

Segmentation multimodale et multiclasse des lésions de la moelle épinière par apprentissage profond

par

Bianca POPA

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE
AVEC MÉMOIRE EN GÉNIE DES TECHNOLOGIES DE L'INFORMATION
M. Sc. A.

MONTRÉAL, LE 27 MAI 2026

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Bianca Popa, 2026



Cette licence Creative Commons signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette oeuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'oeuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE:

M. Luc Duong, directeur de mémoire
Département de génie logiciel et TI à l'École de technologie supérieure

Mme. Sylvie Ratté, codirectrice
Département de génie logiciel et TI à l'École de technologie supérieure

M. Jean-Marc Mac-Thiong, codirecteur externe
Centre de recherche de l'Hôpital du Sacré-Cœur de Montréal

Mme. Camille Coti, présidente du jury
Département de génie logiciel et TI à l'École de technologie supérieure

Mme. Valérie Hayot-Sasson, membre du jury
Département de génie logiciel et TI à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 30 AVRIL 2026

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je remercie en premier lieu mon directeur de recherche, Luc Duong, pour son encadrement tout au long de ce projet, qui m'a permis de mener ce travail à son terme. Je remercie également ma codirectrice, Sylvie Ratté, pour ses conseils éclairés et son soutien tout au long de ce projet.

Ma gratitude va également à Jean-Marc Mac-Thiong et à toute son équipe au Centre de recherche de l'Hôpital du Sacré-Cœur de Montréal, qui ont rendu possible l'accès aux données des patients et avec qui les échanges ont nourri ce travail. Une mention particulière à Lara Chawa.

À mes amis et à mes amours : merci de me pousser à m'améliorer continuellement et de me soutenir au quotidien.

Enfin, je dédie ces remerciements à ma famille et à ma mère, pour tout.

Ce projet a été rendu possible grâce au soutien financier du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) ainsi que de la Bourse d'excellence de l'École de technologie supérieure.

Segmentation multimodale et multiclasse des lésions de la moelle épinière par apprentissage profond

Bianca POPA

RÉSUMÉ

Cette étude porte sur la segmentation automatique des deux composantes pathologiques des lésions traumatiques de la moelle épinière—œdème et hémorragie—à partir d'IRM multimodale, et sur leur quantification pronostique. La base de données comprend 148 patients recrutés au Centre de recherche de l'Hôpital du Sacré-Cœur de Montréal, avec annotations manuelles sur 502 coupes sagittales (140 avec T1/T2 sagittal, 122 avec T2 axial). Un pipeline intègre prétraitement, segmentation en cascade, biomarqueurs par niveau vertébral et explicabilité.

Deux contributions ont été développées : une segmentation multiclasse en cascade qui surpasse l'état de l'art sur notre cohorte et distingue œdème et hémorragie, et une analyse pronostique montrant que les biomarqueurs différenciés apportent une valeur prédictive à moyen et long terme au-delà du volume total et de la baseline clinique.

La première contribution est une cascade de deux U-Net 2D exploitant les plans sagittal et axial. Le modèle sagittal (T1, T2, STIR, masque de moelle) produit une segmentation initiale ; ses prédictions sont reportées en axial et servent d'entrée au second modèle. Un terme de perte topologique impose la relation hémorragie entourée d'œdème. Comparée à SciSeg (Spinal Cord Toolbox), notre approche obtient une amélioration médiane du Dice de 0,250 en axial (0,540 vs 0,268, $P < 0,001$) et 0,054 en sagittal (0,575 vs 0,461, $P < 0,001$), validation 5-fold. L'œdème est bien segmenté (Dice 0,557), l'hémorragie limitée (Dice 0,135) par le déséquilibre de classes.

La deuxième contribution quantifie les biomarqueurs pronostiques par niveau vertébral (aire, volume, ratio hémorragie/lésion) et évalue leur valeur prédictive pour l'évolution du score ASIA. Modèles emboîtés : baseline clinique (M1), + volume total (M2), + biomarqueurs différenciés (M3), comparés à plusieurs temps dans 39 patients avec suivi longitudinal. Dès 3 mois post-opératoires, les biomarqueurs différenciés améliorent significativement la prédiction de la récupération neurologique par rapport au volume total seul ($P = 0,028$), avec un effet encore plus marqué à 6 et 12 mois ($P < 0,001$), tandis qu'aucun gain n'est observé en post-opératoire immédiat ($P = 0,258$). L'analyse SHAP montre un rôle des caractéristiques contextuelles pour l'hémorragie ; un prototype relie les métriques aux IRM et segmentations.

Mots-clés: segmentation, apprentissage profond, lésions médullaires, IRM multimodale, explicabilité

Multimodal and Multiclass Spinal Cord Injury Segmentation using Deep Learning : Towards Automated Quantification of Prognostic Metrics with Clinical Explainability

Bianca POPA

ABSTRACT

This study addresses automatic segmentation of the two pathological components of traumatic spinal cord lesions—edema and hemorrhage—from multimodal MRI, and their quantification for prognostic assessment. The database comprises 148 patients recruited at the Centre de recherche de l’Hôpital du Sacré-Cœur de Montréal, with manual annotations on 502 sagittal slices (140 with T1/T2 sagittal, 122 with axial T2). A pipeline integrates preprocessing, cascade segmentation, vertebral-level biomarkers and explainability.

Two main contributions were developed : a supervised multiclass cascade that significantly outperforms the state of the art on this cohort and distinguishes edema from hemorrhage, and a prognostic analysis showing that differentiated biomarkers add predictive value for medium- and long-term neurological outcome beyond total lesion volume and clinical baseline.

The first contribution is a cascade of two 2D U-Nets exploiting complementary sagittal and axial information. The sagittal model (T1, T2, STIR, cord mask) produces an initial segmentation; its predictions are resampled to axial space as an extra input channel for the second model, with a topology loss enforcing hemorrhage surrounded by edema. Compared to SciSeg (Spinal Cord Toolbox), our method achieves median Dice improvement of 0.250 in the axial plane (0.540 vs 0.268, $P < 0.001$) and 0.054 in the sagittal plane (0.575 vs 0.461, $P < 0.001$), with 5-fold cross-validation. Edema segmentation is satisfactory (Dice 0.557); hemorrhage is limited (Dice 0.135) due to class imbalance and rarity.

The second contribution quantifies prognostic biomarkers per vertebral level (area, volume, hemorrhage-to-lesion ratio) and evaluates their predictive value for ASIA improvement. Nested linear models—clinical baseline (M1), plus total lesion volume (M2), plus differentiated biomarkers (M3)—were compared at several follow-up times in 39 patients with longitudinal data. From 3 months post-surgery onward, differentiated biomarkers significantly improved prediction of neurological recovery compared to total volume alone ($P = 0.028$), with even stronger effects at 6 and 12 months ($P < 0.001$), while no gain was observed at the immediate post-operative assessment ($P = 0.258$). SHAP analysis indicated contextual features for hemorrhage detection; a visualization prototype links metrics to axial MRI and segmentations.

Keywords: segmentation, deep learning, spinal cord injury, multimodal MRI, explainability

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 REVUE DE LA LITTÉRATURE	5
1.1 Lésions de la moelle épinière	5
1.1.1 Classification des lésions médullaires	6
1.2 Rôle de l'IRM dans l'évaluation des LME	9
1.2.1 IRM conventionnelle	9
1.2.2 IRM quantitative	11
1.2.3 Évolution des lésions médullaires traumatiques observées par IRM	12
1.3 Biomarqueurs d'imagerie et leur impact clinique	14
1.3.1 Longueur de la lésion	14
1.3.2 Volume de la lésion	16
1.3.3 Œdème médullaire	16
1.3.4 Hémorragie intramédullaire	16
1.3.5 Compression médullaire maximale (MSCC)	17
1.3.6 Compromission maximale du canal (MCC)	17
1.3.7 Ratio de dommage axial maximal	17
1.4 Apprentissage profond pour la segmentation	19
1.4.1 Défis de la segmentation des lésions médullaires	22
1.5 Méthodes de segmentation de la moelle et des lésions	24
CHAPITRE 2 MÉTHODOLOGIE	31
2.1 Vue d'ensemble de l'approche proposée	31
2.1.1 Objectifs et hypothèses de recherche	31
2.1.2 Architecture générale du système	32
2.2 Description et préparation des données	33
2.2.1 Cohorte de patients et protocoles d'acquisition IRM	33
2.2.2 Annotation des données et définition des classes	33
2.2.3 Transformation des annotations sagittales vers l'espace axial	34
2.2.4 Prétraitement des images multimodales	35
2.3 Segmentation multimodale et multiclasse des lésions médullaires	35
2.3.1 Configuration nnU-Net pour la segmentation multimodale	35
2.3.2 Entraînement des modèles en cascade	37
2.3.2.1 Modèle sagittal	37
2.3.2.2 Modèle axial	38
2.3.3 Fonction de perte et optimisation	38
2.3.4 Augmentation de données	41
2.4 Explicabilité du modèle de segmentation	42
2.4.1 Analyse SHAP des prédictions	42
2.4.2 Validation et interprétation des résultats SHAP	44

2.5	Quantification des métriques pronostiques	45
2.5.1	Extraction des biomarqueurs axiaux par niveau vertébral	45
2.5.2	Système de visualisation et aide à la décision clinique	47
2.6	Métriques d'évaluation et validation	47
2.6.1	Métriques de segmentation	47
2.6.2	Méthode de validation croisée	49
2.6.3	Tests statistiques et comparaison à l'état de l'art	50
CHAPITRE 3 RÉSULTATS ET DISCUSSION		53
3.1	Performance de la segmentation automatique	53
3.1.1	Résultats de la segmentation multimodale	53
3.1.1.1	Performance du modèle de segmentation	53
3.1.1.2	Impact des différentes modalités sur la segmentation	55
3.1.2	Analyse de la segmentation multiclasse	57
3.1.2.1	Performance par classe (œdème vs hémorragie)	57
3.1.2.2	Analyse des cas difficiles	58
3.1.2.3	Comparaison avec l'évaluation manuelle	63
3.2	Explicabilité des résultats de segmentation	66
3.2.1	Analyse des valeurs de Shapley et attribution des contributions	67
3.2.2	Cartes d'activation et importance des modalités	67
3.2.3	Interprétation des décisions du modèle	69
3.3	Quantification des métriques pronostiques et validation clinique	70
3.3.1	Caractérisation des biomarqueurs axiaux par niveau vertébral	70
3.3.2	Prédiction du pronostic neurologique	72
3.3.3	Système de visualisation et aide à la décision clinique	75
3.4	Limitations et perspectives	77
3.4.1	Limitations de la segmentation automatique	77
3.4.2	Limitations de la quantification et de l'explicabilité	78
3.4.3	Perspectives d'amélioration	79
CONCLUSION ET RECOMMANDATIONS		81
ANNEXE I	DÉTAILS TECHNIQUES DU PRÉTRAITEMENT DES DONNÉES	83
ANNEXE II	CONFIGURATION DÉTAILLÉE DE L'AUGMENTATION DE DONNÉES	87
ANNEXE III	DÉTAILS TECHNIQUES DE L'ANALYSE SHAP	89
ANNEXE IV	RÉSULTATS COMPLÉMENTAIRES DE SEGMENTATION	93
ANNEXE V	CAS DIFFICILES SUPPLÉMENTAIRES	99
ANNEXE VI	ANALYSE STATISTIQUE DÉTAILLÉE : VALEUR PRONOSTIQUE DES BIOMARQUEURS LÉSIONNELS	105

ANNEXE VII SCRIPTS ET REPRODUCTIBILITÉ111
BIBLIOGRAPHIE115

LISTE DES TABLEAUX

		Page
Tableau 1.1	Comparaison des architectures de segmentation	23
Tableau 2.1	Résumé des hyperparamètres du modèle sagittal	37
Tableau 2.2	Résumé des hyperparamètres du modèle axial	38
Tableau 3.1	Comparaison des performances entre notre modèle et SciSeg pour les métriques Dice et coefficient de Jaccard dans les plans axial et sagittal	53
Tableau 3.2	Performances de segmentation selon les configurations de modalités d'imagerie	55
Tableau 3.3	Performances de segmentation par classe pour le modèle en cascade (résultats finaux axiaux)	57
Tableau 3.4	Statistiques descriptives des biomarqueurs axiaux par niveau vertébral pour l'ensemble de la cohorte (n = 113 patients, 746 observations patient-niveau)	70
Tableau 3.5	Distribution des biomarqueurs par niveau vertébral (statistiques calculées sur toutes les observations, incluant les valeurs nulles)	71
Tableau 3.6	Comparaison des modèles de régression pour la prédiction de l'amélioration AIS (delta = score à la période – AIS préopératoire) selon la période de suivi	73

LISTE DES FIGURES

		Page
Figure 1.1	Innervation normale de la moelle épinière. Schéma montrant les principaux segments médullaires et leurs fonctions correspondantes. Les nerfs spinaux émergent de chaque segment pour innerver des territoires cutanés (dermatomes) et des groupes musculaires spécifiques	8
Figure 1.2	IRM des lésions médullaires traumatiques. (A) IRM cervicale, coupe sagittale T2, montrant une contusion médullaire hémorragique au niveau C3-C4. Tirée de Daghfous (2018). (B) IRM cervicale, coupe sagittale T2, montrant une contusion médullaire œdémateuse au niveau C4-C5. Tirée de Lamrani (2017). (C) IRM cervicale, coupe sagittale T2, montrant une moelle épinière saine pour comparaison	10
Figure 1.3	Évolution d’une LMT cervicale traumatique. Les IRM en T2 (plans sagittal et axial) illustrent l’évolution d’une lésion médullaire chez une patiente de 63 ans (incomplète, AIS D) en (A) phase aiguë (1 jour post-LMT), (B) phase subaiguë (1 mois post-LMT) et (C) phase chronique (~24 mois post-LMT). Tirée de Seif, Wheeler-Kingshott, Cohen-Adad, Flanders & Freund (2019).	13
Figure 1.4	Évaluation de la longueur de la lésion médullaire. Gauche : IRM postopératoire pondérée en T2 montrant une lésion intramédullaire cervicale. Droite : Schéma illustrant les principaux paramètres lésionnels mesurables sur l’IRM, notamment la longueur de la lésion (LL), ainsi que d’autres indices comme l’aire de la lésion (LA), sa largeur (LW) ou la présence de ponts tissulaires épargnés dans les plans sagittaux dorsal (DB) et ventral (VB)	15
Figure 1.5	Architecture U-Net pour la segmentation des images biomédicales	21
Figure 1.6	Processus d’entraînement en trois phases du modèle SCIseg pour la segmentation automatique des lésions médullaires	27
Figure 2.1	Architecture en cascade sagittal → axial avec entrées T1/T2	32
Figure 2.2	Exemple d’annotation manuelle dans ITK–SNAP (T2 sagittal) : œdème (vert) et hémorragie (rouge)	34
Figure 2.3	Schéma U-Net (encodeur–goulot–décodeur avec connexions de saut) ...	36

Figure 2.4	Hémorragie centrale entourée par l'œdème (illustration du principe topologique)	40
Figure 2.5	Schéma du calcul des valeurs SHAP (DeepExplainer) : prédiction $f(\text{test})$, baseline $\mathbb{E}[f(\text{ref})]$, et cartes SHAP pour l'œdème et l'hémorragie	44
Figure 2.6	Système de visualisation axiale des lésions : coupe T2 avec contours d'œdème (Label 1) et d'hémorragie (Label 2), et niveau vertébral indiqué	48
Figure 3.1	Cas 1 - Images sagittales montrant la résolution réduite (384×384), les artéfacts d'imagerie (lignes verticales) et la petite taille de la lésion œdémateuse par rapport à la moelle épinière	60
Figure 3.2	Cas 25 - Images sagittales illustrant la résolution réduite (384×384) et la petite taille de la lésion œdémateuse	60
Figure 3.3	Cas 89 - Images sagittales montrant un cadre d'image incomplet avec la lésion située dans la partie inférieure de la moelle épinière, limitant le contexte de moelle saine visible en dessous de la lésion	61
Figure 3.4	Cas 136 - Images sagittales montrant une résolution extrêmement faible (240×320×15) et des artéfacts d'imagerie importants, rendant l'image très pixélisée	62
Figure 3.5	Cas 136 - Images axiales montrant un décalage du masque de vérité terrain par rapport à l'emplacement réel de la moelle épinière, illustrant les défis de transformation pour les images de faible qualité	62
Figure 3.6	Cas 87 - Segmentation sagittale montrant la détection précise de quatre foyers d'hémorragie distincts (rouge), incluant des zones punctiformes de type pétéchial, alternant avec de l'œdème (vert)	64
Figure 3.7	Cas 87 - Segmentation axiale. Bien que la distinction œdème/hémorragie soit correcte, une petite zone hémorragique (visible en haut à droite de la moelle) n'a pas été détectée, illustrant les limites de la projection axiale	65
Figure 3.8	Cas 114 - Segmentation sagittale montrant une lésion avec hémorragie centrale. On note une sous-estimation de l'extension de l'hémorragie (rouge) et une sur-estimation correspondante de l'œdème (vert), liée au déséquilibre de classe	65

Figure 3.9	Cas 114 - Segmentation axiale confirmant la sous-estimation de la zone hémorragique centrale observée en sagittal	66
Figure 3.10	Cartes SHAP pour le patient 114, classes 1 (œdème, au milieu) et 2 (hémorragie, à droite). Pour l'œdème : valeurs positives concentrées au canal rachidien, négatives dans la moelle saine. Pour l'hémorragie : prédominance de valeurs négatives, canal rachidien négatif, valeurs positives dispersées dans la région endommagée. Échelle : 95 ^e percentile des valeurs absolues	68
Figure 3.11	Comparaison du R^2 ajusté des trois modèles de régression selon la période de suivi. Le modèle différencié M3 (biomarqueurs séparés) surpasse significativement les modèles M1 (clinique) et M2 (lésion complète) pour la prédiction du pronostic à moyen et long terme (3 mois, 6 mois, 1 an), alors que les trois modèles présentent des performances similaires en post-opératoire immédiat	74
Figure 3.12	Interface de visualisation axiale pour le patient 114, tranche 14 (niveau C5). L'image T2 axiale (centre) montre la moelle épinière avec les contours de lésion superposés : œdème en vert, hémorragie en rouge. Le panneau de droite affiche les métriques par tranche (aire d'œdème : 34,49 mm ² , aire d'hémorragie : 23,73 mm ² , aire totale : 58,22 mm ² , ratio hémorragie/lésion : 0,4076, ratio de dommage : 0,7603) et les métriques agrégées par niveau vertébral (volume œdème : 832,65 mm ³ , volume hémorragie : 101,38 mm ³ , ratio niveau : 0,1085)	76

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AIS	ASIA Impairment Scale (échelle de déficience ASIA)
ASIA	American Spinal Injury Association
CNN	Convolutional Neural Network
Dice	Coefficient de Dice (similarity coefficient)
DTI	Imagerie par tenseur de diffusion (Diffusion Tensor Imaging)
FDR	Taux de fausses découvertes (False Discovery Rate)
FLAIR	Fluid-Attenuated Inversion Recovery
IA	Intelligence Artificielle
IC	Intervalle de confiance
IoU	Intersection over Union (indice de Jaccard)
IRM	Imagerie par Résonance Magnétique
LME	Lésions de la Moelle Épineuse
LMT	Lésions Médullaires Traumatiques
MCC	Compromission Maximale du Canal
MSCC	Compression Médullaire Maximale
MTI	Imagerie par transfert de magnétisation (Magnetization Transfer Imaging)
nnU-Net	no-new U-Net
ReLU	Rectified Linear Unit
SCIM	Spinal Cord Independence Measure
SGD	Descente de gradient stochastique (Stochastic Gradient Descent)
SHAP	SHapley Additive exPlanations
STIR	Short Tau Inversion Recovery
TE	Temps d'écho
TR	Temps de répétition

LISTE DES SYMBOLES ET UNITÉS DE MESURE

α	Paramètre (déformation élastique, augmentation)
β	Coefficient de régression
γ	Paramètre de transformation gamma
σ	Paramètre (déformation élastique, bruit ou flou gaussien)
ρ	Coefficient de corrélation de Spearman
ε	Terme d'erreur (modèle de régression)
R^2	Coefficient de détermination
p, P	Valeur p (probabilité, test statistique)

INTRODUCTION

Les lésions traumatiques de la moelle épinière peuvent entraîner des déficits neurologiques permanents et altérer la qualité de vie des patients. L'évaluation de ces lésions par imagerie par résonance magnétique (IRM) permet d'établir un pronostic et de guider les décisions thérapeutiques. Les lésions médullaires traumatiques présentent une hétérogénéité pathologique, comprenant deux composantes distinctes : l'œdème, caractérisé par un hypersignal T2 et représentant l'inflammation, et l'hémorragie, caractérisée par un hyposignal T2 et indiquant une destruction tissulaire irréversible. Ces deux composantes ont des significations pronostiques différentes, l'œdème étant potentiellement réversible tandis que l'hémorragie indique des dommages sévères.

L'annotation manuelle des lésions médullaires sur les images IRM est un processus long et sujet à la variabilité inter-observateurs. De plus, l'extraction manuelle des métriques quantitatives (volumes, aires, ratios) pour chaque composante lésionnelle et par niveau vertébral est fastidieuse. Le développement d'outils automatisés de segmentation et de quantification permet de faciliter l'évaluation clinique et d'effectuer des analyses pronostiques basées sur des biomarqueurs quantitatifs.

Les approches actuelles de segmentation automatique des lésions médullaires, telles que SciSeg (SpinalCordToolbox), segmentent généralement la lésion comme une classe unique sans distinguer l'œdème de l'hémorragie. La quantification séparée de ces composantes pourrait améliorer la prédiction de l'évolution neurologique des patients. De plus, les performances de ces approches sur les lésions traumatiques restent limitées, notamment en raison de la complexité de la topographie lésionnelle et de la variabilité des caractéristiques d'acquisition IRM entre les patients.

Ce mémoire propose une approche de segmentation multiclasse permettant de distinguer l'œdème et l'hémorragie dans les lésions médullaires traumatiques, puis de quantifier leurs

caractéristiques physiques et géométriques par niveau vertébral. L'objectif est de développer un système automatisé qui segmente les deux composantes lésionnelles et fournit des biomarqueurs quantitatifs pour l'évaluation pronostique.

La première contribution de ce travail consiste en une architecture de segmentation en cascade exploitant les informations complémentaires des plans sagittal et axial. Un premier modèle U-Net 2D, entraîné sur les images sagittales avec des entrées multimodales (T1, T2, STIR), effectue une segmentation initiale dans le plan sagittal qui offre une meilleure visualisation longitudinale des lésions. Les prédictions de ce modèle sont ensuite transformées vers l'espace axial et servent de canal d'entrée supplémentaire à un deuxième modèle, qui produit la segmentation finale dans le plan axial, offrant une perspective plus directe sur l'anatomie de la moelle épinière pour l'évaluation clinique.

L'approche intègre également un terme topologique dans la fonction de perte, basé sur les connaissances médicales selon lesquelles l'hémorragie est généralement localisée au centre de la lésion et entourée par l'œdème. Cette contrainte anatomique guide le modèle à respecter la topologie pathologique attendue. L'évaluation est effectuée sur une cohorte de 148 patients recrutés au Centre de recherche de l'Hôpital du Sacré-Cœur de Montréal, avec comparaison à SciSeg pour situer les performances.

La deuxième contribution concerne la quantification des biomarqueurs pronostiques pour chaque composante lésionnelle par niveau vertébral. À partir des segmentations automatiques, nous extrayons des métriques incluant l'aire de section transversale maximale et moyenne, le volume, et le ratio hémorragie/lésion pour chaque niveau vertébral. Ces biomarqueurs quantitatifs, calculés sur 113 patients couvrant 10 niveaux vertébraux (C1 à T3), fournissent une caractérisation de la distribution spatiale des lésions. L'objectif est d'évaluer si l'identification et la quantification distinctes de l'œdème et de l'hémorragie — notamment la présence et le volume d'hémorragie ainsi que le ratio hémorragie/lésion — améliorent significativement la prédiction de l'évolution

neurologique par rapport à une approche de lésion complète, au-delà des variables cliniques seules.

Pour faciliter l'interprétation clinique de ces biomarqueurs, nous avons développé un prototype de visualisation permettant de lier les métriques calculées aux images IRM axiales et aux segmentations. Ce prototype offre une interface de navigation entre les tranches axiales avec affichage simultané des métriques par tranche et par niveau vertébral.

La troisième contribution est l'analyse de l'explicabilité du modèle de segmentation à l'aide de l'approche SHAP (SHapley Additive exPlanations). Cette analyse permet d'identifier quelles régions de l'image contribuent le plus aux prédictions d'œdème et d'hémorragie, validant que le modèle se base sur des caractéristiques anatomiquement cohérentes et identifiant d'éventuels biais dans les prédictions.

Ce mémoire est organisé en trois chapitres principaux. Le premier chapitre présente la revue de littérature, couvrant l'anatomie de la moelle épinière et des lésions médullaires traumatiques, les paramètres IRM pertinents, les facteurs pronostiques, ainsi que les approches existantes de segmentation automatique et les architectures d'apprentissage profond utilisées en imagerie médicale.

Le deuxième chapitre détaille la méthodologie développée. Il présente la préparation des données (cohorte de 148 patients, annotation manuelle de 502 coupes sagittales, transformation des annotations vers l'espace axial, prétraitement multimodale), l'architecture des modèles de segmentation (configuration nnU-Net, entraînement en cascade avec modalités T1, T2, STIR, fonction de perte combinant Dice, entropie croisée et terme topologique), l'analyse d'explicabilité SHAP, la quantification des biomarqueurs pronostiques par niveau vertébral, et les métriques d'évaluation utilisées (validation croisée 5-fold, comparaison statistique avec SciSeg).

Le troisième chapitre présente les résultats et la discussion. Il rapporte les performances de segmentation comparées à SciSeg, l'analyse des résultats par modalité et par classe (œdème vs hémorragie), l'analyse détaillée des cas difficiles et des cas exemplaires, l'analyse d'explicabilité SHAP, la quantification des biomarqueurs par niveau vertébral, et le système de visualisation développé. La section de limitations et perspectives discute les contraintes de l'approche et les axes d'amélioration futurs.

CHAPITRE 1

REVUE DE LA LITTÉRATURE

1.1 Lésions de la moelle épinière

Les lésions traumatiques de la moelle épinière (LMT) surviennent à la suite d'un impact physique externe, comme un accident de voiture, une chute ou une blessure sportive. Ces lésions affectent plus fréquemment les hommes (79,8 %) que les femmes (20,2 %) (Ahuja *et al.*, 2017). Le profil d'âge des personnes touchées par une lésion médullaire traumatique suit une distribution bimodale : une première concentration entre 15 et 29 ans, et une seconde, plus petite mais en augmentation, chez les individus de plus de 50 ans. Les lésions surviennent le plus souvent au niveau de la colonne cervicale (60 %), suivie par les régions thoracique (32 %) et lombo-sacrée (9 %). Les accidents de la route sont la principale cause des lésions traumatiques en Amérique du Nord, représentant 38 % des blessures entre 2010 et 2014, bien que ce chiffre soit en déclin. Les chutes viennent en deuxième position avec 31 % des blessures, suivies des blessures sportives (10–17 %).

Les phases de lésion traumatique comprennent plusieurs stades : la phase aiguë (moins de 48 heures après la blessure), caractérisée par une inflammation et une hémorragie ; la phase subaiguë (de 48 heures à 14 jours), marquée par une réponse inflammatoire continue et le début de la cicatrisation ; la phase intermédiaire (de 14 jours à 6 mois) ; et la phase chronique (au-delà de 6 mois), où des modifications structurelles permanentes se produisent, telles que la formation de cavités syringomyéliques et de tissus cicatriciels (Ahuja *et al.*, 2017). Ces différentes phases de lésion soulignent l'importance des soins rapides et continus pour minimiser les dommages à long terme.

Les lésions non-traumatiques de la moelle épinière peuvent résulter de diverses conditions médicales, telles que les infections, les tumeurs, les maladies dégénératives et les troubles vasculaires. Ces lésions diffèrent par leur étiologie mais peuvent également entraîner une dysfonction neurologique significative. Les principales causes incluent les infections, par exemple,

la myélite transverse, une inflammation de la moelle épinière souvent due à une infection virale ou bactérienne ; les tumeurs, comme les tumeurs intramédullaires ou extramédullaires qui compressent la moelle épinière ; les maladies dégénératives, comme la sclérose en plaques, qui provoque la démyélinisation des fibres nerveuses ; et les troubles vasculaires, incluant les infarctus médullaires causés par une interruption de l'apport sanguin.

Les conséquences des lésions médullaires, qu'elles soient traumatiques ou non-traumatiques, incluent la paralysie, soit la perte de la fonction motrice en dessous du niveau de la lésion ; la perte sensorielle, soit la diminution ou absence de sensation en dessous du niveau de la lésion ; et la dysfonction autonome, qui a un impact sur les fonctions corporelles régulées par le système nerveux autonome, telles que la respiration, la digestion et la régulation de la température. En plus des impacts physiques, les lésions médullaires ont des conséquences psychologiques et socio-économiques importantes. Les patients peuvent éprouver des difficultés à reprendre une vie indépendante et active, ce qui entraîne souvent des coûts élevés pour les soins et le soutien à long terme (Ahuja *et al.*, 2017).

1.1.1 Classification des lésions médullaires

Pour évaluer et classifier la gravité des lésions médullaires, l'American Spinal Injury Association (ASIA) a développé l'échelle ASIA Impairment Scale (AIS), largement utilisée dans le monde entier. Cette échelle, basée sur un examen neurologique standardisé, permet de déterminer le niveau et l'étendue de la lésion médullaire (Kirshblum *et al.*, 2011).

Afin de mieux comprendre les principes de cette classification, il est essentiel de connaître l'organisation segmentaire de la moelle épinière et les fonctions associées à chaque niveau médullaire (Figure 1.1). La moelle épinière est divisée en segments cervicaux (C1 à C8), thoraciques (T1 à T12), lombaires (L1 à L5), sacrés (S1 à S5) et un segment coccygien. Chaque segment innerve des territoires cutanés (dermatomes) et des groupes musculaires spécifiques par l'intermédiaire des nerfs spinaux. Les nerfs cervicaux contrôlent les mouvements de la tête, du cou, des bras et des mains. Les nerfs thoraciques innervent le tronc et l'abdomen, tandis que les

nerfs lombaires et sacrés contrôlent les jambes, les pieds et les fonctions génito-sphinctériennes. Une lésion médullaire interrompt les voies motrices et sensibles en dessous du niveau lésionnel, causant une paralysie et une perte de sensibilité plus ou moins étendue selon le niveau et la sévérité de l'atteinte (Cavadini, 2019).

Le score ASIA classe les lésions médullaires en cinq catégories :

Grade A (Complète) : Aucune fonction motrice ou sensorielle préservée dans les segments sacrés S4-S5.

Grade B (Incomplète sensorielle) : Fonction sensorielle préservée en dessous du niveau neurologique, y compris les segments sacrés S4-S5, mais aucune fonction motrice.

Grade C (Incomplète motrice) : Fonction motrice préservée en dessous du niveau neurologique, avec plus de la moitié des muscles clés ayant un grade inférieur à 3/5.

Grade D (Incomplète motrice) : Fonction motrice préservée en dessous du niveau neurologique, avec au moins la moitié des muscles clés ayant un grade supérieur ou égal à 3/5.

Grade E (Normal) : Fonctions motrices et sensorielles normales.

Le score ASIA permet d'orienter les décisions thérapeutiques et d'anticiper le degré de récupération. Les patients classés AIS A présentent généralement une récupération fonctionnelle limitée, tandis que ceux classés AIS B ou C ont un potentiel de récupération plus élevé, notamment pour la fonction motrice en dessous du niveau lésionnel (Roberts, Leonard & Cepela, 2017).

L'évaluation ASIA initiale est effectuée dès que l'état du patient le permet, idéalement dans les 72 heures post-traumatiques. Des réévaluations sont ensuite réalisées à 3, 6 et 12 mois pour suivre l'évolution neurologique. Ces données, associées aux résultats de l'IRM, contribuent à affiner le pronostic et à ajuster les stratégies de réadaptation.

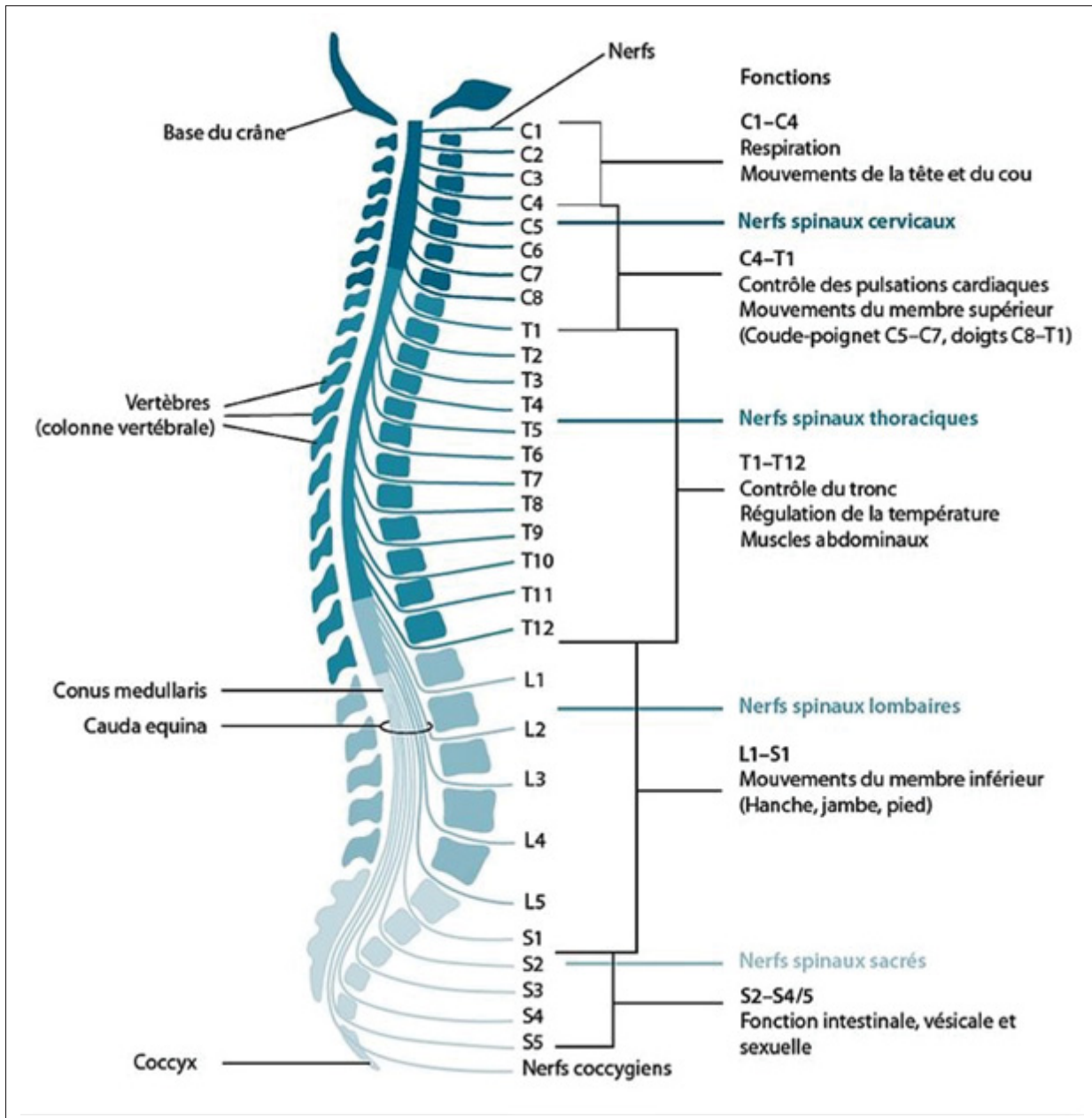


Figure 1.1 Innervation normale de la moelle épinière. Schéma montrant les principaux segments médullaires et leurs fonctions correspondantes. Les nerfs spinaux émergent de chaque segment pour innover des territoires cutanés (dermatomes) et des groupes musculaires spécifiques

1.2 Rôle de l'IRM dans l'évaluation des LME

L'IRM est l'outil de choix pour l'évaluation des lésions médullaires en raison de sa capacité à fournir des images détaillées des structures médullaires et des tissus environnants. Il existe deux types d'IRM pour la visualisation des lésions médullaires : l'IRM conventionnelle et l'IRM quantitative. Les deux jouent un rôle crucial dans le diagnostic et le pronostic des lésions médullaires traumatiques et non traumatiques.

1.2.1 IRM conventionnelle

L'IRM conventionnelle fournit des informations détaillées sur la structure de la moelle épinière après une lésion. Elle permet de visualiser la compression extrinsèque, souvent due à des fragments osseux ou des disques intervertébraux déplacés, ainsi que la perturbation du complexe discoligamentaire, qui stabilise la colonne vertébrale. De plus, l'IRM révèle les dommages intramédullaires tels que l'hémorragie, l'œdème, les cavités kystiques post-traumatiques et les ponts tissulaires entourant les kystes (David *et al.*, 2019; Freund *et al.*, 2019).

L'IRM conventionnelle repose sur l'excitation magnétique des protons d'hydrogène dans les tissus. Les séquences pondérées en T1 et T2, les plus couramment utilisées, offrent des contrastes complémentaires pour caractériser les anomalies. La séquence T1, utilisant des temps de répétition (TR) et d'écho (TE) courts, est particulièrement utile pour visualiser l'anatomie et les hémorragies aiguës, qui apparaissent comme des zones d'hyposignal. La séquence T2, avec des TR et TE plus longs, est plus sensible pour détecter les œdèmes et les inflammations, apparaissant en hypersignal. D'autres paramètres importants incluent le plan de coupe (axial, sagittal, coronal), la résolution spatiale déterminée par la taille du voxel, et le contraste ajusté par les paramètres d'acquisition. Des séquences spéciales comme T2*, STIR, ou FLAIR peuvent également être utilisées pour mettre en évidence des aspects spécifiques des lésions.

Le contraste T1 est particulièrement utile pour visualiser les hémorragies aiguës, apparaissant comme des zones d'hyposignal T1, et pour observer les détails anatomiques. En revanche, un signal normal en pondération T1 peut parfois masquer des lésions à la phase aiguë.

Le contraste T2, quant à lui, est plus sensible pour détecter les œdèmes et les inflammations. Une lésion œdémateuse ou ischémique apparaîtra comme un hypersignal T2 en raison de l'accumulation de liquide, tandis qu'une hémorragie aiguë se traduit par un hyposignal T2, en particulier dans les séquences écho de gradient T2* (sensible à la présence de désoxyhémoglobine). Les hémorragies, moins fréquentes, sont souvent associées à un pronostic plus sévère, alors que les œdèmes et les lésions ischémiques peuvent également être graves, mais sont plus précocement visibles en pondération T2 (Daghfous, 2018).

Les images suivantes illustrent deux types de lésions médullaires traumatiques identifiées par IRM, en séquence pondérée T2, permettant de distinguer les contusions hémorragiques des contusions œdémateuses.

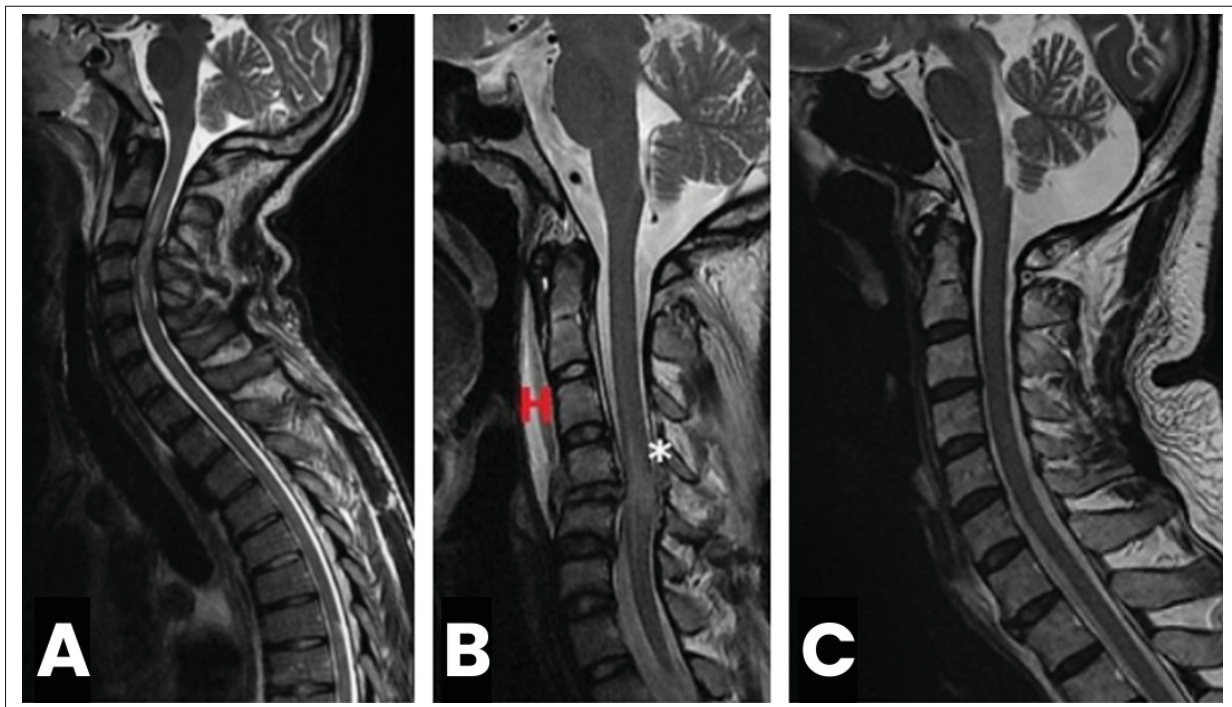


Figure 1.2 IRM des lésions médullaires traumatiques. (A) IRM cervicale, coupe sagittale T2, montrant une contusion médullaire hémorragique au niveau C3-C4. Tirée de Daghfous (2018). (B) IRM cervicale, coupe sagittale T2, montrant une contusion médullaire œdémateuse au niveau C4-C5. Tirée de Lamrani (2017). (C) IRM cervicale, coupe sagittale T2, montrant une moelle épinière saine pour comparaison

Les séquences IRM T2 sagittale et axiale au niveau de la lésion permettent de déterminer l'étendue de la lésion intramédullaire et sa longueur, ainsi que de détecter la présence d'hémorragie, d'œdème et de compression de la moelle épinière ; la vue sagittale visualise en outre les ponts tissulaires préservés dans le plan midsagittal. L'IRM T1, acquise au-dessus du niveau lésionnel et couvrant la moelle cervicale et le cerveau, permet de détecter une atrophie de ces structures. Enfin, l'IRM T2* de la moelle cervicale et lombaire, à distance du site de la lésion, détecte l'atrophie de la matière grise et blanche de la moelle épinière (Freund *et al.*, 2019).

1.2.2 IRM quantitative

L'IRM quantitative, en revanche, se concentre sur des mesures spécifiques des propriétés des tissus, telles que la densité, la diffusion et la perfusion. Ces mesures permettent une évaluation plus fine des changements microscopiques et physiopathologiques dans la moelle épinière. Les techniques quantitatives comme l'imagerie par diffusion (DTI) et l'imagerie par transfert de magnétisation (MTI) peuvent fournir des informations détaillées sur l'intégrité des fibres nerveuses et la démyélinisation. Les mesures d'IRM quantitative sont particulièrement utiles pour suivre l'évolution des lésions et pour évaluer l'efficacité des traitements, car elles peuvent détecter des changements subtils qui ne sont pas visibles sur les images conventionnelles.

Les mesures d'IRM conventionnelles et quantitatives, obtenues tôt après une lésion de la moelle épinière, sont prédictives de l'issue clinique. Cependant, dans le cadre de ce projet, nous nous concentrerons sur les données d'IRM conventionnelle pour la segmentation et la quantification des lésions. Ce choix est motivé non seulement par la disponibilité des données, mais aussi par la robustesse et la maturité des techniques d'IRM conventionnelle dans la pratique clinique. L'IRM conventionnelle permet d'obtenir des images détaillées qui sont largement utilisées et bien validées pour l'analyse des lésions médullaires, ce qui en fait une base solide pour l'application des techniques de segmentation automatique.

1.2.3 Évolution des lésions médullaires traumatiques observées par IRM

Après une lésion médullaire traumatique, les premières heures sont marquées par des changements visibles à l'IRM, notamment avec les séquences T2 et T1. Sur les images T2, prises en coupe sagittale et axiale, on observe souvent une hyperintensité—une zone qui apparaît plus lumineuse que les tissus environnants—ce qui indique la présence d'œdème et l'extension de la lésion dans la moelle épinière. Cette zone lumineuse peut s'étendre vers le haut et le bas de la moelle, en fonction de la gravité de la blessure. En parallèle, les images T1 montrent généralement une hypointensité—une zone plus sombre—qui reflète la présence d'hémorragies, un indicateur clé de la gravité immédiate de la lésion (Freund *et al.*, 2019).

Un outil diagnostique et pronostique proposé pour classifier les anomalies intramédullaires observées sur les scans pondérés en T2 est le score BASIC (Brain and Spinal Cord Injury Classification), qui identifie cinq patrons distincts d'anomalies de signal. Ces patrons varient de l'absence d'anomalies aux lésions les plus sévères, combinant œdème et hémorragie. La validité pronostique du score BASIC a été démontrée chez les patients souffrant de lésions cervicales et thoraciques aiguës, et la taille des lésions intramédullaires mesurées sur les séquences T2 est un bon prédicteur de la récupération, notamment en lien avec la gravité de la lésion et l'issue de la décompression chirurgicale (Talbot *et al.*, 2015).

Au fil du temps, les lésions médullaires traumatiques évoluent : les signes initiaux d'œdème et d'hémorragie laissent progressivement place à une extension plus diffuse de la lésion au sein de la moelle épinière. Environ un mois après la blessure, une cavité kystique post-traumatique se forme chez la plupart des patients. Ces kystes sont entourés de ponts tissulaires préservés, que l'on peut visualiser sur les images T2 midsagittales. Ces ponts tissulaires jouent un rôle important, car leur préservation peut être un indicateur potentiel de la récupération fonctionnelle à long terme (Freund *et al.*, 2019; David *et al.*, 2019).

Les lésions médullaires ne se limitent pas au site de la blessure initiale. Une atrophie progressive de la moelle épinière et du cerveau est également observable à l'IRM conventionnelle, surtout dans les séquences pondérées en T1. Ces changements sont directement liés à la gravité des

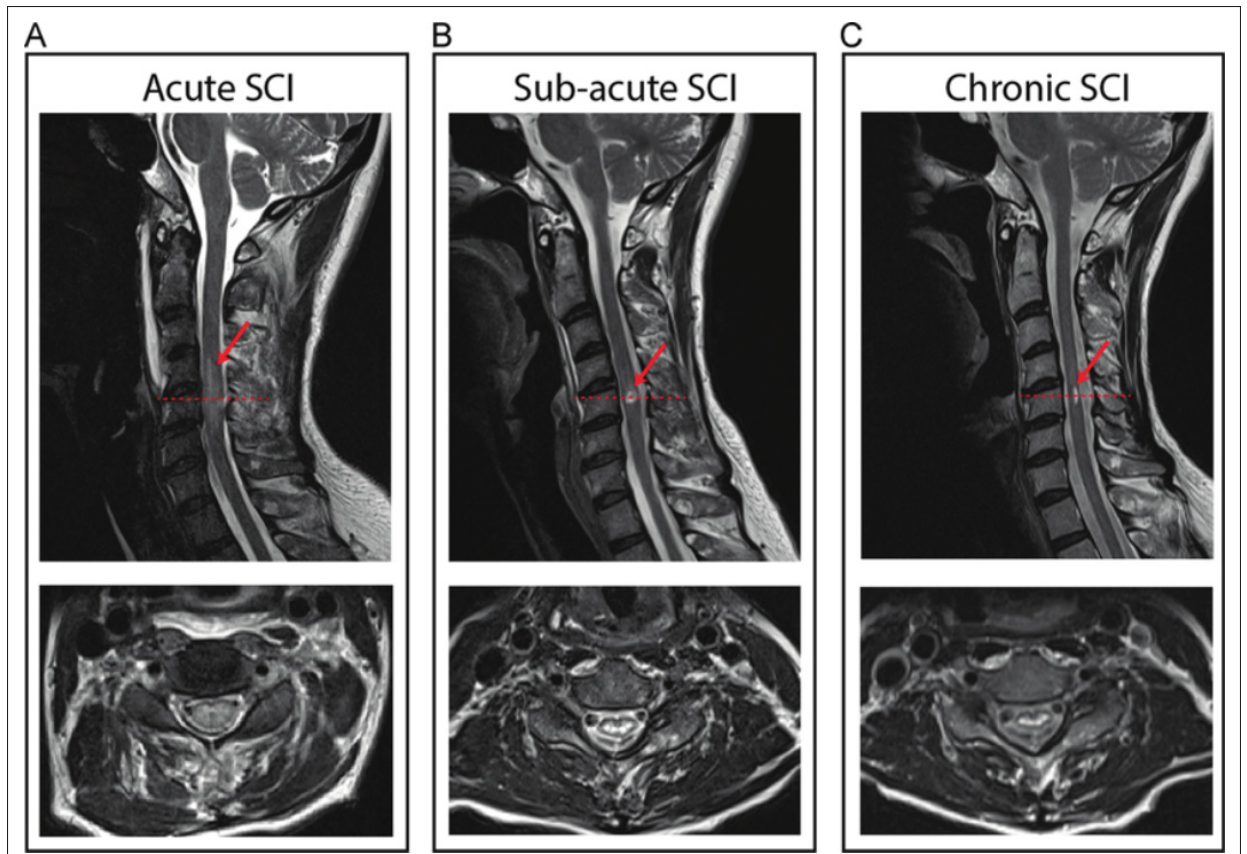


Figure 1.3 Évolution d'une LMT cervicale traumatique. Les IRM en T2 (plans sagittal et axial) illustrent l'évolution d'une lésion médullaire chez une patiente de 63 ans (incomplète, AIS D) en (A) phase aiguë (1 jour post-LMT), (B) phase subaiguë (1 mois post-LMT) et (C) phase chronique (~24 mois post-LMT). Tirée de Seif *et al.* (2019).

symptômes moteurs (comme la faiblesse musculaire) et sensoriels (comme la perte de sensation) chez les patients. L'atrophie du cerveau commence généralement dans les premiers mois après la lésion et peut se poursuivre pendant au moins deux ans, touchant des régions importantes comme le cortex moteur primaire et le thalamus (Freund *et al.*, 2019; David *et al.*, 2019).

L'utilisation de l'IRM conventionnelle, en particulier les séquences pondérées en T1 et T2, est un outil essentiel dans le diagnostic et le pronostic des LMT. Suivre l'évolution de l'œdème, de l'hémorragie et de la nécrose, ainsi que l'apparition de kystes post-traumatiques, permet d'affiner l'évaluation clinique et de mieux planifier le traitement. La combinaison de différentes

séquences IRM améliore la précision de l'évaluation et offre une visualisation plus détaillée des structures et des lésions de la moelle épinière, ce qui est déterminant pour le pronostic et les décisions thérapeutiques.

1.3 Biomarqueurs d'imagerie et leur impact clinique

Les biomarqueurs d'imagerie obtenus à partir des séquences IRM pondérées en T1 et T2 jouent un rôle essentiel dans la prédiction des résultats cliniques pour les patients atteints de lésions médullaires. Plusieurs études ont mis en évidence l'importance de ces biomarqueurs et leur corrélation avec les issues neurologiques (Miyanji, Furlan, Aarabi, Arnold & Fehlings, 2007; Gupta *et al.*, 2014; Rutges *et al.*, 2017; Martineau, Goulet, Richard-Denis & Mac-Thiong, 2019; Dobran *et al.*, 2023). Il est important de noter que ces biomarqueurs d'imagerie sont généralement utilisés en complément de l'échelle de déficience ASIA (AIS), qui reste le critère clinique le plus important dans le pronostic des lésions spinales.

1.3.1 Longueur de la lésion

La longueur de la lésion médullaire est un biomarqueur clé. Une analyse de régression linéaire a démontré que les patients présentant une lésion plus courte ont de meilleurs scores au questionnaire SCIM (Spinal Cord Independence Measure) un an après la blessure (Dobran *et al.*, 2023). Gupta *et al.* (2014) ont également montré que la longueur de la lésion était significativement plus importante chez les patients avec des lésions médullaires complètes par rapport à ceux avec des lésions incomplètes. Cela indique que les lésions plus longues sont associées à des issues neurologiques plus sévères.

En effet, la longueur de la lésion préopératoire (IMLL pour IntraMedullary Lesion Length) est un prédicteur significatif des résultats à long terme. Une étude de Martineau *et al.* (2019) a confirmé que l'IMLL, mesurée sur des images sagittales pondérées en T2, peut contribuer à établir des pronostics pour les patients atteints de lésions traumatiques de la moelle épinière.

Comme illustré dans la Figure 1.4 (gauche), la longueur de la lésion se mesure sur l'IRM en séquence pondérée T2, en incluant le site de nécrose centrale, l'hémorragie et l'œdème périlésionnel. Cette longueur est mesurée en identifiant les tranches les plus crâniennes et caudales où la lésion est visible, puis en additionnant les épaisseurs de coupe entre ces extrémités, en tenant compte de l'angle de la moelle épinière (De Leener *et al.*, 2017).



Figure 1.4 Évaluation de la longueur de la lésion médullaire. Gauche : IRM postopératoire pondérée en T2 montrant une lésion intramédullaire cervicale. Droite : Schéma illustrant les principaux paramètres lésionnels mesurables sur l'IRM, notamment la longueur de la lésion (LL), ainsi que d'autres indices comme l'aire de la lésion (LA), sa largeur (LW) ou la présence de ponts tissulaires épargnés dans les plans sagittaux dorsal (DB) et ventral (VB)

1.3.2 Volume de la lésion

Bien que moins souvent mesuré, le volume de la lésion est un biomarqueur important, souvent corrélé à la gravité des dommages et aux résultats neurologiques. Le volume d'une lésion est mesuré en additionnant tous les voxels de la zone touchée, puis en multipliant cette somme par les dimensions des voxels ($dx * dy * dz$) (De Leener *et al.*, 2017). L'étude de McCoy *et al.* (2019) a montré que les volumes de lésion, calculés par des réseaux de neurones convolutifs (CNN) pour la segmentation automatique des lésions sur des images IRM T2, étaient significativement associés aux scores moteurs des patients à l'admission et à la sortie. Cette association renforce l'importance du volume de la lésion comme indicateur pronostique.

1.3.3 Œdème médullaire

L'œdème médullaire est un autre biomarqueur important. Gupta *et al.* (2014) ont trouvé que l'œdème était un indicateur de lésions incomplètes, avec un potentiel de récupération neurologique. L'œdème reflète des dommages qui peuvent être, dans une certaine mesure, réversibles, et est souvent associé à des issues plus favorables par rapport à d'autres types de dommages comme l'hémorragie. Dans l'étude de Gupta *et al.* (2014), les patients ont été évalués pour les déficits neurologiques en utilisant le système de notation ASIA (American Spinal Injury Association), qui évalue les fonctions motrices et sensorielles. L'œdème a été identifié comme un prédicteur pour le score ASIA de base, mais n'est pas nécessairement indicatif du résultat à long terme, car il peut indiquer des lésions réversibles.

1.3.4 Hémorragie intramédullaire

L'hémorragie intramédullaire est un indicateur de pronostic particulièrement défavorable. La présence d'hémorragie dans la moelle épinière indique des dommages significatifs à un niveau cellulaire, souvent irréversibles. Gupta *et al.* (2014) ont observé que l'hémorragie était le facteur qualitatif le plus significatif pour indiquer la sévérité des lésions médullaires, et était associée à des lésions complètes sans récupération neurologique. La présence d'une hémorragie

intramédullaire est également un indicateur de pronostic défavorable selon Dobran *et al.* (2023), qui souligne que l'hémorragie est liée à des dommages significatifs et à une récupération neurologique limitée.

1.3.5 Compression médullaire maximale (MSCC)

La compression médullaire maximale (MSCC) a été largement étudiée et montre une corrélation étroite avec les résultats neurologiques initiaux. Bien que la MSCC soit un bon prédicteur du statut neurologique de base, elle n'est pas toujours un indicateur fiable pour le pronostic à long terme, car la compression peut souvent être atténuée par des interventions chirurgicales ou d'autres traitements (Gupta *et al.*, 2014). Dobran *et al.* (2023) ont également montré que la MSCC est significativement associée aux résultats neurologiques, mais que cette corrélation est plus forte pour le statut initial que pour le pronostic à long terme.

1.3.6 Compromission maximale du canal (MCC)

La compromission maximale du canal (MCC) est un autre facteur qualitatif souvent étudié en relation avec les résultats neurologiques initiaux. Bien que la MCC puisse corrélérer avec le statut neurologique de base, elle n'est pas nécessairement utile pour prédire le pronostic à long terme. Gupta *et al.* (2014) ont confirmé cette observation, indiquant que bien que des facteurs qualitatifs tels que MCC et MSCC puissent être liés au statut neurologique initial, ils ne sont pas des prédicteurs fiables pour le pronostic à long terme. Cela s'explique par le fait que la compression aiguë du canal peut souvent être soulagée par des interventions médicales, permettant ainsi une certaine récupération neurologique.

1.3.7 Ratio de dommage axial maximal

Le ratio de dommage axial maximal est un biomarqueur essentiel dans l'évaluation des lésions médullaires, indiquant la proportion de la surface de la moelle épinière endommagée sur une coupe axiale. Ce ratio se mesure en identifiant la coupe où la lésion est la plus étendue sur les

séquences IRM T2, puis en calculant la surface de la lésion par rapport à la surface totale de la moelle (De Leener *et al.*, 2017). Un ratio élevé est associé à une atteinte neurologique plus sévère et à un potentiel de récupération fonctionnelle limité, car il reflète la destruction des tissus dans les voies motrices et sensorielles (Smith *et al.*, 2021).

Il est important de noter que l'utilisation de ces biomarqueurs d'imagerie en combinaison avec l'évaluation clinique initiale (score AIS) peut améliorer la précision des prédictions pronostiques. Par exemple, Martineau *et al.* (2019) ont trouvé qu'en utilisant la présence d'hémorragie intramédullaire et le grade AIS initial, ils pouvaient prédire le score moteur AMS avec un coefficient de détermination R-carré de 0,662, ce qui représente une amélioration par rapport à l'utilisation du grade AIS seul (R-carré de 0,636).

De même, la prédiction de la probabilité qu'un patient atteigne un grade AIS de D ou E lors du suivi a été améliorée en combinant la présence d'hémorragie intramédullaire avec le grade AIS initial (index-c de 0,903 contre 0,873 pour le grade AIS seul). Cependant, il est important de noter que l'ajout de critères IRM n'améliore pas toujours les prédictions. Par exemple, dans la prédiction de l'amélioration d'au moins un grade de l'AIS, l'utilisation des critères IRM a légèrement diminué la performance du modèle par rapport à l'utilisation du grade AIS initial seul.

Actuellement, la plupart des études emploient une quantification manuelle de ces biomarqueurs dérivés de l'IRM, ce qui est une tâche subjective et fastidieuse. D'où l'importance de développer des méthodes de quantification automatiques de ces métriques. Des outils automatisés permettent une évaluation plus rapide et plus cohérente des biomarqueurs, améliorant ainsi la précision du pronostic et facilitant les décisions thérapeutiques.

Ces biomarqueurs fournissent des indications précieuses pour le pronostic des patients atteints de lésions médullaires. La longueur de la lésion, le volume de la lésion, l'œdème, l'hémorragie intramédullaire et la compression médullaire maximale peuvent être facilement évalués par IRM, permettant ainsi aux cliniciens de prévoir les résultats et d'adapter les plans de traitement en conséquence. Cependant, le grade AIS préopératoire reste l'indicateur le plus fiable pour

prédire l'évolution des lésions médullaires. Les biomarqueurs d'imagerie doivent être considérés comme des outils complémentaires plutôt que comme des substituts à l'évaluation clinique.

1.4 Apprentissage profond pour la segmentation

L'apprentissage profond a significativement amélioré les capacités de segmentation des images médicales grâce à sa capacité à apprendre des représentations complexes à partir des données brutes. Les réseaux de neurones profonds, en particulier les réseaux de neurones convolutionnels (CNN), sont les principaux outils utilisés dans ce domaine.

Les CNN sont particulièrement adaptés au traitement des images car ils exploitent la structure spatiale inhérente à ces données. Les images sont composées de pixels organisés selon une grille régulière, et les CNN tirent parti de cette structure en appliquant des filtres de convolution à différentes échelles. Ces filtres balaient l'image et apprennent automatiquement à détecter des caractéristiques pertinentes à différents niveaux d'abstraction.

Lors des premières couches d'un CNN, les filtres de convolution capturent des caractéristiques de bas niveau telles que les bords, les coins et les textures simples. Au fur et à mesure que l'on progresse dans les couches plus profondes du réseau, les filtres combinent ces caractéristiques de bas niveau pour former des représentations plus complexes et abstraites, capables de détecter des objets entiers ou des parties d'objets.

Cette approche hiérarchique permet aux CNN d'apprendre efficacement des représentations riches et informatives directement à partir des images brutes. Contrairement aux méthodes traditionnelles, les CNN n'ont pas besoin que les caractéristiques pertinentes soient extraites manuellement par des experts humains. C'est cette capacité à apprendre automatiquement des caractéristiques à différentes échelles qui rend les CNN si puissants pour l'analyse d'images, y compris les tâches de segmentation (Lombaert, 2023).

Les CNN utilisés pour les tâches de segmentation sont composés de couches de convolution et de pooling, mais ne comportent généralement pas de couches entièrement connectées comme

on les trouve dans les architectures de classification d'images. Au lieu de cela, les CNN de segmentation utilisent des couches de convolution transposée (dé-convolution) pour générer des masques de segmentation avec des activations sigmoïdes qui produisent une carte de probabilité binaire pour chaque pixel de l'image, indiquant la présence ou l'absence de la classe d'intérêt.

L'architecture U-Net, introduite par Ronneberger, Fischer & Brox (2015), est devenue une référence pour la segmentation en imagerie médicale. Ce réseau de neurones convolutionnel se distingue par sa capacité à capturer à la fois les caractéristiques globales et les détails fins d'une image. Le U-Net tire son nom de sa forme en U, qui se compose de deux parties : un chemin contractif pour extraire les caractéristiques de l'image par convolution et pooling, et un chemin expansif pour générer les masques de segmentation via des dé-convolutions. Les connexions de saut entre ces deux parties permettent de préserver les détails tout en réintégrant les informations spatiales essentielles, améliorant ainsi la précision de la segmentation.

Cette architecture est particulièrement adaptée aux images médicales, où la précision et la capacité à traiter des structures complexes sont essentielles. Le U-Net est également efficace même avec un nombre limité d'images annotées, ce qui est souvent le cas en imagerie médicale. De plus, il permet un entraînement rapide et des prédictions en temps réel, facilitant son application clinique. La Figure 1.5 illustre l'architecture U-Net, montrant les étapes du traitement des images, de l'entrée à la production du masque de segmentation.

Isensee, Jaeger, Kohl, Petersen & Maier-Hein (2021) ont proposé nnU-Net, une méthode auto-configurante basée sur l'architecture U-Net. nnU-Net ajuste automatiquement sa configuration en fonction des caractéristiques du jeu de données et de la tâche spécifique, ce qui réduit considérablement le temps et les efforts nécessaires pour optimiser les paramètres. Cette approche a montré que les détails de la configuration du pipeline ont plus d'impact sur la performance que les variations architecturales. En effet, l'analyse du classement du défi KiTS 2019 a révélé que les modifications architecturales couramment utilisées, telles que les connexions résiduelles, les connexions denses, les mécanismes d'attention ou les convolutions dilatées, ne sont pas nécessaires pour obtenir de bonnes performances.

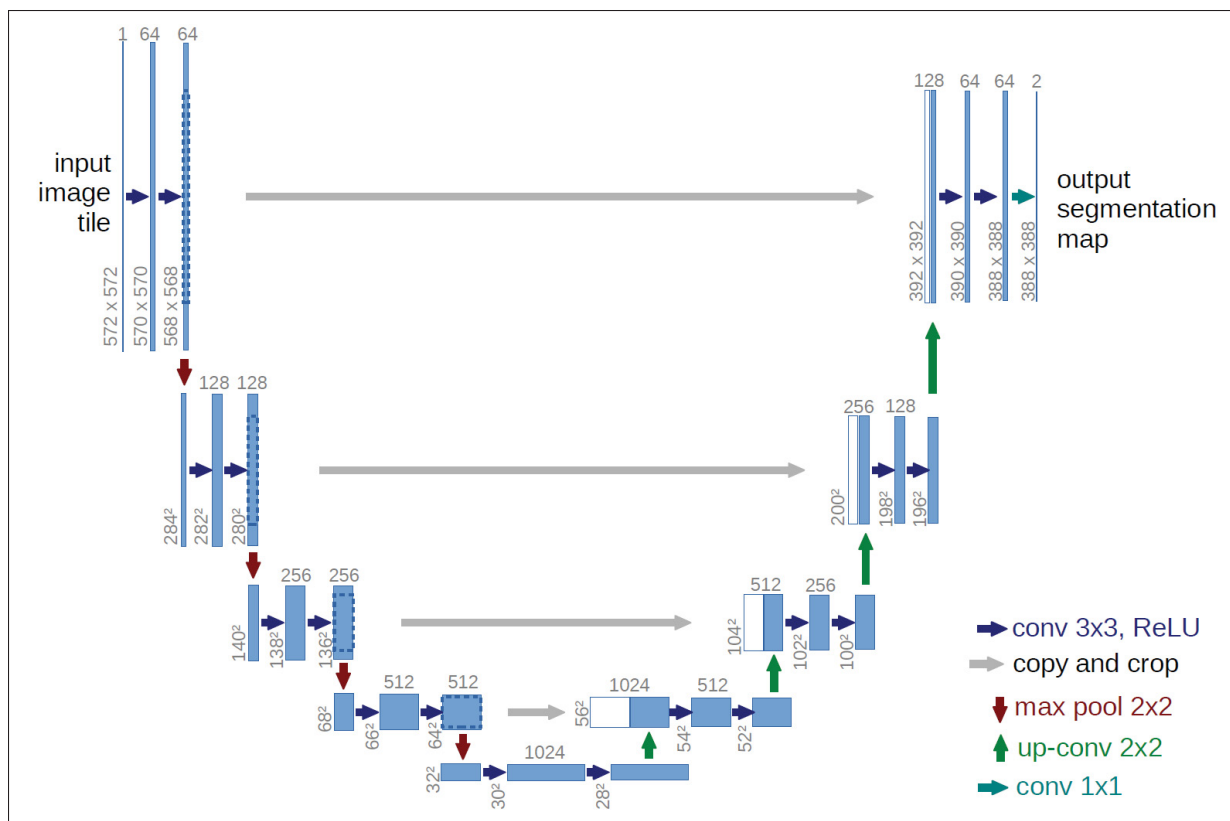


Figure 1.5 Architecture U-Net pour la segmentation des images biomédicales

nnU-Net propose trois configurations de modèles. Le U-Net 2D traite chaque coupe indépendamment, ce qui le rend adapté aux volumes de grande taille, mais au détriment du contexte inter-tranches. Le U-Net 3D opère sur l'ensemble du volume à pleine résolution pour une segmentation cohérente dans les trois dimensions, au prix d'une plus grande demande en mémoire GPU. Le U-Net 3D Cascade combine un premier modèle à résolution réduite pour une segmentation initiale et un second à pleine résolution pour l'affiner, alliant ainsi efficacité computationnelle et précision. Ces configurations permettent à nnU-Net de s'adapter automatiquement aux spécificités des données et aux contraintes matérielles.

Cependant, il est important de noter que malgré l'efficacité démontrée de nnU-Net, de nouvelles architectures continuent d'être proposées, souvent avec des revendications de performances supérieures. Une étude récente, nnU-Net Revisited : A Call for Rigorous Validation in 3D

Medical Image Segmentation (Isensee *et al.*, 2024), a mis en lumière les limites de nombreuses de ces nouvelles approches. Cette étude a montré que, malgré l'introduction de modèles basés sur les transformeurs et d'autres architectures innovantes, les CNN basés sur U-Net, tels que nnU-Net et ses variantes, continuent de produire des performances optimales. Cette supériorité est particulièrement vraie lorsque les modèles sont correctement configurés et adaptés aux ressources matérielles modernes.

L'étude d'Isensee *et al.* (2024) a également révélé que les affirmations de supériorité des nouvelles architectures sont souvent biaisées par des validations inadéquates, des ensembles de données insuffisants et des configurations matérielles inégales. En réalité, la combinaison d'un modèle U-Net bien configuré, tel que nnU-Net, avec un matériel adapté, offre toujours les meilleurs résultats en segmentation d'images médicales. Par conséquent, bien que les transformeurs puissent offrir des avantages potentiels pour capturer des relations complexes à longue distance dans les images, leur adoption en imagerie médicale reste limitée par la nécessité de grands ensembles de données annotées, qui sont souvent difficiles à obtenir.

Pour une comparaison plus claire des différentes architectures de segmentation, le tableau 1.1 résume leurs principaux avantages et inconvénients :

En résumé, les réseaux de neurones convolutionnels de type U-Net et ses variantes, telles que nnU-Net, restent la référence pour la segmentation des images médicales en raison de leur efficacité et de leur adaptabilité aux caractéristiques spécifiques des jeux de données. Les architectures basées sur les transformeurs offrent des avantages potentiels, mais leur adoption est limitée par la nécessité de grands ensembles de données annotées de haute qualité, ce qui reste un défi majeur en imagerie médicale.

1.4.1 Défis de la segmentation des lésions médullaires

Malgré les avancées significatives dans le domaine de la segmentation d'images médicales par apprentissage profond, la segmentation des lésions médullaires présente des défis uniques qui compliquent le développement de modèles efficaces. La variabilité des lésions en termes de

Tableau 1.1 Comparaison des architectures de segmentation

Architecture	Type	Avantages	Inconvénients
U-Net	CNN	Efficace même avec peu de données annotées, rapide à entraîner	Moins performant pour capturer des relations à longue distance
nnU-Net	CNN	Auto-configuration, performances de pointe, adaptabilité	Peut nécessiter un matériel plus moderne pour un fonctionnement optimal
MedNeXt	CNN + Transformer	Performances améliorées sur certaines tâches	Temps d'entraînement accru, nécessite des configurations spécifiques
SwinUNETR	Transformer	Captures de relations complexes à longue distance	Nécessite de grandes quantités de données annotées, plus difficile à entraîner
Auto3DSeg	Framework CNN/Transformer	Promet une configuration automatique comme nnU-Net	Performance encore inférieure à celle de nnU-Net sur la plupart des tâches

taille, forme et intensité, selon la phase de la blessure et son étiologie, rend difficile la création d'un modèle unique capable de segmenter efficacement tous les types de lésions. Cette difficulté est exacerbée par la rareté des données annotées, limitant la capacité des modèles à généraliser efficacement.

Le déséquilibre des classes, où les lésions occupent généralement une petite partie de l'image IRM, peut biaiser les modèles vers une sous-segmentation. Les artefacts d'imagerie, souvent causés par les implants chirurgicaux chez les patients atteints de lésions traumatiques, compliquent davantage la segmentation précise. La variabilité des protocoles d'acquisition d'IRM entre les centres médicaux ajoute une couche supplémentaire de complexité, affectant la performance des modèles de segmentation.

Les frontières floues entre les lésions et les tissus sains, ainsi que les changements temporels dans l'apparence des lésions, rendent difficile la définition précise des contours et la segmentation cohérente sur des scans longitudinaux. La structure anatomique complexe de la moelle épinière,

fine et allongée, entourée de liquide céphalo-rachidien et d'os, complique la distinction entre les lésions et les structures environnantes. Dans certains cas, particulièrement dans les phases subaiguës ou chroniques, le faible contraste entre les lésions et les tissus sains rend la segmentation automatique plus ardue.

Enfin, la variabilité inter-observateurs dans les annotations manuelles des lésions médullaires complique l'établissement d'une référence fiable pour l'entraînement et l'évaluation des modèles. Ces défis soulignent la nécessité de développer des approches de segmentation robustes et adaptatives. Les méthodes actuelles visent à surmonter ces obstacles pour fournir des segmentations précises dans diverses conditions cliniques et d'imagerie.

1.5 Méthodes de segmentation de la moelle et des lésions

Il existe divers outils basés sur l'apprentissage profond pour segmenter les structures anatomiques dans les scans IRM. Dans ce travail, nous distinguons les algorithmes de segmentation générale des outils spécialisés pour la moelle épinière et les lésions médullaires. La spécificité des modèles par rapport aux jeux de données d'entraînement est cruciale pour leur capacité à s'adapter à différentes modalités ou types de structures. Les modèles d'apprentissage automatique, lorsqu'ils sont entraînés sur des jeux de données spécifiques, peuvent rencontrer des difficultés lorsqu'ils sont appliqués à des contextes différents. Par exemple, un modèle conçu pour détecter des lésions traumatiques à un stade aigu peut ne pas fonctionner aussi bien pour segmenter des lésions chroniques ou subaiguës. De plus, les performances des modèles peuvent varier selon les modalités IRM utilisées pour l'entraînement, ce qui affecte leur capacité à traiter des scans obtenus avec des protocoles différents.

Nous introduisons d'abord MRSegmentator et TotalSegmentator MRI, des exemples avancés de modèles de segmentation générale basés sur l'architecture nnU-Net. Conçus pour segmenter une grande variété de structures anatomiques dans les scans IRM et CT, ces modèles montrent des performances robustes dans divers contextes cliniques. Leur capacité à traiter des données de différentes modalités d'imagerie et à segmenter de multiples structures avec une précision

raisonnable constitue un avantage majeur. Cependant, ils ne sont pas adaptés à des tâches spécifiques comme la segmentation des lésions médullaires. Par exemple, MRSegmentator peut segmenter 40 classes de structures, y compris la colonne vertébrale, mais n'a pas été conçu pour segmenter la moelle épinière ou les lésions. De même, TotalSegmentator MRI peut segmenter 59 structures anatomiques, y compris la moelle épinière, mais pas les lésions. Dans les deux cas, l'incorporation de scans CT dans l'entraînement a amélioré la généralisabilité des modèles, leur permettant de mieux s'adapter aux variations des modalités d'imagerie et des protocoles utilisés (D'Antonoli *et al.*, 2024; Häntze *et al.*, 2024).

Les outils de segmentation automatique de la moelle épinière et des lésions intramédullaires ont connu des avancées significatives ces dernières années. Parmi les solutions existantes, Deepseg et BASICseg, basés sur l'architecture U-Net, ont démontré des performances robustes malgré la variabilité des paramètres d'imagerie et des conditions cliniques. Deepseg, conçu pour la segmentation chez les patients atteints de sclérose en plaques (SEP), atteint un coefficient de Dice de 60 %, avec une sensibilité de 83 % et une précision de 77 % pour la détection des lésions (Gros *et al.*, 2019). BASICseg présente des performances similaires pour la segmentation des lésions médullaires traumatiques en phase aiguë (McCoy *et al.*, 2019).

Masse-Gignac, Flórez-Jiménez, Mac-Thiong & Duong (2023) ont proposé une autre approche innovante pour segmenter la moelle épinière traumatisée. Leur méthode analyse séquentiellement les scans dans les plans axial et sagittal via deux réseaux convolutifs distincts. Le processus comporte deux étapes : d'abord, un réseau traite les coupes axiales pour une segmentation préliminaire ; ensuite, un second réseau analyse les coupes sagittales, intégrant les segmentations axiales réarrangées, pour affiner la segmentation finale.

L'architecture utilisée pour les réseaux axiaux et sagittaux est basée sur le U-Net original avec plusieurs adaptations, notamment l'ajout de portes d'attention précédant la concaténation des connexions de saut. Cette méthode a obtenu des résultats significativement meilleurs que Deepseg 2D, avec un coefficient de Dice moyen de 0,95 contre 0,88 pour Deepseg 2D ($p < 0,001$). D'autres métriques comme le taux de vrais positifs, la distance de surface moyenne et la

distance de Hausdorff maximale ont également montré des résultats significativement meilleurs pour l'approche proposée. La combinaison des informations axiales et sagittales est le principal atout de cette approche, permettant une segmentation précise des moelles épinières touchées par des lésions traumatiques (Masse-Gignac *et al.*, 2023).

Néanmoins, les méthodes de segmentation décrites ci-dessus partagent une limitation : elles sont principalement conçues pour délimiter la moelle épinière, plutôt que pour identifier spécifiquement les lésions.

Pour répondre à ces limitations, de nouveaux outils ont été développés, comme SCIseg, introduit en 2024, qui se concentre sur la segmentation des lésions hyperintenses sur des images IRM pondérées en T2 chez les patients atteints de lésions médullaires traumatiques. SCIseg a été conçu pour surmonter les défis posés par la variabilité des lésions en fonction des phases de la blessure et les artefacts d'image induits par les implants chirurgicaux.

En s'appuyant sur l'architecture auto-configurante nnU-Net, SCIseg a été développé à travers trois phases distinctes. Dans la première phase, un modèle de base a été entraîné sur 142 images pondérées en T2 provenant de deux sites, utilisant une stratégie de segmentation régionale. La deuxième phase a introduit une stratégie d'apprentissage actif avec intervention humaine. Cette approche consiste à utiliser le modèle de base pour générer des prédictions sur de nouvelles images non étiquetées, puis à faire vérifier et corriger manuellement ces prédictions par des experts. Ces données corrigées sont ensuite réintégrées dans l'ensemble d'entraînement. Cette méthode permet d'améliorer progressivement les performances du modèle en le confrontant à des cas plus complexes et en bénéficiant de l'expertise humaine. Dans cette phase, 40 scans axiaux supplémentaires du site 1 ont été intégrés après correction manuelle. Enfin, la troisième phase a incorporé 14 scans sagittaux isotropes d'un troisième site, portant le total à 196 scans. Cette approche progressive, combinant des données multi-sites et multi-résolutions, visait à créer un modèle robuste capable de segmenter simultanément la moelle épinière et les lésions sur diverses images IRM, améliorant ainsi sa capacité de généralisation à différents types d'acquisitions.

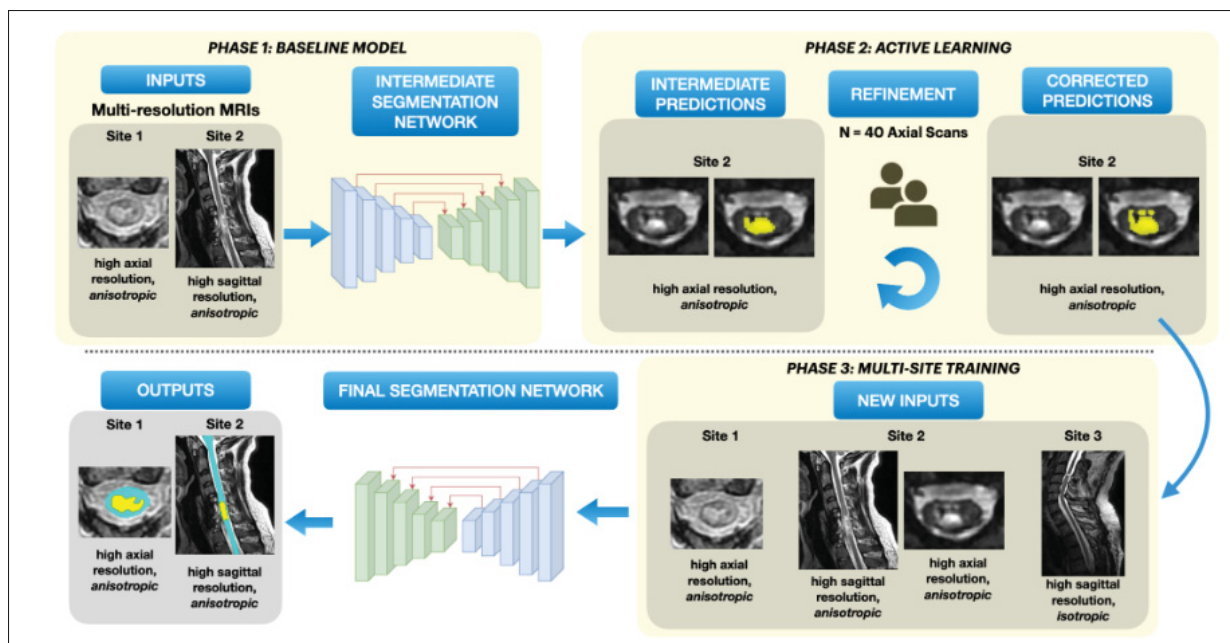


Figure 1.6 Processus d'entraînement en trois phases du modèle SCIseg pour la segmentation automatique des lésions médullaires

SCIseg a démontré de bonnes performances pour la segmentation des lésions, avec un coefficient de Dice moyen de $0,60 \pm 0,29$, atteignant $0,76 \pm 0,16$ sur les données d'un des sites. Bien que comparable à Deepseg 2D en termes de Dice moyen, SCIseg 3D se distingue par une réduction significative des erreurs de volume relatif, offrant une segmentation plus précise et fiable, avec moins de sous-segmentation. D'autres métriques comme la distance de surface moyenne ($0,72 \pm 2,19$), la valeur prédictive positive par lésion ($0,68 \pm 0,43$), la sensibilité par lésion ($0,82 \pm 0,36$) et le score F1 ($0,68 \pm 0,42$) ont également été évaluées.

De plus, SCIseg a établi une corrélation statistiquement significative entre les biomarqueurs quantitatifs des lésions et les scores cliniques, avec des performances comparables à l'annotation manuelle, confirmant sa fiabilité pour l'extraction de biomarqueurs cliniques pertinents (Enamundram *et al.*, 2024).

Cependant, les auteurs ont également reconnu certaines limitations de leur approche. Ils ont noté que le modèle pourrait bénéficier d'informations complémentaires sur les caractéristiques des

lésions provenant d'autres contrastes IRM. De plus, les lésions ont été considérées comme des entités uniques et indépendantes, ce qui peut ne pas refléter la complexité de certaines lésions multifocales. Enfin, les auteurs ont suggéré que les modèles de segmentation pourraient être améliorés en utilisant des masques de vérité terrain plus détaillés, où l'œdème et l'hémorragie seraient traités comme des classes distinctes, permettant une caractérisation plus fine des lésions.

Plus tard en 2024, SCISegV2 a étendu ces capacités en proposant une segmentation universelle des lésions intramédullaires, intégrant les différentes phases et étiologies des lésions. SCISegV2 a exploré deux stratégies d'entraînement : la segmentation hiérarchique à partir de l'image pondérée en T2 seule, et la segmentation des lésions à partir d'une entrée à 2 canaux combinant l'image pondérée en T2 et le masque de segmentation de la moelle épinière.

L'évaluation de SCISegV2 a révélé que la stratégie à 2 canaux améliorerait considérablement les performances, en particulier pour les images préopératoires aiguës. Cependant, cette dépendance aux masques de segmentation de la moelle épinière peut être un inconvénient potentiel. Malgré cela, SCISegV2 surpasse SCISeg sur la plupart des sites de test, sauf ceux contenant principalement des lésions traumatiques, pour lesquels SCISegV1 reste le plus performant.

Outre la segmentation des lésions, SCISegV2 permet de quantifier automatiquement la largeur des ponts tissulaires adjacents aux lésions, un biomarqueur clé pour évaluer la récupération fonctionnelle. Les performances de SCISegV2 ont confirmé que les ponts tissulaires quantifiés automatiquement ne différaient pas significativement de ceux mesurés manuellement, renforçant la confiance dans cet outil pour l'extraction de biomarqueurs cliniquement pertinents (Karthik *et al.*, 2024).

En conclusion, bien que les récentes avancées dans le domaine de la segmentation automatique des lésions médullaires aient conduit au développement d'outils performants comme SCISeg et SCISegV2, cette revue de littérature a mis en lumière plusieurs lacunes importantes dans l'état actuel de la recherche.

Premièrement, les méthodes actuelles, y compris SCIseg, ne permettent pas une segmentation en composantes spécifiques des LMT. Cette limitation empêche une analyse détaillée de la structure interne des lésions, qui pourrait fournir des informations cruciales pour le pronostic et le traitement. La capacité à distinguer entre la nécrose, l'hémorragie et l'œdème au sein d'une même lésion reste un défi non résolu.

Deuxièmement, la quantification manuelle des biomarqueurs souffre d'une importante variabilité inter-observateurs. Les méthodes automatiques actuelles, telles que celles de Spinal Cord Toolbox, ne sont pas suffisamment précises pour quantifier les différentes composantes des lésions médullaires traumatiques. Cette imprécision peut biaiser l'évaluation des lésions, impactant potentiellement le diagnostic et le suivi des patients. Il est nécessaire d'augmenter la précision et la granularité de ces méthodes automatiques, en visant une concordance avec l'annotation manuelle experte, ce qui permettrait simultanément de réduire la variabilité inter-observateurs.

Troisièmement, il existe un manque flagrant d'utilisation de l'IRM multimodale dans les approches actuelles de segmentation. La plupart des méthodes, dont SCIseg, se concentrent principalement sur une seule séquence IRM (généralement T2), négligeant ainsi les informations complémentaires que pourraient apporter d'autres séquences comme T1, T2* ou FLAIR. Cette sous-utilisation des données disponibles limite potentiellement la précision et la robustesse des segmentations obtenues.

Enfin, l'automatisation des processus de segmentation et d'analyse reste insuffisante pour un usage clinique routinier. Les méthodes actuelles nécessitent souvent une intervention manuelle significative, que ce soit pour la préparation des données, l'ajustement des paramètres, ou la validation des résultats. Cette dépendance à l'expertise humaine limite l'applicabilité à grande échelle de ces outils dans un contexte clinique où le temps et les ressources sont souvent limités.

Ces lacunes mettent en évidence un besoin urgent de développer des approches plus complètes, précises et automatisées pour la segmentation et l'analyse des LMT. Le chapitre suivant présentera la méthodologie que nous avons élaborée pour répondre à ces défis.

CHAPITRE 2

MÉTHODOLOGIE

2.1 Vue d'ensemble de l'approche proposée

2.1.1 Objectifs et hypothèses de recherche

Pour rappel, les deux objectifs principaux du projet sont de segmenter les deux composantes des lésions traumatiques de la moelle épinière, soit l'œdème et l'hémorragie, puis de quantifier et de visualiser, pour chacune de ces composantes, leurs caractéristiques physiques et géométriques afin d'aider à établir des pronostics plus fiables.

Cette section présente la méthodologie développée pour atteindre ces objectifs, en détaillant les étapes de préparation des données, l'architecture des modèles de segmentation, et les approches d'évaluation. L'ensemble des développements a été réalisé en Python 3.9 avec PyTorch 2.5.1. L'entraînement des modèles a été effectué sur deux cartes graphiques NVIDIA Quadro RTX 6000 (24 Go VRAM chacune) avec CUDA 12.0.

Dans ce cadre, nous formulons trois hypothèses qui guident l'ensemble de la méthodologie : (H1) qu'un modèle d'apprentissage profond segmente de façon fiable et reproductible l'œdème et l'hémorragie sur IRM, avec des performances supérieures à un baseline standard et comparables à l'accord inter-experts ; (H2) que la combinaison de modalités IRM en entrée (p. ex. T1, T2, STIR) améliore significativement la segmentation par rapport à l'utilisation d'une seule modalité, et qu'elle peut être encore renforcée par l'exploitation conjointe des plans sagittal et axial ; et (H3) que l'identification et la quantification distinctes de l'œdème et de l'hémorragie — notamment la présence et le volume d'hémorragie ainsi que le ratio hémorragie/lésion — améliorent significativement la prédiction de l'évolution neurologique (AIS, scores moteurs) par rapport à une approche de lésion complète (masque unique ou volume total), au-delà des variables cliniques seules.

2.1.2 Architecture générale du système

L'approche proposée repose sur une architecture en cascade de deux modèles 2D U-Net spécialisés. Le premier modèle, entraîné sur les images sagittales, utilise des entrées multimodales (T1, T2, STIR) pour segmenter les composantes lésionnelles dans le plan sagittal. Les prédictions de ce modèle sont ensuite transformées vers l'espace axial et servent de canal d'entrée supplémentaire au deuxième modèle, qui effectue la segmentation finale dans le plan axial.

Cette architecture en cascade permet de combiner les avantages des deux orientations d'imagerie : les vues sagittales offrent une meilleure visualisation longitudinale des lésions, tandis que les vues axiales ont été choisies pour la visualisation finale car elles offrent une perspective plus directe sur l'anatomie de la moelle épinière. L'utilisation de modalités IRM multiples (T1, T2, STIR) au niveau sagittal exploite les caractéristiques complémentaires de chaque séquence pour améliorer la précision de la segmentation.

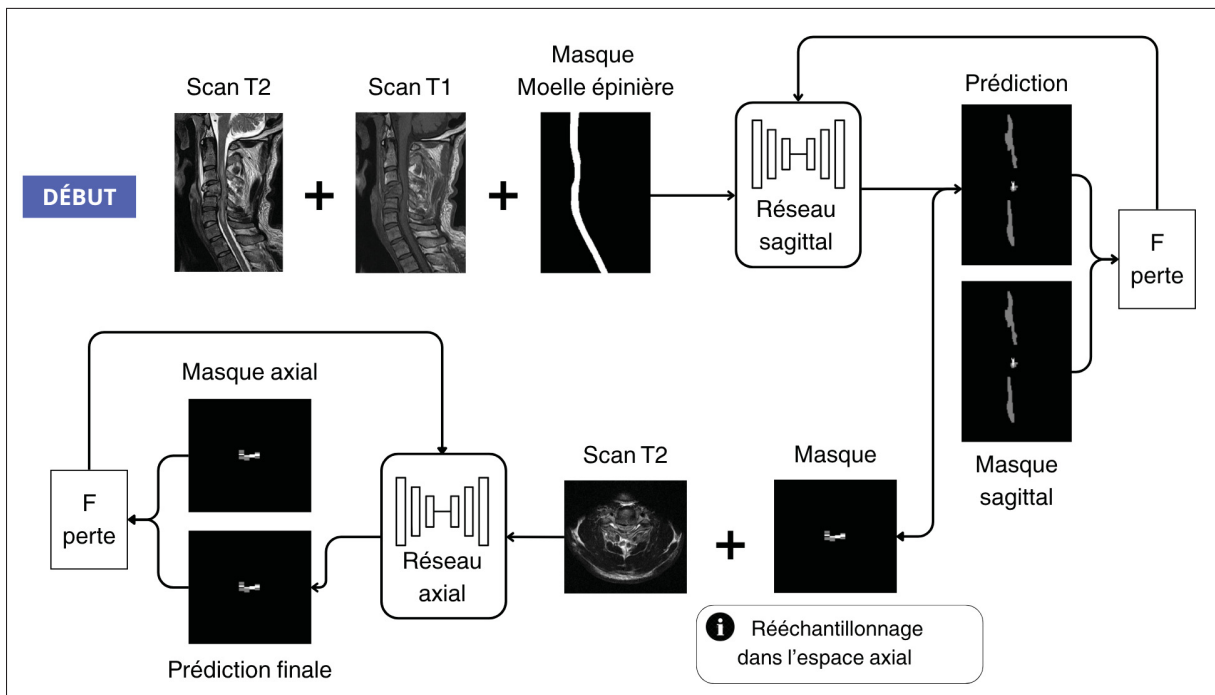


Figure 2.1 Architecture en cascade sagittale → axiale avec entrées T1/T2

2.2 Description et préparation des données

2.2.1 Cohorte de patients et protocoles d'acquisition IRM

Les données utilisées dans cette étude proviennent d'une cohorte de 148 patients ayant subi une lésion traumatique de la moelle épinière, recrutés au Centre de recherche de l'Hôpital du Sacré-Cœur de Montréal. Toutes les images sont préopératoires et présentent des résolutions et orientations hétérogènes. Pour chaque patient, un protocole d'imagerie par résonance magnétique (IRM) multi-séquences a été réalisé, comprenant des images acquises dans les plans sagittal et axial avec différentes pondérations : T1, T2, T2-star (T2*), STIR (Short Tau Inversion Recovery) et FLAIR (Fluid-Attenuated Inversion Recovery).

La disponibilité des séquences varie entre les patients, avec 148 patients disposant d'images sagittales T2, 142 patients d'images sagittales T1, 105 patients d'images sagittales STIR, 127 patients d'images axiales T2 et 16 patients d'images axiales T1. Pour l'entraînement du modèle de segmentation sur les vues sagittales, les données de 140 patients disposant à la fois des séquences T1 et T2 ont été utilisées, tandis que 122 patients disposant d'une séquence T2 axiale ont été inclus pour le modèle axial.

2.2.2 Annotation des données et définition des classes

L'annotation manuelle des lésions médullaires a été effectuée sur les images sagittales pondérées en T2 à l'aide du logiciel ITK-Snap. Le processus consistait à "peindre" manuellement, pour chaque coupe pertinente, les pixels correspondant aux régions pathologiques. Au total, 502 coupes sagittales contenant des lésions ont ainsi été annotées sur l'ensemble de la cohorte, dont 172 coupes présentaient une composante d'hémorragie.

Deux classes de tissus lésionnels ont été identifiées : la classe 1 correspond à l'œdème, caractérisé par un hypersignal (intensité de pixel élevée) par rapport au tissu médullaire sain, tandis que la classe 2 correspond à l'hémorragie, caractérisée par un hyposignal (intensité de pixel faible).

Afin de garantir la qualité et la cohérence des annotations, l'ensemble des segmentations a été révisé et validé par deux observateurs experts de manière indépendante.

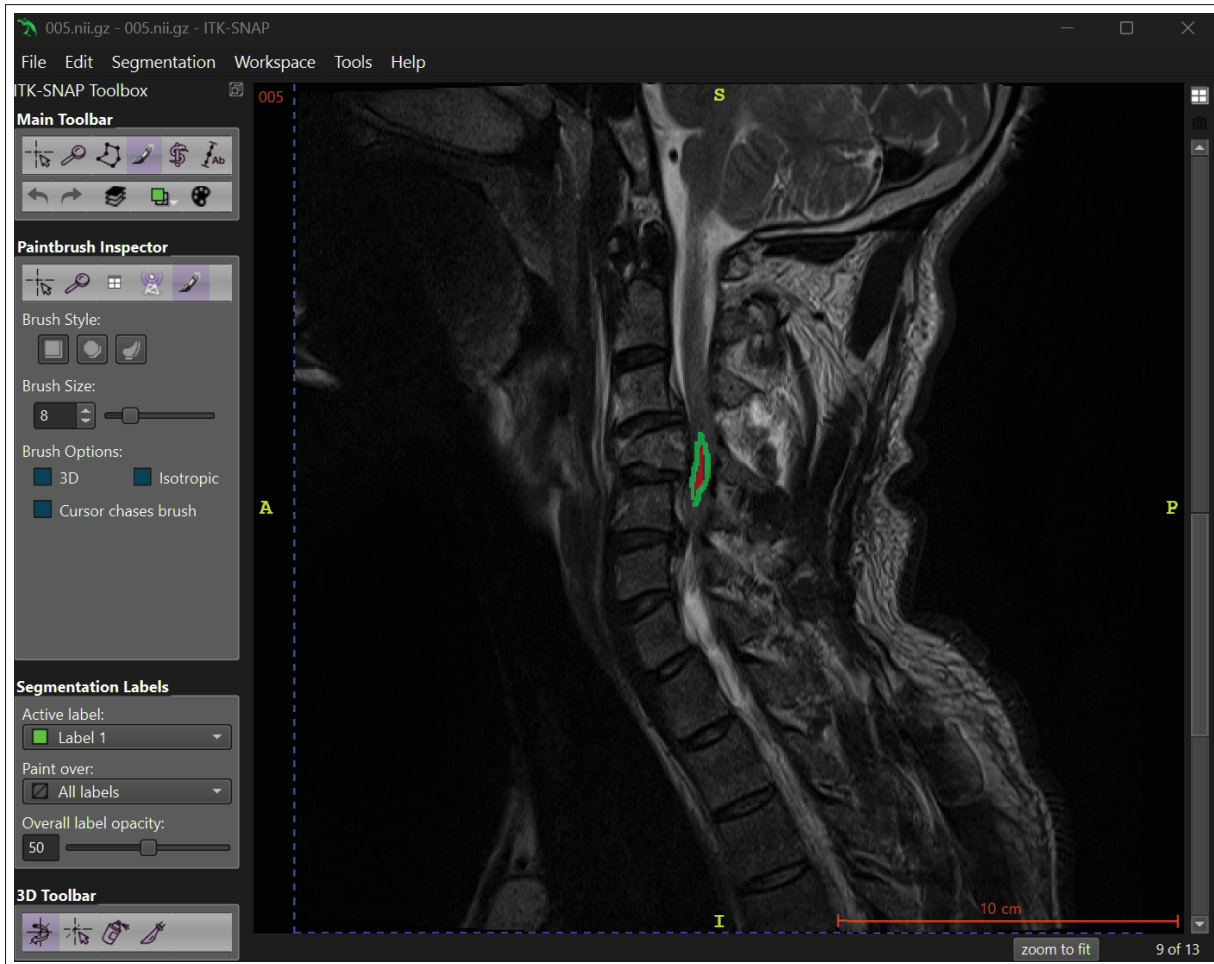


Figure 2.2 Exemple d'annotation manuelle dans ITK-SNAP (T2 sagittal) : œdème (vert) et hémorragie (rouge)

2.2.3 Transformation des annotations sagittales vers l'espace axial

Pour l'entraînement du modèle de segmentation axiale, il était nécessaire de transformer les annotations manuelles réalisées sur les images sagittales vers l'espace de coordonnées des images axiales. Cette transformation a été effectuée en deux étapes principales. D'abord, un enregistrement rigide a été réalisé entre les images sagittales T2 et axiales T2 pour chaque patient à l'aide de l'outil *sct_register_multimodal* de SpinalCordToolbox, utilisant

une métrique d'information mutuelle (MI) pour aligner les deux orientations d'imagerie. Les masques d'annotation sagittaux ont ensuite été transformés vers l'espace axial en appliquant la transformation d'enregistrement calculée à l'étape précédente, en utilisant l'outil *sct_apply_transfo* avec une interpolation du plus proche voisin pour préserver l'intégrité des labels de segmentation. Cette approche a également été appliquée aux prédictions du modèle sagittal pour générer un canal d'entrée supplémentaire lors de l'entraînement du modèle axial, permettant ainsi une fusion d'informations entre les deux orientations d'imagerie.

2.2.4 Prétraitement des images multimodales

Afin de préparer les données pour l'entraînement des modèles de segmentation, une série d'étapes de prétraitement a été appliquée pour standardiser les images et optimiser les performances.

La première étape a consisté à générer des masques de la moelle épinière pour chaque patient à l'aide de l'outil *sct_deepseg_sc* de SpinalCordToolbox. Ces masques ont servi de canal d'entrée supplémentaire lors de l'entraînement du modèle sagittal, en plus des images T1 et T2. L'objectif était de fournir au réseau une contrainte spatiale explicite pour l'aider à localiser la région d'intérêt.

Ensuite, les images ont été préparées en trois étapes : les zones vides autour de la moelle ont été retirées, toutes les images ont été mises à la même échelle spatiale, et les niveaux de gris ont été normalisés pour que les valeurs soient comparables d'un patient à l'autre. Ces étapes ont été automatisées par le framework nnU-Net. Les détails techniques sont présentés à l'annexe I.

2.3 Segmentation multimodale et multiclasse des lésions médullaires

2.3.1 Configuration nnU-Net pour la segmentation multimodale

Pour l'entraînement des modèles de segmentation, nous avons utilisé le framework nnU-Net avec ses paramètres par défaut et l'entraînement standard. nnU-Net automatise la configuration

des architectures U-Net en s'adaptant dynamiquement aux caractéristiques du jeu de données, éliminant ainsi le besoin de réglage manuel des hyperparamètres.

Nous utilisons une U-Net simple. L'encodeur (4 étages) réduit progressivement la taille de l'image tout en augmentant le nombre de cartes de caractéristiques ($\approx 32, 64, 128$, puis 256). À chaque étage, deux convolutions 3×3 sont appliquées puis la taille est réduite ; chaque bloc est normalisé (InstanceNorm) et activé avec Leaky ReLU. Au centre du réseau (le « goulot »), deux convolutions 3×3 apprennent des représentations plus riches (≈ 512 cartes). Le décodeur reconstruit ensuite l'image pas à pas : on agrandit par convolution transposée 2×2 , on concatène la connexion de saut correspondante de l'encodeur, puis on applique de nouveau deux convolutions 3×3 en divisant par deux le nombre de cartes à chaque fois. Enfin, une convolution 1×1 projette vers C classes et une activation Softmax convertit ces scores en probabilités multi-classes.

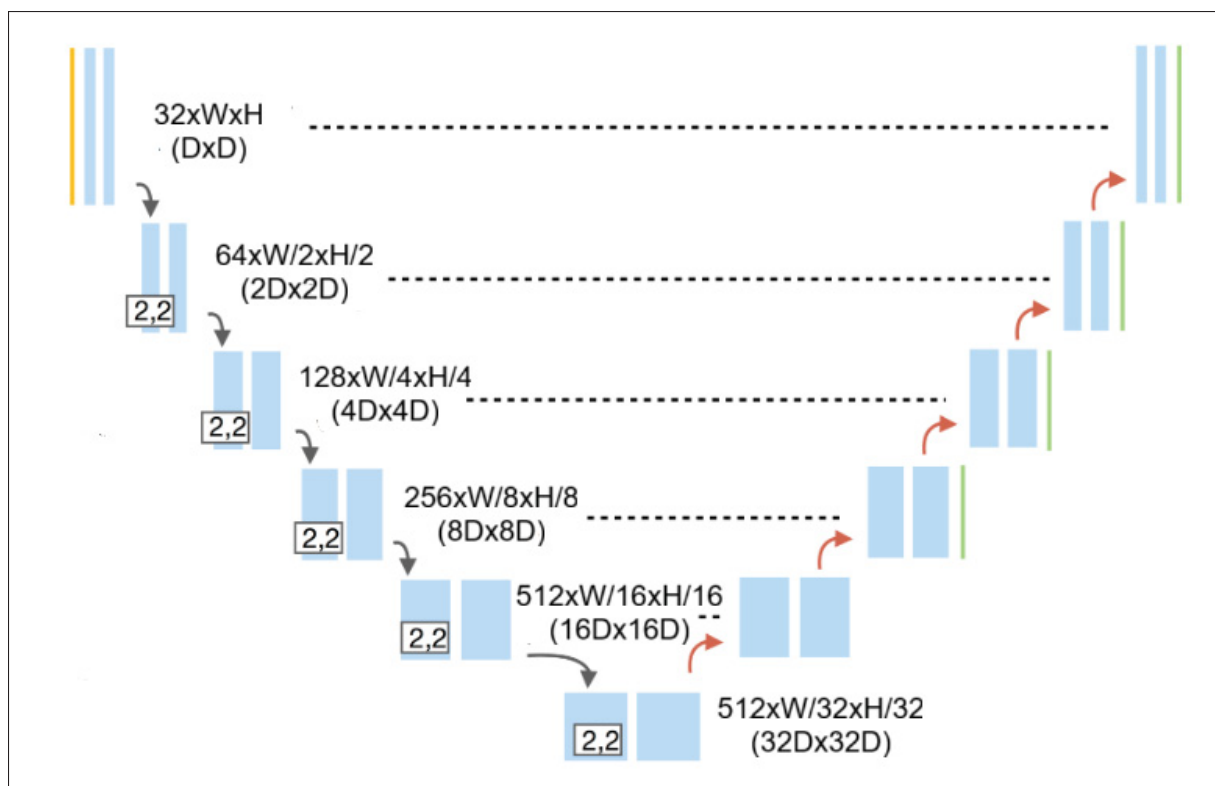


Figure 2.3 Schéma U-Net (encodeur-goulot-décodeur avec connexions de saut)

L'entraînement s'effectue sur 1000 époques avec une taille de lot minimale de 2, utilisant la descente de gradient stochastique avec un taux d'apprentissage initial de 0.01 et un momentum Nesterov de 0.99. La fonction de perte reprend la formulation standard de nnU-Net (Dice + entropie croisée, avec Dice calculé au niveau du lot) à laquelle nous ajoutons un terme topologique; la définition détaillée et la motivation sont présentées en section 2.3.3. Le prétraitement (normalisation Z-score et rééchantillonnage) est celui décrit en section 2.2.4.

2.3.2 Entraînement des modèles en cascade

2.3.2.1 Modèle sagittal

Le modèle sagittal a été entraîné sur 140 patients avec une configuration 2D utilisant les modalités T2 + T1 + masque de la moelle épinière, qui s'est révélée optimale après tests comparatifs. Les paramètres incluent une taille de lot de 16 (nombre d'exemples/patches traités en parallèle), des patches de 448×448 pixels (sous-images 2D découpées pour l'apprentissage) et une résolution de 0.558×0.558 mm par pixel (taille physique représentée par un pixel). Nous avons utilisé un split 80/20 : 28 des 140 patients ($\approx 20\%$) ont été réservés à la validation/test, les 112 restants à l'entraînement.

Tableau 2.1 Résumé des hyperparamètres du modèle sagittal

Paramètre	Valeur
Modèle	U-Net 2D
Modalités d'entrée	T2 + T1 + masque de la moelle
Taille de lot	16
Taille des patches	448 × 448 px
Résolution (mm/pixel)	0.558 × 0.558
Optimisateur	SGD (momentum de Nesterov 0.99)
Taux d'apprentissage	0.01
Nombre d'époques	1000
Normalisation	Z-score (cf. 2.2.4); InstanceNorm dans le réseau
Activation	Leaky ReLU
Fonction de perte	Dice + entropie croisée + terme topologique (cf. 2.3.3)
Split entraînement/validation	112 / 28 patients (80/20)

2.3.2.2 Modèle axial

Le modèle axial a ensuite été entraîné sur 122 patients en utilisant les images axiales natives en haute résolution et les prédictions sagittales du premier modèle comme canal d'entrée supplémentaire pour segmenter les lésions dans l'espace axial. Les paramètres incluent une taille de lot de 32 (exemples traités en parallèle), des patches de 320×320 pixels (fenêtres d'apprentissage) et une résolution de 0.5625×0.5625 mm par pixel. Nous avons utilisé un split 80/20 : 24 des 122 patients ($\approx 20\%$) pour la validation/test, les 98 restants pour l'entraînement.

Tableau 2.2 Résumé des hyperparamètres du modèle axial

Paramètre	Valeur
Modèle	U-Net 2D
Entrées	Axial natif + prédiction sagittale (canal supplémentaire)
Taille de lot	32
Taille des patches	320 × 320 px
Résolution (mm/pixel)	0.5625 × 0.5625
Optimisateur	SGD (momentum de Nesterov 0.99)
Taux d'apprentissage	0.01
Nombre d'époques	1000
Normalisation	Z-score (cf. 2.2.4); InstanceNorm dans le réseau
Activation	Leaky ReLU
Fonction de perte	Dice + entropie croisée + terme topologique (cf. 2.3.3)
Split entraînement/validation	98 / 24 patients (80/20)

2.3.3 Fonction de perte et optimisation

La fonction de perte utilisée combine trois termes : la perte de Dice, l'entropie croisée et un terme topologique adapté aux connaissances médicales du domaine. La fonction de perte totale s'exprime comme suit :

$$L_{total} = L_{Dice} + L_{CE} + L_{topo} \quad (2.1)$$

où L_{Dice} est la perte de Dice, L_{CE} l'entropie croisée et L_{topo} le terme topologique, chacun avec un poids de 1. Basé sur les connaissances médicales, l'hémorragie est généralement localisée au centre de la lésion au niveau de l'épicentre du trauma et est entourée par l'œdème qui représente l'inflammation. Le terme topologique vérifie donc l'alignement des centres de masse des deux composantes lésionnelles.

Pour chaque prédiction, nous calculons d'abord le centre de masse de l'œdème et de l'hémorragie séparément. Le centre de masse d'une région segmentée est calculé en normalisant d'abord les coordonnées spatiales dans l'intervalle $[0, 1]$. Cette normalisation est essentielle pour maintenir l'équilibre entre les différents termes de la fonction de perte composite, garantissant que le terme topologique contribue de manière proportionnelle sans dominer l'optimisation par rapport aux termes de Dice et d'entropie croisée. Le centre de masse normalisé est calculé comme suit :

$$\mathbf{C} = \left(\frac{\sum_{i=1}^n \frac{x_i}{W-1} \cdot I_i}{\sum_{i=1}^n I_i}, \frac{\sum_{i=1}^n \frac{y_i}{H-1} \cdot I_i}{\sum_{i=1}^n I_i} \right) \quad (2.2)$$

où (x_i, y_i) sont les coordonnées du pixel i , I_i est l'intensité binaire du pixel (1 si le pixel appartient à la classe, 0 sinon), W et H sont respectivement la largeur et la hauteur de l'image (correspondant à la taille des patches d'entraînement), et n est le nombre total de pixels dans l'image. Le terme topologique n'est calculé que si les deux classes (œdème et hémorragie) sont présentes dans la vérité terrain, évitant ainsi des pénalités non désirées. Dans ce cas, nous évaluons la distance euclidienne entre le centre de masse de l'œdème dans la vérité terrain et le centre de masse de l'hémorragie prédite :

$$L_{topo} = \sqrt{\left\| \mathbf{C}_{\text{œdème}}^{\text{vérité}} - \mathbf{C}_{\text{hémorragie}}^{\text{préd}} \right\|_2^2 + \varepsilon} \quad (2.3)$$

où $\mathbf{C}_{\text{œdème}}^{\text{vérité}}$ est le centre de masse de l'œdème dans la vérité terrain, $\mathbf{C}_{\text{hémorragie}}^{\text{préd}}$ est le centre de masse de l'hémorragie prédite, et $\varepsilon = 10^{-6}$ est un terme de stabilité numérique.

Cette pénalité encourage le modèle à prédire des structures lésionnelles où l'hémorragie est spatialement proche du centre de l'œdème, respectant ainsi l'anatomie pathologique attendue où l'hémorragie centrale est entourée par l'inflammation œdémateuse, comme illustré dans la Figure 2.4.

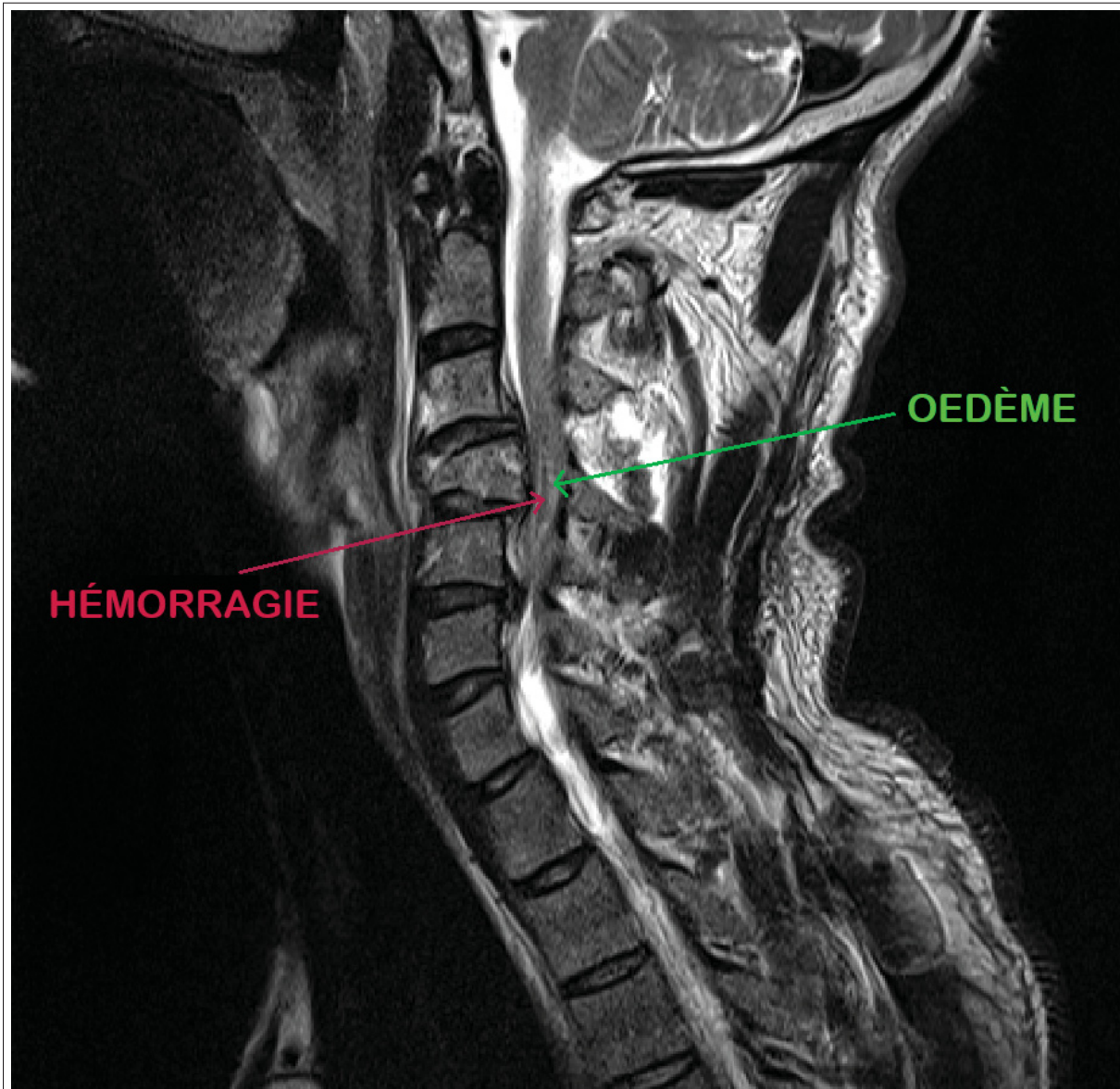


Figure 2.4 Hémorragie centrale entourée par l'œdème (illustration du principe topologique)

La distribution moyenne des pixels dans notre dataset présente un déséquilibre important entre les classes. L'analyse de 148 patients révèle une distribution moyenne de 99.97% pour le fond, 0.029% pour l'œdème et 0.0016% pour l'hémorragie. Pour favoriser l'apprentissage des classes minoritaires, nous avons implémenté une stratégie de balancement des classes de type "inverse", où les poids sont proportionnellement inverses à la distribution de chaque classe. Cette approche permet au modèle de mieux apprendre les caractéristiques des lésions malgré leur faible représentation dans les données.

2.3.4 Augmentation de données

L'augmentation de données a été réalisée en utilisant la configuration agressive de nnU-Net (nnUNetTrainerDA5), qui applique une série de transformations géométriques et d'intensité pour augmenter artificiellement la diversité du dataset d'entraînement. Cette approche est particulièrement importante pour les données médicales où la taille des datasets est souvent limitée.

Les transformations géométriques incluent des rotations dans les trois axes avec des angles aléatoires, des changements d'échelle (facteur de 0.7 à 1.43), et des transformations de miroir selon les axes appropriés. Pour les données 2D, des rotations de $\pm 180^\circ$ sont appliquées, tandis que pour les données 3D, des rotations plus limitées de $\pm 30^\circ$ sont utilisées pour préserver la cohérence anatomique.

Les transformations d'intensité comprennent l'ajout de bruit gaussien, des modifications de luminosité et de contraste, des transformations gamma, et la simulation de résolution réduite. Des transformations locales sont également appliquées, incluant des gradients de luminosité additifs, des transformations gamma locales, et des transformations de netteté. Ces augmentations sont appliquées avec des probabilités variables (généralement entre 0.1 et 0.5) pour chaque échantillon et chaque canal, permettant une diversité maximale tout en préservant la plausibilité anatomique des images.

2.4 Explicabilité du modèle de segmentation

2.4.1 Analyse SHAP des prédictions

Dans cette étude, nous avons utilisé les valeurs de Shapley (SHAP) pour identifier quelles régions (pixels) dans les scans T2 en particulier contribuent le plus aux prédictions d'œdème et d'hémorragie. SHAP est une approche basée sur la théorie des jeux pour expliquer les prédictions de modèles d'apprentissage automatique.

SHAP attribue à chaque caractéristique (pixel) sa contribution marginale à la prédiction finale. Pour comprendre le principe, considérons un exemple simple de régression linéaire : un modèle prédit le prix d'une maison en fonction de sa taille et de son âge. Si le prix moyen des maisons dans le dataset est de 100 000\$ et qu'une maison spécifique vaut 200 000\$, SHAP décompose cette différence de +100 000\$ en montrant que la grande taille contribue +80 000\$ et le jeune âge contribue +20 000\$.

Le baseline représente l'espérance des prédictions du modèle sur un ensemble de données de référence. Dans notre cas d'imagerie médicale, ce baseline correspond à la moyenne des prédictions du modèle sur les images d'entraînement. SHAP calcule la contribution de chaque pixel en évaluant systématiquement comment la sortie du modèle change lorsque ce pixel est modifié par rapport au baseline. Chaque pixel reçoit une valeur SHAP positive s'il augmente la sortie du modèle (par exemple la probabilité prédite dans notre cas), ou négative s'il la diminue.

Cette approche se formalise par l'équation :

$$f(x) - \mathbb{E}[f(x)] = \sum_{i=1}^M \phi_i \quad (2.4)$$

où $f(x)$ représente la prédiction sur l'image test, $\mathbb{E}[f(x)]$ l'espérance des prédictions (baseline), ϕ_i la valeur SHAP du pixel i , et M le nombre total de pixels.

Dans notre cas spécifique, nous avons utilisé l'implémentation DeepExplainer de SHAP, qui est conçue pour les modèles d'apprentissage profond. Deux choix de sélection de tranches sont à distinguer. Pour la distribution de référence (*background*), nous avons utilisé un échantillon de patients tirés aléatoirement de l'ensemble d'entraînement (*seed* fixe pour la reproductibilité), et pour chaque patient nous avons extrait la tranche sagittale médiane, c'est-à-dire la tranche centrale le long de l'axe ayant la plus petite dimension du volume. Ce choix fournit un baseline anatomiquement cohérent à coût computationnel raisonnable. Pour le patient analysé (image test à expliquer), nous avons en revanche sélectionné la tranche présentant la plus grande surface annotée d'œdème (classe 1) dans la vérité-terrain, avec un repli sur la tranche présentant la plus grande surface d'hémorragie (classe 2) si l'œdème est absent. Ce choix maximise le contenu lésionnel de l'image expliquée et évite d'analyser une coupe sans lésion. Ces images de référence et de test sont ensuite utilisées par DeepExplainer pour calculer respectivement l'espérance des prédictions du modèle (baseline) et la prédiction à expliquer.

Le processus d'analyse suit plusieurs étapes (Figure 2.5). Premièrement, le modèle sagittal entraîné est chargé avec ses poids optimaux et configuré en mode évaluation. Les images de référence sont sélectionnées parmi les patients d'entraînement, puis chargées et normalisées selon les mêmes paramètres que ceux utilisés lors de l'entraînement (Z-score par canal). L'outil d'explication DeepExplainer est alors initialisé avec ce modèle entraîné et les images de référence.

Pour chaque image test, les valeurs SHAP sont calculées selon la formulation mathématique présentée ci-dessus. DeepExplainer utilise les images de référence pour calculer le baseline. Chaque pixel reçoit une valeur SHAP qui indique sa contribution à l'écart entre la prédiction de l'image test et le baseline, générant ainsi une carte d'attribution pixel-par-pixel.

Les classes d'œdème (classe 1) et d'hémorragie (classe 2) sont évaluées séparément, permettant d'identifier les régions spécifiques qui contribuent à chaque composante de lésion. Cette approche permet de valider que le modèle se base sur des caractéristiques anatomiquement cohérentes et d'identifier d'éventuels biais dans les prédictions.

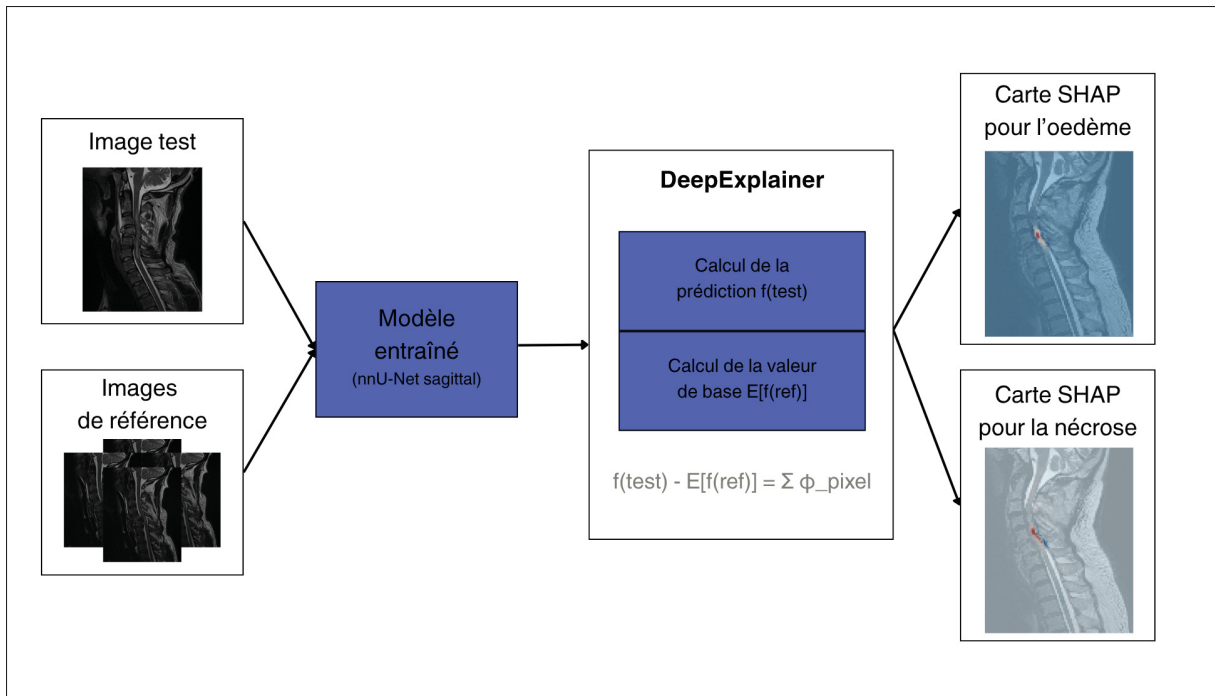


Figure 2.5 Schéma du calcul des valeurs SHAP (DeepExplainer) : prédiction $f(\text{test})$, baseline $\mathbb{E}[f(\text{ref})]$, et cartes SHAP pour l'œdème et l'hémorragie

2.4.2 Validation et interprétation des résultats SHAP

Nous avons choisi SHAP parmi les méthodes d'explicabilité disponibles (LIME, Grad-CAM, Integrated Gradients) pour sa stabilité. SHAP produit des explications cohérentes entre différentes images, ce qui signifie que des pixels similaires reçoivent des valeurs d'attribution similaires. En imagerie médicale, cette fiabilité aide les cliniciens à comprendre les prédictions du modèle et à prendre des décisions thérapeutiques. Contrairement à LIME qui peut varier selon les perturbations, SHAP offre des résultats reproductibles. Cependant, nous avons dû limiter le nombre d'échantillons de référence en raison du coût computationnel élevé.

Nous avons validé les explications SHAP de trois façons : 1) nous avons comparé les cartes d'attribution aux annotations manuelles des experts pour vérifier la correspondance avec les zones anatomiquement pertinentes ; 2) nous avons effectué une analyse de cohérence en calculant la corrélation entre les valeurs SHAP de pixels similaires dans différentes images pour valider

la stabilité ; 3) nous avons réalisé des tests de sensibilité en perturbant légèrement les images d'entrée et en observant les changements dans les cartes SHAP correspondantes.

Les valeurs SHAP positives indiquent une contribution à l'augmentation de la probabilité de prédiction, les valeurs négatives une contribution à la diminution. L'amplitude reflète l'importance relative du pixel dans la décision du modèle. Pour les lésions médullaires, des valeurs plus élevées indiquent de plus forts indices de présence d'œdème ou d'hémorragie respectivement.

2.5 Quantification des métriques pronostiques

2.5.1 Extraction des biomarqueurs axiaux par niveau vertébral

À partir des segmentations, nous extrayons, pour chaque niveau vertébral, des biomarqueurs spécifiques à chaque composante lésionnelle (œdème, hémorragie) dans le plan axial. Plus précisément, nous calculons (i) l'aire de section transversale de l'œdème et de l'hémorragie à chaque niveau, (ii) le volume par niveau en intégrant l'aire axiale le long de l'axe cranio-caudal, et (iii) le ratio hémorragie/lésion défini comme $\text{hémorragie}/(\text{œdème}+\text{hémorragie})$. Ces quantités sont obtenues à partir des masques de segmentation alignés sur l'espace axial, avec l'étiquetage des niveaux vertébraux fourni par SpinalCordToolbox. L'objectif est d'évaluer la valeur pronostique incrémentale de ces biomarqueurs — en particulier la contribution spécifique de l'hémorragie via son aire/volume par niveau et son ratio relatif — par rapport aux variables cliniques seules et à une mesure agrégée de lésion complète.

Pour évaluer cette valeur pronostique, nous construisons trois modèles de régression linéaire emboîtés, permettant de tester l'hypothèse H3 selon laquelle la quantification différenciée de l'œdème et de l'hémorragie améliore la prédiction du pronostic neurologique par rapport à une approche de lésion complète. La variable dépendante utilisée est le score AIS (ASIA Impairment Scale), une échelle standardisée de classification neurologique des lésions médullaires allant de A (lésion complète, absence de fonction motrice et sensitive sous le niveau lésionnel) à E (fonction

normale). Les grades intermédiaires B, C et D correspondent à des lésions incomplètes avec préservation partielle des fonctions sensibles et/ou motrices. Cette échelle constitue la référence internationale pour l'évaluation du pronostic neurologique après traumatisme médullaire.

Le premier modèle (M1, clinique) utilise uniquement le score AIS préopératoire comme variable prédictive, représentant la baseline clinique actuelle où le pronostic est estimé à partir de l'évaluation neurologique initiale : $M1 : Y = \beta_0 + \beta_1 \times AIS_{pre} + \varepsilon$. Le deuxième modèle (M2, lésion complète) ajoute le volume lésionnel total (œdème + hémorragie) comme variable prédictive, représentant l'approche traditionnelle où la lésion est quantifiée de manière globale sans distinction entre ses composantes : $M2 : Y = \beta_0 + \beta_1 \times AIS_{pre} + \beta_2 \times Volume_{total} + \varepsilon$. Le troisième modèle (M3, différencié) remplace le volume total par les biomarqueurs différenciés — volume d'œdème, volume d'hémorragie, et ratio hémorragie/lésion — permettant d'évaluer la contribution spécifique de chaque composante lésionnelle : $M3 : Y = \beta_0 + \beta_1 \times AIS_{pre} + \beta_2 \times Volume_{oedeme} + \beta_3 \times Volume_{necrose} + \beta_4 \times Ratio_{necrose/lesion} + \varepsilon$. La variable dépendante Y représente le score AIS à différentes périodes de suivi (post-opératoire immédiat, 3 mois, 6 mois, 1 an), permettant d'évaluer la valeur prédictive des biomarqueurs à court, moyen et long terme.

La comparaison des modèles s'effectue par le coefficient de détermination ajusté (R^2 ajusté), qui pénalise l'ajout de variables non informatives, et par un test F de modèles emboîtés. Ce test évalue si l'amélioration du R^2 entre M2 et M3 est statistiquement significative, c'est-à-dire si les biomarqueurs différenciés apportent une information prédictive au-delà du volume total. Le test F compare la réduction de la somme des carrés résiduels entre les deux modèles selon la formule $F = \frac{(SSR_{M2} - SSR_{M3}) / (df_{M2} - df_{M3})}{SSR_{M3} / df_{M3}}$, où SSR représente la somme des carrés résiduels et df les degrés de liberté. Une valeur $p < 0,05$ indique que le modèle différencié M3 explique significativement plus de variance que le modèle de lésion complète M2.

2.5.2 Système de visualisation et aide à la décision clinique

L'architecture en cascade (sagittal puis axial) génère des prédictions finales dans l'espace axial. Cette approche a été choisie car la visualisation axiale offre la perspective la plus directe sur l'anatomie de la moelle épinière.

La visualisation axiale permet d'identifier les structures médullaires affectées (substance grise versus substance blanche), d'évaluer la déformation de la moelle, et de localiser la lésion par rapport aux niveaux vertébraux. Comparée à la vue sagittale, elle est préférable pour l'évaluation clinique car elle permet d'identifier précisément les voies neurologiques affectées et d'évaluer l'étendue des dommages.

Le processus de génération des visualisations utilise un modèle de référence qui redresse la ligne centrale de la colonne vertébrale sur l'axe vertical. Les lésions segmentées et les niveaux vertébraux sont d'abord alignés sur ce modèle standard. L'image axiale T2 haute résolution est ensuite enregistrée sur l'image sagittale correspondante. Les masques de lésion et de niveaux vertébraux sont transformés dans l'espace axial en appliquant la matrice de transformation calculée. Les visualisations finales sont générées à partir des coupes axiales avec superposition des étiquettes de niveaux vertébraux et des contours des composantes lésionnelles (œdème et hémorragie). Ce processus s'appuie sur les outils d'étiquetage, de transformation et le modèle de référence fournis par SpinalCordToolbox.

2.6 Métriques d'évaluation et validation

2.6.1 Métriques de segmentation

L'évaluation des modèles sagittal et axial utilise les métriques standardisées du framework nnUNet, calculées pour chaque patient de l'ensemble de validation puis moyennées par classe de lésion (œdème et hémorragie). Une moyenne globale est également calculée pour évaluer la performance d'ensemble.

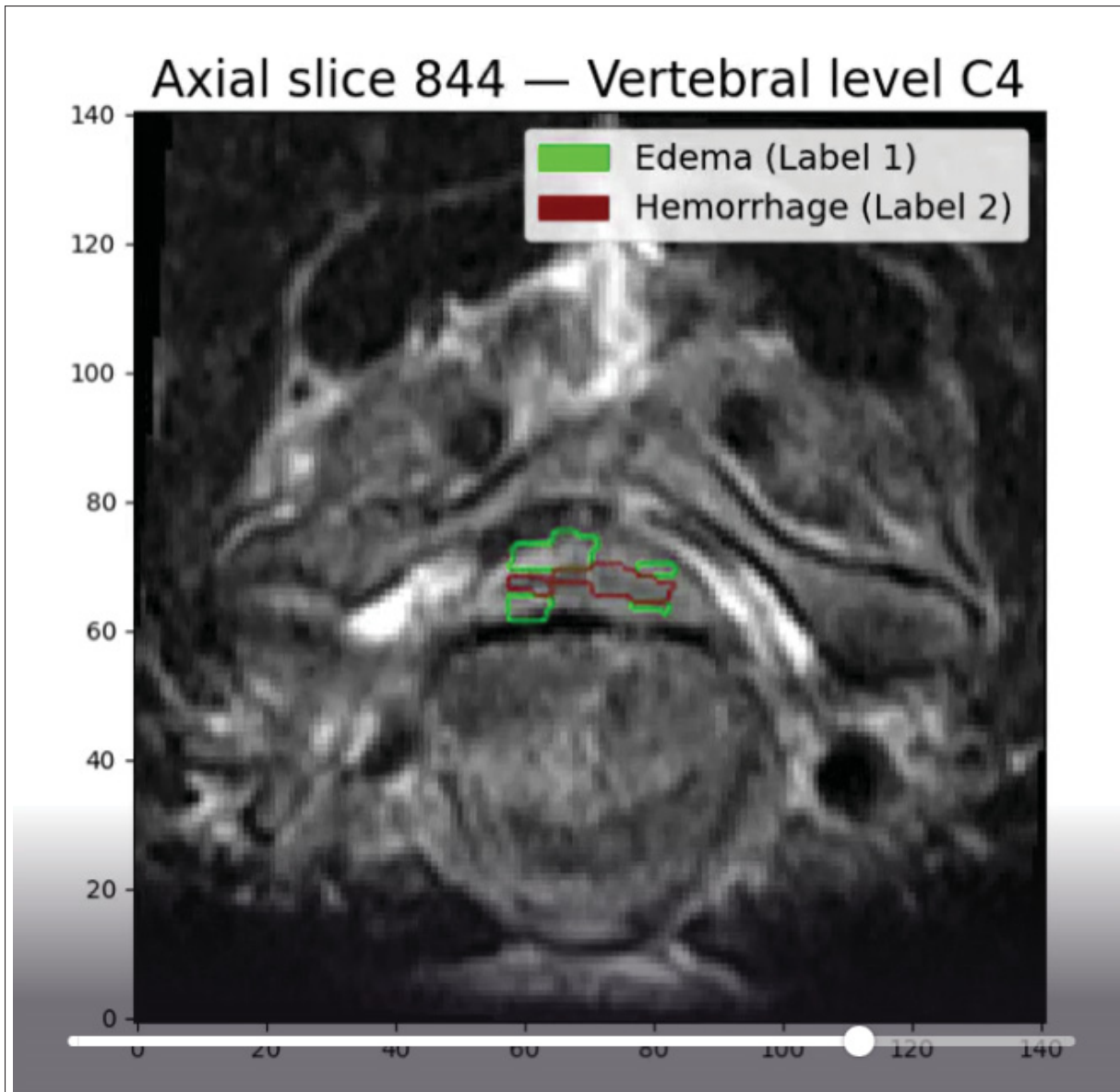


Figure 2.6 Système de visualisation axiale des lésions : coupe T2 avec contours d'œdème (Label 1) et d'hémorragie (Label 2), et niveau vertébral indiqué

Le coefficient Dice mesure le chevauchement entre la prédiction et la vérité terrain, en pénalisant de façon équilibrée les faux positifs et les faux négatifs. Il reste stable pour les petites structures, ce qui est pertinent pour les lésions médullaires, très petites par rapport à la moelle, et pour l'hémorragie, encore plus réduite. Il est aussi moins sensible au fort déséquilibre entre fond et lésion, fréquent en imagerie médullaire. La formule est :

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN} \quad (2.5)$$

En complément, le coefficient IoU (ou Jaccard) évalue la superposition de manière plus stricte, avec une sensibilité accrue aux contours. Il pénalise fortement les pixels mal classés sur les bords et devient ainsi pertinent pour les petites structures comme l'hémorragie, où quelques pixels d'erreur comptent beaucoup. Comme son dénominateur inclut l'union des surfaces prédite et de référence, toute surestimation ou omission de contour dégrade immédiatement le score.

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN} \quad (2.6)$$

Enfin, les métriques de classification binaire incluent les vrais positifs (TP), faux positifs (FP), vrais négatifs (TN) et faux négatifs (FN). Ces métriques permettent de calculer la sensibilité et la spécificité. Pour la segmentation des lésions médullaires, la sensibilité mesure la capacité du modèle à détecter correctement les pixels de lésion, tandis que la spécificité évalue sa capacité à éviter les faux positifs dans les régions saines de la moelle épinière.

2.6.2 Méthode de validation croisée

L'évaluation des modèles sagittal et axial utilise une validation croisée 5-fold. Cette méthode divise l'ensemble de données en cinq sous-ensembles de taille égale. Chaque sous-ensemble sert alternativement d'ensemble de validation tandis que les quatre autres constituent l'ensemble d'entraînement.

Cette approche maximise l'utilisation des données d'entraînement et fournit une estimation robuste des performances. La répartition des données est stratifiée pour maintenir une distribution équilibrée des classes de lésions (œdème et hémorragie) dans chaque sous-ensemble.

Pour chaque sous-ensemble, les métriques de segmentation sont calculées sur l'ensemble de validation correspondant. Les résultats finaux sont obtenus en calculant la moyenne et l'écart-type des métriques sur les cinq sous-ensembles.

2.6.3 Tests statistiques et comparaison à l'état de l'art

Pour situer nos performances par rapport à l'état de l'art, nous comparons nos modèles à SciSeg (SpinalCordToolbox), un outil de segmentation automatique des lésions hyperintenses en T2 basé sur nnU-Net. SciSeg est entraîné sur un large ensemble multi-site multi-fabricants et segmente la moelle ainsi qu'une classe unique de lésion chez des patients présentant des lésions médullaires (traumatiques et apparentées). Comme SciSeg ne renvoie qu'une seule classe alors que nos modèles séparent l'œdème de l'hémorragie, nous fusionnons nos deux classes en un masque de lésion complète (œdème + hémorragie), que nous comparons à la sortie binaire de SciSeg.

Afin d'éviter tout biais d'échantillonnage, les comparaisons sont réalisées sur les mêmes patients (paire par paire). Pour chaque patient, nous calculons Dice et IoU (et, en annexe, sensibilité et spécificité), puis nous résumons les résultats au niveau du groupe.

Nous évaluons, séparément pour chaque métrique et pour chaque sortie (sagittal et axial), si nos modèles diffèrent de SciSeg. Les métriques testées sont le Dice et l'IoU, calculés sur ce masque de lésion complète. Pour chaque patient, nous calculons la différence appariée entre notre score et celui de SciSeg, puis nous réalisons un test bilatéral pour vérifier si la différence moyenne s'écarte de zéro, c'est-à-dire si l'écart moyen est significatif.

Pour déterminer si cet écart est statistiquement significatif, nous vérifions d'abord si la distribution des différences appariées entre méthodes, patient par patient, suit une loi normale (test de Shapiro–Wilk). Si la normalité n'est pas rejetée, nous utilisons un test t apparié et nous rapportons la taille d'effet correspondante ainsi qu'un intervalle de confiance à 95 % de la différence moyenne. Si la normalité est rejetée, nous appliquons le test des rangs signés de Wilcoxon et nous rapportons sa taille d'effet, ainsi qu'un intervalle de confiance à 95 % de la médiane de la différence estimé par bootstrap.

Nous testons quatre comparaisons (métrique \times sortie) : Dice (sagittal), Dice (axial), IoU (sagittal) et IoU (axial). Comme ces quatre tests sont réalisés en parallèle, nous ajustons les valeurs p avec la correction de Benjamini–Hochberg (FDR 5 %) afin de réduire le risque de faux positifs (résultats déclarés significatifs à tort). Nous présentons également les résultats par classe (œdème et hémorragie) pour nos modèles, sans comparaison statistique directe avec SciSeg qui ne produit qu’une seule classe. Enfin, seuls les patients pour lesquels les deux méthodes produisent un masque valide sont inclus dans les analyses, et tous les traitements statistiques sont réalisés avec des scripts Python en environnement contrôlé et reproductible.

CHAPITRE 3

RÉSULTATS ET DISCUSSION

3.1 Performance de la segmentation automatique

3.1.1 Résultats de la segmentation multimodale

3.1.1.1 Performance du modèle de segmentation

Les performances de notre modèle ont été comparées à celles de SciSeg pour la segmentation de la lésion complète (fusion des classes œdème et hémorragie) sur un total de 122 patients pour les évaluations axiales et 140 patients pour les évaluations sagittales. Les résultats sont présentés dans le Tableau 3.1 et montrent une supériorité statistiquement significative de notre modèle sur SciSeg pour toutes les métriques évaluées.

Tableau 3.1 Comparaison des performances entre notre modèle et SciSeg pour les métriques Dice et coefficient de Jaccard dans les plans axial et sagittal

Métrique	Plan	n	Notre modèle	SciSeg	Diff. méd.	IC 95%	p ajusté	Taille d'effet
Dice	Axial	122	0,540 ± 0,193	0,268 ± 0,217	0,250	[0,185–0,307]	< 0,001	r = -0,80
Jaccard	Axial	122	0,392 ± 0,175	0,175 ± 0,163	0,216	[0,154–0,243]	< 0,001	r = -0,80
Dice	Sagittal	140	0,575 ± 0,187	0,461 ± 0,234	0,054	[0,025–0,102]	< 0,001	r = -0,53
Jaccard	Sagittal	140	0,424 ± 0,164	0,328 ± 0,189	0,056	[0,025–0,098]	< 0,001	r = -0,53

Note : Les valeurs sont présentées sous forme de moyenne ± écart-type. La différence médiane correspond à la médiane des différences appariées entre notre modèle et SciSeg. Les intervalles de confiance (IC) à 95% sont calculés par bootstrap pour la médiane. Les p-values ont été ajustées selon la méthode de Benjamini-Hochberg (FDR). Tous les tests statistiques ont été effectués à l'aide du test de Wilcoxon pour données appariées.

Dans le plan axial, notre modèle a démontré des performances nettement supérieures à SciSeg. Pour le coefficient de Dice, notre modèle a obtenu une moyenne de 0,540 ± 0,193 (médiane : 0,577) contre 0,268 ± 0,217 (médiane : 0,278) pour SciSeg, représentant une amélioration médiane de 0,250 (intervalle de confiance à 95% [IC 95%] : 0,185 - 0,307). Pour le coefficient de Jaccard, notre modèle a atteint une moyenne de 0,392 ± 0,175 (médiane : 0,406) comparativement à 0,175 ± 0,163 (médiane : 0,161) pour SciSeg, soit une amélioration médiane de 0,216 (IC

95% : 0,154 - 0,243). Les tests statistiques (test de Wilcoxon pour données appariées) ont révélé des différences hautement significatives pour ces deux métriques ($p < 0,001$ après correction de Benjamini-Hochberg), avec des tailles d'effet importantes ($r \approx -0,80$), indiquant une supériorité de notre approche.

Dans le plan sagittal, notre modèle a également montré une supériorité statistiquement significative, bien que l'ampleur de l'amélioration soit plus modérée que dans le plan axial. Pour le coefficient de Dice, notre modèle a obtenu une moyenne de $0,575 \pm 0,187$ (médiane : 0,619) contre $0,461 \pm 0,234$ (médiane : 0,515) pour SciSeg, représentant une amélioration médiane de 0,054 (IC 95% : 0,025 - 0,102). Pour le coefficient de Jaccard, notre modèle a atteint une moyenne de $0,424 \pm 0,164$ (médiane : 0,449) comparativement à $0,328 \pm 0,189$ (médiane : 0,347) pour SciSeg, soit une amélioration médiane de 0,056 (IC 95% : 0,025 - 0,098). Les tests statistiques (test de Wilcoxon) ont confirmé la significativité de ces différences ($p < 0,001$ après correction de Benjamini-Hochberg), avec des tailles d'effet modérées à importantes ($r \approx -0,53$).

Pour permettre une comparaison équitable avec SciSeg, qui produit un masque binaire de lésion, les prédictions de notre modèle ont été transformées en masques de lésion complète en fusionnant les voxels étiquetés comme lésion (classe 1) ou hémorragie (classe 2) pour former un masque binaire unique.

Toutes les comparaisons ont été effectuées à l'aide de tests statistiques appariés, car chaque patient a été évalué par les deux méthodes. La normalité des distributions des différences a été testée à l'aide du test de Shapiro-Wilk, et dans tous les cas, les distributions se sont révélées non normales ($p < 0,05$), justifiant l'utilisation du test non paramétrique de Wilcoxon pour données appariées. Une correction pour tests multiples a été appliquée selon la méthode de Benjamini-Hochberg (FDR), et toutes les différences sont demeurées significatives après cette correction (p ajusté $< 0,001$ pour toutes les métriques).

Ces résultats démontrent que notre modèle de segmentation multimodale surpasse significativement SciSeg, l'état de l'art actuel, dans les deux plans d'imagerie, avec des améliorations particulièrement marquées dans le plan axial (hypothèse H1).

3.1.1.2 Impact des différentes modalités sur la segmentation

Pour évaluer l'impact de l'ajout de modalités d'imagerie supplémentaires sur les performances de segmentation (hypothèse H2), nous avons comparé trois configurations d'entraînement : (1) une configuration utilisant uniquement la séquence T2 (148 patients), (2) une configuration utilisant les séquences T2 et T1 (140 patients), et (3) une configuration utilisant les séquences T2, T1 et STIR (99 patients). Il est important de noter que le nombre de patients disponibles diminue avec l'ajout de modalités supplémentaires, car toutes les modalités ne sont pas disponibles pour tous les patients de la cohorte.

Tableau 3.2 Performances de segmentation selon les configurations de modalités d'imagerie

Configuration	n	Dice	Coefficient de Jaccard
T2 seul	148	0,555 ± 0,182	0,404 ± 0,156
T2 + T1	140	0,566 ± 0,184	0,414 ± 0,158
T2 + T1 + STIR	99	0,552 ± 0,165	0,397 ± 0,142

Note : Les valeurs sont présentées sous forme de moyenne ± écart-type. n : nombre de patients utilisés pour l'entraînement.

Les résultats montrent une amélioration légère mais non significative des performances avec l'ajout de la modalité T1. La configuration T2+T1 atteint un coefficient de Dice moyen de $0,566 \pm 0,184$ (médiane : 0,613) et un coefficient de Jaccard de $0,414 \pm 0,158$ (médiane : 0,442), comparativement à $0,555 \pm 0,182$ (médiane : 0,588) et $0,404 \pm 0,156$ (médiane : 0,416) respectivement pour la configuration T2 seul. Cette amélioration représente une augmentation relative d'environ 2,0% pour le coefficient de Dice et 2,5% pour le coefficient de Jaccard.

En revanche, l'ajout de la séquence STIR (configuration T2+T1+STIR) ne semble pas apporter d'amélioration supplémentaire, avec des performances légèrement inférieures (Dice : $0,552 \pm 0,165$, médiane : 0,592; Jaccard : $0,397 \pm 0,142$, médiane : 0,421) comparativement à la configuration T2+T1. Cette observation doit cependant être interprétée avec prudence, car la configuration T2+T1+STIR utilise seulement 99 patients pour l'entraînement, soit 33% de moins que la configuration T2 seul (148 patients) et 29% de moins que la configuration T2+T1 (140 patients).

Pour évaluer statistiquement les différences entre les trois configurations, nous avons effectué un test de Kruskal-Wallis (test non paramétrique pour comparer plusieurs groupes indépendants), suivi de comparaisons par paires à l'aide du test de Mann-Whitney U. La normalité des distributions a été testée à l'aide du test de Shapiro-Wilk, et toutes les distributions se sont révélées non normales ($p < 0,05$), justifiant l'utilisation de tests non paramétriques.

Le test de Kruskal-Wallis n'a révélé aucune différence statistiquement significative entre les trois configurations, ni pour le coefficient de Dice ($H = 1,20$, $p = 0,548$) ni pour le coefficient de Jaccard ($H = 1,20$, $p = 0,548$). De même, les comparaisons par paires, corrigées pour tests multiples selon la méthode de Benjamini-Hochberg (FDR), n'ont indiqué aucune différence significative, avec une valeur p ajustée identique de 0,562 pour toutes les comparaisons et les deux métriques.

L'absence de différences statistiquement significatives entre les configurations peut s'expliquer par plusieurs facteurs. Premièrement, la réduction progressive du nombre de patients disponibles avec l'ajout de modalités supplémentaires ($148 \rightarrow 140 \rightarrow 99$) peut contrebalancer les bénéfices potentiels de l'information supplémentaire fournie par les modalités T1 et STIR. En effet, l'apprentissage profond nécessite généralement une quantité suffisante de données pour généraliser efficacement, et la réduction de 33% du nombre de cas d'entraînement entre la configuration T2 seul et T2+T1+STIR peut limiter la capacité du modèle à exploiter pleinement l'information multimodale.

Deuxièmement, l'amélioration observée avec l'ajout de T1, bien que non significative statistiquement, suggère que cette modalité apporte une information complémentaire utile pour la segmentation. La séquence T1 peut aider à mieux distinguer certaines structures anatomiques et à améliorer la délimitation des lésions, ce qui expliquerait la tendance à l'amélioration observée.

Enfin, l'absence d'amélioration avec l'ajout de STIR peut être attribuée à la fois à la réduction du nombre de cas d'entraînement et à la redondance potentielle de l'information fournie par cette séquence par rapport à T2 et T1. La séquence STIR, bien qu'utile pour la détection de certaines

pathologies, peut ne pas apporter d'information suffisamment discriminative supplémentaire pour justifier la perte de données d'entraînement associée.

Ces résultats suggèrent que, dans le contexte de cette étude où la disponibilité des modalités est limitée, la configuration T2+T1 offre le meilleur compromis entre le nombre de cas disponibles et les performances de segmentation, avec une amélioration légère mais cohérente par rapport à la configuration T2 seul.

3.1.2 Analyse de la segmentation multiclasse

3.1.2.1 Performance par classe (œdème vs hémorragie)

Notre approche de segmentation multiclasse utilise une architecture en cascade (section 2.3.2), où un modèle sagittal fournit des prédictions initiales qui sont ensuite utilisées comme entrée supplémentaire pour un modèle axial, produisant les segmentations finales (hypothèse H2). Le modèle permet de distinguer deux classes de tissus lésionnels : l'œdème (classe 1) et l'hémorragie (classe 2), comme décrit dans la section 2.2.2. L'évaluation des performances par classe a été effectuée en utilisant la validation croisée 5-fold (section 2.6.2) sur les résultats finaux du modèle axial.

Tableau 3.3 Performances de segmentation par classe pour le modèle en cascade (résultats finaux axiaux)

Classe	Structure	Dice	Coefficient de Jaccard	TP	FP	FN
1	Œdème	0,557	0,404	578,6	602,6	174,2
2	Hémorragie	0,135	0,092	11,9	14,5	34,5

Note : Les valeurs présentées sont les moyennes calculées sur l'ensemble des patients évalués en validation croisée 5-fold. TP : Vrais positifs, FP : Faux positifs, FN : Faux négatifs. Les sensibilité, spécificité, matrice de confusion et résultats par sous-ensemble sont détaillés en annexe IV.

Les résultats de segmentation par classe montrent une différence marquée entre les deux structures lésionnelles. La classe 1 (œdème) atteint un coefficient de Dice de 0,557 et un coefficient de Jaccard de 0,404, démontrant une capacité de détection et de délimitation satisfaisante de cette structure. En moyenne, le modèle détecte 578,6 voxels d'œdème correctement (vrais positifs),

avec 602,6 faux positifs et 174,2 faux négatifs. Le nombre relativement élevé de faux positifs suggère une tendance à la sur-segmentation de l'œdème, ce qui peut être attribué aux défis de la transformation des annotations sagittales vers l'espace axial (section 2.2.3) ou à la sensibilité du modèle aux variations d'intensité dans les images axiales.

En revanche, la classe 2 (hémorragie) présente des performances nettement inférieures, avec un coefficient de Dice de 0,135 et un coefficient de Jaccard de 0,092. Cette différence importante peut s'expliquer par plusieurs facteurs : (1) l'hémorragie est une structure plus rare dans la cohorte (seulement 172 coupes sur 502 contenaient de l'hémorragie selon la section 2.2.2), ce qui limite la quantité de données d'entraînement disponibles pour cette classe ; (2) l'hémorragie présente un hyposignal par rapport au tissu médullaire sain, ce qui peut la rendre plus difficile à distinguer du bruit ou d'autres structures hypointenses ; (3) la taille généralement plus petite des régions d'hémorragie par rapport à l'œdème rend la segmentation plus sensible aux erreurs de délimitation, particulièrement lors de la transformation des annotations sagittales vers l'espace axial.

La différence de performance entre les deux classes est attendue compte tenu de la nature des données : l'œdème est plus fréquent, plus volumineux et présente un contraste plus marqué (hypersignal) par rapport au tissu sain, facilitant sa détection et sa segmentation. L'hémorragie, en revanche, présente des défis supplémentaires liés à sa rareté, sa petite taille et son hyposignal, qui compliquent sa détection précise, particulièrement dans le contexte d'une architecture en cascade où les annotations de référence sont transformées depuis l'espace sagittal.

3.1.2.2 Analyse des cas difficiles

Pour mieux comprendre les limites de notre modèle et identifier les facteurs contribuant aux erreurs de segmentation, nous avons effectué une analyse systématique des cas présentant les performances les plus faibles. L'analyse des cas difficiles a été effectuée en calculant les métriques de segmentation pour chaque patient de la cohorte. Les patients ont été classés selon plusieurs critères : (1) les cas avec un coefficient de Dice combiné (lésion complète) égal à zéro,

indiquant une absence totale de détection ; (2) les cas avec un Dice combiné très faible ($< 0,2$) ; (3) les cas avec un Dice spécifique à l'œdème (classe 1) inférieur à $0,3$; et (4) les cas avec un Dice spécifique à l'hémorragie (classe 2) inférieur à $0,1$, lorsque cette classe est présente. Cette approche systématique permet d'identifier les patients pour lesquels le modèle rencontre des difficultés particulières et d'analyser les caractéristiques communes à ces cas.

Plusieurs facteurs récurrents contribuent aux erreurs de segmentation. Dans la majorité des cas, la difficulté principale provient d'une résolution spatiale réduite des images sagittales par rapport à la résolution standard observée dans la cohorte. Alors que la plupart des images sagittales présentent une résolution de 512×512 pixels, certains cas difficiles, notamment les cas 1 et 25, présentent une résolution réduite de $384 \times 384 \times 13$ voxels. Cette réduction de résolution limite la quantité d'information spatiale disponible pour le modèle et peut compromettre la précision de la segmentation, particulièrement pour les petites structures lésionnelles.

Les cas 1 et 25 illustrent plusieurs défis communs aux cas difficiles (Figure 3.1). Premièrement, ces cas présentent une résolution spatiale réduite ($384 \times 384 \times 13$) comparativement à la résolution standard de 512×512 observée dans la majorité de la cohorte. Cette réduction de résolution affecte la qualité des détails anatomiques et peut compromettre la précision de la segmentation, particulièrement pour les structures de petite taille.

Deuxièmement, le cas 1 présente des artefacts d'imagerie visibles sous forme de lignes verticales (hachures) qui traversent la moelle épinière. Ces artefacts peuvent être causés par des problèmes d'acquisition IRM, tels que des interférences électromagnétiques, des mouvements du patient, ou des défauts dans la séquence d'acquisition. Ces artefacts introduisent des variations d'intensité non physiologiques qui peuvent confondre le modèle et conduire à des erreurs de segmentation.

Troisièmement, dans les cas 1 et 25, la région d'œdème est très petite par rapport à la taille totale de la moelle épinière visible dans l'image. Cette caractéristique pose un défi particulier car le modèle a été entraîné sur une distribution de tailles de lésions qui peut ne pas inclure suffisamment d'exemples de lésions très petites. La détection de structures de très petite taille

nécessite une précision spatiale élevée, qui peut être compromise par la résolution réduite et les artefacts d'imagerie.

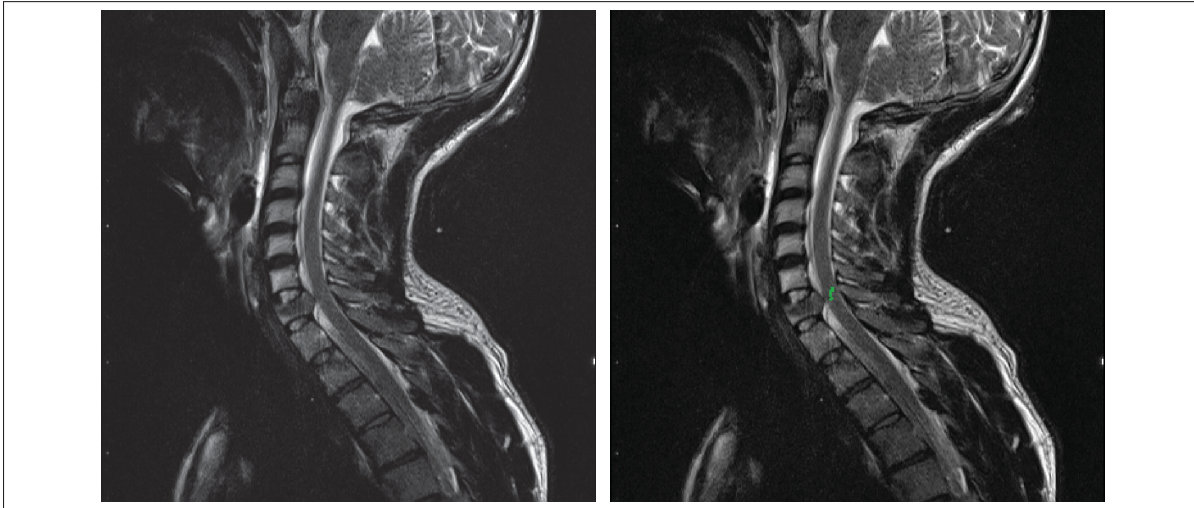


Figure 3.1 Cas 1 - Images sagittales montrant la résolution réduite (384×384), les artefacts d'imagerie (lignes verticales) et la petite taille de la lésion œdémateuse par rapport à la moelle épinière

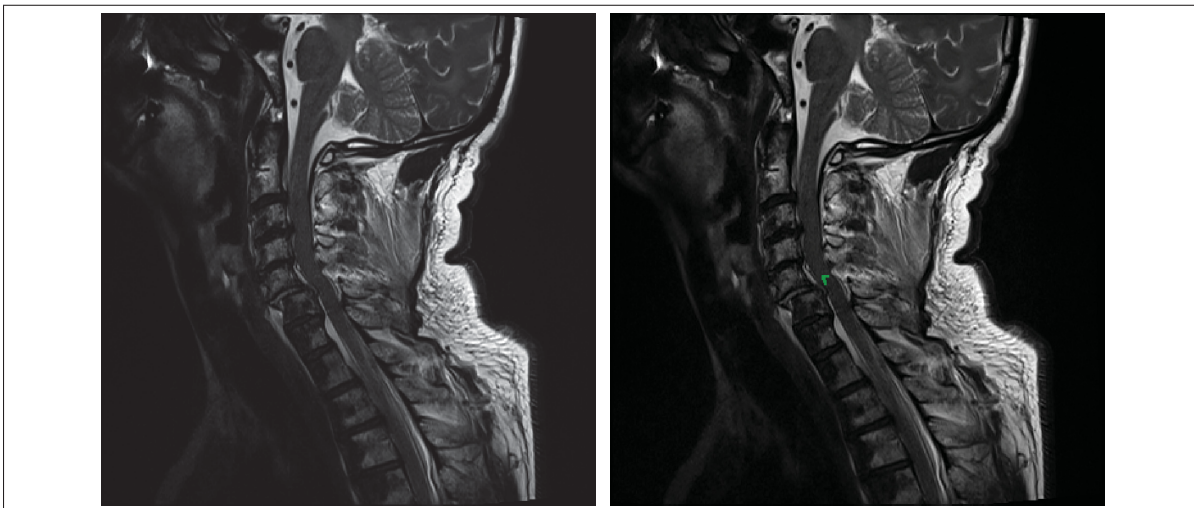


Figure 3.2 Cas 25 - Images sagittales illustrant la résolution réduite (384×384) et la petite taille de la lésion œdémateuse

Le cas 89 présente une caractéristique particulière qui peut affecter les performances du modèle (Figure 3.3). La zone de lésion (œdème) est située dans la partie inférieure de la moelle épinière, et le cadre de l'image est "coupé" ou plus court que les autres images de la cohorte. Cette

configuration signifie que bien qu'il y ait de la moelle épinière saine visible au-dessus de la lésion, il n'en reste que très peu en dessous car la partie inférieure de la moelle est tronquée par la limite de l'image, contrairement à la majorité des images d'entraînement.

Le modèle apprend, à partir de la distribution majoritaire des données d'entraînement, qu'il devrait y avoir des parties de moelle saine au-dessus et au-dessous du niveau de la lésion. Lorsque le contexte spatial attendu n'est pas entièrement présent, comme dans le cas 89, le modèle peut avoir des difficultés à identifier correctement la lésion.

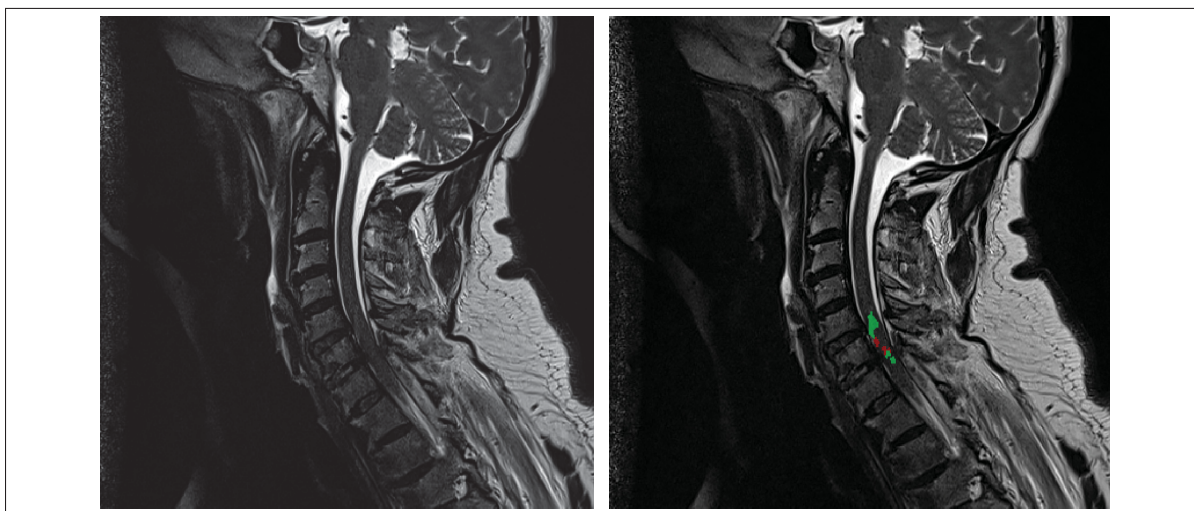


Figure 3.3 Cas 89 - Images sagittales montrant un cadre d'image incomplet avec la lésion située dans la partie inférieure de la moelle épinière, limitant le contexte de moelle saine visible en dessous de la lésion

Le cas 136 représente l'exemple le plus extrême de dégradation de qualité d'image (Figure 3.4). La résolution sagittale est extrêmement faible ($240 \times 320 \times 15$ voxels), soit environ la moitié de la résolution standard. Cette réduction drastique de résolution rend l'image très pixélisée dans le plan sagittal, compromettant significativement la qualité des détails anatomiques.

De plus, ce cas présente des artefacts d'imagerie importants qui dégradent encore davantage la qualité de l'image. Ces artefacts, combinés à la très faible résolution, rendent la segmentation particulièrement difficile, même pour le modèle sagittal qui est directement entraîné sur ces images.

Dans le plan axial, la résolution n'est pas non plus optimale, et on peut observer que même le masque de vérité terrain (ground truth) est décalé par rapport à l'emplacement réel de la moelle épinière (Figure 3.5). Ce décalage illustre les défis associés à la transformation des annotations depuis l'espace sagittal vers l'espace axial lorsque la qualité des images source est compromise.



Figure 3.4 Cas 136 - Images sagittales montrant une résolution extrêmement faible ($240 \times 320 \times 15$) et des artefacts d'imagerie importants, rendant l'image très pixélisée

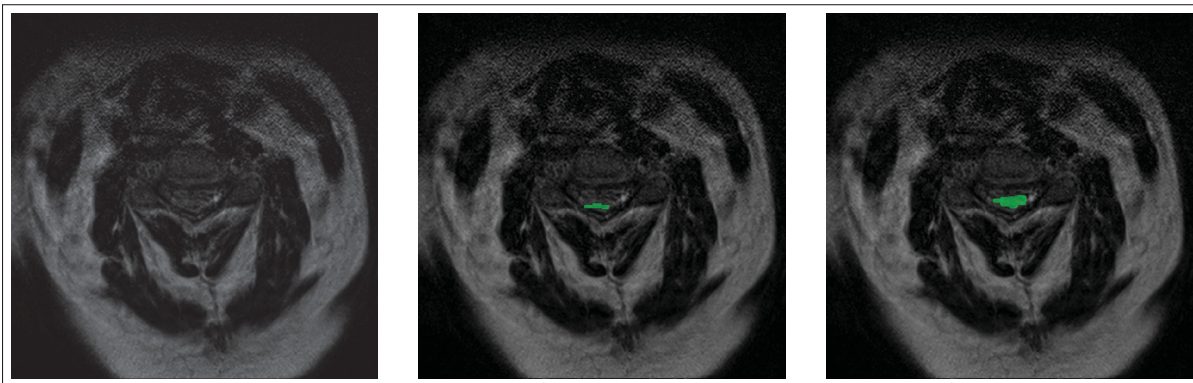


Figure 3.5 Cas 136 - Images axiales montrant un décalage du masque de vérité terrain par rapport à l'emplacement réel de la moelle épinière, illustrant les défis de transformation pour les images de faible qualité

Comme décrit dans la section 2.2.3, les annotations manuelles réalisées sur les images sagittales sont transformées vers l'espace axial à l'aide d'un enregistrement rigide utilisant une métrique d'information mutuelle. Cette transformation affine, bien qu'essentielle pour l'entraînement du

modèle axial, introduit des erreurs d'alignement qui sont amplifiées lorsque les images source présentent une résolution réduite.

Pour les images sagittales de basse résolution, la précision de l'enregistrement est limitée par la quantité d'information spatiale disponible. Les erreurs d'alignement résultantes se propagent lors de la transformation des masques d'annotation vers l'espace axial, conduisant à des décalages entre les annotations transformées et l'emplacement réel des lésions dans les images axiales. Ces erreurs d'alignement affectent non seulement la qualité des annotations de référence utilisées pour l'entraînement du modèle axial, mais également la précision des prédictions du modèle sagittal qui sont transformées et utilisées comme canal d'entrée supplémentaire.

Dans les cas où la résolution sagittale est très faible, comme le cas 136, ces erreurs d'alignement peuvent être suffisamment importantes pour que les masques transformés soient visuellement décalés par rapport aux structures anatomiques réelles, comme observé dans la Figure 3.5. Cette limitation souligne l'importance de maintenir une résolution d'acquisition suffisante pour garantir la qualité des annotations et la précision des transformations entre espaces d'imagerie. Des cas difficiles supplémentaires (Dice < 0,2) sont présentés en annexe V.

3.1.2.3 Comparaison avec l'évaluation manuelle

Au-delà des métriques quantitatives, une évaluation qualitative visuelle a été réalisée sur les cas présentant les meilleurs scores de segmentation afin d'analyser la pertinence clinique des prédictions (hypothèse H1). Cette analyse permet de vérifier si le modèle capture correctement la topologie complexe des lésions de la moelle épinière.

Le cas 87 illustre la capacité du modèle à appréhender des architectures lésionnelles hétérogènes (Figure 3.6). Sur les coupes sagittales, on observe une présentation multifocale de l'hémorragie. Le modèle réussit remarquablement à identifier quatre zones distinctes d'hémorragie (en rouge) séparées par des plages d'œdème (en vert). Il détecte avec précision non seulement les foyers hémorragiques majeurs situés près de l'épicentre traumatique, mais également les foyers plus distaux et punctiformes, rappelant l'aspect de pétéchies hémorragiques fréquemment observées

en clinique. Cette performance indique que le modèle ne se limite pas à une segmentation focale, mais a acquis une compréhension globale du contexte lésionnel, lui permettant de segmenter des zones discontinues.

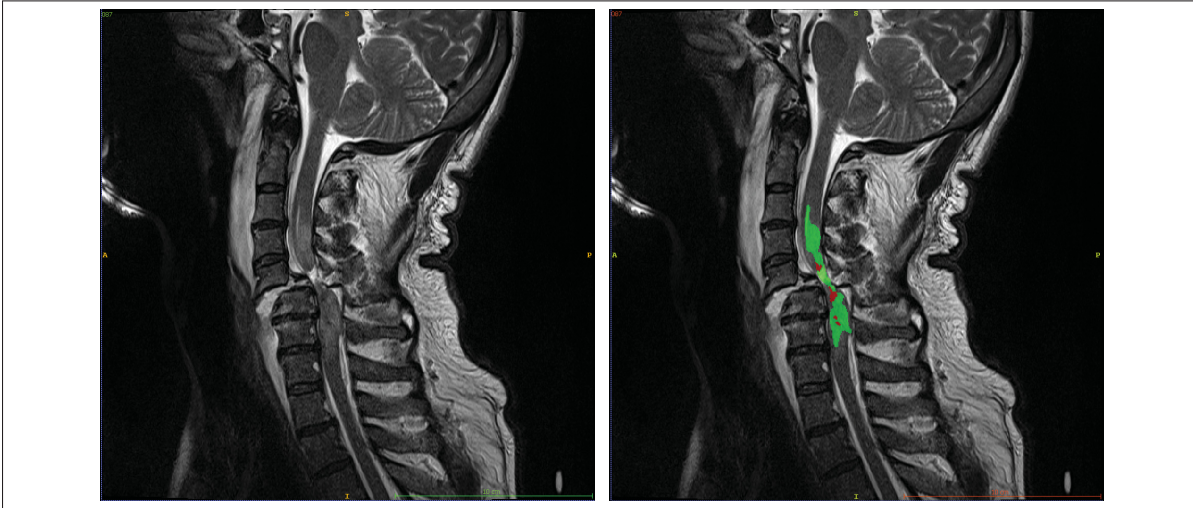


Figure 3.6 Cas 87 - Segmentation sagittale montrant la détection précise de quatre foyers d'hémorragie distincts (rouge), incluant des zones punctiformes de type pétéchiial, alternant avec de l'œdème (vert)

Dans le plan axial pour ce même patient (Figure 3.7), la délimitation entre l'œdème périphérique et le cœur hémorragique est globalement bien respectée. Cependant, une inspection minutieuse révèle que le modèle a omis une petite zone hémorragique située dans la région dorso-latérale droite de la moelle. Cette omission, absente sur la prédiction sagittale, est probablement attribuable aux artefacts et pertes d'information induits par la transformation affine nécessaire pour projeter les prédictions dans l'espace axial.

Le cas 114 présente un phénotype lésionnel différent, caractérisé par une zone d'hémorragie centrale étendue (hématomyélie) entourée d'œdème, une configuration classique dans les contusions médullaires sévères (Figure 3.8). Sur les coupes sagittales, le modèle identifie correctement la présence de cette hémorragie centrale. Toutefois, on note une discordance volumétrique : le modèle tend à sous-estimer l'étendue réelle de l'hémorragie au profit de l'œdème environnant. Cliniquement, la zone hémorragique réelle est plus vaste que celle prédite.

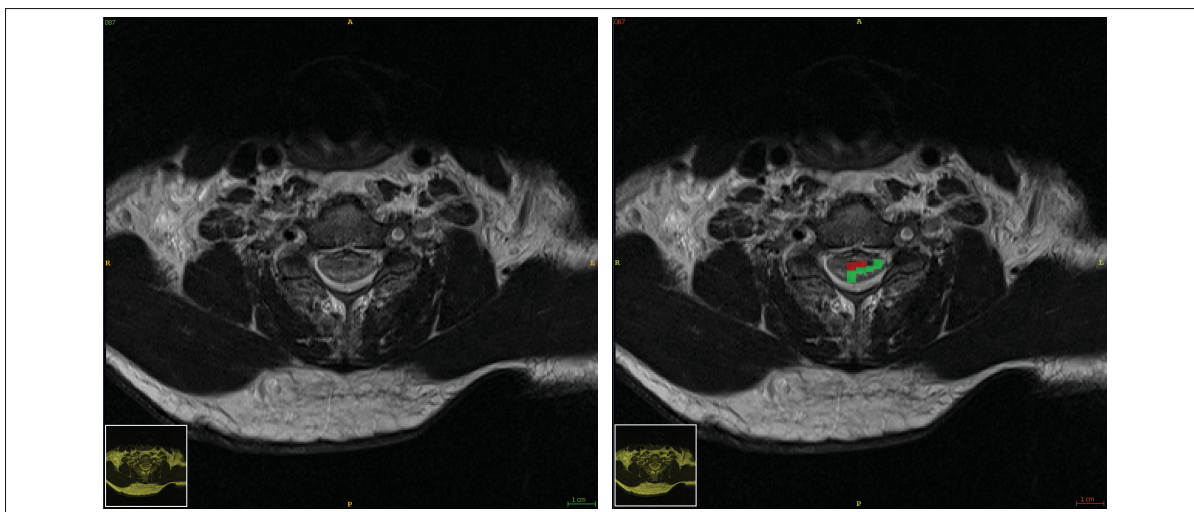


Figure 3.7 Cas 87 - Segmentation axiale. Bien que la distinction œdème/hémorragie soit correcte, une petite zone hémorragique (visible en haut à droite de la moelle) n'a pas été détectée, illustrant les limites de la projection axiale

Ce biais de segmentation s'explique probablement par le déséquilibre de classe inhérent aux données d'apprentissage. L'œdème étant volumétriquement prépondérant et plus fréquent que l'hémorragie dans la cohorte d'entraînement, le modèle adopte un comportement conservateur, favorisant la classe majoritaire (œdème) dans les zones de transition incertaines.

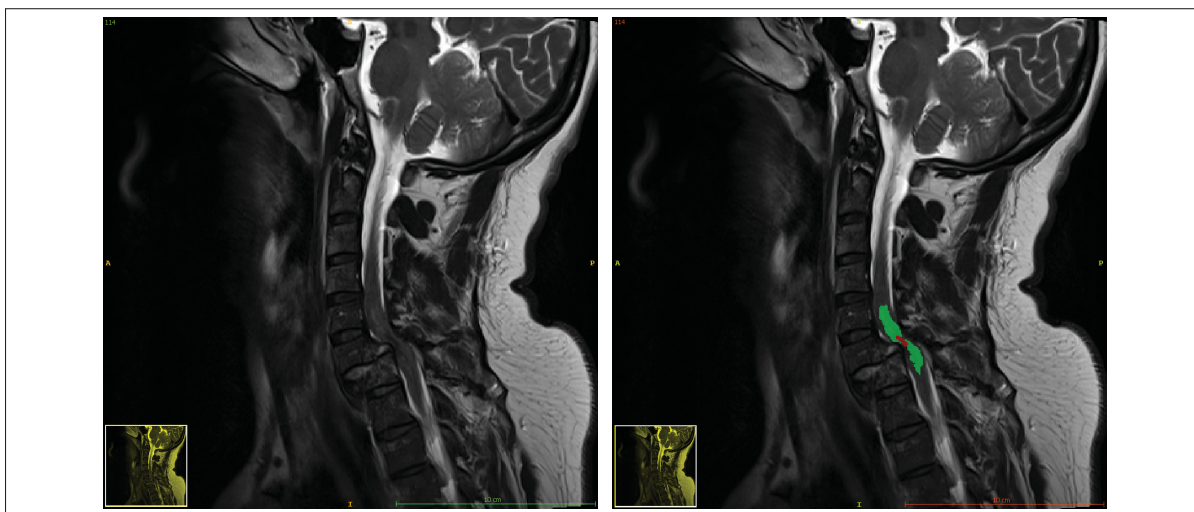


Figure 3.8 Cas 114 - Segmentation sagittale montrant une lésion avec hémorragie centrale. On note une sous-estimation de l'extension de l'hémorragie (rouge) et une sur-estimation correspondante de l'œdème (vert), liée au déséquilibre de classe

Cette tendance à la sous-estimation de l'hémorragie se confirme dans le plan axial (Figure 3.9). De manière générale, dans les deux cas analysés, la précision anatomique et la fidélité des contours lésionnels apparaissent supérieures dans le plan sagittal comparativement au plan axial, confirmant l'impact de la résolution native et des transformations géométriques sur la qualité finale de la segmentation.

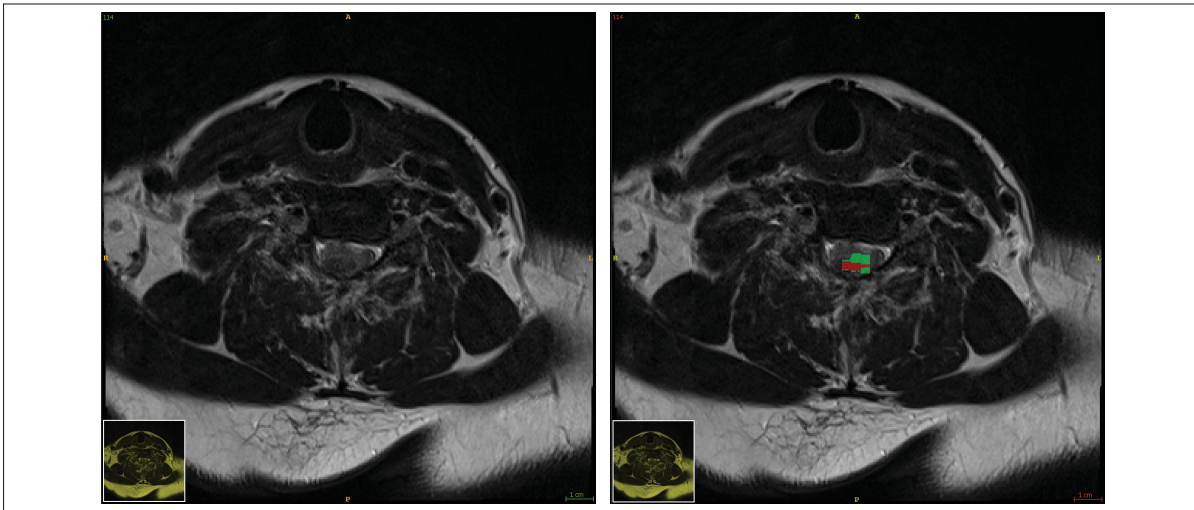


Figure 3.9 Cas 114 - Segmentation axiale confirmant la sous-estimation de la zone hémorragique centrale observée en sagittal

3.2 Explicabilité des résultats de segmentation

Pour comprendre les décisions prises par notre modèle de segmentation, nous avons utilisé l'approche SHAP (SHapley Additive exPlanations) avec l'explainer DeepExplainer. Cette méthode permet d'attribuer une valeur de contribution à chaque voxel de l'image d'entrée, indiquant dans quelle mesure ce voxel a influencé la prédiction du modèle vers une classe donnée. L'analyse a été menée sur le patient 114, un cas présentant à la fois de l'œdème et de l'hémorragie, en utilisant 40 patients de la cohorte d'entraînement comme distribution de référence (background), un nombre restreint par les limites de mémoire GPU mais suffisant pour une estimation stable.

3.2.1 Analyse des valeurs de Shapley et attribution des contributions

L'analyse quantitative des valeurs SHAP révèle des dynamiques différentes pour les deux classes lésionnelles. Pour la classe 1 (œdème), les valeurs SHAP varient de -20,92 à +60,47 (moyenne : 0,28), tandis que pour la classe 2 (hémorragie), la plage est nettement plus étendue, allant de -183,09 à +56,83 (moyenne : -0,86).

Cette différence d'échelle, particulièrement dans les valeurs négatives, suggère que le modèle utilise une stratégie de suppression beaucoup plus forte pour la classe d'hémorragie. Comme l'hémorragie est une structure plus petite et plus rare que l'œdème, le modèle semble s'appuyer davantage sur des preuves négatives (valeurs bleues très basses) pour éviter les faux positifs dans les tissus sains, alors que la détection de l'œdème repose sur un équilibre plus modéré entre contributions positives et négatives. Les valeurs maximales positives sont cependant comparables (~60 vs ~57), indiquant que l'intensité de l'évidence locale pour confirmer la présence de la lésion est similaire pour les deux classes.

3.2.2 Cartes d'activation et importance des modalités

Les cartes de chaleur SHAP générées pour le patient 114 (Figure 3.10) illustrent la distribution spatiale des contributions. Sur l'image T2, l'œdème apparaît en hyper-signal (blanc) tandis que l'hémorragie centrale se présente généralement en hypo-signal (foncé).

Visuellement, on observe que les valeurs positives (vert) sont fortement concentrées dans la région cervicale, spécifiquement dans la zone comprise entre la moelle épinière et le canal vertébral, correspondant à la localisation de la lésion. Les deux classes exploitent fortement les contours et bordures des structures anatomiques visibles sur l'IRM. Cependant, des différences spatiales importantes émergent entre les deux classes. Pour la classe 1 (œdème), la carte centrale montre des valeurs positives principalement concentrées au niveau du canal rachidien, tandis que la moelle épinière saine présente des valeurs négatives (mauve), suggérant que le modèle utilise le contraste entre le canal rachidien et la moelle épinière pour identifier l'œdème. L'échelle de visualisation de cette carte utilise le 95e percentile des valeurs absolues comme limite

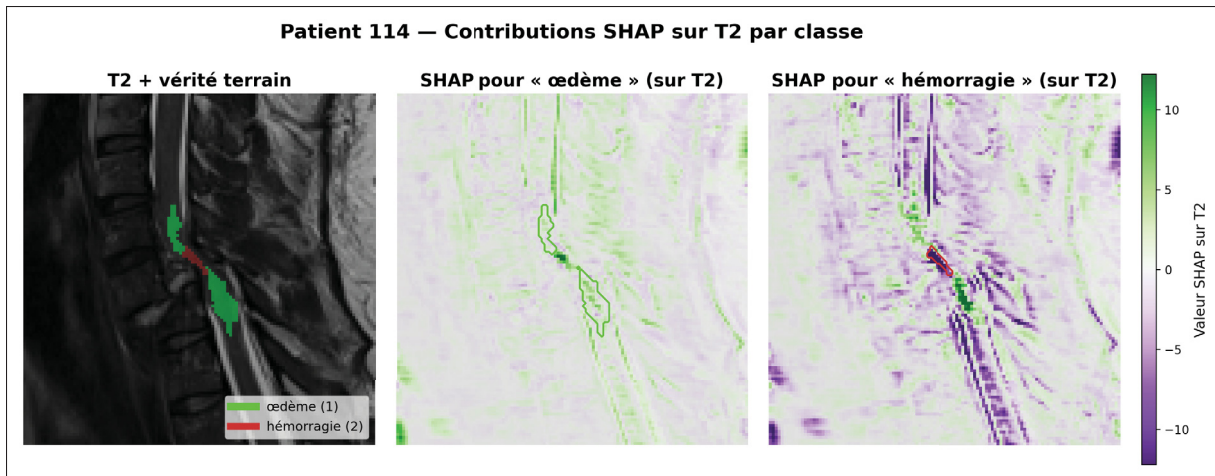


Figure 3.10 Cartes SHAP pour le patient 114, classes 1 (œdème, au milieu) et 2 (hémorragie, à droite). Pour l'œdème : valeurs positives concentrées au canal rachidien, négatives dans la moelle saine. Pour l'hémorragie : prédominance de valeurs négatives, canal rachidien négatif, valeurs positives dispersées dans la région endommagée. Échelle : 95^e percentile des valeurs absolues

maximale (et minimale en négatif), plutôt que la valeur absolue maximale. Cette approche permet d'améliorer le contraste visuel en excluant les 5% de valeurs les plus extrêmes.

Pour la classe 2 (hémorragie), la distribution des contributions est notablement différente. On observe une prédominance de valeurs négatives, avec le canal rachidien présentant des contributions négatives, contrairement à la classe 1 où il est positif. Les valeurs positives sont beaucoup plus dispersées spatialement, mais on peut néanmoins identifier quelques zones positives dans la région où le canal est endommagé. Fait notable, malgré ces différences dans la distribution des contributions, les cartes d'activation pour l'œdème et l'hémorragie mettent en évidence des régions anatomiques similaires comme étant déterminantes pour la prédiction, malgré le fait que ces deux tissus aient des apparences radiologiques opposées sur la séquence T2.

3.2.3 Interprétation des décisions du modèle

La similarité frappante entre les cartes SHAP de l'œdème et de l'hémorragie, en dépit de leurs signatures d'intensité opposées (hyper-signal vs hypo-signal) et des différences dans la distribution des contributions, offre un aperçu précieux du fonctionnement du modèle. Cette observation suggère que le réseau de neurones ne se base pas uniquement sur l'intensité locale des pixels pour segmenter l'hémorragie, mais qu'il exploite fortement les caractéristiques contextuelles.

Dans les lésions traumatiques de la moelle épinière, l'hémorragie (cœur de la lésion) est presque systématiquement entourée d'une zone d'œdème plus vaste. Il est donc probable que le modèle ait appris à utiliser la présence d'un large hyper-signal (l'œdème) comme un "proxy" ou un indicateur contextuel fort pour localiser le complexe lésionnel dans son ensemble. L'œdème agit comme un "phare" (beacon) : sa détection facile permet au modèle de focaliser son attention sur la zone lésée. Une fois la région identifiée grâce au fort signal de l'œdème, le modèle infère la présence probable d'hémorragie au sein de ce complexe. Cette dépendance au contexte explique pourquoi les régions anatomiques similaires contribuent à la détection des deux classes, même si les mécanismes diffèrent : pour l'œdème, le canal rachidien présente des contributions positives, tandis que pour l'hémorragie, il est négatif, suggérant une stratégie d'exclusion des régions saines plutôt qu'une identification directe. Cette dépendance au contexte explique également pourquoi la segmentation de l'hémorragie est plus difficile et présente des performances inférieures (Dice 0,135), car le modèle doit déduire sa présence à partir de caractéristiques environnementales plutôt que de s'appuyer exclusivement sur un contraste local distinct qui pourrait être bruité ou ambigu.

3.3 Quantification des métriques pronostiques et validation clinique

3.3.1 Caractérisation des biomarqueurs axiaux par niveau vertébral

Les biomarqueurs axiaux par niveau vertébral sont extraits à partir des segmentations automatiques pour un total de 113 patients, couvrant 10 niveaux vertébraux (niveaux 1 à 10). Les niveaux vertébraux correspondent aux vertèbres cervicales C1 (niveau 1), C2 (niveau 2), C3 (niveau 3), C4 (niveau 4), C5 (niveau 5), C6 (niveau 6), C7 (niveau 7), et aux vertèbres thoraciques T1 (niveau 8), T2 (niveau 9) et T3 (niveau 10). Les métriques calculées incluent l'aire de section transversale maximale et moyenne, le volume, et le ratio hémorragie/lésion pour chaque composante lésionnelle (œdème et hémorragie) à chaque niveau vertébral, comme décrit dans la section 2.5.1.

Tableau 3.4 Statistiques descriptives des biomarqueurs axiaux par niveau vertébral pour l'ensemble de la cohorte (n = 113 patients, 746 observations patient-niveau)

Biomarqueur	Moyenne \pm ÉT	Médiane	Min	Max	Obs. non nulles
Aire max œdème (mm ²)	27,53 \pm 15,64	27,23	0,46	74,67	330/746 (44,2%)
Aire max hémorragie (mm ²)	9,38 \pm 6,26	9,49	1,58	23,73	25/746 (3,4%)
Volume œdème (mm ³)	266,97 \pm 212,92	225,01	1,83	911,25	330/746 (44,2%)
Volume hémorragie (mm ³)	45,30 \pm 38,06	37,97	5,70	150,36	25/746 (3,4%)
Ratio hémorragie/lésion	0,0870 \pm 0,0719	0,0556	0,0081	0,3220	25/746 (3,4%)

Note : ÉT = écart-type. Les statistiques sont calculées uniquement sur les observations non nulles (combinaisons patient-niveau où la lésion est présente). Le ratio hémorragie/lésion est défini comme volume hémorragie / (volume œdème + volume hémorragie). Sur l'ensemble des 746 observations, 330 (44,2%) présentent de l'œdème et 25 (3,4%) contiennent de l'hémorragie.

Sur l'ensemble des 746 combinaisons patient-niveau analysées, 330 (44,2%) présentent de l'œdème, tandis que seulement 25 (3,4%) contiennent de l'hémorragie. Cette distribution reflète la prévalence plus élevée de l'œdème par rapport à l'hémorragie dans les lésions médullaires traumatiques. Parmi les 113 patients, 18 (15,9%) présentent au moins un niveau vertébral avec de l'hémorragie détectée.

Lorsque l'œdème est présent, l'aire de section transversale maximale varie de 0,46 à 74,67 mm², avec une moyenne de 27,53 \pm 15,64 mm² et une médiane de 27,23 mm². L'aire maximale

d'hémorragie, lorsqu'elle est présente, est nettement plus faible, allant de 1,58 à 23,73 mm², avec une moyenne de $9,38 \pm 6,26$ mm² et une médiane de 9,49 mm². Le volume d'œdème par niveau vertébral, calculé uniquement pour les niveaux où l'œdème est présent, présente une distribution avec une moyenne de $266,97 \pm 212,92$ mm³ (médiane : 225,01 mm³) et des valeurs maximales atteignant 911,25 mm³. Le volume d'hémorragie, calculé uniquement pour les niveaux où l'hémorragie est présente, est quant à lui plus restreint, avec une moyenne de $45,30 \pm 38,06$ mm³ (médiane : 37,97 mm³) et un maximum de 150,36 mm³.

Le ratio hémorragie/lésion, calculé uniquement pour les niveaux où l'hémorragie est présente, présente une moyenne de $0,0870 \pm 0,0719$ (médiane : 0,0556), avec des valeurs allant de 0,0081 à 0,3220. Cette distribution indique que lorsque l'hémorragie est présente, elle représente en moyenne environ 8,7% du volume lésionnel total, bien que cette proportion puisse varier considérablement entre les patients, allant de moins de 1% à plus de 32%.

Tableau 3.5 Distribution des biomarqueurs par niveau vertébral (statistiques calculées sur toutes les observations, incluant les valeurs nulles)

Niveau	n	Aire max œdème (mm ²)		Aire max hémorragie (mm ²)		Volume œdème (mm ³)		Volume hémorragie (mm ³)		Ratio hémorragie/lésion	
		Moy.	Méd.	Moy.	Méd.	Moy.	Méd.	Moy.	Méd.	Moy.	Méd.
1 (C1)	31	3,01	0,00	0,00	0,00	33,28	0,00	0,00	0,00	0,0000	0,0000
2 (C2)	73	2,76	0,00	0,00	0,00	23,60	0,00	0,00	0,00	0,0000	0,0000
3 (C3)	104	11,22	0,00	0,12	0,00	104,83	0,00	0,44	0,00	0,0000	0,0000
4 (C4)	112	21,49	21,45	0,23	0,00	225,77	144,47	1,04	0,00	0,0000	0,0000
5 (C5)	113	22,89	22,78	0,75	0,00	228,40	177,69	3,93	0,00	0,0100	0,0000
6 (C6)	112	16,62	11,07	0,94	0,00	154,04	62,29	4,47	0,00	0,0100	0,0000
7 (C7)	106	6,41	0,00	0,05	0,00	51,18	0,00	0,24	0,00	0,0000	0,0000
8 (T1)	80	1,12	0,00	0,00	0,00	8,38	0,00	0,00	0,00	0,0000	0,0000
9 (T2)	14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,0000	0,0000
10 (T3)	1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,0000	0,0000

Note : n = nombre d'observations patient-niveau pour chaque niveau vertébral. Les statistiques (moyennes et médianes) sont calculées sur toutes les observations, incluant les valeurs nulles (niveaux sans lésion). Cette approche reflète la prévalence réelle des lésions à chaque niveau dans la cohorte.

L'analyse de la distribution spatiale des lésions révèle une prédominance des volumes d'œdème aux niveaux vertébraux centraux (C4, C5 et C6), avec des volumes moyens de 225,77 mm³, 228,40 mm³ et 154,04 mm³ respectivement (médianes : 144,47 mm³, 177,69 mm³ et 62,29 mm³). Cette distribution correspond à la localisation typique des lésions médullaires traumatiques, qui affectent préférentiellement les segments cervicaux moyens. Les niveaux

vertébraux périphériques (C1-C2 et T1-T3) présentent des volumes moyens nettement plus faibles, allant de 0,00 à 33,28 mm³, reflétant la rareté des lésions à ces niveaux.

La distribution de l'hémorragie suit une tendance similaire, bien que sa prévalence soit beaucoup plus faible. L'hémorragie est principalement présente aux niveaux C3-C7, avec des volumes moyens variant de 0,44 mm³ (C3) à 4,47 mm³ (C6). Le niveau C6 présente le volume moyen d'hémorragie le plus élevé (4,47 mm³), suivi du niveau C5 (3,93 mm³). Les niveaux C1-C2 et T1-T3 ne présentent aucune hémorragie dans cette cohorte.

Le ratio hémorragie/lésion, calculé sur toutes les observations, est très faible à tous les niveaux (moyennes $\leq 0,01$), reflétant la rareté de l'hémorragie dans cette cohorte. Les valeurs non nulles sont observées uniquement aux niveaux C5 et C6, avec des moyennes de 0,01 dans les deux cas.

Ces biomarqueurs quantitatifs par niveau vertébral fournissent une caractérisation détaillée de la distribution spatiale des lésions médullaires, permettant une analyse fine de la topographie lésionnelle et de ses variations entre patients. La quantification séparée de l'œdème et de l'hémorragie offre la possibilité d'évaluer la contribution spécifique de chaque composante lésionnelle aux mesures pronostiques, comme décrit dans la section 2.5.1.

3.3.2 Prédiction du pronostic neurologique

Pour évaluer si l'identification et la quantification séparées de l'œdème et de l'hémorragie améliorent la prédiction du pronostic neurologique par rapport à une approche de lésion complète (hypothèse H3), nous avons analysé la relation entre les biomarqueurs IRM et les scores AIS à différentes périodes de suivi. L'analyse a été réalisée sur un sous-ensemble de 39 patients pour lesquels les données cliniques longitudinales et les biomarqueurs d'imagerie étaient disponibles.

Les résultats révèlent un pattern temporel marqué dans la valeur prédictive des biomarqueurs différenciés. En post-opératoire immédiat ($n = 39$), les trois modèles ont un pouvoir prédictif limité et similaire (R^2 ajusté de 0,05 à 0,10 pour l'amélioration AIS), et l'ajout des biomarqueurs différenciés (M3) n'apporte qu'une amélioration marginale non significative (gain = +0,03, $p =$

Tableau 3.6 Comparaison des modèles de régression pour la prédiction de l'amélioration AIS (delta = score à la période – AIS préopératoire) selon la période de suivi

Période	n	R ² M1 (Clinique)	R ² M2 (Lésion complète)	R ² M3 (Différencié)	Gain R ² (M3–M2)	Test F (p)
Post-opératoire	39	0,05	0,06	0,10	+0,03	0,258
3 mois	20	0,04	0,14	0,45	+0,30	0,028*
6 mois	17	-0,07	-0,14	0,70	+0,84	<0,001***
1 an	22	-0,04	-0,10	0,56	+0,66	<0,001***

Note : R² ajusté = coefficient de détermination ajusté. M1 = modèle clinique (AIS préopératoire uniquement). M2 = modèle lésion complète (AIS préopératoire + volume total). M3 = modèle différencié (AIS préopératoire + volume œdème + volume hémorragie + présence hémorragie + ratio hémorragie/lésion). Le test F évalue si l'amélioration de M3 par rapport à M2 est statistiquement significative. *p < 0,05 ; ***p < 0,001. Les coefficients de régression, la matrice de corrélation entre biomarqueurs et le détail du test F sont reportés en annexe VI.

0,258). Le score neurologique préopératoire n'explique donc qu'une faible part de la variance de l'amélioration à court terme.

En revanche, pour les périodes de suivi à moyen et long terme, les biomarqueurs différenciés montrent une valeur prédictive substantielle et statistiquement significative. À 3 mois (n = 20), le modèle différencié M3 atteint un R² ajusté de 0,45, soit un gain de +0,30 par rapport au modèle de lésion complète M2 (p = 0,028). À 6 mois (n = 17), l'écart est très significatif : R² ajusté de 0,70 pour M3 contre des R² négatifs pour M1 et M2, soit un gain de +0,84 (p < 0,001). À 1 an (n = 22), le modèle M3 conserve un R² ajusté de 0,56, avec un gain de +0,66 par rapport à M2 (p < 0,001).

Les valeurs de R² négatives observées pour les modèles M1 et M2 aux périodes de 6 mois et 1 an indiquent que ces modèles simples performant moins bien qu'une prédiction par la moyenne, suggérant que le score AIS préopératoire et le volume total perdent leur pouvoir prédictif pour le pronostic à long terme. En revanche, la quantification différenciée de l'œdème et de l'hémorragie capture des informations pronostiques pertinentes qui se manifestent principalement dans l'évolution neurologique à moyen et long terme.

Ces résultats sont cohérents avec l'hypothèse H3 (valeur pronostique des biomarqueurs différenciés à moyen et long terme). La distinction entre œdème et hémorragie, rendue

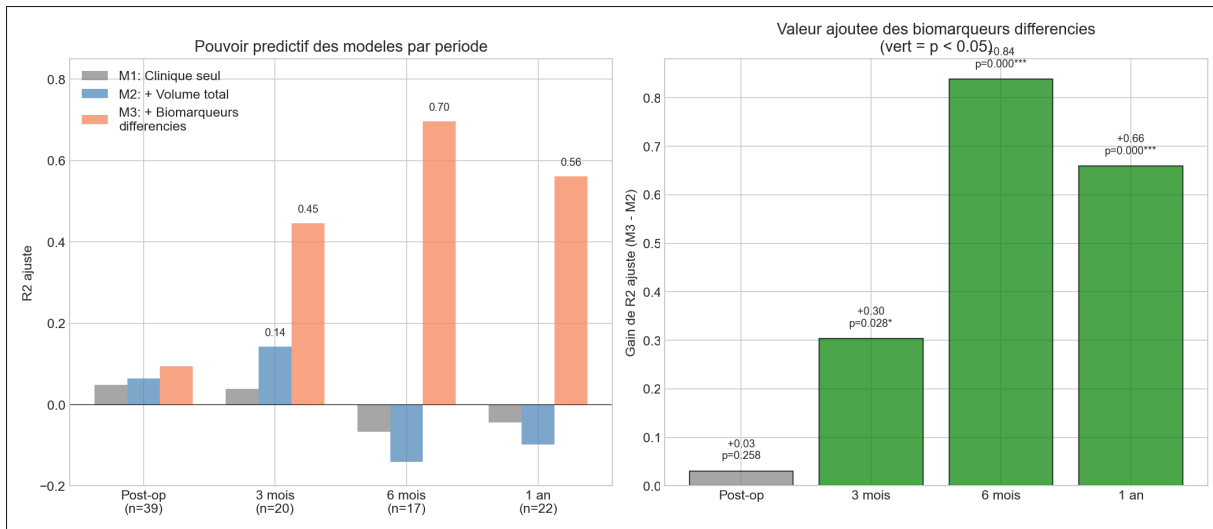


Figure 3.11 Comparaison du R^2 ajusté des trois modèles de régression selon la période de suivi. Le modèle différencié M3 (biomarqueurs séparés) surpasse significativement les modèles M1 (clinique) et M2 (lésion complète) pour la prédiction du pronostic à moyen et long terme (3 mois, 6 mois, 1 an), alors que les trois modèles présentent des performances similaires en post-opératoire immédiat

possible par notre approche de segmentation différenciée, fournit des biomarqueurs dont la valeur pronostique se révèle principalement au-delà de la phase aiguë. Cette observation est cliniquement cohérente : l'œdème représente une composante potentiellement réversible de la lésion, tandis que l'hémorragie reflète des dommages tissulaires irréversibles. Leur quantification séparée permet donc de mieux estimer le potentiel de récupération neurologique à long terme que la simple mesure du volume lésionnel global.

Ces résultats doivent cependant être interprétés avec prudence en raison de plusieurs limitations. Premièrement, la taille d'échantillon diminue considérablement avec le temps de suivi : de 39 patients en post-opératoire immédiat à seulement 17 patients à 6 mois, ce qui limite la puissance statistique et la généralisabilité des résultats pour les périodes de suivi les plus longues. Cette attrition reflète les difficultés inhérentes au suivi longitudinal des patients traumatisés médullaires. Deuxièmement, les biomarqueurs utilisés dans cette analyse sont dérivés des segmentations automatiques, dont les performances pour l'hémorragie sont limitées (Dice : 0,135, section 3.1.2.1). Les erreurs de segmentation, notamment la sous-estimation systématique

de l'hémorragie observée dans l'analyse des cas difficiles (section 3.1.2.3), se propagent directement dans les métriques quantitatives et peuvent affecter la précision des modèles prédictifs. Une validation sur une cohorte indépendante plus large, avec des segmentations manuelles de référence, serait nécessaire pour confirmer ces résultats préliminaires.

3.3.3 Système de visualisation et aide à la décision clinique

Pour faciliter l'interprétation des biomarqueurs quantitatifs, nous avons développé un prototype de visualisation permettant de lier les métriques calculées aux images IRM axiales et aux segmentations. Ce prototype offre une interface de navigation entre les tranches axiales avec affichage simultané des métriques par tranche et par niveau vertébral.

Le choix de la visualisation dans le plan axial repose sur plusieurs avantages fondamentaux pour l'évaluation clinique des lésions médullaires. Le plan axial offre la perspective la plus directe sur l'anatomie de la moelle épinière, permettant aux cliniciens d'identifier précisément les structures médullaires affectées, notamment la distinction entre substance grise et substance blanche. Cette orientation permet également d'évaluer la déformation ou la compression de la moelle épinière, ainsi que de localiser la lésion par rapport aux niveaux vertébraux.

La Figure 3.12 illustre l'interface de visualisation pour le patient 114, tranche axiale 14, correspondant au niveau vertébral C5. L'image T2 axiale centrale montre la moelle épinière en coupe transversale, avec les contours de lésion superposés : l'œdème apparaît en vert (contour extérieur) et l'hémorragie en rouge (zone centrale). Le panneau d'information à droite affiche les métriques quantitatives calculées pour cette tranche spécifique et pour l'ensemble du niveau vertébral C5.

Pour cette tranche spécifique, les métriques révèlent une lésion importante avec une aire totale de 58,22 mm², composée de 34,49 mm² d'œdème et de 23,73 mm² d'hémorragie. Le ratio hémorragie/lésion de 0,4076 (40,76%) indique une proportion significative d'hémorragie dans cette tranche, ce qui est cohérent avec la présence d'une hématomyélie centrale visible sur l'image. Le ratio de dommage de 0,7603 (76,03%) indique que la lésion occupe plus des trois

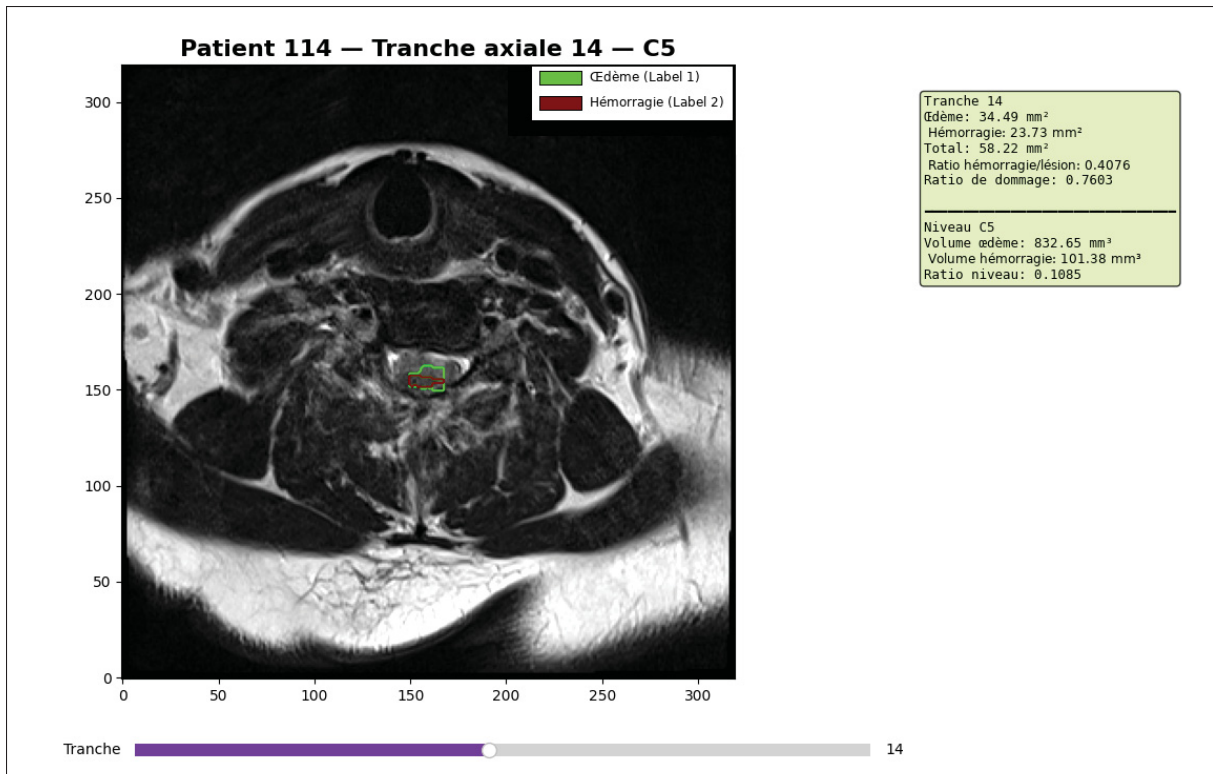


Figure 3.12 Interface de visualisation axiale pour le patient 114, tranche 14 (niveau C5).

L'image T2 axiale (centre) montre la moelle épinière avec les contours de lésion superposés : œdème en vert, hémorragie en rouge. Le panneau de droite affiche les métriques par tranche (aire d'œdème : 34,49 mm², aire d'hémorragie : 23,73 mm², aire totale : 58,22 mm², ratio hémorragie/lésion : 0,4076, ratio de dommage : 0,7603) et les métriques agrégées par niveau vertébral (volume œdème : 832,65 mm³, volume hémorragie : 101,38 mm³, ratio niveau : 0,1085)

quarts de la surface du niveau vertébral C5 dans cette tranche, reflétant l'étendue importante de la lésion.

Les métriques agrégées pour l'ensemble du niveau C5 montrent un volume total d'œdème de 832,65 mm³ et un volume d'hémorragie de 101,38 mm³, avec un ratio hémorragie/lésion de 0,1085 (10,85%) au niveau vertébral. Cette différence entre le ratio par tranche (40,76%) et le ratio par niveau (10,85%) illustre la variabilité spatiale de la composition lésionnelle : la tranche 14 présente une concentration particulièrement élevée d'hémorragie, tandis que les autres tranches du niveau C5 contiennent principalement de l'œdème.

Bien que la visualisation axiale bidimensionnelle offre une perspective précise et cliniquement pertinente, l'extension vers une visualisation tridimensionnelle (3D) pourrait apporter des bénéfices supplémentaires pour la compréhension de la topographie lésionnelle et l'aide à la décision clinique.

3.4 Limitations et perspectives

3.4.1 Limitations de la segmentation automatique

Il est important de noter que la comparaison des performances entre notre modèle et SciSeg a été effectuée sur la même cohorte de patients que celle utilisée pour l'entraînement de notre modèle en validation croisée 5-fold. Bien que la validation croisée garantisse que chaque patient n'est évalué que par des modèles qui n'ont pas été entraînés sur ses données spécifiques, il reste que notre modèle a été optimisé sur cette cohorte particulière, ce qui introduit un biais en sa faveur lors de la comparaison avec SciSeg. SciSeg, quant à lui, a été développé et validé sur des cohortes différentes, ce qui peut expliquer en partie les différences de performance observées. Une comparaison plus équitable nécessiterait l'évaluation des deux méthodes sur une cohorte de test complètement indépendante, n'ayant été utilisée ni pour l'entraînement de notre modèle ni pour celui de SciSeg.

Comme discuté en section 3.1.2.2, la transformation des annotations depuis l'espace sagittal vers l'espace axial introduit des erreurs d'alignement (amplifiées en cas de résolution réduite ou d'artéfacts), affectant la qualité des références et pouvant conduire à des décalages visibles (cas 136, Figure 3.AB).

Le déséquilibre de classe (hémorragie rare : 3,4 % des observations, 172 coupes sur 502) et la tendance à sous-estimer l'hémorragie au profit de l'œdème, déjà évoqués en 3.1.2.1 et 3.1.2.3, contribuent aux performances différenciées entre classes (Dice 0,135 vs 0,557).

La qualité variable des images (résolutions réduites, artefacts, cadres tronqués), illustrée par les cas difficiles en 3.1.2.2, compromet la segmentation et la précision de l'enregistrement sagittal-axial.

Enfin, la dépendance au contexte pour la détection de l'hémorragie (section 3.2.3) rend sa segmentation plus difficile, le modèle inférant l'hémorragie à partir du contexte plutôt qu'à partir d'un contraste local direct.

Par ailleurs, l'architecture en cascade sagittal \rightarrow axial soulève une question méthodologique concernant le rôle du masque sagittal transformé fourni en entrée au modèle axial. Bien que ce masque ne constitue qu'un canal parmi d'autres et que l'image T2 axiale demeure l'entrée principale, il est légitime de se demander si le modèle axial n'apprend pas simplement à reproduire le masque sagittal transformé plutôt qu'à exploiter véritablement l'information contenue dans les images axiales. De plus, lors de l'entraînement, les masques sagittaux fournis au modèle axial sont générés par un modèle sagittal qui a été entraîné sur les mêmes patients (dans le même *fold*). Ces masques sont donc plus précis que ceux qu'on obtiendrait sur un patient inconnu en conditions réelles, ce qui pourrait conduire le modèle axial à surestimer la fiabilité de ce canal d'entrée. Une stratégie de génération *out-of-fold* des masques sagittaux, où chaque masque serait produit par un modèle n'ayant jamais vu le patient correspondant, permettrait de mieux simuler les conditions d'inférence réelles et de réduire ce biais potentiel.

3.4.2 Limitations de la quantification et de l'explicabilité

Concernant l'analyse d'explicabilité, l'analyse SHAP a été réalisée en utilisant seulement 40 patients de la cohorte d'entraînement comme distribution de référence (background), un nombre restreint par les limites de mémoire GPU (section 3.2). Bien que ce nombre soit suffisant pour une estimation stable, une analyse plus complète avec un échantillon de référence plus large pourrait fournir des aperçus supplémentaires sur le comportement du modèle.

En ce qui concerne la quantification des biomarqueurs, les métriques calculées sont directement dépendantes de la qualité des segmentations automatiques. Par conséquent, les erreurs de

segmentation, telles que la sous-estimation de l'hémorragie ou les erreurs de délimitation, se propagent directement dans les métriques quantitatives (volumes, aires, ratios), affectant leur précision et leur fiabilité pour l'évaluation pronostique. De plus, la quantification est effectuée à partir des segmentations axiales, qui sont obtenues après transformation depuis l'espace sagittal, introduisant potentiellement des erreurs supplémentaires liées à l'enregistrement et à la transformation géométrique (section 3.1.2.2).

Une autre limitation réside dans la variabilité spatiale importante observée au sein d'un même niveau vertébral. Comme illustré dans l'exemple du patient 114 (section 3.3.3), le ratio hémorragie/lésion peut varier considérablement entre les tranches individuelles (40,76% pour la tranche 14) et l'agrégation au niveau vertébral (10,85% pour l'ensemble du niveau C5). Cette variabilité soulève des questions sur la représentativité des métriques agrégées par niveau et sur la meilleure échelle d'analyse pour l'évaluation pronostique.

Enfin, la quantification des métriques pour l'hémorragie est basée sur un nombre très limité d'observations (seulement 25 combinaisons patient-niveau sur 746, soit 3,4% selon le Tableau 3.4). Cette rareté limite la fiabilité statistique des métriques calculées pour cette classe et peut introduire des biais dans les analyses pronostiques basées sur ces biomarqueurs, particulièrement pour les niveaux vertébraux où l'hémorragie est absente ou très rare.

3.4.3 Perspectives d'amélioration

Plusieurs axes d'amélioration peuvent être identifiés à partir des limitations observées. Premièrement, l'amélioration de la généralisabilité du modèle pourrait être obtenue par l'augmentation de la taille et de la diversité de la cohorte d'entraînement. L'ajout de patients provenant de différents sites d'acquisition permettrait de mieux capturer la variabilité inter-institutionnelle des protocoles d'imagerie et des caractéristiques des images, améliorant ainsi la robustesse du modèle. L'inclusion d'un plus grand nombre de cas présentant de l'hémorragie permettrait d'améliorer les performances de segmentation pour cette classe. L'augmentation de la diversité des données d'entraînement pour inclure davantage d'exemples avec des résolutions variées et

des artefacts d'imagerie, ainsi que l'utilisation de techniques d'augmentation de données qui simulent ces variations, contribueraient également à améliorer la robustesse du modèle face aux cas difficiles.

Deuxièmement, l'amélioration de la qualité des images d'entrée et de leur traitement pourrait bénéficier de l'implémentation de techniques de prétraitement spécifiques pour la correction des artefacts ou l'amélioration de la résolution, particulièrement pour les cas de faible qualité. L'amélioration des algorithmes d'enregistrement pour les images de basse résolution pourrait réduire les erreurs d'alignement lors de la transformation entre espaces sagittal et axial, limitant ainsi la propagation de ces erreurs dans les annotations de référence et les prédictions du modèle.

Troisièmement, l'analyse SHAP révèle que le modèle exploite implicitement les régions anatomiques environnantes (canal rachidien, structures vertébrales) pour ses prédictions. L'intégration explicite de ces structures anatomiques comme contraintes supplémentaires pourrait potentiellement améliorer la précision de la localisation et de la délimitation des lésions. De plus, l'analyse SHAP montre que le modèle s'appuie fortement sur le contexte plutôt que sur les caractéristiques intrinsèques de l'hémorragie (hyposignal T2) pour sa détection. L'exploration de stratégies pour mieux exploiter les caractéristiques intrinsèques de l'hémorragie pourrait réduire la dépendance au contexte et améliorer la détection directe de cette classe.

Enfin, malgré l'équilibrage des classes, le modèle présente une tendance systématique à sous-estimer la taille des lésions, particulièrement pour l'hémorragie. L'exploration de modifications des fonctions de perte, telles que l'ajustement des poids de classe ou l'utilisation de fonctions de perte adaptatives, pourrait permettre de corriger ce biais conservateur. L'analyse comparative des performances sagittales et axiales révèle que les deux vues présentent un comportement similaire de sous-estimation. L'amélioration de l'intégration multi-vues dans l'architecture en cascade pourrait potentiellement améliorer la détection des contours de lésion en exploitant la complémentarité des informations spatiales disponibles dans chaque plan.

CONCLUSION ET RECOMMANDATIONS

Le premier objectif de ce mémoire était de développer une approche automatisée de segmentation des deux composantes des lésions médullaires traumatiques, soit l'œdème et l'hémorragie. Nos recherches ont mené à une architecture en cascade de deux modèles U-Net 2D exploitant les informations complémentaires des plans sagittal et axial, avec intégration d'un terme topologique dans la fonction de perte basé sur les connaissances médicales. Cette approche s'est révélée significativement supérieure à SciSeg, l'état de l'art actuel, avec une amélioration médiane du coefficient de Dice de 0,250 dans le plan axial ($p < 0,001$) et de 0,054 dans le plan sagittal ($p < 0,001$). Le modèle parvient à distinguer les deux composantes lésionnelles, avec des performances satisfaisantes pour l'œdème (Dice : 0,557) et plus limitées pour l'hémorragie (Dice : 0,135), reflétant la rareté de cette dernière dans la cohorte. La contribution principale de ce mémoire est donc l'élaboration de cette stratégie de segmentation multiclasse en cascade.

Le deuxième objectif de ce mémoire était de quantifier et de visualiser les caractéristiques physiques et géométriques de chaque composante lésionnelle par niveau vertébral afin d'améliorer l'évaluation pronostique. À partir des segmentations automatiques, nous avons extrait des biomarqueurs quantitatifs incluant l'aire de section transversale, le volume et le ratio hémorragie/lésion pour chaque niveau vertébral, calculés sur 113 patients couvrant 10 niveaux vertébraux. Un prototype de visualisation a été développé pour lier ces métriques aux images IRM axiales et aux segmentations. L'analyse d'explicabilité SHAP a révélé que le modèle exploite les caractéristiques contextuelles plutôt que les caractéristiques intrinsèques de l'hémorragie pour sa détection. L'analyse de la valeur pronostique de ces biomarqueurs différenciés a révélé que la distinction entre œdème et hémorragie n'améliore pas significativement la prédiction du pronostic neurologique à court terme (post-opératoire immédiat), où le score AIS préopératoire demeure le principal déterminant. En revanche, pour les périodes de suivi à moyen et long terme (3 mois, 6 mois, 1 an), les biomarqueurs différenciés démontrent une valeur prédictive

substantielle et statistiquement significative, avec des gains de R^2 ajusté allant de +0,30 à +0,85 par rapport à une approche de lésion complète ($p < 0,05$).

Cette étude présente certaines limitations. La comparaison avec SciSeg présente un biais potentiel, notre modèle ayant été optimisé sur la même cohorte utilisée pour l'évaluation. Les erreurs d'alignement introduites par la transformation des annotations depuis l'espace sagittal vers l'espace axial, particulièrement pour les images de faible résolution, ainsi que le déséquilibre de classe important (l'hémorragie représente seulement 3,4% des observations) limitent les performances pour cette classe.

Les connaissances et techniques développées pourraient être étendues à d'autres contextes. L'architecture en cascade pourrait être adaptée à d'autres pathologies médullaires (sclérose en plaques, tumeurs médullaires) ou à d'autres systèmes anatomiques présentant des relations spatiales connues, comme les tumeurs cérébrales. Le terme topologique intégré dans la fonction de perte, basé sur des connaissances médicales concernant les relations spatiales, pourrait être généralisé à d'autres pathologies présentant des relations anatomiques similaires.

Des études futures pourraient explorer l'extension temporelle de cette approche en intégrant des analyses longitudinales pour suivre l'évolution des lésions médullaires au fil du temps et évaluer la réponse au traitement. L'intégration de données cliniques supplémentaires (scores neurologiques, variables démographiques) avec les biomarqueurs quantitatifs extraits pourrait permettre de développer des modèles prédictifs plus robustes pour l'évolution neurologique. L'application de cette approche à d'autres modalités d'imagerie, telles que l'imagerie par tenseur de diffusion (DTI) ou l'imagerie de perfusion, pourrait fournir des informations complémentaires sur la microstructure et la fonctionnalité des tissus médullaires. Enfin, l'adaptation de cette méthodologie à d'autres systèmes anatomiques présentant des défis similaires de segmentation multiclasse et de quantification hiérarchique ouvrirait de nouvelles perspectives pour l'application de l'apprentissage profond en imagerie médicale.

ANNEXE I

DÉTAILS TECHNIQUES DU PRÉTRAITEMENT DES DONNÉES

1. Pipeline de prétraitement nnU-Net

Le prétraitement des données a été réalisé automatiquement par le framework nnU-Net, qui adapte dynamiquement les paramètres aux caractéristiques spécifiques du jeu de données. Le pipeline comprend les étapes suivantes, exécutées séquentiellement pour chaque volume d'imagerie.

Tableau-A I-1 Paramètres de prétraitement calculés automatiquement par nnU-Net

Paramètre	Modèle sagittal	Modèle axial
Espacement cible (mm)	0,558 × 0,558	0,5625 × 0,5625
Taille des patches (pixels)	448 × 448	320 × 320
Méthode de calcul espacement	Médiane de la cohorte	Médiane de la cohorte
Interpolation images	Spline cubique (ordre 3)	Spline cubique (ordre 3)
Interpolation masques	Plus proche voisin	Plus proche voisin
Normalisation	Z-score par canal	Z-score par canal

2. Rééchantillonnage spatial

L'espace cible pour le rééchantillonnage a été déterminé en calculant la médiane des espacements de voxels sur l'ensemble de la cohorte d'entraînement. Cette approche permet d'obtenir une résolution représentative de la majorité des images tout en évitant les valeurs extrêmes.

Pour les images, une interpolation spliniale du troisième ordre (cubique) a été utilisée, offrant un bon compromis entre la préservation des détails et la suppression des artéfacts d'interpolation. Pour les masques de segmentation, une interpolation par plus proche voisin a été appliquée afin de préserver l'intégrité des labels discrets et d'éviter l'introduction de valeurs intermédiaires non valides.

3. Normalisation Z-score

La normalisation Z-score a été appliquée indépendamment à chaque canal d'entrée (T1, T2, masque de moelle pour le modèle sagittal ; T2 axial et prédiction sagittale pour le modèle axial). Pour chaque image, la transformation suivante est appliquée :

$$I_{\text{norm}} = \frac{I - \mu}{\sigma} \quad (\text{A I-1})$$

où I représente l'image originale, μ la moyenne des intensités de l'image, et σ l'écart-type. Cette normalisation centre les données autour de zéro avec une variance unitaire, facilitant la convergence du réseau lors de l'entraînement.

Tableau-A I-2 Statistiques de normalisation moyennes observées sur la cohorte

Modalité	Moyenne (μ)	Écart-type (σ)	Min normalisé	Max normalisé
T2 sagittal	162,0 ± 62,6	196,9 ± 75,0	-0,82 ± 0,10	7,1 ± 1,8
T1 sagittal	182,5 ± 63,6	214,5 ± 67,5	-1,18 ± 0,22	6,9 ± 2,1
T2 axial	129,7 ± 49,8	161,8 ± 57,5	-0,80 ± 0,10	6,5 ± 1,9

Note : Les valeurs sont présentées sous forme de moyenne ± écart-type calculées sur l'ensemble de la cohorte.

4. Recadrage (Cropping)

Le recadrage automatique identifie la boîte englobante minimale contenant toutes les régions non nulles de l'image. Cette étape réduit significativement la taille des données en éliminant les zones d'arrière-plan sans information anatomique pertinente, tout en préservant l'intégralité du contenu médical.

ANNEXE II

CONFIGURATION DÉTAILLÉE DE L'AUGMENTATION DE DONNÉES

1. Transformations géométriques

L'augmentation de données utilise la configuration agressive de nnU-Net (nnUNetTrainerDA5), qui applique une série de transformations pour augmenter artificiellement la diversité du dataset.

Tableau-A II-1 Transformations géométriques appliquées lors de l'entraînement

Transformation	Paramètres	Probabilité
Rotation (2D)	$\pm 180^\circ$	0,2
Rotation (3D)	$\pm 30^\circ$ par axe	0,2
Changement d'échelle	Facteur 0,7–1,43	0,2
Miroir horizontal	–	0,5
Miroir vertical	–	0,5
Déformation élastique	$\sigma = 0-900, \alpha = 9-13$	0,3

2. Transformations d'intensité

Tableau-A II-2 Transformations d'intensité appliquées lors de l'entraînement

Transformation	Paramètres	Probabilité
Bruit gaussien	$\sigma = 0-0,1$	0,1
Flou gaussien	$\sigma = 0,5-1,0$	0,2
Ajustement luminosité	Facteur 0,75-1,25	0,15
Ajustement contraste	Facteur 0,75-1,25	0,15
Transformation gamma	$\gamma = 0,7-1,5$	0,3
Gamma inversé	$\gamma = 0,7-1,5$	0,05
Simulation basse résolution	Facteur 0,5-1,0	0,25

3. Transformations locales

En plus des transformations globales, des modifications locales sont appliquées pour simuler des variations spatiales d'intensité :

Tableau-A II-3 Transformations locales appliquées lors de l'entraînement

Transformation	Description	Probabilité
Gradient de luminosité additif	Gradient linéaire aléatoire sur l'image	0,5
Gamma local	Transformation gamma variable spatialement	0,3
Netteté locale	Filtre de netteté avec masque aléatoire	0,2

ANNEXE III

DÉTAILS TECHNIQUES DE L'ANALYSE SHAP

1. Configuration de DeepExplainer

L'analyse d'explicabilité a été réalisée avec la bibliothèque SHAP (version 0.41.0), utilisant l'explainer DeepExplainer conçu pour les réseaux de neurones profonds.

Tableau-A III-1 Paramètres de configuration de l'analyse SHAP

Paramètre	Valeur
Explainer	DeepExplainer
Nombre de patients de référence (background)	40
Sélection des tranches	Tranche sagittale médiane par patient
Taille des images d'entrée	256 × 256 pixels
Normalisation	Z-score (identique à l'entraînement)
Classes analysées	Classe 1 (œdème), Classe 2 (nécrose)
GPU utilisé	NVIDIA Quadro RTX 6000 (24 Go VRAM)

2. Sélection des images de référence

Les 40 patients de référence ont été sélectionnés aléatoirement parmi l'ensemble d'entraînement. Pour chaque patient, la tranche sagittale médiane a été extraite, car elle contient généralement la plus grande quantité d'informations anatomiques pertinentes, incluant la moelle épinière dans son extension cranio-caudale maximale.

Le nombre de 40 patients a été déterminé comme un compromis entre :

- la stabilité des estimations SHAP (qui augmente avec le nombre d'échantillons de référence);
- les contraintes de mémoire GPU (chaque image de référence augmente la consommation mémoire).

Des tests préliminaires avec 20, 40 et 60 patients de référence ont montré une convergence satisfaisante des valeurs SHAP à partir de 40 patients.

3. Processus de calcul des valeurs SHAP

Le calcul des valeurs SHAP suit le processus suivant :

1. **Chargement du modèle** : Le modèle sagittal entraîné est chargé avec ses poids optimaux et configuré en mode évaluation (eval mode).
2. **Préparation des références** : Les 40 images de référence sont chargées, normalisées (Z-score) et redimensionnées à 256×256 pixels.
3. **Initialisation de l'explainer** : DeepExplainer est initialisé avec le modèle et les images de référence, calculant automatiquement le baseline $E[f(x)]$.
4. **Calcul des attributions** : Pour chaque image test, les valeurs SHAP sont calculées pour chaque pixel et chaque classe de sortie.
5. **Post-traitement** : Les cartes SHAP sont normalisées pour la visualisation en utilisant le 95^e percentile des valeurs absolues comme limite d'échelle.

4. Pseudocode de l'analyse SHAP

```
import shap
import torch

# 1. Charger le modèle entraîné
model = load_trained_model("sagittal_model_weights.pth")
model.eval()

# 2. Préparer les images de référence (background)
background_patients = random_sample(training_patients, n=40)
background_images = []
for patient in background_patients:
    slice_median = extract_median_sagittal_slice(patient)
    slice_normalized = zscore_normalize(slice_median)
    slice_resized = resize(slice_normalized, (256, 256))
    background_images.append(slice_resized)
background_tensor = torch.stack(background_images)

# 3. Initialiser DeepExplainer
explainer = shap.DeepExplainer(model, background_tensor)

# 4. Calculer les valeurs SHAP pour une image test
test_image = preprocess_test_image(patient_114)
shap_values = explainer.shap_values(test_image)

# shap_values[0] : contributions pour la classe 0 (fond)
# shap_values[1] : contributions pour la classe 1 (œdème)
# shap_values[2] : contributions pour la classe 2 (hémorragie)
```

```
# 5. Visualiser les résultats
for class_idx in [1, 2]: # œdème et hémorragie
    shap_map = shap_values[class_idx][0]
    vmax = np.percentile(np.abs(shap_map), 95)
    plot_shap_overlay(test_image, shap_map, vmin=-vmax, vmax=vmax)
```

5. Validation de la stabilité des explications SHAP

La stabilité des explications SHAP a été évaluée qualitativement selon trois critères :

1. **Cohérence inter-configurations** : Les cartes SHAP ont été générées avec différents nombres d'échantillons de référence (background) pour le même patient. Les régions à forte attribution restent cohérentes entre les configurations, ce qui suggère une stabilité des explications.
2. **Sensibilité aux paramètres** : La variation du nombre d'échantillons de référence (de 5 à 55) n'a pas modifié qualitativement les zones de forte contribution, indiquant une robustesse raisonnable de l'analyse.
3. **Correspondance anatomique** : Vérification visuelle que les régions à forte contribution correspondent aux zones anatomiquement pertinentes (lésion, moelle épinière, canal rachidien).

ANNEXE IV

RÉSULTATS COMPLÉMENTAIRES DE SEGMENTATION

1. Sensibilité et spécificité par classe

Tableau-A IV-1 Sensibilité et spécificité de la segmentation par classe (modèle axial final)

Classe	Sensibilité	Spécificité	Précision	F1-Score
Œdème (classe 1)	0,607 ± 0,225	1,000 ± 0,000	0,546 ± 0,200	0,529 ± 0,181
Hémorragie (classe 2)	0,091 ± 0,182	1,000 ± 0,000	0,643 ± 0,311	0,105 ± 0,197

Note : Les valeurs sont présentées sous forme de moyenne ± écart-type calculées sur l'ensemble des 122 patients en validation croisée à 5 sous-ensembles. Pour l'hémorragie, les métriques sont calculées uniquement sur les patients présentant de l'hémorragie dans la vérité terrain ($n = 52$). La précision de l'hémorragie est calculée sur les 15 patients ayant des prédictions non nulles.

La sensibilité pour l'œdème (0,607) indique que le modèle détecte correctement environ 61 % des pixels d'œdème. En revanche, la sensibilité pour l'hémorragie est nettement plus faible (0,091), reflétant la difficulté à détecter cette classe minoritaire. Les spécificités très élevées ($\approx 1,000$) pour les deux classes indiquent que le modèle génère très peu de faux positifs dans les régions saines. La précision élevée de l'hémorragie (0,643) montre que, lorsque le modèle prédit de l'hémorragie, il a raison dans environ 64 % des cas, mais il ne la détecte que très rarement.

2. Matrices de confusion

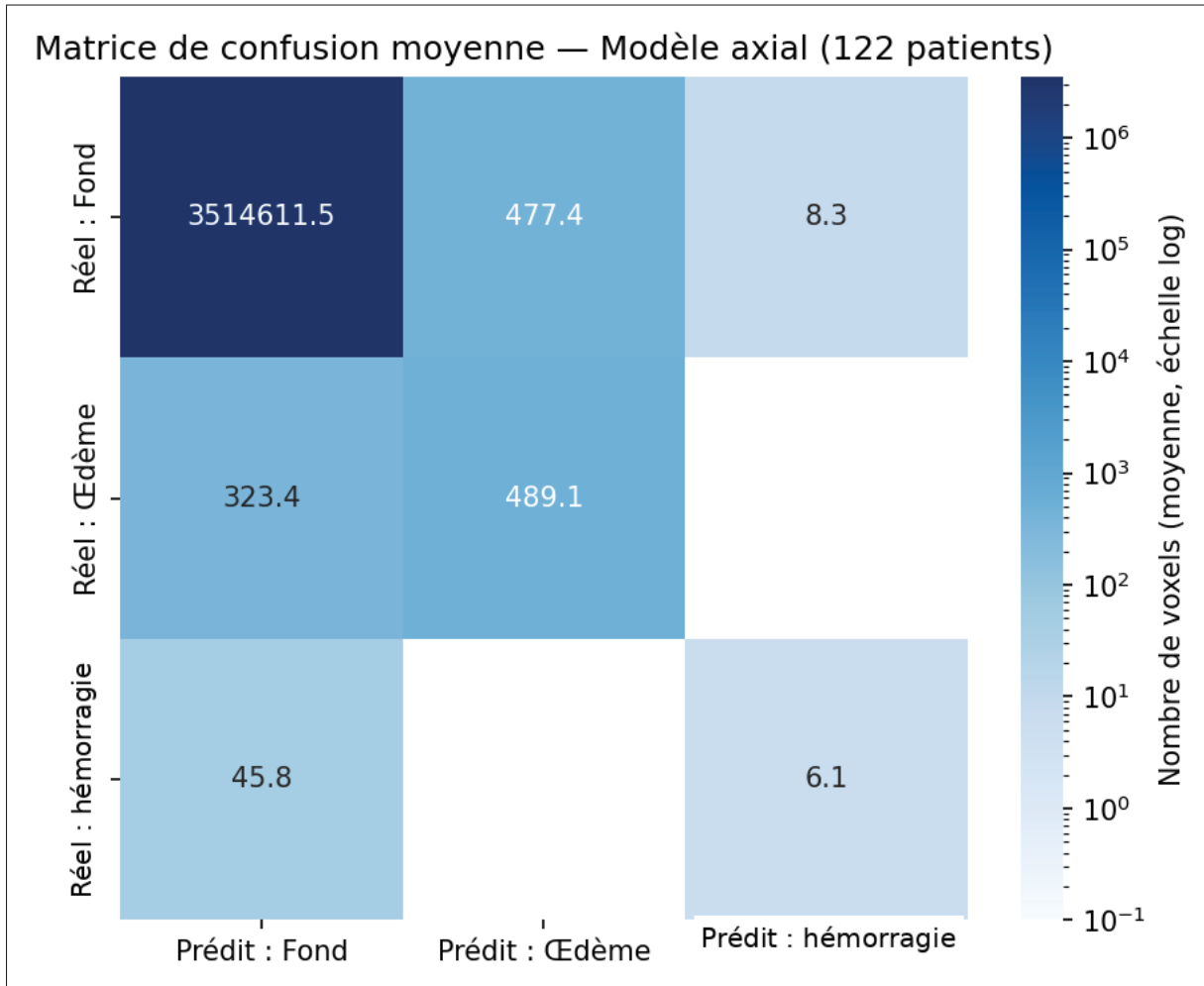


Figure-A IV-1 Matrice de confusion moyenne pour le modèle axial (valeurs en nombre de voxels, 122 patients)

Note : Valeurs moyennes calculées sur l'ensemble des 122 patients à partir des métriques one-vs-rest par sous-ensemble. Les termes croisés œdème↔hémorragie sont approximés à zéro (confusion inter-classes lésionnelles négligeable).

3. Résultats par sous-ensemble de la validation croisée

Tableau-A IV-2 Performances de segmentation par sous-ensemble (coefficient de Dice)

Sous-ensemble	<i>n</i>	Dice Œdème	Dice Hémorragie	Dice Moyen
1	25	0,548 ± 0,164	0,076 ± 0,151	0,406 ± 0,168
2	25	0,536 ± 0,166	0,191 ± 0,243	0,470 ± 0,197
3	24	0,445 ± 0,226	0,111 ± 0,204	0,367 ± 0,209
4	24	0,533 ± 0,165	0,077 ± 0,198	0,401 ± 0,166
5	24	0,586 ± 0,142	0,024 ± 0,063	0,497 ± 0,186
Moyenne	122	0,529 ± 0,181	0,096 ± 0,191	0,428 ± 0,192

Note : Les valeurs par sous-ensemble sont présentées sous forme de moyenne ± écart-type calculées sur les patients du sous-ensemble correspondant. Le Dice moyen est la moyenne foreground (nanmean des classes œdème et hémorragie par patient).

4. Courbes d'apprentissage

Les figures suivantes présentent les courbes d'apprentissage générées par nnU-Net pour la validation croisée à 5 sous-ensembles, pour les modèles sagittal et axial. Chaque figure comprend trois sous-graphiques : (1) l'évolution de la perte d'entraînement (loss_tr, bleu) et de validation (loss_val, rouge) ainsi que le pseudo Dice (vert) sur 1000 époques, (2) la durée par époque, et (3) l'évolution du taux d'apprentissage (polynomial decay).

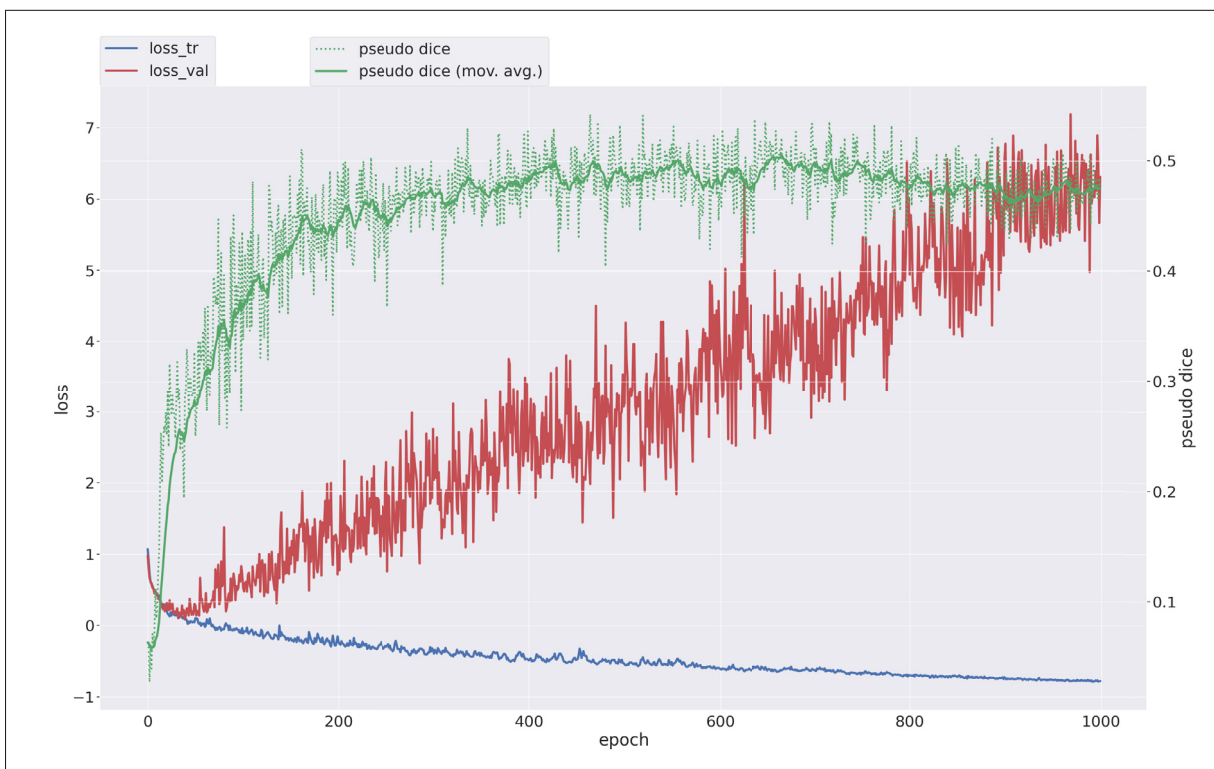


Figure-A IV-2 Courbe d'apprentissage du modèle sagittal (sous-ensemble 1, représentatif). De haut en bas : (1) perte d'entraînement (bleu), perte de validation (rouge) et pseudo Dice (vert) sur 1000 époques ; (2) durée par époque (≈ 52 s) ; (3) taux d'apprentissage (polynomial decay de 0,01 à ≈ 0)

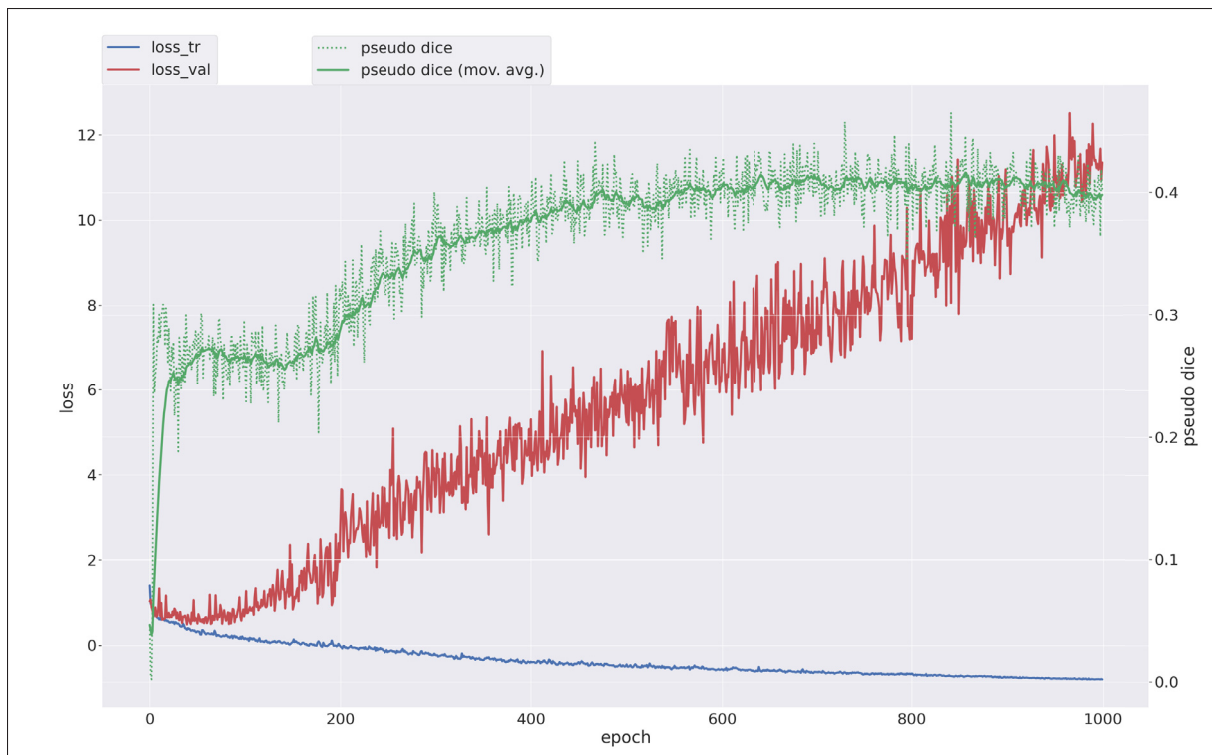


Figure-A IV-3 Courbe d'apprentissage du modèle axial (sous-ensemble 1, représentatif). De haut en bas : (1) perte d'entraînement (bleu), perte de validation (rouge) et pseudo Dice (vert) sur 1000 époques ; (2) durée par époque (≈ 158 s) ; (3) taux d'apprentissage (polynomial decay de 0,01 à ≈ 0)

Note : Les quatre autres sous-ensembles présentent des profils de convergence similaires. La perte d'entraînement décroît de manière monotone tandis que la perte de validation présente des fluctuations typiques. Le pseudo Dice (moyenne mobile en vert) confirme la convergence progressive du modèle.

5. Distribution des scores Dice par patient

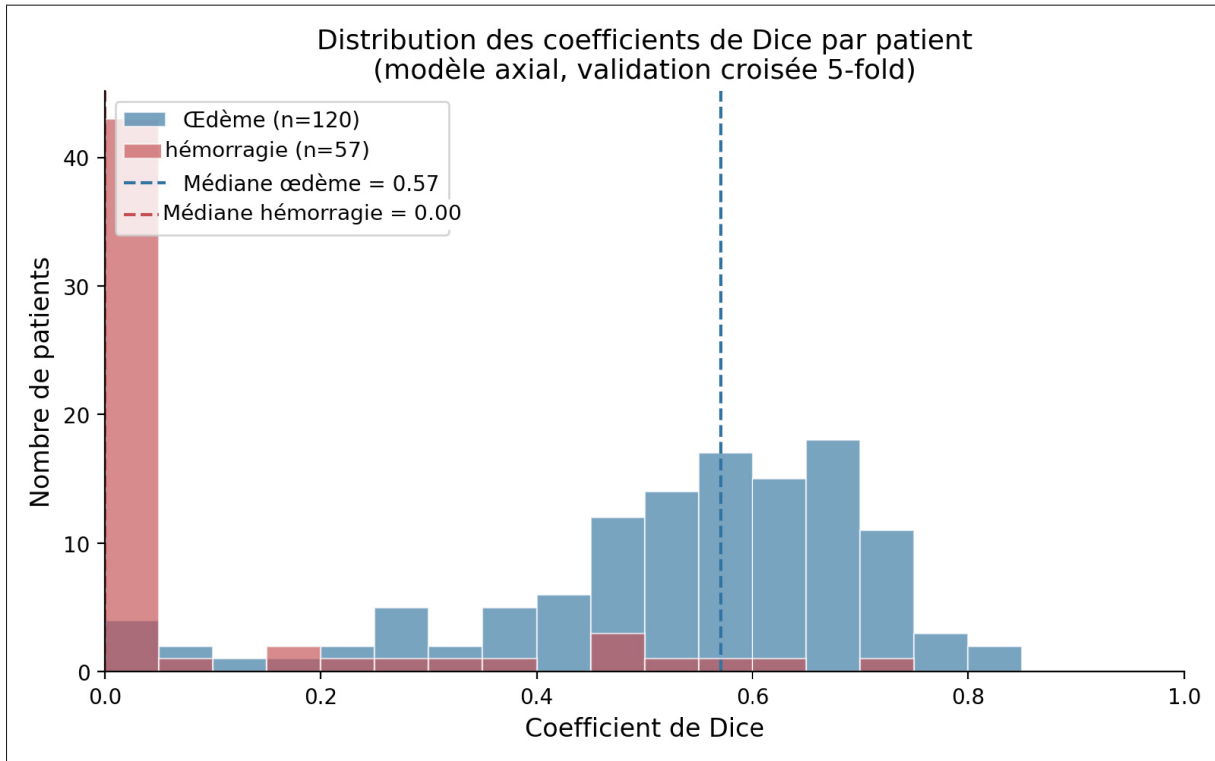


Figure-A IV-4 Distribution des coefficients de Dice pour l'œdème (bleu, $n = 120$) et l'hémorragie (rouge, $n = 57$) sur l'ensemble de la cohorte (122 patients, modèle axial). Les lignes pointillées indiquent la médiane pour chaque classe (œdème : 0,57 ; hémorragie : 0,00). Les patients sans lésion d'une classe donnée (Dice = NaN) sont exclus de la distribution correspondante

ANNEXE V

CAS DIFFICILES SUPPLÉMENTAIRES

1. Galerie des cas difficiles

Cette section présente des cas supplémentaires présentant des performances de segmentation faibles, complémentaires aux cas 1, 25, 89 et 136 analysés en section 3.1.2.2.

Tableau-A V-1 Caractéristiques des cas difficiles supplémentaires

Cas	Dice comb.	Dice œdème	Dice hémorragie	Problème principal
55	0,000	0,000	NaN	Hypersignal atypique (artéfact possible)
148	0,000	0,000	NaN	Discordance de localisation avec la vérité terrain
128	0,043	0,085	0,000	Hémorragie non détectée (faible contraste)
56	0,127	0,253	0,000	Hémorragie prédite mais absente de la vérité terrain
83	0,369	0,369	NaN	Annotation de référence imprécise
137	0,384	0,384	NaN	Discordance de localisation

Note : Les cas 1, 25, 89 et 136 sont analysés en détail dans la section 3.1.2.2. Les valeurs de Dice proviennent de la validation croisée 5-fold du modèle axial. NaN indique l'absence d'hémorragie dans la vérité terrain.

2. Cas supplémentaires détaillés

2.1 Cas 55 : Hypersignal lésionnel atypique (Dice = 0,000)

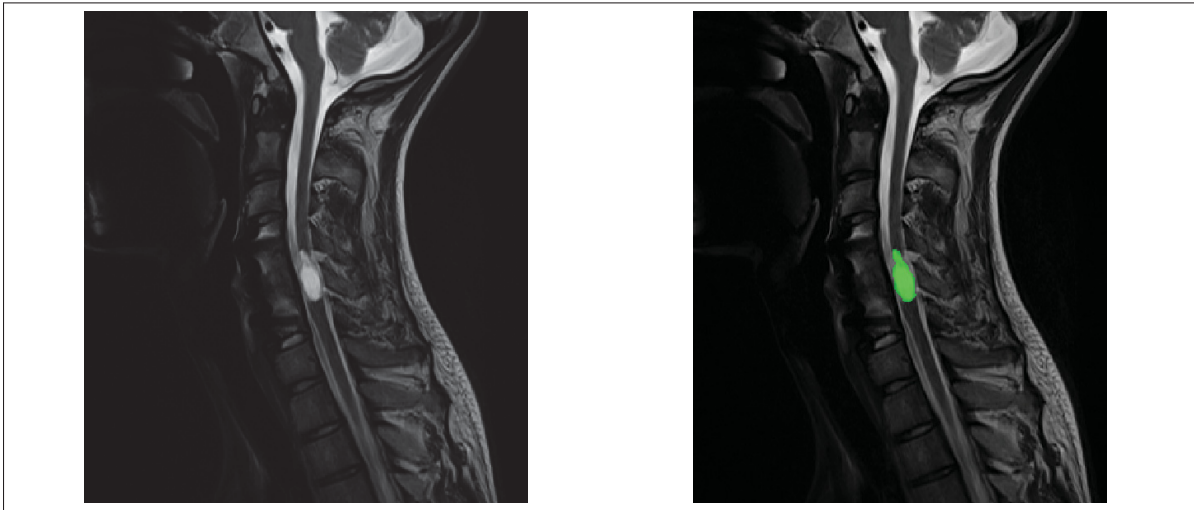


Figure-A V-1 Cas 55 — Image sagittale T2 (gauche) et masque de vérité terrain (droite). La lésion présente un hypersignal très marqué, nettement supérieur à celui habituellement observé dans les lésions œdémateuses de la cohorte, possiblement attribuable à un artefact d'imagerie. Aucun pixel n'a été prédit par le modèle (Dice = 0,000), suggérant que cette intensité atypique se situe en dehors de la distribution apprise lors de l'entraînement

2.2 Cas 83 : Annotation de référence imprécise (Dice = 0,369)

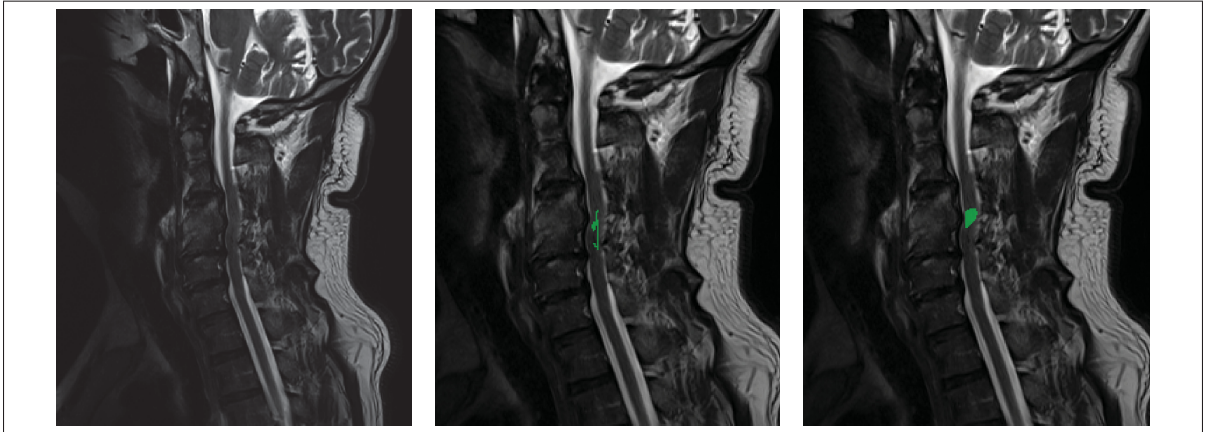


Figure-A V-2 Cas 83 — De gauche à droite : image sagittale T2, masque de vérité terrain et prédiction du modèle. Le masque de vérité terrain présente une annotation peu précise des contours de la lésion. La prédiction du modèle apparaît visuellement de meilleure qualité et mieux délimitée que l'annotation de référence. Le score Dice modéré (0,369) reflète donc davantage une discordance avec l'annotation qu'une réelle erreur de segmentation, illustrant une limite inhérente à l'évaluation par Dice lorsque la vérité terrain est elle-même imprécise

2.3 Cas 128 : Hémorragie non détectée (Dice œdème = 0,085 ; Dice hémorragie = 0,000)

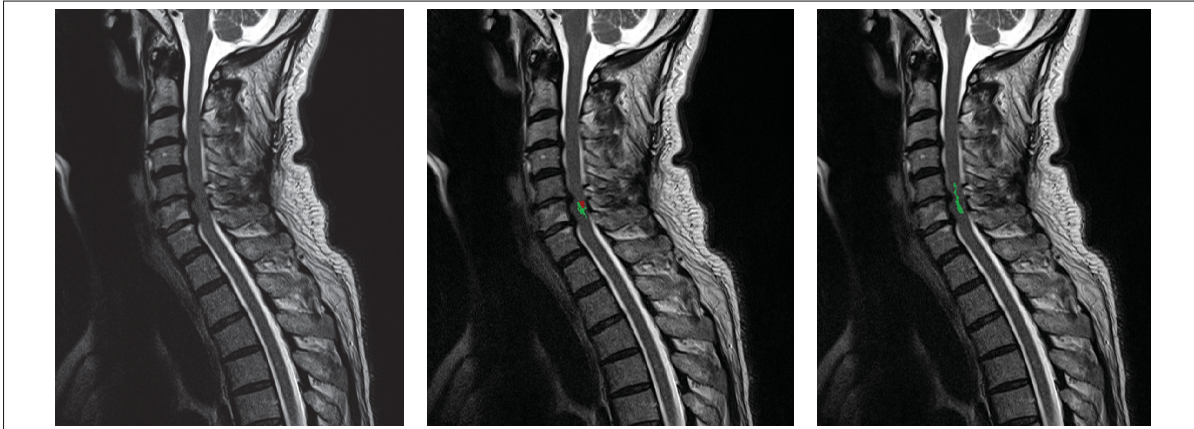


Figure-A V-3 Cas 128 — Vues sagittales. De gauche à droite : image T2, masque de vérité terrain (œdème en vert, hémorragie en rouge) et prédiction du modèle

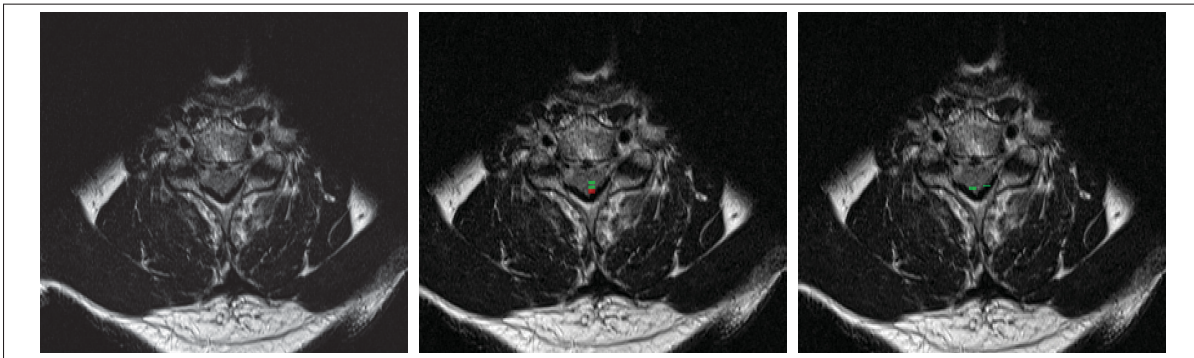


Figure-A V-4 Cas 128 — Coupes axiales. De gauche à droite : image T2, masque de vérité terrain et prédiction du modèle. L'hémorragie, présente dans la vérité terrain, est absente de la prédiction. La distinction entre œdème et hémorragie est particulièrement difficile dans ce cas en raison de la faible différence de contraste entre les deux tissus, rendant la classification correcte de la composante hémorragique quasi impossible pour le modèle

2.4 Cas 148 : Discordance de localisation avec la vérité terrain (Dice = 0,000)

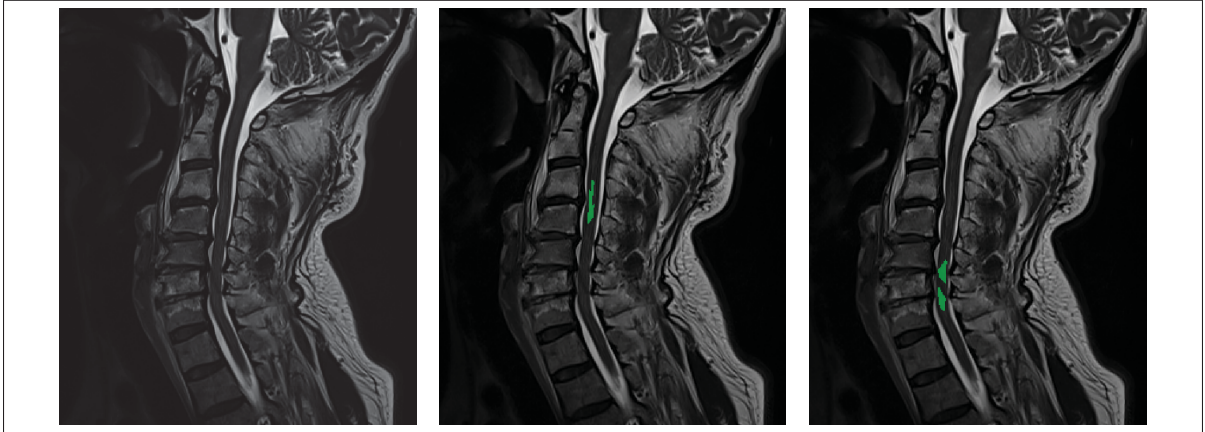


Figure-A V-5 Cas 148 — De gauche à droite : image sagittale T2, masque de vérité terrain et prédiction du modèle. La vérité terrain localise l'œdème dans une région supérieure de la moelle épinière, éloignée du site de traumatisme vertébral visible. La prédiction du modèle, en revanche, situe la lésion au niveau de la zone où l'on observe des dommages vertébraux, ce qui est anatomiquement plus cohérent. Ce cas illustre une possible erreur dans l'annotation de référence, le Dice de 0,000 ne reflétant pas nécessairement une défaillance du modèle mais plutôt une divergence avec une vérité terrain potentiellement erronée

3. Statistiques de résolution par patient

Tableau-A V-2 Distribution des résolutions d'images sagittales dans la cohorte ($n = 140$)

Résolution (pixels)	Nombre de patients	Pourcentage	Dice moyen (sagittal)
512×512	65	46,4 %	0,459 ± 0,177
448×448	7	5,0 %	0,372 ± 0,229
384×384	66	47,1 %	0,469 ± 0,207
320×320 ou moins	2	1,4 %	0,267 ± 0,223

Note : Aucune corrélation significative entre la résolution sagittale et le Dice du modèle sagittal (Spearman : $\rho = -0,02$, $p = 0,80$).

Tableau-A V-3 Distribution des résolutions d'images axiales dans la cohorte ($n = 120$)

Résolution (pixels)	Nombre de patients	Pourcentage	Dice moyen (axial)
512×512	22	18,3 %	0,409 ± 0,226
320×320 ou moins	98	81,7 %	0,432 ± 0,183

Note : La majorité des images axiales ont une résolution de 320×320. Aucune corrélation significative entre la résolution axiale et le Dice du modèle axial (Spearman : $\rho = -0,01$, $p = 0,88$).

ANNEXE VI

ANALYSE STATISTIQUE DÉTAILLÉE : VALEUR PRONOSTIQUE DES BIOMARQUEURS LÉSIONNELS

1. Coefficients de régression par modèle et période

Tableau-A VI-1 Coefficients de régression ($\beta \pm$ erreur standard) pour le modèle M1 (clinique). Variable dépendante : amélioration AIS (delta)

Période	β_0 (Intercept)	β_1 (AIS_pre)	R ² ajusté
Post-opératoire	0,78 \pm 0,36	-0,18 \pm 0,10	0,05
3 mois	0,87 \pm 0,41	-0,15 \pm 0,12	0,04
6 mois	0,21 \pm 0,58	0,01 \pm 0,16	-0,07
1 an	0,22 \pm 0,47	0,04 \pm 0,13	-0,04

Tableau-A VI-2 Coefficients pour le modèle M2 (lésion complète)

Période	β_0	β_1 (AIS_pre)	β_2 (Vol_total)	R ² ajusté
Post-op.	1,17 \pm 0,47	-0,23 \pm 0,11	-0,0003 \pm 0,0002	0,06
3 mois	1,57 \pm 0,55	-0,28 \pm 0,13	-0,0004 \pm 0,0002	0,14
6 mois	0,14 \pm 0,82	0,02 \pm 0,19	0,0000 \pm 0,0003	-0,14
1 an	0,29 \pm 0,68	0,03 \pm 0,16	-0,0000 \pm 0,0003	-0,10

Tableau-A VI-3 Coefficients pour le modèle M3 (biomarqueurs différenciés)

Période	β_1 (AIS_pre)	β_2 (Vol_œd.)	β_3 (Vol_nécr.)	β_4 (Ratio)	R ² aj.
Post-op.	-0,37 \pm 0,13	-0,0002 \pm 0,0003	-0,004 \pm 0,005	-6,03 \pm 6,06	0,09
3 mois	-0,31 \pm 0,14	-0,0007 \pm 0,0003	-0,002 \pm 0,005	-5,92 \pm 4,87	0,45
6 mois	-0,78 \pm 0,19	0,0004 \pm 0,0003	-0,015 \pm 0,005	-13,63 \pm 4,15	0,70
1 an	-0,73 \pm 0,19	0,0002 \pm 0,0003	-0,010 \pm 0,004	-14,94 \pm 4,40	0,56

Note : Le modèle M1 (AIS pré uniquement) a un pouvoir prédictif limité (R² ajusté \leq 0,05 au post-opératoire, proche de 0 ou négatif au-delà). L'ajout du volume total (M2) n'améliore pas substantiellement la prédiction. La différenciation œdème/hémorragie (M3) améliore le R² ajusté à partir de 3 mois (0,45 à 3 mois, 0,70 à 6 mois, 0,56 à 1 an), avec un volume d'hémorragie et un ratio hémorragie/lésion significatifs ($p < 0,05$) aux périodes 3 mois, 6 mois et 1 an. Le volume d'œdème reste non significatif à toutes les périodes.

2. Matrices de corrélation entre biomarqueurs

Tableau-A VI-4 Matrice de corrélation de Spearman entre les biomarqueurs ($n = 39$)

	AIS_pre	Vol_total	Vol_œdème	Vol_hémorragie	Ratio
AIS_pre	1,00	-0,22	-0,20	-0,47	-0,47
Vol_total	-0,22	1,00	1,00	0,35	0,34
Vol_œdème	-0,20	1,00	1,00	0,33	0,31
Vol_hémorragie	-0,47	0,35	0,33	1,00	1,00
Ratio	-0,47	0,34	0,31	1,00	1,00

Note : La forte corrélation Vol_total–Vol_œdème ($\rho = 1,00$) s'explique par la prépondérance de l'œdème.

Vol_hémorragie et Ratio sont fortement corrélés car le ratio en dérive.

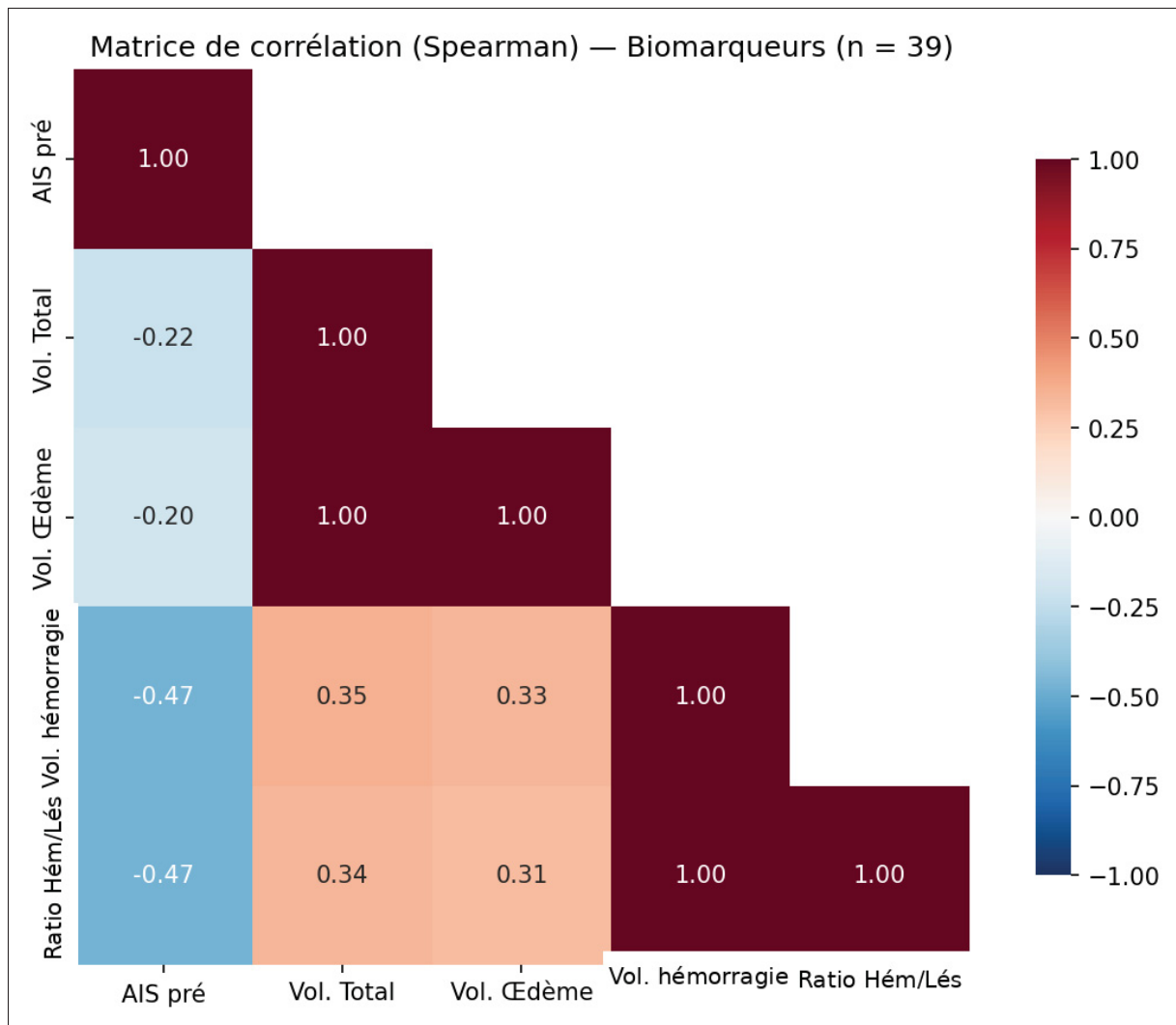


Figure-A VI-1 Matrice de corrélation de Spearman entre les biomarqueurs (AIS préopératoire, volumes œdème et hémorragie, ratio hémorragie/lésion) ($n = 39$)

3. Test F pour modèles emboîtés

Tableau-A VI-5 Résultats du test F comparant le modèle à lésion complète (M2) et le modèle différencié (M3)

Période	SSR_M2	SSR_M3	F	p-value
Post-opératoire	18,05	16,01	1,41	0,258
3 mois	5,02	2,67	4,11	0,028
6 mois	7,05	1,47	13,91	< 0,001
1 an	9,03	3,04	10,53	< 0,001

Note : SSR = somme des carrés résiduels. À partir de 3 mois, la différenciation œdème/hémorragie apporte une amélioration significative.

ANNEXE VII

SCRIPTS ET REPRODUCTIBILITÉ

1. Environnement logiciel

Tableau-A VII-1 Versions des logiciels et bibliothèques utilisés

Logiciel/Bibliothèque	Version
Python	3.9.16
PyTorch	2.5.1
CUDA	12.0
cuDNN	8.9.2
nnU-Net	2.2
SpinalCordToolbox	6.0
NumPy	1.24.3
SciPy	1.10.1
scikit-learn	1.2.2
SHAP	0.41.0
Matplotlib	3.7.1
Pandas	2.0.1
nibabel	5.1.0
SimpleITK	2.2.1

2. Commandes SpinalCordToolbox utilisées

2.1 Segmentation de la moelle épinière

```
# Segmentation automatique de la moelle épinière sur image T2 sagittale
sct_deepseg_sc -i t2_sag.nii.gz -c t2 -o sc_mask.nii.gz
```

2.2 Enregistrement sagittal vers axial

```
# Enregistrement rigide entre images sagittale et axiale
sct_register_multimodal \
  -i t2_sag.nii.gz \
  -d t2_ax.nii.gz \
  -iseg sc_mask_sag.nii.gz \
  -dseg sc_mask_ax.nii.gz \
  -param step=1,type=im,algo=rigid,metric=MI \
  -o t2_sag_registered.nii.gz \
  -owarp warp_sag2ax.nii.gz
```

2.3 Application de la transformation aux masques

```
# Transformation du masque de segmentation vers l'espace axial
sct_apply_transfo \
  -i segmentation_sag.nii.gz \
  -d t2_ax.nii.gz \
  -w warp_sag2ax.nii.gz \
  -x nn \
  -o segmentation_ax.nii.gz
```

2.4 Étiquetage des niveaux vertébraux

```
# Détection et étiquetage des niveaux vertébraux
sct_label_vertebrae \
  -i t2_sag.nii.gz \
  -s sc_mask.nii.gz \
  -c t2 \
  -o vertebral_levels.nii.gz
```

3. Configuration d'entraînement nnU-Net

```
# Prétraitement des données
nnUNetv2_plan_and_preprocess -d DATASET_ID --verify_dataset_integrity

# Entraînement du modèle sagittal (5-fold cross-validation)
nnUNetv2_train DATASET_ID 2d FOLD --npz -tr nnUNetTrainerDA5

# Entraînement du modèle axial
nnUNetv2_train DATASET_ID_AXIAL 2d FOLD --npz -tr nnUNetTrainerDA5

# Inférence
nnUNetv2_predict -i INPUT_FOLDER -o OUTPUT_FOLDER \
  -d DATASET_ID -c 2d -f all
```


BIBLIOGRAPHIE

- Ahuja, C. S., Wilson, J. R., Nori, S., Kotter, M. R., Druschel, C., Curt, A. & Fehlings, M. G. (2017). Traumatic spinal cord injury. *Nature Reviews Disease Primers*, 3(1), 17018. doi : 10.1038/nrdp.2017.18.
- Cavadini, T. (2019). *Exploration du processus d'évaluation cognitive des événements de vie, des compétences émotionnelles et de la santé mentale d'hommes tout-venant : une contribution à l'établissement de normes longitudinales destinées au suivi de sujets blessés médullaires*. (Mémoire de maîtrise, Université de Genève, Genève, Suisse).
- Daghfous, L. R. M. (2018). Traumatisme du rachis cervical bas et IRM cervicale en urgence. Repéré à <https://ebulletin.radiologie.fr/vendredi/traumatisme-du-rachis-cervical-bas-irm-cervicale-urgence>.
- D'Antonoli, T. A., Berger, L. K., Indrakanti, A. K., Vishwanathan, N., Weiß, J., Jung, M. & Wasserthal, J. (2024). TotalSegmentator MRI : Sequence-Independent Segmentation of 59 Anatomical Structures in MR images. *arXiv preprint arXiv :2405.19492v1*. Repéré à <https://arxiv.org/abs/2405.19492v1>.
- David, G., Mohammadi, S., Martin, A. R., Cohen-Adad, J., Weiskopf, N., Thompson, A. & Freund, P. (2019). Traumatic and nontraumatic spinal cord injury : pathological insights from neuroimaging. *Nature Reviews Neurology*, 15(12), 718–731. doi : 10.1038/s41582-019-0270-5.
- De Leener, B., Lévy, S., Dupont, S. M., Fonov, V. S., Stikov, N., Louis Collins, D., Callot, V. & Cohen-Adad, J. (2017). SCT : Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *NeuroImage*, 145, 24–43. doi : 10.1016/j.neuroimage.2016.10.009.
- Dobran, M., Aiudi, D., Liverotti, V., Fasinella, M. R., Lattanzi, S., Melchiorri, C., Polonara, G. & Iacoangeli, M. (2023). Prognostic MRI parameters in acute traumatic cervical spinal cord injury. *European Spine Journal*, 32(5), 1584–1590. doi : 10.1007/s00586-023-07560-4.
- Enamundram, N. K., Valosek, J., Smith, A. C., Pfyffer, D., Schading-Sassenhausen, S., Farner, L., Lebet, A., Rohan, M., Burns, A. S., Fehlings, M. G. et al. (2024). SCISeg : Automatic Segmentation of T2-weighted Hyperintense Lesions in Spinal Cord Injury. *medRxiv*. doi : 10.1101/2024.01.03.24300794.

- Freund, P., Seif, M., Weiskopf, N., Friston, K., Fehlings, M. G., Thompson, A. J. & Curt, A. (2019). MRI in traumatic spinal cord injury : from clinical assessment to neuroimaging biomarkers. *The Lancet Neurology*, 18(12), 1123–1135. doi : 10.1016/S1474-4422(19)30138-3.
- Gros, C., De Leener, B., Badji, A., Maranzano, J., Eden, D., Dupont, S. M., Talbott, J., Zhuoquiong, R., Liu, Y., Granberg, T. et al. (2019). Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *NeuroImage*, 184, 901–915. doi : 10.1016/j.neuroimage.2018.09.081.
- Gupta, R., Bathen, M. E., Smith, J. S., Levi, A. D., Bhatia, N. N. & Steward, O. (2014). Magnetic resonance imaging findings in acute cervical spinal cord injury correlate with neurological outcome. *Neurosurgery*, 75(5), 557–564. doi : 10.1227/NEU.0000000000000477.
- Häntze, H., Xu, L., Dorfner, F. J., Donle, L., Truhn, D., Aerts, H., Wasserthal, J. & Bressemer, K. K. (2024). MRSegmentator : Robust Multi-Modality Segmentation of 40 Classes in MRI and CT Sequences. *arXiv preprint arXiv :2405.06463*. doi : 10.48550/arXiv.2405.06463.
- Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. (2021). nnU-Net : a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203–211. doi : 10.1038/s41592-020-01008-z.
- Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K. & Jaeger, P. F. (2024). nnU-Net Revisited : A Call for Rigorous Validation in 3D Medical Image Segmentation. *arXiv preprint arXiv :2404.09556*. doi : 10.48550/arXiv.2404.09556.
- Karthik, E. N., Valošek, J., Farner, L., Pfyffer, D., Schading-Sassenhausen, S., Leuret, A., Rohan, M., Burns, A. S., Fehlings, M. G., Martin, A. R. et al. (2024). SCISegV2 : A Universal Tool for Segmentation of Intramedullary Lesions in Spinal Cord Injury. *arXiv preprint arXiv :2407.17265*. doi : 10.48550/arXiv.2407.17265.
- Kirshblum, S. C., Burns, S. P., Biering-Sorensen, F., Donovan, W., Graves, D. E., Jha, A., Johansen, M., Jones, L., Krassioukov, A., Mulcahey, M. et al. (2011). International standards for neurological classification of spinal cord injury (revised 2011). *The Journal of Spinal Cord Medicine*, 34(6), 535–546. doi : 10.1179/204577211X13207446293695.
- Lamrani, S. [Image IRM, cas clinique]. (2017). Contusion médullaire œdémateuse cervicale.
- Lombaert, H. [Notes de cours, Département de génie logiciel et des TI, École de technologie supérieure. Moodle. <https://ena.etsmtl.ca/course/view.php?id=19443#section-1>]. (2023). Réseaux de neurones à convolution.

- Martineau, J., Goulet, J., Richard-Denis, A. & Mac-Thiong, J.-M. (2019). The relevance of MRI for predicting neurological recovery following cervical traumatic spinal cord injury. *Spinal Cord*, 57(10), 866–873. doi : 10.1038/s41393-019-0295-z.
- Masse-Gignac, N., Flórez-Jiménez, S., Mac-Thiong, J.-M. & Duong, L. (2023). Attention-gated U-Net networks for simultaneous axial/sagittal planes segmentation of injured spinal cords. *Journal of Applied Clinical Medical Physics*, e14123. doi : 10.1002/acm2.14123.
- McCoy, D. B., Dupont, S. M., Gros, C., Cohen-Adad, J., Huie, J. R., Ferguson, A. R., Tetreault, L. A., Blocker, S. J., Sankaran, S., Pestilli, F. et al. (2019). Convolutional neural network-based automated segmentation of the spinal cord and contusion injury : deep learning biomarker correlates of motor impairment in acute spinal cord injury. *American Journal of Neuroradiology*, 40(4), 737–744. doi : 10.3174/ajnr.A6020.
- Miyanji, F., Furlan, J. C., Aarabi, B., Arnold, P. M. & Fehlings, M. G. (2007). Acute cervical traumatic spinal cord injury : MR imaging findings correlated with neurological outcome—prospective study with 100 consecutive patients. *Radiology*, 243(3), 820–827. doi : 10.1148/radiol.2433060583.
- Roberts, T. T., Leonard, G. R. & Cepela, D. J. (2017). Classifications In Brief : American Spinal Injury Association (ASIA) Impairment Scale. *Clinical Orthopaedics and Related Research*, 475(5), 1499–1504. doi : 10.1007/s11999-016-5133-4. Epub 2016 Nov 4.
- Ronneberger, O., Fischer, P. & Brox, T. (2015). U-Net : Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint arXiv :1505.04597*. doi : 10.48550/arXiv.1505.04597.
- Rutges, J. P., Kwon, B. K., Heran, M., Ailon, T., Street, J. T. & Dvorak, M. F. (2017). A prospective serial MRI study following acute traumatic cervical spinal cord injury. *European Spine Journal*, 26(9), 2324–2332. doi : 10.1007/s00586-017-5097-4.
- Seif, M., Wheeler-Kingshott, C., Cohen-Adad, J., Flanders, A. & Freund, P. (2019). Guidelines for the conduct of clinical trials in spinal cord injury : Neuroimaging biomarkers. *Spinal Cord*, 57. doi : 10.1038/s41393-019-0309-x.
- Smith, A. C., Albin, S. R., O'Dell, D. R., Berliner, J. C., Dungan, D., Sevigny, M., Thomas, C. K., Palacios, E., Ferguson, A. R., Salegio, E. A. et al. (2021). Axial MRI biomarkers of spinal cord damage to predict future walking and motor function : a retrospective study. *Spinal Cord*, 59(6), 693–699. doi : 10.1038/s41393-020-00561-w.

Talbott, J. F., Whetstone, W. D., Readdy, W. J., Ferguson, A. R., Bresnahan, J. C., Saigal, R., Hawryluk, G. W., Beattie, M. S., Manley, G. T. & Dhall, S. S. (2015). The Brain and Spinal Injury Center score : a novel, simple, and reproducible method for assessing the severity of acute cervical spinal cord injury with axial T2-weighted MRI findings. *Journal of Neurosurgery : Spine*, 23(4), 495–504. doi : 10.3171/2015.1.SPINE141033.