problem with grid search is that its computational cost grows exponentially with the number of hyperparameters. If there are m hyperparameters, each taking at most n values, then the number of training and evaluation trials required grows as $\mathcal{O}(n^m)$.” [143] Random search is often more efficient and effective in high-dimensional spaces [144]. Bayesian optimization further improves efficiency by using a probabilistic model to select hyperparameters that are likely to improve performance [145]. Nested cross-validation is crucial when hyperparameter tuning is involved, as it prevents data leakage between training and validation processes and provides an unbiased estimate of model performance [137]: tuning should always be performed within a cross-validation loop.

2.3.3 Validation and the Role of Data Sampling Strategies

Whether prediction models are developed using regression or machine learning approaches, a critical concept is the model validation process, which is directly influenced by data sampling strategies. Both internal and external validation are essential to develop well-calibrated and generalizable prediction models capable of producing reliable and robust inferences. “Internal validation is the process of determining [...] validity for the setting where the development data originated from.” (see Chapter 17 in [146]) “A further aspect is the external validity (or “generalizability”/“transportability”) of the prediction model to populations that are “plausibly related” [147]” [146]. “We learn about external validity by testing the model in samples from other settings than the model development setting.” [146] In that context, appropriate data sampling strategies are essential to ensure efficient use of available data and to match the analytical approach with the study design and computational resources. A classical step in building prediction models is splitting the dataset into subsets for training and testing. The most common approach is the split-sample method, typically partitioning 70% for training and 30% for testing. The training set can be further divided into development and internal validation subsets, while the test set serves for external validation. Although intuitive, this approach reduces statistical power and inefficiently uses data [105, 146]. Single splits are sensitive to sampling variability and may yield over-penalized, biased performance estimates, especially in small or imbalanced datasets [146]. “Lucky” or “unlucky” splits can lead to overly optimistic or pessimistic results. Cross-validation methods, including *leave-one-out* (PRESS) cross-validation [148], offer more robust alternatives by maximizing data use and stabilizing estimates [146, 149–151]. However, they may not fully capture uncertainty from processes like automated variable selection [146]. Bootstrapping, involving repeated sampling with replacement, is another internal validation technique recommended for its robustness, with about 200 iterations often sufficient to stabilize estimates [146]. Yet, bootstrapping demands greater computational resources and typically requires automated modeling strategies [146]. The combination of internal validation with missing data imputation remains debated, though distinguishing between ideal and pragmatic model performance has been proposed to emphasize practical relevance over theory [146, 152–154]. Importantly, none of these internal validation methods replace the need for independent external validation, which assesses a model’s generalizability to new patients with potentially different predictor distributions [146, 155]. While internal validation relies on random splits within the same dataset, external validation involves distinct patient populations and is crucial for confirming model transportability. Greater generalizability strengthens both scientific validity and clinical applicability [146].

2.3.4 Performance Assessment

Assessing prediction and classification model performance involves several considerations. “The distance between the predicted outcome \hat{y} and actual outcome y is central to quantify overall model performance from a statistical perspective” [132] [2001 version]. For continuous outcomes, this residual is $y - \hat{y}$; for binary outcomes, \hat{y} [is] the predicted probability p [146]. Goodness-of-fit reflects a model’s ability to fit data, with better models showing smaller residuals [146]. Key concepts include discrimination—how well a model separates those with and without the outcome—and calibration, the agreement between observed and predicted outcomes [146]. Performance metrics depend on the outcome type. For continuous outcomes, R^2 (coefficient of determination) is common [146], though defining a “good” R^2 is challenging [156]. For binary outcomes, the area under the receiver operating characteristic curve (AUC) is widely used; it ranges from 0.5 (random) to 1.0 (perfect), with 0.7 acceptable, 0.8 modest, and 0.9 excellent [146, 157]. The Brier score [158] is also appropriate but influenced by outcome incidence, unlike AUC. Accuracy measures the proportion of correct predictions, providing a general indicator of a model’s overall performance [159]. Confidence intervals can be estimated by resampling [146]. Calibration plots are recommended for moderate evaluation of calibration [146, 160]. Interpretability of the model and predictions helps for its adoption. Tools like SHAP (SHapley Additive exPlanations) values [161] and permutation feature importance [127] enhance interpretability by quantifying individual predictor contributions, improving transparency and trust in predictive analytics.

2.3.5 Previous Work in Acute Cervical Traumatic Spinal Cord Injury

Before developing new models, transfer learning [162]—leveraging prior related work—is important [105]. In acute tSCI, imaging and non-imaging studies propose tools predicting neurological and non-neurological outcomes [43, 163–167], with growing use of machine learning and AI reflecting clinical interest [90, 168–178]. Numerous studies assess MRI biomarkers for prognosis [21–23, 25, 26, 58, 179], but many confuse prediction with association, and MRI biomarkers’ role remains debated. Large cohorts often exclude imaging biomarkers [178, 180], while imaging-focused studies usually involve small, mono- or bicentric samples [58, 171, 179], limiting generalizability—likely due to disease prevalence and extraction costs. Some use deep learning for automated MRI analysis but often exclude core clinical data [181]. Differences in cohort characteristics (e.g. injury severity) further limit transportability. Prediction timeframes vary, targeting short-term [181] or long-term outcomes [176, 177]. To our knowledge, on a multicenter dataset of this size, no study has combined yet clinical and imaging biomarkers on preoperative MRI using similar approach: this gap motivates the present work.

MRI Sequences in Acute Traumatic Spinal Cord Injury Assessment

The selection of MRI sequences is an important consideration in an imaging research study. In the context of acute tSCI, reviewing the protocols used in previous studies helps establish the optimal imaging analysis framework. Additionally, it is important to distinguish between preoperative and postoperative MRI acquisition, as the presence of metallic surgical hardware in postoperative settings can introduce significant artifacts that significantly affect image quality and interpretation. For example, Freund et al. [14] focus on the role of MRI in the postoperative evaluation of tSCI, which may not be directly applicable to acute preoperative cases. In early tSCI, MRI is typically performed under time constraints, leading to shortened protocols, and the appearance of spinal cord lesions evolves dynamically over time [14, 182, 183]. These contextual differences must be considered when selecting MRI sequences to use.

Kurpad et al. [26] note that standard sequences typically include sagittal and axial T2-weighted MRI, with additional sequences such as sagittal T1-weighted and short tau inversion recovery (STIR) acquired when feasible. Kumar et al. [17] provide a detailed overview of a typical MRI protocol for acute spinal cord injury. Their suggested protocol includes sagittal T1-weighted and T2-weighted spin echo sequences, sagittal STIR, and sagittal and axial T2* gradient recalled echo (GRE) sequences. T1-weighted images are primarily used to visualize anatomical structures and bone fractures. STIR sequences are particularly sensitive for detecting soft tissue and ligamentous injuries due to their uniform fat suppression and strong sensitivity to edema. While fat-suppressed T2-weighted sequences may also reveal edema, STIR sequences offer more consistent signal suppression of adipose tissue. T2-weighted images are well-suited for identifying intramedullary cord edema, and T2* gradient recalled echo sequences are particularly effective for detecting hemorrhage within or around the spinal cord. Recent developments have introduced diffusion tensor imaging (DTI) since it can detect microstructural trauma-related changes in the spinal cord that are not visible using conventional sequences, but further studies are needed before routine usage [184].

Ideally, MRI should be performed within 72 hours of injury. During this period, edema-related T2 hyperintensity enhances the visibility of affected ligaments, which appear hypointense under normal conditions. Beyond this window, the resolution of edema and hemorrhage can diminish the sensitivity of MRI to detect key features such as ligamentous injuries. The value of baseline MRI in acute tSCI has also been explored [26], where a single study by Papadopoulos et al. [185] in 2002 investigated how standardized MRI protocols might influence surgical outcomes. In brief, MRI in acute tSCI should combine shortened acquisitions that enable the identification of spine structures and cord changes, as detailed above.

CHAPTER 3 METHODOLOGY

A thorough literature review and teamwork collaboration support this project completion. The main study design is presented first, followed by the methods that address the objectives.

3.1 Study Design

This master thesis work relates to the imaging substudy of the *Rick Hansen Spinal Cord Injury Registry* study [186] (RHSCIR) coordinated by the Praxis Institute [187]. The RHSCIR imaging substudy is a multicenter retrospective cohort initiative that aims to establish a pan-Canadian MRI repository among RHSCIR participants, leveraging tSCI research.

In that context, 341 participants from 9 centers enrolled in the RHSCIR study between January 1, 2015 - August 31, 2021, have been selected. 284 participants from 7 centers had MRI data available for this thesis. The study follows the standards established by the principles of the Declaration of Helsinki. The Praxis Institute team coordinates data collection after each center obtained approval from its respective institutional ethics board and signed a data sharing agreement. The STROBE Statement (strengthening the reporting of observational studies in epidemiology) [188], a checklist for reporting cohort studies, is provided in Appendix C. All imaging, clinical data, and related processing scripts are available on `spineimage.ca` [29], a secure platform hosted by the *Digital Research Alliance of Canada* [189] in Victoria, British Columbia, Canada. MRI data extracted from the participating sites is curated by local teams (coordination by the Praxis Institute), and standardized to the *Brain Imaging Data Structure* [190] (BIDS) format. Corresponding clinical data for all imaging substudy participants is available through linkage with the RHSCIR study.

3.1.1 Inclusion and Exclusion Criteria

Inclusion criteria are acute cervical tSCI of any severity (i.e., AIS A-D), preoperative MRI acquired in the 72 hours following injury, availability of clinical data in the RHSCIR (with more specifically availability of motor scores and level of injury as per *International Standards for Neurological Classification of Spinal Cord Injury* (ISNCSCI) examination available at baseline [within 72 hours post-injury], and at rehabilitation discharge), date and time of injury available, and enrolled and discharged in a RHSCIR site between January 1, 2015 - August 31, 2021. Exclusion criteria are participants that do not meet inclusion criteria, participants with penetrating trauma, prior spine surgery and non diagnostic MRI.

3.1.2 Sample Size

The sample size estimation may involve additional steps than conventional statistical approaches. First, multiple outcomes-e.g. continuous and dichotomized-are predicted using various statistical models. So, usual sample size statistical softwares like G*Power [191] are avoided. Second, the imaging substudy also supports the development of segmentation models (not covered in this thesis), which reuses the same dataset. Therefore, a pragmatic approach is adopted to assess whether the available sample size is adequate for the intended analyses. As discussed by Riley et al. [192], appropriate sample sizes for prediction models depend on outcome incidence and model complexity. For example, a binary outcome with 30% prevalence requires approximately 323 participants to maintain 5% error and 80% statistical power. If the prevalence is 50%, the required sample increases to 385; for 10% prevalence, 139 participants are needed. In summary, although sample size requirements vary depending on the outcome and modeling approach, the current dataset falls within an acceptable range to achieve sufficient statistical power for the primary analyses presented in this thesis.

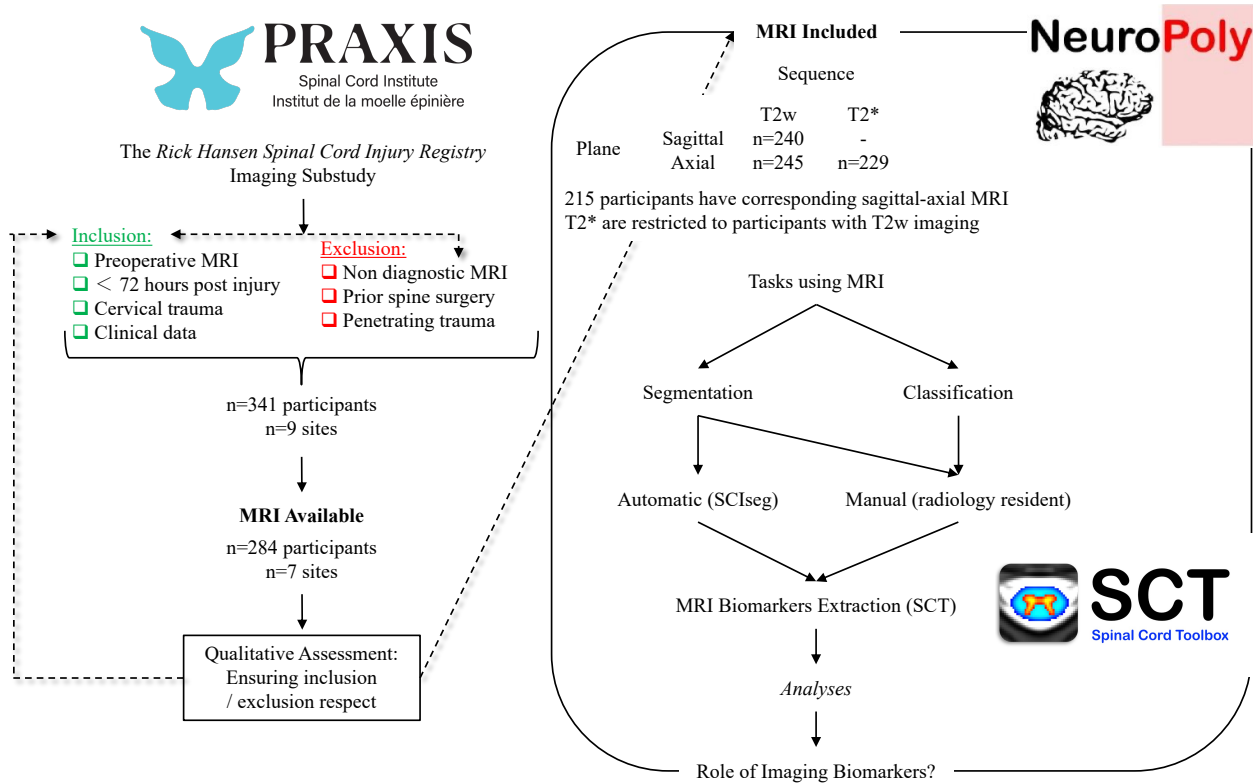


Figure 3.1 Rick Hansen Spinal Cord Injury Registry Imaging Substudy and Thesis Flowchart

The MRI sequences used in this study include sagittal and axial T2-weighted and T2* acquisitions, which are selected based on 1) their consistent use across centers; 2) their utility for imaging spinal cord lesions, where T2-weighted sequences highlight intramedullary edema and T2* sequences highlight hemorrhage.

3.2 Objective 1: To Develop a Software Ecosystem Enhancing Manual Annotation Tasks in Imaging Datasets

The first objective focuses on building a custom platform that facilitate manual ground-truth segmentation and classification tasks, from which MRI metrics/biomarkers are extracted.

3.2.1 Software Selection for Efficient Segmentation and Classification

Manual segmentation is highly time-consuming, making user-friendly navigation critical. Even minor delays—e.g. reopening the viewer for each scan—can disrupt focus and reduce efficiency, especially when working with large datasets. The user experience plays a critical role as time constraints and usability challenges can affect both the accuracy and completeness of the generated ground-truth data—a foundational step that impacts all subsequent analyses. In this context, 3D Slicer is selected as an initial platform due to several advantages: 1) support for developing customized modules and workflows; 2) widespread adoption in imaging research, benefiting from an active user community and continuous updates; and 3) minimal Python skills required to build and deploy modules, thereby facilitating collaborative development and integration. This allows for a hybrid solution that enhances user experience—particularly for clinicians generating ground-truths—while also producing structured outputs for downstream analysis by technical teams.

Development of SlicerCART

SlicerCART—standing for *Slicer Configurable Annotation & Review Tool*—has been developed to lower the barrier to dataset creation, aimed to provide a time-efficient, open-source platform for manual segmentation and classification across large datasets, including appropriate tools for quality control and structured data output. The module is adapted from the original ICH_SEGMENTER 3D Slicer module, developed by Dr. Laurent Létourneau-Guillon’s team. ICH_SEGMENTER proposed useful graphical user interface (GUI) features, including a case list for facilitating navigation. However, it suffered from usability issues and bugs (as with other open-source software) that limited its functionality. Furthermore, it lacked several features necessary for efficient use: customizable case lists from external files, the ability to resume from previous tasks, dynamic GUI behavior tailored to individual user preferences (e.g. mouse interactions or keyboard shortcuts), and versioning settings. Additionally, the original module was built for scans using the NRRD [193] format, whereas the current project involves NIfTI [194] format files.

Before beginning SlicerCART development, preliminary feasibility testing was conducted to evaluate integration and development potential. 3D Slicer includes an embedded Python console operating within its own virtual environment, which may cause challenges in development tasks. Specifically, we examined: 1) Integration of the Slicer Python environment with common IDEs (e.g. PyCharm, VS Code); 2) Reading external files from within Slicer Python scripts; 3) File output capabilities via GUI interaction; 4) GUI modification using Qt Designer; 5) Executing shell scripts from the module. After our review, our findings were that: A) Integration with IDEs was limited due to internal interactions with Slicer’s C++ source code, making classical debugging impractical; B) Feasibility of reading from and writing to external files via GUI-driven Python functions; C) Modification of the GUI using Qt Designer is practical; D) Feasibility of executing shell scripts from the module. Based on these tests, it was concluded that building a module like SlicerCART was feasible.

SlicerCART is developed using Python, following PEP 8 [195] coding standards, and is currently based on 3D Slicer version 5.6.2 and Mac OS, although it is theoretically operating system agnostic, but currently not fully tested in the Windows or Linux environment. It incorporates the *Brain Imaging Data Structure* [190] (BIDS) standardized format. The module is hosted in a public GitHub repository under the Neuropoly organization, which uses pull requests for open-source development and version control. At the time of this thesis submission, the repository includes the following components:

| Filename/Folder | Description |
|-----------------|--|
| README.md | Introductory guide for installing and using SlicerCART |
| CONTRIBUTING.md | Developer guide for contributing to the module |
| doc/ | Subfolder containing detailed documentation about usage and workflow capabilities |
| src/ | Source folder containing the following subfolders: |
| Resources/ | Includes images, logos, and UI layout files such as SlicerCART.ui |
| scripts/ | Each Python file contains one class (except InteractingClasses.py , which includes multiple classes for GUI customization) These files cannot interact between themselves and are named after the class they contain |
| utils/ | Contains helper functions, such as initialization scripts (e.g. environment checks) and shared utilities reused across scripts in scripts/ for debugging and maintenance |
| SlicerCART.py | Main module script that orchestrates GUI behavior and task execution |

Table 3.1 Description of Central Files and Folders in the SlicerCART Repository

UI = User Interface; GUI = Graphical User Interface.

The modular structure is designed to facilitate code maintainability, collaborative development, and extensibility. A visual summary of the code structure is shown in Figure 3.2. Table 3.2 details the purpose of each class, while Table 3.3 describes complementary files.

Code Design Structure for Slicer Configurable Annotation & Review Tool

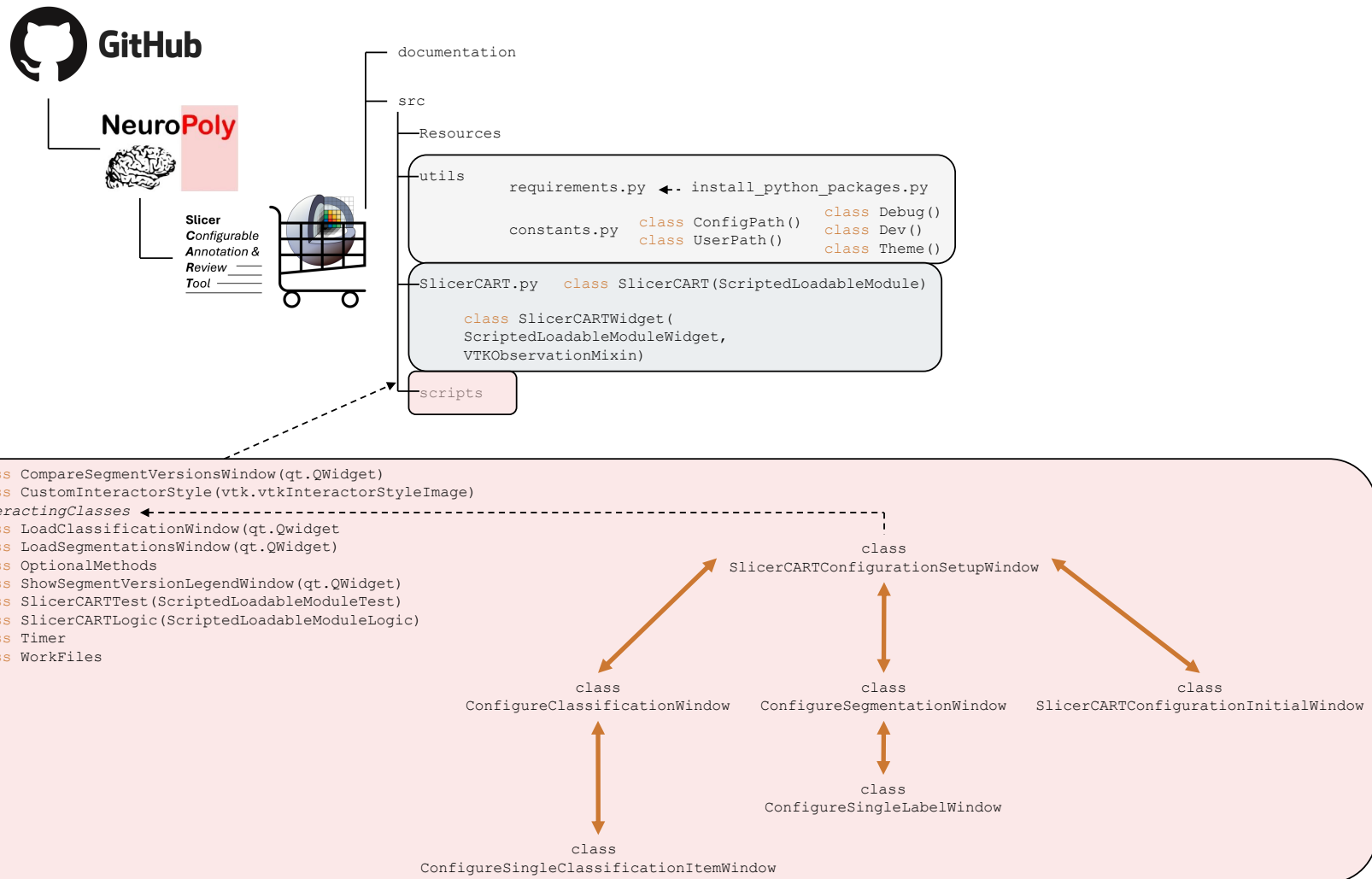


Figure 3.2 Overview of SlicerCART Module Code Structure

| Class Name | Description |
|--|---|
| <code>SlicerCART(ScriptedLoadableModule)</code> | Launches SlicerCART module in 3D Slicer |
| <code>SlicerCARTWidget(ScriptedLoadableModuleWidget, VTKObservationMixin)</code> | Sets up (main class) and controls the working environment and user interface for the SlicerCART module in 3D Slicer |
| <code>CompareSegmentVersionsWindow(qt.QWidget)</code> | Manages comparison and editing of previous segmentation masks |
| <code>CustomInteractorStyle(vtk.vtkInteractorStyleImage)</code> | Defines customized mouse and keyboard shortcuts |
| <code>LoadClassificationWindow(qt.QWidget)</code> | Manages loading of previous classification labels and versions |
| <code>LoadSegmentationsWindow(qt.QWidget)</code> | Manages loading of previous segmentation masks and versions |
| <code>OptionalMethods()</code> | Provides optional methods for extension or customization |
| <code>ShowSegmentVersionLegendWindow(qt.QWidget)</code> | Provides a window to choose which segmentation label version to display |
| <code>SlicerCARTTest(ScriptedLoadableModuleTest)</code> | Placeholder for unit tests |
| <code>SlicerCARTLogic(ScriptedLoadableModuleLogic)</code> | Manages module logic and interaction within 3D Slicer |
| <code>Timer()</code> | Measures and logs execution time for individual tasks |
| <code>WorkFiles()</code> | Handles manipulation of case lists to continue/facilitate the workflow from previous work sessions. |
| <code>SlicerCARTConfigurationSetupWindow()</code> | Manages display of the configuration options menu |
| <code>ConfigureClassificationWindow()</code> | Manages classification labels group configuration (checkboxes, comboboxes) |
| <code>ConfigureSingleClassificationItemWindow()</code> | Adds a specific classification label type (e.g. adding a unique checkbox) |
| <code>ConfigureSegmentationWindow()</code> | Manages segmentation label group configuration (checkboxes, comboboxes) |
| <code>ConfigureSingleLabelWindow()</code> | Adds an individual segmentation label (e.g. adding a unique label mask) |
| <code>SlicerCARTConfigurationInitialWindow()</code> | Configures the initial segmentation configuration window |
| <code>ConfigPath()</code> | Controls access to configuration files depending on the operational context |
| <code>UserPath()</code> | Manages user-specific path settings for workflow resumption |
| <code>Debug()</code> | Facilitates debugging of SlicerCART module |
| <code>Dev()</code> | Provides utilities for development and repetitive tasks (e.g. message boxes) |
| <code>Theme()</code> | Manages the color theme for the module interface |

Table 3.2 Description of SlicerCART Module Core Classes

| Filename | Description |
|---|--|
| <code>requirements.py</code> | Contains all import modules required for proper functioning of SlicerCART module. It ensures all dependencies are clearly defined. |
| <code>install_python_packages.py</code> | Checks for missing Python packages and installs them automatically. Useful for ensuring a consistent working environment across systems. |

Table 3.3 Complementary Python Files that Support SlicerCART

Key SlicerCART development features focus on usability, flexibility, and workflow continuity. First, the installation process is designed to be straightforward: users install 3D Slicer, download the module repository, then load it by adding its local path in Slicer, and are ready to begin. Upon launching SlicerCART, a built-in function automatically checks for additional Python packages that are not included by default in the 3D Slicer virtual environment. If any required packages are missing, the user is notified and automatic installation begins. Second, the module offers a high level of customization, supporting segmentation and/or classification tasks tailored to the specific needs of a project. Each project is currently limited to a single imaging modality (either MRI or CT), and users can configure default viewer settings, as well as define customized segmentation masks and classification labels. Third, SlicerCART supports resuming previous work, enabling users to continue annotation from where they left off. This functionality is designed for segmentation tasks and is driven by two automatically generated files: `working_list.yaml`, which contains the full list of cases for a given project task, and `remaining_cases.yaml`, which is updated after each annotation session to reflect the cases yet to be processed. The updated `remaining_cases.yaml` is cross-referenced with `working_list.yaml` to identify the next case in sequence. This mechanism is illustrated in Figure 4.7, presented in the Results Chapter. If workflow discrepancies are detected—such as mismatches between the filenames in the working list and those in the selected imaging dataset—a pop-up message alerts the user to potential incompatibilities. In such cases, the user is either given the option to override the warning or is redirected to resolve the issue if it is deemed critically incompatible. An example of such a pop-up window is shown in Figure 3.3. Although this resume functionality currently applies only to segmentation tasks, it could be extended to classification tasks in future developments. The `UserPath` and `ConfigPath` classes manage hidden files and previously saved configurations by treating the output folder as a user-defined directory. Within this folder, a `_conf` subfolder is automatically generated to store the configuration used in the current project, enabling seamless task resumption under consistent settings. Additionally, for each output folder, the paths to the associated volume folders (which contain the imaging data) are saved in hidden files on the user’s system. This mechanism allows the project to be transported and resumed on different machines without losing references to its original configuration structure.

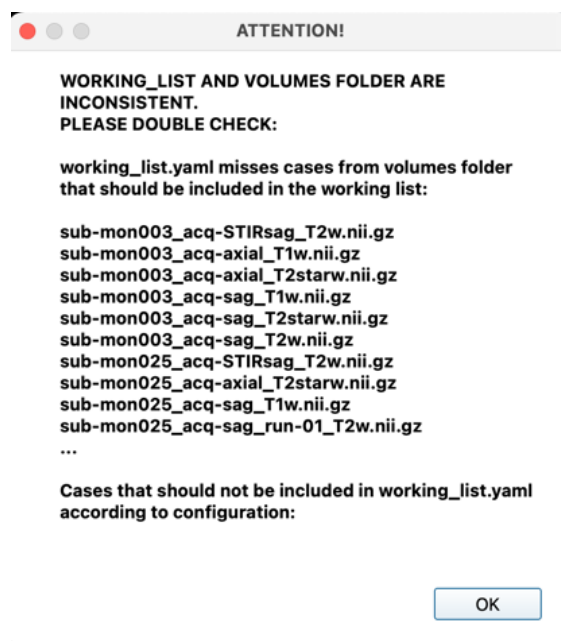


Figure 3.3 Example of a Pop-Up Window Notifying Incompatibility Work Environment

Additional considerations arise when designing a custom module in 3D Slicer. This thesis work focuses on NIfTI segmentation mask files; therefore, before saving each case, an automatic conversion to binary labelmap is done to ensure further compatibility. Moreover, 3D Slicer interpolates images by default to enhance visual quality. To accommodate different use cases, it is important to include the ability to toggle between interpolated and non-interpolated (raw) image data. This distinction is particularly important in contexts such as 3D segmentation model development, where masks must remain binary and aligned precisely with the voxel grid. In contrast, clinicians often rely on post-processed, interpolated images for improved anatomical visualization. As a result, users are advised to carefully verify that segmentations correspond to the intended anatomical regions, since annotations made on interpolated images may introduce slight discrepancies when applied to raw image data.

In this thesis, the success of SlicerCART development is evaluated through a proof of concept involving manual segmentation and classification tasks conducted on imaging data from the *Rick Hansen Spinal Cord Injury Registry* (RHSCIR) imaging substudy. According to a predefined standard operating procedure, the average time required to complete the segmentation or classification tasks for each participant—regardless of the viewing orientation—is expected to remain under five minutes (H1). Based on the outcomes of this proof-of-concept, future evaluations of success may incorporate user feedback from both internal collaborators (e.g. NeuroPoly Lab) and external users (e.g. Slicer community), as well as metrics such as the number of downloads, sustained usage, and frequency of updates to the module.

3.2.2 Manual Annotation: Standard Operating Procedure

In an effort to standardize the practices surrounding imaging annotation, a manual segmentation protocol is proposed to ensure reproducibility in the RHSCIR imaging substudy. The standard operating procedure for both segmentation and classification tasks is described:

Preliminary Steps

MRI data is accessible through `spineimage.ca` (coordinated by the Praxis Institute), which allows users to download MRI sequences locally via `git-annex`. Users must ensure that no spaces are used in folder or filenames, and folder paths from root folder do not begin with numbers, as this may prevent recognition by 3D Slicer’s Python environment. Once data is downloaded locally, the user installs 3D Slicer and SlicerCART. Installation instructions must be followed carefully to enable proper usage on macOS (N.B.: other operating systems are currently untested). The user should maintain a consistent annotation environment—ideally using the same monitor, ambient lighting, and time of day for each session. A 5-minute break should be taken after every 50-minute session. An initial qualitative review is then conducted to identify suitable cases for manual segmentation based on the inclusion/exclusion criteria outlined in subsection 3.1.1. For each T2-weighted MRI and view acquisition, cases are assessed to suit for lesion contour delineation, intramedullary edema and/or hemorrhage as distinct classes. Cases with other lesions or relevant findings are flagged. While the same annotation tool is used for both segmentation and classification, initial manual annotation was performed using the initial `ICH_SEGMENTER` module. The following steps are then undertaken:

1. Extract all sagittal and axial T2-weighted sequence filenames using Python scripts (e.g. `sub-ott005_acq-axial_run-01_T2w.nii.gz`).
2. Create an output folder (recommended: one output folder per view-volume folder, e.g. `site_03_output_sagittal`).
3. Launch SlicerCART within 3D Slicer.
4. Select the corresponding volume folder (e.g. `site_03`) containing the MRI data.
5. Select the output folder (e.g. `site_03_output_sagittal`) related to the volume folder (e.g. `site_03`).
6. Navigate through each T2-weighted sequence and manually assess if it respects the inclusion criteria.

- 6.1. For subjects with multiple sequences, compare all available images and select the one with the best quality;
- 6.2. If no sequence is suitable, the subject is excluded.
7. Double check excluded cases, and confirm with a neuroradiologist (pending).
8. Repeat the procedure for the second view, and for all sites.
9. Use scripts to extract the included sagittal and axial suitable cases for segmentation.

Segmentation Tasks

From the sagittal and axial T2-weighted included cases, the following steps are undertaken:

10. Create new output folders for storing manual segmentation masks (reusing previous folders is not recommended).
11. Open SlicerCART in 3D Slicer.
12. Select “New Configuration” from the pop-up menu and define settings for the project, including the segmentation label names (for this work: **edema** and **hemorrhage**).
13. Select the appropriate volume and output folders.
14. Replace the **working_list** and **remaining_list** YAML files in the output folders, so they now contain only the included cases (e.g. by copy-pasting lists in both files).
15. Select again the output folder, so SlicerCART loads only the included cases.
16. Fill in annotator name, degree, and revision step (e.g. Step-0 for initial annotations).
17. For each included case:
 - 17.1. Perform an initial visual scan to identify lesion location and decide if one or both classes (**edema**, **hemorrhage**) should be segmented;
 - 17.2. Verify the correct label is selected (double-check even if automatic);
 - 17.3. Manually draw lesion contours; adjust if necessary;
 - 17.4. Ensure that contours for each label are closed shapes and present on all relevant slices (subjective);
 - 17.5. Save the segmentation mask(s) (*N.B. Segmentation time per case is automatically saved in a CSV file through SlicerCART functionalities*).

Ensure the following conditions for reproducible segmentation are met:

- Annotations are performed using the same screen in the annotator’s environment;
- Sessions occur under stable low lighting, between 9:00 and 19:00;
- A 5-minute break is taken after every 50-minute session.

After completing all cases:

18. Create a backup copy of the original data.
19. Run a script to fill masks (available from `spineimage.ca`, authored by Jan Valošek).
20. Reload and review all masks to ensure completeness. If empty, manually fill them.
21. Upload (e.g. through Python scripts) the final segmentation masks to `spineimage.ca` under the `derivatives` folder. Ensuring filenames naming convention are respected.

Classification Tasks

From the sagittal T2-weighted included cases, manual classification tasks are performed:

22. Create new folders to store classification labels.
23. Open SlicerCART in 3D Slicer.
24. Select “New Configuration” and define the label names to use. In this study, the labels were chosen based on the author’s exploratory insight and are listed in Table 3.4.
25. Select the corresponding volume and output folders.
26. Provide annotator details (name, degree, and revision step).
27. Review each case (*N.B. Classification time and labels are saved in CSV files*):
 - 27.1. Select checkboxes for present labels;
 - 27.2. Select dropdown options;
 - 27.3. Fill out free-text boxes when required.
28. At the end, double check cases by reviewing comments and ensure consistency between annotations and text notes.

All annotation times are saved automatically using SlicerCART functionalities based on datetime difference and 3D Slicer observer nodes that allow for task completion tracking.

3.3 Objective 2: To Measure Variability of MRI Biomarkers Extracted from Manual Segmentation

From the manual segmentation masks and classification labels obtained in **Objective 1**, MRI Biomarkers are extracted. This section covers both quantitative and qualitative imaging biomarkers selection process for predictive modeling and reliability assessment in tSCI MRI.

3.3.1 Quantitative Metrics

Quantitative MRI metrics presented in Figure 3.4 are selected based on literature and the available `sct_analyze_lesion` function in the *Spinal Cord Toolbox* [196] (SCT). These metrics are generated by Jan Valošek (postdoctoral fellow at NeuroPoly Lab) using both manual and automatic lesion and manually corrected spinal cord segmentation masks on sagittal and axial T2-weighted MRI¹. Lesion masks are postprocessed by intersecting them with the spinal cord masks to correct potential oversegmentation. MRI metrics from automatic segmentation are generated only for participants with successful lesion segmentation using “SCIseg” [197]. All metrics available—whether from automatic or manual lesion segmentation—imply a non-empty automatic spinal cord segmentation from SCT contrast-agnostic model [198] (v2.5). Manual corrections are subsequently applied for sagittal (Jan Valošek) and axial (third-year diagnostic radiology resident) images. Values are saved in a single CSV file for tabular merging. High-order radiomic features and tissue bridges [68] are not extracted.

3.3.2 Qualitative Labels

Qualitative MRI metrics (or “MRI labels”) presented in Figure 3.5 are categorical variables—for example, presence or absence of spinal cord transection (see Table 3.4 for full listing). The selection process occurs in two steps. First, relevant biomarkers are identified through a literature review, highlighting key features such as intramedullary hemorrhage, frequently discussed in tSCI imaging (N.B. Hemorrhage is discussed separately in subsection 3.3.3). Second, additional potential labels are identified during the initial quality control phase of MRI data, prior to lesion segmentation. These are based on radiological assessment—e.g. the presence or absence of likely chronic severe degenerative spine changes. An initial list of T2-weighted sagittal MRI labels—excluding intramedullary hemorrhage—is then established. A third-year diagnostic radiology resident manually classifies each participant for each labels listed in Table 3.4, using SlicerCART with subsequent CSV files merging.

¹Except from automatic segmentation masks on axial T2-weighted MRI, generated by this thesis author.

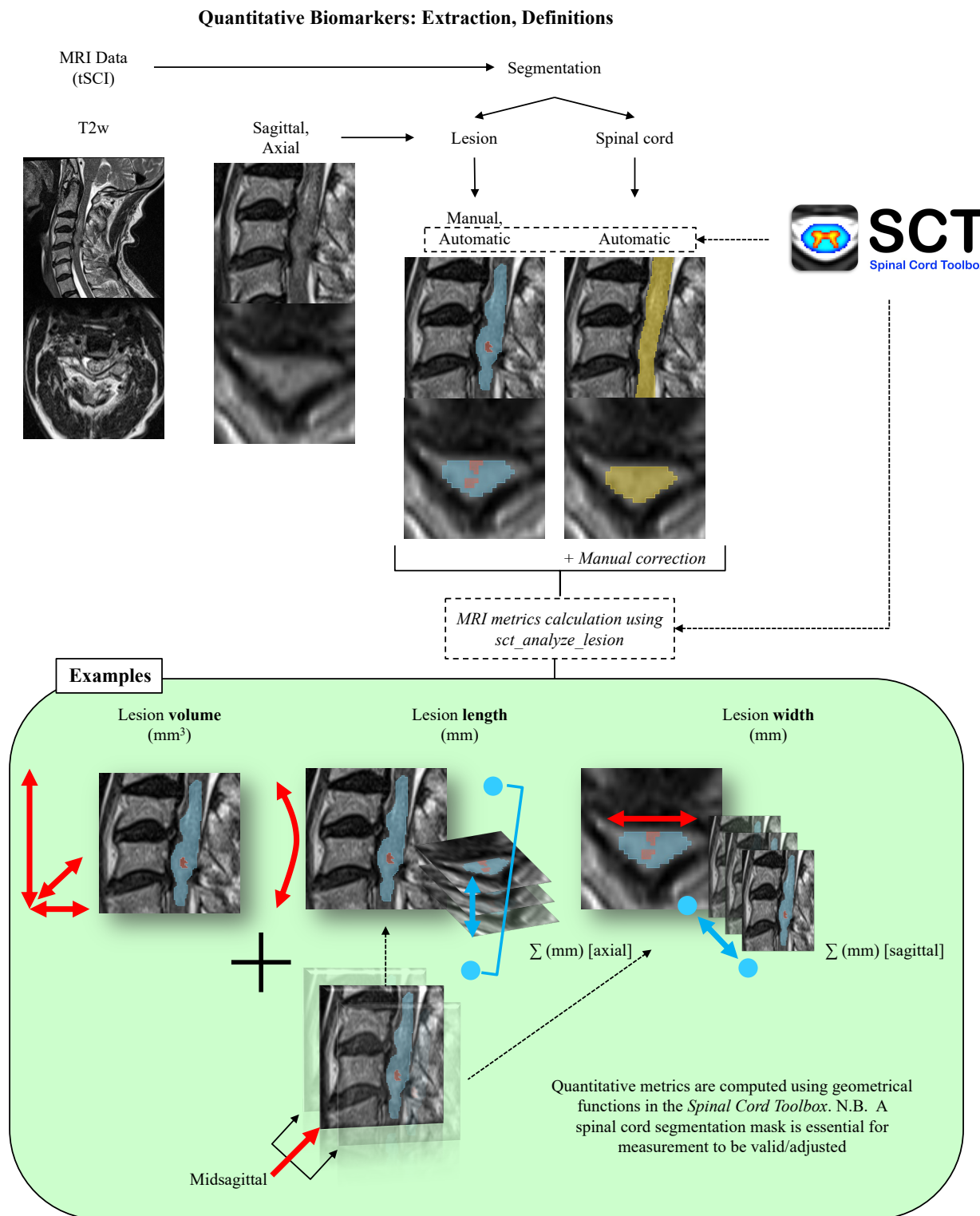


Figure 3.4 Illustrated Examples of Quantitative MRI Metrics Measured

From MRI data available using sagittal and axial T2-weighted sequences, segmentation masks are obtained (prior step) manually and automatically. Spinal cord masks are obtained using the Spinal Cord Toolbox models, which enable morphometrics computation. Metrics such as total lesion volume (mm^3), length (mm) and width (mm) can then be extracted. Additional measurements for lesion length and width are taken specifically on the midsagittal slices (this applies also for axial sequences). Number of lesions (e.g. 1, 2 or 3), maximal axial damage ratio, and lesion max diameter are also included.

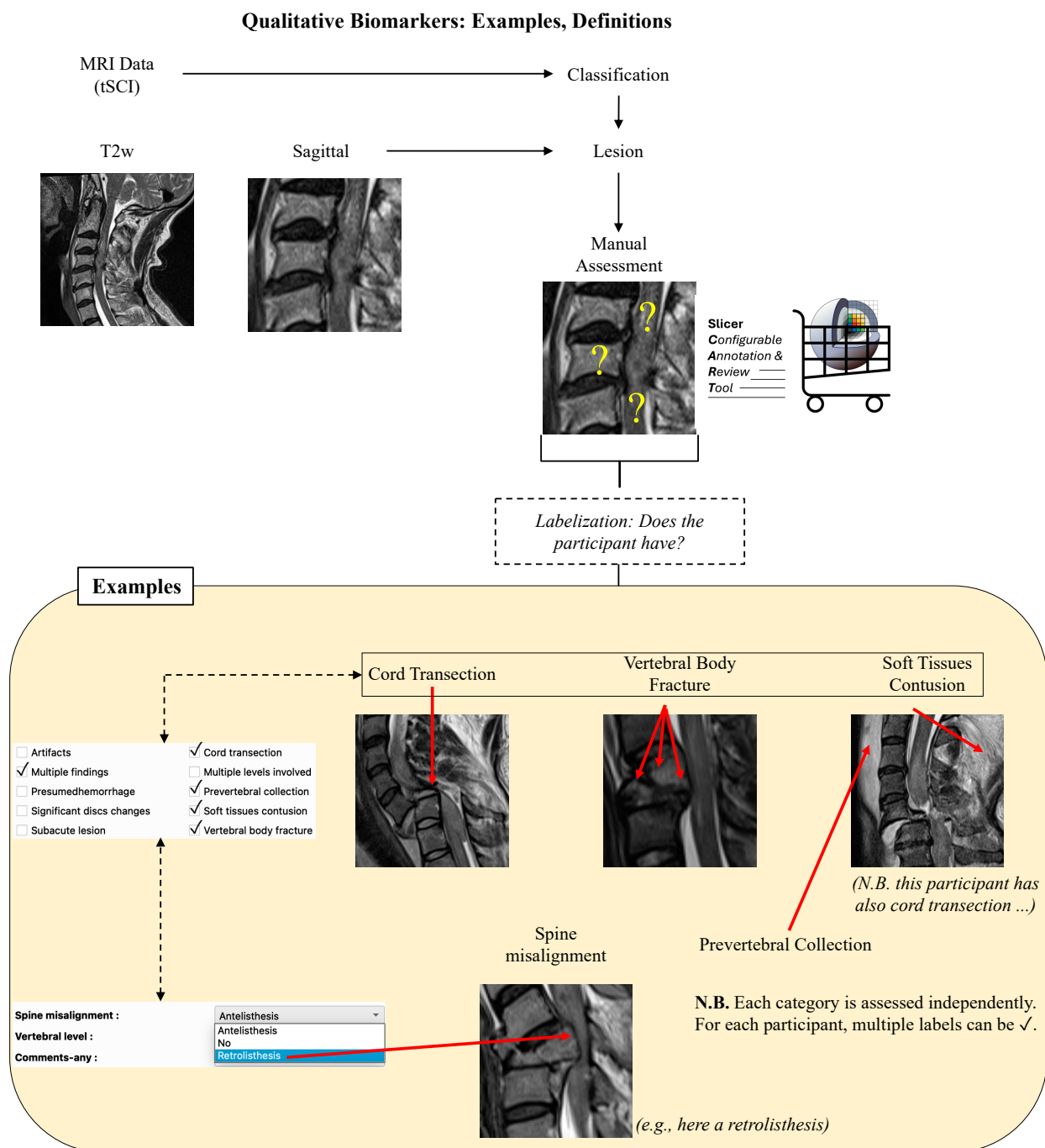


Figure 3.5 Illustrated Examples of Qualitative MRI Labels Classified

From MRI data available using sagittal T2-weighted sequences. Classification labels are obtained manually by assessment of a third-year radiology resident using SlicerCART.

| Variable | Category | Description |
|------------------------------|-----------------------|--|
| Accentuated positioning | dichotomous | Presence of abnormal cervical cyphosis or exaggerated lordosis (yes or no) |
| Artifacts | dichotomous | Presence of any kind of artifacts in the image (yes or no) |
| Congenital spine variants | dichotomous | Presence of congenital spinal cord anatomical variant (e.g., Klippel-Feil syndrome, yes or no) |
| Cord transection | dichotomous | Presence of spinal cord transection (yes or no) |
| Disc buldging | dichotomous | Presence of intracanalicular disc bulding at level of injury (yes or no) |
| Multiple findings | dichotomous | Flag indicating the patient has ≥ 3 significant findings from radiological assessment (yes or no) |
| Multiple levels involved | dichotomous | Intramedullary lesion > 1 vertebral body expected height (yes or no) |
| Other miscellaneous findings | dichotomous | Presence of tSCI-unrelated findings (yes or no) |
| Prevertebral collection | dichotomous | Presence of fluid accumulation in prevertebral space (yes or no) |
| Prevertebral changes | dichotomous | Presence of any other finding in prevertebral space (other than collection, yes or no) |
| Severe spine changes | dichotomous | Presence of likely chronic/prior to injury spinal degenerative changes (yes or no) |
| Soft tissue contusion | dichotomous | Presence of traumatic infiltration of the neck soft tissues (yes or no) |
| Subacute lesion | dichotomous | Suspicion of subacute tSCI lesion (yes or no) |
| Vertebral body fracture | dichotomous | Vertebral body fracture detected on sagittal T2-weighted MRI (yes or no) |
| Vertebral body misalignment | categorical - nominal | Absence or presence of ante or retrolisthesis (yes or no) |

Table 3.4 Listing of MRI Labels for Sagittal T2-Weighted Manual Classification Tasks

3.3.3 Hemorrhage in Acute Traumatic Spinal Cord Injury

Intramedullary hemorrhage is an imaging biomarker that can be qualitative (presence/absence) or quantitative (volume), offering various usage scenarios. Hemorrhage is often associated with worst outcomes, but its role in meta-analyses remains debated. A precise definition enables cross-study comparisons; classification and segmentation agreement clarify its use.

Definition of Hemorrhage

In this work, intramedullary hemorrhage is defined distinctly based on the MRI sequence. Hemorrhage is suspected from intramedullary T2-weighted hypointensity relative to normal-appearing spinal cord, and confirmed on T2* imaging by the presence of blooming artifacts. Additional considerations include lesion location. Figure 3.6 illustrates its applied definition:

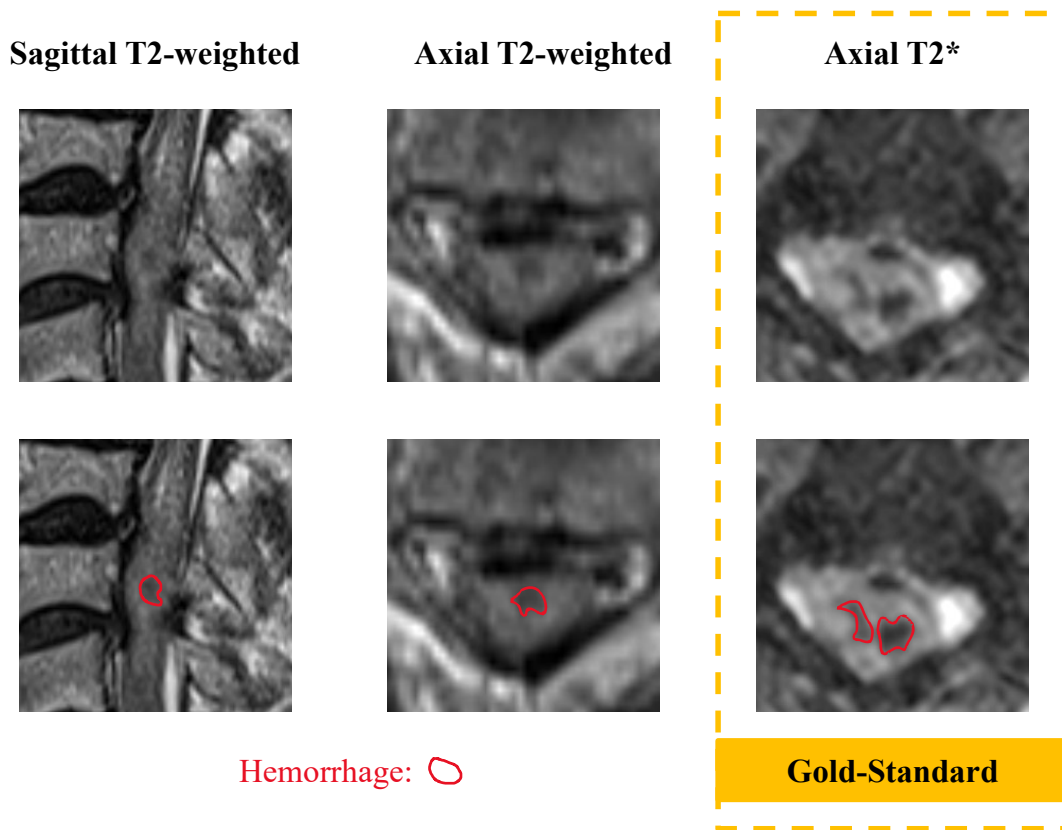


Figure 3.6 Suspicion and Confirmation of Intramedullary Hemorrhage on MRI

Hemorrhagic components are suspected from sagittal and axial T2-weighted magnetic resonance imaging, and confirmed on axial T2*. From the same participant in the *Rick Hansen Spinal Cord Injury Registry*.

3.3.4 Reliability Assessment

Assessing the correlation and variability of both quantitative and qualitative MRI biomarkers is essential to evaluate their reproducibility and applicability across settings. As manual tSCI lesion segmentations are considered the ground truth from which quantitative biomarkers are extracted, evaluating the reproducibility of manual segmentation is a critical first step. In this context, quantitative values of axial and sagittal MRI-derived metrics are correlated on a per-subject basis using scatter plots, and separately for manual and automatic segmentation. Only participants with both sagittal and axial measurements are included, with sample sizes reported. Depending on data normality, either Pearson or Spearman correlation coefficients ([H2.1](#)) are computed. Central tendency and dispersion are reported as median (interquartile range). This analysis highlights the variability of MRI metrics across MRI views in manual segmentation and facilitates direct comparison with metrics from automatic methods.

To address intra-rater variability, sagittal T2-weighted MRI from 127 individuals (single site) have been manually segmented twice by a third-year radiology resident using SlicerCART, six months apart under identical working conditions. Lesion segmentation masks included both edema—i.e. intramedullary T2-weighted hyperintensity relative to normal-appearing spinal cord—and, when present, suspected hemorrhage—see subsection 3.3.3. Dice similarity coefficients ([H2.2](#)) are computed for both labels as one lesion and separately. Total lesion volume is also calculated, as it influences the reliability of the Dice coefficient—smaller lesions tend to yield over-penalized Dice [199]. Results are reported using mean (min; max), median (interquartile range), 95% confidence intervals (calculated with bootstrapping, $n = 10,000$, and numpy [200] library in Python using [`random_default_rng` function with 42 seeds]), and histograms (50 bins from 0.0 to 1.0) to visualize Dice coefficients distribution for each class.

Classification reliability is also assessed. Using the same 127 individuals, the reproducibility of hemorrhage identification on T2-weighted MRI is evaluated by counting non-empty hemorrhage masks across both segmentation sessions. Hemorrhage identification reliability is additionally assessed on axial T2* MRI across two sessions, three months apart under identical conditions, using SlicerCART. Cohen’s kappa ([H2.3](#)) is used to measure agreement, reflecting reproducibility, and 95% confidence intervals are calculated using bootstrapping ($n = 1,000$, numpy [200] library in Python using [`np.random_default_rng` function with 42 seeds])). Classification of hemorrhage for both methods and sessions are reported in contingency tables. Sensitivity and specificity of hemorrhage identification on sagittal T2-weighted imaging are calculated against corresponding axial T2* MRI assessments at both sessions, considered as the reference standard. Inter-rater agreement, for example with a board-certified neuroradiologist with ≥ 10 years of experience, is not part of this thesis.

3.4 Objective 3: To Develop Outcome Prediction and Classification Models From Clinical and MRI Metrics

The following section incorporates guidance from two separate one-hour consultations with a professional statistician working at the University of Montreal, in addition to the concepts learned from the literature review (Section 2.3 in Chapter 2). The current protocol is adapted from the unregistered *Rick Hansen Spinal Cord Injury Registry* (RHSCIR) imaging substudy. Descriptive statistics help in understanding the cohort characteristics and set the conditions in which models apply. Inferential statistics/machine learning guide prediction models development and implementation. Python 3.10 scripts serve for all data manipulations. Variables with $\geq 5.0\%$ of missing values are removed from the final analysis—except for MRI metrics.

3.4.1 Descriptive Statistics: Contextualization

A retrospective descriptive observational cohort substudy has been conducted to describe demographics, clinical and MRI data related to the tSCI cohort. This approach addresses and discloses most relevant features used in assessing the generalizability of prediction models across different datasets. The Praxis Institute team provided demographic and clinical data in a Excel file. MRI parameters have been extracted from BIDS [190] metadata available.

Descriptive features selection relies on literature, clinical expertise and data availability. The described variables are listed in Table 3.5, which exclude MRI biomarkers explored in this work. Outcomes of interest are separately described and include *American Impairment Scale* (AIS) [11] grade, total motor score, and upper (UEMS) and lower (LEMS) extremity motor scores. Those outcomes are described at rehabilitation discharge, alongside with their changes from baseline. The length of stay and discharge destination are described, but not predicted in subsequent prediction modeling. *Functional Independence Measure* (FIM) [201] and *Spinal Cord Independence Measure* (SCIM) [46, 201] are not addressed. The incidence of first delirium or urinary tract infections are not studied.

Continuous described variables are reported as mean (min ; max) and median (interquartile range); categorical variables, including dichotomous ones, as count (%). No site stratification is performed in the descriptive analyses. Reported percentages reflect conditional missing values—data absent only under specific circumstances. Descriptive analysis of outcomes includes non-stratified histograms using various bin size covering the full range of continuous data, and bar plots or pie charts for categorical outcomes. Data description summary through tables, pie charts and bar plots provide an overview of the cohort characteristics.

To assess cohort selection bias ([H3.1](#)), a subgroup analysis by the Praxis Institute compares participants with available MRI data to those without, based on age, age group, mechanism of injury, initial AIS grade, rostral single neurological level of injury at admission, proportions of high complete, low complete, and incomplete tetraplegia, and surgical management status.

| Variable | Category | Description |
|-------------------------------------|-----------------------|--|
| Site | categorical - nominal | Center location of participant |
| Sex | dichotomous | (male or female) |
| Age at injury | continuous | Age at time of injury (years) |
| Age at injury - grouped | categorical - ordinal | Age group that participant belongs to (years) |
| Comorbidity count | continuous | Prior comorbidity (count) |
| Mechanisms of injury | categorical - nominal | Mechanism that caused the injury (≥ 0) |
| Time from injury to initial exam | continuous | Time from injury to initial neurological exam (hours) |
| Time from injury to surgery | continuous | Time from injury to surgical decompression (hours) |
| AIS grade initial | categorical - ordinal | AIS grade at initial neurological exam (A-D) |
| Single neurological level initial | categorical - ordinal | Rostral neurological level involved at initial neurological exam (C1-C8) |
| Neurological status initial | categorical - ordinal | Tetraplegia status (high complete, low complete or incomplete) |
| Upper extremity motor score initial | continuous | Upper extremity motor score at initial neurological exam (0-50) |
| Lower extremity motor score initial | continuous | Lower extremity motor score at initial neurological exam (0-50) |
| Motor score initial | continuous | Total motor score at initial neurological exam (0-100) |
| Surgical management | dichotomous | Participant had surgery (yes or no) |
| Intra-operative adverse event | dichotomous | Participant had intra-operative adverse event (yes or no) |
| Final discharge destination | categorical - nominal | Destination at rehabilitation discharge |
| MRI manufacturer | categorical - nominal | Manufacturer of MRI data |
| MRI field strength | dichotomous | 1.5 or 3.0 (Tesla) |
| MRI acquisition type | dichotomous | Acquisition dimensions (2D or 3D) |

Table 3.5 Listing of Demographic, Clinical, and MRI Features

AIS = American Spinal Injury Association Impairment Scale; MRI = Magnetic Resonance Imaging. Anal contraction and sensation reflex, bulbocavernous reflex, presence of cauda equina syndrome and other neurological deficit are excluded due to missing information.

3.4.2 Predictive Modeling and Classification

Data Preparation and Encoding

Demographics and clinical data listed in Table 3.5 are extracted from the Praxis Excel file. Final AIS grade, total motor score, upper and lower extremity motor score, and their changes from baseline, are also extracted. Time from injury to initial exam, intra-operative adverse event, and final discharge destination are not. Jan Valošek (post-doctoral fellow at NeuroPoly Lab) generates most MRI metrics using the *Spinal Cord Toolbox* [196] (SCT) from both automatic and manual lesion segmentations (see Section 3.3). MRI labels listed in Table 3.4 are obtained through manual classification tasks. Additionally, BASIC-Score [59] is manually assessed on all included axial T2-weighted MRI. Count of non-empty hemorrhage masks for sagittal (second segmentation session) and axial T2-weighted MRI are performed and added to final data through Python scripts. All data are merged to a final tabular CSV file. From the available columns, scripts encode variables based on their type, sometimes with manual input (e.g. categorical data). One hot-encoding [202] is performed for categorical variables. The encoded data is saved into a new final CSV file for model training and testing. No imputation is done for missing data within the model training pipeline.

Input Variables Selection

The optimal input metrics combination for accurately predicting a given outcome is unknown. With ~300 participants, including all variables in one prediction model while maintaining a ratio of ≥ 10 observations per variable is unfeasible. Exploring all combinations of input variables is computationally unrealistic (see Chapter 3 in [132]). Therefore, manual pre-selection identifies clinically relevant metrics based on their ability to answer to ≥ 1 of:

1. Is the metric a demographic variable available before the clinical exam and/or MRI?
2. Is it part of the Gold Standard clinical exam?
3. Is it available at the time of MRI acquisition?
4. Is it readily accessible on the MRI console, for clinicians?
5. Can it be categorized (e.g. yes/no)?

An exception is made for manual MRI metrics, which serve as ground-truth when evaluating prediction capabilities of automatically derived metrics. Selected input variables are grouped into 7 categories based on metric type—demographics, clinical, MRI (manual vs automatic, sagittal vs axial), and MRI labels—to discriminate the best predictive feature categories.

Dichotomization of Outcomes of Interest

Some authors advise against dichotomizing outcomes [102,103,105] due to the associated loss of information and the conceptual distinction between classification and prediction. However, dichotomization simplifies statistical modeling by enabling the use of logistic regression and facilitating the interpretability of overall classification performance assessment. Even critics like Harrell et al. [102] acknowledge its widespread use, but highlight the importance of understanding its limitations. Accordingly, a single dichotomization threshold is applied (> 0 versus ≤ 0), distinguishing participants with improvement from stability or worsening. We acknowledge that the concept of *minimal clinically important difference* [203] is not directly applied. Dichotomized outcomes are added as new columns in the encoded CSV file.

Models Selection, Training, Testing, and Validation

This work evaluates the prediction of AIS grade, total motor score, upper and lower extremity motor score at rehabilitation discharge, and their changes from baseline (see Table 3.6). Classification tasks are also conducted to distinguish participants who show improvement at rehabilitation discharge from those who show either no change or clinical deterioration. For AIS grade (A to E), where E is recovery to normal (i.e., less severe) and A complete lesion (i.e., most severe), ordinal non-linear intervals are considered. The number of input variables is likely higher than 10 per observation, so feature selection approaches are essential. From the literature review, transfer learning is not retained. Accordingly, all models are developed using the scikit-learn or MORD [204] libraries in Python [205]. Models development is divided in two parts: 1) the first for prediction modeling using full available data; 2) the second for prediction modeling using machine learning, where the dataset is randomly split three times using an 80:20 ratio. For each split, 5-fold cross-validation is applied to the training set, and model validation is performed on the remaining 20% unseen test data.

An initial feature selection is performed through independent univariate analyses between each input variable and each outcome (H3.2). The results are summarized in tabular format, and the top 15 associations identified, along with the initial severity, are reported for descriptive purposes. Variables showing a statistically significant association at a conservative threshold of $p \leq 0.10$ are further selected through *Least Absolute Shrinkage and Selection Operator* [115] (LASSO) at $\alpha = 0.10$ (ensuring considerable dimensionality reduction). Those with non-null coefficients are retained for model development. For the first approach (“statistics”), subsequent LASSO, linear, and elastic net regressors from `scikit-learn` [205] (LASSO and logistic regression for dichotomized outcomes; ordinal logistic regression for ordinal outcomes using MORD [204]) are trained on the full dataset using each predefined

hyperparameter configuration (see Appendix D for full listing). It is important to note that these full-data models are limited to no more than three distinct hyperparameters, in accordance with Goodfellow et al. [143], and each with three candidate values. For each continuous outcome, the best-performing model is identified based on the lowest *Prediction Sum of Squares* (PRESS) statistic [206], the area under the curve (AUC) for binary outcomes, and accuracy for ordinal outcome. The corresponding coefficient of determination R^2 , Brier score and accuracy are also reported when applicable (H3.3). This provides an overview of the best possible predictive performance (H3.3), while acknowledging the risk of overfitting. As these are generalized linear models, the top 10 coefficients by absolute magnitude are additionally visualized using horizontal bar plots, to facilitate interpretability (H3.3.1).

For the second approach (“machine learning”), the same initial feature selection approach is applied (univariate analyses [H3.2] + LASSO at $\alpha = 0.10$). The selected features are then incorporated into machine learning–based models, including random forest, LASSO, elastic net, and gradient boosting algorithms (XGBoost, LightGBM), as well as linear regression models. For classification tasks, selected features from the same process are used to train machine learning–based classifiers such as random forest, gradient boosting (XGBoost, LightGBM), and support vector machines. Logistic regression models are also developed. Neural networks, K-nearest neighbors, and linear discriminant analysis are excluded due to concerns regarding limited sample size, high dimensionality, and unmet assumptions of normality, respectively. Model training is performed through hyperparameter optimization using random search across arbitrarily predefined hyperparameter grids for each model (see Appendix D). A 5-fold cross-validation is applied during this process. Among all cross-validation folds, the best-performing model is selected based on the coefficient of determination (R^2) for continuous outcomes, the area under the curve (AUC) for dichotomized outcomes and accuracy for ordinal outcomes. The selected model is then recalibrated on the full training set and subsequently validated on the remaining 20% of unseen data. For each model for each outcome, the performance is tracked for the three splits using the average R^2 from scikit-learn [205] for continuous outcomes and average AUC for dichotomized outcomes. Heatmaps using Seaborn library [207] illustrate the average performance of each model per outcome, for prediction and classification tasks. For each outcome, the performance results for the best model are based on R^2 for continuous outcomes and AUC for dichotomized outcomes. The DeLong test [208] for comparing AUC is not applied here. Accuracy and Brier scores are reported when applicable. The best model performance results are visualized using: 1) Predicted versus observed value plots (receiver operating characteristic curves for dichotomized outcomes, including 95% confidence intervals from bootstrapping; confusion matrix for ordinal outcome); 2) Calibration curves (or pseudo-calibration curves for continuous outcomes, gen-

erated by plotting the mean predicted value within each of 10 equally sized, ordinaly ranked bins), and 3) SHapley Additive exPlanations (SHAP) [209] summary plots or Feature Permutation Importance plots, depending on the model (to interpret feature importance, [H3.3.1](#)). Regardless of model performance, these plots are generated automatically in the modeling pipeline, so caution is advised. These visualizations support the identification of the most discriminative, calibrated and contributing features related to prognosis and outcome prediction. TRIPOD+AI 2024 [112] and PROBAST [107] check adherence to model development guidelines (see Appendix E and F, respectively). The prediction models, including scripts, are available upon request and through spineimage.ca. An exploratory subgroup analysis, stratified by initial AIS grade, is conducted using hierarchical logistic regression to assess the added value of MRI biomarkers over clinical examination alone in classifying participants who improve versus not. AUC are compared using DeLong test. Across all variable categories, the number of baseline features included is determined by the lowest event rate among the subgroups, and this same number of features is retained across all to ensure comparability.

| Outcome | Considered type | Range of values | Prediction regressors | Classifier developed |
|------------------------------------|-----------------|-----------------|-----------------------|----------------------|
| AIS (final) | ordinal | [E, A] | - | ✓ |
| AIS change | continuous | [-3, 3] | ✓ | - |
| AIS change binarized | dichotomous | {Yes, No} | - | ✓ |
| Total motor score (final) | continuous | [0, 100] | ✓ | - |
| Total motor score change | continuous | [-100, 100] | ✓ | - |
| Total motor score change binarized | dichotomous | {Yes, No} | - | ✓ |
| UEMS (final) | continuous | [0, 50] | ✓ | - |
| UEMS change | continuous | [-50, 50] | ✓ | - |
| UEMS change binarized | dichotomous | {Yes, No} | - | ✓ |
| LEMS (final) | continuous | [0, 50] | ✓ | - |
| LEMS change | continuous | [-50, 50] | ✓ | - |
| LEMS change binarized | dichotomous | {Yes, No} | - | ✓ |

Table 3.6 Listing of Outcomes Used for Model Development

AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper extremity motor score; LEMS = Lower extremity motor score. AIS grade is considered as a continuous outcomes and linear severity between initial categories is assumed: thereby, AIS values range are from AIS E (recovery to normal) to AIS A (complete lesion, most severe), considering ordinal non-linear intervals across categories. Change corresponds to the change of score between initial admission (acute) and measure at rehabilitation discharge (final). The “Classifier developed” and “Prediction regressors” columns indicate the models developed.

Summary of Classification and Prediction Models Development

All using Python scripts, All training data recorded
TRIPOD and PROBAST grid filled ✓

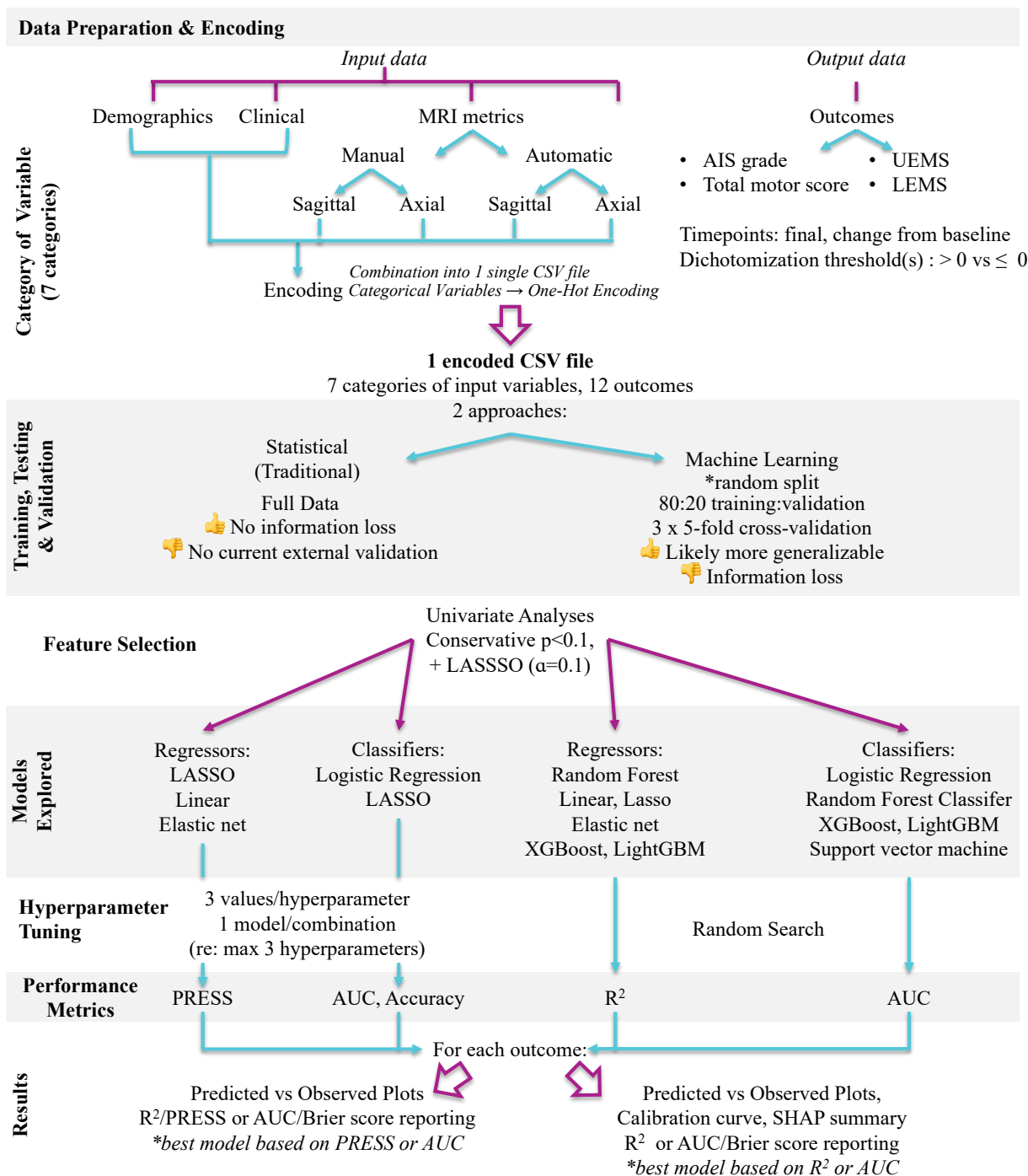


Figure 3.7 Outcome-Dependent Prediction Models Development Summary (Objective 3)

TRIPOD = transparent reporting of a multivariable prediction model for individual prognosis or diagnosis [106,112]; PROBAST = prediction model risk-of-bias assessment tool [107]; MRI = Magnetic Resonance Imaging; AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper extremity motor score; LEMS = Lower extremity motor score; LASSO = Least Absolute Shrinkage and Selection Operator; PRESS = Prediction Sum of Squares; AUC = area under the receiver operating characteristic curve; R^2 = coefficient of determination. Ordinal logistic regression is applied for ordinal outcome.

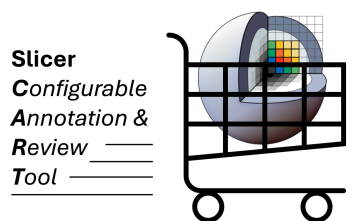
CHAPTER 4 RESULTS

This chapter reports realizations and findings from this thesis, with teamwork-related inputs specified where applicable. First, it addresses the development output of SlicerCART and its usage for manual segmentation and classification. Next, the MRI metrics extracted from automatic and manual segmentation are compared, where intra-rater variability assesses reliability of manual labeling. Finally, it outlines the results of prediction modeling using clinical and MRI data. Detailed interpretation is provided in the Discussion (see Chapter 5).

4.1 SlicerCART

Figures 4.1 to 4.7 illustrate key outputs from SlicerCART development. As a proof of concept, the module has been applied to manual segmentation and classification tasks within the RHSCIR imaging substudy, using tSCI MRI data. Intramedullary cervical lesion segmentation masks were manually delineated on 240 sagittal T2-weighted and 245 axial T2-weighted MRIs, with total segmentation times of 04:49:11 and 07:32:13 (HH:MM:SS), respectively. This corresponds to an average of 1 minute and 12 seconds for sagittal MRI and 1 minute and 50 seconds for axial MRI. Corresponding spinal cord segmentation masks were generated using the contrast-agnostic model [198] from the *Spinal Cord Toolbox* [196], followed by quality control and manual corrections in 241 axial MRI (automatic segmentation failed in 4 participants), requiring 03:06:14 (average of 46 seconds per case). Manual classification of 240 participants using sagittal T2-weighted MRI for 15 labels (see Table 3.4 in Chapter 3) took 02:57:57 (average of 45 seconds per case). Binary classification of axial T2* hemorrhage (306 acquisitions), considering participants with > 1 sequence) took 00:31:23 at baseline and 00:56:35 at 3 months (average of 06 and 11 seconds per case, respectively). Times reflect visualization only, excluding data processing ([H1](#)).

Several core features have been implemented by the author to ensure usability, including: segmentation and classification workflows, refactoring for robust label configuration, support for task resumption, version control, editable previous version segmentation masks, interpolation toggling, customizable mouse controls, and basic documentation. An Ni Wu (initial ICH_SEGMENTER), Delphine Pilon and Kuan Yi Wang have been contributing directly or indirectly to the module (e.g. instantiation of the dynamic graphical user interface used in customization of segmentation and classification labels [refined by the thesis author], and automatic time recording per task). SlicerCART remains an ongoing project.



Imaging Research. Made Simple.

Optimized Annotation

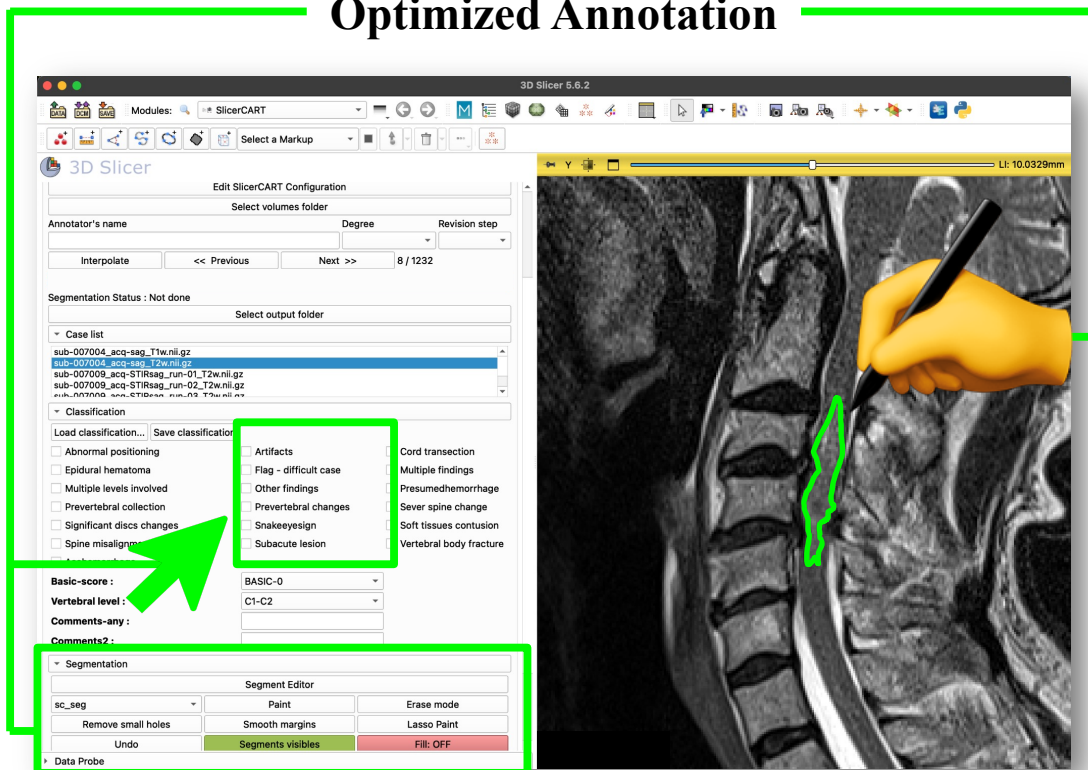


Figure 4.1 Overview

SlicerCART supports efficient imaging data navigation, labeling, and segmentation. This figure shows the module graphical user interface at launching. SlicerCART requires installation of 3D Slicer, an open-source software. Instructions are provided in the documentation. This module was built upon a prototype from Dr. Laurent Létourneau-Guillon's team; the current version has been significantly refractored by the author.

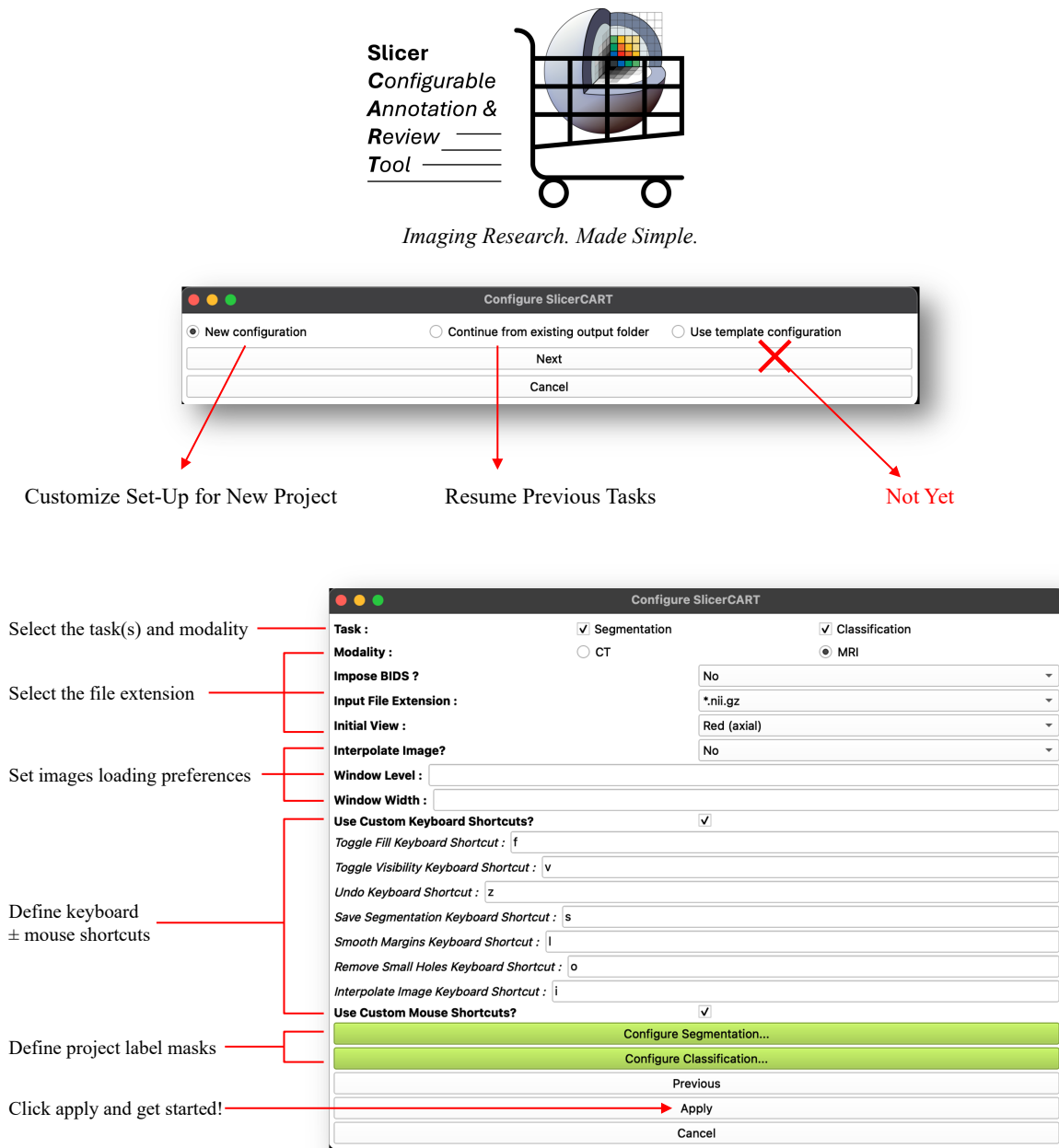
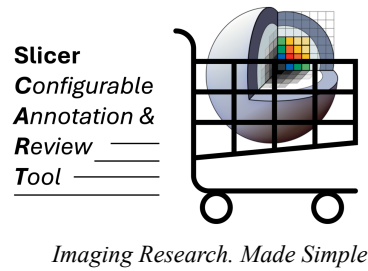


Figure 4.2 Configuration Set-Up

SlicerCART can be configured to accommodate task-specific requirements and user preferences, depending on the intended use. Some configuration examples are illustrated. This initial interface was developed by Delphine Pilon and Kuan Yi Wang during summer 2024, and was subsequently refactored by the author to ensure proper working. While further improvements are planned, the current configuration setup is usable.



SlicerCART supports project-specific and task-specific label configurations

Segmentation

Classification

Configure Segmentation

| | | Name | Value | Colour |
|---|-------------|------------|-------|--------|
| 1 | Edit Remove | edema | 1 | |
| 2 | Edit Remove | hemorrhage | 2 | |

Add Label

Display timer during segmentation? ☐

Apply

Cancel

Configure Classification

| | | Label |
|---|--------|----------------------|
| 1 | Remove | Abnormal positioning |
| 2 | Remove | Artifacts |
| 3 | Remove | Cord transection |
| 4 | Remove | Epidural hematoma |

Add Checkbox

| | | Label | |
|---|--------|-----------------|---------|
| 1 | Remove | Basic-score | BASIC-0 |
| 2 | Remove | Vertebral level | C1-C2 |

Add Drop Down

| | | Label |
|---|--------|--------------|
| 1 | Remove | Comments-any |
| 2 | Remove | Comments2 |

Add Text Field

Save

Cancel

Configure Label

Name :

Value :

N.B. Label value is not editable and will be assigned automatically. Enter any integer between 0 and 255 in RGB to select label color.

Colour : R G B

Save

Cancel

Configure Classification

Item Name :

Adjust the DropDown count. Then, edit the option names. If no text, option number will be used as default. N.B. Options will be saved sorted.

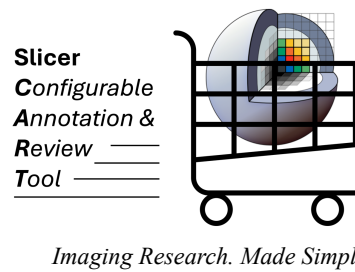
Number of options:

Save

Cancel

Figure 4.3 Label Selection

The module allows users to add, remove, and customize labels for manual segmentation and classification. Interface elements such as combo boxes, drop-down menus, and text fields can be configured based on the specific requirements of the task.



Configuration set-up ensures that labels are compatible with module utilization

Segmentation

Classification

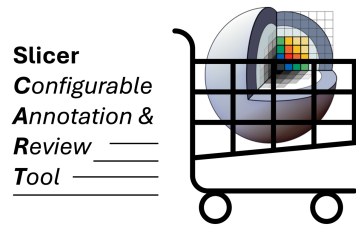
Checkboxes and Comment boxes are added, and label name is automatically adapted to meet SlicerCART encoding standards

A text bar indicates the color code validity
A compatible label value is automatically assigned

A dynamic graphical user interface adapts from the number of options, enabling combo boxes creation

Figure 4.4 Label Standardization

In models utilizing segmentation masks, proper labeling is crucial. The author reinforced the label saving and selection process for both segmentation and classification tasks, ensuring standardized formatting.



Imaging Research. Made Simple.

A volume and output folder enable to set-up the work environment. Annotator's name, degree and revision step facilitate tracking.

Select volumes folder

Annotator's name: Degree: Revision step:

Interpolate << Previous Next >> 18 / 429

ximebouthillier/gitmax/data_confid/praxis/site_006/sub-mon031/anat/sub-mon031_acq-sag_T2w.nii.gz

Segmentation Status : Not done

Select output folder

A case list makes navigation efficient

Case list

- sub-mon031_acq-axial_T2starw.nii.gz
- sub-mon031_acq-axial_T2w.nii.gz
- sub-mon031_acq-sag_T2w.nii.gz
- sub-mon031_acq-sag_run-01_T1w.nii.gz
- sub-mon031_acq-sag_run-02_T1w.nii.gz

The Classification window displays user-configured labels relevant to the case displayed in the viewer

Classification

Load classification... Save classification

☐ Abnormal positioning
 ☐ Artifacts
 ☐ Cord transection

☐ Flag - difficult case
 ☐ Multiple findings
 ☐ Multiple levels involved

☐ Prevertebral collection
 ☐ Prevertebral changes
 ☐ Significant discs changes

☐ Soft tissues contusion
 ☐ Subacute lesion
 ☐ Vertebral body fracture

Spine misalignment : Antelithesis

Vertebral level : C1-C2

Comments-any :

The Segmentation window centralizes annotation tools in one location

Segmentation

Segment Editor

Paint Erase mode

Remove small holes Smooth margins Lasso Paint

Undo Segments visibles Fill: OFF

Place a measurement line

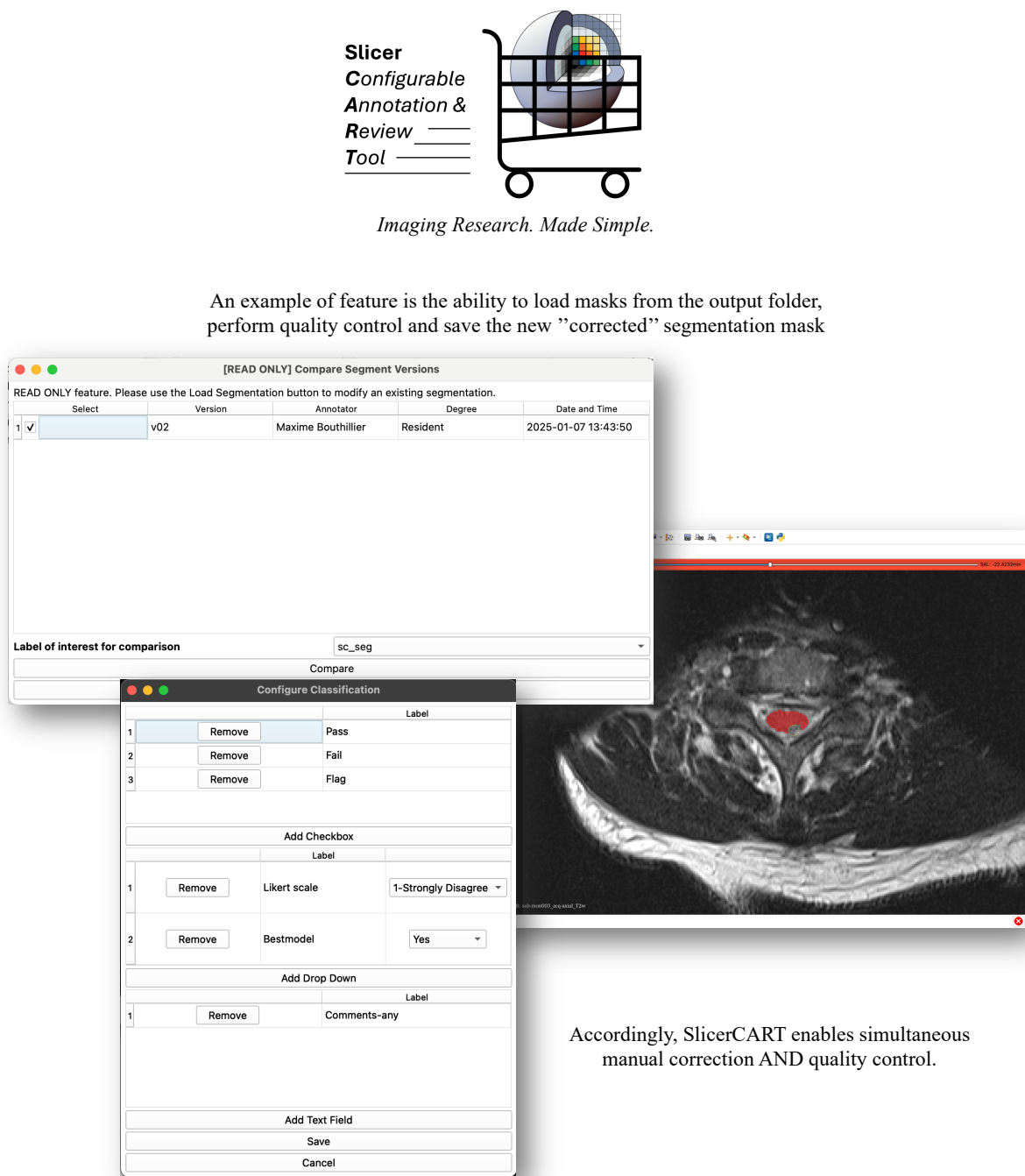
Save segmentation

Load segmentation Load latest masks

Compare segment versions

Figure 4.5 User Interface

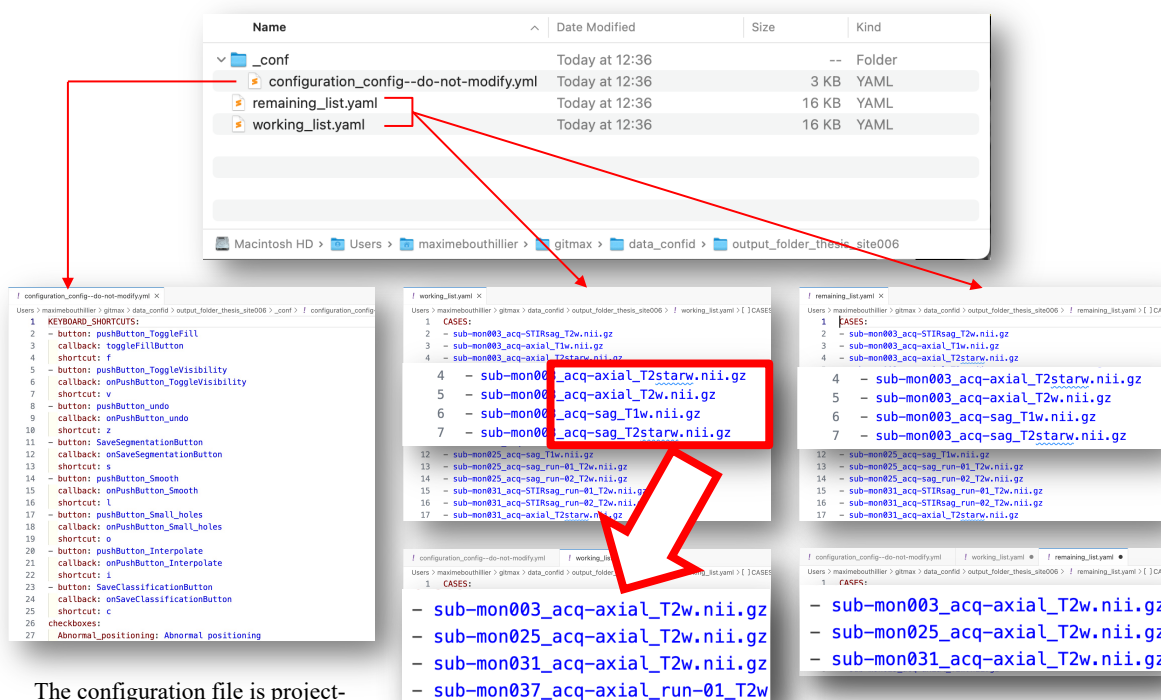
Once the configuration is complete, one essential step remains before starting: the user must select the volume folder (where imaging data are located) and output folder (where masks, processed data, and other task-related files will be saved). The annotator name, degree and revision step can be added for tracking purposes, with the date and time automatically logged at each case-version save. The Case list facilitates efficient navigation through large datasets. The Classification window displays the configured labels. The Segmentation window centralizes key tools for annotation.





Imaging Research. Made Simple.

Selecting an output folder per task/project enables multiple configuration usages.



The configuration file is project-specific, enabling tailored setups.

Work lists are saved in YAML files where the remaining list contains the cases to task. Lists can be customized by the user (e.g. filter and keep only T2w sagittal images). This enables to resume previous tasks in further sessions.



Figure 4.7 Work Environment

When a new output folder is selected, SlicerCART automatically generates a configuration file and worklist files. The configuration file serves as a future reference for resuming the task later—especially useful if SlicerCART has been used for another project with a different setup in the meantime. The worklist can be customized to include or exclude specific cases based on filenames, allowing targeted analysis and streamlined navigation within large datasets. **Important:** the remaining list must be \leq than the working list. If the files are incompatible, the user will be prompted to verify them.

4.2 Imaging Biomarkers Variability

Figure 4.8 presents MRI-derived metric values for each participant, from axial and sagittal T2-weighted MRI, using both manual and automatic lesion segmentation methods ([H2.1](#)):

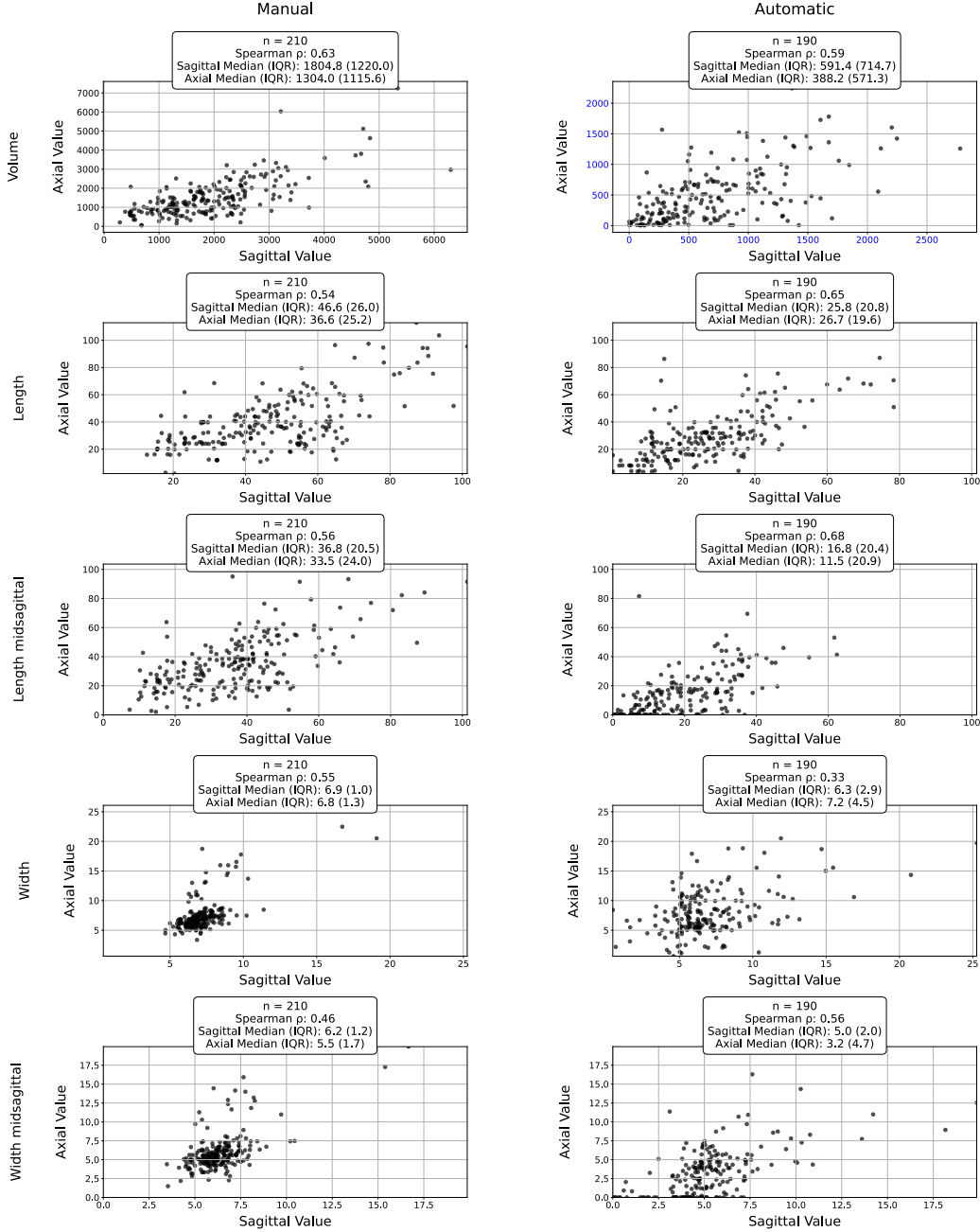
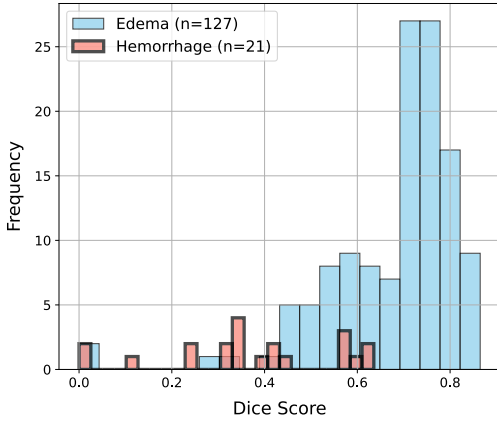


Figure 4.8 MRI Metrics Values Calculated on Axial Versus Sagittal View for Manual and Automatic Segmentation Methods

Volume units in mm^3 , else in mm. Reminder: MRI metrics from manual segmentations are derived from postprocessed masks, constrained to the spinal cord region.

The intra-rater variability for manual lesion segmentation and hemorrhage classification is assessed for 127 participants from a single site. On sagittal T2w MRI, the mean (CI: 95% confidence interval) Dice coefficient for edema is 0.68 (0.65-0.70); for hemorrhage: 0.37 (0.29-0.45); for total lesion: 0.69 (0.66-0.71) (H2.2). Without any postprocessing, the mean manual lesion volume is 2339 mm³ (95% CI: 2101–2573) initially, and 2526 mm³ (95% CI: 2288–2768) at 6 months. Figure 4.9 shows the distribution of Dice coefficients, and descriptive statistics:



Dice Coefficients

| Description | Edema | Hemorrhage | Whole Lesion |
|-----------------|-------------------|-------------------|-------------------|
| Mean (min; max) | 0.68 (0.00; 0.86) | 0.37 (0.00; 0.64) | 0.69 (0.00; 0.87) |
| Median (IQR) | 0.72 (0.16) | 0.35 (0.25) | 0.73 (0.17) |
| 95% CI | 0.36 - 1.02 | -0.05 - 0.93 | 0.66 - 0.71 |

Lesion Volume (mm³)

| Description | Initial | 6-months |
|-----------------|------------------|----------------|
| Mean (min; max) | 2339 (173; 7629) | 2526 (0; 8359) |
| Median (IQR) | 2032 (1775) | 2328 (1548) |
| 95% CI | 2101-2573 | 2288-2768 |

Figure 4.9 Distribution of Dice Scores for Non-Empty Masks at 6-months Interval

IQR = Interquartile range; CI = Confidence interval (derived via bootstrapping). Whole lesion refers to the combination of the edema and hemorrhage masks for each lesion, prior to Dice coefficient computation. Lesion volume refers to the raw manual segmentation volume, without any postprocessing.

Hemorrhage was initially diagnosed on sagittal T2-weighted MRI in 21 participants, then 31 at 6 months (kappa: 0.66; CI: 0.49-0.84). On axial T2* MRI (n=115), hemorrhage was classified in 27 and 31 participants at initial and final sessions, respectively (kappa: 0.77; CI: 0.61-0.90) (H2.3). Sensitivity and specificity for hemorrhage detection using sagittal T2w versus axial T2* imaging are 0.56 and 0.86 initially, and 0.55 and 0.98 on second evaluation. Table 4.1 shows contingency for manual classification of hemorrhage at specific intervals:

| | | 6-months Interval (non-empty masks, n) | |
|---------------------------------|-----|---|----|
| | | Yes | No |
| Initial (non-empty masks, n) | Yes | 19 | 12 |
| | No | 2 | 94 |

| | | 3-months Interval (hemorrhage, n) | |
|----------------------------|-----|--------------------------------------|----|
| | | Yes | No |
| Initial (hemorrhage, n) | Yes | 24 | 3 |
| | No | 7 | 81 |

Table 4.1 Intra-Rater Reliability in Hemorrhage Binary Classification 6-months Apart on Sagittal T2-weighted MRI and 3-months Apart on Axial T2* MRI

4.3 Predictive Modeling and Outcomes Classification

The role of imaging biomarkers in acute preoperative tSCI is assessed through the development of prediction and classification models targeting AIS grade, total motor score, and upper and lower extremity motor scores. The scores at final rehabilitation discharge and their changes from baseline are studied. Two preliminary steps are undertaken: first, a descriptive observational substudy characterizes and clarifies the cohort profile, which defines the applicability of the prediction modeling findings; second, univariate analyses (using $p = 0.05$) are performed to identify significant associations between outcomes and Demographics, Clinical and MRI variables at initial evaluation, final rehabilitation discharge and their changes from baseline. Finally, the performance of statistical and machine learning models is presented, highlighting their predictive accuracy, transportability and most contributing features.

4.3.1 RHSCIR Imaging Substudy Cohort Profile

Among the 341 participants enrolled from 9 level 1 and 2 trauma centers (discharged from hospital between January 1, 2015 and August 31, 2021), 24% are females. The mean age is 56 (16; 92) years and median 59 (26): 32% are between 61 and 75 years old. Falls (55%) is the leading mechanism of injury, followed by transport (21%) and sports (18%). 75% have initial incomplete tetraplegia, 12% high complete tetraplegia and 12% low complete tetraplegia. 87% have surgical management. Participants stay on average 93 days (1; 465) [median of 70 (128)] before final discharge from rehabilitation, and 65% are discharged to a private residence. Cohort characteristics are reported in Figure 4.10. A subgroup analysis (done by the Praxis Institute team) between 281 eligible participants with MRI and 54 eligible without MRI shows that participants with acute preoperative MRI differ from eligible participants without MRI in their age ($p=0.033$) and initial severity mechanism (AIS D vs AIS A, $p=0.003$), but have no different distributions for sex ($p=0.209$), age group ($p=0.138$), mechanism of injury ($p=0.147$) or surgical management ($p=0.468$) (see Table 4.11) [H3.1].

Regarding severity of injury, 49% have initial *American Spinal Injury Association Impairment Scale* (AIS) grade D and 14% AIS grade A. 64% have no AIS grade conversion at rehabilitation discharge (see Table 4.12). At initial evaluation, the mean total motor score, upper and lower extremity motor score are 49, 25 and 24, respectively. At final rehabilitation discharge, the means are 67, 35 and 33, respectively. Regarding their changes from baseline, the means are 18, 11 and 08 respectively (see Figure 4.13 for visual distributions and detailed descriptions).

| Description | Total |
|--|--------------|
| Number of participants | 341 |
| Number of females (%) | 81 (23.8) |
| Mean age | 55.9 |
| (min; max) | (16.0; 92.0) |
| Median age | 59.0 |
| (IQR) | (26.0) |
| Age groups | |
| 16 - 30 | 47 (13.8) |
| 31 - 45 | 49 (14.4) |
| 46 - 60 | 89 (26.2) |
| 61 - 75 | 108 (31.8) |
| 76+ | 47 (13.8) |
| No prior comorbidity (%) | 144 (42.5) |
| Mechanism of injury (%) | |
| Fall | 186 (54.7) |
| Transport | 72 (21.2) |
| Sports | 61 (17.9) |
| Other traumatic cause | 17 (5.0) |
| Assault - blunt | 3 (0.9) |
| Time (hours) from injury to initial exam | |
| Mean | 17.3 |
| (min; max) | (0.0; 72.0) |
| Median | 10.0 |
| (IQR) | (18.0) |
| Surgical management (%) | 297 (87.1) |
| No intra-operative adverse event (%) | 263 (97.0) |
| Length of stay (days) | |
| Mean | 92.9 |
| (min; max) | (1.0; 465.0) |
| Median | 70.0 |
| (IQR) | (128.0) |
| Final discharge destination | |
| Private residence | 223 (65.4) |
| Hospital Facility | 78 (22.9) |
| Nursing home/Long-term care | 17 (5.0) |
| Assisted/Group residence | 10 (2.9) |
| Morgue | 9 (2.6) |
| Other | 4 (1.2) |

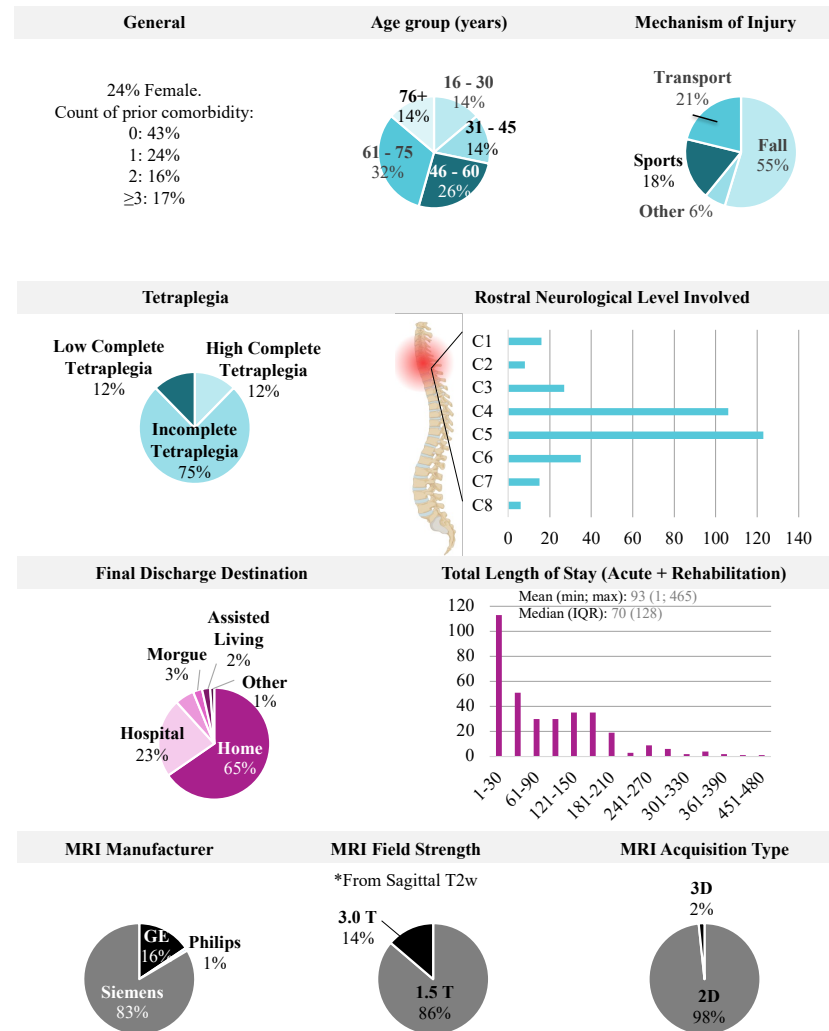


Figure 4.10 Imaging Substudy Cohort Characteristics and Visual Summary

Count (%), mean and median are shown for the whole imaging substudy cohort. The reported % considers missing data.

| Description | Eligible MRI + | Eligible MRI - | p-value |
|-----------------------------|----------------|----------------|---------|
| Female (%) | 22.0 | 31.0 | 0.209 |
| Mean age (years) | 57.0 | 51.0 | 0.033 |
| MOI (%) | | | 0.147 |
| Fall | 57.0 | 48.0 | |
| Transport | 22.0 | 26.0 | |
| Sports | 17.0 | 20.0 | |
| Other | 5.0 | 6.0 | |
| Initial AIS grade (%) | | | 0.003 |
| A | 24.0 | 46.0 | |
| B | 8.0 | 2.0 | |
| C | 21.0 | 22.0 | |
| D | 48.0 | 30.0 | |
| Initial tetraplegia (%) | | | 0.002 |
| Incomplete | 77.0 | 54.0 | |
| High complete | 12.0 | 26.0 | |
| Low complete | 11.0 | 20.0 | |
| Initial level of injury (%) | | | 0.009 |
| C1 | 5.0 | 2.0 | |
| C2 | 2.0 | 4.0 | |
| C3 | 9.0 | 0.0 | |
| C4 | 28.0 | 54.0 | |
| C5 | 36.0 | 28.0 | |
| C6 | 13.0 | 7.0 | |
| C7 | 4.0 | 2.0 | |
| C8 | 1.0 | 2.0 | |
| T1 | 1.0 | 2.0 | |
| Surgical management (%) | 85.0 | 80.0 | 0.468 |

Figure 4.11 Subgroup Analysis: MRI versus No MRI

MOI = Mechanism of Injury; AIS = *American Spinal Injury Association Impairment Scale*.
This table has been produced by the Praxis Institute team.

| | | Final (n) | | | | |
|-------------|---|-----------|----|----|-----|---|
| Initial (n) | | A | B | C | D | E |
| | A | 48 | 17 | 15 | 4 | 0 |
| | B | 0 | 5 | 4 | 14 | 0 |
| | C | 0 | 8 | 9 | 51 | 0 |
| | D | 1 | 0 | 3 | 155 | 7 |

Figure 4.12 Initial and Final *American Spinal Injury Association Impairment Scale* (AIS) Grades at Initial Evaluation and Final Discharge from Rehabilitation

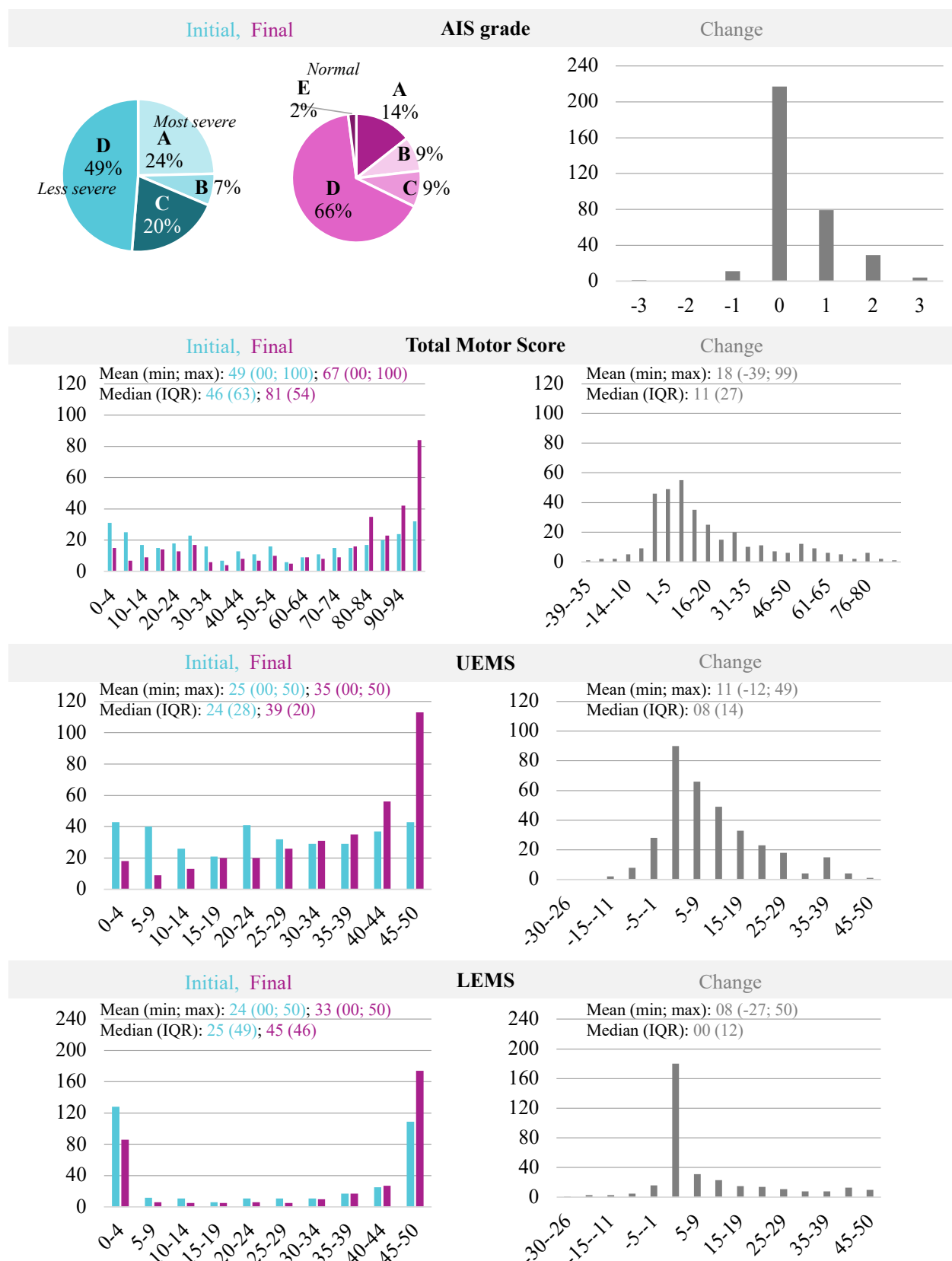


Figure 4.13 Rick Hansen Imaging Substudy Cohort Profile: Severity Assessment

AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score; IQR = Interquartile Range.

4.3.2 Univariate Analyses

Detailed univariate analysis results examining associations between demographic, clinical, and MRI-based variables with neurological outcomes—including initial evaluation and final rehabilitation discharge AIS grade, total motor score, upper and lower extremity motor scores, and their changes from baseline (both absolute and binary > 0 versus ≤ 0)—are presented in Appendix G (H3.2). Figures 4.14, 4.15, 4.16, and 4.17 illustrate the top 15 most significant associations identified through univariate testing (effect sizes are not considered in this ranking). Table 4.2 reports the frequency of significant associations across variable categories and provides the legend for the color coding used in the subsequent tables:








| Color | Category | Initial | Final | Change | Total Count | Variables (n) |
|---|---------------------|---------|-------|--------|-------------|---------------|
|  | Demographics | 6 | 6 | 4 | 16 | 4 |
|  | Clinical | 31 | 28 | 44 | 103 | 10 |
|  | MRI Manual Sag | 24 | 26 | 5 | 55 | 7 |
|  | MRI Manual Axial | 28 | 29 | 2 | 59 | 8 |
|  | MRI Automatic Sag | 24 | 27 | 10 | 61 | 7 |
|  | MRI Automatic Axial | 28 | 28 | 8 | 64 | 8 |
|  | MRI Labels Sag | 28 | 28 | 17 | 73 | 22 |
| | Total | 169 | 172 | 90 | | |

Table 4.2 Frequency of Significant Association Grouped Per Variable Category Using $p = 0.05$

Category corresponds to the variable category. **Initial** corresponds to the initial evaluation timepoint (acute, at admission). **Final** corresponds to the final evaluation timepoint (at rehabilitation discharge). **Change** corresponds to the change in score from baseline. **Total count** corresponds to the number of significant association. **Variables (n)** corresponds to the number of variable per category. Each of the **Initial**, **Final**, and **Change** columns includes outcomes related to AIS grade, total motor score, and upper and lower extremity motor scores.

Among the significant associations grouped per variable category using $p = 0.05$, features contributing to initial severity are drawn from Clinical (31), MRI Manual Axial (28), MRI Labels Sag (28), MRI Automatic Axial (28), MRI Manual Sag (24), and MRI Automatic Sag (24), and Demographics (6). For final severity, they are drawn from Clinical (28), MRI Manual Sag (26), MRI Automatic Sag (27), MRI Manual Axial (29), MRI Automatic Axial (28), and MRI Labels Sag (28). Regarding change in severity, there are from Clinical (44), MRI Labels Sag (17), MRI Automatic Sag (10), MRI Automatic Axial (8), and Demographics (4), and MRI Manual-derived biomarkers (7). While MRI-based features (manual and automatic) are often associated with initial and final severity, they are not likely associated with change from baseline, in contrast to clinical variables.

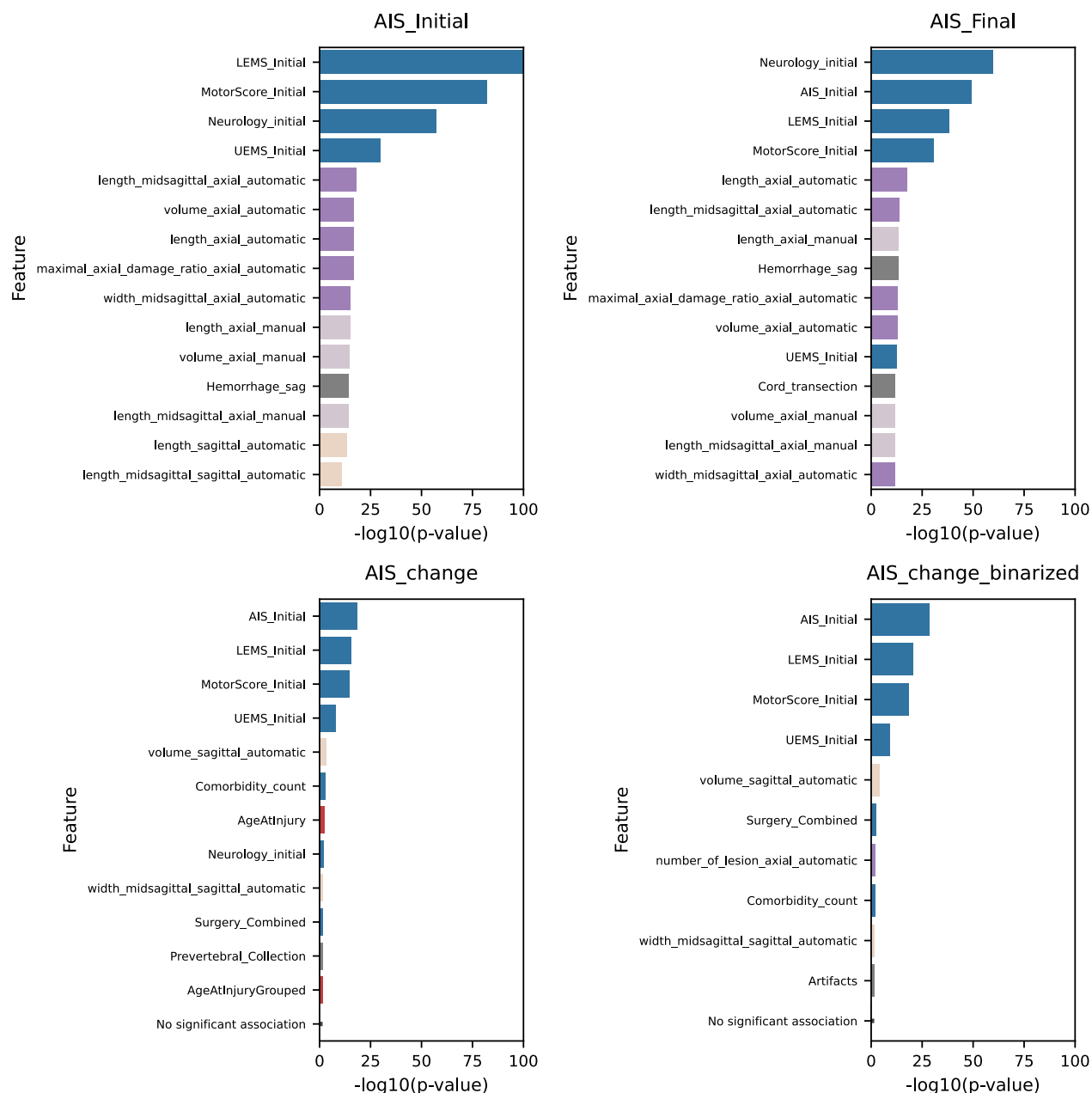


Figure 4.14 Top 15 Most Significant Associations for AIS Grade

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score. Initial corresponds to the initial evaluation timepoint (acute, at admission). Final corresponds to the final evaluation timepoint (at rehabilitation discharge). Change corresponds to the change in score from baseline. **See Table 4.2 for color-coding legend.**

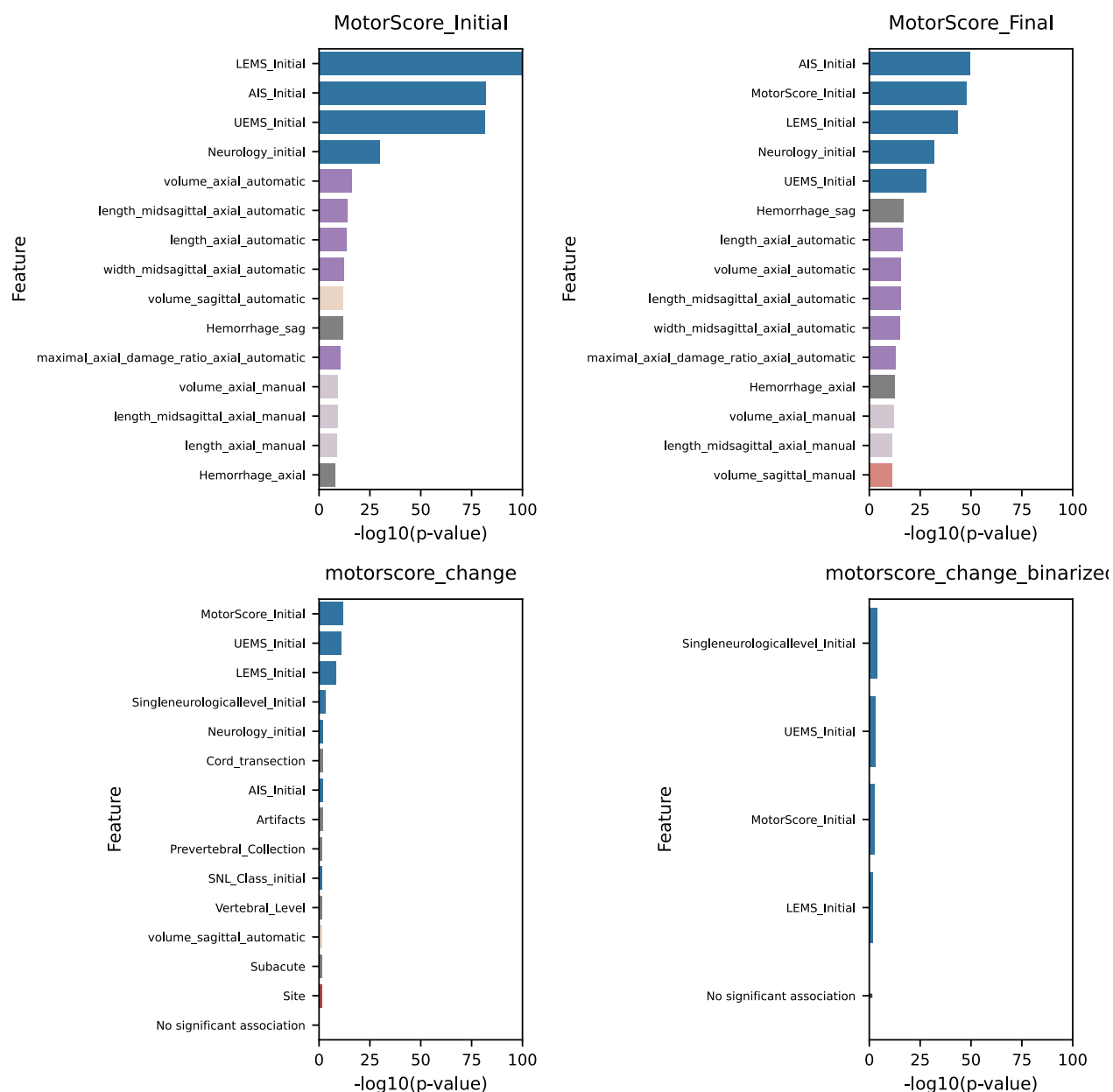


Figure 4.15 Top 15 Most Significant Associations for Total Motor Score

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score. Initial corresponds to the initial evaluation timepoint (acute, at admission). Final corresponds to the final evaluation timepoint (at rehabilitation discharge). Change corresponds to the change in score from baseline. **See Table 4.2 for color-coding legend.**

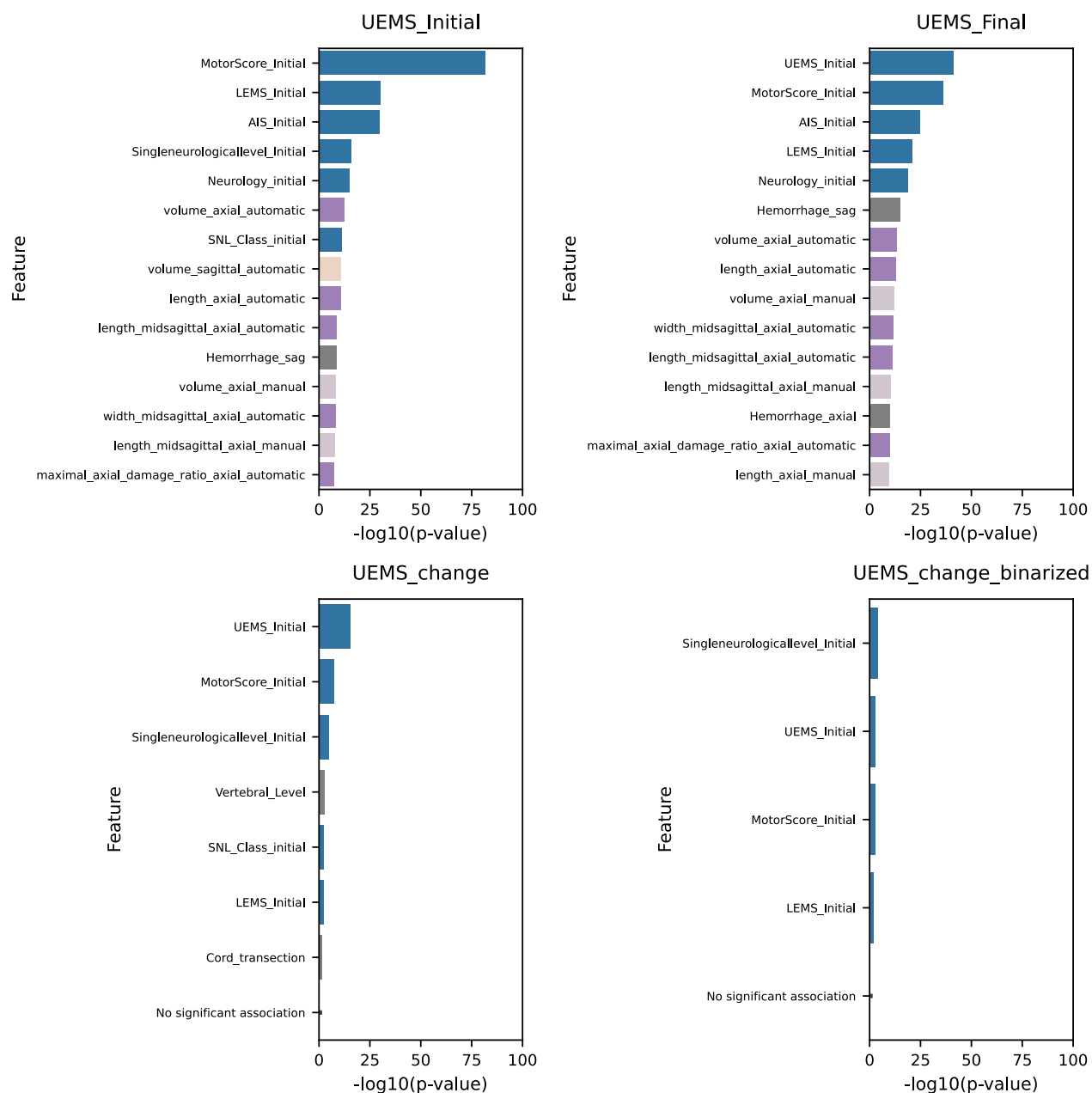


Figure 4.16 Top 15 Most Significant Associations for Upper Extremity Motor Score

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score. Initial corresponds to the initial evaluation timepoint (acute, at admission). Final corresponds to the final evaluation timepoint (at rehabilitation discharge). Change corresponds to the change in score from baseline. **See Table 4.2 for color-coding legend.**

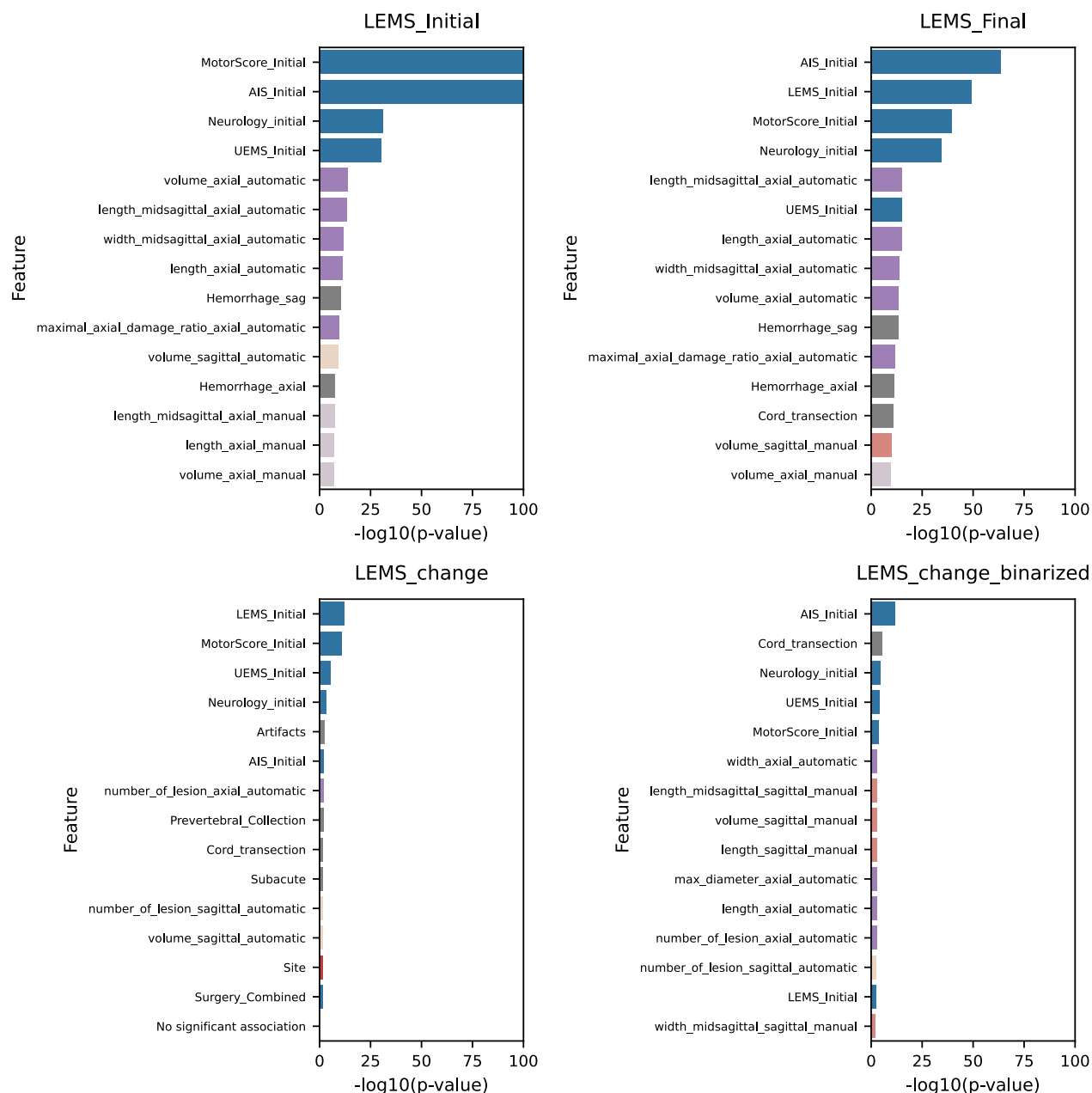


Figure 4.17 Top 15 Most Significant Associations for Lower Extremity Motor Score

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = *Upper Extremity Motor Score*; LEMS = *Lower Extremity Motor Score*. Initial corresponds to the initial evaluation timepoint (acute, at admission). Final corresponds to the final evaluation timepoint (at rehabilitation discharge). Change corresponds to the change in score from baseline. **See Table 4.2 for color-coding legend.**

4.3.3 Model Performance Evaluation

Data from a total of 272 participants from 7 sites have been included for predictive modeling. MRI metrics from automatic segmentations were available for 219 sagittal and 222 axial T2-weighted MRI. MRI metrics from manual segmentations were available for 237 sagittal and 243 axial T2-weighted MRI. Among the participants, 190 had metrics from both sagittal and axial automatic segmentations, while 210 had both from sagittal and axial manual segmentations. Modeling performance is presented through the following figures ([H3.3](#)):

- Figure 4.18, which shows the predictive results using the full dataset for modeling;
- Figure 4.19, which displays model feature coefficients from full data modeling, using variable category color-coding as defined in Table 4.2;
- Figure 4.20, which summarizes the average model performance across all prediction and classification models;
- Figures 4.21, 4.22, 4.23, and 4.24, which present the best-performing models for:
 - *American Spinal Injury Association Impairment Scale* (AIS) grade;
 - Total motor score;
 - Upper extremity motor score (UEMS);
 - Lower extremity motor score (LEMS);

For each outcome, predictions are made for the score values at rehabilitation discharge as well as their changes from baseline. In addition, classification models evaluate whether participants improved or worsened in these scores (≤ 0 vs. > 0).

- Figure 4.25 and Table 4.3, which assess the added value of MRI biomarkers in predicting changes in these scores (≤ 0 vs. > 0) beyond clinical examination, for the full cohort as well as stratified by initial *American Spinal Injury Association Impairment Scale* (AIS) grade: A or B, and C or D.

Interpretation is provided in the Discussion (see Chapter 5).

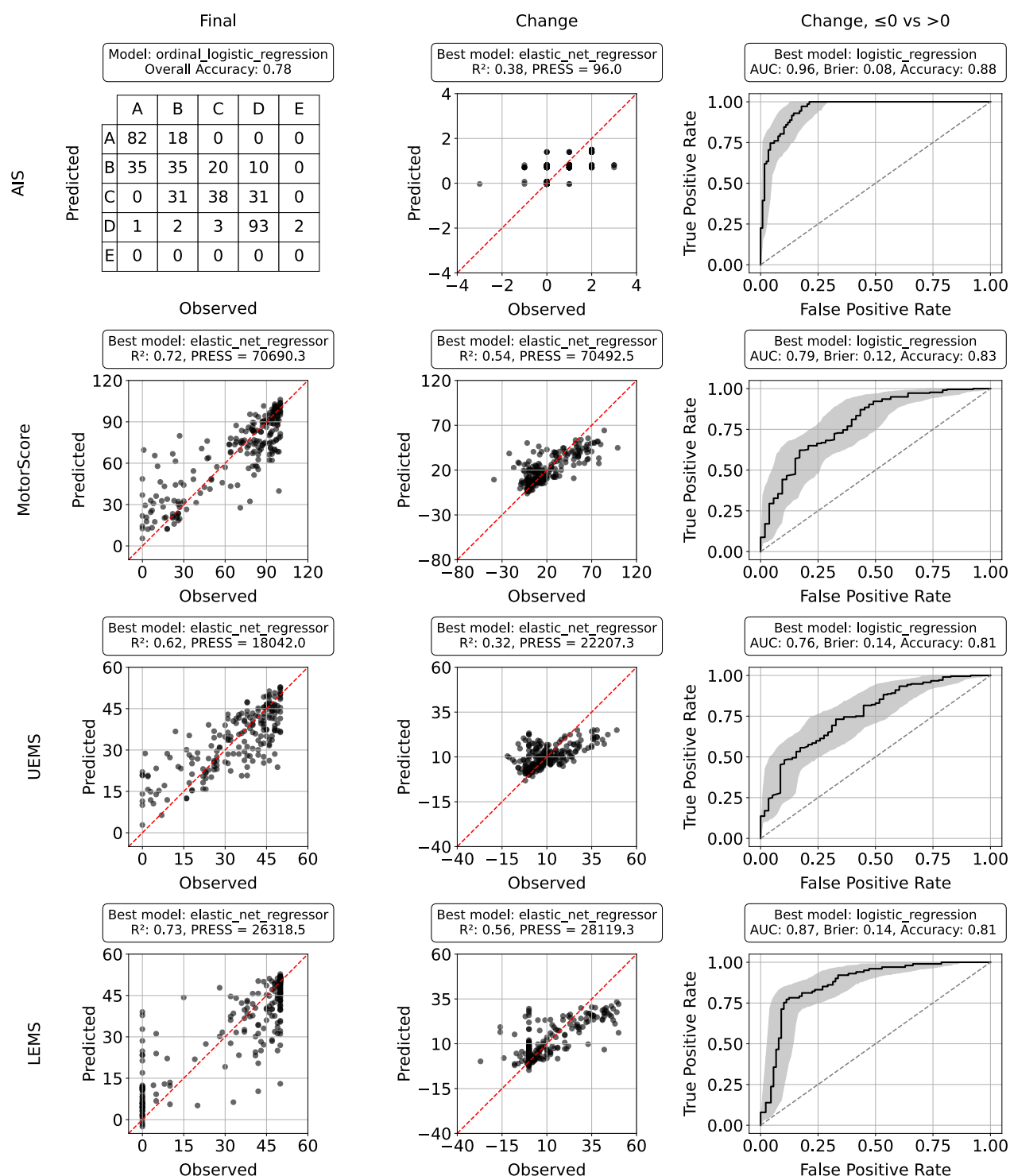


Figure 4.18 Outcomes Baseline Modeling Results Using Full Available Data at Final Discharge, Change from Baseline, and Change ≤ 0 Versus > 0 in Scores

AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score; AUC = Area Under the Receiver Operating Characteristic Curve; R^2 = Coefficient of Determination; PRESS = Predicted Residual Error Sum of Squares; Brier = Brier Score. The final AIS grade classification matrix shows the percentage (%) of correct predictions relative to the observed classes. Red dashed line indicates perfect prediction; black dashed line in AUC plots shows random performance and shaded areas mark 95% bootstrap confidence intervals. Results reflect the best model across LASSO, linear and elastic net regressor for continuous outcomes, logistic regression and LASSO for dichotomized outcomes, and ordinal logistic regression for final AIS grade. Predicted versus Observed plots.

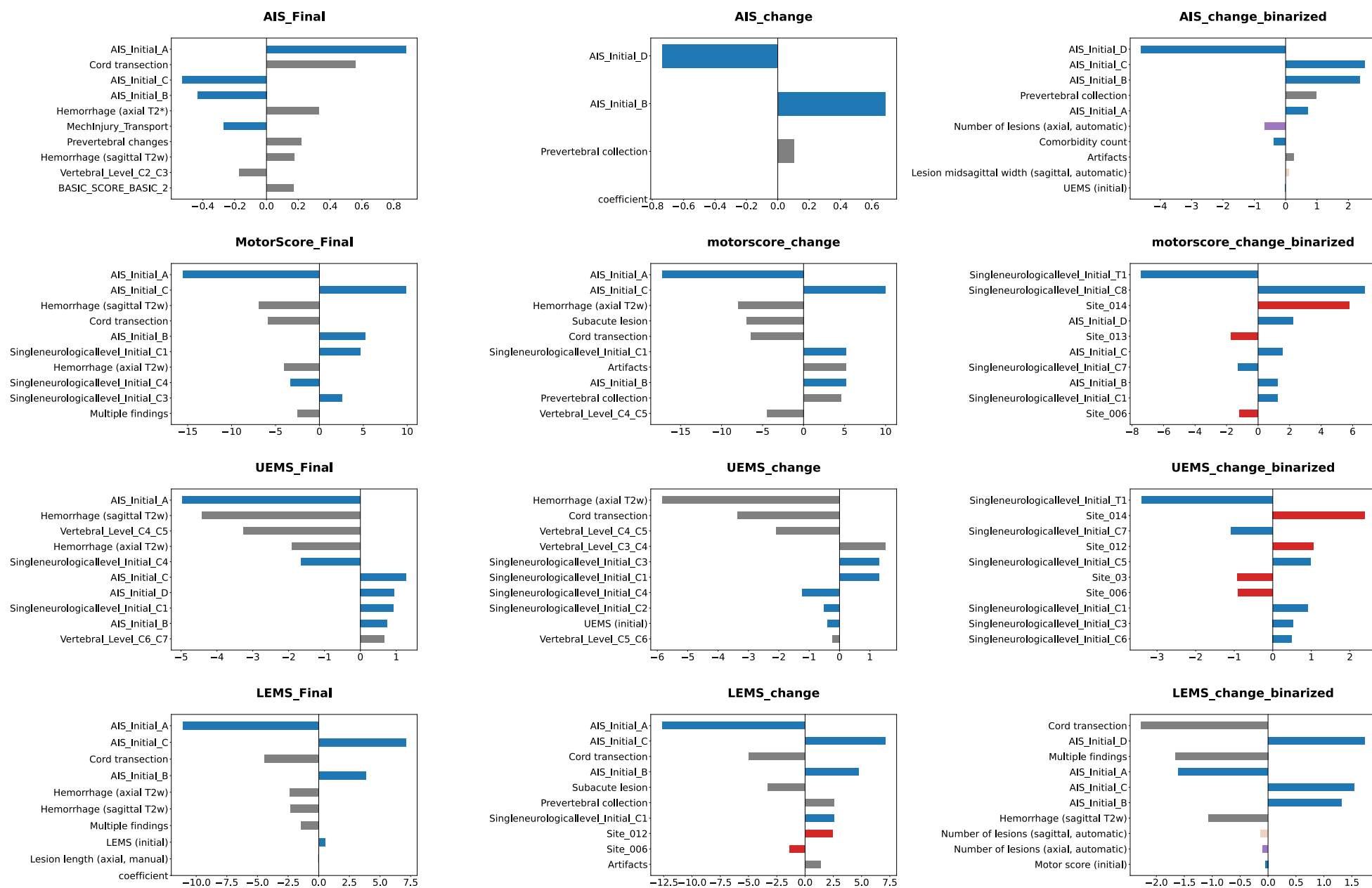


Figure 4.19 Top 10 Feature Coefficients Magnitude in Predictive Modeling Using Full Available Data for Outcomes Prediction at Final Discharge, Change from Baseline, and Change (≤ 0 versus > 0) in Scores

AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score.

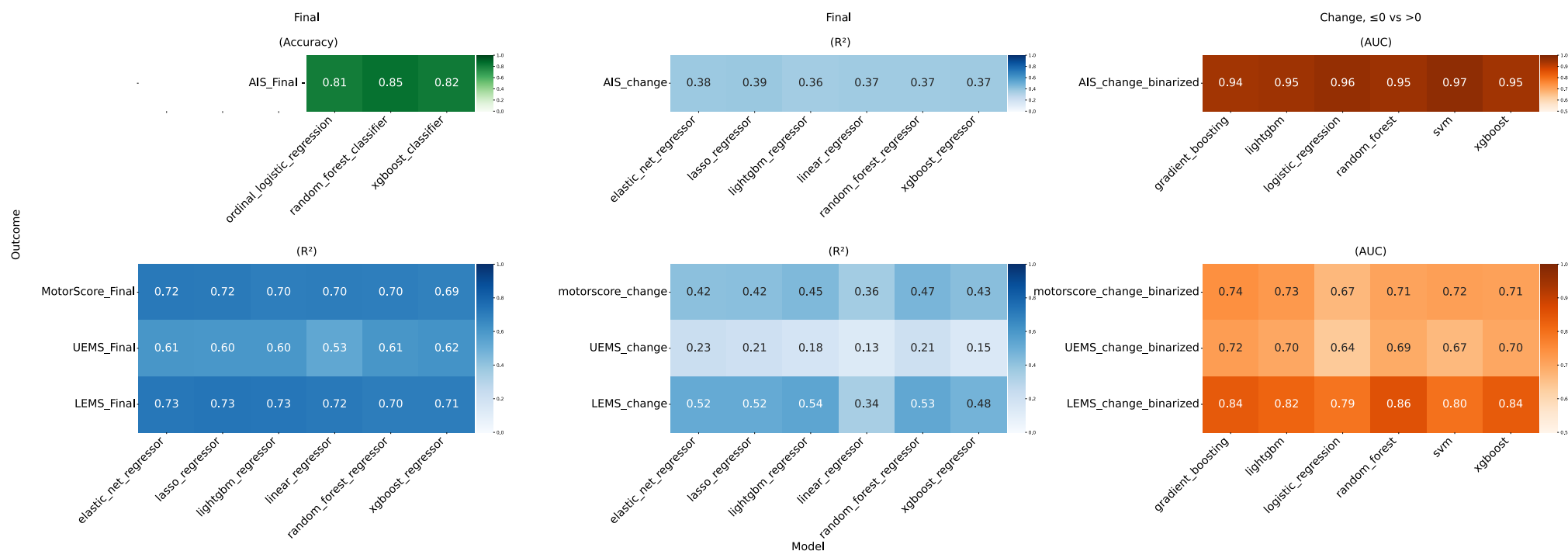


Figure 4.20 Model Performance for Outcomes Prediction and Classification at Final Discharge, Change from Baseline, and Change (≤ 0 Versus > 0)

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score; Accuracy = Overall proportion of correct classification across all classes; R^2 = Coefficient of Determination; AUC = Area Under the Receiver Operating Characteristic Curve. Final corresponds to the rehabilitation discharge score value. Change corresponds to the change in such score from baseline evaluation. In each heatmap, the x axis refers to the model, and the y axis refers to the outcome.

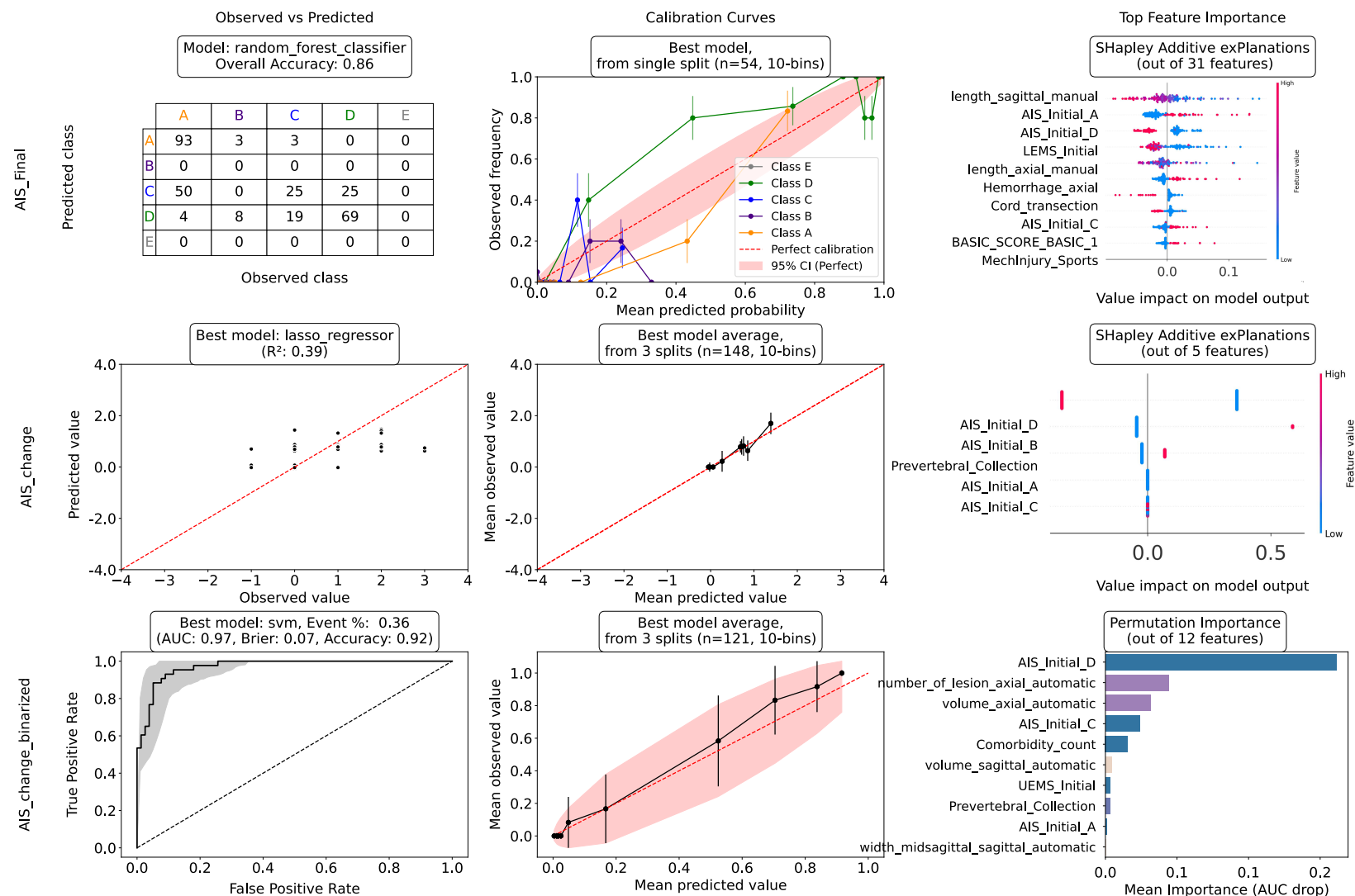


Figure 4.21 Assessment of *American Spinal Injury Association Impairment Scale* Grade Prediction and Classification at Final Discharge, Change from Baseline, and Change from Baseline ≤ 0 Versus > 0

AIS = *American Spinal Injury Association Impairment Scale*. AIS grades are modeled as categorical ordinal values ranging from E (normal) to A (most severe), considering non-linear intervals between severity categories. Matrix values correspond to the percentage (%) of correctly predicted classes. Red dashed line indicate perfect prediction; black dashed line in receiver operating characteristic curves show random performance. Receiver operating characteristic curves shaded areas mark 95% bootstrap confidence intervals. Calibration curves show 95% confidence intervals as error bars (values) and shaded red zone (perfect line). Results are from the best model across $3 \times 80:20$ random splits using 5-fold cross-validation.

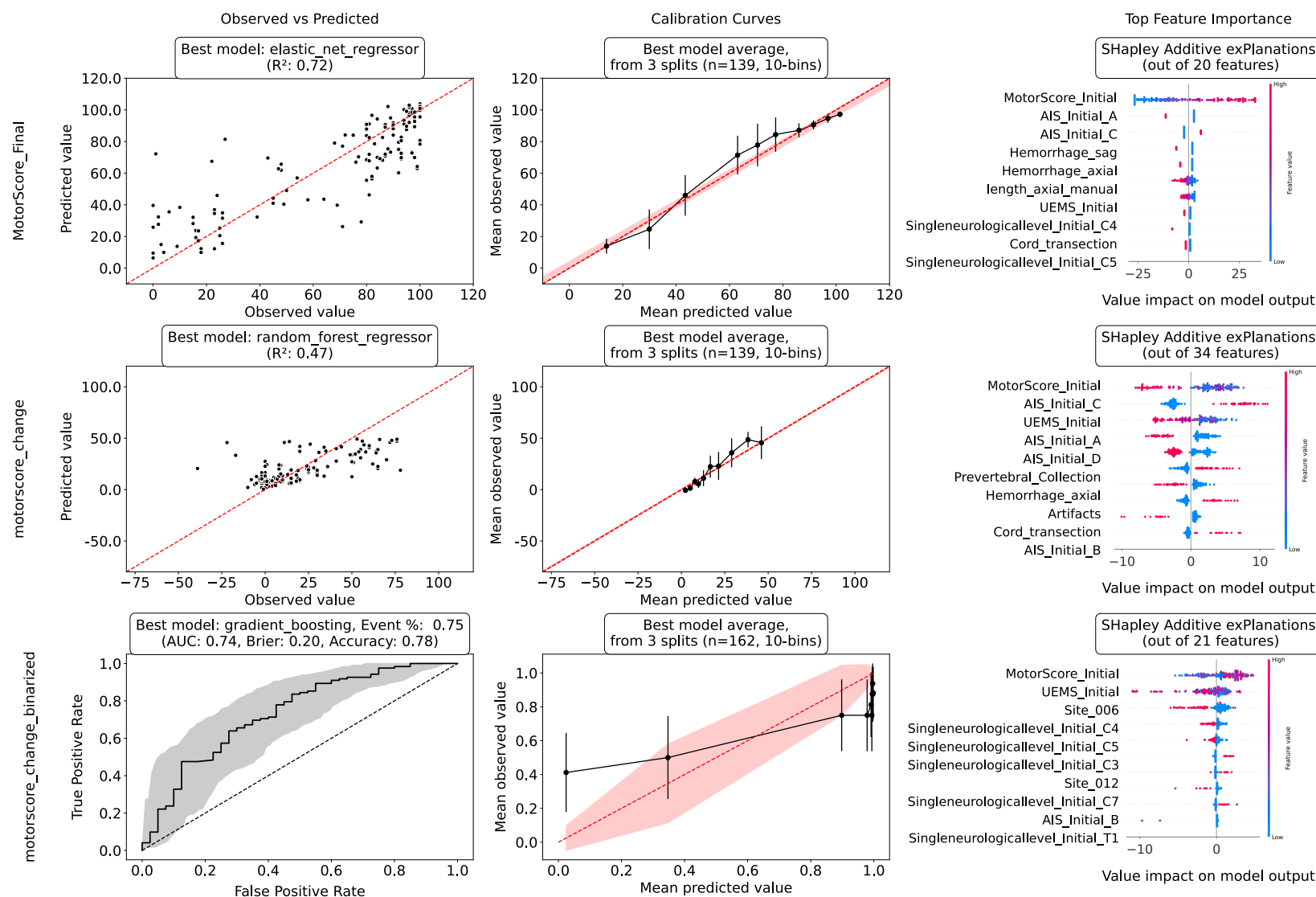


Figure 4.22 Assessment of Total Motor Score Prediction at Final Discharge, and Change from Baseline

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = *Upper Extremity Motor Score*; LEMS = *Lower Extremity Motor Score*. Red dashed line indicate perfect prediction; black dashed line in receiver operating characteristic curves show random performance. Receiver operating characteristic curves shaded areas mark 95% bootstrap confidence intervals. Calibration curves show 95% confidence intervals as error bars (values) and shaded red zone (perfect line). Results are from the best model across $3 \times 80:20$ random splits using 5-fold cross-validation.

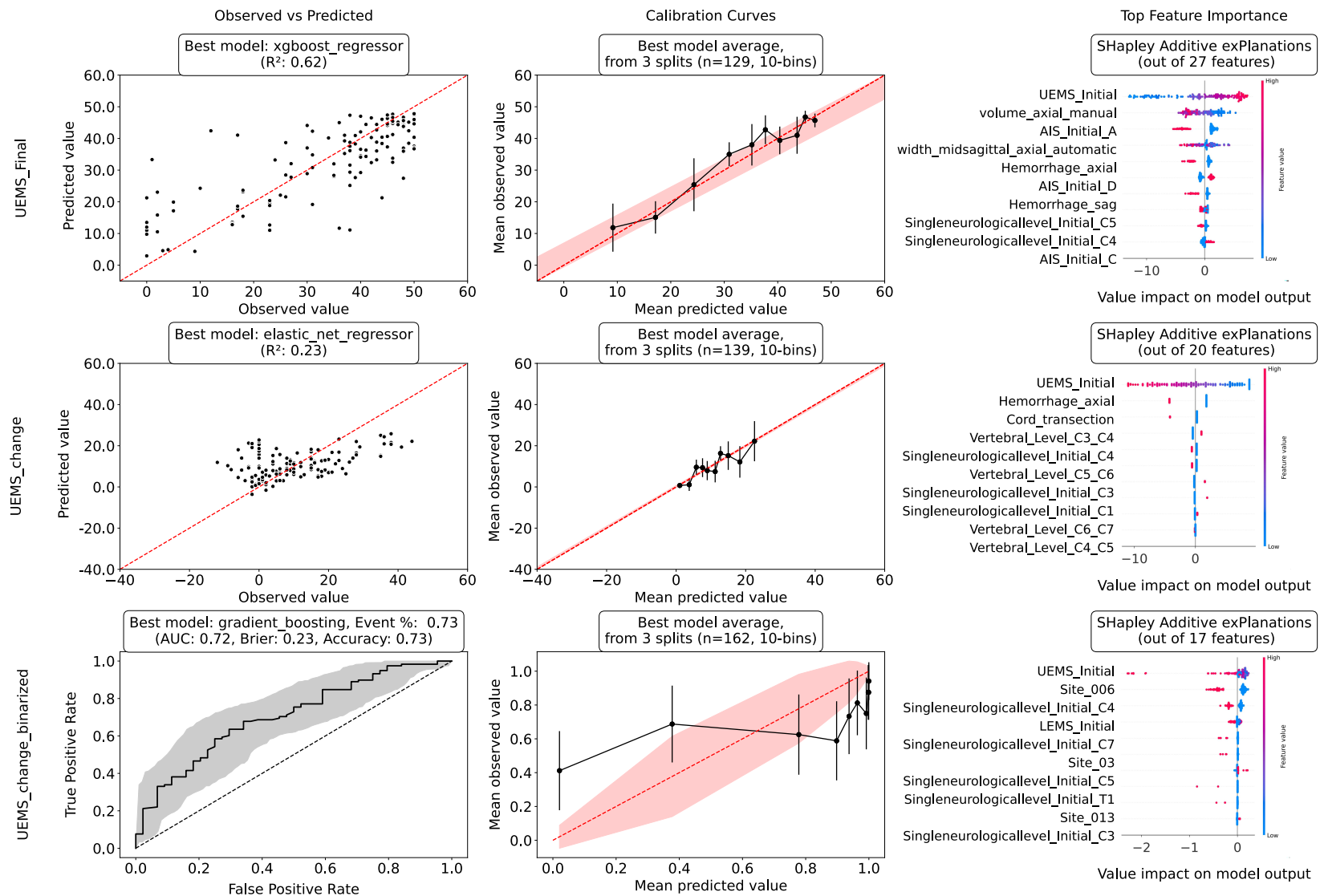


Figure 4.23 Assessment of Upper Extremity Motor Score Prediction at Final Discharge, and Change from Baseline

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = *Upper Extremity Motor Score*; LEMS = *Lower Extremity Motor Score*. Red dashed line indicate perfect prediction; black dashed line in receiver operating characteristic curves show random performance. Receiver operating characteristic curves shaded areas mark 95% bootstrap confidence intervals. Calibration curves show 95% confidence intervals as error bars (values) and shaded red zone (perfect line). Results are from the best model across $3 \times 80:20$ random splits using 5-fold cross-validation.

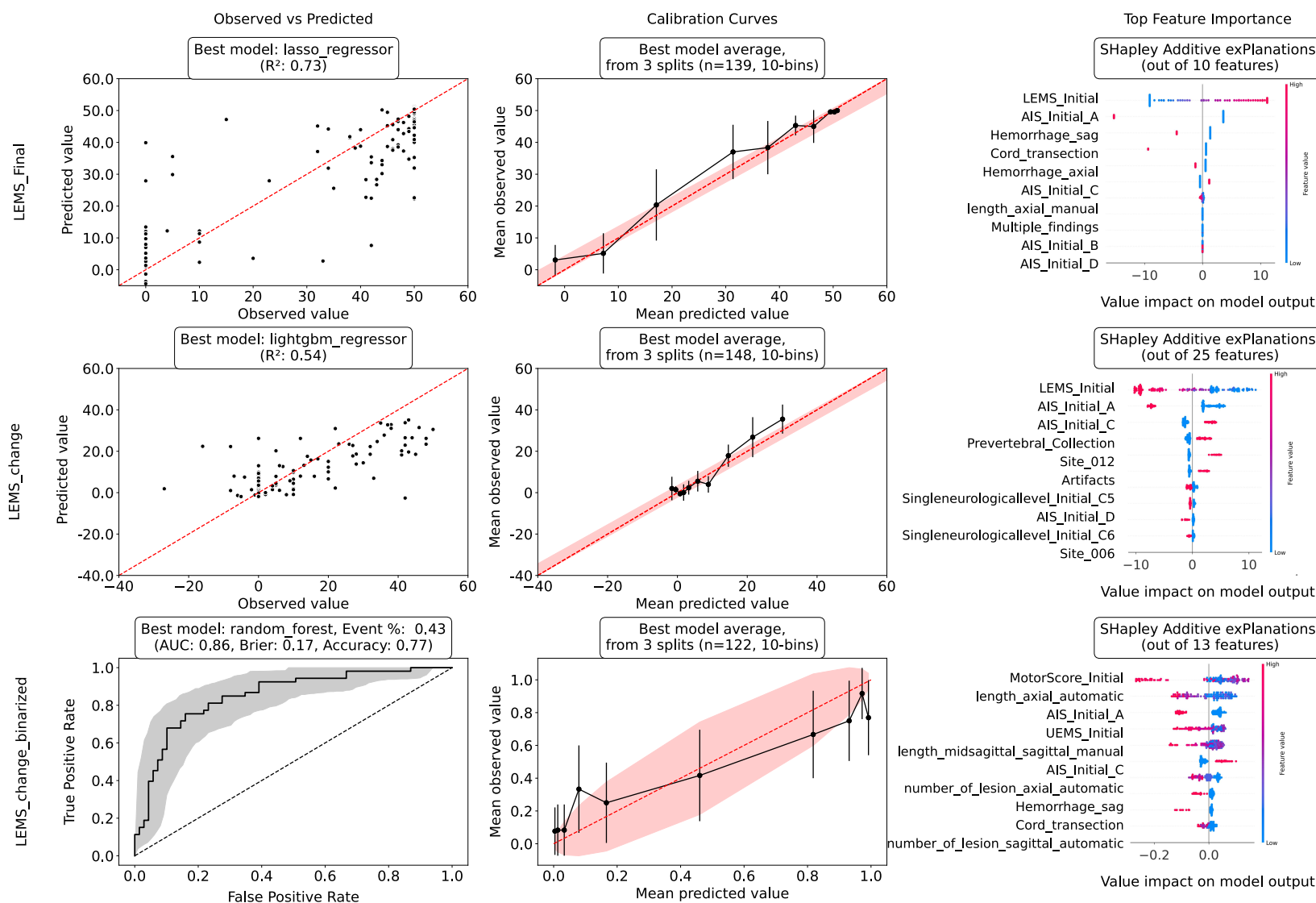


Figure 4.24 Assessment of Lower Extremity Motor Score Prediction at Final Discharge, and Change from Baseline

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = *Upper Extremity Motor Score*; LEMS = *Lower Extremity Motor Score*. Red dashed line indicate perfect prediction; black dashed line in receiver operating characteristic curves show random performance. Receiver operating characteristic curves shaded areas mark 95% bootstrap confidence intervals. Calibration curves show 95% confidence intervals as error bars (values) and shaded red zone (perfect line). Results are from the best model across $3 \times 80:20$ random splits using 5-fold cross-validation.

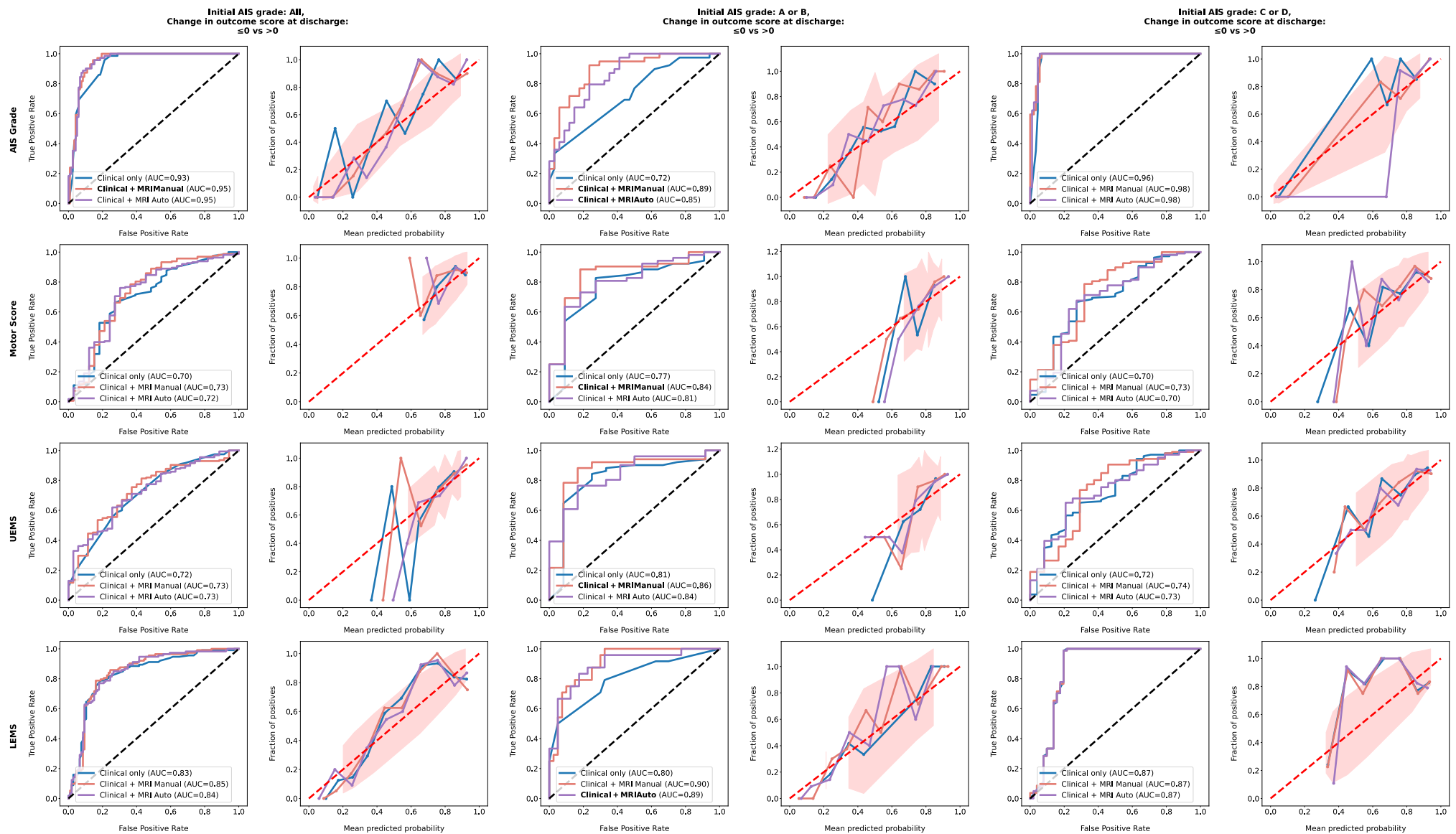


Figure 4.25 The Role of Magnetic Resonance Imaging Biomarkers in Acute Preoperative Cervical Traumatic Spinal Cord Injury

AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score; AUC = Area Under the Receiver Operating Characteristic Curve. Receiver operating characteristic curves are shown for the best combination (based on AUC) of two clinical features at initial admission predicting each outcome at rehabilitation discharge, along with the best combination of two imaging biomarkers derived from manual and automatic lesion segmentation on preoperative MRI. **Combinations in bold** reflect a statistically significant AUC difference ($p \leq 0.05$) over clinical features alone, based on the DeLong test. Calibration curves display 95% confidence intervals as a shaded red zone around the perfect calibration line. Calibration curve colors correspond to the ROC curve legend.

| Outcome | Initial AIS grade: All | Initial AIS grade: A or B | Initial AIS grade: C or D |
|--|--|---|---|
| American Spinal Injury Association Impairment Scale Grade | n = 189 Clinical: Comorbidity_count, AIS_Initial MRI_Manual: length_axial_manual, length_midsagittal_sagittal_manual MRI_Automatic: length_axial_automatic, length_sagittal_automatic | n = 73 Clinical: Comorbidity_count, Neurology_initial MRI_Manual: length_axial_manual, width_axial_manual MRI_Automatic: width_axial_automatic, length_midsagittal_axial_automatic | n = 126 Clinical: Comorbidity_count, AIS_Initial MRI_Manual: length_sagittal_manual, width_midsagittal_sagittal_manual MRI_Automatic: maximal_axial_damage_ratio_axial_automatic, length_sagittal_automatic |
| Total Motor Score | n = 196 Clinical: Neurology_initial, MotorScore_Initial MRI_Manual: width_midsagittal_axial_manual, number_of_lesion_sagittal_manual MRI_Automatic: number_of_lesion_axial_automatic, maximal_axial_damage_ratio_axial_automatic | n = 63 Clinical: Comorbidity_count, Singleneurologicallevel_Initial MRI_Manual: width_midsagittal_axial_manual, number_of_lesion_sagittal_manual MRI_Automatic: max_diameter_axial_automatic, number_of_lesion_sagittal_automatic | n = 130 Clinical: Singleneurologicallevel_Initial, MotorScore_Initial MRI_Manual: volume_axial_manual, width_sagittal_manual MRI_Automatic: width_sagittal_automatic, max_diameter_sagittal_automatic |
| Upper Extremity Motor Score | n = 190 Clinical: Singleneurologicallevel_Initial, Neurology_initial MRI_Manual: width_axial_manual, width_midsagittal_axial_manual MRI_Automatic: width_axial_automatic, width_midsagittal_sagittal_automatic | n = 63 Clinical: Comorbidity_count, Singleneurologicallevel_Initial MRI_Manual: width_midsagittal_axial_manual, number_of_lesion_sagittal_manual MRI_Automatic: width_midsagittal_axial_automatic, volume_sagittal_automatic | n = 130 Clinical: Singleneurologicallevel_Initial, MotorScore_Initial MRI_Manual: volume_axial_manual, width_sagittal_manual MRI_Automatic: width_sagittal_automatic, max_diameter_sagittal_automatic |
| Lower Extremity Motor Score | n = 219 Clinical: Neurology_initial, MotorScore_Initial MRI_Manual: volume_sagittal_manual, number_of_lesion_sagittal_manual MRI_Automatic: volume_sagittal_automatic, length_sagittal_automatic | n = 64 Clinical: AIS_Initial, Singleneurologicallevel_Initial MRI_Manual: length_sagittal_manual, number_of_lesion_sagittal_manual MRI_Automatic: length_axial_automatic, number_of_lesion_sagittal_automatic | n = 132 Clinical: Neurology_initial, LEMS_Initial MRI_Manual: number_of_lesion_sagittal_manual, width_midsagittal_sagittal_manual MRI_Automatic: length_axial_automatic, width_axial_automatic |

Table 4.3 Summary of Sample Size and Selected Predictive Features by Outcome and Subpopulation Based on *American Spinal Injury Association Impairment Scale* Grade at Initial Admission

Each row represents a specific outcome, while columns correspond to subgroups defined by baseline AIS grade. Within each cell, the sample size (n) and the selected features per category (Clinical, MRI Manual, MRI Automatic) are reported. A maximum of two features are selected per category due to the limited number of change events in the smallest outcome subgroup.

CHAPTER 5 DISCUSSION

Each stage of this work presents distinct strengths, limitations, and future perspectives. The following discussion is guided by the project’s objectives, with thematic cross-referencing where appropriate. It begins with SlicerCART, followed by insights into MRI biomarkers variability, and concludes with interpretation of the predictive modeling results. This chapter aims to provide a cohesive understanding of the contributions and implications of this thesis.

5.1 SlicerCART

As a reminder, high quality annotations are important for imaging analysis model development. Often, they require manual input, which is a highly time-consuming task. Currently, no efficient open-source solution (see Appendix B) has been found relevant to standardize and streamline the annotation process. In that context, SlicerCART was built to facilitate manual annotation tasks, and as a proof-of-concept, annotation times were recorded. Manual segmentation of tSCI lesions using SlicerCART took on average 1 minute and 12 seconds for sagittal MRI and 1 minute and 50 seconds for axial MRI—both well below the hypothesized 5-minute threshold per participant (H1). This threshold was defined based on the author’s initial experiments during the manual segmentation process. While these times cannot be directly compared, they may serve as a future reference for similar projects involving tSCI lesion segmentation. It is important to take these times into account in future studies to assess whether annotation time can be reduced—particularly given that the current standard operating procedure, annotation platform, and workflow were developed by the author, who was also actively involved in project execution. In that context, annotation times may differ for external users performing the task independently. Lesion delineation is also likely influenced by annotator experience; in this study, annotations are performed by a radiology resident, and the resulting segmentations may differ from those produced by non-medical users (due to an unknown and user-dependent learning curve, difficult to estimate without explicit testing) or by board-certified radiologists. Furthermore, although published data on manual segmentation of tSCI lesions remain limited, it is reasonable to hypothesize that reproducibility would range from moderate to low, due to frequent high inter-rater variability observed across various pathologies in manual segmentation. From a subjective, manual imaging analysis perspective, tSCI lesions are medium-sized and generally unique. In contrast, multiple sclerosis lesions (in the brain or spinal cord) are typically small or very small (per lesion) and multiples, whereas tumors (in the brain or spinal cord) are

likely larger (single or multiple). Thus, segmentation times should not be directly compared across pathologies. However, future projects in related fields—where literature of manual lesion segmentation times may be more reported—could measure annotation time and use it to quantitatively assess the time-efficiency of SlicerCART or other annotation tool while collecting other confounding variables such as the annotator’s experience and level of training. It is also worth noting that annotation time was likely influenced by the chosen manual segmentation approach, which involved delineating only the lesion contours rather than drawing the full lesion mask. This likely contributed to the reduced annotation time in this project. The remaining part of the lesions were automatically filled using Python scripts. While such a feature is currently available in 3D Slicer through tools like Lasso and Scissors, it has not yet been implemented in SlicerCART. Classification tasks required roughly 45 seconds per participant; however, the annotator had prior experience manipulating the same dataset during the segmentation sessions, which were completed before the classification tasks. This familiarity, combined with qualitative assessments, informed the author’s guidance for exploratory MRI label definitions that contributed to the classification tasks. As a result, an observer bias [210] may have been introduced, potentially reducing task completion time. However, classification times for hemorrhage on axial T2* imaging were nearly twice as long when repeated three months later—contradicting partially this assumption. Although the working conditions remained consistent, this longer labeling time may be attributed to the author’s attention during the second session, informed by prior experience.

5.1.1 Scope of realizations

Beyond these metrics, it is arguably more important to remind and define the intended scope and use cases for SlicerCART. As highlighted in the literature review (see Appendix B), many tools exist, but most are either inefficient or too specialized and task-specific. It is therefore crucial to consider what characteristics lead to this situation and how software/-modules/applications should be developed moving forward. The author briefly presented SlicerCART to the 3D Slicer core development team (as of May 20th, 2025—the day of this submission, coincidentally!), where many of whom have been contributors to the modules assessed in Appendix B. The discussion highlighted the need for a module like SlicerCART, but also recognized the fragmentation across multiple parallel software efforts aimed at targeting similar task domains that often end up focusing on a specific project. While the idea of developing core functionalities that can benefit the community is commendable, the time and resource investment required in the development and long-term maintenance of such tools were brought up as potential barriers. Additionally, questions were raised about the transportability of SlicerCART—specifically, whether it is limited to neuroimaging. In

fact, SlicerCART has been designed with adaptability in mind, making it suitable for diverse imaging settings (tSCI in MRI is one example, but it can be applied to any radiological or non-radiological field using NIfTI or NRRD format in imaging analysis, similar to its parent module initially focused on intracerebral hemorrhage in computed tomography).

So, how was SlicerCART designed to be generic? The reason lies more fundamentally in the author’s belief that any imaging analysis study should be grounded in four basic components:

- **Image visualization:** Can the user clearly view the images he intends to analyze?
- **Navigation:** Can the user move effortlessly through all images to be annotated?
- **Annotation:** Does the user have the ability to access the full spectrum of manual and semi-automated annotation tools for analyses, quality control, and other purposes?
- **Revision:** Can the user efficiently modify or revise previous annotations while tracking version changes?

These characteristics (not exhaustive, however) are likely key components to most project workflows and should define the scope of SlicerCART’s usage. Many of the existing tools cited (within 3D Slicer) attempt to address these aspects, but not in a complementary or unified manner. For example, the SlicerCaseIterator module [211] is designed to navigate forward through cases but lacks centralized functionalities for image annotation or labeling, and does not allow for custom configurations (e.g. label definitions) directly from the GUI. Similarly, mpReview is tailored for multiparametric MRI—but would it function on CT scans?

In this context, an even generic, customizable module within 3D Slicer (or other platforms) is essential—and this is precisely how SlicerCART’s functionalities were targeted during its development. SlicerCART introduces several innovative features. It supports project-specific configuration (e.g. for segmentation and classification labels), and the ability to resume from previous tasks (enabling to change project intermitantly): such capabilities are rarely unified in existing solutions. The configuration is accessible via the graphical user interface (GUI), making it usable for both technical and non-technical users. Such features foster collaboration, reusability, and broader applicability. Custom working lists (e.g. based on contrast type) and user-defined keyboard shortcuts further enhance flexibility, accommodating individual user or project-specific preferences or requirements. In brief, the module’s main advantage lies in its capacity to adapt to diverse workflows and study requirements.

5.1.2 Future Developments

To further improve SlicerCART, several features are targeted for development. SlicerCART currently lacks effective support for simultaneous multi-contrast image management—an option available in ITK-Snap [212]. While ITK-Snap offers multi-contrast navigation, it lacks, in return, a dedicated case list for optimized navigation, so the trade-off between time efficiency and cross-referencing between multiple contrasts must be weighed for each project. SlicerCART currently does not support DICOM images but the functionality is available within 3D Slicer and could be implemented. Another limitation is that painted pixels in Slicer sometimes appear slightly offset from the user-intended location, possibly due to internal rendering algorithms. This may lead to over-segmentation (as in current manual tSCI lesion segmentations) and warrants further investigation; potentially, a correction function could address this limitation. Future directions include: 1) to enhance utilization robustness across different users; 2) deploying the module internally at the NeuroPoly Lab and affiliated research facilities at the University of Montreal to evaluate its feature in real annotation environment, using standardized assessment (e.g. the NASA task load index [213]); and 3) disseminating it more broadly within the research community. At that stage, module success will be evaluated based on adoption and usage—though this remains a future goal.

5.2 MRI Biomarkers

5.2.1 Axial Versus Sagittal

From the manual segmentation masks generated using SlicerCART, MRI biomarkers are extracted. Notably, the computed values differ depending on whether they are derived from sagittal or axial masks, which contradicts theoretical expectations—values should be equivalent across views. This discrepancy is evident in Figure 4.8, and it also extends to automatic segmentation, where measurements likewise vary by view. With the exception of lesion volume ($\rho = 0.63$), secondary hypothesis (H2.1)—which postulates good agreement (correlation coefficient $\rho \geq 0.60$) between sagittal and axial measurements—is rejected for manual segmentations. In contrast, the hypothesis is supported when using automatic segmentations, where most quantitative metrics meet or exceed the 0.60 threshold, with the exception of lesion width, which shows poor correlation ($\rho = 0.33$). These results suggest that automatic measurements help overcome the reproducibility limitations inherent in manual segmentation masks MRI biomarker extraction; however, they also exhibit intrinsic variability, which is theoretically unexpected due to its standardized, model-driven nature.

5.2.2 Automatic Versus Manual

Another important observation is the discrepancy in MRI metric values between manual and automatic approaches. For example, there is a substantial difference in the scale (highlighted in blue in Figure 4.8) of lesion volume values: the median lesion volume for manual sagittal segmentations is $\sim 1800 \text{ mm}^3$, compared to $\sim 1300 \text{ mm}^3$ for axial; in contrast, automatic segmentations report lower median volumes of $\sim 600 \text{ mm}^3$ and $\sim 400 \text{ mm}^3$, respectively. Additionally, for each manual metric, the median values derived from manual segmentation masks are consistently higher than those obtained from automatic segmentation masks. While manual segmentation sagittal masks consistently yield higher or similar median values than axial masks across all metrics, this pattern is not observed with automatic measurements.

It is relevant to note that SCISeg [197], used for automatic segmentation, is optimized for intramedullary hyperintense T2 lesions and may under-segment hemorrhagic (hypointense) areas, explaining the lower automatic volume measurements. Such an example is shown in Figure 5.1. Conversely, manual segmentation may overestimate lesion size due to limitations in precise voxel alignment during annotation in Slicer, as stated above. The inclusion of a few imaging substudy subjects in the SCISeg training set is negligible, as their manual masks were created in a non-controlled setting by various users for few subjects. Although calculating Dice coefficients between manual and automatic segmentations for all subjects could estimate overlap reproducibility between the two segmentation methods, the primary objective of this study is rather to evaluate the impact of quantitative metrics derived from the segmentation masks in predictive modeling (e.g. feature importance, then compare “manual” versus “automatic” metrics, see [Objective 3](#)). In addition, Dice coefficient may be over-penalized in smaller volumes [199], which may relate to tSCI lesion depending on the segmentation.

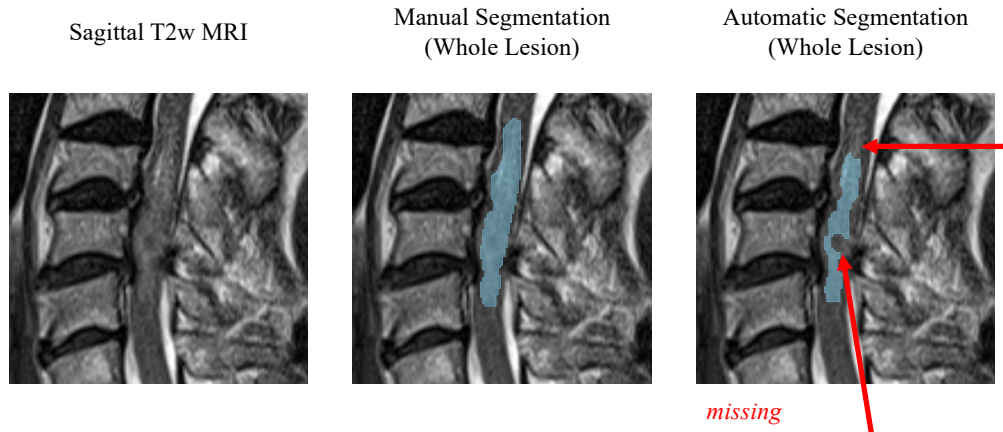


Figure 5.1 Comparison of Manual and Automatic Segmentation on Sagittal T2-weighted MRI
Volume units in mm^3 , else in mm.

5.2.3 Intra-Rater Variability

On the other hand, manual segmentation reproducibility has been evaluated on a subset of participants to assess for intra-rater variability ([H2.2](#), $n = 127$) due to its usual reference role. The mean Dice coefficient for the lesion as a whole (single segmentation class) is 0.69 (95% CI: 0.66–0.71), which is within the hypothesis acceptable range of 0.70. However, large discrepancies are observed between the mean Dice for edema (0.68) and hemorrhage (0.37). This may be due to the fact that the hemorrhage volume appeared much lower in comparison to edema (making Dice unstable) and/or suggest that hemorrhage manual segmentation on T2-weighted imaging is unreliable. Consequently, the targeted development of automatic intramedullary hemorrhage segmentation tools may be of limited utility unless other modalities such as T2-gradient sequences with better depiction of hemorrhage are used. Combining edema and hemorrhage into a single lesion mask may enhance segmentation consistency, particularly in regions initially labeled as edema but later reclassified as hemorrhage (or vice versa). This likely contributes to the slightly higher Dice scores observed for the overall lesion. It is to note that the whole lesion mean volume (see [Figure 4.9](#)) differs between initial (2339 mm³) versus at 6-months (2526 mm³). The fact that both average volumes are higher than the final mean volumes (1805 mm³) after post-processing (constraining to the spinal cord masks) suggests that these manual segmentation masks were over-segmenting the lesion (potentially also attributed to the Slicer delineation issue, mentioned above).

Classification of hemorrhage on sagittal T2-weighted imaging has a good kappa (Altman [214]) of 0.66 (95% CI: 0.49–0.84), which confirms secondary hypothesis [H2.3](#). Classification agreement for axial T2* images is good to very good (Altman [214]) with a kappa of 0.77 (95% CI: 0.61–0.90), supporting its clinical relevance in hemorrhage identification on tSCI MRI. Nevertheless, these results must be interpreted with caution and have intrinsic limitations. For instance, if a lesion was correctly classified at the initial time point but misclassified later, the results would not be representative. Furthermore, as the annotator was a radiology resident, variability in comparison to a board-certified radiologist remains unaddressed. Repeating the same tasks under consistent conditions (e.g. at 6- and 3-month intervals) by a radiologist would provide a more comprehensive picture of both intra- and inter-rater variability. It would also help isolate intra-rater variability related to experience—e.g, a radiology resident (few years of training) versus a radiologist (e.g. ≥ 10 years of experience).

5.2.4 Intramedullary Hemorrhage and Outcomes

In addition, univariate analyses (Appendix G) reveal that hemorrhage classification from sagittal T2-weighted imaging is significantly associated with three out of four outcomes—i.e., AIS grade, total motor score, upper and lower extremity motor score-, excluding the initial upper extremity motor score. This applies at initial and final time points. Classification from axial T2-weighted imaging shows significant associations ($p < 0.001$) with all initial and final outcomes severity, as do T2* axial classification and BASIC scores. These findings support the reliability of axial-based classification for hemorrhage detection and are consistent with the literature, as the BASIC score is a known MRI-based severity classification system and T2* signal loss corresponds to blooming artifacts from hemorrhagic components. Sensitivity and specificity for hemorrhage detection using sagittal T2w versus axial T2* imaging are 0.56 and 0.86 initially, and 0.55 and 0.98 on repeated evaluation. Given the more consistent axial T2-weighted imaging hemorrhage significant association with outcomes, sensitivity and specificity could be calculated from non-empty axial segmentation hemorrhage masks using axial T2* imaging. It is also important to note that none of these classification labels show a significant association with changes in outcome severity, even after dichotomization. Effect sizes warrant consideration for a more comprehensive understanding (we remind here that those univariate analyses serve as an initial feature selection step in prediction modeling).

5.2.5 Future Analyses

Ultimately, other imaging analysis methods—such as lesion cropping followed by higher-order radiomics extraction, as applied by Okimatsu et al. [181] for short-term prognostication or by Shimizu et al. [215] in postoperative tSCI imaging—could be transported to preoperative imaging analysis to further improve predictive performances.

5.3 Predictive Modeling

5.3.1 Cohort Profile: Selection Bias?

Prior to incorporation of such metrics in prediction models, another aspect is to understand the cohort characteristics, which define the applicability of prediction modeling findings. Among RHSCIR imaging substudy participants, key characteristics include the mechanism of injury and initial severity. Falls (55%) is the leading mechanism of injury, followed by transport (21%) and sports (18%). 49% have initial AIS grade D, 14% AIS grade A, and lower proportions for grades B and C. Therefore, the relevance of imaging biomarkers in this

study primarily applies to patients with AIS D that likely experienced a fall. The average age is 56 years, with 76% male and 32% aged between 61 and 75, which is comparable to the wider RHSCIR data [216]. On imaging, it can be hypothesized that these individuals likely exhibit chronic degenerative changes of the spine, which may influence rehabilitation outcomes or stratified recovery potential. In this cohort, 64% have no AIS grade conversion at rehabilitation discharge. In the context of differentiating traumatic versus non-traumatic spinal cord injuries (see Section 2.1 in Chapter 2), it is also relevant to consider whether the fall is sufficient to classify the individual as having sustained a tSCI, as opposed to an injury more related to degenerative cervical myelopathy [217], therefore categorized as non-tSCI.

By contrast, the *United States Spinal Cord Injury Statistical Center's 2023* report [218] (thousands of participants) mentions that 43% of acute tSCI patients present with AIS grade A at admission (versus 19% with AIS D), highlighting population differences and expected recovery trajectories with the actual imaging substudy cohort. In the United States, 32% of injuries are caused by falls and 37% by transport-related incidents. Their reported average age is 44 years old [218]. The sex distribution, however, is comparable (78% male) to the present study (76%). These differences suggest that findings from this imaging substudy may not be directly generalizable to all tSCI populations. Sensitivity analyses on subgroups with different socio-demographics and clinical characteristics would strengthen the findings. Moreover, a subgroup analysis (H3.1) from the Praxis Institute team supports that individuals undergoing preoperative MRI tend to be older and less severely injured, reflecting a selection bias. This highlights the need for population-based studies to support the generalization of predictive modeling and to guide its application in relevant clinical scenarios. Additionally, it is important to note that the subgroup analysis conducted by the Praxis Institute only included eligible participants (i.e., those with available clinical data as an inclusion criterion), and not all individuals from the *Rick Hansen Spinal Cord Injury Registry* (RHSCIR). This distinction is important, as the broader RHSCIR population may also differ.

Another important consideration is the clinical context in which MRI is performed. As discussed in the literature review (see “Role of Magnetic Resonance Imaging” in Subsection 2.1.4), specific indications exist for MRI assessment in the acute tSCI setting. These indications may partly explain the observed population characteristics between the subgroup that undergoes MRI and the subgroup that does not. From a clinical utility perspective, MRI can assist in evaluating the spinal cord and the extent of compression. In cases of severe injury, patients may go directly to surgery, whereas less straightforward cases may undergo MRI as part of a diagnostic workup to guide surgical management. Such contextual factors are essential to consider when interpreting potential differences between imaging substudy cohorts and those in broader, unrestricted tSCI research studies.

5.3.2 Associations are not Predictions

Univariate analyses reveal that clinical variables consistently show the strongest associations across different outcomes—at initial evaluation, rehabilitation discharge (final), change from baseline, and binarized outcomes—while MRI-derived metrics, whether from manual or automated segmentations, are primarily associated with initial and final severity, but not with changes from baseline (almost all). Hypothesis [H3.2](#), which postulates that imaging biomarkers are associated with initial, final, and change scores, is then partially supported: associations with initial and final severity are observed, but not with changes in severity from baseline. From the exploratory manual classification labels, with the exception of spinal cord transection—which tends to be additionally associated with change in severity from baseline—only a few labels show consistent associations with initial severity, final severity, or change from baseline. This suggests that MRI classification labels could be either restricted or expanded based also on diagnostic considerations, such as the identification of ligamentous injuries (and consequently assess the prognostic value of those diagnostic characteristics). While a trend suggests that most significant associations with outcomes come from clinical data, Top-15 Figures 4.14 to 4.17 should be interpreted with caution, as they are intended for descriptive purposes only. Effect sizes are not considered, and the number of variables per category differs. Nonetheless, these associations provide insight into which variables are most likely to be included in the prediction models training pipeline, given that univariate results guide the initial feature selection. It is important to recall that associations and predictions are distinct concepts and often confused [76]. Accordingly, prediction modeling has been done in two parts: 1) applying generalized linear models to full data, and 2) using machine learning methods with cross-validation and train-test splits, with performance averaging.

5.3.3 Performance Assessment

Regarding Hypothesis [H3.3](#) (achieving $R^2 \geq 0.70$, AUC or $Accuracy \geq 0.80$), using the full dataset for prediction at final discharge, the models meet the R^2 threshold for continuous outcomes in 2 out of 3 cases, and achieves an accuracy of 0.78 for the ordinal classification of AIS grade, which is in the range of the hypothesized threshold of 0.80. In contrast, no model achieves the target R^2 threshold for predicting absolute changes. Classification tasks fulfill the hypothesized threshold in 2 out of 4 tasks where models achieve excellent discrimination for AIS change and lower extremity motor score (LEMS) (AUC of 0.96 and 0.87, respectively), and acceptable to good AUC [146] for upper extremity motor score (UEMS) and total motor score (0.76 and 0.79, respectively). Accuracy is similar across all outcomes, i.e. ≥ 0.80 . Brier scores are harder to interpret and might be better suited for comparison

with external datasets. Interpreting R^2 , AUC and accuracy on the full dataset is limited by potential overfitting, calibration challenges, and the absence of clear reference standards (for R^2). In this clinical and technical context, unlike in the social sciences where interpretive thresholds exist [219], defining a “good” R^2 is difficult. Here, R^2 was computed using `scikit-learn` [220], though its interpretation remains nontrivial. Chicco et al. [221] mention a value of $R^2 = 0.756$ as very good, and $R^2 = 0.535$ as good through examples, but this project’s context differs. Here, an insightful observation is that R^2 values are good or very good for predicting final severity, but performance systematically drops for predicting changes (e.g. $R^2 = 0.72$ for final motor score versus 0.54 for its change). This suggests that predicting final severity categories is more feasible than capturing individual-level expected changes, thereby limiting potential individual stratification (as it remains a major known challenge in traumatic spinal cord injury [15]). Given that the models are trained on the full data without external validation, those performances cannot generalize, but provide insight into the potential capabilities of predictive modeling in tSCI.

Regarding feature importance, it must be acknowledged that interpretation depends on model performance: before looking at most contributing features, the question we must ask is “Does the prediction model perform well?” (although good may depend on the context). From the color-coding in Figure 4.19, the majority of top predictors come from clinical examination (blue) or qualitative manual MRI assessment (gray). MRI biomarkers (both derived from automatic and manual segmentations), appear seldomly ([H3.3.1](#)). Those graphs show that characteristics such as the AIS grade (clinical examination), the level of injury (whether from clinical neurological examination or from the vertebral level assessed on MRI), and presence of hemorrhage (MRI) appear often in the prediction models across all outcomes. This is not surprising. In binary classification, categorical nominative features like site are included, possibly due to unreliable discrimination, calibration, or selection bias through different populations or practices across sites. Does that mean MRI biomarkers are not predictors? The answer to this question depends on many considerations, including the answer to another question whether there is any added value of MRI biomarkers, beyond clinical examination. Here, it is to note that LASSO regression has been used alongside univariate analysis to reduce dimensionality, which may prevent biomarkers for being used by the models (e.g. by assigning them a coefficient of zero during feature selection).

In summary, initial statistical modeling on full data shows that good results can be achieved in predicting final severity, while predicting their changes from baseline remains systematically less accurate. MRI biomarkers derived from manual or automatic segmentations rarely emerge as top contributors, prompting the use of hierarchical regression—using baseline clinical variables—to assess the added value of quantitative MRI biomarkers (detailed in 5.3.4).

Machine Learning Models

The sample size may be suboptimal for robust, generalizable machine-learning models, but exploratory machine learning-based modeling is still valuable. As a reminder, machine learning models used 3 random 80:20 train-test splits, with 5-fold cross-validation on the training set (instead of 10-fold, due to sample size constraints). This helps provide an estimate of the models' generalization ability to unseen data. However, while cross-validation mitigates overfitting, it reduces training set size and lowers statistical power. Figure 4.20 shows that machine learning performance for final severity prediction is generally moderate to good, but prediction of change in outcomes performs poorly across all models, except for LEMS. Machine learning models achieve the $R^2 \geq 0.70$ threshold for total motor score and LEMS at final rehabilitation discharge, but none for predicting changes in scores from baseline. While classifiers of AIS change and LEMS change models show excellent discrimination ($AUC \geq 0.95$ and 0.80 , respectively), acceptable performances are achieved for motor score and UEMS. For AIS grade ordinal classification, the models respect the hypothesized accuracy threshold of 0.80). Thus, hypothesis [H3.3](#) is partially supported, warranting examination of [H3.3.1](#).

AIS grade

Ordinal classification of AIS grade at final rehabilitation discharge provides an overall accuracy of 0.86 , with 93% with final AIS grade A and 69% of AIS D being correctly classified. Nevertheless, the model tends to underclassify final AIS A (overestimation of recovery) while overclassifying the severity of final AIS (underestimation of recovery). This might be due to the distribution of AIS grade conversion (see Figure 4.13), initial and final severity, and the impact of AIS A and AIS D classification on expected recovery. In such classifier, the most contributing feature is the lesion length from manual segmentation on sagittal MRI (lesion length from axial manual segmentation also plays a contributing role), but sparsity remains high (for both). Initial AIS grade, particularly grades A and D, also play a significant role in model predictions, where AIS A tends to increase the prediction and AIS D decreases it. Additional MRI-derived classification features—such as hemorrhage, cord transection, and BASIC score—also influence the model. Prediction performance for AIS grade change remains modest ($R^2 = 0.39$), with AIS grade emerging as the dominant predictive feature. It is important to consider that in the scope of this thesis, the change in AIS grade has been considered continuous (e.g. going from A to B = $+1$ and from C to D = $+1$), but that does not likely reflect the reality since the conversion of an AIS A to B or C is clinically more relevant than a conversion from C to D or D to E. The key clinical value of such a work is trying to predict which AIS grade A or B will have improved between initial evaluation and discharge, to guide management. While predictive performance of that change is mod-

est, the methodological consideration remains an additional limitation. For that outcome, some categories receive zero coefficients, due to LASSO being the best-performing model for this outcome. This reflects the distinction between association and predictive utility of variables [76]. Support Vector Machine (SVM) model achieves excellent discrimination ($AUC = 0.97$) and good calibration for classifying participants as improving or worsening, where AIS grade plays a role (expected) alongside with automatic-derived MRI biomarkers. This is a promising and exciting result within the scope of this thesis. The fact that the model achieves high discrimination and calibration promotes its use in external dataset with likely a good performance expected, and the fact that it includes automatic-derived MRI biomarkers in its contributing features (with an overall total of 12 features) promotes a wider use. Using automatically derived features may facilitate broader clinical adoption, given the time-consuming nature of manual segmentation (*reminder: here, manual segmentations were used here as a reference ground truth against which automatic methods were evaluated*). The next step about that result is to test it on the remaining incoming participants of the RHSCIR imaging substudy, and other independent datasets.

Total Motor Score

Prediction of the continuous total motor score at final rehabilitation discharge achieves strong performance ($R^2 = 0.72$, supporting H3.3), with well-estimated calibration. The initial motor score is the most influential feature (and across all related outcomes), though its contribution shows sparsity in the model output. Additional predictors such as initial AIS grade and the presence of intramedullary hemorrhage on T2-weighted MRI (both axial and sagittal) contribute, but to a lesser extent. In contrast, performance declines when predicting the change in motor score from baseline, dropping to $R^2 = 0.47$. Classification models for improvement versus worsening show acceptable discrimination, but poor calibration. This limits their potential for external application—further constrained by the appearance of sites among the top contributing features, which hinders generalizability. While MRI data (but not quantitative biomarkers) contribute in the prediction of the final total motor score, baseline performance thresholds are not met for predicting score changes, highlighting the need for alternative modeling strategies or methodological considerations in future studies.

UEMS

Moderate performance ($R^2 = 0.62$) with well-estimated calibration is achieved for predicting UEMS at final rehabilitation discharge. However, performance drops for predicting changes in UEMS ($R^2 = 0.23$), and while the classification of change achieves acceptable discrimination ($AUC = 0.72$), calibration is poor. Across all related outcomes, initial UEMS consistently emerges as the top contributing feature, though sparsity remains high in both final and

change prediction models. For the final UEMS prediction, a diverse range of features contributes significantly—such as initial AIS grade, lesion volume from manual segmentation, and presence of intramedullary hemorrhage—supporting the potential of multimodal modeling for this outcome. As with total motor score, baseline predictive thresholds are not met for UEMS change, highlighting the need for alternative strategies in future studies.

LEMS

Prediction of LEMS at final rehabilitation discharge achieves strong performance ($R^2 = 0.72$, supporting H3.3), with well-estimated calibration. Initial LEMS is a key driver of model output, along with initial AIS grade A, presence of hemorrhage on both sagittal and axial T2-weighted MRI, and cord transection—identified as non-null contributors. SHapley Additive exPlanations summary reveals that features such as AIS grades B, C, and D, alongside with lesion length from axial manual segmentation have zero coefficients, indicating no contribution. Performance moderately declines for predicting the change in LEMS ($R^2 = 0.54$), and clinical features remain the main contributors. Binary classification of participants who improve versus worsen shows good discrimination ($AUC = 0.86$) with moderate calibration, warranting validation on external datasets. Interestingly, in that model, initial total motor score—rather than LEMS—is the top contributing feature, and several MRI-derived biomarkers (e.g., lesion length) also influence predictions. These findings are encouraging within the scope of this thesis and support the contribution MRI metrics in prediction modeling.

Summary

Overall, machine learning-based prediction results are mixed, with both promising and negative findings. The most successful models are related to AIS grade change classification, followed by LEMS prediction alongside with its change classification. Strong performance is also achieved in total motor score prediction with well-estimated calibration, suggesting potential applicability in external datasets. While clinical variables are often in the top predictive features across models and outcomes, the frequent presence of qualitative and quantitative MRI biomarkers supports their relevance in acute preoperative tSCI. To further assess their added value, exploratory analyses evaluating the incremental benefit of MRI features over baseline clinical exams are discussed in subsection 5.3.4. Unlike traditional models trained on full datasets, these machine learning models show similar or superior performance to classical statistical approaches with theoretical better approximates of real-world application, thanks to their evaluation using cross-validations and independent dataset splits. Future modeling efforts would benefit from larger datasets and the inclusion of additional features such as maximum cord compression, canal compromise, radiomics (e.g., via PyRadiomics [222]), and biochemical markers (e.g., blood tests). Finally, while timing differences between clinical

exams, MRI, and surgery may influence model utility, in this study, time to surgery was not significantly associated with final outcomes or their changes, thus excluded during feature selection. Those timing could be assessed in further work using different considerations. Once predictive performances reach acceptable levels and robustness, decision curve analysis could guide the establishment of thresholds, towards real-world implementation.

5.3.4 The Role of Imaging Biomarkers

Figure 4.25 and Table 4.3 evaluate the added value of MRI biomarkers in predicting binary changes in outcome scores (≤ 0 vs. > 0) using classical logistic regression, beyond clinical examination alone. The analysis focuses on participants with initial AIS grades A and B, as these represent the most severely injured subgroups and where outcome prediction is most clinically relevant. However, stratification leads to a reduction in statistical power, primarily due to the current cohort's distribution—with half have an initial AIS grade D. Among participants with AIS A or B ($n \approx 85$), 39 do not improve by discharge, and 46 show conversion, forming a balanced but modest subgroup that still supports exploratory analyses. The same approach is concomitantly used on the full cohort and on AIS C and D subgroup, for comparison. To ensure model parsimony, a maximum of two features per category (Clinical, MRI Manual, MRI Automatic) are included. Notably, results from these logistic regressions exhibit patterns consistent with earlier models. Logistic models show good calibration when predicting AIS grade and LEMS in the full dataset and particularly in the AIS A-B subgroup, but not in the C-D subgroup. In AIS A-B participants, combining MRI biomarkers with clinical features significantly improved AUC for predicting AIS grade change ($AUC = 0.89$ for Clinical+Manual MRI and 0.85 for Clinical+Automatic MRI) compared to clinical features alone ($AUC = 0.72$). Similarly, for LEMS change, AUCs increased to 0.90 and 0.89 (Clinical+Manual and +Automatic MRI, respectively), versus 0.80 for clinical-only models. While a statistically significant improvement is also observed in the full cohort using manual MRI features, the baseline AUC of 0.93 questions the clinical relevance of this gain—given that incremental improvements beyond $AUC \geq 0.90$ are rarely meaningful [146, 223]. In terms of feature selection, lesion length and width are the most frequently retained MRI predictors—more than lesion volume—highlighting their relevance in tSCI. These findings underscore the need to: 1) increase the number of AIS A and B participants to enable machine learning directly on this subgroup with the above mentioned benefits, and 2) validate these results on external datasets: here, the objective is prediction rather than explanation [76], thus reducing emphasis on collinearity diagnostics. Despite current limitations, these exploratory results highlight the potential utility of quantitative MRI biomarkers in acute tSCI and justify further investigation in this field ([H3.3.1](#)).

CHAPTER 6 CONCLUSION

The primary objective of this work is to enhance the prediction of neurological outcomes at rehabilitation discharge in acute tSCI, using both clinical and preoperative MRI data. Targeted outcomes include AIS grade, total motor score, upper extremity motor score (UEMS), and lower extremity motor score (LEMS). In addition to prediction at discharge, a key goal is to assess changes in these scores from the acute baseline examination—an aspect that could support patient stratification for clinical trials aimed at the crucial period following the acute injury that may have a disproportionate impact on final prognosis. Although imaging biomarkers have been widely studied in tSCI, their predictive value remains debated. The author notes the frequent conflation of association studies—predominantly done in imaging biomarkers related research—with true predictive modeling, which is comparatively rare and often lacks calibration analysis. While association studies may show strong correlations, they do not necessarily translate into reliable predictions. In contrast, this work adopts prediction-based approaches, which produced mixed yet informative results:

1. **Models show promising results for AIS and LEMS change classification, and good results for and LEMS and total motor score prediction.**

Excellent discriminatory capabilities with good calibration are achieved through machine-learning for classifying participants that improve or worsen their AIS grade and LEMS at discharge relative to baseline. Strong predictive performances are obtained for final motor score and LEMS, with well-estimated calibration. Similar trends are observed using classical statistical methods, supporting the potential for complete external validation. However, across all outcomes, performance consistently drops when predicting score changes from baseline. Some models provide also modest performance or uncalibrated classification. While clinical features often appear as the top predictors, MRI data whether qualitative or quantitative also appear in the top-performing machine-learning models (though less so in statistical only ones), suggesting a role of imaging biomarkers in acute tSCI outcomes prediction and classification.

2. **Subgroup analyses suggest that MRI biomarkers are most valuable in patients most severely injured, i.e. those with AIS grade A or B at admission.**

When stratifying the cohort based on initial AIS grade (A–B vs. C–D), models incorporating MRI biomarkers show improved discriminatory capabilities—measured by area under the ROC curve—compared to models based solely on clinical features. This indicates a potential added value of MRI biomarkers for outcome prediction and clas-

sification in acute preoperative tSCI. However, the limited sample size of participants with AIS grade A or B restricts statistical power. Larger, dedicated cohorts focusing on this subgroup are necessary to validate these findings and enable the development of robust machine learning models. Importantly, predictive modeling in this subgroup is clinically relevant due to the expected recovery and functional resulting disabilities.

3. The role of imaging biomarkers is often misunderstood.

This work supports prior evidence of significant associations between MRI metrics and clinical outcomes. Clinical variables consistently show stronger associations and top contributions across prediction models. Nevertheless, MRI biomarkers frequently emerge as key contributors in top-performing models. While their individual effects may be modest, their combined inclusion can enhance the overall performance of prediction models. This highlights the role of imaging biomarkers as complementary tools rather than standalone predictors. Whether these biomarkers are derived from manual or automatic segmentation influences their clinical usefulness.

4. Manual segmentation of tSCI lesions exhibits moderate variability; automatic methods are more consistent but may underestimate lesions.

Our analysis of manual segmentations reveal moderate inter-observer variability in lesion delineation, which affects the reproducibility of derived quantitative MRI features. In contrast, automatic segmentations tend to be more consistent across views, but also tend to underestimate lesion extent. These findings underscore the importance of developing and validating reliable segmentation protocols for tSCI lesion analysis.

5. We developed and introduced SlicerCART: a new 3D Slicer module for dataset navigation, segmentation, classification and quality-control tasks.

A key contribution of this thesis is the development of SlicerCART-*Slicer Configurable Annotation and Review Tool*-, a new 3D Slicer module tailored to support manual segmentation and classification in imaging studies. This tool bridges clinical and research workflows by providing an efficient interface for annotation and quality control. Future plans include internal and external deployment to the broader research community.

In summary, this work offers outcome-oriented insights into the role of imaging biomarkers in tSCI, emphasizing the importance of distinguishing between predictive and associative studies. It highlights both the potential and the current limitations of MRI in predicting rehabilitation outcomes and introduces a practical tool that may benefit other imaging analysis studies, beyond tSCI. The integration of prediction modeling, critical evaluation of biomarkers, and tool development serves to inform ongoing research and clinical applications.

REFERENCES

- [1] C. S. Ahuja, J. R. Wilson, S. Nori, M. R. N. Kotter, C. Druschel, A. Curt, and M. G. Fehlings, “Traumatic spinal cord injury,” *Nat. Rev. Dis. Primers*, vol. 3, no. 1, p. 17018, Apr. 2017.
- [2] J. Bennett, J. M. Das, and P. D. Emmady, “Spinal cord injuries,” in *StatPearls*. Treasure Island (FL): StatPearls Publishing, Jan. 2025.
- [3] H. Krueger, V. K. Noonan, L. M. Trenaman, P. Joshi, and C. S. Rivers, “The economic burden of traumatic spinal cord injury in canada,” *Chronic Dis. Inj. Can.*, vol. 33, no. 3, pp. 113–122, Jun. 2013.
- [4] V. Y. Ma, L. Chan, and K. J. Carruthers, “Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the united states: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain,” *Arch. Phys. Med. Rehabil.*, vol. 95, no. 5, pp. 986–995.e1, May 2014.
- [5] C. S. Rivers, N. Fallah, V. K. Noonan, D. G. Whitehurst, C. E. Schwartz, J. A. Finkelstein, B. C. Craven, K. Ethans, C. O’Connell, B. C. Truchon, C. Ho, A. G. Linassi, C. Short, E. Tsai, B. Drew, H. Ahn, M. F. Dvorak, J. Paquet, M. G. Fehlings, and L. Noreau, “Health conditions: Effect on function, health-related quality of life, and life satisfaction after traumatic spinal cord injury. a prospective observational registry cohort study,” *Arch. Phys. Med. Rehabil.*, vol. 99, no. 3, pp. 443–451, Mar. 2018.
- [6] P. Lude, P. Kennedy, M. L. Elfström, and C. S. Ballert, “Quality of life in and after spinal cord injury rehabilitation: a longitudinal multicenter study,” *Top. Spinal Cord Inj. Rehabil.*, vol. 20, no. 3, pp. 197–207, Aug. 2014.
- [7] D. Cardile, A. Calderone, R. De Luca, F. Corallo, A. Quartarone, and R. S. Calabrò, “The quality of life in patients with spinal cord injury: Assessment and rehabilitation,” *J. Clin. Med.*, vol. 13, no. 6, p. 1820, Mar. 2024.
- [8] T. Tian, S. Zhang, and M. Yang, “Recent progress and challenges in the treatment of spinal cord injury,” *Protein Cell*, vol. 14, no. 9, pp. 635–652, Sep. 2023.

- [9] J. J. van Middendorp, B. Goss, S. Urquhart, S. Atresh, R. P. Williams, and M. Schuetz, "Diagnosis and prognosis of traumatic spinal cord injury," *Global Spine J.*, vol. 1, no. 1, pp. 1–8, Dec. 2011.
- [10] "American spinal injury association," <https://asia-spinalinjury.org/>, accessed: 2025-5-10.
- [11] R. Rupp, F. Biering-Sørensen, S. P. Burns, D. E. Graves, J. Guest, L. Jones, M. S. Read, G. M. Rodriguez, C. Schuld, K. E. Tansey-Md, K. Walden, and S. Kirshblum, "International standards for neurological classification of spinal cord injury: Revised 2019," *Top. Spinal Cord Inj. Rehabil.*, vol. 27, no. 2, pp. 1–22, 2021.
- [12] M. Youseffard, M. Hosseini, A. Oraii, A. R. Vaccaro, and V. Rahimi-Movaghar, "Severity assessment of impairment in spinal cord injury; a systematic review on challenging points about international standards for neurological classification of spinal cord injury," *Frontiers in Emergency Medicine*, vol. 6, no. 2, pp. e22–e22, Feb. 2022.
- [13] S. Kirshblum, M. Schmidt Read, and R. Rupp, "Classification challenges of the 2019 revised international standards for neurological classification of spinal cord injury (IS-NCSCI)," *Spinal Cord*, vol. 60, no. 1, pp. 11–17, Jan. 2022.
- [14] P. Freund, M. Seif, N. Weiskopf, K. Friston, M. G. Fehlings, A. J. Thompson, and A. Curt, "MRI in traumatic spinal cord injury: from clinical assessment to neuroimaging biomarkers," *Lancet Neurol.*, vol. 18, no. 12, pp. 1123–1135, Dec. 2019.
- [15] D. W. Cadotte and M. G. Fehlings, "Will imaging biomarkers transform spinal cord injury trials?" *Lancet Neurol.*, vol. 12, no. 9, pp. 843–844, Sep. 2013.
- [16] D. Mahanes, S. Muehlschlegel, K. E. Wartenberg, V. Rajajee, S. A. Alexander, K. M. Busl, C. J. Creutzfeldt, G. V. Fontaine, S. E. Hocker, D. Y. Hwang, K. S. Kim, D. Madzar, S. Mainali, J. Meixensberger, P. N. Varelas, C. Weimar, T. Westermaier, and O. W. Sakowitz, "Guidelines for neuroprognostication in adults with traumatic spinal cord injury," *Neurocrit. Care*, vol. 40, no. 2, pp. 415–437, Apr. 2024.
- [17] Y. Kumar and D. Hayashi, "Role of magnetic resonance imaging in acute spinal trauma: a pictorial review," *BMC Musculoskelet. Disord.*, vol. 17, p. 310, Jul. 2016.
- [18] R. Izzo, T. Popolizio, R. F. Balzano, A. M. Pennelli, A. Simeone, and M. Muto, "Imaging of cervical spine traumas," *Eur. J. Radiol.*, vol. 117, pp. 75–88, Aug. 2019.

- [19] L. M. Shah and J. S. Ross, “Imaging of spine trauma,” *Neurosurgery*, vol. 79, no. 5, pp. 626–642, Nov. 2016.
- [20] B. M. Ellingson, N. Salamon, and L. T. Holly, “Imaging techniques in spinal cord injury,” *World Neurosurg.*, vol. 82, no. 6, pp. 1351–1358, Dec. 2014.
- [21] A. M. Tarawneh, D. D’Aquino, A. Hilis, A. Eisa, and N. A. Quraishi, “Can MRI findings predict the outcome of cervical spinal cord injury? a systematic review,” *Eur. Spine J.*, vol. 29, no. 10, pp. 2457–2464, Oct. 2020.
- [22] K. G. M. Ridia, P. Astawa, M. F. Deslivia, C. Santosa, and S. D. Savio, “A systematic review of scoring system based on magnetic resonance imaging parameters to predict outcome in cervical spinal cord injury,” *Spine Surg. Relat. Res.*, vol. 7, no. 1, pp. 1–12, Jan. 2023.
- [23] M. Dobran, D. Aiudi, V. Liverotti, M. R. Fasinella, S. Lattanzi, C. Melchiorri, A. Iacoangeli, S. Campa, and G. Polonara, “Prognostic MRI parameters in acute traumatic cervical spinal cord injury,” *Eur. Spine J.*, vol. 32, no. 5, pp. 1584–1590, May 2023.
- [24] M. Seif, C. A. Gandini Wheeler-Kingshott, J. Cohen-Adad, A. E. Flanders, and P. Freund, “Guidelines for the conduct of clinical trials in spinal cord injury: Neuroimaging biomarkers,” *Spinal Cord*, vol. 57, no. 9, pp. 717–728, Sep. 2019.
- [25] J. Martineau, J. Goulet, A. Richard-Denis, and J.-M. Mac-Thiong, “The relevance of MRI for predicting neurological recovery following cervical traumatic spinal cord injury,” *Spinal Cord*, vol. 57, no. 10, pp. 866–873, Oct. 2019.
- [26] S. Kurpad, A. R. Martin, L. A. Tetreault, D. J. Fischer, A. C. Skelly, D. Mikulis, A. Flanders, B. Aarabi, T. E. Mroz, E. C. Tsai, and M. G. Fehlings, “Impact of baseline magnetic resonance imaging on neurologic, functional, and safety outcomes in patients with acute traumatic spinal cord injury,” *Global Spine J.*, vol. 7, no. 3 Suppl, pp. 151S–174S, Sep. 2017.
- [27] H. Blumenfeld, *Neuroanatomy through Clinical Cases*, d. Edition, Ed. Sunderland, Massachusetts: Sinauer, Inc. Publishers, 2010.
- [28] L. Thau, V. Reddy, and P. Singh, “Anatomy, central nervous system,” in *StatPearls*. Treasure Island (FL): StatPearls Publishing, Jan. 2025.
- [29] “The alliance,” <https://alliancecan.ca/en/about/alliance>, accessed: 2025-4-20.

- [30] N. P. Thorogood, V. K. Noonan, X. Chen, N. Fallah, S. Humphreys, N. Dea, B. K. Kwon, and M. F. Dvorak, “Incidence and prevalence of traumatic spinal cord injury in canada using health administrative data,” *Front. Neurol.*, vol. 14, p. 1201025, Jul. 2023.
- [31] W. Ding, S. Hu, P. Wang, H. Kang, R. Peng, Y. Dong, and F. Li, “Spinal cord injury: The global incidence, prevalence, and disability from the global burden of disease study 2019,” *Spine (Phila. Pa. 1976)*, vol. 47, no. 21, pp. 1532–1540, Nov. 2022.
- [32] A. Singh, L. Tetreault, S. Kalsi-Ryan, A. Nouri, and M. G. Fehlings, “Global prevalence and incidence of traumatic spinal cord injury,” *Clin. Epidemiol.*, vol. 6, pp. 309–331, Sep. 2014.
- [33] L. Guttman, “I. organisation of spinal units. history of the national spinal injuries centre, stoke mandeville hospital, aylesbury,” *Paraplegia*, vol. 5, no. 3, pp. 115–126, Nov. 1967.
- [34] H. L. Frankel, D. O. Hancock, G. Hyslop, J. Melzak, L. S. Michaelis, G. H. Ungar, J. D. Vernon, and J. J. Walsh, “The value of postural reduction in the initial management of closed injuries of the spine with paraplegia and tetraplegia. I: Part I,” *Paraplegia*, vol. 7, no. 3, pp. 179–192, Nov. 1969.
- [35] R. M. Kretzer, “A clinical perspective and definition of spinal cord injury,” *Spine (Phila. Pa. 1976)*, vol. 41, p. S27, Apr. 2016.
- [36] V. Krishna, H. Andrews, A. Varma, J. Mintzer, M. S. Kindy, and J. Guest, “Spinal cord injury: how can we improve the classification and quantification of its severity and prognosis?” *J. Neurotrauma*, vol. 31, no. 3, pp. 215–227, Feb. 2014.
- [37] S. Kirshblum, A. Botticello, J. Benedetto, J. Donovan, R. Marino, S. Hsieh, and N. Wagaman, “A comparison of diagnostic stability of the ASIA impairment scale versus frankel classification systems for traumatic spinal cord injury,” *Arch. Phys. Med. Rehabil.*, vol. 101, no. 9, pp. 1556–1562, Sep. 2020.
- [38] J. W. Fawcett, A. Curt, J. D. Steeves, W. P. Coleman, M. H. Tuszynski, D. Lammertse, P. F. Bartlett, A. R. Blight, V. Dietz, J. Ditunno, B. H. Dobkin, L. A. Havton, P. H. Ellaway, M. G. Fehlings, A. Privat, R. Grossman, J. D. Guest, N. Kleitman, M. Nakamura, M. Gaviria, and D. Short, “Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP panel: spontaneous recovery after spinal cord

- injury and statistical power needed for therapeutic clinical trials,” *Spinal Cord*, vol. 45, no. 3, pp. 190–205, Mar. 2007.
- [39] A. S. Burns, R. J. Marino, A. E. Flanders, and H. Flett, “Clinical diagnosis and prognosis following spinal cord injury,” *Handb. Clin. Neurol.*, vol. 109, pp. 47–62, 2012.
 - [40] W. P. Coleman and F. H. Geisler, “Injury severity as primary predictor of outcome in acute spinal cord injury: retrospective results from a large multicenter clinical trial,” *Spine J.*, vol. 4, no. 4, pp. 373–378, Jul. 2004.
 - [41] J. F. Ditunno, “Outcome measures: evolution in clinical trials of neurological/functional recovery in spinal cord injury,” *Spinal Cord*, vol. 48, no. 9, pp. 674–684, Sep. 2010.
 - [42] P. J. Brown, R. J. Marino, G. J. Herbison, and J. F. Ditunno, Jr, “The 72-hour examination as a predictor of recovery in motor complete quadriplegia,” *Arch. Phys. Med. Rehabil.*, vol. 72, no. 8, pp. 546–548, Jul. 1991.
 - [43] J. J. van Middendorp, A. J. F. Hosman, A. R. T. Donders, M. H. Pouw, J. F. Ditunno, Jr, A. Curt, A. C. H. Geurts, H. Van de Meent, and E.-S. S. Grp, “A clinical prediction rule for ambulation outcomes after traumatic spinal cord injury: a longitudinal cohort study,” *Lancet*, vol. 377, no. 9770, pp. 1004–1010, Mar. 2011.
 - [44] “(new) ISNCSCI 2019 revision released,” <https://asia-spinalinjury.org/isncsci-2019-revision-released/>, May 2019, accessed: 2025-5-10.
 - [45] R. A. Keith, C. V. Granger, B. B. Hamilton, and F. S. Sherwin, “The functional independence measure: a new tool for rehabilitation,” *Adv. Clin. Rehabil.*, vol. 1, pp. 6–18, 1987.
 - [46] A. Catz, M. Itzkovich, E. Agranov, H. Ring, and A. Tamir, “SCIM–spinal cord independence measure: a new disability scale for patients with spinal cord lesions,” *Spinal Cord*, vol. 35, no. 12, pp. 850–856, Dec. 1997.
 - [47] D. Lammertse, D. Dungan, J. Dreisbach, S. Falci, A. Flanders, R. Marino, E. Schwartz, and National Institute on Disability and Rehabilitation, “Neuroimaging in traumatic spinal cord injury: an evidence-based review for clinical practice and research: Report of the national institute on disability and rehabilitation research spinal cord injury measures meeting,” *J. Spinal Cord Med.*, vol. 30, no. 3, pp. 205–214, 2007.

- [48] “ASIA 2025 - spinal cord injury data challenge!” <https://asia-spinalinjury.org/asia-2025-spinal-cord-injury-data-challenge/>, Jan. 2025, accessed: 2025-5-10.
- [49] P. Mputu Mputu, M. Beauséjour, A. Richard-Denis, and J.-M. Mac-Thiong, “Early predictors of neurological outcomes after traumatic spinal cord injury: A systematic review and proposal of a conceptual framework,” *Am. J. Phys. Med. Rehabil.*, vol. 100, no. 7, pp. 700–711, Jul. 2021.
- [50] T. Dalkilic, N. Fallah, V. K. Noonan, S. S. Elizei, K. Dong, L. Belanger, L. Ritchie, A. Tsang, E. Bourassa-Moreau, M. K. S. Heran, S. J. Paquette, T. Ailon, N. Dea, J. Street, C. G. Fisher, M. F. Dvorak, and B. K. Kwon, “Predicting injury severity and neurological recovery after acute cervical spinal cord injury: A comparison of cerebrospinal fluid and magnetic resonance imaging biomarkers,” *J. Neurotrauma*, vol. 35, no. 3, pp. 435–445, Feb. 2018.
- [51] E. Ydens, I. Palmers, S. Hendrix, and V. Somers, “The next generation of biomarker research in spinal cord injury,” *Mol. Neurobiol.*, vol. 54, no. 2, pp. 1482–1499, Mar. 2017.
- [52] B. Guan, G. Li, R. Zheng, Y. Fan, L. Yao, L. Chen, S. Feng, and H. Zhou, “A critical appraisal of clinical practice guidelines for diagnostic imaging in the spinal cord injury,” *Spine J.*, vol. 23, no. 8, pp. 1189–1198, Aug. 2023.
- [53] C. S. Ahuja, G. D. Schroeder, A. R. Vaccaro, and M. G. Fehlings, “Spinal cord injury—what are the controversies?” *J. Orthop. Trauma*, vol. 31 Suppl 4, pp. S7–S13, Sep. 2017.
- [54] D. M. Schaefer, A. E. Flanders, J. L. Osterholm, and B. E. Northrup, “Prognostic significance of magnetic resonance imaging in the acute phase of cervical spine injury,” *J. Neurosurg.*, vol. 76, no. 2, pp. 218–223, Feb. 1992.
- [55] A. Matsushita, T. Maeda, E. Mori, I. Yuge, O. Kawano, T. Ueta, and K. Shiba, “Can the acute magnetic resonance imaging features reflect neurologic prognosis in patients with cervical spinal cord injury?” *Spine J.*, vol. 17, no. 9, pp. 1319–1324, May 2017.
- [56] D. W. Cadotte, J. R. Wilson, D. Mikulis, P. W. Stroman, S. Brady, and M. G. Fehlings, “Conventional MRI as a diagnostic and prognostic tool in spinal cord injury: a systemic review of its application to date and an overview on emerging MRI methods,” *Expert Opin. Med. Diagn.*, vol. 5, no. 2, pp. 121–133, Feb. 2011.

- [57] N. S. Mahmood, R. Kadavigere, K. R. Avinash, and V. R. Rao, "Magnetic resonance imaging in acute cervical spinal cord injury: a correlative study on spinal cord changes and 1 month motor recovery," *Spinal Cord*, vol. 46, no. 12, pp. 791–797, Jun. 2008.
- [58] F. Miyanji, J. Furlan, B. Aarabi, P. Arnold, and M. Fehlings, "Acute cervical traumatic spinal cord injury: MR imaging findings correlated with neurologic outcome—prospective study with 100 consecutive patients," *Radiology*, vol. 243, no. 3, pp. 820–827, Jun. 2007.
- [59] J. F. Talbott, W. D. Whetstone, W. J. Readdy, A. R. Ferguson, J. C. Bresnahan, R. Saigal, G. W. J. Hawryluk, M. S. Beattie, M. C. Mabray, J. Z. Pan, G. T. Manley, and S. S. Dhall, "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," *J. Neurosurg. Spine*, vol. 23, no. 4, pp. 495–504, Oct. 2015.
- [60] J. J. Smith, A. G. Sorensen, and J. H. Thrall, "Biomarkers in imaging: realizing radiology's future," *Radiology*, vol. 227, no. 3, pp. 633–638, Jun. 2003.
- [61] E. O. Mensah, J. I. Chalif, B. R. Johnston, E. Chalif, T. Parker, S. Izzy, Z. He, R. Saigal, M. G. Fehlings, and Y. Lu, "Traumatic spinal cord injury: a review of the current state of art and future directions - what do we know and where are we going?" *N. Am. Spine Soc. J.*, vol. 22, no. 100601, p. 100601, Jun. 2025.
- [62] M. Seif, G. David, E. Huber, K. Vallotton, A. Curt, and P. Freund, "Cervical cord neurodegeneration in traumatic and non-traumatic spinal cord injury," *J. Neurotrauma*, vol. 37, no. 6, pp. 860–867, Mar. 2020.
- [63] M. G. Fehlings, S. C. Rao, C. H. Tator, G. Skaf, P. Arnold, E. Benzel, C. Dickman, B. Cuddy, B. Green, P. Hitchon, B. Northrup, V. Sonntag, F. Wagner, and J. Wilberger, "The optimal radiologic method for assessing spinal canal compromise and cord compression in patients with cervical spinal cord injury: Part II: Results of a multicenter study," *Spine*, vol. 24, no. 6, p. 605, Mar. 1999.
- [64] N. Mummaneni, J. F. Burke, A. M. DiGiorgio, L. H. Thomas, X. Duong-Fernandez, M. Harris, L. U. Pascual, A. R. Ferguson, J. Russell Huie, J. Z. Pan, D. D. Hemmerle, V. Singh, A. Torres-Espin, C. Omondi, N. Kyritsis, P. R. Weinstein, W. D. Whetstone, G. T. Manley, J. C. Bresnahan, M. S. Beattie, J. Cohen-Adad, S. S. Dhall, and J. F. Talbott, "Injury volume extracted from MRI predicts neurologic outcome in acute

- spinal cord injury: A prospective TRACK-SCI pilot study,” *J. Clin. Neurosci.*, vol. 82, pp. 231–236, Dec. 2020.
- [65] B. Aarabi, C. A. Sansur, D. M. Ibrahimi, J. M. Simard, D. S. Hersh, E. Le, C. Diaz, J. Massetti, and N. Akhtar-Danesh, “Intramedullary lesion length on postoperative magnetic resonance imaging is a strong predictor of ASIA impairment scale grade conversion following decompressive surgery in cervical spinal cord injury,” *Neurosurgery*, vol. 80, no. 4, pp. 610–620, Apr. 2017.
- [66] A. C. Smith, S. R. Albin, D. R. O’Dell, J. C. Berliner, D. Dungan, M. Sevigny, C. Draganich, J. M. Elliott, and K. A. Weber, “Axial MRI biomarkers of spinal cord damage to predict future walking and motor function: a retrospective study,” *Spinal Cord*, vol. 59, no. 6, pp. 693–699, Jun. 2021.
- [67] E. Huber, P. Lachappelle, R. Sutter, A. Curt, and P. Freund, “Are midsagittal tissue bridges predictive of outcome after cervical spinal cord injury?” *Ann. Neurol.*, vol. 81, no. 5, pp. 740–748, May 2017.
- [68] D. Pfyffer, A. C. Smith, K. A. Weber, 2nd, A. Grillhiesl, O. Mach, C. Draganich, J. C. Berliner, C. Tefertiller, I. Leister, D. Maier, J. M. Schwab, A. Thompson, A. Curt, and P. Freund, “Prognostic value of tissue bridges in cervical spinal cord injury: a longitudinal, multicentre, retrospective cohort study,” *Lancet Neurol.*, vol. 23, no. 8, pp. 816–825, Aug. 2024.
- [69] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. W. L. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, and R. J. Gillies, “Radiomics: the process and the challenges,” *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1234–1248, Nov. 2012.
- [70] R. J. Gillies, P. E. Kinahan, and H. Hricak, “Radiomics: Images are more than pictures, they are data,” *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016.
- [71] M. E. Mayerhoefer, A. Materka, G. Langs, I. Häggström, P. Szczypiński, P. Gibbs, and G. Cook, “Introduction to radiomics,” *J. Nucl. Med.*, vol. 61, no. 4, pp. 488–495, Apr. 2020.
- [72] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A. G. Morganti, and M. Bellomi, “Radiomics: the facts and the challenges of image analysis,” *Eur. Radiol. Exp.*, vol. 2, no. 1, p. 36, Dec. 2018.

- [73] M. R. Tomaszewski and R. J. Gillies, “The biological meaning of radiomic features,” *Radiology*, vol. 298, no. 3, pp. 505–516, Mar. 2021.
- [74] S. S. F. Yip and H. J. W. L. Aerts, “Applications and limitations of radiomics,” *Phys. Med. Biol.*, vol. 61, no. 13, pp. R150–66, Jul. 2016.
- [75] J. C. Furlan, A. Kailaya-Vasan, B. Aarabi, and M. G. Fehlings, “A novel approach to quantitatively assess posttraumatic cervical spinal canal compromise and spinal cord compression: a multicenter responsiveness study: A multicenter responsiveness study,” *Spine (Phila. Pa. 1976)*, vol. 36, no. 10, pp. 784–793, May 2011.
- [76] G. Shmueli, “To explain or to predict?” *Stat. Sci.*, vol. 25, no. 3, pp. 289–310, Aug. 2010.
- [77] T. V. Varga, K. Niss, A. C. Estampador, C. B. Collin, and P. L. Moseley, “Association is not prediction: A landscape of confused reporting in diabetes - a systematic review,” *Diabetes Res. Clin. Pract.*, vol. 170, no. 108497, p. 108497, Dec. 2020.
- [78] A. M. Tarawneh, D. D’Aquino, A. Hilis, A. Eisa, and N. A. Quraishi, “Can MRI findings predict the outcome of cervical spinal cord injury? a systematic review,” *Eur. Spine J.*, vol. 29, no. 10, pp. 2457–2464, Oct. 2020.
- [79] A. Bozzo, J. Marcoux, M. Radhakrishna, J. Pelletier, and B. Goulet, “The role of magnetic resonance imaging in the management of acute spinal cord injury,” *J. Neurotrauma*, vol. 28, no. 8, pp. 1401–1411, Aug. 2011.
- [80] M. Mathew, W. C. Mezue, M. C. Chikani, A. O. Jimoh, E. O. Uche, and M. B. Mathew, “Correlation of quantitative MRI parameters with neurological outcome in acute cervical spinal cord injury,” *J. West Afr. Coll. Surg.*, vol. 11, no. 1, pp. 5–10, Jun. 2022.
- [81] J. R. Wilson, R. G. Grossman, R. F. Frankowski, A. Kiss, A. M. Davis, A. V. Kulkarni, J. S. Harrop, B. Aarabi, A. Vaccaro, C. H. Tator, M. Dvorak, C. I. Shaffrey, S. Harkema, J. D. Guest, and M. G. Fehlings, “A clinical prediction model for long-term functional outcome after traumatic spinal cord injury based on acute clinical and imaging factors,” *J. Neurotrauma*, vol. 29, no. 13, pp. 2263–2271, Sep. 2012.
- [82] S. Sharif and M. Y. Jazaib Ali, “Outcome prediction in spinal cord injury: Myth or reality,” *World Neurosurg.*, vol. 140, pp. 574–590, Aug. 2020.

- [83] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [84] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2017, pp. 2980–2988.
- [85] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang, “Deep learning: A primer for radiologists,” *Radiographics*, vol. 37, no. 7, pp. 2113–2131, Nov. 2017.
- [86] T. Jiang, J. L. Gradus, and A. J. Rosellini, “Supervised machine learning: A brief primer,” *Behav. Ther.*, vol. 51, no. 5, pp. 675–687, Sep. 2020.
- [87] Y. Chen, M. Mancini, X. Zhu, and Z. Akata, “Semi-supervised and unsupervised deep visual learning: A survey,” Aug. 2022.
- [88] F. Renard, S. Guedria, N. D. Palma, and N. Vuillerme, “Variability and reproducibility in deep learning for medical image segmentation,” *Sci. Rep.*, vol. 10, no. 1, p. 13724, Aug. 2020.
- [89] L. Joskowicz, D. Cohen, N. Caplan, and J. Sosna, “Automatic segmentation variability estimation with segmentation priors,” *Med. Image Anal.*, vol. 50, pp. 54–64, Dec. 2018.
- [90] I. G. L. N. A. Artha Wiguna, Y. Kristian, M. F. Deslivia, R. Limantara, D. Cahyadi, I. A. Liando, H. A. Hamzah, K. Kusuman, D. Dimitri, M. Anastasia, and I. K. Suyasa, “A deep learning approach for cervical cord injury severity determination through axial and sagittal magnetic resonance imaging segmentation and classification,” *Eur. Spine J.*, vol. 33, no. 11, pp. 4204–4213, Nov. 2024.
- [91] A. P. Krishnan, Z. Song, D. Clayton, L. Gaetano, X. Jia, A. de Crespigny, T. Bengtsson, and R. A. D. Carano, “Joint MRI T1 unenhancing and contrast-enhancing multiple sclerosis lesion segmentation with deep learning in OPERA trials,” *Radiology*, vol. 302, no. 3, pp. 662–673, Dec. 2021.
- [92] C. P. Behrenbruch, S. Petroudi, S. Bond, J. D. Declerck, F. J. Leong, and J. M. Brady, “Image filtering techniques for medical image post-processing: an overview,” *Br. J. Radiol.*, vol. 77 Spec No 2, no. suppl_2, pp. S126–32, 2004.
- [93] T. Rädtsch, A. Reinke, V. Weru, M. D. Tizabi, N. Heller, F. Isensee, A. Kopp-Schneider, and L. Maier-Hein, “Quality assured: Rethinking annotation strategies in imaging AI,”

- in *Lecture Notes in Computer Science*, ser. Lecture notes in computer science. Cham: Springer Nature Switzerland, 2025, pp. 52–69.
- [94] T. Rädtsch, A. Reinke, V. Weru, M. D. Tizabi, N. Schreck, A. E. Kavur, B. Pekdemir, T. Roß, A. Kopp-Schneider, and L. Maier-Hein, “Labelling instructions matter in biomedical image analysis,” *Nat. Mach. Intell.*, vol. 5, no. 3, pp. 273–283, Mar. 2023.
 - [95] B. Freeman, N. Hammel, S. Phene, A. Huang, R. Ackermann, O. Kanzheleva, M. Hutson, C. Taggart, Q. Duong, and R. Sayres, “Iterative quality control strategies for expert medical image labeling,” *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 9, pp. 60–71, Oct. 2021.
 - [96] S. J. Francis, “Automatic lesion identification in MRI of multiple sclerosis patients,” 2004.
 - [97] E. A. Ashton, C. Takahashi, M. J. Berg, A. Goodman, S. Totterman, and S. Ekholm, “Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI,” *J. Magn. Reson. Imaging*, vol. 17, no. 3, pp. 300–308, Mar. 2003.
 - [98] L. Zhang, R. Tanno, M.-C. Xu, C. Jin, J. Jacob, O. Ciccarelli, F. Barkhof, and D. C. Alexander, “Disentangling human error from the ground truth in segmentation of medical images,” *arXiv [cs.CV]*, Jul. 2020.
 - [99] National Academies of Sciences, Engineering, Medicine, Policy, G. Affairs, Committee on Science, Engineering, Medicine, Public Policy, Board on Research Data, Information, Division on Engineering, Physical Sciences, Committee on Applied, T. Statistics, Board on Mathematical Sciences, Analytics, Division on Earth, Life Studies, Nuclear, Radiation Studies Board, Division of Behavioral, Social Sciences, Education, Committee on National Statistics, Board on Behavioral, Cognitive, Sensory Sciences, Committee on Reproducibility, and Replicability, “Summary,” in *Reproducibility and Replicability in Science*. National Academies Press (US), May 2019.
 - [100] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool,” *BMC Med. Imaging*, vol. 15, p. 29, Aug. 2015.
 - [101] “GBM6953EE,” <https://moodle.polymtl.ca/enrol/index.php?id=2871>, accessed: 2025-5-12.
 - [102] F. E. Harrell, *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*, 2nd ed., ser. Springer series in statistics. Cham, Switzerland: Springer International Publishing, Aug. 2015.

- [103] L. Egidi, “Prediction isn’t everything, but everything is prediction,” <https://statmodeling.stat.columbia.edu/2024/01/10/prediction-isnt-everything-but-everything-is-prediction/>, accessed: 2025-4-15.
- [104] Institut de la statistique du Québec, “Classifications and definitions,” <https://statistique.quebec.ca/en/institut/methodology/classifications-and-definitions>, accessed: 2025-4-15.
- [105] O. Efthimiou, M. Seo, K. Chalkou, T. Debray, M. Egger, and G. Salanti, “Developing clinical prediction models: a step-by-step guide,” *BMJ*, vol. 386, p. e078276, Sep. 2024.
- [106] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, “Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement,” *BMC Med.*, vol. 13, no. 1, p. 1, Jan. 2015.
- [107] R. F. Wolff, K. G. M. Moons, R. D. Riley, P. F. Whiting, M. Westwood, G. S. Collins, J. B. Reitsma, J. Kleijnen, S. Mallett, and PROBAST Group†, “PROBAST: A tool to assess the risk of bias and applicability of prediction model studies,” *Ann. Intern. Med.*, vol. 170, no. 1, pp. 51–58, Jan. 2019.
- [108] H. Hemingway, P. Croft, P. Perel, J. A. Hayden, K. Abrams, A. Timmis, A. Briggs, R. Udumyan, K. G. M. Moons, E. W. Steyerberg, I. Roberts, S. Schroter, D. G. Altman, R. D. Riley, and PROGRESS Group, “Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes,” *BMJ*, vol. 346, no. feb05 1, p. e5595, Feb. 2013.
- [109] R. D. Riley, J. A. Hayden, E. W. Steyerberg, K. G. M. Moons, K. Abrams, P. A. Kyzas, N. Malats, A. Briggs, S. Schroter, D. G. Altman, H. Hemingway, and PROGRESS Group, “Prognosis research strategy (PROGRESS) 2: prognostic factor research,” *PLoS Med.*, vol. 10, no. 2, p. e1001380, Feb. 2013.
- [110] E. W. Steyerberg, K. G. M. Moons, D. A. van der Windt, J. A. Hayden, P. Perel, S. Schroter, R. D. Riley, H. Hemingway, D. G. Altman, and PROGRESS Group, “Prognosis research strategy (PROGRESS) 3: prognostic model research,” *PLoS Med.*, vol. 10, no. 2, p. e1001381, Feb. 2013.
- [111] A. D. Hingorani, D. A. v. d. Windt, R. D. Riley, K. Abrams, K. G. M. Moons, E. W. Steyerberg, S. Schroter, W. Sauerbrei, D. G. Altman, H. Hemingway, and PROGRESS Group, “Prognosis research strategy (PROGRESS) 4: stratified medicine research,” *BMJ*, vol. 346, no. feb05 1, p. e5793, Feb. 2013.

- [112] G. S. Collins, K. G. M. Moons, P. Dhiman, R. D. Riley, A. L. Beam, B. Van Calster, M. Ghassemi, X. Liu, J. B. Reitsma, M. van Smeden, A.-L. Boulesteix, J. C. Camaradou, L. A. Celi, S. Denaxas, A. K. Denniston, B. Glocker, R. M. Golub, H. Harvey, G. Heinze, M. M. Hoffman, A. P. Kengne, E. Lam, N. Lee, E. W. Loder, L. Maier-Hein, B. A. Mateen, M. D. McCradden, L. Oakden-Rayner, J. Ordish, R. Parnell, S. Rose, K. Singh, L. Wynants, and P. Logullo, “TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods,” *BMJ*, vol. 385, p. e078378, Apr. 2024.
- [113] M.-P. Sylvestre, “*MSO6061-A-A23 - Introduction à la biostatistique*,” Chapitre 8: Régression : terminologie et équation d’une droite, 2023.
- [114] M. Bland, *An Introduction to Medical Statistics*, 2015.
- [115] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the royal statistical society series b-methodological*, vol. 58, no. 1, pp. 267–288, 1996.
- [116] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 42, no. 1, p. 80, Feb. 2000.
- [117] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, Apr. 2005.
- [118] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, Sep. 1936.
- [119] EliteDataScience, “Modern machine learning algorithms: Strengths and weaknesses,” <https://elitedatascience.com/machine-learning-algorithms>, May 2017, accessed: 2025-4-15.
- [120] “MeSH - PubMed basics,” <https://www.nlm.nih.gov/oet/ed/pubmed/mesh/mod00/01-000.html>, accessed: 2025-4-15.
- [121] “Machine learning - MeSH - NCBI,” <https://www.ncbi.nlm.nih.gov/mesh/2010029>, accessed: 2025-4-15.
- [122] R. P. França, A. C. Borges Monteiro, R. Arthur, and Y. Iano, “An overview of deep learning in big data, image, and signal processing in the modern digital age,” in *Trends in Deep Learning Methodologies*. Elsevier, Jan. 2021, pp. 63–87.
- [123] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, “Machine learning in medicine: a practical introduction,” *BMC Med. Res. Methodol.*, vol. 19, no. 1, p. 64, Mar. 2019.

- [124] R. C. Deo, “Machine learning in medicine,” *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015.
- [125] L. Lo Vercio, K. Amador, J. J. Bannister, S. Crites, A. Gutierrez, M. E. MacDonald, J. Moore, P. Mouches, D. Rajashekar, S. Schimert, N. Subbanna, A. Tuladhar, N. Wang, M. Wilms, A. Winder, and N. D. Forkert, “Supervised machine learning tools: a tutorial for clinicians,” *J. Neural Eng.*, vol. 17, no. 6, p. 062001, Nov. 2020.
- [126] W. A. Belson, “Matching and prediction on the principle of biological classification,” *J. R. Stat. Soc. Ser. C. Appl. Stat.*, vol. 8, no. 2, p. 65, Jun. 1959.
- [127] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [128] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [129] J. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
- [130] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [131] E. Fix and J. L. Hodges, join(’ ’), “Discriminatory analysis. nonparametric discrimination: Consistency properties,” *Int. Stat. Rev.*, vol. 57, no. 3, p. 238, Dec. 1989.
- [132] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*, 2nd ed., ser. Springer series in statistics. New York, NY: Springer, Dec. 2009.
- [133] “sklearn.ensemble,” <https://scikit-learn.org/stable/api/sklearn.ensemble.html>, accessed: 2025-4-17.
- [134] “1. supervised learning,” https://scikit-learn.org/stable/supervised_learning.html, accessed: 2025-4-17.
- [135] “2. unsupervised learning,” https://scikit-learn.org/stable/unsupervised_learning.html, accessed: 2025-4-17.
- [136] C. L. Andaur Navarro, J. A. A. Damen, T. Takada, S. W. J. Nijman, P. Dhiman, J. Ma, G. S. Collins, R. Bajpai, R. D. Riley, K. G. M. Moons, and L. Hooft, “Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review,” *BMJ*, vol. 375, p. n2281, Oct. 2021.

- [137] G. Cawley and N. L. C. Talbot, “On over-fitting in model selection and subsequent selection bias in performance evaluation,” *J. Mach. Learn. Res.*, vol. 11, pp. 2079–2107, Mar. 2010.
- [138] H. Liu, “Feature selection,” in *Encyclopedia of Machine Learning*. Boston, MA: Springer US, 2011, pp. 402–406.
- [139] X. Zhang, “Regularization,” in *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US, 2016, pp. 1–5.
- [140] ElisseeffAndré, “An introduction to variable and feature selection,” *J. Mach. Learn. Res.*, Mar. 2003.
- [141] Y. Saeys, I. Inza, and P. Larrañaga, “A review of feature selection techniques in bioinformatics,” *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Oct. 2007.
- [142] S. Friedrich, A. Groll, K. Ickstadt, T. Kneib, M. Pauly, J. Rahnenführer, and T. Friede, “Regularization approaches in clinical biostatistics: A review of methods and their applications,” *Stat. Methods Med. Res.*, vol. 32, no. 2, pp. 425–440, Feb. 2023.
- [143] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [144] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 281–305, Feb. 2012.
- [145] J. Snoek, H. Larochelle, and R. P. Adams, “Practical bayesian optimization of machine learning algorithms,” *arXiv [stat.ML]*, Jun. 2012.
- [146] E. W. Steyerberg, *Clinical prediction models: A practical approach to development, validation, and updating*, 2nd ed., ser. Statistics for Biology and Health. Cham, Switzerland: Springer Nature, Aug. 2019.
- [147] A. C. Justice, K. E. Covinsky, and J. A. Berlin, “Assessing the generalizability of prognostic information,” *Ann. Intern. Med.*, vol. 130, no. 6, pp. 515–524, Mar. 1999.
- [148] D. M. Allen, “The relationship between variable selection and data agumentation and a method for prediction,” *Technometrics*, vol. 16, no. 1, p. 125, Feb. 1974.
- [149] S. Borra and A. Di Ciaccio, “Measuring the prediction error. a comparison of cross-validation, bootstrap and covariance penalty methods,” *Comput. Stat. Data Anal.*, vol. 54, no. 12, pp. 2976–2989, Dec. 2010.

- [150] J.-H. Kim, “Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap,” *Comput. Stat. Data Anal.*, vol. 53, no. 11, pp. 3735–3745, Sep. 2009.
- [151] B. Efron, “The estimation of prediction error: Covariance penalties and cross-validation,” *J. Am. Stat. Assoc.*, vol. 99, no. 467, pp. 619–632, Sep. 2004.
- [152] M. Schomaker and C. Heumann, “Bootstrap inference when using multiple imputation,” *Stat. Med.*, vol. 37, no. 14, pp. 2252–2266, Jun. 2018.
- [153] Y. Vergouwe, P. Royston, K. G. M. Moons, and D. G. Altman, “Development and validation of a prediction model with missing predictor data: a practical approach,” *J. Clin. Epidemiol.*, vol. 63, no. 2, pp. 205–214, Feb. 2010.
- [154] A. M. Wood, P. Royston, and I. R. White, “The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data: The estimation and use of predictions for the assessment,” *Biom. J.*, vol. 57, no. 4, pp. 614–632, Jul. 2015.
- [155] E. W. Steyerberg, “Validation in prediction research: the waste by data splitting,” *J. Clin. Epidemiol.*, vol. 103, pp. 131–133, Nov. 2018.
- [156] A. Gupta, T. S. Stead, and L. Ganti, “Determining a meaningful R-squared value in clinical medicine,” *Academic Medicine & Surgery*, Oct. 2024.
- [157] J. A. Hanley and B. J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve,” *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.
- [158] G. W. Brier, “Verification of forecasts expressed in terms of probability,” *Mon. Weather Rev.*, vol. 78, no. 1, pp. 1–3, Jan. 1950.
- [159] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.
- [160] P. C. Austin and E. W. Steyerberg, “Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers,” *Stat. Med.*, vol. 33, no. 3, pp. 517–535, Feb. 2014.
- [161] S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *arXiv [cs.AI]*, May 2017.

- [162] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [163] S. Javeed, J. K. Greenberg, J. K. Zhang, C. F. Dibble, J. M. Khalifeh, Y. Liu, T. J. Wilson, L. J. Yang, Y. Park, and W. Z. Ray, “Derivation and validation of a clinical prediction rule for upper limb functional outcomes after traumatic cervical spinal cord injury,” *JAMA Netw. Open*, vol. 5, no. 12, p. e2247949, Dec. 2022.
- [164] D. Najafali, M. Pozin, A. Naik, B. MacInnis, N. Subbarao, S. L. Zuckerman, and P. M. Arnold, “Early predictors and outcomes of american spinal injury association conversion at discharge in surgical and nonsurgical management of sports-related spinal cord injury,” *World Neurosurg.*, vol. 171, pp. e93–e107, Nov. 2022.
- [165] Z. Qiu, F. Wang, Y. Hong, J. Zhang, H. Tang, X. Li, S. Jiang, Z. Lv, S. Liu, S. Chen, and J. Liu, “Clinical predictors of neurological outcome within 72 h after traumatic cervical spinal cord injury,” *Sci. Rep.*, vol. 6, p. 38909, Dec. 2016.
- [166] S. El Sammak, G. D. Michalopoulos, N. Arya, A. R. Bhandarkar, F. M. Moinuddin, R. Jarrah, Y. U. Yolcu, A. Shoushtari, and M. Bydon, “Prediction model for neurogenic bladder recovery one year after traumatic spinal cord injury,” *World Neurosurg.*, vol. 179, pp. e222–e231, Aug. 2023.
- [167] G. Scivoletto, M. Torre, M. Iosa, M. R. Porto, and M. Molinari, “Prediction model for the presence of complications at admission to rehabilitation after traumatic spinal cord injury,” *Top. Spinal Cord Inj. Rehabil.*, vol. 24, no. 2, pp. 151–156, Nov. 2017.
- [168] N. Dietz, Vaitheesh Jaganathan, V. Alkin, J. Mettelle, M. Boakye, and D. Drazin, “Machine learning in clinical diagnosis, prognostication, and management of acute traumatic spinal cord injury (SCI): A systematic review,” *J. Clin. Orthop. Trauma*, vol. 35, no. 102046, p. 102046, Dec. 2022.
- [169] O. Khan, J. H. Badhiwala, G. Grasso, and M. G. Fehlings, “Use of machine learning and artificial intelligence to drive personalized medicine approaches for spine care,” *World Neurosurg.*, vol. 140, pp. 512–518, Aug. 2020.
- [170] D. Kapoor and C. Xu, “Spinal cord injury AIS predictions using machine learning,” *eNeuro*, vol. 10, no. 1, Jan. 2023.
- [171] T. Shimizu, K. Suda, S. Maki, M. Koda, S. Matsumoto Harmon, M. Komatsu, M. Ota, H. Ushirozako, A. Minami, M. Takahata, N. Iwasaki, H. Takahashi, and M. Yamazaki,

- “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,” *J. Clin. Neurosci.*, vol. 107, pp. 150–156, Nov. 2022.
- [172] S. Maki, T. Furuya, T. Inoue, A. Yunde, M. Miura, Y. Shiratani, Y. Nagashima, J. Maruyama, Y. Shiga, K. Inage, Y. Eguchi, S. Orita, and S. Ohtori, “Machine learning web application for predicting functional outcomes in patients with traumatic spinal cord injury following inpatient rehabilitation,” *J. Neurotrauma*, vol. 41, no. 9-10, pp. 1089–1100, May 2024.
- [173] Y. Facchinello, M. Beauséjour, A. Richard-Denis, C. Thompson, and J.-M. Mac-Thiong, “Use of regression tree analysis for predicting the functional outcome after traumatic spinal cord injury,” *J. Neurotrauma*, vol. 38, no. 9, pp. 1285–1291, May 2021.
- [174] Z. DeVries, M. Hoda, C. S. Rivers, A. Maher, E. Wai, D. Moravek, A. Stratton, S. Kingwell, N. Fallah, J. Paquet, P. Phan, and RHSCIR Network, “Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients,” *Spine J.*, vol. 20, no. 2, pp. 213–224, Sep. 2019.
- [175] C. Kato, O. Uemura, Y. Sato, and T. Tsuji, “Decision tree analysis accurately predicts discharge destination after spinal cord injury rehabilitation,” *Arch. Phys. Med. Rehabil.*, vol. 105, no. 1, pp. 88–94, Sep. 2023.
- [176] T. Belliveau, A. M. Jette, S. Seetharama, J. Axt, D. Rosenblum, D. Larose, B. Houlihan, M. Slavin, and C. Larose, “Developing artificial neural network models to predict functioning one year after traumatic spinal cord injury,” *Arch. Phys. Med. Rehabil.*, vol. 97, no. 10, pp. 1663–1668.e3, May 2016.
- [177] N. Fallah, V. K. Noonan, Z. Waheed, C. S. Rivers, T. Plashkes, M. Bedi, M. Etminan, N. P. Thorogood, T. Ailon, E. Chan, N. Dea, C. Fisher, R. Charest-Morin, S. Paquette, S. Park, J. T. Street, B. K. Kwon, and M. F. Dvorak, “Development of a machine learning algorithm for predicting in-hospital and 1-year mortality after traumatic spinal cord injury,” *Spine J.*, vol. 22, no. 2, pp. 329–336, Feb. 2022.
- [178] G. Fan, H. Liu, S. Yang, L. Luo, M. Pang, B. Liu, L. Zhang, L. Han, L. Rong, and X. Liao, “Early prognostication of critical patients with spinal cord injury: A machine learning study with 1485 cases,” *Spine (Phila. Pa. 1976)*, vol. 49, no. 11, pp. 754–762, Nov. 2023.

- [179] S. Shabani, M. Kaushal, M. Budde, and S. N. Kurpad, “Correlation of magnetic resonance diffusion tensor imaging parameters with american spinal injury association score for prognostication and long-term outcomes,” *Neurosurg. Focus*, vol. 46, no. 3, p. E2, Mar. 2019.
- [180] K. Shibahashi, M. Nishida, Y. Okura, and Y. Hamabe, “Epidemiological state, predictors of early mortality, and predictive models for traumatic spinal cord injury: A multicenter nationwide cohort study,” *Spine (Phila. Pa. 1976)*, vol. 44, no. 7, pp. 479–487, Apr. 2019.
- [181] S. Okimatsu, S. Maki, T. Furuya, T. Fujiyoshi, M. Kitamura, T. Inada, M. Aramomi, T. Yamauchi, T. Miyamoto, T. Inoue, A. Yunde, M. Miura, Y. Shiga, K. Inage, S. Orita, Y. Eguchi, and S. Ohtori, “Determining the short-term neurological prognosis for acute cervical spinal cord injury using machine learning,” *J. Clin. Neurosci.*, vol. 96, pp. 74–79, Feb. 2022.
- [182] B. G. Leypold, A. E. Flanders, and A. S. Burns, “The early evolution of spinal cord lesions on MR imaging following traumatic spinal cord injury,” *AJNR Am. J. Neuro-radiol.*, vol. 29, no. 5, pp. 1012–1016, May 2008.
- [183] J. Chandra, F. Sheerin, L. Lopez de Heredia, T. Meagher, D. King, M. Belci, and R. J. Hughes, “MRI in acute and subacute post-traumatic spinal cord injury: pictorial review,” *Spinal Cord*, vol. 50, no. 1, pp. 2–7, Jan. 2012.
- [184] O. A. Zaninovich, M. J. Avila, M. Kay, J. L. Becker, R. J. Hurlbert, and N. L. Martirosyan, “The role of diffusion tensor imaging in the diagnosis, prognosis, and assessment of recovery and treatment of spinal cord injury: a systematic review,” *Neurosurg. Focus*, vol. 46, no. 3, p. E7, Mar. 2019.
- [185] S. Papadopoulos, N. Selden, D. Quint, N. Patel, B. Gillespie, and S. Grube, “Immediate spinal cord decompression for cervical spinal cord injury: feasibility and outcome,” *J. Trauma*, vol. 52, no. 2, pp. 323–332, Feb. 2002.
- [186] V. K. Noonan, B. K. Kwon, L. Soril, M. G. Fehlings, R. J. Hurlbert, A. Townson, M. Johnson, M. F. Dvorak, and RHSCIR Network, “The rick hansen spinal cord injury registry (RHSCIR): a national patient-registry,” *Spinal Cord*, vol. 50, no. 1, pp. 22–27, Jan. 2012.
- [187] “Home - praxis spinal cord institute - moving SCI knowledge into action,” <https://praxisinstitute.org/>, Aug. 2019, accessed: 2025-4-29.

- [188] “STROBE - strengthening the reporting of observational studies in epidemiology,” <https://www.strobe-statement.org/>, accessed: 2025-4-29.
- [189] “Home page,” <https://www.alliancecan.ca/en/node/10>, accessed: 2025-4-29.
- [190] “BIDS - the brain imaging data structure,” <https://bids.neuroimaging.io/index.html>, accessed: 2025-3-27.
- [191] F. Faul, “G*Power,” <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>, Nov. 2024, accessed: 2025-4-29.
- [192] R. D. Riley, J. Ensor, K. I. E. Snell, F. E. Harrell, Jr, G. P. Martin, J. B. Reitsma, K. G. M. Moons, G. Collins, and M. van Smeden, “Calculating the sample size required for developing a clinical prediction model,” *BMJ*, vol. 368, p. m441, Mar. 2020.
- [193] “nrrd,” <https://teem.sourceforge.net/nrrd/>, accessed: 2025-5-12.
- [194] “NIFTI: — neuroimaging informatics technology initiative,” <https://nifti.nimh.nih.gov/>, Jan. 2005, accessed: 2025-5-12.
- [195] “PEP 8 – style guide for python code,” <https://peps.python.org/pep-0008/>, accessed: 2025-4-29.
- [196] B. De Leener, S. Lévy, S. M. Dupont, V. S. Fonov, N. Stikov, D. Louis Collins, V. Callot, and J. Cohen-Adad, “SCT: Spinal cord toolbox, an open-source software for processing spinal cord MRI data,” *Neuroimage*, vol. 145, no. Pt A, pp. 24–43, Jan. 2017.
- [197] E. Naga Karthik, J. Valošek, A. C. Smith, D. Pfyffer, S. Schading-Sassenhausen, L. Farner, K. A. Weber, 2nd, P. Freund, and J. Cohen-Adad, “SCIseg: Automatic segmentation of intramedullary lesions in spinal cord injury on T2-weighted MRI scans,” *Radiol. Artif. Intell.*, vol. 7, no. 1, p. e240005, Jan. 2025.
- [198] S. Bédard, E. N. Karthik, C. Tsagkas, E. Pravatà, C. Granziera, A. Smith, K. A. Weber, Ii, and J. Cohen-Adad, “Towards contrast-agnostic soft segmentation of the spinal cord,” *Med. Image Anal.*, vol. 101, no. 103473, p. 103473, Apr. 2025.
- [199] A. Reinke, M. D. Tizabi, M. Baumgartner, M. Eisenmann, D. Heckmann-Nötzel, A. E. Kavur, T. Radsch, C. H. Sudre, L. Acion, M. Antonelli, T. Arbel, S. Bakas, A. Benis, F. Buettner, M. J. Cardoso, V. Cheplygina, J. Chen, E. Christodoulou, B. A. Cimini, K. Farahani, L. Ferrer, A. Galdran, B. van Ginneken, B. Glocker, P. Godau,

- D. A. Hashimoto, M. M. Hoffman, M. Huisman, F. Isensee, P. Jannin, C. E. Kahn, D. Kainmueller, B. Kainz, A. Karargyris, J. Kleesiek, F. Kofler, T. Kooi, A. Kopp-Schneider, M. Kozubek, A. Kreshuk, T. Kurc, B. A. Landman, G. Litjens, A. Madani, K. Maier-Hein, A. L. Martel, E. Meijering, B. Menze, K. G. M. Moons, H. Müller, B. Nichyporuk, F. Nickel, J. Petersen, S. M. Rafelski, N. Rajpoot, M. Reyes, M. A. Riegler, N. Rieke, J. Saez-Rodriguez, C. I. Sánchez, S. Shetty, R. M. Summers, A. A. Taha, A. Tiulpin, S. A. Tsaftaris, B. Van Calster, G. Varoquaux, Z. R. Yaniv, P. F. Jäger, and L. Maier-Hein, “Understanding metric-related pitfalls in image analysis validation,” *Nat. Methods*, vol. 21, no. 2, pp. 182–194, Feb. 2024.
- [200] “NumPy,” <https://numpy.org/>, accessed: 2025-5-17.
- [201] K. Anderson, S. Aito, M. Atkins, F. Biering-Sørensen, S. Charlifue, A. Curt, J. Ditunno, C. Glass, R. Marino, R. Marshall, M. J. Mulcahey, M. Post, G. Savic, G. Scivoletto, A. Catz, and Functional Recovery Outcome Measures Work Group, “Functional recovery measures for spinal cord injury: an evidence-based review for clinical practice and research: Report of the national institute on disability and rehabilitation research spinal cord injury measures meeting,” *J. Spinal Cord Med.*, vol. 31, no. 2, pp. 133–144, 2008.
- [202] J. T. Hancock and T. M. Khoshgoftaar, “Survey on categorical data for neural networks,” *J. Big Data*, vol. 7, no. 1, pp. 1–41, Dec. 2020.
- [203] X. Wu, J. Liu, L. G. Tanadini, D. P. Lammertse, A. R. Blight, J. L. K. Kramer, G. Scivoletto, L. Jones, S. Kirshblum, R. Abel, J. Fawcett, E. Field-Fote, J. Guest, B. Levinson, D. Maier, K. Tansey, N. Weidner, W. G. Tetzlaff, T. Hothorn, A. Curt, and J. D. Steeves, “Challenges for defining minimal clinically important difference (MCID) after spinal cord injury,” *Spinal Cord*, vol. 53, no. 2, pp. 84–91, Feb. 2015.
- [204] “rankdata — SciPy v1.15.3 manual,” <https://docs.scipy.org/doc/scipy-1.15.3/reference/generated/scipy.stats.rankdata.html>, accessed: 2025-6-22.
- [205] “scikit-learn,” <https://scikit-learn.org/stable/>, accessed: 2025-4-12.
- [206] “10.5 - information criteria and PRESS,” <https://online.stat.psu.edu/stat501/lesson/10/10.5>, accessed: 2025-5-17.
- [207] “seaborn.heatmap — seaborn 0.13.2 documentation,” <https://seaborn.pydata.org/generated/seaborn.heatmap.html>, accessed: 2025-4-20.

- [208] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, “Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach,” *Biometrics*, vol. 44, no. 3, pp. 837–845, Sep. 1988.
- [209] “An introduction to explainable AI with shapley values — SHAP latest documentation,” https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html, accessed: 2025-5-17.
- [210] Z. Liao, S. Hu, Y. Xie, and Y. Xia, “Modeling annotator preference and stochastic annotation error for medical image segmentation,” *Med. Image Anal.*, vol. 92, no. 103028, p. 103028, Feb. 2024.
- [211] J. van Griethuysen, “SlicerCaseIterator: Simple scripted module to batch process patients in 3D slicer.”
- [212] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig, “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.
- [213] “NASA task load index,” <https://digital.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/nasa-task-load-index>, accessed: 2025-5-20.
- [214] D. G. Altman, *Practical Statistics for Medical Research*. London: Chapman and Hall, 1991.
- [215] T. Shimizu, K. Inomata, K. Suda, S. Matsumoto Harmon, M. Komatsu, M. Ota, H. Ushirozako, A. Minami, S. Maki, T. Endo, K. Yamada, N. Iwasaki, H. Takahashi, M. Yamazaki, and M. Koda, “A multimodal machine learning model integrating clinical and MRI data for predicting neurological outcomes following surgical treatment for cervical spinal cord injury,” *Eur. Spine J.*, Apr. 2025.
- [216] “A look at spinal cord injury in canada in 2020,” <https://praxisinstitute.org/a-look-at-spinal-cord-injury-in-canada-in-2020/>, Nov. 2022, accessed: 2025-5-16.
- [217] J. Milligan, K. Ryan, M. Fehlings, and C. Bauman, “Degenerative cervical myelopathy: Diagnosis and management in primary care,” *Can. Fam. Physician*, vol. 65, no. 9, pp. 619–624, Sep. 2019.

- [218] “Reports and stats,” <https://sites.uab.edu/nscisc/reports-and-stats/>, accessed: 2025-5-16.
- [219] P. K. Ozili, “The acceptable R-square in empirical modelling for social science research,” *SSRN Electron. J.*, Jun. 2022.
- [220] “r2_score,” https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html, accessed: 2025-5-16.
- [221] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput. Sci.*, vol. 7, no. e623, p. e623, Jul. 2021.
- [222] “Welcome to pyradiomics documentation! — pyradiomics v3.1.0rc2.post5+g6a761c4 documentation,” <https://pyradiomics.readthedocs.io/en/latest/>, accessed: 2025-5-18.
- [223] M. J. Pencina, R. B. D’Agostino, K. M. Pencina, A. C. J. W. Janssens, and P. Greenland, “Interpreting incremental value of markers added to risk prediction models,” *Am. J. Epidemiol.*, vol. 176, no. 6, pp. 473–481, Sep. 2012.
- [224] “ChatGPT,” <https://chatgpt.com/>, accessed: 2025-5-17.

**APPENDIX A DETAILED ACKNOWLEDGMENTS ABOUT THE RICK
HANSEN SPINAL CORD INJURY REGISTRY (RHSCIR)
SITES AND PARTICIPANTS**

In details, I would like to thank the RHSCIR participants and network, including all the participating local RHSCIR sites: Vancouver General Hospital, Foothills Hospital, Royal University Hospital, Toronto Western Hospital, St. Michael's Hospital, Sunnybrook Health Sciences Centre, Hamilton General Hospital, The Ottawa Hospital Civic Campus, Hôpital de l'Enfant Jésus, Hôpital du Sacre Coeur de Montréal, QEII Health Sciences Centre, and Saint John Regional Hospital. It is important to note that the Rick Hansen Spinal Cord Injury Registry and related work are supported by funding from the Praxis Spinal Cord Institute through the Government of Canada and the Province of British Columbia. For more information about RHSCIR, please visit <https://www.praxisinstitute.org>.

APPENDIX B EXISTING OPEN-SOURCE SOLUTIONS FOR IMAGING ANALYSIS

This appendix summarizes the research strategy—developed with the guidance of Julien Chevrier, librarian at Polytechnique Montréal—used to identify existing open-source solutions for segmentation and classification tasks relevant to the *Rick Hansen Spinal Cord Injury Registry* imaging substudy. The strategy consists of three complementary approaches:

1. **Review of existing 3D Slicer modules or common open source software:** Table B.1 presents a non-systematic review of 3D Slicer modules (or widely used softwares) that could meet the requirements for segmentation and classification. Predefined inclusion criteria were applied to assess compatibility between available tools and project-specific expectations;
2. **Literature overview:** Table B.4 summarizes publications that describe open-source solutions for segmentation and/or classification. These were identified through targeted literature searches;
3. **Web-based code search:** Figure B.2 illustrates the outcome of an informatics-driven search for relevant code repositories and tools available online.

These three components were pursued in a complementary manner, reflecting the heterogeneous nature of the information sought. Notably, this multi-pronged investigation did not identify any existing open-source solution that fully met the specific requirements of this study or anticipate the needs for efficient annotations in further projects. The absence of such an open-source solution suggests that these tools are either not widely adopted within the research community or remain insufficiently developed—thereby reinforcing the need to design a novel, customized approach.

| Module/Software | Desc | Dev | Install | Sufficient Docx | NIfTI/NRRD Format | Case List | Custom Features | Resume Tasks | Version | Combined Seg Class | Main Adv | Main Inconv | General Comments |
|--------------------------|---|-----|---------|-----------------|-------------------|-----------|-----------------|--------------|---------|--------------------|------------------|---|--|
| <i>In 3D Slicer:</i> | | | | | | | | | | | | | |
| SlicerCaseIterator | “streamline segmentation of image datasets by handling loading and saving” | - | - | - | ✓ | ✓ | - | ✓ | ✓ | - | Similar purpose | Inactive; Missing key features; Requires CSV file; Instructions lacking | Similar purpose, not enough developed |
| MONAILabel Reviewer | “tool for research groups to check the quality of the segmentation” | ✓ | ✓ | ✓ | - | ✓ | - | - | - | - | GPU Integration; | Cloud-based; No custom | - |
| Segmentation Review | “tool for clinicians who need to quickly and efficiently review segmentations” | - | ✓ | ✓ | ✓ | ✓ | - | - | - | - | Similar purpose | Only for review and no editing; Only forward | Not pragmatic, limited functionalities |
| mpReview | ‘module that facilitates review and annotation (segmentation) of multi-parametric imaging datasets’ | - | ✓ | ✓ | - | - | - | - | - | - | - | Only DICOM; Not customizable | nan |
| SlicerCleverSegmentation | “Semi-automatic segmentation of anatomical structures” | - | ✓ | ✓ | ✓ | - | ✓ | - | - | - | - | - | Enhance segmentation features |

Table B.1 Listing of Existing Open-Source 3D Slicer Module or Software for Manual Segmentation and/or Classification Tasks (Page 1)

Predefined Criteria: Desc = Description of the module or software; Dev = Is the platform being actively developed?; Install = Is the installation intuitive?; Sufficient Docx = Is there sufficient documentation?; NIfTI/NRRD Format = Does the module or software support NIfTI and/or NRRD format; Case List = Is there a case list?; Custom Features = Does the user have the possibility to custom features (e.g. keyboard shortcuts)?; Resume Tasks = Does the user have the ability to resume previous tasks? Version = Does the module support version control? Combined Seg Class = Does the module enable both segmentation and classification tasks? Main Adv = Main advantages of the module; Main Inconv = Main inconvenients.

| Module/Software | Desc | Dev | Install | Sufficient Docx | NIfTI/NRRD Format | Case List | Custom Features | Resume Tasks | Version | Combined Seg Class | Main Adv | Main Inconv | General Comments |
|-------------------------------------|--|-----|---------|-----------------|-------------------|-----------|-----------------|--------------|---------|--------------------|---------------|-----------------------------------|--|
| Chest Imaging Platform | "Chest Imaging Platform" | - | ✓ | ✓ | - | - | - | - | - | - | - | For thoracic imaging | - |
| Other platforms: | - | - | - | - | - | - | - | - | - | - | - | - | - |
| FSLeves | "FSLeves is a viewer for neuroimaging data" | ✓ | - | ✓ | ✓ | - | - | - | - | - | Used in neuro | Lack many features | Not optimal for large datasets |
| ITK-SNAP | "software application to segment structures biomedical images" | ✓ | ✓ | ✓ | ✓ | - | - | - | - | - | Large support | No case list | - |
| OHIF | "open source, web-based, medical imaging platform" | ✓ | - | ✓ | - | ✓ | ✓ | - | - | - | Online | Online | Difficult to use |
| FreeSurfer | "software package for the analysis and visualization of structural and functional neuroimaging data" | ✓ | - | - | - | - | ✓ | - | - | - | - | Only for neuro, install deficient | Basic |
| ImageJ | "designed for scientific multidimensional images" | ✓ | ✓ | ✓ | ✓ | - | - | - | - | ✓ | Support | For pathology | Good alternative, but not for large datasets |
| Medical Imaging Interaction Toolkit | "a powerful and free application to view, process, and segment medical images" | ✓ | ✓ | ✓ | ✓ | - | - | - | - | ✓ | Good solution | Unability to custom | Good alternative |

Table B.2 Listing of Existing Open-Source 3D Slicer Module or Software for Manual Segmentation and/or Classification Tasks (Page 2)

Predefined Criteria: Desc = Description of the module or software; Dev = Is the platform being actively developed?; Install = Is the installation intuitive?; Sufficient Docx = Is there sufficient documentation?; NIfTI/NRRD Format = Does the module or software support NIfTI and/or NRRD format; Case List = Is there a case list?; Custom Features = Does the user have the possibility to custom features (e.g. keyboard shortcuts)?; Resume Tasks = Does the user have the ability to resume previous tasks? Version = Does the module support version control? Combined Seg Class = Does the module enable both segmentation and classification tasks? Main Adv = Main advantages of the module; Main Inconv = Main inconvenients.

| Module/Software | Desc | Dev | Install | Sufficient Docx | NIfTI/NRRD Format | Case List | Custom Features | Resume Tasks | Version | Combined Seg Class | Main Adv | Main Inconv | General Comments |
|--------------------------|--|-----|---------|-----------------|-------------------|-----------|-----------------|--------------|---------|--------------------|----------------------------|-----------------------------|------------------|
| rilcontour | "analysis tools to facilitate medical imaging research" | - | - | - | ✓ | - | ? | ? | ? | ? | Similar purpose | Does not work intuitively | - |
| icy.bioimageanalysis.org | "An open community platform for bioimage informatics" | - | - | - | ? | ? | ? | ? | ? | ? | - | Not still developed | - |
| DIPlib | "provide a one-stop library and development environment for quantitative image analysis" | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | Unable to make it work, C++ | - |
| Weasis | "web-based DICOM viewer" | ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ | - | Similar to clinical viewer | For DICOM | - |

Table B.3 Listing of Existing Open-Source 3D Slicer Module or Software for Manual Segmentation and/or Classification Tasks (Page 3)

Predefined Criteria: Desc = Description of the module or software; Dev = Is the platform being actively developed?; Install = Is the installation intuitive?; Sufficient Docx = Is there sufficient documentation?; NIfTI/NRRD Format = Does the module or software support NIfTI and/or NRRD format; Case List = Is there a case list?; Custom Features = Does the user have the possibility to custom features (e.g. keyboard shortcuts)?; Resume Tasks = Does the user have the ability to resume previous tasks? Version = Does the module support version control? Combined Seg Class = Does the module enable both segmentation and classification tasks? Main Adv = Main advantages of the module; Main Inconv = Main inconvenients.

Literature Overview.
Although code and/or software research might require different than conventional strategies, exploration has been done in literature to see what scientific papers have been published related to tools optimizing manual annotation and workflow in medical imaging. [Blue corresponds to thesis author's opinion.](#)

RESEARCH_DATABASE
Research_equation
YYYY-MM-DD: N Results, n selected (*with link to access them*), x presented

WEB OF SCIENCE
1- TS= ("manual segmentation" OR "manual annotation" OR (data* AND annotation) OR label*) AND (platform* OR software* OR module* OR extension* OR package*) AND ("Open Source" OR "open source" OR "free" OR "free software") AND ("Medical Imaging" OR "Radiology")
2024-11-02: 52 results, 7 selected, 3 presented

| Result Name (Article, Conference, etc.) (with hyperlink or not accessible) | Summary | Opinion |
|--|---|--|
| Image annotation and curation in radiology: an overview for machine learning practitioners | <p>Narrative review about platform used for annotation and curation in radiology European Radiology Experimental — for both technical and clinicians!</p> <p>Article discusses (following this order) about those solutions:</p> <ol style="list-style-type: none">imageJSlicer<ol style="list-style-type: none">Thanks to its features, user-friendly interface, and scripting capabilities, it is one of the most employed solutions to generate ground truth data for the development of AI tools for image processingITK-Snap<ol style="list-style-type: none">The built-in semi-automatic tool is based on active contours methods; in addition, recent versions of ITK-Snap offer a registration feature to improve the management of multimodal images, as well as a distributed segmentation service aimed at allowing cloud-based segmentation through algorithms provided by the developer community on the InternetIcyDIPlib <p><i>Interesting sentences:</i></p> <p>““Garbage in, garbage out” summarises well the importance of high-quality data in machine learning and artificial intelligence. All data used to train and validate models should indeed be consistent, standardised, traceable, correctly annotated, and de-identified, considering local regulations””</p> | <p>-Recent article: feb 2024 -Authors are researchers from Zurich -Interesting that some groups review workflow for medical image annotation/curation</p> <p>-Learned from this article:</p> <ul style="list-style-type: none">Existence of ImageJ (looks great, see review above!)Agreement that Slicer is largely used (↑ adoptability for a good-solution) |

Table B.4 Summary of Literature Review for Open-Source Solutions in Manual Segmentation and/or Classification Tasks (Page 1)

| Result Name (Article, Conference, etc.) (with hyperlink or not accessible) | Summary | Opinion |
|---|--|--|
| | <p>“In the last years, many works focused on the development of image annotation tools to speed up the process [1–3].”</p> <ul style="list-style-type: none"> References mention: <ol style="list-style-type: none"> RIL-Contour package Visual attention mechanism supporting [...] automatic image annotation (not relevant here) caBIG annotation and image markup project (workflow proposal) <p>“Some of the presented tools have been embedded in web applications that can be also easily used by healthcare professionals without a technical background”</p> <p>“The scientific research community has always shown great interest in free software and substantially contributed to the development of publicly available software for several applications. Image processing, including tasks that are more relevant for AI research such as anonymisation, curation, segmentation, and annotation of medical images, is no exception.</p> | |
| Interactive segmentation framework of the Medical Imaging Interaction Toolkit | <p>Describes a package to enhance segmentations in medical imaging: Medical Imaging Interaction Toolkit (MITK) and InteractiveSegmentation.</p> <p>Mentions existence of interactive segmentation methods (relatively recent at the time of the manuscript i.e. 2008): Graph Cut and Random Walker</p> <p>“The presented application InteractiveSegmentation is a program intended for use by medical end-users for the interactive segmentation of medical volumetric images from modalities such as CT, MRI or 3D ultrasound. The possibility to handle images with a time dimension distinguishes InteractiveSegmentation from all applications mentioned above”</p> | <p>-Manuscript from 2008 (i.e. before Slicer?!) – shows the long term questioning about optimal segmentation workflow</p> <p>-Presents a package with supposed open-source software (MITK: endless opening, las version June 2024; InteractiveSegmentation: not found)</p> <p>-Looks quite basic (Fig. 1), no case Iterator</p> <p>-Eventually, we should consider ultrasound as well in the 3D Slicer module: open an issue on Github or too early?</p> |
| An Open-Source Label Atlas Correction Tool and Preliminary Results on Huntingtons Disease Whole-Brain MRI Atlases | <p>Goal: ↓ required time for annotations</p> <p>How: proposes a process, introduces a LabelAtlasEditor in Slicer. Assessment: test their solution on Huntingtons Disease patients scans</p> <p>“This project proposes an efficient process to drastically reduce the time necessary for manual revision in order to improve atlas label quality introduce the LabelAtlasEditor tool, a SimpleITK-based open-source label atlas correction tool distributed within the image visualization software 3D Slicer.”</p> <p>“LabelAtlasEditor incorporates several 3D Slicer widgets into one consistent interface and provides label-specific correction tools, allowing for rapid identification, navigation, and modification of the small, disconnected erroneous labels within an atlas.”</p> | <p>2016 article (ok)</p> <p>LabelAtlasEditor test in Slicer LabelAtlasEditor in Slicer — not found.</p> <p>Interesting to note that the manuscript that proposes a solution tests it (if a manuscriptf from the slicer module, as discussed before, it should be tested with use cases!)</p> |

Table B.5 Summary of Literature Review for Open-Source Solutions in Manual Segmentation and/or Classification Tasks
(Page 2)

2- TS=((3dslicer OR "3d slicer" OR slicer3d OR "slicer 3d") AND (module* OR extension* OR librar* OR package*) AND (segment* OR label*))

*This research equation has been proposed by Polymtl librarian, M. Julien Chevrier.

2024-11-02: 102 results, 13 selected, 2 presented (others however shortly described).

| Result Name (Article, Conference, etc.) (with hyperlink or not accessible) | Summary | Opinion |
|--|--|---|
| SlicerRT: Radiation therapy research toolkit for 3D Slicer | 2012 module. 292 citations. Still active (commits 3 weeks ago). Effectiveness shown across multiple use cases. Towards RT which is quite specific. | Looks like a SlicerCART finality, but aimed at specific Radiotherapy tasks! Unable to make the module work. |
| 3D Slicer as a Tool for Interactive Brain Tumor Segmentation | Basic module of Slicer, mentioned to decrease annotation time. | Interesting to look how the manuscript is organized and built-on: a paper presenting SlicerCART should definitely have a use case(s) assessment component! Module not found. |
| Others ... | | |
| A Novel Tool for Supervised Segmentation Using 3D Slicer | 2018. Last commit 5 years ago. Module is not really displayed in the manuscript nor has a name to be found! | Supervised segmentation toolbox not found in the official slicer extension. |
| An open-source framework for pulmonary fissure completeness assessment | N/A | Unable to make the module work. |
| Chest_imaging_platform in Slicer extension | N/A | Unable to make the module work. |
| SlicerHeart: An open-source computing platform for cardiac image analysis and modeling | N/A | Unable to make the module work. |
| Robust semi-automatic segmentation method: an expert assistant tool for muscles in CT and MR data | N/A | Doubt about the package utility in the segmentation tasks. Unable to make the module work. |
| Evaluating free segmentation tools for CBCT-derived models: Cost-effective solutions | N/A | Compared open source with commercial alternatives, and assessed the different measures |
| TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning | For processing, but not manual annotations. | |
| 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network | One of the initial Slicer release papers (>7K citations). | |

Table B.6 Summary of Literature Review for Open-Source Solutions in Manual Segmentation and/or Classification Tasks
(Page 3)

Not accessibles:

- Bender: An Open Source Software for Efficient Model Posing and Morphing
- DeepInfer: open-source deep learning deployment toolkit for image-guided therap
From the abstract, we understand that they deployed a toolkit named DeepInfer: looks like similar to MONAILabel, and oriented more towards Radiotherapy applications. Not found in Slicer official extensions. <https://github.com/DeepInfer>
- Evaluation of 3D slicer as a medical virtual reality visualization platform
Oriented towards virtual reality. However, recognizes the lack of open source software in imaging and computer vision sciences (from the abstract)

Key summary points for an eventual Review section:

- Optimizing workflow for annotations in medical imaging is a longstanding issue (from > 1990s e.g. imageJ)
- In recent years/months (2024), active research groups continue to explore open source solutions that enable efficient imaging annotation (automated models need ground-truth data, which is often provided from manual annotations!)
- Many open-source solutions exist (list few examples), but often lack basic features or support that makes them inconvenient for the inexperienced user (ex clinicians asked to complete manual annotation tasks as considered experts)
- In the open-source solutions, Slicer is one of the most used (mention the reasons why: user-friendly, active community, customizable, possibility to build-up extensions, etc.), but lacks basic feature for an efficient dataset annotation (e.g. mention absence of viewer). Also, although many Slicer modules have been developed around the same idea (optimizing annotations), they are in parallele and not well supported, leading to abandon them after its release (e.g. mention NIL-Countour). Some of them are generic but also related to specific tasks e.g. radiotherapy (e.g SlicerRT). Our module proposes an all-included solution to optimize workflow of manual annotations, considering users without technical background (mention an example where 2 phds mention the importance of their software that should be use by clinicians, but that require terminal commands which are not easily understood by clinicians without background information)
- Review of our Approach: State that in other cases, the module's utilities are probably limited since not enough visibility! (less likely to have active support since a strength of open source models are involvement from the community and generally the more the better)
- Many slicer modules that have been developed are not easy to install or make work, decreasing the scope/reason why they have been created (give examples of inability to make different modules working!)
- Describe the review we did among existing solutions (including literature non-systematic, code research and application sknown by their reputation and discussions)
- Emphasize the importance of assessing the module (e.g. decreasing annotation times, by using at least one or multiple examples!)

Table B.7 Summary of Literature Review for Open-Source Solutions in Manual Segmentation and/or Classification Tasks
(Page 4)

Code available – Overview from various platforms.

GITHUB (used the topic:3d-slicer-extension in research bar)

2024-11-02: 218 results, 5 selected mentioned below (some are duplicate from literature and search through Slicer Extension Manager, mentioned in the first pages)

MedSAM: Slicer Plugin from Meta Segment Anything Project (also related, but by 1 individual: <https://github.com/mazurowski-lab/SlicerSegmentWithSAM>)

SegmentationReview: mentioned above

SegmentRegistration: mentioned above?

mpReview: mentioned above

SlicerSegmentationVerification: last commit 2 months ago. 1 follow-up. Shows interest at least elsewhere.

Not relevant for this purpose, but otherwise importantly related to Slicer:

Project MONAILabel

Project SlicerJupyter

SlicerRT (mentioned above)

SlicerVirtualReality (mentioned above)

SlicerHeart (mentioned above)

SlicerDebuggingTools

SlicerTorchIO (mentioned above)

SlicerMarkupsToModel

SlicerDevelopmentToolbox

SlicerPipelines

SlicerDcm2nii

SlicerNeuro: installs tools commonly needed for neuroimaging

SlicerLegacyModules

3dslicer_totalsegmentator_setup

SlicerImportGifti: commit last year

LS-SEG: LS-Seg is a 3D Slicer Extension developed to ease lumbar spine segmentation on MR images

N.B. When searching “manual-segmentation” as a topic in github, only 4 results are shown.

SOURCEFORGE (Platform similar to github, where you can research projects):

“manual segmentation”

2024-11-02: 6 results, no selected nor presented

(Research of open source projects related to manual segmentation led to ITK-SNAP Medical image Segmentation Tool only (no other project relevant mentioned))

OpenAIRE (Platform to find code and/or software, referred by polymtl Librarian, M. Julien Chevrier)

Tried, but does not seem to work (blank page for me, unable to see results)

To try (if it works):

- Advanced search using the terms: 3dslicer OR “3d slicer” OR slicer3D OR “slicer 3d”
- Limit the search to software
- MB: try with “manual segmentation” OR “manual annotation”?

Figure B.1 Results for Available Code Searching Related to Manual Segmentation and/or Classification Tools (Page 1)

Polymtl librarian M. Julien Chevrier suggested to try PapersWithCode using 3dslicer → provides only few results, related to big companies behind open source projects (e.g. SAM slicer extension, MONAI label) — *To avoid.*
 Polymtl librarian M. Julien Chevrier suggested to try site:github.com 3dslicer segmentation|label in google → provides results related to automated segmentation. Nevertheless: introduction to segmentation, brain tumor example,

List of excluded, but proposed solutions.

| Description | Reasons of exclusion |
|-------------|------------------------------|
| Flywheel | Based on a lucrative company |

Random.

Related results / information found without specifically looking for it.

| Result | Description |
|---|---|
| https://www.sciencedirect.com/science/article/pii/S016926072300442X | Progressive interactive segmentation: method how to perform segmentation? (seems semi-automated?) |
| https://supervisely.com/blog/complete-overview-and-comparison-of-manual-segmentation-approaches/ | Summary of segmentation methods approach in computer vision (interesting!) |
| https://arxiv.org/abs/2408.02635 | Segment anything from meta? |

Figure B.2 Results for Available Code Searching Related to Manual Segmentation and/or Classification Tools (Page 2)

APPENDIX C STROBE STATEMENT

STROBE Statement—checklist of items that should be included in reports of observational studies

| | Item No. | Recommendation | Page No. | Relevant text from manuscript |
|------------------------------|----------|--|----------|--|
| Title and abstract | 1 | (a) Indicate the study's design with a commonly used term in the title or the abstract | | See Chapter Methodology, in "Study Design" |
| | | (b) Provide in the abstract an informative and balanced summary of what was done and what was found | | See thesis abstract N.B. Because of the actual thesis nature, the title does not fulfill the criteria. |
| Introduction | | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | | See Introduction and Literature Review chapters. |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | | See Introduction, where objectives and hypotheses are listed. |
| Methods | | | | |
| Study design | 4 | Present key elements of study design early in the paper | | See Chapter Methodology |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | | See Chapter Methodology, in "Study Design" |
| Participants | 6 | (a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants | | See Chapter Methodology, in "Study Design" |
| | | (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case | | See Chapter Methodology, in "Study Design" |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | | See Chapter Methodology, in "Objective 3 ..." |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | | See Chapter Methodology, in "Objective 3 ..." See Chapter Methodology, in "Objective 3 ...", especially about subgroup analysis (by the Praxis Institute) and descriptive substudy. |
| Bias | 9 | Describe any efforts to address potential sources of bias | | |
| Study size | 10 | Explain how the study size was arrived at | | See Chapter Methodology, in "Study Design"-sample size |

Continued on next page

Figure C.1 STROBE Statement Filled for the *Rick Hansen Spinal Cord Injury Registry* Imaging Substudy and Thesis (Page 1)

| | | | |
|------------------------|-----|---|---|
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | See Chapter Methodology, in "Objective 3 ..." |
| Statistical methods | 12 | (a) Describe all statistical methods, including those used to control for confounding | See Chapter Methodology |
| | | (b) Describe any methods used to examine subgroups and interactions | See Chapter Methodology |
| | | (c) Explain how missing data were addressed | See Chapter Methodology |
| | | (d) Cohort study—If applicable, explain how loss to follow-up was addressed Case-control study—If applicable, explain how matching of cases and controls was addressed Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy | See Chapter Methodology |
| | | (e) Describe any sensitivity analyses | Not done for now (not in the scope of this thesis) |
| | | Results | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | See Chapter Methodology, in "Study Design" and "Objective 3..." |
| | | (b) Give reasons for non-participation at each stage | See Chapter Methodology, in "Study Design" and "Objective 3..." |
| | | (c) Consider use of a flow diagram | See Chapter Methodology, in "Study Design" |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | See Chapter Results, in cohort descriptive substudy |
| | | (b) Indicate number of participants with missing data for each variable of interest | See table in cohort descriptive substudy in Results |
| | | (c) Cohort study—Summarise follow-up time (eg, average and total amount) | N/A |
| Outcome data | 15* | Cohort study—Report numbers of outcome events or summary measures over time | N/A |
| | | Case-control study—Report numbers in each exposure category, or summary measures of exposure | N/A |
| | | Cross-sectional study—Report numbers of outcome events or summary measures | N/A |
| Main results | 16 | (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | See Chapter Results |
| | | (b) Report category boundaries when continuous variables were categorized | See Chapter Results |
| | | (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | N/A |

Continued on next page

Continued on next page

Figure C.2 STROBE Statement Filled for the *Rick Hansen Spinal Cord Injury Registry* Imaging Substudy and Thesis (Page 2)

| | | | |
|--------------------------|----|--|---|
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | See Chapter Results, in cohort descriptive substudy |
| Discussion | | | |
| Key results | 18 | Summarise key results with reference to study objectives | See Chapter Results |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | See Chapter Discussion |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | See Chapter Discussion and Conclusion |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | See Chapter Discussion |
| Other information | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | See Acknowledgements |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

Figure C.3 STROBE Statement Filled for the *Rick Hansen Spinal Cord Injury Registry* Imaging Substudy and Thesis (Page 3)

APPENDIX D MODEL HYPERPARAMETERS

For ordinal outcome prediction, models trained on full data are explored for each of those hyperparameters using MORD [204] (values arbitrarily generated using ChatGPT [224]):

```
models_and_params = {  
    'ordinal_logistic_regression': {  
        'base_model': mord.LogisticAT(alpha=1.0), # alpha is the  
            regularization parameter  
        'param_dist': {  
            'alpha': [0.01, 0.1, 1, 10]  
        }  
    },  
}
```

For continuous outcomes prediction, models trained on full data are explored for each of those hyperparameters using scikit-learn [205] (values arbitrarily generated using ChatGPT [224]):

```
model_hyperparams =  
{  
    'lasso_regressor': {  
        'base_model': Lasso(max_iter=10000),  
        'param_dist': {  
            'alpha': [0.01, 0.1, 1]  
        }  
    },  
    'linear_regressor': {  
        'base_model': LinearRegression(),  
        'param_dist': {} # No hyperparameters to tune  
    },  
    'elastic_net_regressor': {  
        'base_model': ElasticNet(max_iter=10000),  
        'param_dist': {  
            'alpha': [0.01, 0.1, 1],  
            'l1_ratio': [0.1, 0.3, 0.5]  
        }  
    }  
}
```

For dichotomized outcomes classification, models trained on full data are explored for each hyperparameter combination using scikit-learn [205] (values generated using ChatGPT [224]):

```
model_hyperparams = {
    'lasso_regressor': {
        'base_model': Lasso(max_iter=10000),
        'param_dist': {
            'alpha': [0.01, 0.1, 1]
        }
    },
    'logistic_regression': {
        'base_model': LogisticRegression(solver='liblinear',
                                          max_iter=1000),
        'param_dist': {
            'C': np.logspace(-4, 4, 20),
            'penalty': ['l1', 'l2']
        }
    }
}
```

For ordinal outcome classification, models trained using cross validation are explored using random search in scikit-learn [205] for the best hyperparameter combination for each of those models (arbitrary values generated using ChatGPT [224]):

```
models_and_params = {
    'ordinal_logistic_regression': {
        'base_model': mord.LogisticAT(alpha=1.0),
        'param_dist': {
            'alpha': [0.01, 0.1, 1, 10]
        }
    },
    'random_forest_classifier': {
        'base_model': RandomForestClassifier(),
        'param_dist': {
            'n_estimators': [100, 200],
            'max_depth': [None, 10, 20],
            'min_samples_split': [2, 5],
            'min_samples_leaf': [1, 2]
        }
    },
    'xgboost_classifier': {
        'base_model': XGBClassifier(
            objective='multi:softprob',
            num_class=5,
            use_label_encoder=False
        ),
        'param_dist': {
            'n_estimators': [100, 200],
            'max_depth': [3, 6],
            'learning_rate': [0.05, 0.1],
            'subsample': [0.8, 1.0],
            'colsample_bytree': [0.8, 1.0]
        }
    }
}
```

For continuous outcomes prediction, models trained using cross validation are explored using random search in scikit-learn [205] for the best hyperparameter combination for each of those models (arbitrary values generated using ChatGPT [224]):

```
model_hyperparams = {
    'random_forest_regressor': {
        'base_model': RandomForestRegressor(),
        'param_dist': {
            'n_estimators': [100, 200, 500],
            'max_depth': [None, 10, 20, 30],
            'min_samples_split': [2, 5, 10],
            'min_samples_leaf': [1, 2, 4],
            'max_features': ['sqrt', 'log2', None],
            'bootstrap': [True, False]
        }
    },
    'linear_regressor': {
        'base_model': LinearRegression(),
        'param_dist': {} # No hyperparams to tune
    },
    'lasso_regressor': {
        'base_model': Lasso(max_iter=10000),
        'param_dist': {
            'alpha': [0.0001, 0.001, 0.01, 0.1, 1, 10]
        }
    },
    'elastic_net_regressor': {
        'base_model': ElasticNet(max_iter=10000),
        'param_dist': {
            'alpha': [0.0001, 0.001, 0.01, 0.1, 1, 10],
            'l1_ratio': [0.1, 0.3, 0.5, 0.7, 0.9]
        }
    },
    'xgboost_regressor': {
        'base_model': XGBRegressor(objective='reg:squarederror',
                                   verbosity=0),
        'param_dist': {
            'n_estimators': [100, 200, 500],
            'max_depth': [3, 6, 10],
```



```
        'learning_rate': [0.01, 0.1, 0.2],
        'subsample': [0.6, 0.8, 1.0],
        'colsample_bytree': [0.6, 0.8, 1.0]
    }
},
'lightgbm_regressor': {
    'base_model': LGBMRegressor(),
    'param_dist': {
        'n_estimators': [100, 200, 500],
        'max_depth': [-1, 10, 20],
        'learning_rate': [0.01, 0.1, 0.2],
        'num_leaves': [31, 50, 100],
        'subsample': [0.6, 0.8, 1.0],
        'colsample_bytree': [0.6, 0.8, 1.0]
    }
}
}
```



```

    'param_dist': {
        'n_estimators': [100, 200],
        'learning_rate': [0.01, 0.1, 0.2],
        'max_depth': [3, 5, 10],
        'subsample': [0.8, 1.0],
        'colsample_bytree': [0.8, 1.0]
    }
},

'lightgbm': {
    'base_model': LGBMClassifier(),
    'param_dist': {
        'n_estimators': [100, 200],
        'learning_rate': [0.01, 0.1, 0.2],
        'num_leaves': [31, 50, 100],
        'max_depth': [-1, 5, 10]
    }
},

'svm': {
    'base_model': SVC(probability=True),
    'param_dist': {
        'C': [0.1, 1, 10],
        'kernel': ['linear', 'rbf'],
        'gamma': ['scale', 'auto']
    }
}
}

```

APPENDIX E TRIPOD-AI, CHECKLISTS

The *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis + Artificial intelligence* (TRIPOD-AI) [112] checklists for abstract (see Figure E.1) and manuscript (see Figure E.2) are filled for this thesis work.



TRIPOD+AI for Abstracts

| Section and item | Checklist item | Checked: |
|---------------------|--|----------------------|
| Title | | |
| 1 | Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted | x |
| Background | | |
| 2 | Provide a brief explanation of the healthcare context and rationale for developing or evaluating the performance of all models | x |
| Objectives | | |
| 3 | Specify the study objectives, including whether the study describes model development, evaluation, or both | x |
| Methods | | |
| 4 | Describe the sources of data | x |
| 5 | Describe the eligibility criteria and setting where the data were collected | x |
| 6 | Specify the outcome to be predicted by the model, including time horizon of predictions in case of prognostic models | x |
| 7 | Specify the type of model, a summary of the model-building steps, and the method for internal validation† | x |
| 8 | Specify the measures used to assess model performance (eg, discrimination, calibration, clinical utility) | x |
| Results | | |
| 9 | Report the number of participants and outcome events | x |
| 10 | Summarise the predictors in the final model† | x (top 10) |
| 11 | Report model performance estimates (with confidence intervals) | on calibration curve |
| Discussion | | |
| 12 | Give an overall interpretation of the main results | x |
| Registration | | |
| 13 | Give the registration number and name of the registry or repository | x |

see Praxis Institute

Figure E.1 TRIPOD-AI 2024 Abstract Checklist Filled for the *Rick Hansen Spinal Cord Injury Registry* Imaging Substudy Thesis Predictive Modeling



Version: 11-January-2024

| Section/Topic | Item | Development / evaluation ¹ | Checklist item | Reported on page |
|---------------------|------|---------------------------------------|--|--------------------------------|
| TITLE | | | | |
| Title | 1 | D;E | Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted | Thesis title |
| ABSTRACT | | | | |
| Abstract | 2 | D;E | See TRIPOD+AI for Abstracts checklist | See Appendix |
| INTRODUCTION | | | | |
| Background | 3a | D;E | Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models | x (1st sentences) |
| | 3b | D;E | Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public) | x |
| | 3c | D;E | Describe any known health inequalities between sociodemographic groups, disabilities related to TSCI; see literature review | see literature review |
| Objectives | 4 | D;E | Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both) | Thesis. Obj3 |
| METHODS | | | | |
| Data | 5a | D;E | Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data | x |
| | 5b | D;E | Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up | x |
| Participants | 6a | D;E | Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres | x (not location except Canada) |
| | 6b | D;E | Describe the eligibility criteria for study participants | x |
| | 6c | D;E | Give details of any treatments received, and how they were handled during model development or evaluation, if relevant | N/A |
| Data preparation | 7 | D;E | Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups | x |
| Outcome | 8a | D;E | Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups | Many! |
| | 8b | D;E | If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors | x |
| | 8c | D;E | Report any actions to blind assessment of the outcome to be predicted | ? |
| Predictors | 9a | D | Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building | x |
| | 9b | D;E | Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors) | x |
| | 9c | D;E | If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors | x |
| Sample size | 10 | D;E | Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation | x |
| Missing data | 11 | D;E | Describe how missing data were handled. Provide reasons for omitting any data. (*from RHSCIR) | x |
| Analytical methods | 12a | D | Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements | x |
| | 12b | D | Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation). | x |
| | 12c | D | Specify the type of model, rationale ² , all model-building steps, including any hyperparameter tuning, and method for internal validation | x |
| | 12d | D;E | Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations ³ N/A | x |
| | 12e | D;E | Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models | x |
| | 12f | E | Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings N/A | x |
| | 12g | E | For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface) | x |
| Class imbalance | 13 | D;E | If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions No | |
| Fairness | 14 | D;E | Describe any approaches that were used to address model fairness and their rationale No | |
| Model output | 15 | D | Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified (improving vs stable or worsening, context of stratify patients based on their own risk) | x |

¹ D=items relevant only to the development of a prediction model; E=items relating solely to the evaluation of a prediction model; D;E=items applicable to both the development and evaluation of a prediction model

² Separately for all model building approaches.

³ TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (eg, at different hospitals or centres). Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018]

Figure E.2 TRIPOD-AI 2024 Checklist Filled for the Rick Hansen Spinal Cord Injury Registry Imaging Substudy Thesis Predictive Modeling (Page 1)



Version: 11-January-2024

| | | | | |
|---|-----|-----|--|--|
| Training versus evaluation | 16 | D;E | Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors | N/A |
| Ethical approval | 17 | D;E | Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent | x |
| OPEN SCIENCE | | | | |
| Funding | 18a | D;E | Give the source of funding and the role of the funders for the present study | See |
| Conflicts of interest | 18b | D;E | Declare any conflicts of interest and financial disclosures for all authors | acknowledgement |
| Protocol | 18c | D;E | Indicate where the study protocol can be accessed or state that a protocol was not prepared | x |
| Registration | 18d | D;E | Provide registration information for the study, including register name and registration number, or state that the study was not registered | x |
| Data sharing | 18e | D;E | Provide details of the availability of the study data | x |
| Code sharing | 18f | D;E | Provide details of the availability of the analytical code ⁴ | x (study design) |
| PATIENT & PUBLIC INVOLVEMENT | | | | |
| Patient & Public Involvement | 19 | D;E | Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement. | Cite contributors and e.g. use of statistician |
| RESULTS | | | | |
| Participants | 20a | D;E | Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful (64%) | yes for classification tasks |
| | 20b | D;E | Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups. | x |
| | 20c | E | For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome). distributions are shown prior to model development | |
| Model development | 21 | D;E | Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation) | |
| Model specification | 22 | D | Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary) ⁵ | x |
| Model performance | 23a | D;E | Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation. | x |
| | 23b | D;E | If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details ⁵ . | N/A |
| Model updating | 24 | E | Report the results from any model updating, including the updated model and subsequent performance | x |
| DISCUSSION | | | | |
| Interpretation | 25 | D;E | Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies | x |
| Limitations | 26 | D;E | Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability | x |
| Usability of the model in the context of current care | 27a | D | Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model | x |
| | 27b | D | Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users | x |
| | 27c | D;E | Discuss any next steps for future research, with a specific view to applicability and generalizability of the model | x |

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378

⁴ This relates to the analysis code, for example, any data cleaning, feature engineering, model building, evaluation.

⁵ This relates to the code to implement the model to get estimates of risk for a new individual.

Figure E.3 TRIPOD-AI 2024 Checklist Filled for the *Rick Hansen Spinal Cord Injury Registry* Imaging Substudy Thesis Predictive Modeling (Page 2)

APPENDIX F PROCAST

PROCAST

(Prediction model study Risk Of Bias Assessment Tool)

Published in Annals of Internal Medicine (freely available):

1. [PROCAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies](#)
2. [PROCAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies. Explanation and Elaboration](#)

What does PROCAST assess?

PROCAST assesses both the *risk of bias* and *concerns regarding applicability* of a study that evaluates (develops, validates or updates) a multivariable diagnostic or prognostic prediction model. It is designed to assess primary studies included in a systematic review.

Bias occurs if systematic flaws or limitations in the design, conduct or analysis of a primary study distort the results. For the purpose of prediction modelling studies, we have defined *risk of bias* to occur when shortcomings in the study design, conduct or analysis lead to systematically distorted estimates of a model's predictive performance or to an inadequate model to address the research question. Model predictive performance is typically evaluated using calibration, discrimination and sometimes classification measures, and these are likely inaccurately estimated in studies with high risk of bias. *Applicability* refers to the extent to which the prediction model from the primary study matches your systematic review question, for example in terms of the participants, predictors or outcome of interest.

A primary study may include the development and/or validation or update of more than one prediction model. A PROCAST assessment should be completed for each distinct model that is developed, validated or updated (extended) for making individualised predictions. Where a publication assesses multiple prediction models, only complete a PROCAST assessment for those models that meet the inclusion criteria for your systematic review. Please note that subsequent use of the term "model" includes derivatives of models, such as simplified risk scores, nomograms, or recalibrations of models.

PROCAST is not designed for all multivariable diagnostic or prognostic studies. For example, studies using multivariable models to identify predictors associated with an outcome but not attempting to develop a model for making individualised predictions are not covered by PROCAST.

PROCAST includes four steps.

| Step | Task | When to complete |
|------|--|---|
| 1 | Specify your systematic review question(s) | Once per systematic review |
| 2 | Classify the type of prediction model evaluation | Once for each model of interest in each publication being assessed, for each relevant outcome |
| 3 | Assess risk of bias and applicability | Once for each development and validation of each distinct prediction model in a publication |
| 4 | Overall judgment | Once for each development and validation of each distinct prediction model in a publication |

If this is your first time using PROCAST, we strongly recommend reading the detailed explanation and elaboration (E&E, see link above) paper and to check the examples on www.procast.org

Step 1: Specify your systematic review question

State your systematic review question to facilitate the assessment of the applicability of the evaluated models to your question. The following table should be completed once per systematic review.

| Criteria | Specify your systematic review question |
|--|--|
| Intended use of model: | N/A. (No systematic review by the author for this project) |
| Participants including selection criteria and setting: | |
| Predictors (used in prediction modelling), including types of predictors (e.g. history, clinical examination, biochemical markers, imaging tests), time of measurement, specific measurement issues (e.g., any requirements/prohibitions for specialized equipment): | |
| Outcome to be predicted: | |

Figure F.1 PROCAST Guidelines – Pages 1 and 2

Step 2: Classify the type of prediction model evaluation

Use the following table to classify the evaluation as model development, model validation or model update, or combination. Different signalling questions apply for different types of prediction model evaluation. If the evaluation does not fit one of these classifications then PROBAST should not be used.

| Classify the evaluation based on its aim | | | |
|--|----------------------------|---------------------|---|
| Type of prediction study | PROBAST boxes to complete | Tick as appropriate | Definition for type of prediction model study |
| Development only | Development | x | Prediction model development without external validation. These studies may include internal validation methods, such as bootstrapping and cross-validation techniques. |
| Development and validation | Development and validation | | Prediction model development combined with external validation in other participants in the same article. |
| Validation only | Validation | | External validation of existing (previously developed) model in other participants. |

This table should be completed once for each publication being assessed and for each relevant outcome in your review. N/A

| | |
|-----------------------|--|
| Publication reference | |
| Models of interest | |
| Outcome of interest | |

Step 3: Assess risk of bias and applicability

PROBAST is structured as four key domains. Each domain is judged for risk of bias (low, high or unclear) and includes signalling questions to help make judgements. Signalling questions are rated as yes (Y), probably yes (PY), probably no (PN), no (N) or no information (NI). All signalling questions are phrased so that “yes” indicates absence of bias. Any signalling question rated as “no” or “probably no” flags the potential for bias; you will need to use your judgement to determine whether the domain should be rated as “high”, “low” or “unclear” risk of bias. The guidance document contains further instructions and examples on rating signalling questions and risk of bias for each domain.

The first three domains are also rated for concerns regarding applicability (low/ high/ unclear) to your review question defined above.

Complete all domains separately for each evaluation of a distinct model. Shaded boxes indicate where signalling questions do not apply and should not be answered.

| DOMAIN 1: Participants | | | |
|--|---|------|------|
| A. Risk of Bias | | | |
| Describe the sources of data and criteria for participant selection: Source of data: RHSCIR Study (imaging substudy); Criteria: see methodology | | | |
| | | Dev | Val |
| 1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data? | | x | x |
| 1.2 Were all inclusions and exclusions of participants appropriate? (prior fulfilling) | | x | x |
| Risk of bias introduced by selection of participants | RISK: (low/ high/ unclear) | high | high |
| Rationale of bias rating: This imaging substudy assesses imaging biomarkers, so it is restricted to a subpopulation of TSCI with preoperative MRI, whereas not all TSCI patients undergo preoperative MRI, introducing selection bias. This is addressed by describing the cohort characteristics and subgroup analysis in the RHSCIR. | | | |
| B. Applicability | | | |
| Describe included participants, setting and dates: See Methodology and Results. In brief, participants have in average 56 years old with the most frequent age group being from 61 to 75 years old. Around half experienced TSCI from fall and almos half have initial severity being AIS grade D. | | | |
| Concern that the included participants and setting do not match the review question | CONCERN: (low/ high/ unclear) | low | |
| Rationale of applicability rating: No concern that included participants do not respect the applicability context, since the goal of the study is to incorporate prediction modeling from imaging biomarkers (and this population has MRI anyway). | | | |

Figure F.2 PROBAST Guidelines – Pages 3 and 4

| DOMAIN 2: Predictors | | | |
|---|----------------------------------|-----|-----|
| A. Risk of Bias | | | |
| List and describe predictors included in the final model, e.g. definition and timing of assessment: Many. See Results and Discussion section. | | | |
| | | Dev | Val |
| 2.1 Were predictors defined and assessed in a similar way for all participants? | | x | x |
| 2.2 Were predictor assessments made without knowledge of outcome data? | (CV) | | x |
| 2.3 Are all predictors available at the time the model is intended to be used? | | x | x |
| Risk of bias introduced by predictors or their assessment | RISK: (low/ high/ unclear) | low | low |
| Rationale of bias rating: See methodology. While a method disclose that there is likely overfitting in the method applied, a complementary approach using random dataset splits and cross-validations provide a comprehensive overview. | | | |
| B. Applicability | | | |
| Concern that the definition, assessment or timing of predictors in the model do not match the review question | CONCERN: (low/ high/ unclear) | low | low |
| Rationale of applicability rating: The model development is targeted to answer the research question. | | | |

| DOMAIN 3: Outcome | | | |
|--|----------------------------------|-----|-----|
| A. Risk of Bias | | | |
| Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: See methodology and Results (for outcomes distribution) | | | |
| | | Dev | Val |
| 3.1 Was the outcome determined appropriately? | | x | x |
| 3.2 Was a pre-specified or standard outcome definition used? | | x | x |
| 3.3 Were predictors excluded from the outcome definition? | | x | x |
| 3.4 Was the outcome defined and determined in a similar way for all participants? | | x | x |
| 3.5 Was the outcome determined without knowledge of predictor information? | | x | x |
| 3.6 Was the time interval between predictor assessment and outcome determination appropriate? | | x | x |
| Risk of bias introduced by the outcome or its determination | RISK: (low/ high/ unclear) | low | low |
| Rationale of bias rating: Designed from the Pan-Canadian RHSCIR Study. Imaging substudy population selected based on availability of data of interest (i.e.g outcome, clinical and MRI data) | | | |
| B. Applicability | | | |
| At what time point was the outcome determined: at rehabilitation discharge (around 90 days) | | | |
| If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: N/A. | | | |
| Concern that the outcome, its definition, timing or determination do not match the review question | CONCERN: (low/ high/ unclear) | low | low |
| Rationale of applicability rating: Designed by clinicians, the RHSCIR imaging substudy is targeted towards clinical relevance (e.g., predict the recovery of individuals at rehabilitation discharge informs about the recovery trajectory) So, no concerns about that. | | | |

Figure F.3 PROBAST Guidelines – Pages 5 and 6

| DOMAIN 4: Analysis | | |
|---|--------------------------------------|-----------------|
| Risk of Bias | | |
| Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor: See Methodology and Results. | | |
| Describe how the model was developed (for example in regards to modelling technique (e.g. survival or logistic modelling), predictor selection, and risk group definition): See Methodology. | | |
| Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): See Methodology. | | |
| Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism: See Methodology. | | |
| Describe any participants who were excluded from the analysis: See Methodology. | | |
| Describe missing data on predictors and outcomes as well as methods used for missing data: No imputation. See Methodology. | | |
| | Dev | Val |
| 4.1 Were there a reasonable number of participants with the outcome? | x | x |
| 4.2 Were continuous and categorical predictors handled appropriately? | x | x |
| 4.3 Were all enrolled participants included in the analysis? No. If missing data, not considered in the analysis. | | |
| 4.4 Were participants with missing data handled appropriately? Yes. | x | x |
| 4.5 Was selection of predictors based on univariable analysis avoided? No, due to mix approach. See methodology. | | |
| 4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately? | x | x |
| 4.7 Were relevant model performance measures evaluated appropriately? | x | x |
| 4.8 Were model overfitting and optimism in model performance accounted for? disclosed. | | |
| 4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? no multivariable analysis done. | | |
| Risk of bias introduced by the analysis | RISK: (low/ high/ unclear) | unclear unclear |
| Rationale of bias rating: Many models, many outcomes. 2 different approach applied, aiming for re-usability in further studies that will clarify if any bias. | | |

Step 4: Overall assessment

Use the following tables to reach overall judgements about risk of bias and concerns regarding applicability of the prediction model evaluation (development and/or validation) across all assessed domains. Complete for each evaluation of a distinct model.

| Reaching an overall judgement about risk of bias of the prediction model evaluation | |
|---|--|
| Low risk of bias | If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains, consider downgrading to high risk of bias. Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation. |
| High risk of bias | If at least one domain is judged to be at high risk of bias. |
| Unclear risk of bias | If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains. |

| Reaching an overall judgement about applicability of the prediction model evaluation | |
|--|--|
| Low concerns regarding applicability | If low concerns regarding applicability for all domains, the prediction model evaluation is judged to have low concerns regarding applicability. |
| High concerns regarding applicability | If high concerns regarding applicability for at least one domain, the prediction model evaluation is judged to have high concerns regarding applicability. |
| Unclear concerns regarding applicability | If unclear concerns (but no "high concern") regarding applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns regarding applicability overall. |

| Overall judgement about risk of bias and applicability of the prediction model evaluation | | |
|---|---|--|
| Overall judgement of risk of bias | RISK: (low/ high/ unclear) | |
| Summary of sources of potential bias: selection bias (population) and overfitting. unclear. Needs more external validation. | | |
| Overall judgement of applicability | CONCERN: (low/ high/ unclear) | |
| Summary of applicability concerns: High. Using "real-world" clinical data with robust methods for "real-world" testing and implementation | | |

Figure F.4 PROBAST Guidelines – Pages 7 and 8

APPENDIX G UNIVARIATE ANALYSES FOR DEMOGRAPHICS, CLINICAL AND MRI DATA

The following pages detail the univariate analyses results for associations between demographic, clinical, and MRI data with initial and final rehabilitation discharge *American Spinal Injury Association Impairment Scale* (AIS) grade, total motor score, and upper and lower extremity motor scores. Changes in score from baseline (both absolute and > 0 versus ≤ 0) are also reported.

| Description | AIS Initial | Motor Score Initial | UEMS Initial | LEMS Initial | AIS Final | Motor Score Final | UEMS Final | LEMS Final | AIS change | Motor Score change | UEMS change | LEMS change |
|--|-------------|---------------------|--------------|--------------|-----------|-------------------|------------|------------|------------|--------------------|-------------|-------------|
| Sex | 0.185 | 0.204 | 0.155 | 0.341 | 0.099 | 0.047 | 0.032 | 0.106 | 0.710 | 0.372 | 0.452 | 0.406 |
| Site | 0.052 | 0.028 | 0.029 | 0.027 | 0.153 | 0.068 | 0.010 | 0.230 | 0.234 | 0.042 | 0.142 | 0.042 |
| Age at injury | 0.011 | 0.242 | 0.746 | 0.040 | 0.517 | 0.421 | 0.594 | 0.101 | 0.004 | 0.571 | 0.819 | 0.472 |
| Age group at injury | 0.018 | 0.102 | 0.319 | 0.050 | 0.694 | 0.020 | 0.036 | 0.022 | 0.035 | 0.149 | 0.170 | 0.214 |
| Mechanism of injury | 0.007 | 0.018 | 0.064 | 0.018 | 0.079 | 0.107 | 0.320 | 0.063 | 0.215 | 0.459 | 0.519 | 0.576 |
| Comorbidity count | 0.264 | 0.568 | 0.207 | 0.986 | 0.373 | 0.272 | 0.191 | 0.419 | 0.001 | 0.495 | 0.947 | 0.259 |
| Surgical management | <0.001 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | 0.005 | <0.001 | 0.021 | 0.059 | 0.220 | 0.042 |
| Time to surgery | <0.001 | 0.331 | 0.272 | 0.479 | <0.001 | 0.562 | 0.599 | 0.588 | 0.499 | 0.597 | 0.435 | 0.807 |
| AIS (initial) | - | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.011 | 0.169 | 0.010 |
| Motor score (initial) | <0.001 | - | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| UEMS (initial) | <0.001 | <0.001 | - | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| LEMS (initial) | <0.001 | <0.001 | <0.001 | - | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.005 | <0.001 |
| Neurological level of injury (initial) | 0.135 | <0.001 | <0.001 | 0.056 | 0.504 | 0.014 | <0.001 | 0.309 | 0.351 | 0.001 | <0.001 | 0.064 |
| High or low cervical (initial) | 0.192 | <0.001 | <0.001 | 0.164 | 0.488 | 0.047 | 0.001 | 0.475 | 0.471 | 0.025 | 0.003 | 0.247 |
| Tetraplegia severity (initial) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.008 | 0.009 | 0.097 | 0.001 |

Table G.1 Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 1)

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = *Upper Extremity Motor Score*; LEMS = *Lower Extremity Motor Score*

| Description | AIS Initial | Motor Score Initial | UEMS Initial | LEMS Initial | AIS Final | Motor Score Final | UEMS Final | LEMS Final | AIS change | Motor Score change | UEMS change | LEMS change |
|-----------------------------------|-------------|---------------------|--------------|--------------|-----------|-------------------|------------|------------|------------|--------------------|-------------|-------------|
| Cord transection | <0.001 | <0.001 | 0.185 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | 0.135 | 0.011 | 0.022 | 0.022 |
| Severe spine changes | 0.025 | 0.060 | 0.167 | 0.054 | 0.014 | 0.019 | 0.155 | 0.007 | 0.460 | 0.617 | 0.938 | 0.397 |
| Soft tissues contusion | 0.093 | 0.098 | 0.372 | 0.054 | 0.086 | 0.129 | 0.159 | 0.164 | 0.738 | 0.787 | 0.591 | 0.404 |
| Vertebral Level | 0.003 | 0.260 | <0.001 | 0.010 | 0.001 | 0.560 | 0.081 | 0.008 | 0.742 | 0.029 | 0.002 | 0.291 |
| Artifacts | 0.095 | 0.151 | 0.626 | 0.010 | 0.868 | 0.712 | 0.105 | 0.584 | 0.067 | 0.014 | 0.187 | 0.004 |
| Prevertebral collection | 0.023 | 0.008 | 0.015 | 0.019 | 0.424 | 0.335 | 0.204 | 0.532 | 0.035 | 0.020 | 0.108 | 0.015 |
| Prevertebral changes | 0.578 | 0.966 | 0.975 | 0.966 | 0.093 | 0.379 | 0.622 | 0.301 | 0.362 | 0.273 | 0.580 | 0.191 |
| Vertebral Body fracture(s) | 0.111 | 0.250 | 0.584 | 0.164 | 0.534 | 0.373 | 0.381 | 0.434 | 0.147 | 0.687 | 0.730 | 0.369 |
| Spine vertebral body misalignment | 0.156 | 0.590 | 0.863 | 0.369 | 0.072 | 0.254 | 0.901 | 0.059 | 0.225 | 0.351 | 0.609 | 0.264 |
| Multiple levels involved | 0.752 | 0.355 | 0.290 | 0.496 | 0.416 | 0.243 | 0.395 | 0.216 | 0.707 | 0.790 | 0.729 | 0.493 |
| Disc bulding | 0.449 | 0.240 | 0.395 | 0.225 | 0.236 | 0.174 | 0.208 | 0.211 | 0.696 | 0.860 | 0.679 | 0.971 |
| Multiple findings | 0.297 | 0.293 | 0.171 | 0.512 | 0.041 | 0.028 | 0.028 | 0.058 | 0.461 | 0.154 | 0.388 | 0.112 |
| Other miscellaneous findings | 0.773 | 0.824 | 0.995 | 0.734 | 0.719 | 0.702 | 0.848 | 0.641 | 0.836 | 0.844 | 0.824 | 0.887 |
| Accentuated positioning | 0.797 | 0.680 | 0.228 | 0.828 | 0.299 | 0.944 | 0.420 | 0.503 | 0.373 | 0.513 | 0.555 | 0.560 |
| Subacute lesion | 0.152 | 0.299 | 0.597 | 0.217 | 0.883 | 0.618 | 0.463 | 0.788 | 0.156 | 0.038 | 0.115 | 0.038 |
| Congenital spine variants | 0.156 | 0.432 | 0.996 | 0.222 | 0.761 | 0.687 | 0.504 | 0.869 | 0.088 | 0.109 | 0.423 | 0.055 |

Table G.1 Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 2)

AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score

| Description | AIS Initial | Motor Score Initial | UEMS Initial | LEMS Initial | AIS Final | Motor Score Final | UEMS Final | LEMS Final | AIS change | Motor Score change | UEMS change | LEMS change |
|-------------------------------------|-------------|---------------------|--------------|--------------|-----------|-------------------|------------|------------|------------|--------------------|-------------|-------------|
| Hemorrhage (axial T2w) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.293 | 0.064 | 0.066 | 0.139 |
| Hemorrhage (sagittal T2w) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.115 | 0.265 | 0.122 | 0.563 |
| Hemorrhage (axial T2*) | <0.001 | <0.001 | 0.003 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.711 | 0.424 | 0.187 | 0.823 |
| BASIC Score | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.852 | 0.695 | 0.353 | 0.622 |
| Lesion volume (axial, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.206 | 0.261 | 0.287 | 0.346 |
| Lesion volume (axial, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.075 | 0.901 | 0.910 | 0.782 |
| Lesion volume (sagittal, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.650 | 0.191 | 0.298 | 0.205 |
| Lesion volume (sagittal, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.031 | 0.081 | 0.041 |
| Lesion length (axial, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.376 | 0.433 | 0.583 | 0.420 |
| Lesion length (axial, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.501 | 0.361 | 0.555 | 0.332 |
| Lesion length (sagittal, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.983 | 0.173 | 0.337 | 0.157 |
| Lesion length (sagittal, automatic) | <0.001 | 0.035 | 0.090 | 0.041 | <0.001 | 0.019 | 0.033 | 0.030 | 0.435 | 0.780 | 0.637 | 0.933 |
| Lesion width (axial, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.858 | 0.719 | 0.899 | 0.508 |

Table G.1 Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 3)

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score

| Description | AIS Initial | Motor Score Initial | UEMS Initial | LEMS Initial | AIS Final | Motor Score Final | UEMS Final | LEMS Final | AIS change | Motor Score change | UEMS change | LEMS change |
|---|-------------|---------------------|--------------|--------------|-----------|-------------------|------------|------------|------------|--------------------|-------------|-------------|
| Lesion width (axial, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.754 | 0.216 | 0.617 | 0.125 |
| Lesion width (sagittal, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.769 | 0.408 | 0.460 | 0.460 |
| Lesion width (sagittal, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.269 | 0.307 | 0.476 | 0.294 |
| Number of lesions (axial, manual) | 0.068 | 0.131 | 0.131 | 0.206 | 0.038 | 0.093 | 0.195 | 0.080 | 0.272 | 0.847 | 0.715 | 0.559 |
| Number of lesions (axial, automatic) | 0.712 | 0.965 | 0.693 | 0.834 | 0.074 | 0.153 | 0.344 | 0.110 | 0.056 | 0.054 | 0.489 | 0.014 |
| Number of lesions (sagittal, manual) | 0.105 | 0.225 | 0.250 | 0.289 | 0.038 | 0.053 | 0.031 | 0.125 | 0.911 | 0.379 | 0.272 | 0.583 |
| Number of lesions (sagittal, automatic) | 0.023 | 0.236 | 0.632 | 0.138 | 0.001 | 0.008 | 0.133 | 0.002 | 0.259 | 0.051 | 0.203 | 0.038 |
| Maximal axial damage ratio (axial, manual) | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.847 | 0.277 | 0.740 | 0.147 |
| Maximal axial damage ratio (axial, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.952 | 0.394 | 0.373 | 0.515 |
| Lesion midsagittal length (axial, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.192 | 0.427 | 0.486 | 0.479 |
| Lesion midsagittal length (axial, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.298 | 0.557 | 0.405 | 0.776 |
| Lesion midsagittal | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.575 | 0.126 | 0.177 | 0.168 |

Table G.1 Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 4)

AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score

| Description | AIS Initial | Motor Score Initial | UEMS Initial | LEMS Initial | AIS Final | Motor Score Final | UEMS Final | LEMS Final | AIS change | Motor Score change | UEMS change | LEMS change |
|---|-------------|---------------------|--------------|--------------|-----------|-------------------|------------|------------|------------|--------------------|-------------|-------------|
| length (sagittal, manual) | | | | | | | | | | | | |
| Lesion midsagittal length (sagittal, automatic) | <0.001 | <0.001 | 0.002 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | 0.139 | 0.903 | 0.849 | 0.742 |
| Lesion midsagittal width (axial, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.752 | 0.547 | 0.628 | 0.568 |
| Lesion midsagittal width (axial, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.298 | 0.339 | 0.221 | 0.574 |
| Lesion midsagittal width (sagittal, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.798 | 0.494 | 0.509 | 0.566 |
| Lesion midsagittal width (sagittal, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.019 | 0.858 | 0.817 | 0.916 |
| Max diameter (axial, manual) | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.707 | 0.457 | 0.840 | 0.312 |
| Max diameter (axial, automatic) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.993 | 0.266 | 0.554 | 0.201 |
| Max diameter (sagittal, manual) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.876 | 0.425 | 0.564 | 0.414 |
| Max diameter (sagittal, automatic) | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.365 | 0.316 | 0.678 | 0.214 |

Table G.1 Univariate Analyses Results for Initial Data and Initial, Final and Change in Severity (Page 5)

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = *Upper Extremity Motor Score*; LEMS = *Lower Extremity Motor Score*

| Description | AIS change binarized | Motor Score change binarized | UEMS change binarized | LEMS change binarized |
|--|----------------------|------------------------------|-----------------------|-----------------------|
| Sex | 0.831 | 0.146 | 0.170 | 0.672 |
| Site | 0.659 | 0.056 | 0.050 | 0.243 |
| Age at injury | 0.134 | 0.261 | 0.329 | 0.447 |
| Age group at injury | 0.256 | 0.265 | 0.531 | 0.231 |
| Mechanism of injury | 0.589 | 0.577 | 0.727 | 0.272 |
| Comorbidity count | 0.013 | 0.482 | 0.750 | 0.306 |
| Surgical management | 0.004 | 0.582 | 0.272 | 0.421 |
| Time to surgery | 0.364 | 0.607 | 0.745 | 0.291 |
| AIS (initial) | <0.001 | 0.089 | 0.150 | <0.001 |
| Motor score (initial) | <0.001 | 0.003 | 0.003 | <0.001 |
| UEMS (initial) | <0.001 | 0.001 | 0.001 | <0.001 |
| LEMS (initial) | <0.001 | 0.024 | 0.019 | 0.005 |
| Neurological level of injury (initial) | 0.711 | <0.001 | <0.001 | 0.278 |
| High or low cervical (initial) | 0.605 | 0.488 | 0.883 | 0.447 |
| Tetraplegia severity (initial) | 0.204 | 0.387 | 0.270 | <0.001 |
| Cord transection | 0.339 | 0.456 | 0.626 | <0.001 |
| Severe spine changes | 0.792 | 0.436 | 0.721 | 0.552 |
| Soft tissues contusion | 0.709 | 0.142 | 0.246 | 0.485 |
| Vertebral Level | 0.685 | 0.743 | 0.658 | 0.359 |
| Artifacts | 0.032 | 0.247 | 0.297 | 0.380 |
| Prevertebral collection | 0.093 | 1.000 | 1.000 | 1.000 |
| Prevertebral changes | 0.219 | 1.000 | 0.738 | 0.245 |
| Vertebral Body fracture(s) | 0.145 | 0.864 | 0.298 | 0.725 |
| Spine vertebral body misalignment | 0.287 | 0.302 | 0.258 | 0.115 |
| Multiple levels involved | 1.000 | 0.348 | 0.341 | 1.000 |
| Disc bulging | 0.731 | 1.000 | 1.000 | 0.744 |
| Multiple findings | 1.000 | 0.370 | 0.391 | 0.032 |
| Other miscellaneous findings | 0.994 | 0.469 | 0.305 | 0.836 |
| Accentuated positioning | 0.731 | 0.402 | 1.000 | 0.744 |

Table G.1 Univariate Analyses Results for ≤ 0 vs 0 Change in Severity (Page 1)

AIS = American Spinal Injury Association Impairment Scale; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score

| Description | AIS change binarized | Motor Score change binarized | UEMS change binarized | LEMS change binarized |
|---|----------------------|------------------------------|-----------------------|-----------------------|
| Subacute lesion | 0.415 | 0.109 | 0.624 | 0.437 |
| Congenital spine variants | 0.530 | 0.377 | 0.410 | 1.000 |
| Hemorrhage (axial T2w) | 0.650 | 1.000 | 0.643 | 0.022 |
| Hemorrhage (sagittal T2w) | 0.187 | 0.802 | 0.805 | 0.013 |
| Hemorrhage (axial T2*) | 0.308 | 0.584 | 0.691 | 0.771 |
| BASIC Score | 0.440 | 0.438 | 0.534 | 0.424 |
| Lesion volume (axial, manual) | 0.300 | 0.859 | 0.690 | 0.066 |
| Lesion volume (axial, automatic) | 0.051 | 0.500 | 0.597 | 0.162 |
| Lesion volume (sagittal, manual) | 0.791 | 0.293 | 0.334 | 0.002 |
| Lesion volume (sagittal, automatic) | <0.001 | 0.102 | 0.194 | 0.458 |
| Lesion length (axial, manual) | 0.323 | 0.618 | 0.666 | 0.105 |
| Lesion length (axial, automatic) | 0.675 | 0.968 | 0.799 | 0.002 |
| Lesion length (sagittal, manual) | 0.826 | 0.286 | 0.270 | 0.002 |
| Lesion length (sagittal, automatic) | 0.392 | 0.155 | 0.162 | 0.216 |
| Lesion width (axial, manual) | 0.306 | 0.842 | 0.948 | 0.052 |
| Lesion width (axial, automatic) | 0.287 | 0.683 | 0.972 | 0.001 |
| Lesion width (sagittal, manual) | 0.941 | 0.432 | 0.292 | 0.028 |
| Lesion width (sagittal, automatic) | 0.206 | 0.742 | 0.628 | 0.012 |
| Number of lesions (axial, manual) | 0.164 | 1.000 | 0.879 | 0.075 |
| Number of lesions (axial, automatic) | 0.013 | 0.637 | 0.837 | 0.003 |
| Number of lesions (sagittal, manual) | 0.603 | 0.299 | 0.269 | 0.966 |
| Number of lesions (sagittal, automatic) | 0.327 | 0.460 | 0.891 | 0.003 |
| Maximal axial damage ratio (axial, manual) | 0.481 | 0.531 | 0.481 | 0.135 |
| Maximal axial damage ratio (axial, automatic) | 0.809 | 0.835 | 0.811 | 0.019 |
| Lesion midsagittal length (axial, manual) | 0.201 | 0.738 | 0.748 | 0.111 |

Table G.1 Univariate Analyses Results for ≤ 0 vs 0 Change in Severity (Page 2)

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = *Upper Extremity Motor Score*; LEMS = *Lower Extremity Motor Score*

| Description | AIS change binarized | Motor Score change binarized | UEMS change binarized | LEMS change binarized |
|---|----------------------|------------------------------|-----------------------|-----------------------|
| Lesion midsagittal length (axial, automatic) | 0.118 | 0.514 | 0.620 | 0.057 |
| Lesion midsagittal length (sagittal, manual) | 0.652 | 0.419 | 0.465 | 0.002 |
| Lesion midsagittal length (sagittal, automatic) | 0.082 | 0.535 | 0.522 | 0.386 |
| Lesion midsagittal width (axial, manual) | 0.815 | 0.510 | 0.501 | 0.030 |
| Lesion midsagittal width (axial, automatic) | 0.140 | 0.852 | 0.831 | 0.041 |
| Lesion midsagittal width (sagittal, manual) | 0.862 | 0.449 | 0.265 | 0.011 |
| Lesion midsagittal width (sagittal, automatic) | 0.028 | 0.247 | 0.289 | 0.329 |
| Max diameter (axial, manual) | 0.281 | 0.809 | 0.680 | 0.029 |
| Max diameter (axial, automatic) | 0.695 | 0.660 | 0.725 | 0.002 |
| Max diameter (sagittal, manual) | 0.606 | 0.204 | 0.189 | 0.159 |
| Max diameter (sagittal, automatic) | 0.240 | 0.691 | 0.472 | 0.019 |

Table G.1 Univariate Analyses Results for ≤ 0 vs 0 Change in Severity (Page 3)

AIS = *American Spinal Injury Association Impairment Scale*; UEMS = Upper Extremity Motor Score; LEMS = Lower Extremity Motor Score