



UNIVERSITÉ DE LILLE
FACULTÉ DE MÉDECINE HENRI WAREMBOURG

Année : 2024

THÈSE POUR LE DIPLÔME D'ÉTAT
DE DOCTEUR EN MÉDECINE

**Étude de la performance d'un outil de segmentation automatisée
dans le suivi des lésions de sclérose en plaques par IRM au sein du
CHU de Lille**

Présentée et soutenue publiquement le 15 octobre à 14h00
au Pôle Recherche
par Sheena-Chloé KOUEMO

JURY

Président :

Monsieur Le Professeur Grégory KUCHCINSKI

Assesseurs :

Madame La Professeure Hélène ZEPHIR

Monsieur Le Docteur Arnaud BEAUGUITTE

Directeur de thèse :

Monsieur Le Docteur Olivier OUTTERYCK

Avertissement

La Faculté n'entend donner aucune approbation aux opinions émises dans les thèses : celles-ci sont propres à leurs auteurs.

Table des matières

Abréviations	6
Table des figures	8
Table des tableaux	8
I. État des connaissances actuelles sur la sclérose en plaques et l'intelligence artificielle	9
A. La sclérose en plaques	9
1. Généralités	9
2. Épidémiologie	9
3. Physiopathologie et immunopathologie	10
3.1 Inflammation et démyélinisation	10
4. Facteurs de risques	11
4.1 Facteurs de risque génétique	11
4.2 Facteurs de risques environnementaux	12
5. Diagnostic	14
5.1 Signes cliniques	14
5.2 Diagnostics différentiels	15
5.3 Formes cliniques	15
5.4 Critères de MacDonald	18
6. Apport de l'IRM	20
6.1 Généralités	20
7. Patterns des lésions en IRM de la SEP	25
7.1 Caractéristiques générales	25
7.2 Localisations	25
8. Pronostic	38
8.1 Espérance de vie et risque de mortalité	38
8.2 Morbidité	39
8.3 Facteurs pronostics	41
9. Traitement	41
9.1 Traitement de poussées	41
9.2 Traitement symptomatique	42
9.3 Traitement de fond	42
B. Intelligence artificielle	43
1. Introduction	43
1.1 L'intelligence artificielle dans le monde de la santé en général	43
1.2 Dans le monde de la radiologie	43
2. Définition de l'intelligence artificielle	45
2.1 Intelligence artificielle	45
2.2 Les différents sous-types de l'intelligence artificielle	47
3. Créer et évaluer un modèle d'IA	53
3.1 Données	53
3.2 Entraînement	53
3.3 Inconvénients possibles	54
3.4 Évaluer le modèle	55
4. Exemple de l'IA appliquée dans la sclérose en plaques	57
4.1 Diagnostic	58
4.2 Pronostic	63
4.3 Traitement	64
C. Objectifs de la thèse	65
II. Article	66
A. Matériels et méthodes : généralités	66
1. Caractéristiques de la population	66
2. Variables recueillies	66

3. Tests statistiques	66
B. Matériels et méthodes : objectifs de l'étude	67
1. Objectif principal	67
1.1 Le radiologue expert	67
1.2 Le logiciel Pixyl ©.....	67
1.3 Test statistique	68
2. Objectifs secondaires	69
2.1 Concordance du nombre de nouvelles lésions et croissantes	69
2.2 Facteurs étiologiques en cas de discordance	70
2.3 L'implication des paramètres techniques sur la détection des lésions croissantes et nouvelles.	73
C. Résultats	74
1. Description de la population et des paramètres extrinsèques et intrinsèques de l'IRM 74	
1.1 Description de la population.....	74
1.2 Description des paramètres extrinsèques et intrinsèques de l'IRM	75
2. Objectif principal	77
3. Objectifs secondaires	78
3.1 Concordance pour la détection de nouvelles lésions et lésions croissantes.....	78
3.2 Concordance pour la détection du nombre de lésions croissantes et nouvelles lésions.	81
3.3 Facteurs étiologiques expliquant une discordance.	84
3.4 Influence des paramètres techniques sur la performance du logiciel Pixyl© à détecter des lésions croissantes et des nouvelles lésions	86
D. Discussion	89
1. Rappel de l'objectif et des résultats de l'étude	89
2. Segmentation/détection des lésions de SEP et évolution des lésions de SEP	90
2.1 Segmentation/détection des lésions de SEP	90
2.2 Évolution des lésions de SEP	91
3. Limites de l'étude	94
3.1 Boîte noire.....	94
3.2 « IA vs expert » et « IA + expert »	95
3.3 Lésions non prises en compte par le logiciel	95
3.4 Concept de lésions croissantes.....	96
4. Solutions envisagées	97
4.1 Modèle d'incertitude.....	97
4.2 Nouveaux paramètres ?	97
E. Conclusion	98
Références	99

Abréviations

3D DIR : Three dimensions –Double Inversion Recupération
3DT2 FLAIR : Three dimensions - Fluid Attenuated Inversion Recovery
ACR : American College Radiology
ADC : Apparent Coefficient Diffusion
ADEM : Acute Disseminated EncephaloMyelitis
ADN : Acide DésoxyriboNucléique
AUC : Aire sous la courbe
AVC : Accident vasculaire cérébral
AVD : Anomalie Veineuse de Développement
BHE : Barrière Hémato-Encéphalique
Ca : Calcium
CADASIL : Cerebral Autosomic Dominant Arteriopathy
CE : Conformité Européenne
Cho : Choline
CHU : Centre Hospitalier Universitaire
CIS : Syndrome Cliniquement Isolé
Clippers : Chronic Lymphocytic Inflammation with Pontine Perivascular Enhancement Responsive to Steroids
CMH : Complexe Majeur d'Histocompatibilité
CNN : Réseau de neurone convolutif
DTI : Diffusion Tensor Imaging
EBNA : Epstein Barr Nuclear Antigène
EBV : Epstein Barr Virus
ECR : European College Radiology
EDSS : Expanded Disability Status Score
EG : Echo de Gradient
FDA : Food and Drug Administration
FP : Faux Positif
GlialCam : Glial Cell Adhesion Molecule
GWAS : Étude d'association pangénomique
HTLV 1 : Human T-Lymphotropic Virus-1
IA : Intelligence Artificielle
IFN- γ : Interferon gamma
IL : Interleukine
inTK : invariant natural killer
IRM : Imagerie par Résonance Magnétique

LB : Lymphocyte B
LCS : Liquide Cérébro-Spinal
LEMP : Leucoencéphalopathie Multiple Progressive
LT : Lymphocyte T
MAG : Myeline Associated Glycoprotein
MOG : Myeline Oligodendrocyte Glycoprotein
MOGAD : Encéphalomyélite associée aux Ac anti-MOG
MTI : Magnetization Transfer Imaging
Na : Sodium
NEDA : No Evidence Disease Activity
Nk : Natural Killer
NMO : Neuromyérite Optique
NORB : Névrite Optique Rétro-Bulbaire
OCT : Tomographie à cohérence optique
OFSEP : Observatoire Français de la Sclérose En Plaques
PACS : Picture Archiving and Communication System
PP : Primaire Progressive
PPP : Parité de pouvoir d'achat
ReLU : Rectified Linear Unit
RF : Radiofréquence
RIS : Syndrome radiologiquement isolé
ROC : Receiver operating characteristic
RR : Forme rémittente / récurrente
SE : Spin Echo
SEP : Sclérose en plaques
SNC : Système Nerveux Central
SP : Secondairement Progressive
STIR : Short Tau Inversion Recovery
SWI : Susceptibility – Weighted – Imaging
tCr : creatine and phosphocreatine
TNF β : Tumor necrosis factor β
UVB : Ultraviolet B
VCAM : Vascular Cell Adhesion Molecule
VIH : Virus Immunodéficience Humaine
VN : Vrai négatif
VPP : Valeur Prédictive Positive
VPN : Valeur Prédictive Négative

Table des figures

Figure 1 : Résumé de la physiopathologie de la SEP ⁽³⁾	11
Figure 2 : Les différents loci HLA et le risque de SEP ⁽¹⁵⁾	12
Figure 3 : Résumé des 3 formes cliniques « historiques » de la SEP ⁽³⁴⁾	18
Figure 4 : Exemple d'atteinte médullaire d'une SEP ⁽⁴⁵⁾	24
Figure 5 : Lésions non-SEP périventriculaires ⁽⁴⁶⁾	26
Figure 6 : Lésions périventriculaires physiologiques "capping" ⁽⁴⁶⁾	27
Figure 7 : Lésions corticales dans le cadre de la SEP ⁽⁴⁶⁾	27
Figure 8 : Lésions corticales ou juxta-corticales non SEP ⁽⁴⁶⁾	28
Figure 9 : Lésions en hypersignal FLAIR infratentorielles typiques de SEP ⁽⁴⁶⁾	29
Figure 10 : Lésions en hypersignal FLAIR et T2 centro-pontine en apport avec une maladie des petits vaisseaux ⁽⁴⁶⁾	30
Figure 11 : Lésions infratentorielles non-SEP ⁽⁴⁶⁾	30
Figure 12 : Lésions médullaires typiques de SEP ⁽⁴⁶⁾	31
Figure 13 : Lésions médullaires non-SEP ⁽⁴⁶⁾	32
Figure 14 : Lésions réhaussées caractéristiques d'une SEP ⁽⁴⁶⁾	32
Figure 15 : Lésions réhaussées non-SEP ⁽⁴⁶⁾	34
Figure 16 : Névrite optique dans le cadre d'une SEP ⁽⁴⁸⁾	35
Figure 17 : Névrite optique dans le cadre d'une NMO et ischémique ⁽⁴⁹⁾⁽⁵⁰⁾	35
Figure 18 : Signe de la veine centrale ⁽⁵²⁾	36
Figure 19 : Anneau paramagnétique d'une lésion de SEP ⁽⁵³⁾	37
Figure 20 : Échelle EDSS ⁽⁵⁶⁾⁽⁵⁷⁾	39
Figure 21 : Le cycle d'une « machine learning » ⁽⁷⁴⁾	46
Figure 22 : Réseaux de neurones dans le « deep learning » ⁽⁷⁴⁾	49
Figure 23 : Couche de convolution dans le réseau de neurones convolutifs ⁽⁸⁹⁾	51
Figure 24 : Illustration du « max pooling »	52
Figure 25 : Résumé du CNN	52
Figure 26 : Illustration du surentraînement et sous-entraînement	55
Figure 27 : Les facettes et multimodalités de l'IA dans le diagnostic de la SEP ⁽⁴⁹⁾	58
Figure 28 : Le fonctionnement du U-net ⁽¹⁰¹⁾	60
Figure 29 : La technique du 3D U-net et segmentation d'une nouvelle lésion ⁽¹⁰⁵⁾	62
Figure 30 : Évolution des lésions de SEP pour un patient ⁽⁹⁶⁾	62
Figure 31 : Exemple de compte-rendu fourni par Pixyl © ⁽¹¹⁸⁾	68
Figure 32 : Exemple d'une discordance sur le nombre de lésions croissantes (raison 3)	71
Figure 33 : Exemple d'une discordance sur le nombre de nouvelles lésions (raison 1)	72
Figure 34 : Différence du nombre de nouvelles lésions détectées par le logiciel Pixyl© et le radiologue expert.	82
Figure 35 : Graphique de Bland-Altman illustrant la comparaison du nombre des nouvelles lésions entre le logiciel et le radiologue-expert	82
Figure 36 : Différence du nombre de lésions croissantes détectées par le logiciel Pixyl© et le radiologue expert	83
Figure 37 : Graphique de Bland-Altman illustrant la comparaison des lésions croissantes entre le logiciel et le radiologue-expert	83
Figure 38 : Répartition par un diagramme en secteur des étiologies de discordance des nouvelles lésions	85
Figure 39 : Répartition par un diagramme en secteur des étiologies des discordances des lésions croissantes	86

Table des tableaux

Tableau 1 : Critères de McDonald d'une forme RR de SEP ⁽³⁰⁾	19
Tableau 2 : Critères de McDonald d'une forme PP de SEP ⁽³⁸⁾	19
Tableau 3 : Résumé des protocoles OFSEP pour la SEP ⁽⁴¹⁾	24
Tableau 4 : Différences entre CNN classique et U-net	60
Tableau 5 : Description de la population	75
Tableau 6 : Paramètres extrinsèques et intrinsèques des IRM réalisées	76
Tableau 7 : Concordance pour la détection d'une activité inflammatoire	78
Tableau 8 : Concordance pour la détection de nouvelles lésions	79
Tableau 9 : Concordance pour la détection des lésions croissantes	80
Tableau 10 : Concordance pour le nombre de nouvelles lésions	81
Tableau 11 : Concordance pour le nombre de lésions croissantes	83
Tableau 12 : Paramètres techniques de l'IRM et discordance pour la détection d'une nouvelle lésion	87
Tableau 13 : Paramètres techniques de l'IRM et discordance pour la détection d'une lésion croissante	88
Tableau 14 : Coefficient Dice et F1 score des algorithmes issus du MSSEG-II challenge ⁽¹²²⁾	92

I. État des connaissances actuelles sur la sclérose en plaques et l'intelligence artificielle

A. La sclérose en plaques

1. Généralités

La sclérose en plaques (SEP) est une maladie inflammatoire chronique auto-immune ayant pour tropisme le système nerveux central (SNC) associant une démyélinisation et une neurodégénérescence. Elle est caractérisée par une accumulation de cellules inflammatoires de distribution périvasculaire au sein de la substance blanche et grise ainsi que de la moelle. L'ensemble engendrant des signes cliniques neurologiques de localisation variable et de durée variable ⁽¹⁾.

Le terme de « *sclérose en plaque disséminée* » a été utilisé pour la première fois en 1868 par le neurologue JM.Charcot qui notait des zones sclérotiques focales bien limitées de la substance blanche correspondant à des plaques démyélinisées sur des analyses post-mortem ⁽²⁾.

2. Épidémiologie

Cette maladie est un phénomène de santé publique étant donné que :

- Sa prévalence est estimée à au moins 2,8 millions de personnes dans le monde avec une répartition suivant un gradient Nord-Sud ⁽³⁾. Environ 12.000 personnes en sont atteintes dans la région des Hauts-de-France.
- C'est l'une des premières causes de handicap du sujet jeune d'autant plus que l'âge moyen du début de la maladie est de 30 ans. ^{(4) (5)}. Il existe également des formes pédiatriques de la maladie (entre 3 à 10% de la SEP) avec une incidence de 0,66 à 1,66 pour 100000 enfants⁽⁶⁾.
- Les coûts de la maladie liés à la fois au traitement et au handicap chronique du patient sont élevés.⁽⁷⁾.

Le sex-ratio de la maladie autrefois de 2 :1 (F :H) est passé à 3 :1. Une des explications seraient l'augmentation du tabagisme féminin, le tabac étant un facteur de risque de développer la maladie ⁽⁵⁾.

3. Physiopathologie et immunopathologie

La SEP est une maladie complexe et multifactorielle. Sa physiopathologie n'est que partiellement comprise.

Des cellules issues de l'immunité innée et adaptative vont attaquer le SNC provoquant une inflammation chronique via une atteinte de la myéline (démýélinisation), un processus de remýélinisation qui va s'amenuiser au fur et à mesure, une destruction axonale et à terme une perte de neurone (neurodégénérescence) sous-tendant le handicap.

3.1 Inflammation et démýélinisation

On peut considérer trois phases successives sur le plan immunopathologique (« inflammation/démýélinisation » ; « remýélinisation » puis « neurodégénérescence ». Cette phase inflammatoire survient surtout durant les phases précoces (notamment rémittente-récurrente) de la SEP mais peut être présente à tous les stades de la SEP. ⁽⁸⁾

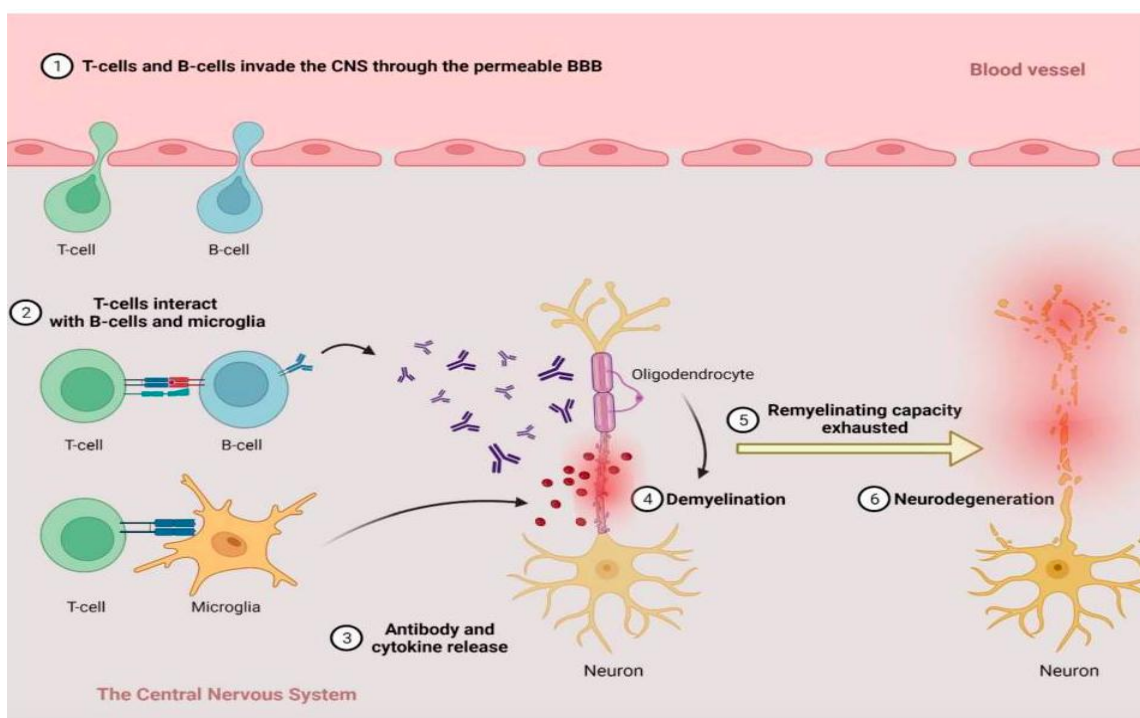
Elle est associée à l'afflux de cellules immunitaires traversant la barrière hémato-encéphalique (BHE), de répartition périveinulaire principalement et proche du liquide céphalo-spinal (LCS), à l'origine de la formation des plaques démýélinisantes identifiées en IRM sous la forme d'hypersignaux T2 ovoïdes et centrés par une veine en séquence SWI-EPI (*susceptibility weight imaging – echo planar imaging*) ⁽⁸⁾. Durant le processus de démýélinisation, l'inflammation va engendrer des plaques caractérisées par une démýélinisation segmentaire et une atteinte axonale ⁽⁸⁾.

Un processus de réparation (remýélinisation) est possible impliquant des cellules diverses comme les cellules souches d'oligodendrocytes qui vont proliférer, migrer et se différencier en cellules matures pour produire de la myéline ⁽⁸⁾ et la microglie, qui pourra sécréter de l'interleukine 4 (IL4) favorisant la survie des oligodendrocytes et la prolifération de leurs cellules souches. ^{(9) (3)}

La remýélinisation est variable d'un patient à l'autre. Ce processus dépendra de l'étendue de la lésion, la durée de la maladie, de l'âge et vraisemblablement de facteurs génétiques. ⁽¹⁰⁾

La neurodégénérescence s'observe dès le début de la maladie mais s'accélère progressivement au cours du temps (plusieurs années). Elle sous-tend l'apparition du handicap chronique et irréversible des patients. Lors de cette phase dégénérative et progressive, est observée une seconde phase inflammatoire médiée par les lymphocytes B qui peuvent s'accumuler dans les méninges et les espaces de Virchow-Robin et induire une atteinte beaucoup plus diffuse intéressant la substance blanche et la substance grise ⁽¹¹⁾. L'extension de la perte neuro-axonale peut aboutir à terme à une atrophie cérébrale et médullaire. ⁽¹²⁾

Figure 1 : Résumé de la physiopathologie de la SEP ⁽³⁾



4. Facteurs de risques

L'étiopathogénie de la SEP est complexe et multifactorielle. Elle implique des facteurs génétiques et environnementaux.

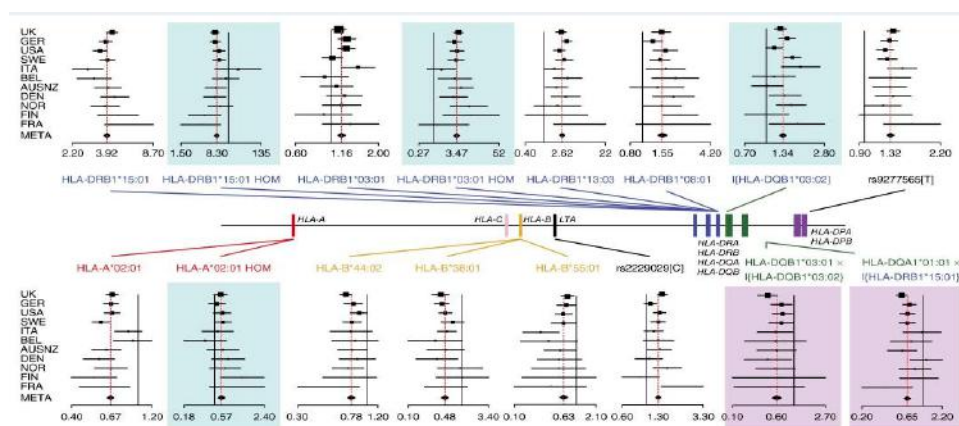
4.1 Facteurs de risque génétique

Le risque de survenue de la SEP chez des sujets d'une même famille est estimé entre 2 et 5% ⁽¹³⁾. Ce pourcentage augmente à environ 20% chez des jumeaux monozygotes. ⁽¹⁴⁾

Une étude d'association pangénomique (GWAS) a identifié plusieurs centaines de variants génétiques s'associant à un risque accru de SEP. Parmi les gènes à risque, 32 sont liés au complexe majeur d'histocompatibilité (CMH) dont le principal est : HLA DRB1 ⁽¹⁵⁾.

Par ailleurs il existe également plusieurs gènes protecteurs du CMH I comme HLA A 0201 ; HLA-B*44:02 ; HLA-B*38:01 ; HLA-B*55:01 qui seraient protecteurs vis-à-vis du risque de développer une SEP ⁽¹⁶⁾.

Figure 2 : Les différents loci HLA et le risque de SEP ⁽¹⁵⁾



4.2 Facteurs de risques environnementaux

4.2.1 Carence en vitamine D

Les ultraviolets B (UVB) participent à la synthèse de la vitamine D et inhibent la production de cytokines pro-inflammatoires impliquées dans la pathogenèse de la SEP. Ainsi, dans les zones géographiques moins ensoleillées (et par extension, susceptibles d'engendrer une carence en vitamine D, le risque de SEP est augmenté (gradient Nord-Sud de la SEP mentionné ci-dessus dans A.2)). Le gradient nord/sud ne s'explique pas à lui seul par la carence en vitamine D et le taux d'ensoleillement. D'autres agents environnementaux, infectieux, notamment viraux pourraient être impliqués ⁽¹⁷⁾.

4.2.2 Epstein Barr Virus (EBV)

L'infection à EBV est associée à un risque multiplié par 32 de SEP d'après une étude longitudinale d'une cohorte de plus de 10 millions de jeunes adultes entre 1993 à 2013. ⁽¹⁸⁾ Plusieurs hypothèses peuvent expliquer le lien entre ces 2 entités :

- Après une primo-infection par EBV, le virus est latent au sein des LB mémoires circulants et des réactivations partielles du virus sont contrôlées par le système immunitaire. Dans certains cas, ce contrôle serait défaillant et favoriserait l'infiltration du SNC. ⁽¹⁹⁾

- Il existerait un mimétisme moléculaire entre les antigènes d'EBV (EBNA 1 et 2 *Epstein Barr Nuclear Antigene*) et une molécule d'adhérence du SNC (GlialCam : *glial cell adhesion molecule*) exprimée par les astrocytes et les oligodendrocytes, et ainsi un risque de réaction croisée ⁽²⁰⁾

L'infection à EBV semble être une condition nécessaire mais pas suffisante au développement d'une pathologie auto-immune comme la SEP.

4.2.3 Tabac

Le risque de SEP chez les sujets fumeurs est 1.5 fois plus élevé que chez les non-fumeurs. ⁽²¹⁾

Plus généralement, le tabac favorise un environnement pro-inflammatoire avec la production de cytokines pro-inflammatoire et un stress oxydatif.

Concernant le rôle du tabac dans la SEP :

- La nicotine et le monoxyde d'azote (NO) sont responsables d'une augmentation de la perméabilité de la BHE favorisant l'entrée des cellules immunitaires auto-réactives (cf A.3) ⁽²²⁾
- Le monoxyde d'azote et le monoxyde de carbone peuvent induire une démyélinisation, une dégénérescence axonale et une perte neuronale ⁽²³⁾
- Il jouerait un rôle dans la méthylation de l'ADN et notamment l'expression des gènes de susceptibilité, et ceci varie proportionnellement à la quantité de tabagisme. ⁽²²⁾

4.2.4 Obésité

Plusieurs études ont mis en évidence un lien entre l'obésité et la SEP où un indice de masse corporelle (IMC) > 30 à l'enfance ⁽²⁴⁾ et à l'adolescence ⁽²⁵⁾ augmente le risque de façon significative de développer la pathologie.

Le tissu adipeux comprend de nombreux types cellulaires. L'obésité modifie l'homéostasie du tissu adipeux en créant un environnement cellulaire davantage pro-inflammatoire impliqué dans la pathogenèse de la SEP et participant au dérèglement du système immunitaire.

4.2.5 Dérégulation du microbiote

Le microbiote, notamment intestinal, participe aux réponses immunitaires innée et adaptative. Une dérégulation du microbiote (augmentation de certaines bactéries comme *Methanobrevibacter* et *Akkermansia*, et diminution d'autres comme *Butyricimonas*) serait constatée chez des sujets atteints de SEP ⁽²⁶⁾. Il s'y associe dans le même temps une diminution des LT régulateurs et une production de LTh17 pro-inflammatoires ⁽²⁷⁾.

5. Diagnostic

Le diagnostic de la SEP repose sur plusieurs faisceaux d'arguments cliniques et paracliniques (IRM et marqueurs dans le LCS) devant répondre à des critères de dissémination spatiale et temporelle qui définissent les critères de MacDonald. La dissémination temporelle ne serait plus nécessaire dans les futurs critères révisés de SEP, présentés au congrès de l'ECTRIMS 2024.

5.1 Signes cliniques

5.1.1 Névrite optique

Elle révèle la SEP dans 1/4 des cas ⁽²⁸⁾.

Dans le cas de la SEP, cette névrite est le plus souvent rétrobulbaire, unilatérale et douloureuse (c'est une douleur péri-orbitaire qui augmente à la mobilisation de l'œil) engendrant notamment une baisse progressive de l'acuité visuelle ne s'étendant pas au-delà de 2 semaines avec une récupération partielle ou complète (80% d'une récupération complète à 6 mois) et une dyschromatopsie rouge/vert. ^{(28) (29)}

5.1.2 Myélite

Elle révèle la SEP dans près de la moitié des cas. Elle a une présentation polymorphe, avec des signes cliniques moteurs déficitaires associés à un syndrome pyramidal et/ou sensitifs de degrés variable, et/ou vésico-sphinctériens. La fréquence des troubles sensitifs observés dans la SEP est associée à une localisation lésionnelle préférentielle cordonale postérieure. ⁽²⁹⁾

Une atteinte médullaire cervicale peut conduire à un signe de Lhermitte qui produit une décharge électrique très brève ressentie le long de la colonne vertébrale lors de la flexion de la nuque ⁽²⁹⁾.

5.1.3 Atteinte du tronc cérébral

Elle est moins fréquente et a également une présentation polymorphe avec des signes cliques divers : diplopie, vertige, déficit sensitif, dysarthrie, paralysie du VI, ophtalmoplégie internucléaire ⁽²⁸⁾.

5.2Diagnostics différentiels

Les diagnostics différentiels sont nombreux ⁽²⁹⁾. Le diagnostic de SEP nécessite d'avoir vérifié l'absence d'arguments pour d'autres pathologies ou atteintes du SNC. Il existe ainsi une série de symptômes/anomalies cliniques et paracliniques (IRM, biologie, etc...) qui peuvent faire évoquer des diagnostics différentiels. A titre d'exemple, en cas de symptomatologie type « névrite optique », l'absence de douleur, la présence d'un œdème papillaire majeur, une atteinte bilatérale d'emblée, l'absence totale de récupération, le caractère étendu et postérieur de la lésion en IRM, doivent faire évoquer de potentiels diagnostics différentiels.

Les diagnostics différentiels comprennent l'ADEM, la MOGAD, la NMOSD, les maladies systémiques avec atteinte du SNC, certains maladies métaboliques (adrénoleucodystrophie et paraparésie progressive de l'adulte par exemple) et génétiques (Leber...), etc ... Il est également important de considérer la possibilité d'observer des hypersignaux T2 de la substance blanche de façon non-spécifique ou parfois dans des contextes dysmétaboliques/vasculaires ^{(30) (31)}.

5.3Formes cliniques

Trois formes cliniques ont été identifiées par *Lublin et al.* en 1996 ⁽³²⁾ : la forme rémittente / récurrente ; la forme secondairement progressive et la forme primaire progressive. Cette classification décrit le phénotype clinique d'un patient mais est finalement peu adaptée lors de la discussion de la mise en place d'une thérapeutique immunomodulatrice ou suppressive. En 2014, les critères de Lublin ont été modifiés en termes de progression et d'activité de la maladie. L'activité de la maladie pouvant se démontrer du point de vue radiologique et/ou clinique ⁽³³⁾.

5.3.1 Forme rémittente / récurrente (RR)

C'est la forme la plus commune représentant 85% des formes de SEP.

Elle est caractérisée par une alternance de phase de poussées et de phase de rémission. La poussée se produit en dehors d'infection ou fièvre, et doit durer au moins 24 heures. Les signes neurologiques sont variés. La poussée peut être monofocale ou plus rarement plurifocale. Les poussées les plus fréquentes sont la névrite optique aiguë unilatérale, la myélite partielle, l'atteinte du tronc cérébral et/ou du cervelet. ⁽³⁴⁾

La phase de rémission peut être totale ou partielle sans argument pour une progression clinique entre les poussées. Cette forme rémittente/récurrente peut être active (cliniquement ou par IRM avec des nouvelles lésions rehaussées ou des lésions qui augmentent de taille), non active, stable ou non stable (sans récupération complète avec aggravation des signes cliniques sur une période donnée). ⁽³⁴⁾

5.3.2 Forme secondairement progressive

Il s'agit de patients qui présentaient une forme RR de la maladie et qui après 10-20 ans de maladie, voient leur handicap augmenter progressivement avec ou sans nouvelle poussée. Cette forme secondairement progressive (SP) peut être considérée stable, en rechute ou en poursuite de progression. Les facteurs de risque de conversion en forme SP sont l'âge élevé lors du début de la phase rémittente/récurrente, le sexe masculin, une augmentation rapide de la charge lésionnelle T2 encéphalique lors des cinq premières années de la maladie et les signes d'atteinte médullaire. ⁽³⁴⁾

5.3.3 Forme primaire progressive (PP)

Cette forme concerne 10% des personnes atteintes de SEP. Elle est caractérisée par une dégradation d'emblée progressive et insidieuse de l'état neurologique. Cette forme pourrait être décrite comme une forme secondairement progressive pour laquelle la phase inflammatoire RR est restée silencieuse. ⁽³⁴⁾

Il existe des formes active (cliniquement ou par IRM), stable, active avec progression (mesurée objectivement via des scores mesurant le handicap comme l'EDSS : *Expanded Disability Status Scale*)), active sans progression. ⁽³⁴⁾

5.3.4 Syndrome cliniquement isolé (CIS)

Le CIS est caractérisé par un symptôme clinique hautement évocateur d'une maladie démyélinisante du SNC mais sans réunir les critères de dissémination spatiale et/ou temporelle de SEP. ⁽³⁴⁾

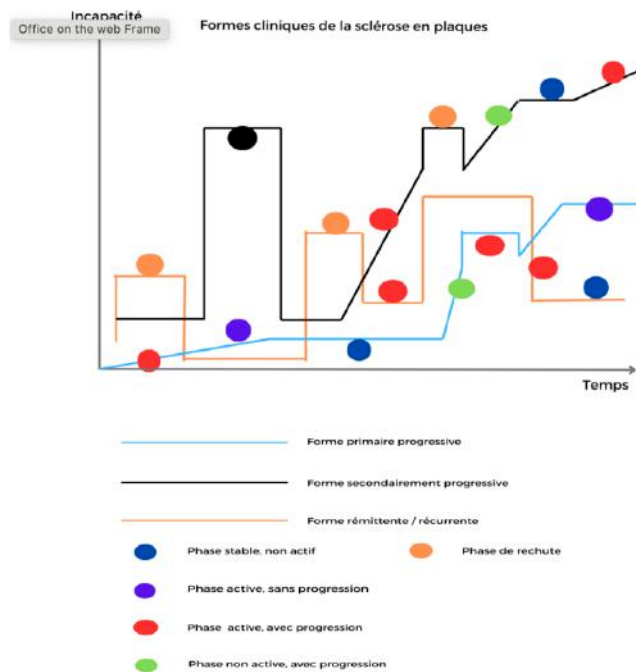
5.3.5 Syndrome radiologiquement isolé (RIS)

Ce syndrome correspond à des anomalies IRM faisant suspecter des lésions démyélinisantes de SEP mais sans signes cliniques (pas d'événement suggestif à l'interrogatoire et examen clinique normal). Différents critères diagnostiques ont été proposés, en reprenant initialement les critères dits de Barkhof (au moins 3 critères sur 4 parmi les suivants ⁽³⁴⁾: (9 lésions en hypersignal T2 ou 1 lésion T1 prenant le contraste, au moins 3 lésions périventriculaires, au moins 1 lésion juxta-corticale) puis en rendant possible le diagnostic de RIS en cas de lésion démyélinisante typique dans une des 4 régions, associée à 2 critères sur 3 parmi les suivants : lésion médullaire, bandes oligoclonales dans le LCS, dissémination temporelle radiologique sur une IRM de suivi ⁽³⁵⁾ . Il existe des facteurs de risques de conversion vers un CIS ⁽³⁴⁾ :

- > 9 lésions en IRM associée à une bande oligoclonale et/ou IgG élevé
- L'atteinte médullaire
- Le sexe masculin

Figure 3 : Résumé des 3 formes cliniques « historiques » de la SEP ⁽³⁴⁾

Adapté du National Multiple Sclerosis Society ⁽³⁴⁾



5.4 Critères de MacDonald

Le diagnostic de la SEP impliquait jusqu'à récemment la présence à la fois d'un critère de dissémination spatiale et temporelle faisant intervenir des signes cliniques, biologiques et IRM, l'ensemble constituant les critères de McDonald. Ils ont été établis pour la première fois en 2001 et actualisés plusieurs fois, la dernière en date en 2017 ⁽³⁶⁾. Une révision des critères diagnostique a été présentée à l'ECTRIMS 2024 ⁽³⁷⁾. Nous décrivons ci-dessous les derniers critères publiés en 2017.

5.4.1 Critères de dissémination spatiale ⁽³⁶⁾

Au moins une lésion symptomatique ou asymptomatique en hypersignal T2 dans 2 de ces 4 régions :

- Périventriculaire
- Corticale / Juxta-corticale
- Sous-tentoriale
- Médullaire

5.4.2 Critères de dissémination temporelle ⁽³⁶⁾

- 1) Une nouvelle lésion en hypersignal T2 ou qui se rehausse sur une IRM de suivi
- 2) Sur une même IRM : la présence d'au moins une lésion qui se rehausse et une qui ne se rehausse pas
- 3) La présence de bandes oligoclonales dans le LCS en l'exclusion d'autres pathologies

5.4.3 Diagnostic de SEP pour les formes RR et PP

5.4.3.1 Forme RR

Tableau 1 : Critères de McDonald d'une forme RR de SEP ⁽³⁰⁾

Adapté de l'article «Multiple Sclerosis: Diagnosis and Differential Diagnosis » par Sami Ömerhoca, Sinem Yazici Akkaş, and Nilüfer Kale İçen ⁽³⁰⁾

Symptômes cliniques	Lésions corrélées en IRM	Données complémentaires
≥ 2	≥ 2	Aucune
≥ 2	1	Dissémination spatiale : apparition d'une nouvelle poussée clinique impliquant une lésion dans un site différent OU par IRM
1	≥ 2	Dissémination temporelle : apparition d'une nouvelle poussée clinique OU par IRM OU par la synthèse intratéchale d'IgG
1	1	Dissémination spatiale : apparition d'une nouvelle poussée clinique impliquant une lésion dans un site différent OU par IRM Dissémination temporelle : apparition d'une nouvelle poussée clinique OU par IRM OU par la présence de bandes oligoclonales surnuméraires dans le LCS

5.4.3.2 Forme primaire progressive

Tableau 2 : Critères de McDonald d'une forme PP de SEP ⁽³⁸⁾

Adapté de Vukusic S « Révisions 2017 des critères diagnostiques pour la Sclérose en Plaques » ⁽³⁸⁾

Clinique	Paraclinique
1 an de progression de la maladie ET	Au moins 2 critères sur 3
	DIS cérébrale : ≥ 1 lésion T2 dans au moins 1 de ces régions : juxta-corticale ou corticale / <u>périventriculaire</u> / <u>sous-tentorielle</u> et médullaire
	DIS médullaire : ≥ 2 lésions T2
	LCS positif : 2 bandes oligoclonales

Lors du congrès de l'ECTRIMS 2024, ont été présentés les futurs critères révisés de SEP. **Le nerf optique** est inclus dans les topographies lésionnelles suggestives de SEP. On en compte dorénavant cinq. Le critère de dissémination temporelle n'est plus nécessaire. Si le patient présente des lésions dans ≤ 3 topographies, des critères supplémentaires tels que des bandes oligoclonales du LCS, signe de la veine centrale, anneau paramagnétique, dissémination temporelle, peuvent être nécessaires. Si le patient est âgé de plus de 50 ans ou s'il présente des facteurs de risque vasculaire significatifs, des critères additionnels sont fortement recommandés (lésion médullaire, bandes LCS, signe de la veine centrale)⁽³⁷⁾.

6. Apport de l'IRM

6.1 Généralités

L'IRM assure un rôle primordial dans la prise en charge de la SEP. Elle est utile lors du diagnostic et lors du suivi afin d'évaluer le pronostic et permet de guider la prise en charge thérapeutique. L'IRM encéphalique a été historiquement incluse pour la première fois dans les critères diagnostique de SEP lors des critères de MacDonald en 2001. Les patients présentant un syndrome clinique isolé typique de SEP (en excluant d'autres pathologies) où les lésions IRM devaient remplir les critères de dissémination spatiale et temporelle.

L'IRM encéphalique est depuis restée une pierre angulaire dans le diagnostic de la SEP. L'amélioration des techniques d'IRM associée à une meilleure compréhension de la physiopathologie de la maladie ont permis avec le temps de réviser au fur et à mesure les critères de McDonald, en allant vers une simplification tout en conservant une bonne sensibilité sans perdre en spécificité.

6.2 Protocoles en IRM

6.2.1 Généralités

L'observatoire Français de la Sclérose en Plaques (OFSEP) étudie et suit une cohorte mère d'environ 90.000 patients en France atteints de SEP. Plus de 40 centres d'imagerie en France contribuent à l'enrichissement de la base de données⁽³⁹⁾. Des protocoles d'IRM standardisés ont été proposés par l'OFSEP selon la chronologie de la maladie et le suivi thérapeutique. A l'heure où les thérapeutiques dans la SEP ont considérablement évolué, il est important que le protocole IRM

comprenne systématiquement des séquences 3D pour détecter toutes les lésions T2 inflammatoires encéphaliques et identifier les signes d'activité, et que le suivi IRM soit régulièrement réalisé ⁽⁴⁰⁾.

6.2.2 Périodicité

L'OFSEP recommande de réaliser une IRM encéphalique lors de la suspicion initiale de SEP puis une (injectée) avant l'introduction d'un nouveau traitement et 6 mois (injectée) après cette introduction. Si absence d'introduction d'un traitement, il est préconisé de réaliser une IRM encéphalique (non injectée) au moins 1 fois tous les 3 ans.

L'IRM médullaire n'est pas nécessaire au diagnostic de la SEP (*cf Critères de McDonald*) mais fréquemment réalisée, étant donné la spécificité des lésions médullaires. Sous traitement, les recommandations OFSEP sont de réaliser une IRM médullaire tous les 3 ans ⁽⁴¹⁾.

6.2.3 Protocoles

6.2.3.1 IRM encéphalique

6.2.3.1.1 Séquences recommandées

Les séquences utilisées de façon standard lors de toutes les phases de la maladie sont les suivantes :

- **Une séquence 3DT1 sans injection** : c'est une séquence anatomique permettant d'évaluer la trophicité cérébrale et de voir les « black holes » (évolution d'une ancienne lésion active vers une nécrose centrale, ces black holes peuvent soit disparaître via un œdème réversible ou une démyélinisation, ou devenir persistante liée à une perte axonale et myélinique importante et irréversible) ⁽⁴²⁾
- **Une séquence diffusion b1000 avec cartographie ADC** (*apparent coefficient diffusion*) (épaisseur de coupe < 3 mm) : permettant des détections précoces de LEMP, d'évaluer la distribution des lésions au cours du temps et d'éliminer les diagnostics différentiels comme des lésions ischémiques aiguës. ⁽⁴²⁾
- **Une séquence 3DT2 FLAIR** : (*three dimensions -Fluid attenuated inversion recovery*) : c'est la séquence clé pour évaluer les critères de dissémination spatiale et temporelle. Le FLAIR permet de mieux détecter les lésions corticales et juxta-corticales (tout comme la séquence 3D DIR). Contrairement au mode 2D

anciennement utilisé, le 3D permet de mieux visualiser les lésions en fosse postérieure. Usuellement, la séquence 3D est réalisée dans le plan sagittal notamment pour explorer une partie de la moelle cervicale. ^{(41) (42)}
Les séquences 3D doivent avoir une résolution de 1 mm³. ⁽⁴¹⁾

6.2.3.1.2 Séquences optionnelles

Certaines séquences jugées optionnelles sont toutefois à recommander lors du diagnostic initial de SEP. :

- **Une séquence T2* ou SWI** : c'est une séquence de sensibilité paramagnétique à la deoxyhémoglobine qui raccourcit la relaxation en T2* faisant apparaître les veines en hyposignal : elle est utilisée lors de l'IRM diagnostique *princeps* de la SEP et n'est pas obligatoire lors du suivi. Cette séquence permet de mettre en évidence le « signe de la veine centrale », les lésions avec anneau paramagnétique et de détecter des microsaignements tout en éliminant certains diagnostics différentiels (par exemple, une vascularite en cas de microsaignements abondants chez un sujet jeune). ^{(42) (43)}
- **Une séquence 3DT1 après injection de gadolinium** 0.1 mmol/kg d'un agent macrocyclique (en spin echo (SE) principalement) permet de remplir le critère de dissémination temporelle lors du diagnostic (à travers la coexistence des lésions rehaussées témoignant d'une lésion récente et non rehaussées témoignant d'une lésion ancienne). Elle permet également d'éliminer certains diagnostics différentiels car usuellement le rehaussement des lésions de SEP est sous forme d'un anneau le plus souvent incomplet. Cette séquence doit également être réalisée 6 mois après l'introduction d'un traitement pour témoigner de son efficacité. ^{(41) (42)}

6.2.3.1.3 Optionnelle

- **Une séquence 3D DIR** : permet de mieux détecter les lésions corticales, du nerf optique et de la moelle épinière cervicale, peut être utilisé en complément de la séquence 3D FLAIR. ⁽⁴²⁾
- **Une séquence axiale 2DTSE DP T2 (ou 3DT2)** : elle permet de mieux visualiser les lésions périventriculaires qui ont de base un signal proche du LCS. La pondération en densité de protons permet de diminuer le signal du LCS et ainsi mieux contraster le signal des lésions périventriculaire du LCS. Elle est à réaliser en option de la séquence 3D FLAIR qui reste plus performante. ⁽⁴²⁾

- **DTI** (*diffusion tensor imaging*) dans au moins 15 directions (avec une résolution d'au moins 2 mm³) : cette séquence permet d'évaluer la structure de la substance blanche en étudiant les mouvements Brownien des particules d'eau. En cas d'atteinte des faisceaux de la substance blanche ou de la membrane axonale (augmentation de la perméabilité de celle-ci) : il y a une augmentation de diffusion des molécules d'eau, de l'ADC et une baisse de la fraction d'anisotropie (et donc une augmentation de l'isotropie où les molécules d'eau diffusent librement dans toutes les directions). ^{(42) (43)}

- **Une spectroscopie** : Celle-ci est rarement réalisée lors d'une suspicion de SEP. Elle pourra être réalisée face à une lésion pseudo-tumorale inaugurale. Elle permet d'étudier certains neurotransmetteurs dont les taux varient selon certaines pathologies cérébrales. Concernant la SEP, la variation pathologique de la concentration de certains neurotransmetteurs peut survenir même lors que la substance blanche ou grise apparaît normale. Chez des sujets atteints de SEP, il a été retrouvé une inversion du rapport entre la concentration du N-acetyl aspartate (produit exclusivement par les neurones donc une baisse de ce neurotransmetteur peut traduire une dégénérescence axonale, une perte neuronale et une dysfonction mitochondriale réversible) et le myo-inositol (qui augmente en cas d'inflammation des cellules gliales). La choline (Cho), impliquée dans le métabolisme de la membrane cellulaire gliale, peut se retrouver augmenter en cas de SEP car l'inflammation, la démyélinisation et la gliose vont augmenter la concentration de ce neurotransmetteur. Il existe également d'autres neurotransmetteurs dont la concentration peut augmenter en cas de SEP comme la creatine and phosphocreatine (tCr) un marqueur de prolifération cellulaire, le glutamate et la glutamine impliqués dans le métabolisme des cellules gliales et enfin le lactate (métabolite du glucose) ⁽⁴⁴⁾

- **MTI** (*magnetization transfer imaging*) : a pour but de quantifier la démyélinisation et la perte axonale ainsi que d'améliorer le contraste des lésions sur les séquences T1 gadolinium et T2 en atténuant le signal de la substance blanche et grise. Elle permet de mesurer les échanges entre les protons libres (eau, graisse) et les protons liés aux macromolécules et de saturer ces derniers via une onde de radiofréquence RF décalée de 1500 Hz par rapport à la fréquence de résonnance des protons libres. ⁽⁴⁵⁾

6.2.3.2 IRM médullaire ⁽⁴²⁾

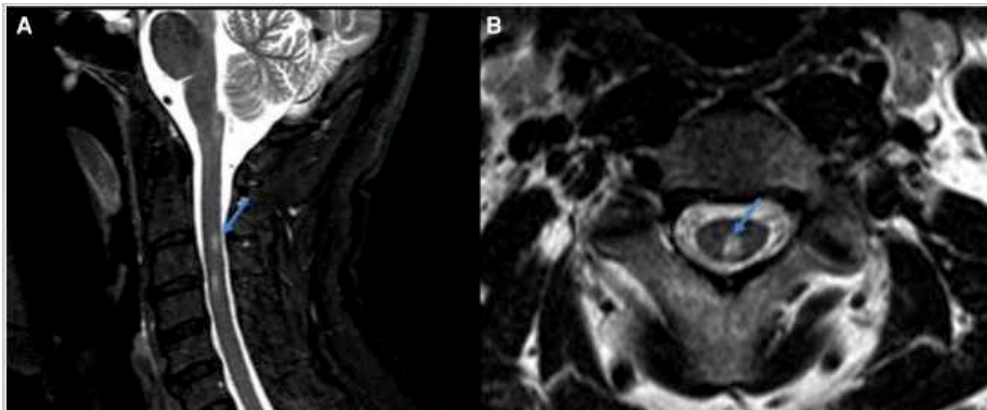
6.2.3.2.1 Séquence obligatoire

- Une séquence sagittale **SE T2**

6.2.3.2.2 Séquence optionnelle

- Une séquence sagittale T1 après injection à réaliser s'il y a la présence d'une lésion et lors de la première IRM médullaire
- Une séquence axiale T1 après injection centrée sur la lésion
- Une séquence axiale T2 EG centrée sur la lésion
- Une séquence sagittale STIR (*short tau inversion recovery*) : plus sensible pour détecter les lésions mais source potentielle d'artefacts.

Figure 4 : Exemple d'atteinte médullaire d'une SEP ⁽⁴⁵⁾



A : Séquence sagittale STIR avec 2 lésions en hypersignal STIR en C3 et C3-C4 // B : Séquence axiale T2 SE avec une lésion à la moelle dorsale en C3

Tableau 3 : Résumé des protocoles OFSEP pour la SEP ⁽⁴¹⁾

Adapté du protocole OFSEP 2018 ⁽⁴¹⁾

	Obligatoire	Optionnel
IRM encéphalique	- 3DT1 sans injection - 3D FLAIR - Axial diffusion b1000 + ADC	- Axial 2D TSE DP/T2 ou 3DT1 - DTI - 2D T2 EG - SWI - 3D DIR
IRM médullaire	- Sagittale T2	- Axial T2 EG - Axial T1 gadolinium - Sagittal STIR
Gadolinium	- Diagnostic initial - Avant l'introduction d'un nouveau traitement - 6 mois après l'introduction d'un nouveau traitement - Si lésion médullaire	- En cas de maladies intercurrentes - Suspicion de LEMP - En cas d'exacerbations

7. Patterns des lésions en IRM de la SEP

7.1 Caractéristiques générales

Les caractéristiques générales des lésions sont les suivantes ⁽⁴⁶⁾ :

- Rondes ou ovoïdes
- Entre 3 mm et 1 cm de diamètre
- Doit être visible sur au moins 2 coupes (pour éliminer les artefacts)
- De répartition encéphalique asymétrique le plus souvent

7.2 Localisations

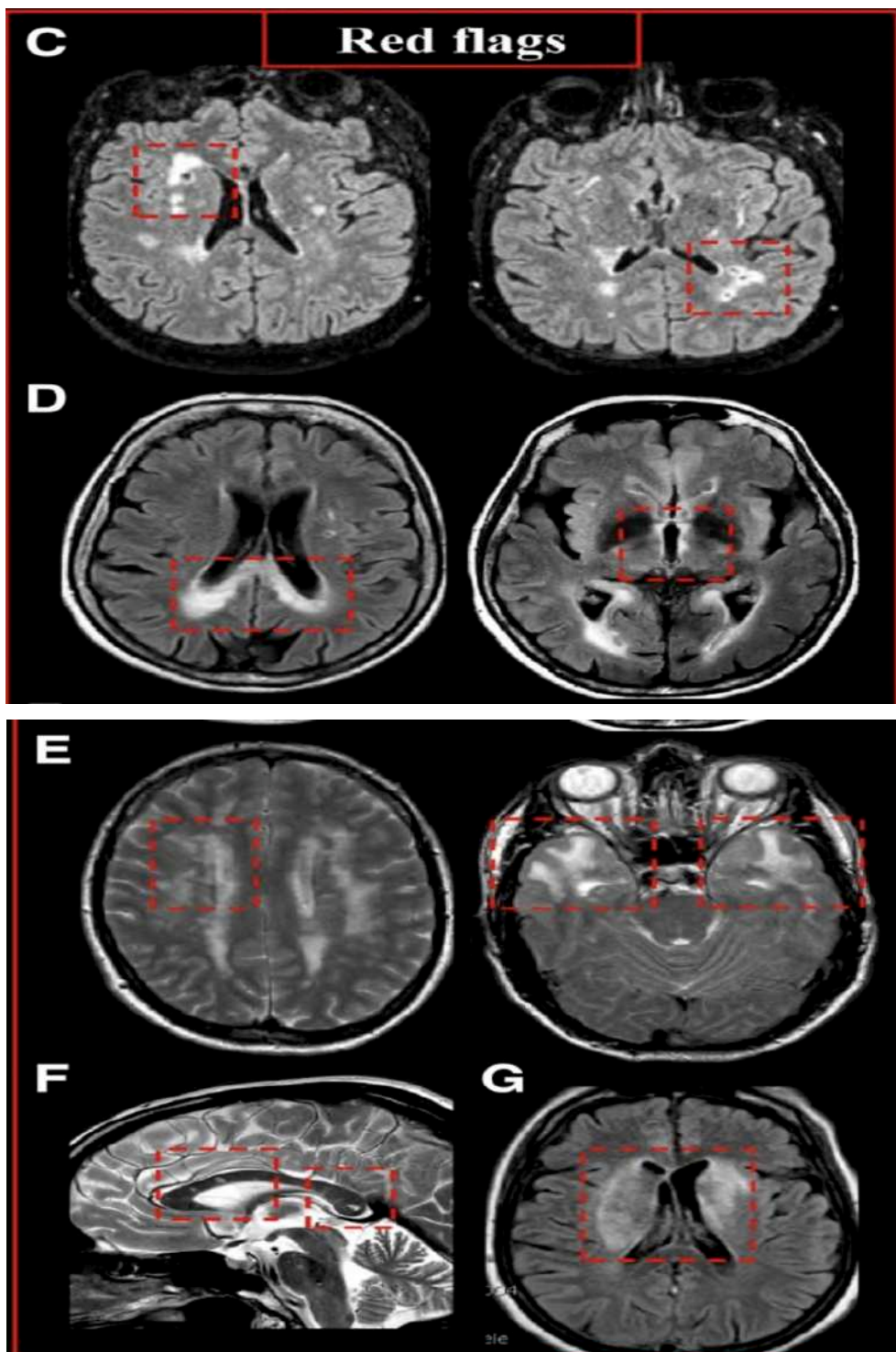
7.2.1 Périventriculaires

Ce sont des lésions hyperintenses en T2 de la substance blanche au contact des ventricules latéraux d'axe perpendiculaire à ceux-ci et/ou au corps calleux.

Il existe cependant plus diagnostics différentiels de ces lésions non-SEP, ce sont des lésions de la substance blanche non périventriculaires (c'est-à-dire séparées du ventricule par de la substance blanche normale) et/ou avec un grand axe non perpendiculaire à l'axe des ventricules ou du corps calleux ⁽⁴⁶⁾ :

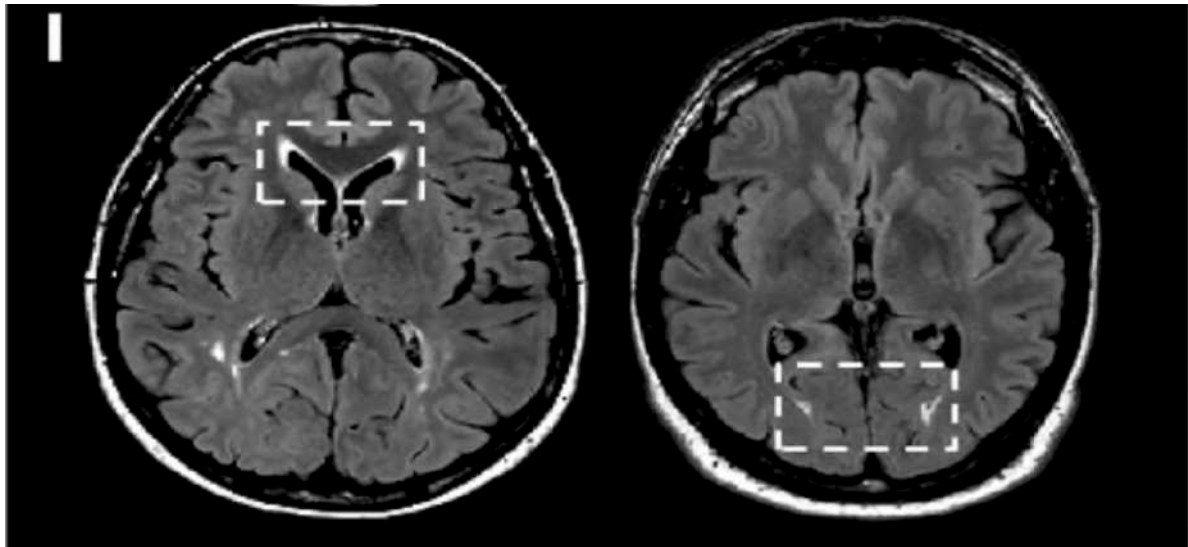
- Multiples lésions de la substance blanche périventriculaire et de la substance grise : maladie des petits vaisseaux
- Extension à la partie postérieure du corps calleux et diencephalique de façon bilatérale : neuromyéélite optique (NMO)
- Multiples lésions affectant la substance blanche, la capsule externe, les lobes temporaux : Cerebral autosomic dominant artériopathy (CADASIL)
- Lésions en boule de neige dans le corps calleux : syndrome de Susac
- Lésions antérieures et postérieures, symétriques, en chapeau des cornes frontales et occipitales en chapeau qui seraient « physiologiques » ⁽⁴⁷⁾.

Figure 5 : : Lésions non-SEP périventriculaires ⁽⁴⁶⁾



C : Multiples hypersignaux FLAIR de la substance blanche périventriculaire et de la substance grise dans le cadre d'une maladie des petits vaisseaux. // D : Hypersignaux FLAIR à la partie postérieure du corps calleux et diencephalique dans le cadre d'une neuromyéélite optique // E : Multiples hypersignaux T2 de la substance blanche profonde, de la capsule externe, des lobes temporaux avec des infarctus sous-corticaux et leucoencéphalopathie // F : Lésions en hypersignal T2 dans le corps calleux en boule de neige dans le cadre du syndrome de Susac // G : Lésions diffuses et extensives en hypersignal FLAIR affectant la substance blanche et grise dans le lupus érythémateux.

Figure 6 : Lésions périventriculaires physiologiques “capping” ⁽⁴⁶⁾



7.2.2 Lésions corticales et juxta-corticales

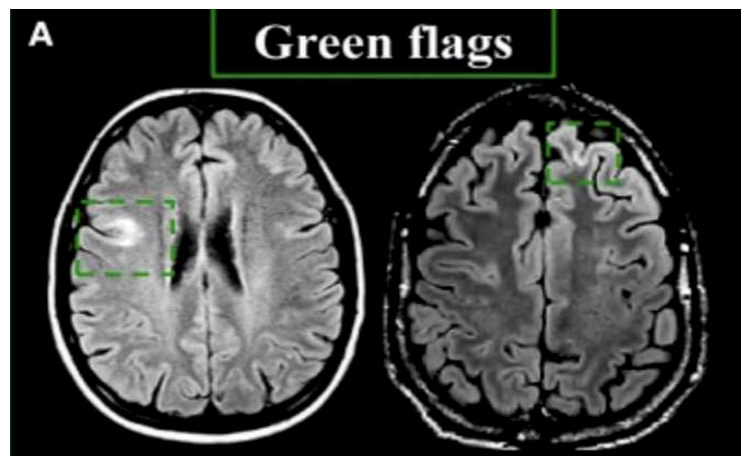
Les lésions juxta-corticales affleurent le cortex sans interposition de substance blanche normale entre les 2. Elles touchent les fibres en U. ⁽⁴⁶⁾

Les lésions corticales pénètrent le cortex et/ou la substance blanche adjacente.

Il existe plusieurs types de lésions corticales ⁽⁴⁶⁾ :

- Type I : Lésions cortico-sous-corticales : affectent la substance blanche et grise
- Type II : Petites lésions périveineuses intra-corticales ne touchant pas la substance blanche et la surface piale
- Type III : Démyélinisation sous-piale (le plus fréquent)
- Type IV : S'étend à toute la largeur du cortex sans atteindre les bords par la substance blanche.

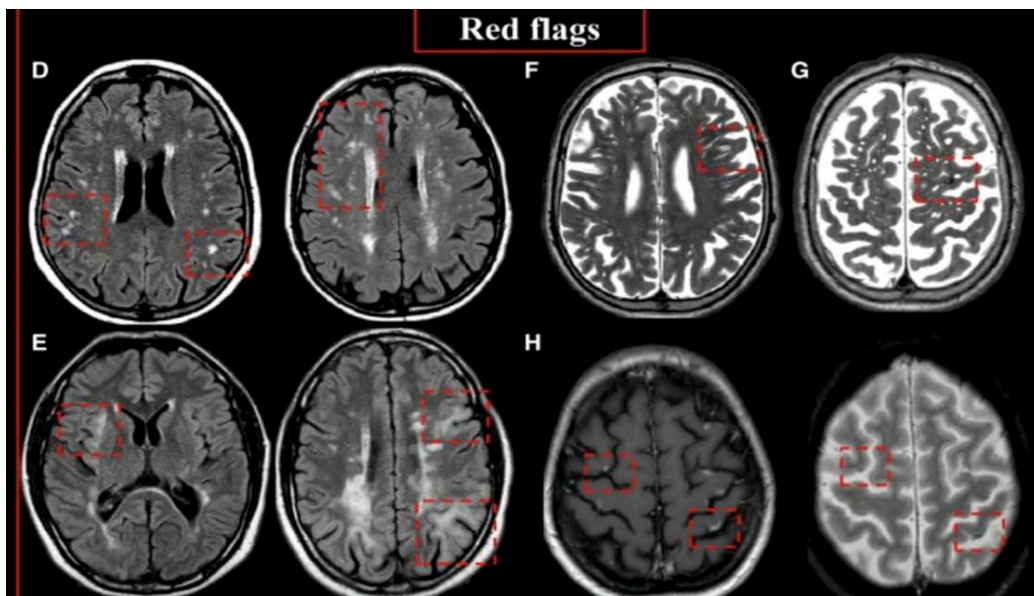
Figure 7 : Lésions corticales dans le cadre de la SEP ⁽⁴⁶⁾



Il existe plusieurs diagnostics différentiels des lésions corticales et juxta-corticales non-SEP : ⁽⁴⁶⁾

- Maladie des petits vaisseaux : multiples lésions de la substance blanche sous-corticale et substance blanche profonde
- Leucoencéphalopathie multiple progressive
- Nécrose laminaire corticale
- Espace de Virchow-Robin dilaté
- Microbleeds
- Vascularite : multiples hyperintensités T1 leptoméningées / corticales avec hypointensité en echo de gradient

Figure 8 : Lésions corticales ou juxta-corticales non-SEP ⁽⁴⁶⁾

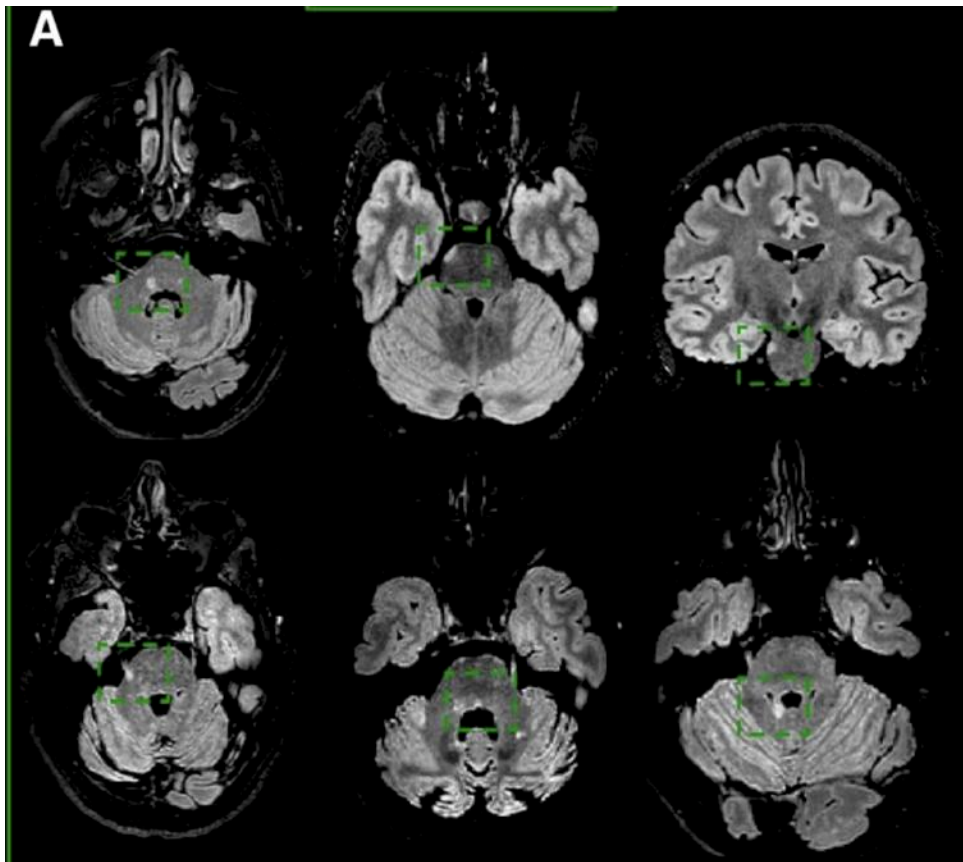


D : Multiples petites lésions en hypersignal FLAIR de la substance blanche et sous-corticale suggérant une maladie des petits vaisseaux // E : Lésions en hypersignal FLAIR de la substance grise-blanc sans interposition de substance blanche normale dans le cadre d'une LEMP // F : Multiples lésions en hypersignal T2 bien limitées qui apparaissent comme des points ou des bandes dans des espaces de Virchow Robin dilaté // G: Hypointensité T2 suggérant des dépôts d'hémosidérine // H : Multiples lésions hyperintenses en T2 et hypointenses en echo gradient leptoméningées / corticales dans le cadre d'une vascularite.

7.2.3 Lésions infratentorielles ⁽⁴⁶⁾

- Tronc cérébral : unique, parfois limitée à une discrète ligne sous-piale
- Pont : contigu avec la citerne, la surface pontique et le tractus trigéminin intra-pontique
- Moelle allongée
- Pédoncules cérébraux
- Substance blanche cérébelleuse : pédoncules cérébelleux moyens et supérieurs

Figure 9 : Lésions en hypersignal FLAIR infratentorielles typiques de SEP ⁽⁴⁶⁾



Les diagnostics différentiels de lésions infratentorielles sont ⁽⁴⁶⁾ :

- MOGAD
- Leucoencéphalopathie multifocale progressive
- NMOSD : en périaqueducal et à la partie postérieure du tronc cérébral adjacent au 4^{ème} ventricule (incluant l'aera postrema et le tractus solitaire)
- Clippers (*chronic lymphocytic inflammation with pontine perivascular enhancement responsive to steroids syndrome*): les lésions sont de type miliaire.

Figure 10 : Lésions en hypersignal FLAIR et T2 centro-pontine en rapport avec une maladie des petits vaisseaux ⁽⁴⁶⁾

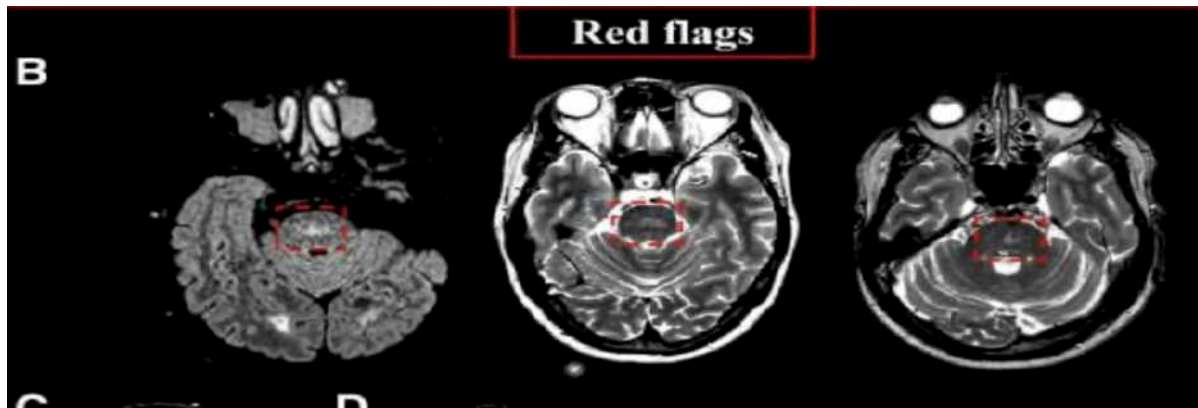
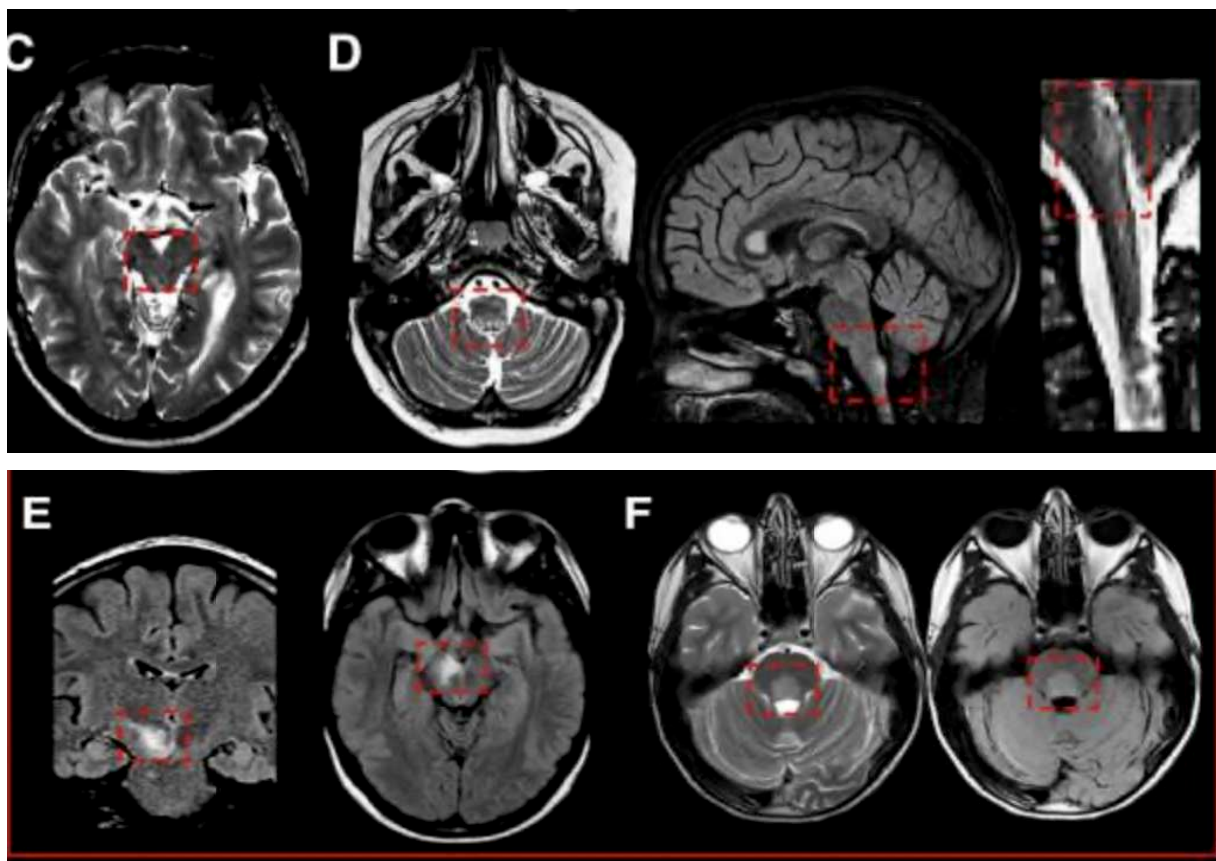


Figure 11 : Lésions infratentorielles non-SEP ⁽⁴⁶⁾



C : Lésion péri-aqueducule en hypersignal flair et T2 avec une lésion médullaire en hypersignal T2 impliquée dans la neuromyéélite optique // D : Lésion en hypersignal T2 de l'aera postrema impliquée dans la neuromyéélite optique // E : Lésion mésencéphalique et diencephalique impliquée dans le syndrome anti-MOG. // F : Lésion ovoïde en hypersignal FLAIR et T2 sur la base du 4ème ventricule dans le cadre d'une maladie de Behçet

7.2.4 Lésions médullaires ⁽⁴⁶⁾

- Sont souvent multiples
- Plus fréquentes en cervical
- Doit être vues dans au moins 2 séquences ou dans 2 plans
- Petites (mais > 3mm)
- Moins de 2 vertèbres
- La plupart à la périphérie de la moelle spinale (cordons latéraux ou postérieurs)
- Doivent être focales et bien limitées
- Hypointense T1 et hyperintense T2

Figure 12 : Lésions médullaires typiques de SEP ⁽⁴⁶⁾

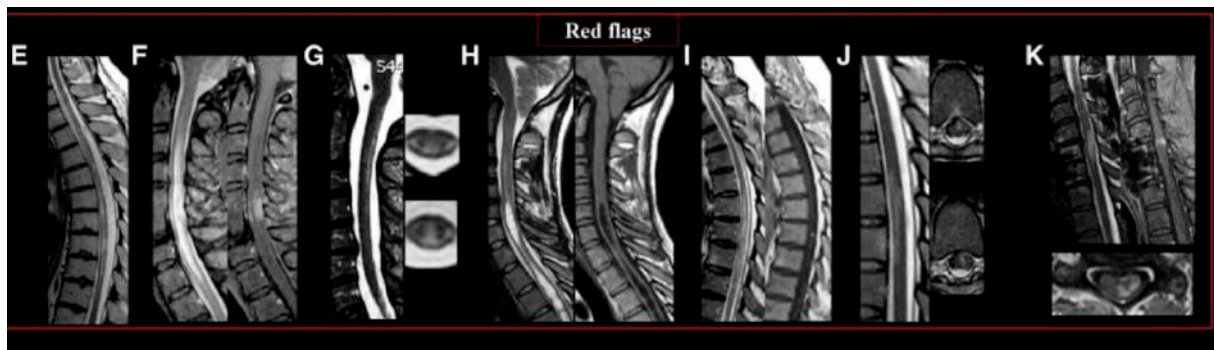


A: Séquence sagittale STIR avec une lésion en hypersignal de la moelle cervicale // B : Séquence sagittale T1 et T2 avec une lésion en hypersignal T2 et hyposignal T1// C : Séquence axiale avec hypersignal centro-latéral en T2 et PSIR (phase sensitive inversion recovery)

Les diagnostics différentiels des lésions médullaires non-SEP ⁽⁴⁶⁾ :

- NMOSD, MOGAD : affectent plus de 3 vertèbres et les lésions se localisent dans la substance grise centrale et à la portion cervico-thoracique pour la NMOSD et la partie caudale de la moelle pour la MOGAD
- Neurosarcoïdose
- Infection
- Néoplasique : leptoméningé ou racines nerveuses
- Cavitité : syringomyélie
- Substance blanche étendue : carence en vitamine B12
- Ischémie : les 2/3 antérieurs de la moelle spinale
- Myélopathie cervicoarthrosique

Figure 13 : Lésions médullaires non-SEP ⁽⁴⁶⁾



E : Séquence sagittale T2 : myélite transverse étendue atteignant plus de 3 vertèbres dans le cadre d'une NMO // F : myélite extensive atteignant plus de 3 vertèbres avec rehaussement leptoméningé et périphérique dans le cadre d'une neurosarcoïdose // G : Hypersignaux T2 postéro-latéraux de la moelle : neurodégénérative // H : Cavité de syringomyélie // I : Hypersignal T2 étendu avec rehaussement tortueux et punctiforme dans le cadre d'une fistule artério-veineuse // J : Lésion en hypersignal dans la partie antérieure de la moelle thoracique s'étendant sur plus de 2 vertèbres dans le cadre d'une myélopathie subaiguë ischémique // K : Lésion en hypersignal T2 se rehaussant en gâteau dans le cadre d'une myélopathie cervicoarthrosique.

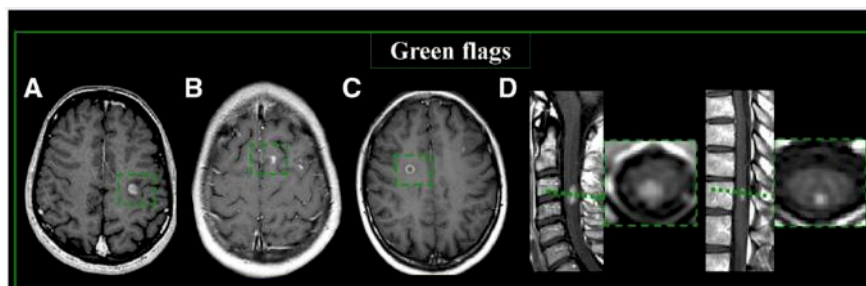
7.2.5 Lésions rehaussées ⁽⁴⁶⁾

- Peuvent témoigner de la dissémination temporelle avec des lésions qui se rehaussent et d'autres non (cf critères révisés de McDonald).

Le rehaussement d'une nouvelle lésion de SEP se voit jusqu'à 4 semaines. Au-delà de 3 mois, il faut savoir évoquer un diagnostic différentiel (sarcoïdose, anomalie veineuse de développement, télangiectasie capillaire).

Une lésion de SEP qui se rehausse après injection de produit de contraste (>5minutes) est une zone d'au moins 3 mm hyperintense en T1 spin echo ou echo-de-gradient rehaussée de façon homogène où s'associe des anomalies de signal en T2 ou T2 FLAIR.

Figure 14 : Lésions réhaussées caractéristiques d'une SEP ⁽⁴⁶⁾

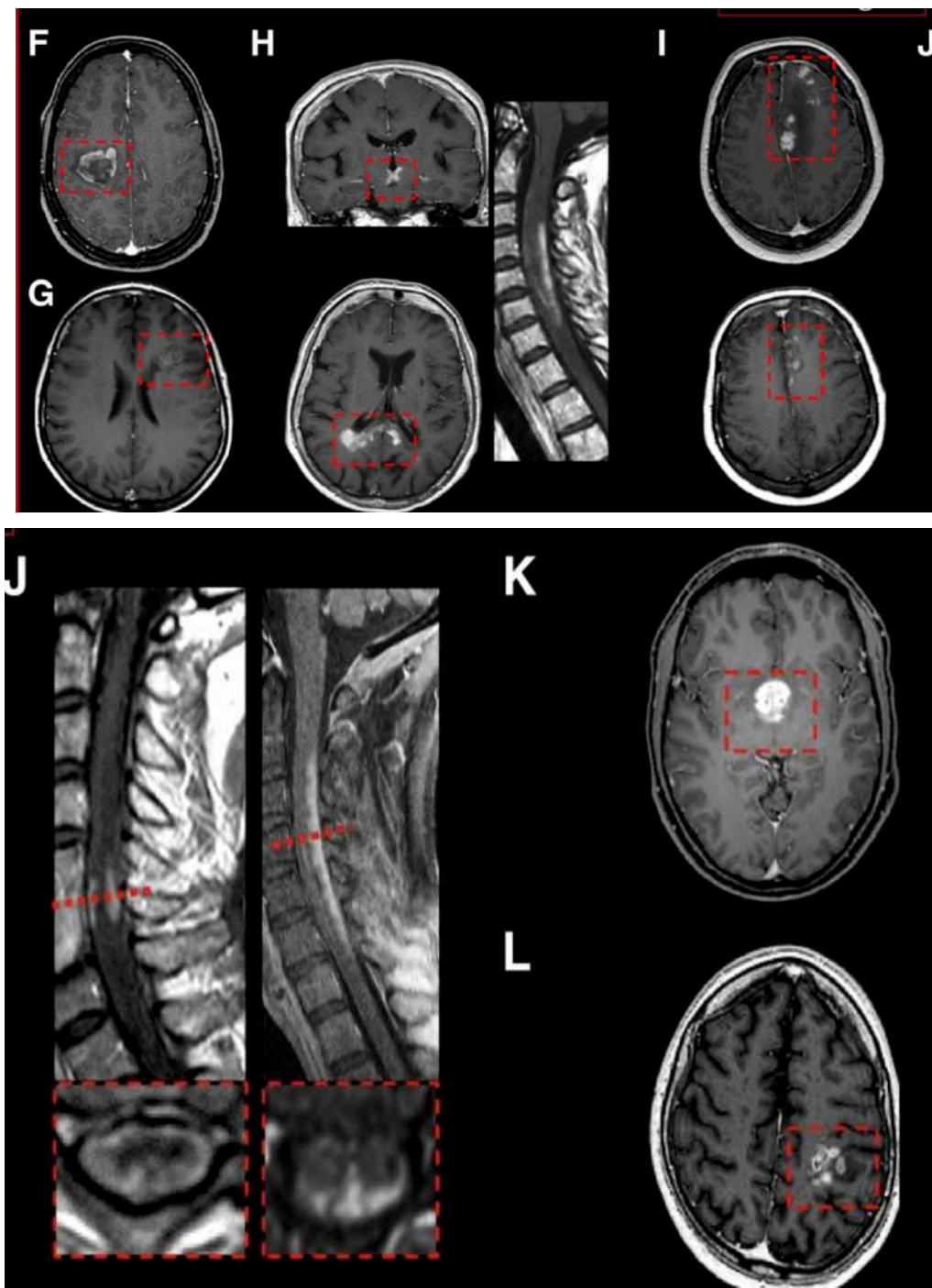


A : Rehaussement nodulaire // B : Rehaussement en anneau ouvert // C : Rehaussement en anneau // D : Rehaussement nodulaire de la moelle cervicale

Les diagnostics différentiels des lésions rehaussées non-SEP sont ⁽⁴⁶⁾

- Rehaussement punctiforme ou miliaire : CLIPPERS, vascularite, LEMP (reconstitution immunitaire), syndrome de Susac
- Rehaussement en bande : Sclérose de Balo
- Rehaussement leptoméningé : les maladies granulomateuses
- Rehaussement strictement cortical : accident ischémique aiguë avec rupture de BHE
- Rehaussement de type patchy : télangiectasie
- Rehaussement subpial associé au signe du trident : neurosarcoïdose, déficit en vitamine B12
- Rehaussement irrégulier et inhomogène : glioblastome
- Rehaussement circonférentiel (SB) complet (« donnut ») ou incomplet (« croissant ») prédominant sous le site de compression maximale : myélopathie cervicoarthrosique

Figure 15 : Lésions rehaussées non-SEP ⁽⁴⁶⁾

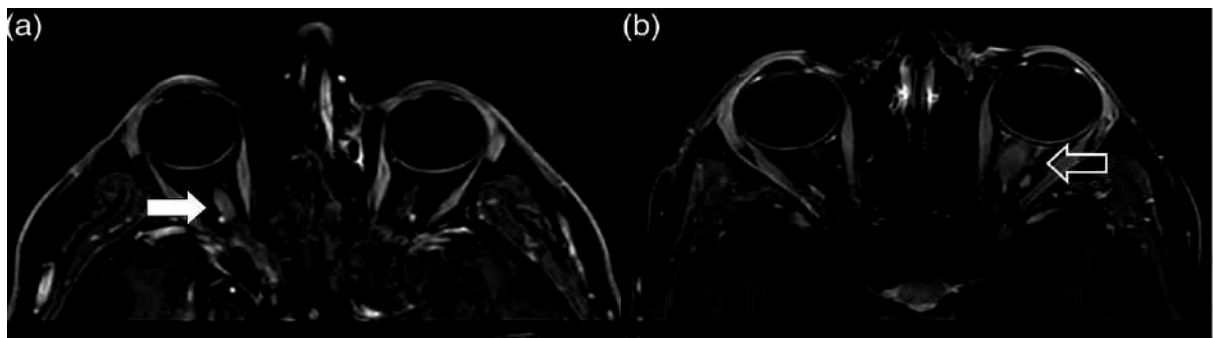


F : Rehaussement inhomogène d'une vaste lésion : lésion inflammatoire démyélinisante idiopathique // G : Rehaussement en bande dans la sclérose de Baló // H : Rehaussement du diencéphalique, du corps calleux et étendue de la moelle spinale dans le cadre de la NMO // I : Rehaussement irrégulier corticale et sous-cortical leptoméningé dans le cadre d'une vascularite // J : Rehaussement pial et leptoméningé avec le signe du trident dans la neurosarcoïdose // K : Rehaussement homogène diencéphalique dans le cadre d'un syndrome paranéoplasique anti-Ma2 // L : Rehaussement irrégulier et inhomogène dans le cadre d'un glioblastome

7.2.6 Lésions du nerf optique

Bien que la NORB représente 25% des premiers événements neurologiques inflammatoires, le nerf optique ne faisait pas partie des topographies lésionnelles évocatrices dans les critères révisés de McDonald 2017. Lors de la prochaine publication des critères révisés en 2024 ou 2025, le nerf optique est inclus dans les critères diagnostic. Dans le cadre de la SEP : les lésions sont hyperintenses en T2, se rehaussent et atteignent le plus souvent la partie antérieure du nerf optique ⁽⁴⁶⁾

Figure 16 : Névrite optique dans le cadre d'une SEP ⁽⁴⁸⁾



A) Séquence axiale T1 Fat-Sat gado : rehaussement de la portion antérieure du nerf optique droit // B) Séquence axiale T1 FS gado : rehaussement de la portion antérieure du nerf optique gauche

Les diagnostics différentiels sont multiples ⁽⁴⁶⁾ :

- Atteinte postérieure et chiasmatique : Ac anti-aquaporine 4 dans la NMO
- Absence d'hypersignal T2 à la phase aiguë : ischémique ou toxique
- Rehaussement des tissus mous péri-nerveux, péri-orbitaire, de l'apex orbitaire ou les sinus caverneux : cause granulomateuse ; tumorale ; infectieux ; MOGAD.

Figure 17 : Névrite optique dans le cadre d'une NMO et ischémique ⁽⁴⁹⁾⁽⁵⁰⁾



Image de droite ⁽⁴⁹⁾ : Séquence axiale T1 FS gado avec rehaussement bilatéral de la partie postérieure des nerfs optiques dans un contexte de NMO // Image de gauche : neuropathie optique ischémique bilatérale avec hypersignal diffusion et restriction à la cartographie ADC ⁽⁵⁰⁾

7.2.7 Signe de la veine centrale

- Il est retrouvé dans 80% des lésions de SEP mais est moins fréquente dans les lésions corticales comparativement à la substance blanche profonde et périventriculaire. ⁽⁴⁶⁾
- C'est une fine ligne hypointense (en SWI ou T2*) linéaire visible dans au moins 1 plan ou punctiforme < 2 mm visible dans au moins 2 plans pénétrant au centre de la lésion ⁽⁴⁶⁾. Ce signe est davantage visible en 3T ou 7T grâce à une augmentation de l'effet de susceptibilité magnétique.
- Le signe de la veine centrale ne peut pas être considéré comme tel pour les lésions de SEP < 3 mm, les lésions confluentes ou les lésions où il y a plusieurs veines au sein de la lésion ⁽⁴⁶⁾.
- 3 méta-analyses (*Bhandari et al.* en 2020, *Castellaro et al.* en 2020 et *Suh et al.* en 2019) avaient démontré la pertinence de ce signe dans le diagnostic de la SEP avec une sensibilité à 95% et une spécificité à 92% ⁽⁵¹⁾

Figure 18 : Signe de la veine centrale ⁽⁵²⁾



Séquence axiale SWI : la flèche rouge désignant le signe de la veine centrale, parallèle au grand axe de la lésion

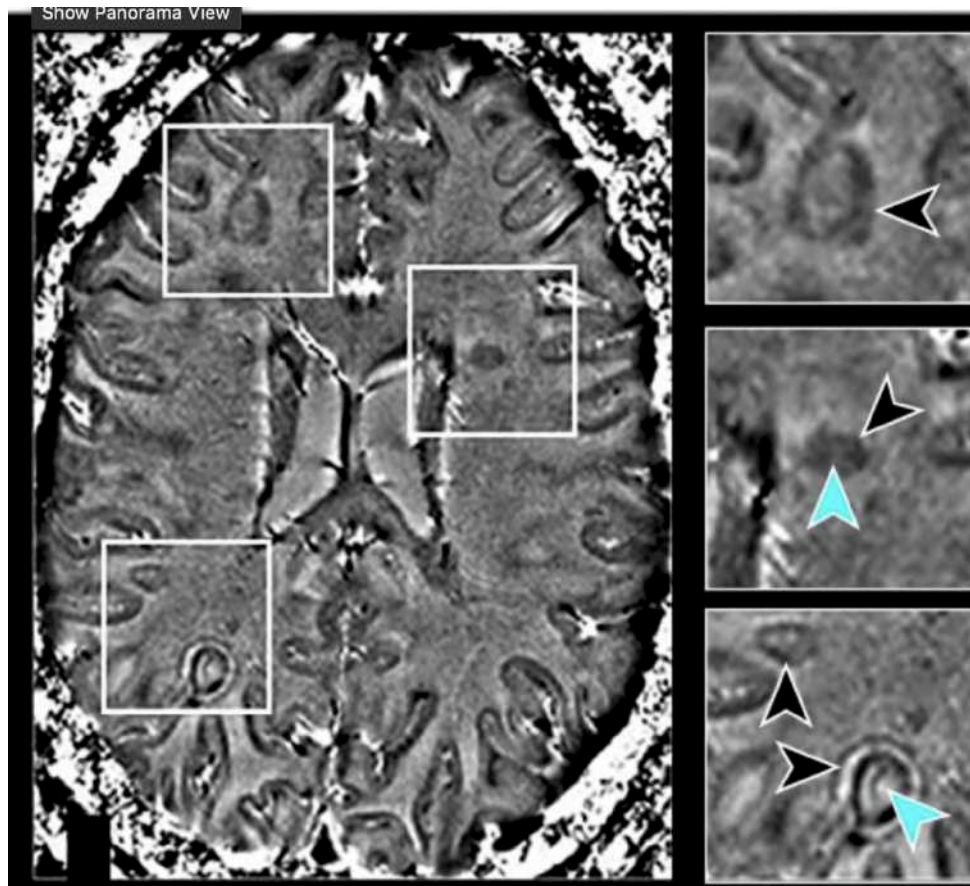
7.2.8 Démyélinisation sous-piale ⁽⁴⁶⁾

- Elle est souvent associée aux lésions corticales et à une inflammation méningée
- Elle est très difficile à détecter même avec des techniques avancées.
- Elle est vraisemblablement mieux identifiée en IRM à ultra-haut champ magnétique.

7.2.9 Anneau paramagnétique ⁽⁴⁶⁾

- Concerne les lésions chroniques actives d'évolution lente avec le temps mais aussi certaines lésions aiguës récentes
- Anneau périphérique comprenant de la microglie/ des macrophages riches en Fer, des LT impliqués dans la perte neuronale et la démyélinisation
- En IRM : il est en hyposignal en T2* ou séquence de susceptibilité magnétique, en périphérie d'une lésion

Figure 19 : Anneau paramagnétique d'une lésion de SEP ⁽⁵³⁾



8. Pronostic

8.1 Espérance de vie et risque de mortalité

D'après une cohorte suivant 1388 patients en Norvège présentant une SEP pendant 60 ans, l'espérance de vie chez les sujets atteints de SEP est raccourcie en moyenne de 7 ans (sans cause spécifique de décès significative mise en évidence), avec une espérance de vie plus courte pour les hommes par rapport aux femmes, comparativement à des sujets indemnes de la maladie : l'âge médian du décès étant de 72 ans chez les sujets atteints de SEP (dont 77 ans pour les femmes et 72 ans pour les hommes) contre 82 ans chez les sujets sains. Le risque de mortalité globale (toutes formes de SEP confondus) est de 2.7 ⁽⁵⁴⁾ Cependant ce risque de mortalité diminue au fur et à mesure des périodes de l'étude où à titre d'exemple, il était de 3 entre 1953 et 1974 contre 0,7 entre 1997 et 2012, et il varie également selon la forme de SEP où il est plus élevé en cas de formes PP (2,4 pour les formes RR versus 3,9 en cas de forme PP). Cette variation peut être imputée à l'introduction des nouvelles thérapeutiques qui ont un impact positif sur les formes RR. ⁽⁵⁴⁾

8.2 Morbidité

L'Expanded Disability Status Score (EDSS) est un score établi par Kurtzke en 1983 évaluant le handicap dans la SEP. C'est un score allant de 0 à 10 avec une incrémentation de 0,5 points ⁽⁵⁵⁾, prenant en compte l'atteinte de certaines fonctions (ou signes fonctionnels (SF)) que sont la fonction pyramidale, cérébelleuse, sensitive, du tronc cérébral, sphinctérienne, visuelle, cérébrale et mentale ⁽⁵⁶⁾. A partir d'un score de 4.0, le périmètre de marche est limité. A partir de 6.0, le patient nécessite une assistance pour se déplacer ⁽⁵⁵⁾.

Figure 20 : Échelle EDSS ⁽⁵⁶⁾ ⁽⁵⁷⁾

Adaptés de National Multiple Sclerosis Society et Heathline

Score	Signes fonctionnels
0	Examen neurologique normal => Tous les SF à 0
1	Absence de handicap fonctionnel avec signes minimes d'atteinte d'une des fonctions => 1 seul SF à 1 sauf le SF mental
1,5	Absence de handicap fonctionnel, signes minimes dans plus <u>d'un SF</u> => Plus de 1 SF à 1 sauf le SF mental)
2	Handicap minime d'un des SF => 1 SF à 2
2,5	Handicap minime dans 2 SF => 2 SF à 2
3	Handicap modéré dans un SF ou atteinte minime de 3 ou 4 fonctions mais malade ambulateur => 3 ou 4 SF à 2
3,5	Totalement ambulateur mais atteinte modérée dans <u>un SF</u> => 1 ou 2 SF à 2 ou 2 SF à 3 ou 5 SF à 0
4	Malade totalement autonome pour la marche vaquant à ses occupations 12h/ jour malgré une gêne fonctionnelle importante. Peut marcher 500 m sans aide ni repos => 1 SF à 4
4,5	Malade autonome pour la marche Vaquant à ses occupations toute la journée Handicap relativement sévère. Peut marcher 300 m sans aide ni repos => 1 SF à 4
5	Peut marcher 200 m sans aide ni repos Handicap suffisamment sévère pour entraver l'activité d'une journée normale => 1 SF à 5
5,5	Peut marcher 100 m sans aide ni repos => 1 SF à 5
6	Peut marcher 100 m avec aide unilatérale (béquille, canne, canne anglaise) et/ou avec repos intermédiaire => Plus de 2 SF à 3 ou +
6,5	Peut marcher 20 m avec aide permanente bilatérale et/ou avec repos intermédiaire => Plus de 2 SF à 3 ou +
7	Incapable de marcher plus de 5m même avec aide Confiné au fauteuil roulant pendant au moins 12h / j et effectue le transfert => Plus d'un SF à 4 ou +
7,5	Incapable de faire quelques pas Strictement confiné au fauteuil roulant A parfois besoin d'aide pour le transfert => Plus d'un SF à 4 ou +
8	Confiné au lit ou au fauteuil et garde l'usage des bras => Pus de 2 SF à 4 ou plus d'un SF à 4 ou +
8,5	Confiné au <u>lit</u> , garde un usage partiel des bras et conserve quelques fonctions élémentaires => Plus d'un SF à 4 ou +
9	Patient grabataire Peut communiquer et manger => Plus d'un SF à 4 ou +
9,5	Patient totalement impotent Ne peut plus manger ni avaler ni communiquer => SF à 4 ou + dans tous les systèmes
10	Décès lié à la SEP

8.3 Facteurs pronostics

A travers quelques éléments cités-ci dessus, il peut être établi des facteurs pronostics de la maladie ⁽⁵⁸⁾

8.3.1 Les facteurs de bons pronostics

- Pas d'handicap après 5 ans
- Une IRM initiale normale
- Récupération complète après le premier épisode neurologique
- Névrite optique ou un symptôme sensitif isolé comme premier signe clinique
- Intervalle élevé avant la prochaine rechute
- Forme rémittente/récurrente

8.3.2 Les facteurs de mauvais pronostics

- Un taux de rechute élevé après 2 à 5 ans
- Handicap important après 5 ans
- IRM initiale anormale
- Atteinte médullaire ou sous-tentorielle à l'IRM

9. Traitement

Les traitements peuvent être divisés en 3 catégories :

- Les traitements d'une poussée
 - Les traitements symptomatiques
 - Les traitements de fond
- :

9.1 Traitement de poussées

Il est basé sur la méthylprednisolone par voie IV ou voie orale (efficacité identique des 2 voies d'administration) pendant 3 à 5 jours qui vont uniquement diminuer la durée de la poussée, sans interférer sur l'évolution de la maladie.

Des échanges plasmatiques sont parfois pratiqués (de façon concomitante à la corticothérapie, ou seule) en cas de poussée sévère ⁽⁵⁸⁾ mais n'ont pas démontré leur efficacité.

9.2 Traitement symptomatique

Les traitements symptomatiques ont pour but d'améliorer la qualité de vie en agissant sur certains signes fonctionnels comme la douleur, la fatigue, la spasticité, les troubles urinaires ou digestifs.

9.3 Traitement de fond

Les traitements de fond ont pour but de modifier l'évolution de la maladie.

Il existe de plus en plus de traitements depuis ces 20 dernières années dont la sélection du traitement administré peut se faire selon plusieurs paramètres comme l'âge du patient, les comorbidités du patient, la balance bénéfique/risque (effets indésirables) et l'activité clinique et radiologique de la maladie.

Ces traitements ont un mode d'action qui diffère en agissant sur un maillon de la chaîne physiopathologique de la SEP (cf A.3), à titre d'exemple (liste non exhaustive) ⁽⁵⁸⁾ :

- Interféron β : c'est un immunomodulateur qui inhibe l'interféron γ et la production de cytokine pro-inflammatoire
- Acétate de glitamer : bloque la présentation des Ag de la myéline (antigène du soi) aux LT
- Teriflunomide : inhibe la pyrimidine impliquée dans la synthèse des leucocytes.
- Dimethyl fumarate : active la voie transcriptionnelle du facteur nucléaire NRF2 (érythroid-derived-2-like 2)
- Fingolimod : modulateur/antagoniste fonctionnel des récepteurs de la sphingosine-1-phosphate, présents sur les lymphocytes induisant leur rétention dans le thymus et les ganglions lymphatiques et empêchant leur sortie et donc de passer la BHE et ainsi pénétrer le SNC.
- Natalizumab : Ac monoclonal humanisé anti-intégrine $\alpha 4$ (présent à la surface des Lc) et va empêcher leur migration du sang vers le SNC. Il peut être la cause de de survenue de LEMP .
- Ocrelizumab : Ac monoclonal humanisé anti-CD20 qui va induire une déplétion des LB (IV semestriel)
- Ofatumumab : Ac monoclonal humain anti-CD20 (sous-cutané mensuel)
- Alemtuzumab : Ac monoclonal humanisé anti-CD52 présent sur les LB et les LT.

Les lésions de SEP sont dans l'immense majorité des cas, « silencieuses » cliniquement lorsqu'elles apparaissent en IRM. Pour autant l'accumulation progressive au cours du temps de ces lésions silencieuses est néfaste et cause de perte neuro-axonale. Il a été démontré qu'un des meilleurs critères prédictifs de progression du handicap était l'augmentation de la charge lésionnelle T2 encéphalique. Les critères de Rio et de Sormani ⁽⁵⁹⁾ ont mis en avant l'intérêt d'envisager un switch thérapeutique précoce en cas d'activité inflammatoire radiologique et/ou clinique, le risque étant maximal quand les deux activités sont présentes. Les critères NEDA (*no evidence of disease activity*) sont un des objectifs thérapeutiques à atteindre si on souhaite que le patient conserve un handicap neurologique le plus stable possible au cours du temps. Il est devenu ainsi justifié et nécessaire de réaliser un suivi radiologique IRM annuel à tous les patients atteints de SEP, de façon associée au suivi neurologique clinique régulier.

B. Intelligence artificielle

1. Introduction

1.1 L'intelligence artificielle dans le monde de la santé en général

L'implantation des nouvelles technologies et par extension le développement de l'intelligence artificielle (IA) concernent divers domaines et spécialités de la santé comme par exemple la cardiologie avec la détection automatisée de la fibrillation atriale à travers une montre ou un téléphone portable connecté, l'endocrinologie avec le monitoring de la glycémie via des téléphones intelligents ou encore en neurologie avec le monitoring et la détection des crises d'épilepsie. ⁽⁶⁰⁾

Aux Etats-Unis, une liste officielle de logiciels d'intelligence artificielle ayant la permission d'être mis sur le marché, est actualisée et autorisée par la *Food and Drug Administration* (FDA) où actuellement 201 algorithmes d'intelligence artificielle sont concernés. ⁽⁶¹⁾

1.2 Dans le monde de la radiologie

Dans le monde de la radiologie, il existe un accroissement des demandes d'imageries où leurs interprétations peuvent être soumises à des erreurs associées à

une variabilité intra et inter-observateur. L'IA pourrait être une aide complémentaire pour essayer de pallier ces problèmes.

Pour le moment, son utilisation est assez mitigée voire controversée, par exemple selon un sondage réalisé par l'ACR (*American College of Radiology*) en 2020 concernant 1427 radiologues, l'utilisation de l'IA concerne 33,5% des radiologues et 94% d'entre eux qui l'utilisent rapportent une inconstance de leur performance ⁽⁶²⁾. Un autre sondage concernant l'ECR en 2018 (*European Society of Radiology*) va également dans ce sens où 40% des 690 radiologues interrogés ont déjà eu affaire à l'IA et la moitié d'entre eux ne souhaitent pas en faire une pratique récurrente ⁽⁶³⁾.

1.2.1 L'apport de l'IA dans l'interprétation d'examen

1.2.1.1 Sénologie

Le cancer du sein est le cancer le plus fréquent chez les femmes ainsi l'IA est utilisée en sénologie pour la détection des cancers du sein et représentent 15% des logiciels d'IA sur le marché ⁽⁶⁴⁾.

D'après une étude rétrospective concernant 170230 mammographies, l'étude de *Kim et al.* avait montré que *Lunit Insight MMG* © améliore la performance de 6 radiologues spécialisés et 6 radiologues non spécialisés pour l'interprétation des mammographies ($AUC \text{ sans IA} = 0,810 \text{ vs } 0,881 \text{ avec IA pour un } p < 0,0001$) ⁽⁶⁵⁾.

1.2.1.2 Imagerie thoracique

Le cancer du poumon est le cancer avec le plus haut taux de mortalité. De fait, de plus en plus de logiciels de conception assistée par ordinateur d'aide au diagnostic (CAD) sont présents sur le marché représentant 27% des logiciels sur le marché ⁽⁶⁴⁾ pour faciliter leur détection et leur suivi.

Sur une cohorte multicentrique réalisée entre avril et novembre 2021, l'étude d'*Ahn et al.* avait montré que *Lunit Insight* © améliore les performances des radiologues pour l'interprétation des radiographies du thorax (nodule, pneumothorax, pneumonie, épanchement pleural) avec une meilleure sensibilité pour chacun de ces paramètres comparativement à la seule interprétation du radiologue ⁽⁶⁶⁾.

1.2.1.3 Imagerie musculo-squelettique

L'étude de *Regnard et al.* avait montré dans une étude rétrospective en 2017 concernant 4774 radiographies qu'un logiciel d'IA (*BoneView* ©) était plus sensible

que le radiologue pour la détection de fracture , d'épanchement du coude, de dislocation ou des lésions osseuses focales mais moins spécifique ⁽⁶⁷⁾.

1.2.1.4 Imagerie neuro-radiologique

Aux Etats-Unis, 36% des logiciels d'IA mis sur le marché concernent la neuroradiologie et la plupart d'entre eux concerne la détection des AVC, des hémorragies et des anévrismes sur des tomodensitométries.

Sur un modèle de comparaison radiologue vs IA, l'étude de *Fasen et al.* concernant l'analyse tomodensitométrique de 474 patients (dont 75 avec une authentique occlusion proximale) avait montré que le logiciel *StrokeViewer* © avait une sensibilité similaire à celle du radiologue mais une spécificité nettement inférieure (88,5% pour l'IA contre 100% pour le radiologue) ⁽⁶⁸⁾

1.2.2 L'apport de l'IA hors champ de l'interprétation des examens

L'intelligence artificielle peut également aider à réduire la dose des examens, à créer des protocoles et améliorer la qualité de l'image ⁽⁶⁹⁾.

En somme l'IA peut apparaître de prime abord comme une menace pour le radiologue d'autant plus que les modèles d'étude s'intéressent davantage à « l'IA vs radiologue » plutôt qu'au modèle « IA vs IA + radiologue » où il est démontré que l'IA apporte une aide complémentaire au radiologue et au clinicien, et améliore ses performances dans le diagnostic des maladies ⁽⁷⁰⁾.

2. Définition de l'intelligence artificielle

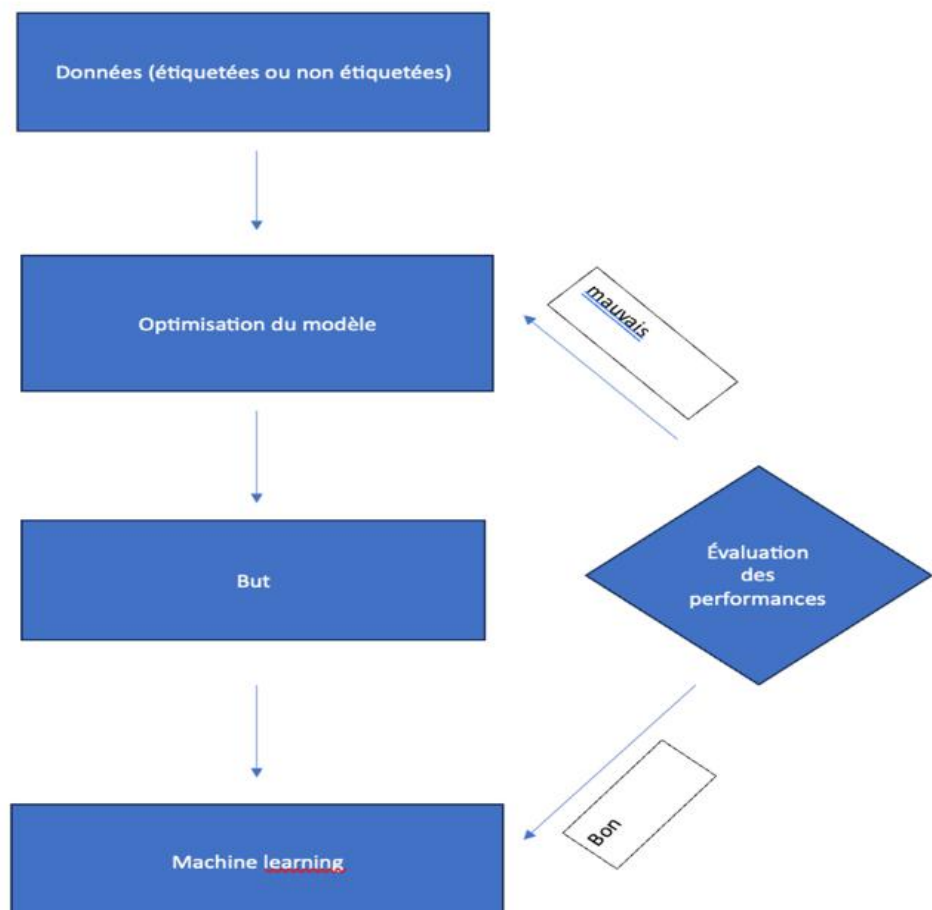
2.1 Intelligence artificielle

La notion d'intelligence artificielle a été introduite pour la première fois par Alan Turing, mathématicien et père de l'informatique théorique, en 1950 via son questionnaire "*est-ce que les ordinateurs peuvent penser?*" où il proposa un test "Imitation Game" rebaptisé plus tard « Test de Turing » consistant à évaluer la capacité d'une machine à se comporter comme un humain. Pour ce faire, un individu extérieur, après avoir posé des questions à une machine et un humain doit évaluer chacune de leurs réponses à l'aveugle et s'il est impossible de distinguer la réponse de la machine de celle de l'humain alors le test est considéré comme réussi. ⁽⁷¹⁾

De fil en aiguille, l'intelligence artificielle peut être définie comme la capacité d'une machine à imiter les fonctions cognitives d'un humain. ⁽⁷²⁾ Au sein de ce grand ensemble constituant l'intelligence artificielle, il y a l'apprentissage automatique ("machine learning"), terme introduit pour la première fois par *J.McCarthy et al.* en 1955, ⁽⁷²⁾ qui élabore des algorithmes et des modèles statistiques permettant d'apprendre des tâches complexes et avec l'expérience et l'entraînement de créer des modèles prédictifs performants via une grande base de données ⁽⁷⁹⁾. Il existe 5 sous-types de machine learning : l'apprentissage supervisé ; l'apprentissage non supervisé ; l'apprentissage semi-supervisé ; l'apprentissage par renforcement et l'apprentissage profond (« deep learning ») ⁽⁷³⁾.

Figure 21 : Le cycle d'une « machine learning » ⁽⁷⁴⁾

Adapté de Sorantin et al. ⁽⁷⁴⁾



2.2 Les différents sous-types de l'intelligence artificielle

2.2.1 L'apprentissage supervisé

L'apprentissage supervisé permet de créer un algorithme à travers plusieurs données labelisées ("ground-truth label") ayant pour but de prédire un modèle avec un résultat précis ⁽⁷⁶⁾. Il existe deux catégories d'algorithmes ⁽⁷⁵⁾ :

- **L'algorithme de régression** permet de prédire une quantité : comme les arbres décisionnels ; la forêt d'arbres décisionnels (combinant les résultats de plusieurs arbres décisionnels) ; la régression logistique (qui permet d'étudier la relation entre plusieurs variables prédictives et une variable binomiale. ⁽⁷²⁾
- **L'algorithme de classification** permet de prédire une classe, comme la machine à vecteurs de support qui classe les données en deux groupes en trouvant un hyperplan avec la marge la plus maximale séparant les 2 groupes ⁽⁷⁷⁾ ; la méthode des "k" plus proches voisins classant un objet non connu selon la proximité (distance minimale) avec d'autres objets connus ⁽⁷⁷⁾.

2.2.2 L'apprentissage non supervisé

Il permet de créer un algorithme à partir des données non labelisées (absence de "ground truth label") afin d'établir un modèle permettant de découvrir par lui-même des informations et des caractéristiques initialement cachées et les regrouper. Il peut être utilisé dans 3 domaines ⁽⁷⁵⁾ :

- **L'association** afin de trouver un lien d'intérêt entre plusieurs variables.
L'algorithme *A priori* repère des caractéristiques récurrentes dans une base de données afin de les associer. ⁽⁷⁸⁾
- **Le regroupement** consistant à créer des groupes homogènes basés sur leurs différences et similarités. Plusieurs algorithmes sont utilisés à cet effet comme le partitionnement de données diffus où chaque donnée peut appartenir à un ou plusieurs "cluster" sachant que le but de final étant que chaque "cluster" ainsi que les éléments qui le composent, possèdent des caractéristiques les plus semblables ⁽⁷⁹⁾ ; le regroupement hiérarchique où les données sont soit regroupées de façon ascendante (d'un groupe de plus en plus large et grand) ou de façon descendante (d'un grand groupe à des groupes plus petits et séparés) ⁽⁸⁰⁾ ; le champ aléatoire de Markov où les variables aléatoires suivent la propriété de Markov où la probabilité d'une variable dépend de son ou ses voisins immédiats. ⁽⁸¹⁾ ; le partitionnement en k moyenne a pour principe de

diviser un nombre n de données en k clusters où chaque donnée appartient au groupe qui lui est le plus proche ⁽⁸²⁾ ; et enfin le modèle de mélange Gaussien qui estime la probabilité qu'une donnée suivant une distribution Gaussienne appartienne à un cluster particulier. ⁽⁸³⁾

- **La réduction de dimension** : visant à réduire le nombre de données si celui-ci est trop haut, tout en conservant leur intégrité autant que possible. Cette réduction de dimension de données se fait avant le traitement de données et avant la conception du modèle ⁽⁷⁵⁾. Il y a par exemple l'analyse en composante principale ou la décomposition en valeurs singulières qui pour ce dernier permet de décomposer une matrice en plusieurs autres afin de réduire / compresser le nombre d'informations tout en préservant au mieux la matrice originale ⁽⁸⁹⁾.

2.2.3 L'apprentissage semi-supervisé

C'est un croisement des deux apprentissages cités ci-dessus où l'algorithme utilise un petit nombre de données labélisées et un grand nombre de données non labélisées, le tout pouvant améliorer la qualité d'apprentissage et obtenir un gain de temps sachant que l'étiquetage des données fait intervenir un tiers ce qui peut être chronophage d'autant plus si ce volume des données est important ⁽⁸⁵⁾.

2.2.4 L'apprentissage par renforcement

L'algorithme apprend de ses erreurs à travers des essais itératifs afin d'obtenir le modèle le plus efficace en l'absence de données labélisées ⁽⁸⁶⁾.

2.2.5 L'apprentissage profond

Il a été introduit en 2006 par *Hinton et al.* ⁽⁸⁷⁾ et permet à un modèle de prendre des décisions grâce à l'apprentissage d'une grande base de données via l'aide d'un réseau de neurones artificiels qui simule celui de l'humain ⁽⁷⁵⁾. Ce dernier est constitué d'une couche d'entrée, une ou plusieurs couches masquées et une couche de sortie.

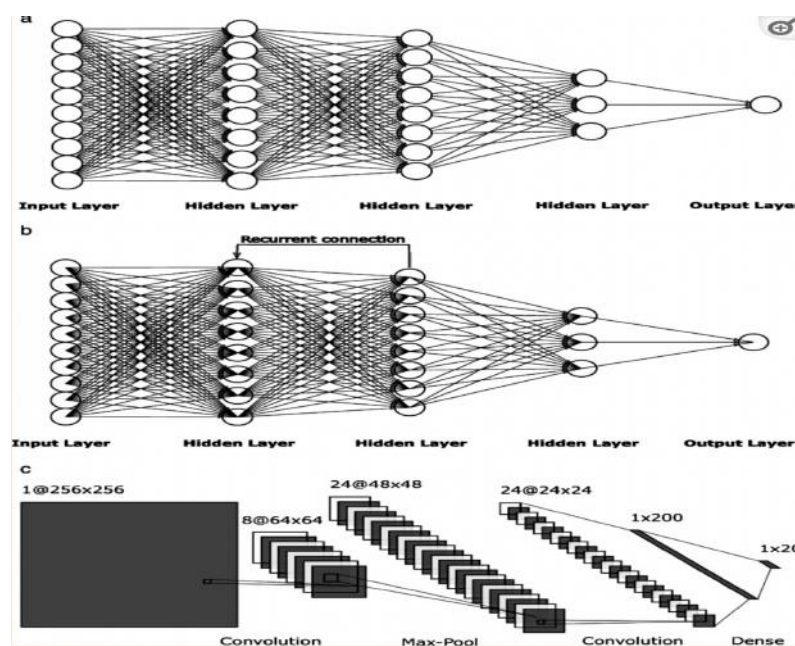
Il existe plusieurs catégories de réseau de neurones artificiels ⁽⁷⁴⁾ :

- **Un réseau de neurone à action directe** où l'information transite dans une direction à sens unique sans mémoire interne de la couche précédente (couche d'entrée → une ou plusieurs couches masquées → couche de sortie)

⁽⁷⁴⁾

- **Un réseau de neurone convolutif (CNN)** permettant la reconnaissance et la classification d'une image grâce à l'agencement des neurones similaires à celui du cortex visuel animal ⁽⁸⁸⁾ (la direction est multidimensionnelle). Il est constitué de 3 sous-couches ⁽⁸⁹⁾ : une couche de convolution, une couche de sous-échantillonnage (ou *pooling*) et une couche entièrement connectée. (cf 2.2.5.1)
- **Un réseau de neurone récurrent** où l'information peut circuler dans deux directions par l'intermédiaire de boucles, c'est-à-dire qu'une information issue d'une couche de neurone précédente peut être utilisée par la suite à la fois pour la couche de neurone en cours et suivante (mémoire interne). Il permet ainsi de prédire des données séquentielles (phrases, chiffres, ...) ⁽⁷⁴⁾.

Figure 22 : Réseaux de neurones dans le « deep learning » ⁽⁷⁴⁾



A : Réseau de neurone à action directe // B : Réseau de neurone récurrent // C : Réseau de neurone convolutif

2.2.5.1 Le réseau de neurone convolutif

Le réseau de neurones convolutifs connaît un essor dans le domaine médical et notamment dans le milieu de la radiologie permettant de classer, détecter, reconstruire et segmenter des images. L'objectif de ce paragraphe est d'en détailler le principe.

Le CNN est constitué de 3 couches :

1) Couche de convolution

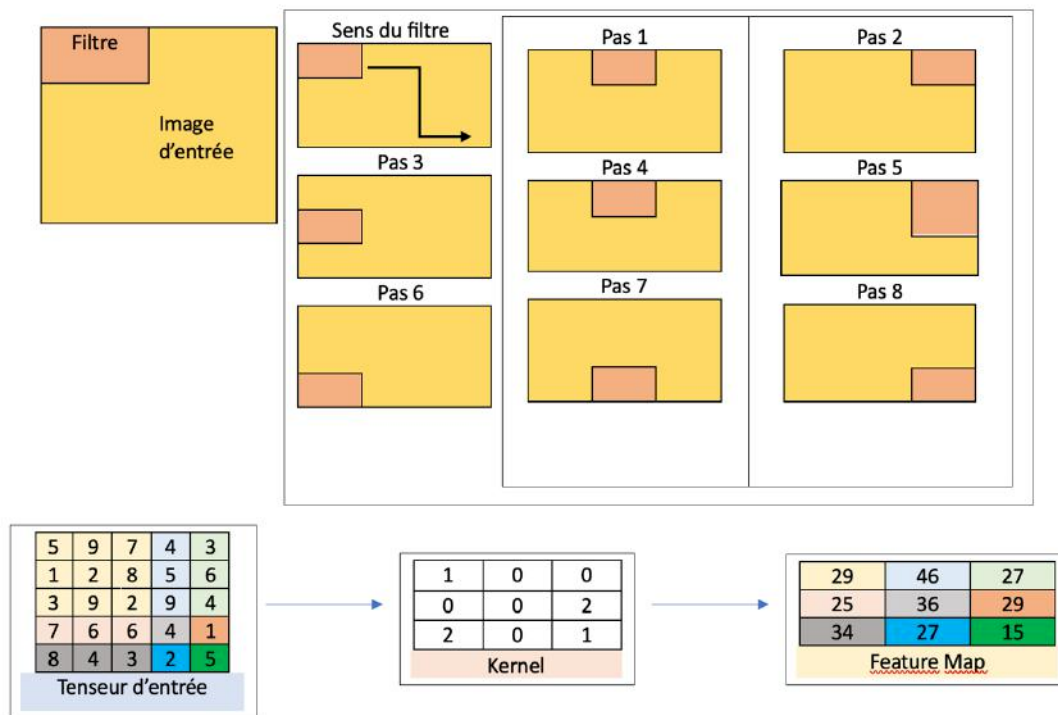
Elle constitue la première couche du CNN où un filtre (ou kernel) d'une certaine dimension (taille) et d'un certain poids (chaque valeur constituant le kernel) se déplace progressivement sur chaque portion de l'image d'entrée (selon un certain pas de convolution : *par exemple se déplace pour chaque pixel, tous les 2 pixels, tous les 3 pixels,...*) afin de produire une nouvelle matrice (équivalent au produit scalaire de la valeur de chaque poids du kernel et la valeur de chaque pixel de l'image) nommée **carte d'activation** (ou *feature map*) contenant des caractéristiques de l'image détectées spécifiquement par un kernel.

Cette opération est répétée pour couvrir l'ensemble de l'image d'entrée avec un nombre de filtre, une taille de filtre, un pas de convolution et un *padding* (*ajout des pixels à chaque bord de l'image pour éviter la réduction de taille après chaque convolution*) définis au préalable avant la conception du modèle (ce sont des hyperparamètres).

Pour chaque couche convolutive, le nombre de filtre est fixé a priori (hyperparamètre) de façon proportionnelle à la complexité de l'image, ce nombre de filtre équivaut au nombre de feature map produit (*par exemple s'il y a 10 filtres, il y aura 10 feature map*). Chaque filtre étant différent, possédant un poids différent appris pendant l'entraînement pour détecter une caractéristique spécifique de l'image (*par exemple un filtre n°1 détectant le bord d'une image ; un filtre n°2 détectant la forme d'une image,...*) et génère une feature map différente. Toutes ces dernières créées seront compilées en un tenseur de sortie à plusieurs dimensions qui deviendra à son tour un nouveau tenseur d'entrée pour les couches suivantes où un nombre plus important de filtres différents y seront appliqués, capturant des caractéristiques de plus en plus complexes, les couches devenant plus profondes ⁽⁹⁰⁾.

Figure 23 : Couche de convolution dans le réseau de neurones convolutifs ⁽⁸⁹⁾

La première image est adaptée de Datascientest. Convolutional Neural Networks. Datascientest.



Le filtre (ici 3x3) se déplace dans un certain sens et selon un pas de convolution défini (un pas de 1 pixel par pixel dans cet exemple) pour parcourir toute l'image d'entrée. Une feature map (ici 3x3) est produite selon le filtre donné pour chaque portion du tenseur d'entrée (5x5).

2) Fonction d'activation non linéaire

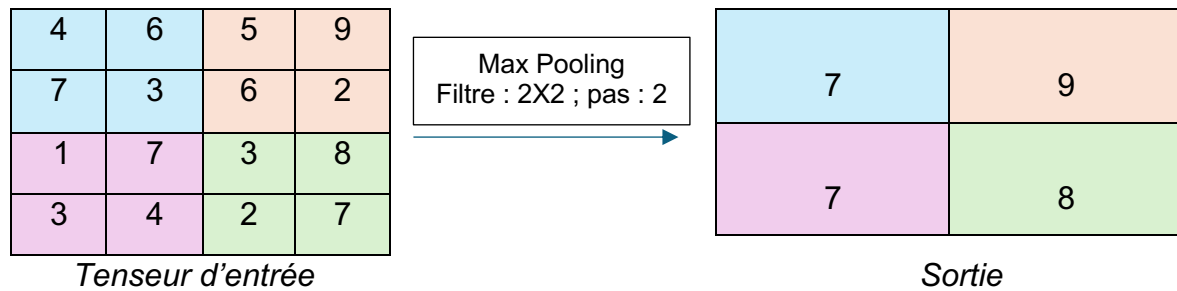
Après chaque couche de convolution, une fonction d'activation non linéaire est appliquée comme la fonction ReLU (Rectified Linear Unit) dans les CNN consistant à remplacer les valeurs négatives par des 0 afin de rendre le modèle non linéaire et plus complexe ⁽⁸⁹⁾.

3) Couche de sous-échantillonnage

C'est une méthode sous-échantillonnage (ou pooling) permettant de **réduire la taille de l'image** tout en conservant les données les plus importantes (via la technique du *max pooling* où la valeur maximale de chaque région est retenue). Il n'y a pas de paramètres appris sur cette couche, en revanche il y a 3 hyperparamètres qui doivent être fixés avant le début de l'algorithme à savoir la taille du filtre, le pas de convolution (le nombre de pixels pour lequel le filtre se déplace à chaque étape) et le

« padding ». Les hyperparamètres les plus souvent utilisés pour cette couche sont une taille de filtre 2x2, avec un pas de 2, sans padding ⁽⁸⁹⁾.

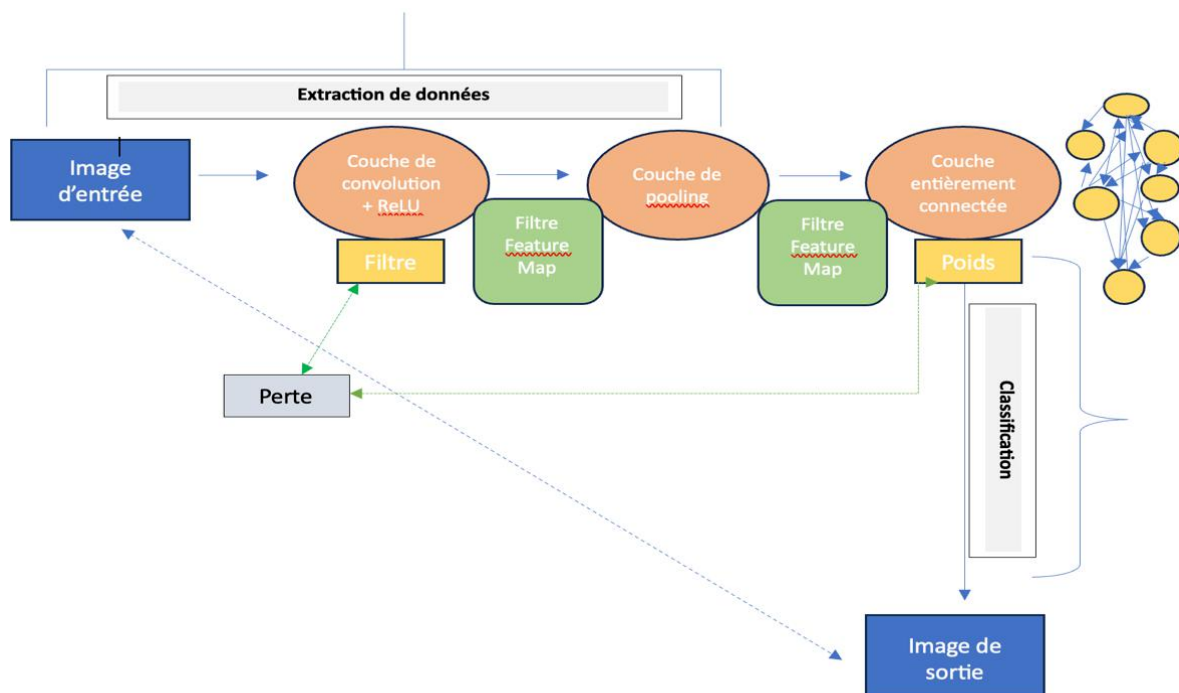
Figure 24 : Illustration du « max pooling »



4) Couche entièrement connectée

Les caractéristiques extraites des couches précédentes sont aplaties et transformées en vecteurs (en une seule dimension) puis prises en charge par une ou plusieurs couches de neurones connectées qui produisent un vecteur de sortie établissant une probabilité de l'image d'appartenir à une classe ⁽⁸⁹⁾.

Figure 25 : Résumé du CNN



3. Créer et évaluer un modèle d'IA

3.1 Données

Afin de générer un modèle robuste d'IA, il faut à la fois une base de données quantitative et qualitative (par exemple peu d'artefacts) ce qui peut certes être chronophage et engendrer un coût important, mais permet au modèle d'IA de filtrer des caractéristiques pertinentes et d'être représentatif afin d'être appliqué dans n'importe quelle population.

Pour les apprentissages profond et supervisé, les données sont labélisées par un collectif d'experts : plus il y a d'experts, mieux les données sont précises ⁽⁹⁰⁾.

3.2 Entraînement

Ces données labélisées sont séparées en 3 échantillons ⁽⁹¹⁾ :

- **Un échantillon d'entraînement** : permettant d'entraîner le modèle. Dans le CNN l'objectif est d'optimiser et actualiser les paramètres que sont le filtre (kernel) et le poids dans les couches de neurones correspondant à savoir respectivement la couche de convolution et la couche entièrement connectée, permettant de minimiser au maximum l'écart entre les données d'entrées labélisées et la prédiction du modèle. Ceci est permis d'une part grâce à la fonction de perte, fonctionnant sur un mode de « propagation avant » avec la mesure de la différence entre les données brutes et les données de sortie du modèle et d'autre part par la descente de gradient : un algorithme d'optimisation fonctionnant sur un mode de rétropropagation où les erreurs (écart) sont utilisées pour ajuster les paramètres de la couche de neurone suivante en mettant à jour le filtre (kernel) et le poids.
- **Un échantillon de validation** : permet d'évaluer le modèle d'entraînement, d'optimiser à posteriori les hyperparamètres (pour rappel c'est un paramètre fixé défini manuellement avant de commencer l'entraînement du modèle par exemple le nombre de couches, le nombre de neurones ou le taux d'apprentissage...) qui peuvent être ajustés par plusieurs techniques :
 - La recherche en grille consistant à essayer toutes les combinaisons possibles des hyperparamètres en utilisant un modèle pour chaque combinaison et en sélectionner le meilleur.

- La recherche aléatoire via la sélection de façon aléatoire de différentes combinaisons d'hyperparamètres et d'entraîner un modèle pour chaque combinaison.
 - L'optimisation Bayésienne via une mise à jour d'une distribution de probabilité sur les hyperparamètres en fonction des résultats obtenus lors de l'entraînement.
 - L'essai itératif visant à tester les différentes valeurs d'hyperparamètres de manière itérative et ajuster selon les résultats obtenus et à la fin de sélectionner le modèle le plus performant.
- **Un échantillon test** : permet d'évaluer à la toute fin le modèle sélectionné une fois l'entraînement terminé (cf 3.4).

3.3 Inconvénients possibles

Un modèle imparfait peut être surentraîné ou sous-entraîné

3.3.1 Surentraînement

Le surentraînement (ou *overfitting*) correspond à un modèle qui excelle dans l'échantillon d'entraînement lié à des données apprises par cœur mais qui est mauvais dans l'échantillon test comprenant des données inconnues (qui n'ont pas pu être apprises). De fait, cela engendre un modèle de prédiction qui n'est pas représentatif et généralisable ⁽⁸⁹⁾ ⁽⁹⁰⁾.

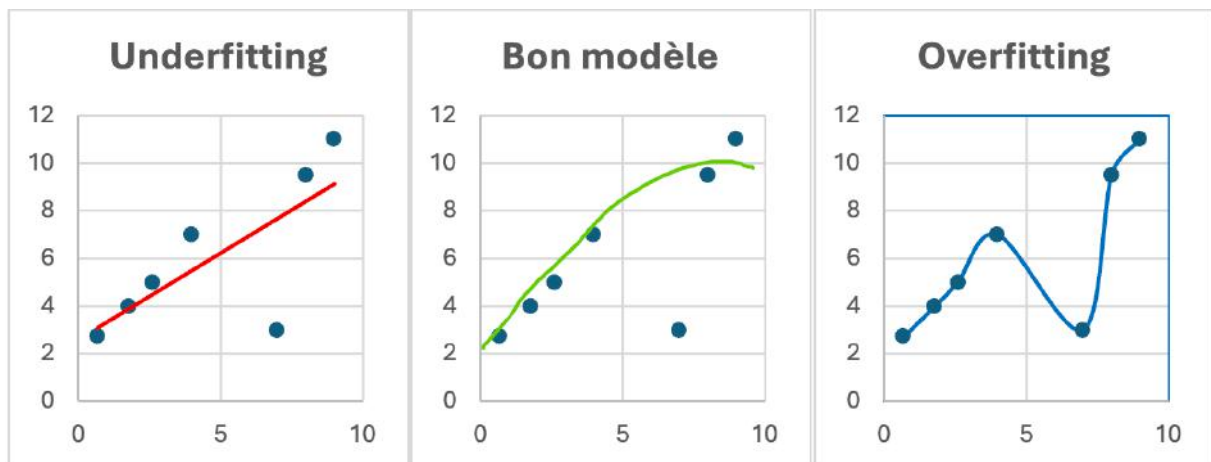
Pour réduire l'overfitting, plusieurs solutions peuvent exister comme :

- Augmenter le nombre de données d'entraînement (même si ceci est plus difficile dans le domaine médical au vu du coût important) afin d'être plus généralisable pour des données inédites.
- Le « dropout » où un certain pourcentage de neurone est ignoré afin d'éviter un surapprentissage de neurones spécifiques.
- La réduction de la complexité architecturale
- La normalisation par lot permettant d'ajuster les activations des neurones dans un réseau pour stabiliser et accélérer l'entraînement
- La régularisation L2 (ou « weight decay ») est une technique pénalisant les poids du modèle en favorisant les poids petits.

3.3.2 Sous-entraînement

Le sous-entraînement (ou *underfitting*), à l'inverse, est un modèle non performant à la fois dans l'échantillon d'entraînement et dans l'échantillon test. Ce problème peut être dû à une mauvaise qualité des données d'entraînement et/ou à une quantité insuffisante de données d'entraînement et/ou à un modèle peu complexe ne pouvant pas capturer les relations présentes dans les données ⁽⁸⁹⁾.

Figure 26 : Illustration du surentraînement et sous-entraînement



3.4 Évaluer le modèle

Voici quelques exemples de paramètres couramment utilisés (non exhaustifs) permettant d'évaluer un modèle d'IA :

3.4.1 Aire sous la courbe

La technique AUC ROC (aire sous la courbe – *Receiver operating characteristic*) permet d'évaluer graphiquement la performance d'un modèle (*par exemple : détection ou non d'une activité inflammatoire radiologique à travers un algorithme de segmentation*). Plus l'AUC de la courbe ROC se rapproche de 1 (considéré comme un modèle parfait), meilleur est le modèle, tout en s'attachant à maximiser les valeurs de la sensibilité et de la spécificité ⁽⁹⁰⁾.

3.4.2 Le coefficient DICE

Le coefficient DICE est un coefficient de **similarité** permettant d'évaluer la performance d'un algorithme de segmentation d'une image (nombre de voxel classé

et appartenant à juste titre ou non à la pathologie étudiée) en comparaison avec une segmentation étiquetée comme Gold Standard le plus souvent réalisé par un collectif d'expert de façon manuelle. Un coefficient à 0 équivaut à une absence totale de similarité entre les 2 modèles (les 2 segmentations ne se recouvrent en aucun point) et un coefficient à 1 montre une délimitation identique entre les 2 méthodes ⁽⁸⁹⁾. Sa formule est la suivante : $DICE = (2 \times VP) / (2 \times (VP + FP + FN))$: VP (vrai positif) représentant le nombre de voxel correctement classé et délimité comme une lésion de SEP par exemple ; FP (faux positif) représentant le nombre de voxel délimité et classé à tort comme une lésion de SEP par exemple ; FN (faux négatif) représentant le nombre de voxel n'étant pas délimité à tort comme une lésion de SEP ⁽⁹⁰⁾.

3.4.3 Kappa de Cohen

Le Kappa de Cohen mesure la **concordance globale** entre deux méthodes de classification en prenant en compte la probabilité qu'un accord soit dû au hasard notamment en cas de classes déséquilibrées.

Par exemple : le kappa de Cohen peut être utilisé afin de mesurer la concordance pour la détection d'une activité inflammatoire radiologique de la SEP par un logiciel d'IA en comparaison avec un radiologue expert, où 2 classes sont considérées : la présence d'une activité inflammatoire ou son absence.

Il se mesure ainsi : $\kappa = (P_0 - P_e) / (1 - P_e)$ où P_0 représente la proportion d'accord observé entre les 2 méthodes et P_e représente la proportion d'accord attendu par le hasard.

Un kappa proche de 1 signe un accord parfait entre les 2 méthodes, en revanche un kappa se rapprochant de 0 signifie que l'accord est proche du hasard ⁽⁹²⁾.

3.4.4 F1 Score

Le F1 score établit la **performance** de prédiction d'un modèle en se focalisant principalement sur une seule classe (classe positive) et sans tenir compte de la proportion de l'accord obtenu par le hasard contrairement au kappa de Cohen.

Les paramètres utilisés sont le rappel (qui peut être un équivalent de la sensibilité : proportion de cas positifs détectés correctement sur tous les cas positifs réels mesurés par le Gold Standard) et la précision (qui peut être un équivalent de la

valeur prédictive positive (VPP) : proportion des cas positifs corrects détectés par le logiciel) où une moyenne harmonique des deux est réalisée.

Il se mesure ainsi : $F1\ score = 2 \times ((précision \times rappel) + (précision + rappel))$

Par exemple : le F1 score peut-être utilisé pour mesurer la performance d'un modèle d'IA détectant une activité inflammatoire radiologique positive de SEP (en comparaison avec le Gold Standard).

Un F1 score proche de 1 signifie que le modèle détecte parfaitement toute activité inflammatoire radiologique positive de la SEP, à contrario un F1 score proche du 0 dénote un taux important de faux positifs ou faux négatifs, le modèle de détection d'une activité inflammatoire radiologique positive de SEP étant imparfait et sous-optimal ⁽⁹²⁾.

3.3.5 Coefficient de Matthews

Il est le reflet de l'évaluation de la **performance globale** d'un modèle en jugeant la qualité des classifications binaires, efficace surtout quand les classes sont déséquilibrées, en prenant compte toutes les classes donc les vrais positifs, les vrais négatifs, les faux positifs et les faux négatifs (contrairement au F1 score).

Il se mesure ainsi : $MCC = \frac{(VP+FP)-(VP+FN)}{\sqrt{(VN+FP)(VN+FN)(VP \times VN)-(FP \times FN)}}$

Un coefficient proche de 1 est une classification parfaite, proche de 0 témoigne d'une classification aléatoire et proche de -1 est une classification totalement inversée.

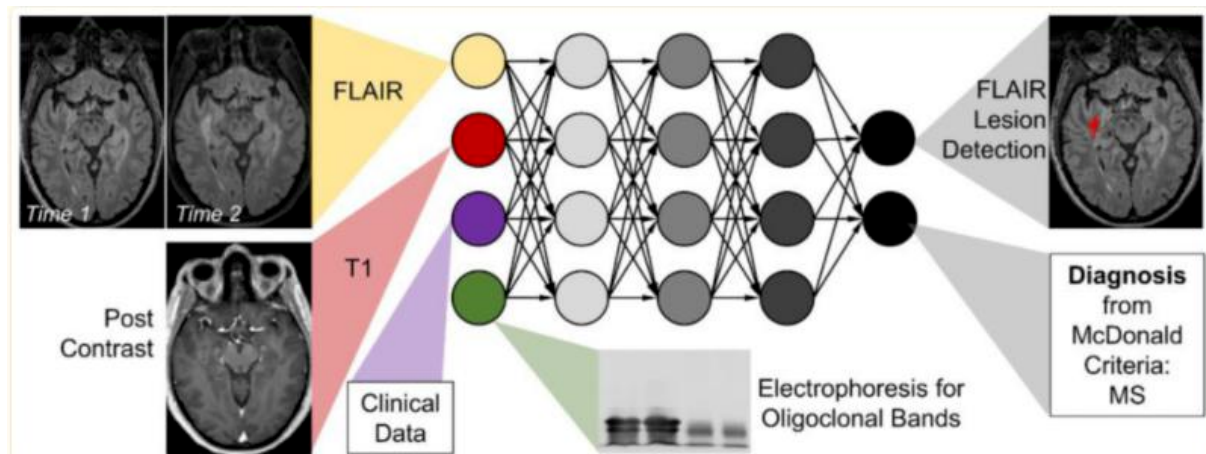
Le coefficient de Matthews évaluera par exemple si le modèle d'IA classe correctement les patients présentant une activité inflammatoire radiologique ⁽⁹²⁾.

4. Exemple de l'IA appliquée dans la sclérose en plaques

La SEP est une pathologie où toutes les facettes de l'intelligence artificielle mentionnées ci-dessus peuvent être appliquées dans le diagnostic, le pronostic et le traitement de la maladie. L'IRM comme vu précédemment est une pierre angulaire dans la pathologie que ce soit au niveau du diagnostic (*cf Critères révisés de McDonald*) ou du pronostic (*cf A.8*). Cependant la détection et la quantification des lésions de SEP peuvent s'avérer être compliquées à cause de la localisation et la taille variables des lésions, l'aspect chronophage pour l'interprétation manuelle (considérée comme le Gold Standard) qui est soumise à une variabilité inter et intra-

observateur. L'intelligence artificielle pourrait apporter une aide complémentaire pour résoudre ces problèmes. Par ailleurs, une méta-analyse réalisée par *Nabizadeh et al.* avait montré une précision de 93% pour le diagnostic de la SEP par l'IA (technique d'apprentissage profond) ⁽⁹³⁾.

Figure 27 : Les facettes et multimodalités de l'IA dans le diagnostic de la SEP ⁽⁴⁹⁾



4.1 Diagnostic

4.1.1 IRM

4.1.1.1 Segmentation

La simple segmentation des lésions de SEP peut être assimilée à une étude transversale basée sur une seule imagerie, impliquée dans le critère de dissémination spatiale. Les apprentissages supervisés et non supervisés étaient utilisés autrefois pour la segmentation des lésions de SEP avant l'essor de l'apprentissage profond dont notamment le CNN, permis grâce à de grandes bases de données se basant une segmentation manuelle des lésions de SEP par des experts (comme l'Open MS data dataset ou le MSSEG2-challenge MICCAI 2021). ⁽⁹⁵⁾

La technique du CNN peut utiliser 2 catégories de segmentation ⁽⁹⁶⁾ :

- La segmentation par patch : le modèle analyse des « patches » de pixels du tenseur d'entrée indépendamment des uns des autres, les classe (*par exemple : lésion SEP ou non SEP*) puis les assemble pour former une segmentation complète de l'image.

- La segmentation sémantique : proposée par *Brosch et al.* en 2015 où le modèle analyse le tenseur d'entrée dans sa globalité puis y classe chaque pixel (*lésion SEP ou non SEP*) de cette image.

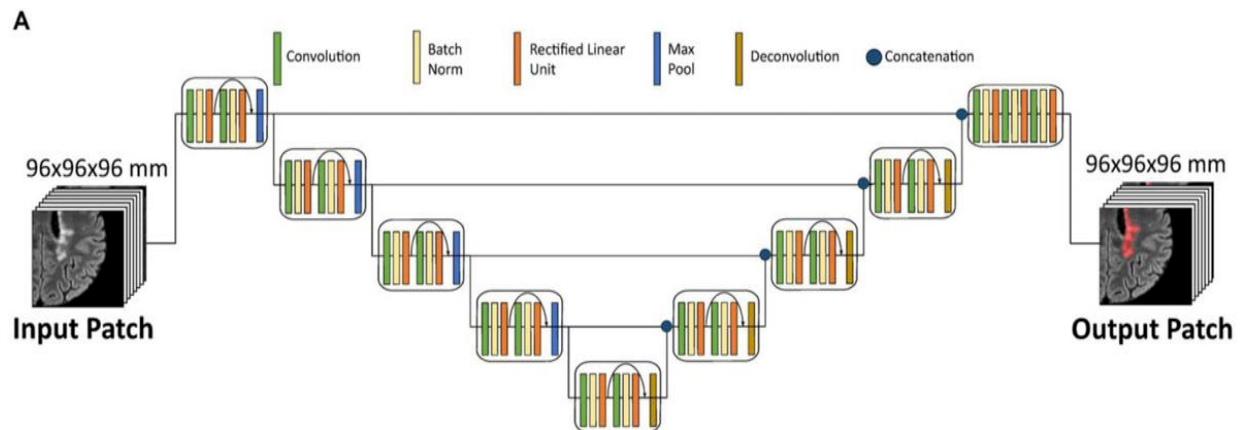
Il existe plusieurs sous-types d'architectures du modèle du réseau de neurone convolutif comme AlexNet, VGG, ResNet, GoogleNet ⁽⁹⁷⁾ mais c'est le U-net qui est la technique la plus utilisée dans le cadre de la SEP et qui a été proposé pour la première fois par *Ronneberger et al.* en 2015 ⁽⁹⁸⁾ .

C'est une technique permettant de segmenter précisément et rapidement les lésions (et est également utilisée pour la segmentation de lésions d'autres pathologies) de SEP sur des séquences FLAIR, T1 avant et/ou après injection et T2.) où chaque pixel sera classé appartenant à une classe spécifique. Il a une forme en U et est composé :

- **D'un encodeur** (*comme un CNN classique*) : permettant d'analyser et d'extraire des caractéristiques importantes de l'image d'entrée tout en réduisant la taille de l'image. Pour ce faire, il existe plusieurs étapes que sont la convolution qui applique des filtres sur l'image permettant de capturer des caractéristiques globales de l'image pour les premières couches de neurones puis de plus en plus fines au fur et à mesure des couches plus profondes ; une fonction de non-linéarité (ReLU) afin d'apprendre des relations plus complexes ; un (max) pooling afin de réduire la taille de l'image
- **D'un décodeur** (*non présent dans un CNN classique*) : permettant de sélectionner les caractéristiques extraites lors de l'encodage et en augmenter la résolution ⁽⁹⁹⁾, via des couches de déconvolution et des couches de connexions de saut afin de conserver les caractéristiques pertinentes et détails importants des couches précédentes qui pourraient être potentiellement perdus dans le processus de réduction d'image. L'image de sortie présente donc les mêmes dimensions que l'image d'entrée sous forme d'une carte de segmentation.

Il existe plusieurs variants du U-net pouvant être en 2D ou en 3D. ⁽¹⁰⁰⁾

Figure 28 : Le fonctionnement du U-net ⁽¹⁰¹⁾



Le fonctionnement d'U-net avec la branche gauche représentant la phase d'encodage et la branche droite la phase de décodage

Tableau 4 : Différences entre CNN classique et U-net

	CNN	U-Net
Architecture	Séquentielle	En forme de U
Décodeur	Non	Oui
Sortie	Vecteurs de classes	Carte de segmentation
Connexions de saut	Non	Oui
Objectif	Classification d'une image	Segmentation

Une technique basée sur le 3D U NET associée à une séquence T2 FLAIR et T1 MP2RAGE (*Magnetization Prepared 2 Rapid Acquisition Gradient Echoes*) serait efficace avec un taux de détection à 76% des petites lésions corticales et un taux de faux positifs à 29% comparativement au gold standard ⁽¹⁰⁰⁾.

D'autres modèles basés sur le neurone convolutif permettent de délimiter les lésions médullaires comme celui établi par Gros *et al.* en 2019 avec un coefficient de Dice de 95% ⁽¹⁰²⁾.

4.1.1.1 Paramètres influençant la qualité de la segmentation

Une étude de Commowick *et al.* analysait de façon rétrospective 53 lésions de SEP à travers plusieurs algorithmes (apprentissage supervisé, non supervisé et apprentissage profond) comparativement à une délimitation manuelle considérée comme gold standard et avait montré que plus la lésion et le volume des lésions sont

petits, plus la segmentation automatisée est moins précise de sorte qu'il est plus difficile de délimiter un nombre élevé de petites lésions comparativement à un petit nombre de grandes lésions.

D'autre part, la même étude avait prouvé qu'une fusion de plusieurs algorithmes permettant la segmentation des lésions offre de meilleurs résultats qu'une utilisation unique d'un algorithme ⁽¹⁰³⁾.

4.1.1.2 Évolution des lésions

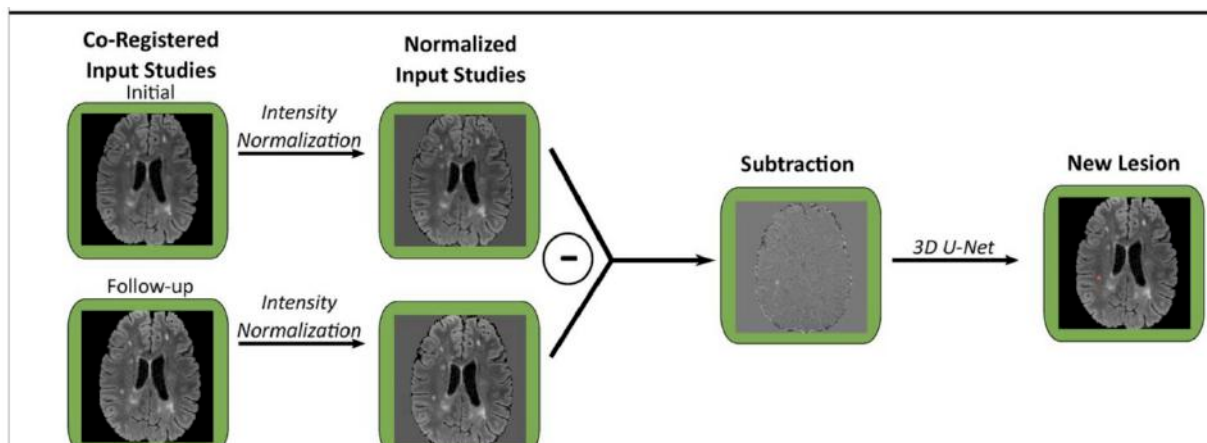
L'évolution des lésions de SEP est un point essentiel impliqué dans le critère de dissémination temporelle et dans l'évaluation de l'efficacité d'une thérapeutique. Contrairement à une simple segmentation s'assimilant plutôt à une analyse transversale, le suivi d'une lésion de SEP peut s'apparenter à une étude longitudinale en comparant 2 imageries pour un même patient où il peut être mis en évidence l'apparition d'une nouvelle lésion, d'une lésion croissante pré-existante, d'une lésion décroissante ou la disparition d'une lésion.

La technique de soustraction usuelle (qui n'est pas une méthode d'apprentissage automatisé) entre 2 IRM permet de mettre en évidence des nouvelles lésions en effectuant la différence des voxels de même topographie pour chaque IRM qui vont s'annuler, démasquant uniquement les nouvelles lésions. C'est ce genre d'outil qui est utilisé au quotidien au CHU de Lille pour comparer annuellement la charge lésionnelle T2 des patients (*outil LoBI de IntellispacePortal*). L'étude de *Moraal et al.* avait montré que cette méthode augmente de 1,7 fois la détection des nouvelles lésions comparativement à l'analyse séparée des images natives. Toutefois cette technique est sensible aux artefacts et aux variations des paramètres techniques d'IRM nécessitant donc un calibrage parfait. ⁽¹⁰⁴⁾ Elle peut être utilisée avec ou sans les techniques de deep learning.

Quelques algorithmes de deep learning commencent peu à peu à émerger permettant de détecter et segmenter les nouvelles lésions comme par exemple celui basé sur le "nn-Unet" (le no new- unet) d'*Isensee et al.* ⁽¹⁰⁵⁾ qui pré-traite les données et ajuste les hyperparamètres tout en conservant l'architecture du U-net afin d'obtenir des résultats davantage satisfaisants. L'inconvénient de ces modèles pour ce suivi longitudinal est le faible nombre de données labélisées qui peut toutefois être compensé par des techniques comme l'apprentissage par transfert (ou « transfert learning ») où des données déjà entraînées d'un ancien modèle sont ré-utilisées pour entraîner le nouveau modèle en question (par exemple les données

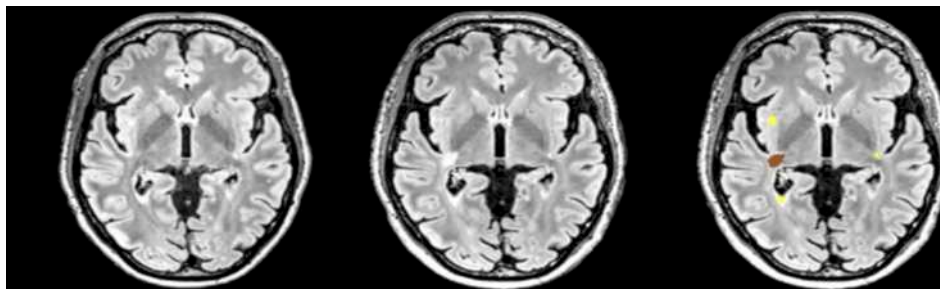
d'un modèle impliqué dans la segmentation des lésions de SEP seront ré-utilisées pour un modèle visant à établir la présence ou la croissance d'une lésion de SEP étant donné que la détection de leur évolution nécessite que le modèle soit déjà capable de base de segmenter/détecter une lésion de SEP) ou encore l'augmentation des données transformées de façon artificielle (*par exemple augmentation du bruit d'une image d'entrée, rotation d'une image, ...*) ⁽¹⁰⁶⁾

Figure 29 : La technique du 3D U-net et segmentation d'une nouvelle lésion ⁽¹⁰⁵⁾



La technique de soustraction s'associant au 3D U-net mettant en évidence les nouvelles lésions

Figure 30 : Évolution des lésions de SEP pour un patient ⁽⁹⁶⁾



En marron : nouvelle lésion // En jaune : lésion stable (basale)

4.1.1.3 Diagnostics différentiels

L'apprentissage supervisé, non supervisé et l'apprentissage profond peuvent permettre de différencier la SEP d'autres pathologies inflammatoires et non inflammatoires ainsi que des sujets « contrôle »

A titre d'exemple (non exhaustif) : l'étude de *Eshagi et al.* en 2016 ⁽¹⁰⁷⁾ avait montré un coefficient de Dice de 80% pour la classification de la NMOSD vs SEP ou encore

l'étude de *Kim et al.* en 2020, en utilisant la technique du neurone convolutif pour différencier une NMOSD et la SEP avait montré un coefficient de Dice de 71% ⁽¹⁰⁸⁾.

4.1.2 Autres biomarqueurs

Comme vu précédemment, d'autres paramètres non inclus dans les critères de McDonald peuvent renforcer le diagnostic de la SEP et sont applicables à l'intelligence artificielle dans leur capacité de détection.

4.1.2.1 Imagerie

Le signe de la veine centrale (*cf A.7.2.7*) est un signe très sensible de la SEP visible en T2* ou SWI.

L'étude de *Maggi et al.* en 2020 avait montré via une étude rétrospective la performance de la méthode du neurone convolutif qui se rapproche des performances humaines (Gold Standard) dans la détection de ce signe avec une précision à 91% et une rapidité plus importante que l'humain (4 secondes versus 40 minutes dans un échantillon test de 33 éléments) ⁽¹⁰⁹⁾.

Cavalière et al. avait prouvé la possibilité de classer les sujets SEP et sains sans symptômes visuels à travers les paramètres de l'OCT (tomographie à cohérence optique), et notamment l'épaississement de la couche maculaire ganglionnaire, en utilisant un algorithme de classification via la méthode de la machine à vecteurs avec un coefficient de Matthew à 0,81, une sensibilité à 89% et une spécificité à 91% et une AUC à 0,97. ⁽¹¹⁰⁾

4.1.2.2 Sérique

Goyal et al. avait appliqué 4 algorithmes (issus de l'apprentissage supervisé, non supervisé et profond) pour détecter la SEP à partir de la concentration de 8 cytokines : 1 β , IL-2, IL-4, IL-8, IL-10, IL-13, IFN γ et TNF β chez 956 sujets atteints de SEP et 199 sujets sains. La forêt d'arbre décisionnel était la technique la plus précise (90,9% avec une sensibilité de 76% et une spécificité de 86% et une AUC de 0,96) ⁽¹¹¹⁾.

4.2 Pronostic

L'apprentissage automatisé permet de prédire l'évolution de la maladie via différentes techniques.

4.2.1 CIS

Bendfeldt et al. en 2019 en utilisant la technique de la machine à vecteur de support basée sur la clinique, des caractères démographiques et IRM (lésions hyperintenses en T2 et le volume de la substance grise) avait montré un taux de prédiction à 70% de conversion à une SEP. ⁽¹¹²⁾

4.2.2 Formes rémittentes-récurrentes

Pinto et al. en 2020 via la technique de la machine à vecteur de support avait montré un taux de prédiction vers une forme secondairement progressive après 2 ans estimée à 86%. ⁽¹¹³⁾. Une autre étude portée sur une large cohorte utilisant la technique du neurone récurrent réalisée par *Seccia et al.* en 2020 avait prédit une conversion à 2 ans vers une forme secondairement progressive avec une sensibilité à 67% et une VPP à 42% ⁽¹¹⁴⁾.

Enfin, une autre étude réalisée par *Roca et al.* en 2020 se basant sur une technique hybride entre l'apprentissage automatisé et l'apprentissage profond (basé sur la méthode du neurone convolutif) et utilisant des paramètres comme l'âge et l'IRM, avait montré une marge d'erreur à 1,7 du score EDSS à 2 ans ⁽¹¹⁵⁾.

4.3 Traitement

La réponse au traitement de la SEP est multifactorielle. De façon conventionnelle, elle est établie par la clinique et par imagerie sur l'évolution du nombre et la taille des lésions.

Cependant, il existe d'autres paramètres non usuellement pris en compte comme l'atrophie corticale.

Pour les formes rémittentes / récurrentes : selon une étude réalisée par *Kanber et al.* en 2019, l'apport de l'apprentissage automatisé sur la prédiction de la réponse au traitement par Natalizumab combinant plusieurs paramètres (l'étude de 144 zones régionales du cerveau et la trophicité cérébrale avant et après traitement) était supérieur aux techniques conventionnelles (AUC = 0,890 VS 0,686 $p < 0,01$) ⁽¹¹⁶⁾.

C. Objectifs de la thèse

La comparaison manuelle des lésions de SEP peut s'avérer être fastidieuse (surtout en cas de charge lésionnelle importante), chronophage et soumise à des variations intra et inter-observateur. Compte-tenu de l'impact de la découverte d'une activité inflammatoire radiologique sur le suivi et le traitement d'un patient atteint de SEP (poursuite/switch thérapeutique), il est primordial que cette comparaison d'IRM puisse être réalisée de façon précise et reproductible par le radiologue. Il est également nécessaire que ce travail comparatif puisse être réalisé en un temps acceptable pendant ou au décours immédiat de l'examen.

Actuellement, beaucoup d'algorithmes utilisant le deep learning sont disponibles permettant de segmenter des lésions de SEP à un temps donné mais peu sont aptes à suivre l'évolution des lésions de sclérose en plaques.

Pixyl.Neuro.MS © est un logiciel d'IA créé en 2015 par une entreprise Grenobloise (*Doyle – Forbes – Dojat*), et ayant obtenu le marquage CE en 2019. Cet outil est disponible au CHU de Lille en démonstration depuis Mars 2024. Le logiciel utilise la technique du CNN permettant en quelques minutes de segmenter des lésions de SEP en établissant un volume des lésions et également d'effectuer une analyse longitudinale permettant d'objectiver l'évolution des lésions (nouvelle ou croissante) à partir de 2 imageries de temporalité différente sur des séquences 3D FLAIR. ⁽¹¹⁷⁾

L'objectif premier de cette étude est d'analyser la concordance de ce logiciel pour la **détection** d'une activité inflammatoire radiologique de la SEP, définie par l'apparition d'une nouvelle lésion et/ou lésion croissante) en comparaison avec le gold standard (segmentation manuelle réalisée par le radiologue expert assisté d'un logiciel de fusion d'image [LoBI Intellispace philips Portal]). En effet, détecter une activité inflammatoire dans le cadre de la SEP est primordial pour évaluer l'efficacité du traitement et pour établir le pronostic de la maladie.

Les objectifs secondaires sont indirectement d'étudier la **segmentation** des lésions du logiciel en analysant la concordance du nombre de lésions croissantes et de nouvelles lésions et d'identifier les différents **facteurs** en cas de discordance.

II. Article

A. Matériels et méthodes : généralités

1. Caractéristiques de la population

Nous proposons une étude observationnelle analytique et descriptive multicentrique reposant sur une cohorte rétrospective nommée MISSYVV (DPO ; DEC22-140 ; CHU Lille) concernant des patients ayant un diagnostic certifié de SEP pour laquelle 106 patients issus de celle-ci ont été sélectionnés au hasard. La période de réalisation des IRM dans notre étude s'étend du 02/11/2016 au 19/07/2024. Chaque patient effectuait 2 IRM selon un intervalle de temps variable, pouvant être dans le même centre ou un centre différent, avec des paramètres techniques pouvant varier (épaisseur de coupe, plan d'acquisition, champ magnétique, marque). Le radiologue expert considéré comme Gold Standard relisait toutes les imageries à l'aveugle du logiciel étudié et recueillait pour chaque patient plusieurs données comme le nombre de lésions croissantes et nouvelles lésion et l'activité inflammatoire radiologique (nombre de lésions croissantes et/ou nouvelles strictement supérieur à 0 quelque soit le nombre exact). Une étude analytique et descriptive des facteurs était établie pour chaque valeur discordante de ces valeurs.

2. Variables recueillies

Parmi les variables recueillies nécessaires à l'étude : les variables qualitatives ont été décrites en termes de fréquences et de pourcentages et les variables quantitatives ont été décrites par la moyenne et l'écart type ou par la médiane et l'intervalle interquartile en cas de distribution non Gaussienne. La normalité des distributions a été vérifiée graphiquement et à l'aide du test de Shapiro-Wilk.

3. Tests statistiques

Trois tests statistiques sont utilisés dans cette étude : le Kappa de Cohen, le test exact de Fisher (en cas d'effectif réel ou théorique < 5) et le test exact du Chi-Deux. Le coefficient Kappa de Cohen pour la concordance s'interprète ainsi : < 0 : discordance ; 0 : aléatoire ; 0-0,2 : aucune ; 0,2-0,4 : minimale (très faible) ; 0,4-0,6 : faible ; 0,6-0,8 : modérée ; 0,8-0,9 : forte (bonne) ; $> 0,9$: très forte (très bonne).

Le niveau de significativité pour le test exact de Fisher et Chi-deux est fixé à 5%. Toutes les analyses statistiques ont été effectuées à l'aide du logiciel SAS (SAS Institute version 9.4).

B. Matériels et méthodes : objectifs de l'étude

1. Objectif principal

Le critère de jugement principal était d'étudier l'évolution des lésions SEP et ainsi affirmer ou infirmer la présence d'une activité inflammatoire (autrement dit l'absence ou la présence d'une nouvelle lésion et/ou une lésion croissante, et ce quel que soit leur nombre exact au-delà d'1 pour ce dernier cas) afin d'analyser la **concordance** pour la détection d'une activité inflammatoire entre le radiologue expert (considéré comme Gold Standard) et le logiciel Pixyl ©.

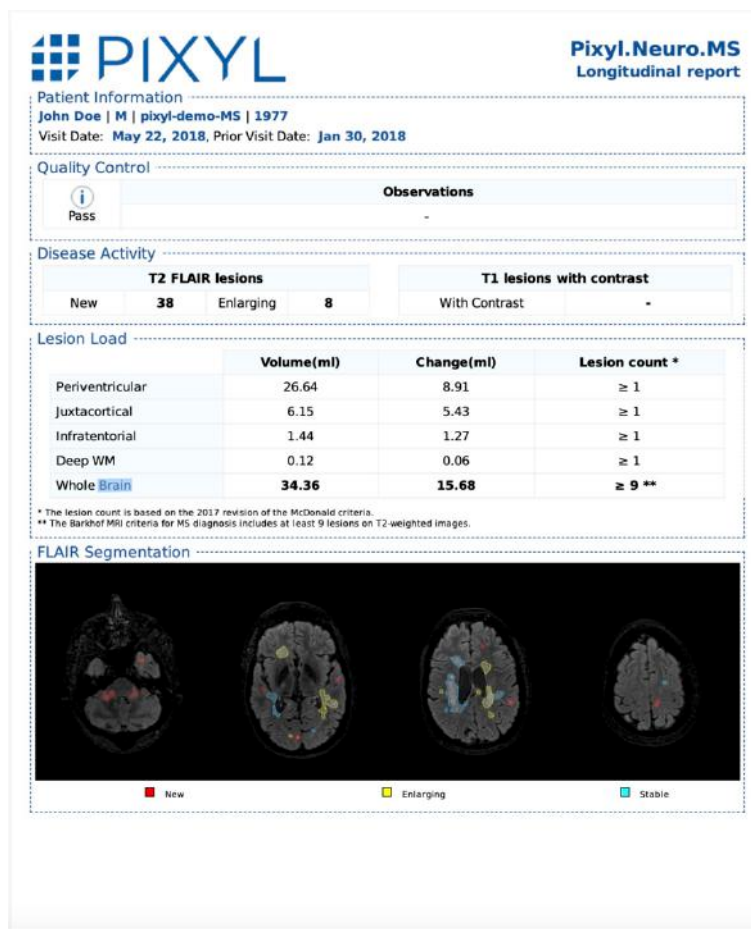
1.1 Le radiologue expert

Le radiologue expert comparait, relisait et interprétait les 2 IRM de temporalités différentes (réalisées à un temps t et à un temps $t+1$) pour un patient donné, à l'aveugle du résultat du logiciel Pixyl ©.

1.2 Le logiciel Pixyl ©

En sélectionnant uniquement les 2 séquences 3D FLAIR (via le PACS : *Picture Archiving and Communication System*) pour les 2 IRM du patient à comparer et en les exportant de façon anonymisée sur le serveur Pixyl ©, ce dernier générerait un résultat écrit en quelques minutes (<10) envoyé sur le PACS indiquant le nombre de lésions croissantes et nouvelles lésions ainsi que le volume des lésions de SEP par région d'intérêt (périventriculaire, corticale, sous-tentorielle, substance blanche profonde). Il produisait également un résultat visuel (également envoyé sur le PACS) avec une image de fusion des 2 IRM où un code couleur mettait en évidence les nouvelles lésions contourées en rouge, les lésions croissantes contourées en jaunes et les lésions stables en bleu.

Figure 31 : Exemple de compte-rendu fourni par Pixyl © ⁽¹¹⁸⁾



1.3 Test statistique

Si le logiciel Pixyl © et le radiologue expert détectaient au moins 1 lésion croissante et/ou 1 nouvelle (et ce quelque soit le nombre exact au-delà de 1), il y avait une concordance attestant une activité inflammatoire.

Si le logiciel Pixyl © et le radiologue expert ne détectaient aucune lésion croissante ni nouvelle lésion, il y avait une concordance sur l'absence d'activité inflammatoire.

A l'inverse si le logiciel Pixyl © détectait au moins 1 lésion croissante et/ou 1 lésion nouvelle a contrario du radiologue expert ne détectant aucune lésion croissante ni aucune nouvelle lésion, et réciproquement ; il y avait une discordance.

Pour répondre au critère de jugement principal, le test statistique utilisé était le coefficient Kappa de Cohen avec un intervalle de confiance à 95%.

- La présence d'une activité inflammatoire (par le logiciel et le radiologue expert) est codée par le chiffre 1.

- L'absence d'activité inflammatoire (par le logiciel et le radiologue expert) est codée par le chiffre 0.
- On répertorie et croise la valeur trouvée pour les 106 patients au sein des 4 cases du tableau :
 - Concordance :
 - Activité inflammatoire radiologue 0 / activité inflammatoire Pixyl © 0
 - Activité inflammatoire radiologue 1 / activité inflammatoire Pixyl © 1
 - Discordance :
 - Activité inflammatoire radiologue 1 / activité inflammatoire Pixyl © 0
 - Activité inflammatoire radiologue 0 / activité inflammatoire Pixyl © 1

2. Objectifs secondaires

Les objectifs secondaires étaient d'étudier la qualité de segmentation du logiciel en analysant la concordance non pas sur la détection d'une activité inflammatoire (au moins une lésion croissante et/ou nouvelle quelque soit le nombre exact) mais sur le nombre exact de lésions croissantes et nouvelles, puis d'en identifier les facteurs en cas de discordance et leur fréquence. Dans un second temps, l'étude étant multicentrique, chaque patient pouvait réaliser leur 2 IRM dans le même centre ou un centre différent, et aussi avec des paramètres techniques (champ magnétique, épaisseur de coupe, plan d'acquisition, marque de l'IRM) qui peuvent différer quel que soit le centre de réalisation de l'examen. L'objectif étant de déterminer si la variabilité des paramètres extrinsèques et intrinsèques pour 2 imageries de suivi influait sur la performance du logiciel à détecter les nouvelles lésions et les lésions croissantes comparativement au Gold Standard.

2.1 Concordance du nombre de nouvelles lésions et croissantes

Pour chaque patient, le radiologue, à l'aveugle du résultat du logiciel, répertoriait le nombre de nouvelles lésions et le nombre de lésions croissantes.

D'autre part en exportant sur le serveur de Pixyl © les séquences 3D FLAIR pour les 2 IRM pour chaque patient, le logiciel générait un compte-rendu envoyé sur le PACS mentionnant également le nombre de nouvelles lésions et lésions croissantes. Les valeurs trouvées étaient classées en 4 catégories : 0, 1, 2 et > 2. Le test statistique utilisé pour analyser la concordance de ces résultats était le Kappa de Cohen avec un intervalle de confiance à 95%.

2.2 Facteurs étiologiques en cas de discordance

2.2.1 Discordance sur le nombre de lésions croissantes

Pour chaque valeur discordante sur les lésions croissantes entre le logiciel et le radiologue expert pour un patient donné, on établit une ou plusieurs raisons pouvant expliquer chaque donnée différente.

Au total, sur les 106 patients, 8 facteurs peuvent expliquer une discordance sur le nombre de lésions croissantes établi par le radiologue expert et le logiciel :

- ⇒ Une segmentation erratique par Pixyl© c'est-à-dire des lésions faussement étiquetées comme des lésions de SEP et segmentées par Pixyl alors qu'elles n'en sont pas une :
 - Artefacts tout de même segmentés par Pixyl © (*raison 1*)
 - Des lésions non-SEP à savoir de la substance blanche normale – segmentées par le logiciel (*raison 2*)
- ⇒ Une absence de segmentation d'une lésion de SEP par le logiciel :
 - Un oubli authentique du logiciel en cas de toutes petites lésions (*raison 7*).
- ⇒ Une lésion de SEP authentique détectée par le logiciel mais avec une segmentation erronée
 - Une lésion étiquetée par le logiciel comme stable alors qu'elle est croissante ce qui signifie une sous-estimation et une segmentation insuffisante par le logiciel (*raison 6*).
 - Une lésion étiquetée par le logiciel comme croissante alors qu'elle est stable donc une sur-estimation par le logiciel concernant surtout les lésions périventriculaires (*raison 3*).
 - Une lésion étiquetée par le logiciel comme croissante alors qu'elle est stable donc une segmentation excessive par le logiciel en cas de charge lésionnelle importante (*raison 4*).
 - Une lésion étiquetée par le logiciel comme croissante alors qu'elle est stable donc une segmentation excessive à cause de paramètres techniques différents et/ou une mauvaise qualité de l'IRM (*raison 5*).
- ⇒ Un oubli/erreur du radiologue (bien que Gold Standard) (*raison 8*)

Figure 32 : Exemple d'une discordance sur le nombre de lésions croissantes (raison 3)

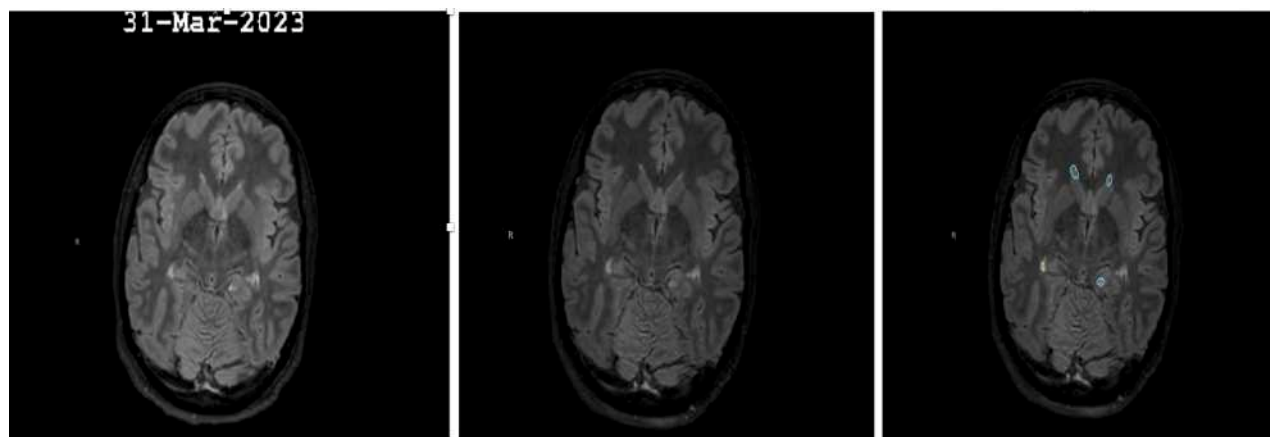


Image de gauche : IRM séquence axiale FLAIR au temps t

IRM du milieu : IRM séquence axiale FLAIR au temps t+1

IRM de droite : IRM Pixyl© au temps t et t+1 avec une délimitation en jaune d'une lésion croissante en regard de la corne temporale droite (périventriculaire) qui en fait semble être stable.

2.2.2 Discordance sur le nombre de nouvelles lésions

On établit une ou plusieurs raisons pouvant expliquer chaque donnée différente pour chaque valeur discordante sur les nouvelles lésions entre le logiciel et le radiologue expert pour un patient donné.

Au total sur les 106 patients, 9 facteurs pouvaient expliquer une discordance sur le nombre de nouvelles lésions établi par le radiologue expert et le logiciel :

- ⇒ Une segmentation erratique par Pixyl© c'est-à-dire des lésions faussement étiquetées comme des lésions de SEP et segmentées par Pixyl alors qu'elles n'en sont pas une :
 - Artefacts tout de même segmentés par Pixyl © (*raison 1*).
 - Des lésions non-SEP à savoir de la substance blanche normale – segmentées par le logiciel (*raison 2*).
- ⇒ Une absence de segmentation d'une nouvelle lésion de SEP par le logiciel :
 - Un oubli authentique du logiciel en cas de toutes petites lésions non vues ou prise pour de la substance blanche normale (*raison 7*).
 - Un oubli authentique du logiciel en cas de charge lésionnelle importante avec une coalescence des lésions pouvant rendre difficile la visualisation d'une nouvelle lésion parmi l'amas de lésions déjà présentes (*raison 8*).

⇒ Une lésion de SEP authentique détectée par le logiciel mais avec une segmentation erronée :

- Une lésion étiquetée par le logiciel comme croissante alors qu'elle est nouvelle ce qui signifie une sous-estimation et une segmentation insuffisante par le logiciel notamment en cas de charge lésionnelle importante pouvant rendre difficile la visualisation d'une nouvelle lésion parmi l'amas de lésions déjà présentes (*raison 6*).
- Une lésion étiquetée par le logiciel comme stable alors qu'elle est nouvelle donc une sous-estimation et une segmentation insuffisante imputée à des paramètres techniques différents (*raison 5*).
- Une lésion étiquetée par le logiciel comme nouvelle alors qu'elle est croissante donc une segmentation excessive par le logiciel en cas de lésion non vue sur la première IRM car petite en taille mais qui augmente considérablement de taille sur l'IRM 2 et donc devient visible par le logiciel qui la considère comme nouvelle (*raison 3*).
- Une lésion étiquetée par le logiciel comme nouvelle alors qu'elle est croissante donc une segmentation excessive à cause de paramètres techniques différents (*raison 4*).

⇒ Un oubli/erreur du radiologue (bien que Gold Standard) (*raison 9*).

Figure 33 : Exemple d'une discordance sur le nombre de nouvelles lésions (raison 1)

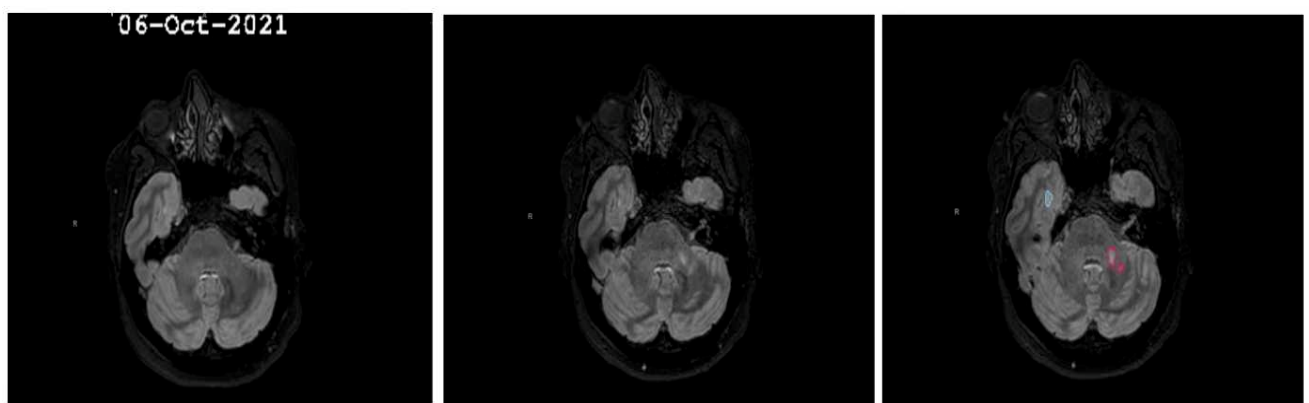


Image de gauche : IRM séquence axiale FLAIR au temps t

IRM du milieu : IRM séquence axiale FLAIR au temps t+1

IRM de droite : IRM Pixyl© au temps t et t+1 avec une délimitation en rouge d'une nouvelle lésion cérébelleuse supérieure gauche qui est en fait un artefact de susceptibilité

2.3 L'implication des paramètres techniques sur la détection des lésions croissantes et nouvelles.

2.3.1 Sur les lésions croissantes

Le but était d'analyser si la différence pour chaque paramètre technique était un facteur pouvant expliquer la discordance de détection des lésions croissantes entre le logiciel et le radiologue expert.

L'IRM t et l'IRM $t+1$ réalisées par chaque patient peuvent avoir des caractéristiques différentes.

Les paramètres techniques recueillis étaient :

- L'épaisseur de coupe : variant entre 0,40 et 4,00 mm
- Le champ magnétique : 1,5 Tesla ou 3 Tesla
- La marque de l'IRM : Philips, GE (General Electric), Siemens
- Le plan d'acquisition pour la séquence 3D : sagittal ou axial
- Le centre de l'IRM : CHU de Lille, Climal, Bois, St Philibert, St Vincent, Seclin, Léonard de Vinci, Roubaix, Cateau-Cambrésis, Centre Hospitalier de Valenciennes, Parc, Clarence, Riaumont, GIMA, Imanord, Godinot Reims, Clermont-Ferrand, Le Havre

On croisait chacun de ces paramètres techniques selon leur similarité (codée 1) ou leur différence (codée 0) avec le nombre de patients présentant une concordance et une discordance pour la détection de lésions croissantes par le logiciel et le radiologue expert.

Pour ce faire, le test statistique utilisé était le test exact du Chi-deux avec un niveau de significativité à 5%.

2.3.2 Sur les nouvelles lésions

Le but était d'analyser si la différence pour chaque paramètre technique était un facteur pouvant expliquer la discordance de détection des nouvelles lésions entre le logiciel et le radiologue expert.

L'IRM t et l'IRM $t+1$ réalisées par chaque patient peuvent avoir des caractéristiques différentes.

Les paramètres techniques recueillis étaient :

- L'épaisseur de coupe : variant entre 0,40 et 4,00 mm
- Le champ magnétique : 1,5 Tesla ou 3 Tesla
- La marque de l'IRM : Philipps, GE (General Electric), Siemens
- Le plan d'acquisition pour la séquence 3D : sagittal ou axial
- Le centre de l'IRM : CHU de Lille, Climal, Bois, St Philibert, St Vincent, Seclin, Léonard de Vinci, Roubaix, Cateau-Cambrésis, Centre Hospitalier de Valenciennes, Parc, Clarence, Riaumont, GIMA, Imanord, Godinot Reims, Clermont-Ferrand, Le Havre

On croisait chacun de ces paramètres techniques selon leur similarité (codée 1) ou leur différence (codée 0) avec le nombre de patients présentant une concordance et une discordance pour la détection de nouvelles lésions par le logiciel et le radiologue expert.

Pour ce faire, le test statistique utilisé était le test exact de Fisher (effectifs théoriques < 5) avec un niveau de significativité à 5%.

C. Résultats

1. Description de la population et des paramètres extrinsèques et intrinsèques de l'IRM

1.1 Description de la population

La cohorte comportait 106 patients dont 40 hommes (37,7%) et 66 femmes (62,3%). Cette distribution selon le sexe était un reflet de la répartition réelle de la SEP avec pour rappel un sex ratio de 3 : 1 (femmes : hommes) (cf I] A.2).

Lors de la réalisation de la 1^{ère} IRM, l'âge moyen des patients inclus est de 39,4 ans ($\pm 10,7$). Le plus jeune patient avait 12 ans et le plus âgé avait 74 ans, lors de la réalisation de la première IRM.

La médiane du délai de réalisation entre les 2 IRM était de 12 mois avec un premier quartile à 11 mois et un troisième quartile à 16 mois. Le délai le plus court entre 2 IRM était de 0,2 mois et le plus long de 115 mois.

L'ensemble de ces données est détaillé dans le *tableau 5* ci-dessous.

Tableau 5 : Description de la population

Sexe	Masculin	40 (37,7%)
	Femme	66 (62,3%)
Age lors de la 1ère IRM (année)	Moyenne ± écart-type	39,4 ± 10,7
	Médiane (Q1 ; Q3)	38,0 (33,0 ; 46,0)
	Minium Maximum	12,0 74,0
Délai entre 2 IRM (mois)	Moyenne ± écart-type	17,0 ± 16,1
	Médiane (Q1 ; Q3)	12,0 (11,0 ; 16,0)
	Minium Maximum	0,2 115,0

1.2 Description des paramètres extrinsèques et intrinsèques de l'IRM

L'étude étant multicentrique, les IRM pour un même patient ou pour des patients différents pouvaient se dérouler dans le même centre ou un centre différent (paramètre intrinsèque). D'autre part, que le centre soit identique ou non, certains paramètres intrinsèques pouvaient tout de même varier comme le champ magnétique, la marque de l'IRM, le plan d'acquisition en 3D et l'épaisseur de coupe.

Un des objectifs secondaires de cette étude était de déterminer si la variabilité de chacun de ces paramètres influait sur la performance du logiciel à détecter des nouvelles lésions et des lésions croissantes.

Concernant l'IRM n°1 réalisée pour un patient donné (sur les 106 patients de l'étude): 83 (78,3%) des IRM n°1 étaient réalisées au CHU de Lille avec un champ magnétique à 3T pour 89 d'entre elles (84%), une épaisseur de coupe à 1,12 mm pour 78 d'entre elles (73,6%) et un plan d'acquisition en sagittal pour la séquence 3D pour 101 patients (95,3%). La marque prédominante était Philips concernant 89 (84%) des IRM n°1 réalisées.

Concernant l'IRM n°2 réalisée pour un patient donné (sur les 106 patients de l'étude) : 90 (84,9%) des IRM n°2 étaient réalisées au CHU de Lille avec un champ magnétique à 3T pour 92 d'entre elles (86,8%), une épaisseur de coupe à 1,12 mm pour 47 d'entre elles (44,3%) et un plan d'acquisition en sagittal pour la séquence 3D pour 64 patients (60,3%). La marque prédominante était Philips concernant 91 (85,8%) des IRM n°2 réalisées.

Au total sur les 212 IRM, 173 (81,6%) ont été réalisées au CHU avec un champ magnétique à 3T pour 181 d'entre elles (85,3%), une épaisseur de coupe à 1,12 mm

pour 125 d'entre elles (58,9%), un plan d'acquisition en sagittal pour la séquence 3D pour 165 patients (77,8%) et la marque Philips pour 180 IRM (84,9%).

En comparant les 2 IRM pour le même patient : le champ magnétique utilisé était identique pour 89 patients (84%), l'épaisseur de coupe est identique pour 47 patients (44,3%), le centre était identique pour 87 patients (82,1%), le plan d'acquisition est identique pour 63 patients (59,4%) et la marque était identique pour 92 patients (86,8%). Au total le protocole des 2 IRM (même séquence et mêmes paramètres cités ci-dessus) était identique pour 31 patients (29,2%).

L'ensemble de ces données sont résumées dans le tableau 6 ci-dessous :

Tableau 6 : Paramètres extrinsèques et intrinsèques des IRM réalisées

IRM Paramètres		IRM 1	IRM 2	Total
Centre	CHU de Lille	83 (78,3%)	90 (84,9%)	173 (81,6%)
	Climal	5 (4,7%)	2 (1,9%)	7 (3,3%)
	Bois	5 (4,7%)	3 (2,8%)	8 (3,8%)
	St-Philibert	1 (0,9%)	1 (0,9%)	2 (0,9%)
	St-Vincent	1 (0,9%)	0 (0,0%)	1 (0,5%)
	Seclin	1 (0,9%)	3 (2,8%)	4 (1,9%)
	Léonard de Vinci	1 (0,9%)	2 (1,9%)	3 (1,4%)
	Roubaix	1 (0,9%)	0 (0,0%)	1 (0,5%)
	Cateau-Cambrésis	1 (0,9%)	1 (0,9%)	2 (0,9%)
	Centre Hospitalier de Valenciennes	1 (0,9%)	1 (0,9%)	2 (0,9%)
	Parc	1 (0,9%)	0 (0,0%)	1 (0,5%)
	Clarence	1 (0,9%)	0 (0,0%)	1 (0,5%)
	Riaumont	1 (0,9%)	0 (0,0%)	1 (0,5%)
	GIMA	1 (0,9%)	0 (0,0%)	1 (0,5%)
	Imanord	0 (0,0%)	2 (1,9%)	2 (0,9%)
	Godinot-Reims	0 (0,0%)	1 (0,9%)	1 (0,5%)
	Clermont-Ferrand	1 (0,9%)	0 (0,0%)	1 (0,5%)
	Le Havre	1 (0,9%)	0 (0,0%)	1 (0,5%)
Champ magnétique (Tesla)	1,5	17 (16,0%)	14 (13,2%)	31 (14,7%)
	3	89 (84,0%)	92 (86,8%)	181 (85,3%)
Épaisseur de coupe (mm)	0,40	1 (0,9%)	0 (0,0%)	1 (0,5%)
	0,57	0 (0,0%)	1 (0,9%)	1 (0,5%)
	0,60	4 (3,8%)	39 (36,8%)	43 (20,3%)
	0,70	2 (1,9%)	2 (1,9%)	4 (1,9%)
	0,72	1 (0,9%)	2 (1,9%)	3 (1,4%)
	0,89	0 (0,0%)	1 (0,9%)	1 (0,5%)
	0,90	1 (0,9%)	0 (0,0%)	1 (0,5%)
	0,95	3 (2,8%)	1 (0,9%)	4 (1,9%)
	1,00	3 (2,8%)	2 (1,9%)	5 (2,4%)
	1,03	0 (0,0%)	1 (0,9%)	1 (0,5%)

	1,05	0 (0,0%)	1 (0,9%)	1 (0,5%)
	1,12	78 (73,6%)	47 (44,3%)	125 (59,0%)
	1,20	6 (5,7%)	3 (2,8%)	9 (4,2%)
	1,28	3 (2,8%)	3 (2,8%)	5 (2,4%)
	1,30	1 (0,9%)	0 (0,0%)	1 (0,5%)
	1,40	2 (1,9%)	1 (0,9%)	3 (1,4%)
	1,50	0 (0,0%)	1 (0,9%)	1 (0,5%)
	1,60	1 (0,9%)	0 (0,0%)	1 (0,5%)
	4,00	0 (0,0%)	1 (0,9%)	1 (0,5%)
Plan d'acquisition	Sagittal	101 (95,3%)	64 (60,3%)	165 (77,8%)
	Axial	5 (4,7%)	42 (39,7%)	47 (22,2%)
Marque	Philipps	89 (84,0%)	91 (85,8%)	180 (84,9%)
	GE	6 (5,7%)	7 (6,6%)	13 (6,1%)
	Siemens	11(10,3%)	8 (7,5%)	19 (9,0%)

2. Objectif principal

L'objectif principal de l'étude était d'analyser la performance du logiciel Pixyl © à identifier une activité inflammatoire radiologique de la SEP autrement dit la capacité de détecter la présence ou l'absence d'au moins une lésion SEP croissante et/ou une nouvelle lésion parmi 2 IRM, sans préjuger du nombre exact de lésions croissantes ou nouvelles détectées, le tout comparativement au radiologue expert (Gold Standard).

Le radiologue expert avait détecté 69 patients ne présentant pas une activité radiologique sur leur IRM contre 37 en présentant une.

Le logiciel Pixyl © avait détecté 27 patients ne présentant pas une activité radiologique sur leur IRM contre 79 en présentant une.

Le radiologue expert et le logiciel étaient concordants pour 60 patients sur la présence ou l'absence d'une activité inflammatoire radiologique (35 vrais positifs (VP) et 25 vrais négatifs (VN)) et discordants pour 46 patients (44 faux positifs (FP) et 2 faux négatifs (FN)). Par conséquent le logiciel était sensible avec une sensibilité estimée à 94,6% ($VP / (VP+FN)$) pour détecter la présence d'une activité inflammatoire mais peu spécifique pour détecter l'absence d'une activité inflammatoire radiologique (spécificité estimée à 32,6% : $VN / (VN + FP)$).

La valeur prédictive positive (VPP) était faible, estimée à 44,3% ($VP / (FP+VP)$) donc quand le logiciel détectait la présence d'une activité inflammatoire radiologique, il a raison dans 44% des cas.

La valeur prédictive négative (VPN) était forte, estimée à 92,6% ($VN/(VN+FN)$) donc quand le logiciel ne détectait pas la présence d'une activité inflammatoire radiologique, il avait raison dans 92,6% des cas.

En utilisant le test statistique Kappa de Cohen, il en résultait un kappa à 0,24 (intervalle de confiance à 95% [0,12 ; 0,37]) équivalent à une concordance minimale (très faible).

Au total : la performance du logiciel Pixyl © semble très faible pour la détection d'une activité inflammatoire radiologique dans le cadre de la SEP.

L'ensemble est résumé dans le *tableau 7* ci-dessous :

Tableau 7 : Concordance pour la détection d'une activité inflammatoire

Activité inflammatoire du radiologue	Activité inflammatoire Pixyl©				
	0	Au moins 1	Total		
0	25	44	69	VN	FP
Au moins 1	2	35	37	FN	VP
Total	27	79	106		
Coefficient kappa	Kappa = 0,24 (intervalle de confiance à 95% [0,12 ; 0,37])				

3. Objectifs secondaires

3.1 Concordance pour la détection de nouvelles lésions et lésions croissantes

La performance du logiciel Pixyl © étant très faible pour la détection d'une activité inflammatoire radiologique de la SEP, ces objectifs secondaires permettaient de déterminer si cette faible concordance était imputée à une performance sous-optimale du logiciel pour la détection uniquement des nouvelles lésions, des lésions croissantes ou les 2.

3.1.1 Concordance pour la détection des nouvelles lésions.

L'objectif était d'analyser la capacité du logiciel à détecter de façon authentique au moins une nouvelle lésion comparativement au Gold Standard, quel que soit le nombre de nouvelles lésions détectées.

Le radiologue-expert avait détecté 29 patients (27,4%) présentant au moins une nouvelle lésion sur leur IRM et le logiciel Pixyl © en avait détecté 31 (29,2%).

Le radiologue-expert avait détecté 77 patients (72,6%) ne présentant pas de nouvelles lésions contre 72 patients pour le logiciel (68%).

Le radiologue expert et le logiciel étaient concordants pour 98 patients sur la présence ou l'absence d'une nouvelle lésion (26 vrais positifs (VP) et 72 vrais négatifs (VN)) et discordants pour 8 patients (5 faux positifs (FP) et 3 faux négatifs (FN)). Par conséquent le logiciel était sensible avec une sensibilité estimée à 89,6% ($VP / (VP+FN)$) pour détecter la présence d'au moins une nouvelle lésion et spécifique pour détecter l'absence de nouvelle lésion (spécificité estimée à 93,5% : $VN / (VN + FP)$).

La valeur prédictive positive (VPP) était élevée, estimée à 83,9% ($VP / (FP+VP)$) donc quand le logiciel détecte la présence d'une nouvelle lésion, il avait raison dans 83,9% des cas.

La valeur prédictive négative (VPN) était forte, estimée à 96% ($VN/(VN+FN)$) donc quand le logiciel ne détectait pas la présence d'au moins une nouvelle lésion, il avait raison dans 96% des cas.

En utilisant le test statistique Kappa de Cohen, il en résultait un kappa à 0,81 (intervalle de confiance à 95% [0,69 ; 0,94]) équivalent à une concordance forte.

Au total : la performance du logiciel Pixyl © semble forte pour la détection de nouvelles lésions dans le cadre de la SEP.

L'ensemble est résumé dans le *tableau 8* ci-dessous :

Tableau 8 : Concordance pour la détection de nouvelles lésions

Nouvelles lésions radiologue	Nouvelles lésions Pixyl ©				
	0	Au moins 1	Total		
0	72	5	77	VN	FP
Au moins 1	3	26	29	FN	VP
Total	75	31	106		
Coefficient kappa	Kappa = 0,81 (intervalle de confiance à 95% [0,69 ; 0,94])				

3.1.2 Concordance pour la détection de lésions croissantes

L'objectif était d'analyser la capacité du logiciel à détecter de façon authentique au moins une lésion croissante comparativement au Gold Standard, quel que soit le nombre de lésions croissantes détectées.

Le radiologue-expert avait détecté 19 patients (27,4%) présentant au moins une lésion croissante sur leur IRM et le logiciel Pixyl © en avait détecté 72 (67,9%). Le radiologue-expert avait détecté 87 patients (82,1%) ne présentant pas de lésions croissantes contre 34 (32,0%) pour le logiciel.

Le radiologue expert et le logiciel étaient concordants pour 45 patients sur la présence ou l'absence d'une lésion croissante (30 vrais positifs (VP) et 15 vrais négatifs (VN)) et discordants pour 61 patients (57 faux positifs (FP) et 4 faux négatifs (FN)). Par conséquent le logiciel est peu sensible avec une sensibilité estimée à 30,6% ($VP / (VP+FN)$) pour détecter la présence d'au moins une lésion croissante et spécifique pour détecter l'absence de lésion croissante (spécificité estimée à 88,2% : $VN / (VN + FP)$).

La valeur prédictive positive (VPP) était élevée, estimée à 78,9% ($VP / (FP+VP)$) donc quand le logiciel détecte la présence d'une lésion croissante, il a raison dans 78,9% des cas.

La valeur prédictive négative (VPN) était faible, estimée à 46,9% ($VN/(VN+FN)$) donc quand le logiciel ne détecte pas la présence d'au moins une lésion croissante, il a raison dans moins de la moitié des cas.

En utilisant le test statistique Kappa de Cohen, il en résultait un kappa à 0,06 (intervalle de confiance à 95% [-0,04 ; 0,16]) équivalent à une concordance nulle.

Au total : la performance du logiciel Pixyl © est proche de l'aléatoire pour la détection de lésions croissantes dans le cadre de la SEP.

Tableau 9 : Concordance pour la détection des lésions croissantes

Lésions croissantes radiologue	Lésions croissantes Pixyl ©				
	0	Au moins 1	Total		
0	30	57	87	VN	FP
Au moins 1	4	15	19	FN	VP
Total	34	72	106		
Coefficient kappa	Kappa = 0,06 (intervalle de confiance à 95% [-0,04 ; 0,16])				

3.2 Concordance pour la détection du nombre de lésions croissantes et nouvelles lésions.

L'objectif était non plus d'étudier la capacité de détection du logiciel mais la qualité de segmentation en analysant cette fois-ci le nombre précis de nouvelles lésions et lésions croissantes détectées par Pixyl © comparativement au Gold Standard.

3.2.1 Concordance pour le nombre de nouvelles lésions détectées.

La médiane du nombre de nouvelles lésions détectées par le radiologue était 0,0 avec un premier quartile à 0,0 et un troisième quartile à 1,0.

La médiane du nombre de nouvelles lésions détectées par Pixyl était 0,0 avec un premier quartile à 0,0 et un troisième quartile à 1,0.

Le nombre de nouvelles lésions détectées ((classé en 4 catégories : 0,1,2 et > 2) par le radiologue expert et le logiciel Pixyl© pour chaque patient ainsi que la différence du nombre de nouvelles lésions détectées entre les 2 sont reportés respectivement dans le *tableau 10* et les *figures 34* et *35*. En utilisant le test statistique du Kappa de Cohen, le coefficient kappa pondéré était à 0,81 avec un intervalle de confiance à 95% [0,72 ; 0,91] équivalant à une concordance forte pour la quantification du nombre de nouvelles lésions.

Au total : la performance du logiciel Pixyl © semble forte pour dénombrer le nombre de nouvelles lésions dans le cadre de la SEP ce qui dénote d'une capacité efficace à soustraire 2 IRM pour révéler les nouvelles lésions.

Tableau 10 : Concordance pour le nombre de nouvelles lésions

Nombre de nouvelles lésions détectées par Pixyl ©	Nombre de nouvelles lésions détectées par le radiologue expert				
	0	1	2	>2	Total
0	72	4	1	0	77
1	3	5	1	2	11
2	0	2	3	0	5
>2	0	0	2	11	13
Total	75	11	7	13	106
Coefficient kappa	Kappa pondéré : 0,81 (intervalle de confiance à 95% [0,72 ; 0,91])				

Figure 34 : Différence du nombre de nouvelles lésions détectées par le logiciel Pixyl© et le radiologue expert.

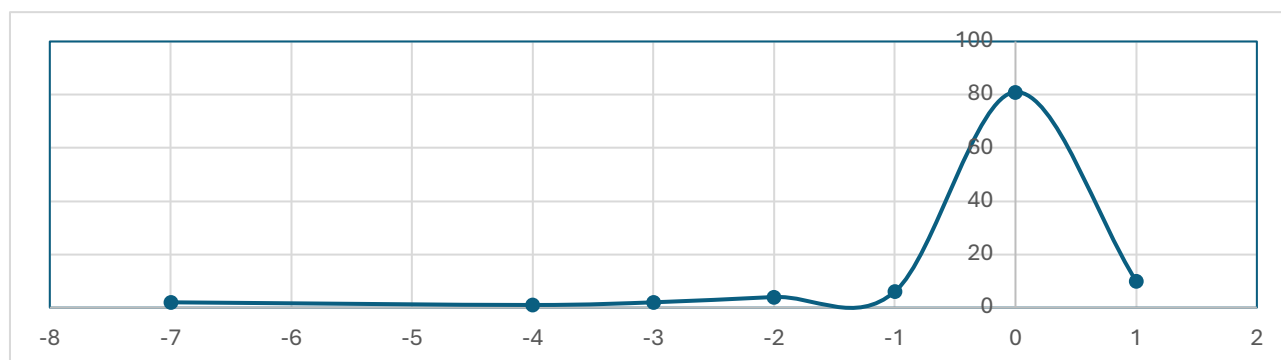
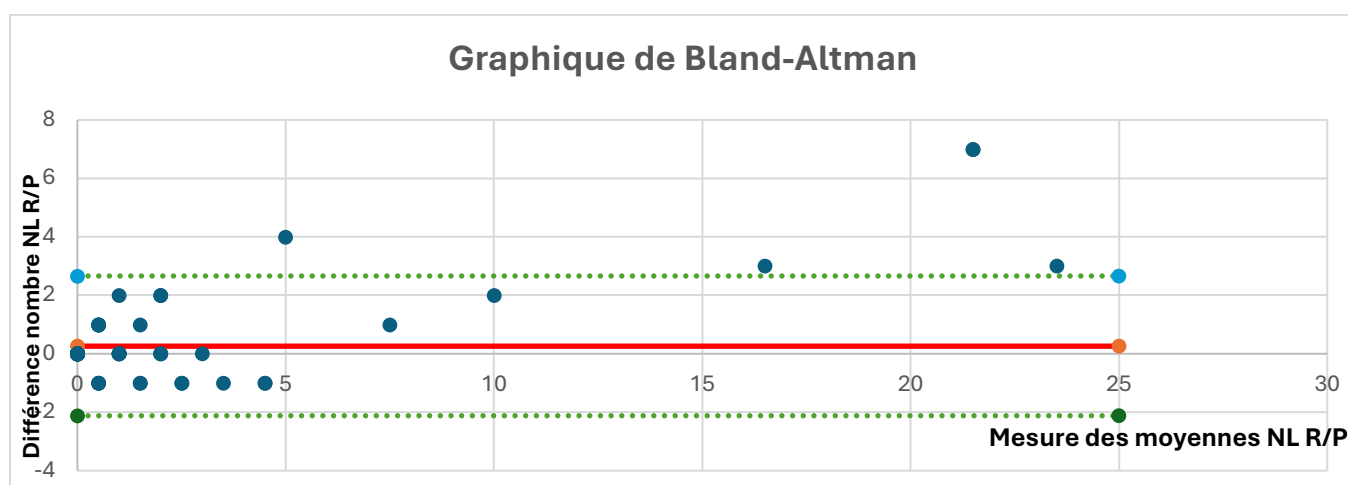


Figure 35 : Graphique de Bland-Altman illustrant la comparaison du nombre des nouvelles lésions entre le logiciel et le radiologue-expert



3.2.2 Concordance pour le nombre de lésions croissantes détectées

La médiane du nombre de lésions croissantes détectées par le radiologue était de 0,0 avec un premier quartile à 0,0 et un troisième quartile à 0,0.

La médiane du nombre de lésions croissantes détectées par Pixyl était de 1,0 avec un premier quartile à 0,0 et un troisième quartile à 4,0.

Le nombre de nouvelles lésions détectées (classé en 4 catégories 0,1,2 et > 2) par le radiologue expert et le logiciel Pixyl © ainsi que la différence du nombre de nouvelles lésions détectées entre les 2 sont reportés respectivement dans le *tableau 11* et les *figure 36 et 37*. En utilisant le test statistique Kappa de Cohen, le coefficient kappa pondéré était à 0,06 avec un intervalle de confiance à 95% [-0,01 ; 0,13] équivalant à une concordance nulle pour la quantification du nombre de lésions croissantes.

Au total : la performance du logiciel Pixyl © semble proche du hasard pour dénombrer le nombre de lésions croissantes dans le cadre de la SEP, étant la conséquence d'une superposition et d'une segmentation inefficace du logiciel pour comparer l'évolution d'une lésion à deux temps différents.

Tableau 11 : Concordance pour le nombre de lésions croissantes

Nombre de lésions croissantes détectées par le radiologue expert	Nombre de lésions croissantes détectées par Pixyl ©				
	0	1	2	>2	Total
0	30	17	9	31	87
1	4	1	1	6	12
2	0	1	0	3	4
>2	0	1	0	2	3
Total	34	20	10	42	106
Coefficient kappa	Kappa pondéré : 0,06 (intervalle de confiance à 95% [-0,01 ; 0,13])				

Figure 36 : Différence du nombre de lésions croissantes détectées par le logiciel Pixyl© et le radiologue expert

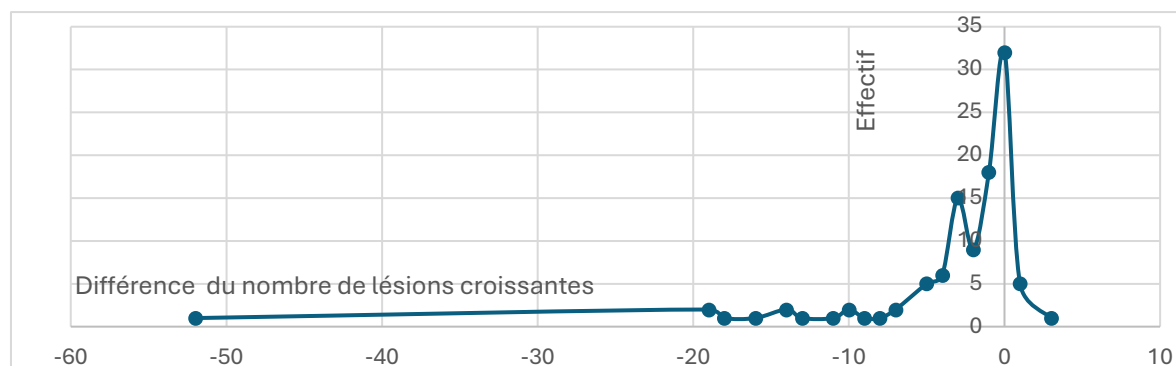
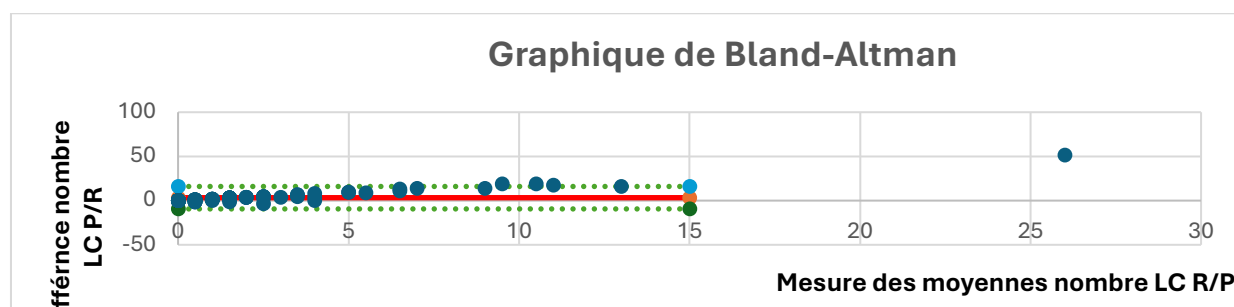


Figure 37 : Graphique de Bland-Altman illustrant la comparaison des lésions croissantes entre le logiciel et le radiologue-expert



3.3 Facteurs étiologiques expliquant une discordance.

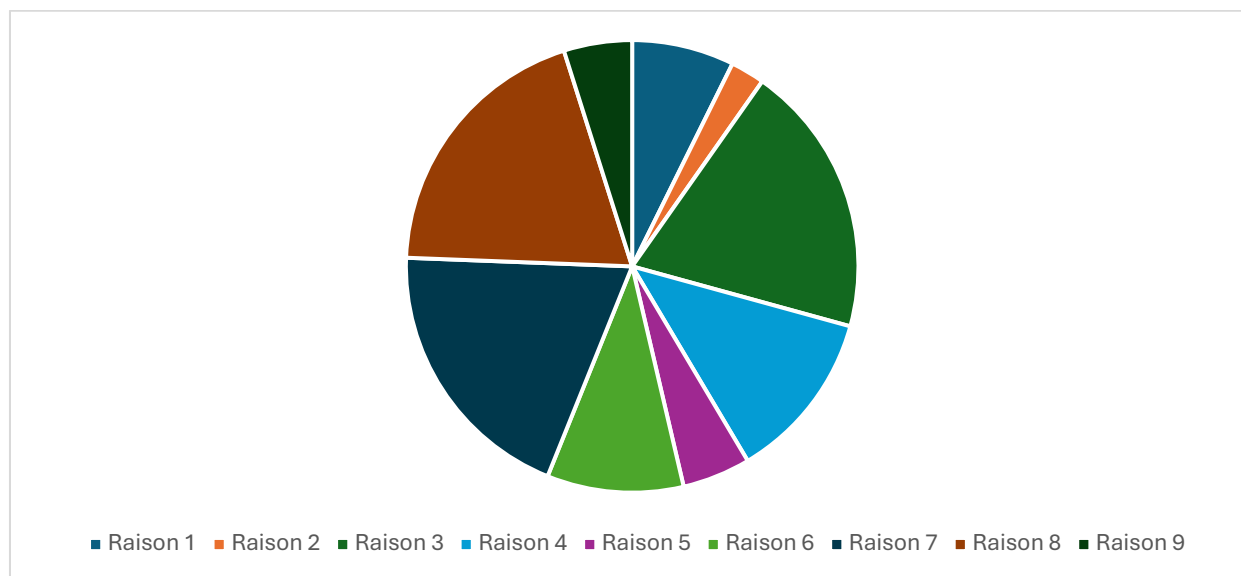
Comme détaillée dans la partie « matériel et méthodes », chaque discordance sur la valeur du nombre de lésions croissantes et nouvelles était tributaire d'une ou plusieurs étiologies.

3.3.1 Facteurs étiologiques d'une discordance pour le nombre des nouvelles lésions

Parmi les discordances sur le nombre de nouvelles lésions détectées s'élevant au nombre de 25 : 8 étaient liées à une segmentation excessive du logiciel délimitant une nouvelle lésion authentique de SEP alors qu'elle était déjà présente sur l'IRM princeps mais de taille minime, ainsi passée inaperçue par le logiciel (*raison 3*) ; 8 étaient liées à un oubli du logiciel ne délimitant pas de lésions de SEP imputée au fait qu'elle soit de petite taille ou considérée comme de la substance blanche normale (*raison 7*) ; 8 étaient liées également à un oubli du logiciel ne délimitant aucune lésion de SEP suite à une charge lésionnelle importante avec une coalescence de plusieurs lésions (*raison 8*) ; 5 étaient liées à une segmentation excessive / surestimation des lésions où le logiciel décrit une lésion comme nouvelle alors qu'elle est croissante, imputée à des paramètres techniques différents (*raison 4*) ; 4 étaient liées à une segmentation insuffisante par le logiciel authentifiant une lésion SEP comme croissante alors qu'elle est nouvelle, le tout est lié à une charge lésionnelle importante et une coalescence de multiples lésions où la détection et l'évolution individuelles des lésions sont moins optimales (*raison 6*) ; 3 étaient liées à une segmentation erratique causée par des artefacts où des lésions non SEP sont faussement identifiées comme des nouvelles lésions (*raison 1*) ; 2 étaient liées à une segmentation insuffisante / sous-estimation des lésions où le logiciel atteste de la croissance d'une lésion alors qu'elle est nouvelle, imputée à des paramètres techniques différents (*raison 5*) ; 2 étaient liées à un oubli du radiologue (bien que Gold Standard), 1 était liée à une segmentation erratique par le logiciel qui délimite faussement une lésion non SEP (substance blanche normale).

La répartition de ces différentes étiologies pour ces discordances du nombre de nouvelles lésions est mentionnée ci-dessous sur la *figure 38*.

Figure 38 : Répartition par un diagramme en secteur des étiologies de discordance des nouvelles lésions



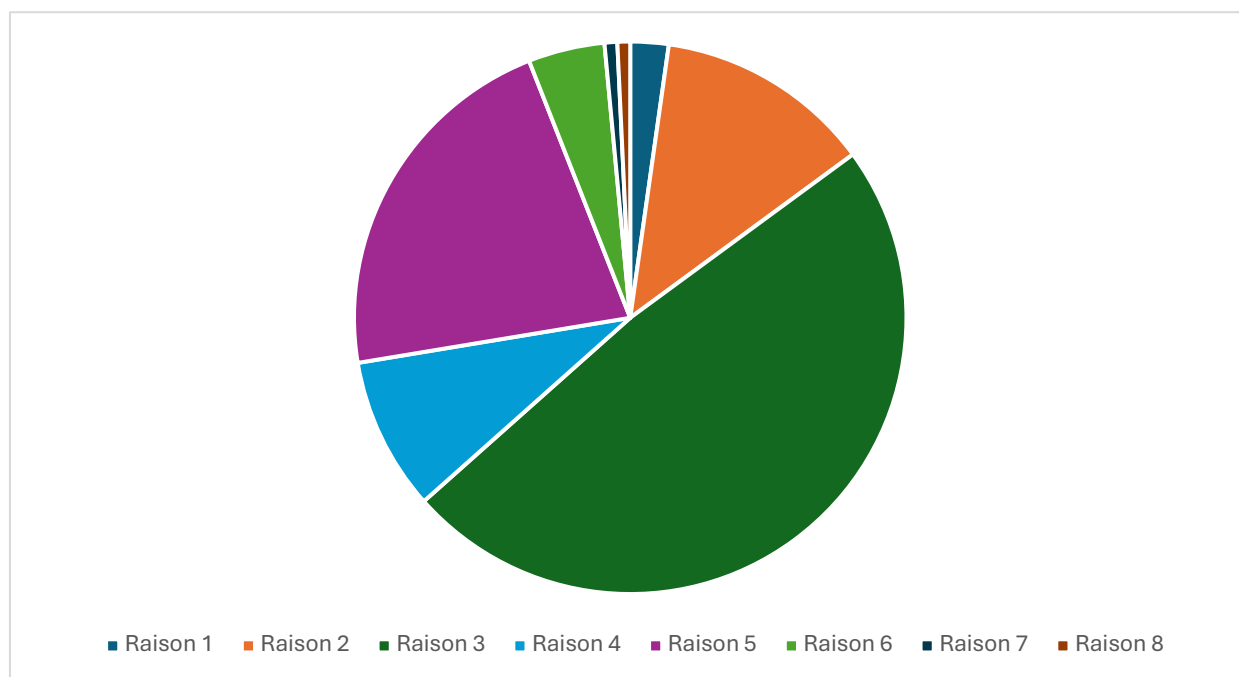
3.3.2 Facteurs étiologiques d'une discordance pour le nombre des nouvelles lésions croissantes

Parmi les 74 lésions croissantes discordantes : 66 étaient liées à une segmentation excessive des lésions étiquetées par le logiciel comme croissante alors qu'elles étaient stables, concernant majoritairement les lésions SEP de topographie périventriculaire (*raison 3*) ; 29 étaient dues à une segmentation excessive d'une lésion étiquetée par Pixyl © comme croissante alors qu'elle était stable, imputée à des paramètres techniques différents (*raison 5*) ; 17 étaient expliquées par une délimitation de lésions non SEP par le logiciel (substance blanche normale) (*raison 2*) ; 12 étaient liées à une segmentation excessive imputée par une charge lésionnelle importante où le logiciel considère une lésion comme croissante alors qu'elle est stable (*raison 4*) ; 6 étaient liées à une délimitation insuffisante des lésions par le logiciel qui décrit une lésion SEP comme stable alors qu'elle était croissante sans topographie préférentielle et sans rapport avec des paramètres techniques qui pourraient différer, en lien avec des lésions de taille millimétrique avec une croissance millimétrique (*raison 6*) ; 3 étaient liées à une segmentation erratique par le logiciel considérant faussement des lésions non SEP comme SEP et croissantes (donc considérée par le logiciel comme peu visible sur l'IRM princeps) en lien avec des artefacts (*raison 1*) ; 1 était liée à un oubli du radiologue Gold Standard (*raison 8*)

et pour finir 1 était liée à un oubli de Pixyl© concernant une lésion millimétrique (*raison 7*).

La répartition des différentes étiologies expliquant la discordance du nombre de lésions croissantes est illustrée par la *figure 39* ci-dessous.

Figure 39 : Répartition par un diagramme en secteur des étiologies des discordances des lésions croissantes



3.4 Influence des paramètres techniques sur la performance du logiciel Pixyl© à détecter des lésions croissantes et des nouvelles lésions

L'objectif était de déterminer si la variabilité de chacun des paramètres intrinsèques de l'IRM que sont : le champ magnétique, l'épaisseur de coupe, le centre, le plan d'acquisition et la marque, influaient sur la capacité du logiciel à détecter les lésions croissantes et les nouvelles lésions (autrement dit à l'origine d'une discordance entre le logiciel et le radiologue pour l'évolution des lésions de SEP)

3.4.1 Influence des paramètres techniques pour la détection des nouvelles lésions

Par le biais d'un test exact de Fisher en comparant 2 groupes (paramètre identique/différent (codé 1/0) et concordance/discordance du nombre de nouvelles

lésions (codé 0/1), aucun des paramètres à savoir le champ magnétique, l'épaisseur de coupe, le centre, le plan d'acquisition et la marque ne constituaient des facteurs de risque significatif expliquant une discordance entre le logiciel et le radiologue expert dans la détection des nouvelles lésions ($p > 0,05$ pour chacune des variables). L'ensemble de ces résultats est illustré dans le *tableau 12* ci-dessous.

Tableau 12 : Paramètres techniques de l'IRM et discordance pour la détection d'une nouvelle lésion

Paramètres techniques		Détection nouvelles lésions		Test	pvalue
		Concordance	Discordance		
	Différent	0	1		
	Identique	N=98	N=8		
Champ magnétique	0	16 (16.3)	1 (12.5)	Fisher exact	1.00
	1	82 (83.7)	7 (87.5)		
Plan d'acquisition	0	42 (42.9)	1 (12.5)	Fisher exact	0.14
	1	56 (57.1)	7 (87.5)		
Centre	0	18 (18.4)	1 (12.5)	Fisher exact	1.00
	1	80 (81.6)	7 (87.5)		
Marque	0	13 (13.3)	1 (12.5)	Fisher exact	1.00
	1	85 (86.7)	7 (87.5)		
Épaisseur de coupe	0	56 (57.1)	3 (37.5)	Fisher exact	0.46
	1	42 (42.9)	5 (62.5)		

3.4.2 Influence des paramètres techniques pour la détection des lésions croissantes

Par le biais d'un test du Khi-deux en comparant les 2 groupes (paramètre identique/différent (codé 1/0) et concordance/discordance du nombre de lésions croissantes (codé 0/1), aucun des paramètres à savoir le champ magnétique, l'épaisseur de coupe, le centre, le plan d'acquisition et la marque ne constituaient des facteurs de risque significatif responsable d'une discordance dans la détection des lésions croissantes ($p > 0,05$). L'ensemble de ces résultats est notifié de façon exhaustive dans le *tableau 13* ci-dessous.

Tableau 13 : Paramètres techniques de l'IRM et discordance pour la détection d'une lésion croissante

Paramètres techniques		Détection lésions croissantes		Test	pvalue
		Concordance	Discordance		
	Différent	0	1		
	Identique	N=45	N=61		
Champ magnétique	0	6 (13.3)	11 (18.0)	Khi-deux	0,51
	1	39 (86.7)	50 (82.0)		
Plan d'acquisition	0	14 (31.1)	29 (47.5)	Khi-deux	0.14
	1	31 (68.9)	32 (52.5)		
Centre	0	6 (13.3)	13 (21.3)	Khi-deux	0,29
	1	39 (86.7)	48 (78.7)		
Marque	0	3 (6.7)	11 (18.0)	Khi-deux	0,09
	1	42 (93.3)	50 (82.0)		
Épaisseur de coupe	0	22 (48.9)	37 (60.7)	Khi-deux	0.23
	1	23 (51.1)	24 (39.3)		

D. Discussion

1. Rappel de l'objectif et des résultats de l'étude

Évaluer l'évolution des lésions de SEP pour un patient est fondamental sachant que cela constitue un marqueur à la fois diagnostique (critère de dissémination temporelle), pronostique (corrélé à la sévérité du handicap) et thérapeutique (efficacité des traitements).

Cependant même pour des opérateurs entraînés, comparer manuellement les lésions de SEP sur deux IRM peut s'avérer fastidieux et chronophage (notamment en cas de charge lésionnelle importante) et est soumis à un risque de variabilité inter et intra-observateur. L'essor de l'intelligence artificielle et plus particulièrement du *deep learning* pourrait apporter une aide complémentaire dans cette tâche grâce à une reproductibilité potentiellement accrue et une restitution des résultats plus rapide. Par conséquent, l'objet de l'étude était d'analyser l'apport d'un logiciel d'intelligence artificielle (Pixyl ©) dans cette démarche en évaluant sa capacité à détecter une activité inflammatoire radiologique (c'est-à-dire à détecter et décompter des lésions de SEP évolutives c'est-à-dire nouvelle et/ou croissante) en comparaison à l'expert et étudier les étiologies en cas de discordance par relecture des masques de segmentation. Il était établi dans notre étude que la concordance entre le logiciel Pixyl© et l'expert pour la détection d'une activité inflammatoire radiologique est proche de l'aléatoire à cause de la très faible concordance pour la détection des lésions croissantes liés à un taux important de faux positifs imputés le plus souvent à une segmentation erratique par le logiciel pour les lésions majoritairement de topographie périventriculaire, une segmentation des lésions non SEP et à des paramètres techniques différents entre les 2 IRM. Cependant la concordance est forte pour la détection et le dénombrement des nouvelles lésions où le logiciel semble très performant à ce niveau.

2. Segmentation/détection des lésions de SEP et évolution des lésions de SEP

2.1 Segmentation/détection des lésions de SEP

Plusieurs techniques de segmentation des lésions de SEP (non supervisée, supervisée, semi-automatique, *machine learning*...) se sont succédées.

Actuellement, la segmentation automatique (avec la technique du *deep learning* et la technique du CNN) des lésions de SEP en un temps donné (sous forme d'une étude transversale) fait l'objet de nombreuses études où plusieurs algorithmes sont évalués dans le but d'être robuste en offrant des paramètres d'évaluation optimaux se rapprochant davantage du Ground Truth établi par la délimitation manuelle. Les paramètres les plus souvent pris en compte sont le coefficient de Dice (entre 0 et 1) étant un reflet du chevauchement des lésions segmentées (voxel par voxel) par le logiciel et le Gold Standard et le F1-score (entre 0 et 1) évaluant la performance de l'algorithme à détecter une authentique lésion à travers une moyenne harmonique entre la précision (équivalent de la VPP) et le rappel (équivalent de la sensibilité).

L'étude de *Commowick et al.* publiée en 2018 récapitulait les essais de 13 algorithmes différents de segmentation des hypersignaux FLAIR des lésions de SEP tous présentés lors d'une conférence en 2016 de la MICCAI (*Medical Image Computing and Computer Assisted Intervention*) sur une base de données de 53 patients atteints de SEP, issus de 3 centres hospitaliers différents avec 4 machines différentes. Parmi les logiciels testés, le logiciel Pixyl© présentait les meilleurs résultats avec un coefficient de Dice à 0,49 et un F1-score à 0,36. ⁽¹⁰³⁾

Le logiciel a ensuite continué à évoluer pour la détection des lésions inflammatoires de SEP en hypersignal T2 dans l'objectif d'être le plus sensible possible, tout en limitant les faux positifs.

L'étude avait également permis de mettre en évidence que les plus petites lésions sont moins bien détectées et qu'un faible volume total des lésions induit une qualité moindre de segmentation, en revanche il n'existe pas de lien entre le nombre de lésions et la qualité de segmentation. ⁽¹⁰³⁾

Au sein des différents algorithmes proposés, le F1-score est souvent plus faible que le coefficient de Dice car les algorithmes se focalisent davantage sur la délimitation/segmentation des lésions plutôt que leur nombre bien que ce dernier soit plus important en pratique clinique.

2.2 Évolution des lésions de SEP

Comme dit précédemment, l'évolution des lésions de SEP est un reflet de l'activité inflammatoire de la maladie et consiste à détecter des nouvelles lésions ou des lésions croissantes entre 2 imageries successives pour un même patient.

A contrario de la segmentation automatisée des lésions de SEP ayant fait l'objet de plusieurs études, l'analyse automatisée des lésions de SEP a fait l'objet de beaucoup moins de publications. En effet, elle est plus difficile à mettre en place, se basant non plus sur une étude transversale pour une seule imagerie mais sur une étude longitudinale en comparant 2 imageries distinctes, s'ajoutant au fait qu'il y ait beaucoup moins de données d'entrées disponibles pour entraîner l'algorithme en comparaison aux données d'une simple segmentation (bien que plusieurs techniques permettent d'augmenter ce nombre). Comparer indépendamment en deux temps distincts chaque imagerie de façon transversale en utilisant les algorithmes de segmentation mentionnés ci-dessus pour en faire une analyse d'étude longitudinale n'est pas réalisable en pratique, de fait d'autres méthodes doivent être utilisées afin de comparer en un seul temps les imageries ⁽¹⁰⁶⁾. Historiquement, les techniques utilisées étaient par exemple la technique de soustraction (consistant à soustraire les voxels similaires sur 2 imageries révélant les nouvelles lésions), le champ de déformation (ajuster et synchroniser 2 images distinctes) ou encore de façon encore plus ancienne la méthode de tests d'hypothèse statistique pour la détection de changement associée à la méthode de normalisation non linéaire d'intensité (ajustant les pixels de façon non linéaire pour améliorer la compatibilité des images). D'après l'étude de Bosc *et al.* ⁽¹¹⁹⁾, les techniques les plus récentes s'appuient sur le *deep learning*, en particulier le CNN et l'architecture U-net.

La conférence MICCAI ayant eu lieu en 2021 a permis à 30 algorithmes d'évaluer leur performance sur la détection et la segmentation des nouvelles lésions FLAIR de sclérose en plaques (MSSEG-II challenge) sur un total de 15 IRM différentes, le tout en comparaison à un *ground truth* constitué de 4 neuroradiologues expert (un sénior expert est consulté en cas de discordance significative sur un cas parmi les experts) puis une synthèse de leur annotation manuelle est réalisée. La cohorte était issue de l'OFSEP et était constituée de 100 patients répartis en 2 groupes : 40 patients étaient inclus dans les données d'entraînement (dont 29 avaient d'authentiques nouvelles lésions) et 60 dans les données de validation (dont 32 patients avaient

d'authentiques nouvelles lésions)⁽¹²⁰⁾. Les paramètres pour évaluer la robustesse du modèle étaient le coefficient Dice, le F1-score, la sensibilité, la spécificité et la valeur prédictive positive. Au total parmi les différents algorithmes testés (le logiciel Pixyl© n'en faisait pas parti), il en résulte qu'ils sont sensibles pour détecter une nouvelle lésion mais engendre beaucoup de faux positifs⁽¹²¹⁾, constat néanmoins peu retrouvé dans notre étude où le logiciel était à la fois sensible et spécifique avec seulement 5 cas de faux positifs sur les 106 patients pour la détection des nouvelles lésions. L'évolution automatisée des nouvelles lésions est complexe sachant qu'une lésion peut varier ou apparaître de façon subtile entre 2 imageries et ainsi passer inaperçue (exemple de l'apparition d'une lésion de petite taille ; ou d'une nouvelle lésion T2 parmi une charge lésionnelle importante) rendant difficile la mise en évidence de l'apparition ou la croissance d'une lésion. De plus, le manque de *data* concernant l'évolution des lésions de SEP peut altérer la performance de l'algorithme, cependant des stratégies sont utilisées pour y pallier comme le transfert learning où des données utilisées pour un ancien modèle sont ré-utilisées (par exemple les données de segmentation en un temps donnée d'une simple IRM) comme données d'entrée pour le modèle étudié, l'augmentation des données (*ou data augmentation*) consistant à les transformer artificiellement (*par exemple rotation d'une image augmentation du bruit, du contraste..*) ou encore la création de données synthétiques (*par exemple la création d'un masque d'une nouvelle lésion à partir d'une seule IRM*). L'étude *Kamraoui et al.* avait montré que ces techniques augmentant et enrichissant les données améliorent la performance des modèles avec une meilleure segmentation et détection des nouvelles lésions.⁽¹⁰⁶⁾

Tableau 14 : Coefficient Dice et F1 score des algorithmes issus du MSSEG-II challenge⁽¹²²⁾

Algorithmes	Dice	F1
Expert 1	0.629	0.709
Expert 3	0.597	0.637
Expert 2	0.535	0.601
Expert 4	0.459	0.549
Proposed	0.510	0.552
MedICL	0.507	0.500
LaBRI-IQDA	0.500	0.515

Algorithmes	Dice	F1
SNAC	0.485	0.514
LaBRI-D&E	0.472	0.504
NVAUTO	0.460	0.461
LaBRI-lw	0.453	0.463
New New Brain	0.451	0.476
ITU	0.443	0.480
Mediaire-B	0.437	0.474
Mediaire-A	0.432	0.524
Empenn	0.424	0.458
McEwan-IM	0.423	0.495
PVG	0.414	0.430
Neuropoly-1	0.411	0.425
IAMLALB	0.411	0.412
LYLE	0.409	0.443
Neuropoly-2	0.403	0.413
SCAN	0.403	0.431
SCA-SimpleUNet	0.400	0.448
I3M	0.398	0.358
Neuropoly-3	0.379	0.416
The NoCoDers	0.365	0.381
Vicorob	0.357	0.396
HufsAIM	0.346	0.402
CMIC	0.330	0.362
MIAL	0.332	0.354
SCA-withPriors	0.223	0.276
LIT	0.214	0.242
IBBM	0.200	0.158

Une autre étude de *Rival et al.* évaluant le logiciel Pixyl © dans le suivi longitudinal des lésions de SEP avait montré que le logiciel permettait de récupérer 4 nouvelles lésions (soit 6,5% de la totalité des lésions) chez 3 patients différents, non visualisées initialement par les experts dont 2 avaient permis la conversion en SEP

⁽¹²³⁾. Dans le cas de notre étude, l'apport du logiciel par une seconde lecture dans la détection des nouvelles lésions à posteriori par le radiologue expert (pour rappel 3 faux négatifs dans l'étude) n'avait pas été étudié.

Par ailleurs, toujours d'après l'étude de *Rival et al.* ⁽¹²³⁾ en 2022 l'association de la lecture manuelle avec le logiciel positivait davantage de nouvelles lésions et/ou croissantes comparativement à la référence ce qui peut être en lien avec cette autre étude de *Méreaux et al.* en 2022 comparant le logiciel Pixyl© et la technique de FUSION (c'est une superposition des 2 IRM et mise en évidence des nouvelles lésions par un filtre couleur) où la relecture des masques de segmentation Pixyl© avaient révélé 75,9% d'artefacts positifs (liés aux variations de contraste, aux modifications morphologiques ou à un alignement imparfait des coupes). ⁽¹²⁴⁾ Ces constatations n'étaient encore une fois pas retrouvées dans notre étude pour la détection et le nombre des nouvelles lésions mais plutôt pour les lésions croissantes. Ces différences sont peut-être dues à des modifications de l'algorithme du logiciel au cours du temps avec une robustesse améliorée pour la détection et segmentation des nouvelles lésions. En revanche, peu d'études dans la littérature se sont focalisées sur la détection et segmentation des lésions croissantes par les techniques d'intelligence artificielle où il y aurait probablement encore moins de data d'entrées pour alimenter les modèles en question.

Si la valeur pronostique de l'apparition d'une ou de plusieurs nouvelles lésions T2 est indiscutable dans la SEP, la valeur pronostique d'une augmentation de taille d'une ou plusieurs lésions reste à démontrer. En effet, la majorité des études publiées considèrent dans « l'augmentation de la charge lésionnelle », à la fois les nouvelles lésions T2 et les lésions qui grossissent. Une lésion inflammatoire peut croître progressivement dans le cadre d'une lésion chronique active mais croître également dans le cadre d'un processus neurodégénératif au pourtour de la lésion. Les deux phénomènes n'ont ainsi pas la même valeur.

3. Limites de l'étude

Plusieurs paramètres peuvent limiter les résultats de notre étude.

3.1 Boîte noire

Le logiciel Pixyl© comme la plupart des logiciels d'intelligence artificielle peut être assimilé comme une « boîte noire » où il est difficile de connaître l'ensemble du

cheminement situé entre les données d'entrée et le résultat de sortie obtenu. Par exemple dans un modèle de CNN, les rôles des différentes couches (convolutive, pooling, entièrement connectée) sont connus, par exemple chaque couche extrayant une caractéristique en question, mais la façon permettant d'assembler l'ensemble des données extraites afin de produire un résultat donné est obscur. De fait, par exemple la prédiction finale d'une lésion croissante ou nouvelle par le logiciel ne peut pas être expliquée par l'étude des mécanismes intrinsèques ⁽¹¹⁷⁾.

3.2 « IA vs expert » et « IA + expert »

L'étude était une analyse de la concordance entre le logiciel et le radiologue expert les mettant en opposition (modèle « Pixyl© vs expert »). Concernant la détection des nouvelles lésions et le nombre de nouvelles lésions de SEP, le logiciel Pixyl© était performant avec respectivement 8 discordances sur les 106 patients pour la détection de nouvelles lésions (5 faux positifs et 3 faux négatifs) et 25 discordances sur les 106 patients pour le nombre de nouvelles lésions. En revanche, la performance du logiciel était très faible concernant la détection des lésions croissantes et le nombre de celles-ci. Il aurait été intéressant d'effectuer une étude complémentaire sur le modèle « Pixyl© + expert » en levant dans un second temps la lecture aveugle de l'expert par rapport au logiciel permettant d'évaluer l'apport que celui-ci pourrait apporter, par exemple en rattrapant des faux positifs ou négatifs à la fois pour les lésions croissantes et nouvelles lésions (*cf* 2.2 où l'étude *Rival et al.* avait permis au logiciel Pixyl© de rattraper 4 nouvelles lésions non visualisées par le consensus d'expert ⁽¹²³⁾). L'interprétation manuelle de l'évolution des lésions de SEP étant chronophage, le temps passé sur l'interprétation d'une IRM avec l'aide du logiciel Pixyl© aurait pu également être étudié toujours sous ce modèle « Pixyl© + expert ». D'ailleurs *Rival et al* avait montré que le temps de lecture avec l'aide de Pixyl© était diminué (temps médian d'analyse estimé à 196s sans l'aide du logiciel contre 152s avec) ⁽¹²³⁾.

3.3 Lésions non prises en compte par le logiciel

Les lésions du nerf optique n'étaient pas visualisées par le logiciel. Au sein de notre étude, 2 patients présentaient un discret hypersignal FLAIR (mais mieux visible en

séquence DIR) du nerf optique dont 1 était apparu pour un patient (asymptomatique) qui, par ailleurs, ne présentait pas d'autre nouvelle lésion à l'étage encéphalique et 1 était stable et déjà présent pour l'autre patient. Au moment de la réalisation de ce travail, le nerf optique ne faisait pas partie des critères diagnostiques de SEP. Pour le suivi IRM, il n'existe aucune recommandation sur le fait d'inclure ou non les lésions du nerf optique dans les critères d'activité de la maladie. Considérant qu'une nouvelle lésion du nerf optique peut être symptomatique ou asymptomatique mais dans tous les cas engendrer une perte neuroaxonale silencieuse, il apparaît justifié de la comptabiliser dans les signes d'activité inflammatoire de la maladie. Par ailleurs, depuis septembre 2024, le nerf optique est inclus dans les nouveaux critères diagnostiques de la SEP.

Les lésions médullaires ne sont pas encore explorées par le logiciel : dans notre étude, il n'y avait pas de lésions médullaires nouvelles sur les coupes sagittales hautes des séquences 3D FLAIR pour le suivi des patients atteints de SEP sans IRM médullaire réalisée ainsi il n'y avait pas d'impact sur l'étude réalisée (pas de risque de faux négatifs).

Le logiciel Pixyl© ne permettait pas de détecter les lésions décroissantes ce qui pourrait être intéressant pour évaluer l'efficacité éventuelle d'un traitement instauré récemment. Dans notre étude, deux patients présentaient des lésions décroissantes mais s'associant de façon concomitante à des nouvelles lésions ce qui témoigne d'une dynamique lésionnelle inflammatoire et dans tous les cas d'une activité inflammatoire radiologique.

3.4 Concept de lésions croissantes

Contrairement à la définition d'une nouvelle lésion caractérisée par l'apparition d'une lésion de plus de 3 mm, il n'y a pas de véritable consensus sur la définition d'une lésion croissante.

Le décompte des lésions croissantes est soumis à une variabilité inter-observateur entre experts encore plus forte que les nouvelles lésions. Ce constat n'avait pas pu être prouvé dans cette étude composée d'un expert unique mais l'étude d'*Altay et al.* avait montré une concordance proche du hasard (entre 0,00 et 0,14) parmi 4 experts différents sur la quantification des lésions croissantes ⁽¹²⁵⁾, preuve d'un concept alambiqué.

La délimitation et la détection d'une lésion croissante peuvent être complexes surtout en cas de charge lésionnelle importante/lésions confluentes.

Dans notre étude, il y avait beaucoup de faux positifs concernant la détection et le nombre des lésions croissantes. La relecture des masques de segmentation Pixyl © de ces lésions pouvait parfois suggérer une progression inframillimétrique non visible à l'œil nu en première analyse. La croissance d'une lésion peut être lente et il serait intéressant de réévaluer la performance du logiciel sur une nouvelle IRM de suivi à distance afin de confirmer ou infirmer avec plus de certitude la croissance lésionnelle, sans pour autant pouvoir préjuger de son origine (processus inflammatoire chronique ou neurodégénérescence).

4. Solutions envisagées

4.1 Modèle d'incertitude

Les résultats rendus par le logiciel étaient plutôt binaires : lésion stable ou non / lésion croissante ou non / nouvelle lésion ou non. Suite à la concordance très faible avec l'expert concernant les lésions croissantes (détection et décompte), il serait pertinent d'intégrer au sein du logiciel un modèle d'incertitude (*sous forme par exemple d'un nouveau code couleur en cas de doute*) pour les lésions douteuses (*par exemple en cas de charge lésionnelle importante la délimitation lésion croissante / stable est moins nette ou encore en cas doute d'une lésion SEP avec de la substance blanche normale*) car dans l'optique du modèle « Pixyl©+ expert » ceci permettrait au radiologue de se focaliser davantage sur les lésions douteuses plutôt que l'ensemble des lésions comprenant celles faisant consensus. ⁽¹²¹⁾

4.2 Nouveaux paramètres ?

4.2.1 Séquences

Bien qu'il eût été établi au sein de notre étude que la variabilité des paramètres techniques n'influait pas sur la détection des nouvelles lésions et croissantes, il serait tout de même intéressant d'alimenter l'algorithme du logiciel Pixyl© avec des nouvelles séquences comme des séquences DIR, T1 voire SWI (signe de la veine centrale) permettant d'améliorer la détection des lésions / d'affiner la segmentation

des lésions notamment croissantes (source de discordance) sachant que la robustesse du « ground truth » étant étroitement liée à la performance du logiciel sachant que les données de ce dernier sont labellisées par des experts.

L'étude de *McKinley et al.* montrait que 28% des nouvelles lésions/croissantes sur des séquences FLAIR pouvaient être manquées (par l'expert) et l'interprétation à l'aide de la séquence DIR pouvait en améliorer la détection ⁽¹²⁶⁾.

D'autre part, l'étude de *Valencia et al.* concluait que l'ajout d'une séquence T1 synthétique à la séquence FLAIR dans les données d'entrées d'un modèle le rendait plus performant pour la détection des nouvelles lésions ⁽¹²⁷⁾.

4.2.2 Champ magnétique

L'IRM 7 tesla permet de détecter plus facilement les lésions de SEP de petite taille, les lésions de topographie corticale grâce à une plus grande résolution spatiale et le signe de la veine centrale. Cela pourrait être pertinent pour l'expert et les modèles d'IA d'intégrer ces IRM à très haut champ magnétique ⁽¹²⁸⁾.

E. Conclusion

En conclusion, notre travail met en avant la très bonne concordance entre le logiciel et l'expert pour la détection des nouvelles lésions de SEP. Notre travail met en avant également une discordance importante dans l'appréciation de la croissance lésionnelle. Le logiciel ne rapporte pas de diminution de taille de lésion mais démontre une tendance nette à surestimer le nombre de lésion croissante en comparaison à l'expert. Indéniablement, le logiciel peut aider le radiologue qui ne dispose pas d'outil de fusion d'image performant, dans la détection de nouvelle(s) lésion(s). Son usage au quotidien pourrait guider l'expert dans son analyse radiologique des lésions potentiellement croissantes, en pointant les lésions potentiellement croissantes, que l'expert confirmerait ou non.

Le développement de logiciel IA de segmentation automatique ne va faire que se poursuivre dans le futur. Il s'agira de savoir les évaluer, les faire progresser pour les rendre utiles au suivi des patients en routine clinique.

Références

1. Haki M, AL-Biati HA, Al-Tameemi ZS, Ali IS, Al-hussaniy HA. Review of multiple sclerosis: Epidemiology, etiology, pathophysiology, and treatment. *Medicine*. 23 févr 2024;103(8):e37297.
2. Herthum H, Hetzer S, Scheel M, Shahryari M, Braun J, Paul F, et al. In vivo stiffness of multiple sclerosis lesions is similar to that of normal-appearing white matter. *Acta Biomater*. 15 janv 2022;138:410-21.
3. Sandi D, Kokas Z, Biernacki T, Bencsik K, Klivényi P, Vécsei L. Proteomics in Multiple Sclerosis: The Perspective of the Clinician. *Int J Mol Sci*. 5 mai 2022;23(9):5162.
4. Doshi A, Chataway J. Multiple sclerosis, a treatable disease. *Clin Med (Lond)*. déc 2016;16(Suppl 6):s53-9.
5. Dobson R, Giovannoni G. Multiple sclerosis – a review. *European Journal of Neurology*. 2019;26(1):27-40.
6. Deiva K. Pediatric onset multiple sclerosis. *Revue Neurologique*. 1 janv 2020;176(1):30-6.
7. Pugliatti M, Rosati G, Carton H, Riise T, Drulovic J, Vécsei L, et al. The epidemiology of multiple sclerosis in Europe. *European Journal of Neurology*. 2006;13(7):700-22.
8. Lubetzki C. Physiopathologie de la sclérose en plaques : actualités. *Bulletin de l'Académie Nationale de Médecine*. 1 juin 2022;206(6):721-6.
9. <https://www.elsevier.com/fr-fr/connect/anatomie-pathologique-de-la-sclerose-en-plaques-et-des-maladies-apparentees>
10. Pons V, Rivest S. Beneficial Roles of Microglia and Growth Factors in MS, a Brief Review. *Front Cell Neurosci*. 23 sept 2020;14:284.
11. Maciak K, Dziedzic A, Saluk J. Remyelination in multiple sclerosis from the miRNA perspective. *Front Mol Neurosci*. 1 juin 2023;16:1199313.
12. Biller A, Pflugmann I, Badde S, Diem R, Wildemann B, Nagel AM, et al. Sodium MRI in Multiple Sclerosis is Compatible with Intracellular Sodium Accumulation and Inflammation-Induced Hyper-Cellularity of Acute Brain Lesions. *Sci Rep*. 10 août 2016;6:31269.
13. Nielsen NM, Westergaard T, Rostgaard K, Frisch M, Hjalgrim H, Wohlfahrt J, et al. Familial Risk of Multiple Sclerosis: A Nationwide Cohort Study. *American Journal of Epidemiology*. 15 oct 2005;162(8):774-8.
14. O'Gorman C, Lin R, Stankovich J, Broadley SA. Modelling Genetic Susceptibility to Multiple Sclerosis with Family Data. *Neuroepidemiology*. 11 oct 2012;40(1):1-12.
15. Ma Q, Shams H, Didonna A, Baranzini SE, Cree BAC, Hauser SL, et al. Integration of epigenetic and genetic profiles identifies multiple sclerosis disease-critical cell types and genes. *Commun Biol*. 30 mars 2023;6(1):1-10.

16. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet.* oct 2015;47(10):1107-13.
17. Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. *Nat Commun.* 19 juin 2018;9(1):2397.
18. Rhead B, Bäärnhielm M, Gianfrancesco M, Mok A, Shao X, Quach H, et al. Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. *Neurology Genetics.* oct 2016;2(5):e97.
19. Soldan SS, Lieberman PM. Epstein–Barr virus and multiple sclerosis. *Nat Rev Microbiol.* 2023;21(1):51-64.
20. Lanz TV, Brewer RC, Ho PP, Moon JS, Jude KM, Fernandez D, et al. Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature.* mars 2022;603(7900):321-7.
21. Wingerchuk DM. Smoking: effects on multiple sclerosis susceptibility and disease progression. *Ther Adv Neurol Disord.* janv 2012;5(1):13-22.
22. Wang Z, Xie J, Wu C, Xiao G. Correlation Between Smoking and Passive Smoking with Multiple Sclerosis and the Underlying Molecular Mechanisms. *Med Sci Monit.* 31 janv 2019;25:893-902.
23. Rejdak K, Eikelenboom MJ, Petzold A, Thompson EJ, Stelmasiak Z, Lazeron RHC, et al. CSF nitric oxide metabolites are associated with activity and progression of multiple sclerosis. *Neurology.* 26 oct 2004;63(8):1439-45.
24. Munger KL, Bentzen J, Laursen B, Stenager E, Koch-Henriksen N, Sørensen TIA, et al. Childhood body mass index and multiple sclerosis risk: a long-term cohort study. *Mult Scler.* sept 2013;19(10):1323-9.
25. Correale J, Marrodan M. Multiple sclerosis and obesity: The role of adipokines. *Front Immunol.* 15 nov 2022;13:1038393.
26. Tremlett H, Waubant E. The multiple sclerosis microbiome? *Ann Transl Med.* févr 2017;5(3):53.
27. Cantoni C, Lin Q, Dorsett Y, Ghezzi L, Liu Z, Pan Y, et al. Alterations of host-gut microbiome interactions in multiple sclerosis. *EBioMedicine.* 27 janv 2022;76:103798.
28. Collège des enseignants de neurologie. *Neurologie.* Elsevier Masson ; 2023.
29. Ford H. Clinical presentation and diagnosis of multiple sclerosis. *Clin Med (Lond).* juill 2020;20(4):380-3.
30. Miller D, Weinshenker B, Filippi M, Banwell B, Cohen J, Freedman M, et al. Differential diagnosis of suspected multiple sclerosis: a consensus approach. *Mult Scler.* nov 2008;14(9):1157-74.
31. Ömerhoca S, Akkaş SY, İçen NK. Multiple Sclerosis: Diagnosis and Differential Diagnosis. *Noro Psikiyatr Ars.* 2018;55(Suppl 1):S1-9.

32. [40.https://radiologyassistant.nl/neuroradiology/multiple-sclerosis/diagnosis-and-differential-diagnosis-3](https://radiologyassistant.nl/neuroradiology/multiple-sclerosis/diagnosis-and-differential-diagnosis-3)
33. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sorensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology*. 2014 Jul 15;83(3):278-86.
34. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology*. avr 1996;46(4):907-11.
35. 35 Lebrun-Frénay C, Okuda DT, Siva A, Landes-Chateau C, Azevedo CJ, Mondot L, et al. The radiologically isolated syndrome: revised diagnostic criteria. *Brain*. 2 mars 2023;146(8):3431-43.
36. Klineova S, Lublin FD. Clinical Course of Multiple Sclerosis. *Cold Spring Harb Perspect Med*. sept 2018;8(9):a028928.
- 37.ECTRIMS 2024 HIGHLIGHTS – THURSDAY, SEPTEMBER 19, 2024 | NeuroSens [Internet]. [cité 2 oct 2024]. Disponible sur: <https://neuro-sens.com/ectrims-2024-highlights-thursday-september-19-2024/>
38. Jankowska A, Chwojncki K, Szurowska E. The diagnosis of multiple sclerosis: what has changed in diagnostic criteria? *Pol J Radiol*. 12 déc 2023;88:e574-81.
39. Vukusic S. Révision 2017 des critères diagnostiques de McDonald pour la Sclérose en Plaques ; 2017.
40. OFSEP - Descriptif de la cohorte [Internet]. [cité 25 août 2024]. Disponible sur: <https://www.ofsep.org/fr/la-cohorte-ofsep/descriptif-de-la-cohorte>
41. New OFSEP recommendations for MRI assessment of multiple sclerosis patients: Special consideration for gadolinium deposition and frequent acquisitions. *Journal of Neuroradiology*. 1 juin 2020;47(4):250-8.
42. Cotton F, Kremer S, Hannoun S, Vukusic S, Dousset V. OFSEP, a nationwide cohort of people with multiple sclerosis: Consensus minimal MRI protocol. *Journal of Neuroradiology*. 1 juin 2015;42(3):133-40.
43. Abdrabou A. Radiopaedia. [cité 25 août 2024]. Diffusion tensor imaging and fiber tractography | Radiology Reference Article | Radiopaedia.org. Disponible sur: <https://radiopaedia.org/articles/diffusion-tensor-imaging-and-fiber-tractography-1?lang=us>
44. Heckova E, Dal-Bianco A, Strasser B, Hangel GJ, Lipka A, Motyka S, et al. Extensive Brain Pathologic Alterations Detected with 7.0-T MR Spectroscopic Imaging Associated with Disability in Multiple Sclerosis. *Radiology* [Internet]. 4 janv 2022 [cité 21 août 2024]; Disponible sur: <https://pubs.rsna.org/doi/10.1148/radiol.210614>
45. Hemond CC, Bakshi R. Magnetic Resonance Imaging in Multiple Sclerosis. *Cold Spring Harb Perspect Med*. mai 2018;8(5):a028969.

46. Filippi M, Preziosa P, Banwell BL, Barkhof F, Ciccarelli O, De Stefano N, et al. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Lancet Neurol*. 2019 Aug;18(8):747-758.
47. Neema M, Guss ZD, Stankiewicz JM, Arora A, Healy BC, Bakshi R. Normal findings on brain fluid-attenuated inversion recovery MR images at 3T. *AJNR Am J Neuroradiol*. 2009 Jun;30(6):1275-1280. doi: 10.3174/ajnr.A1553.
48. Darakdjian M, Chaves H, Hernandez J, Cejas C. MRI pattern in acute optic neuritis: Comparing multiple sclerosis, NMO and MOGAD. *Neuroradiol J*. 2023 Jun;36(3):267-272. doi: 10.1177/19714009221124308. Epub 2022 Sep 3.
49. Kim HJ, Paul F, Lana-Peixoto MA, Tenenbaum S, Asgari N, Palace J, et al. MRI characteristics of neuromyelitis optica spectrum disorder. *Neurology*. 17 mars 2015;84(11):1165-73.
50. Bender B, Heine C, Danz S, Bischof F, Reimann K, Bender M, et al. Diffusion restriction of the optic nerve in patients with acute visual deficit. *Journal of Magnetic Resonance Imaging*. 2014;40(2):334-40.
51. Kolb H, Al-Louzi O, Beck ES, Sati P, Absinta M, Reich DS. From pathology to MRI and back: Clinically relevant biomarkers of multiple sclerosis lesions. *NeuroImage: Clinical*. 1 janv 2022;36:103194.
52. Al-Zandi SH, Fayadh NAH, Al-Waely NKN. Central vein sign detected by SWI at 3 T MRI as a discriminator between multiple sclerosis and leukoaraiosis. *The Egyptian Journal of Radiology and Nuclear Medicine*. 1 mars 2018;49(1):158-64.
53. Suthiphosuwat S, Sati P, Absinta M, Guenette M, Reich DS, Bharatha A, Oh J. Paramagnetic Rim Sign in Radiologically Isolated Syndrome. *AJNR Am J Neuroradiol*. 2020 Mar;41(3):454-458. doi: 10.3174/ajnr.A6421.
54. 60 Lunde HMB, Assmus J, Myhr KM, Bø L, Grytten N. Survival and cause of death in multiple sclerosis: a 60-year longitudinal population study. *J Neurol Neurosurg Psychiatry*. août 2017;88(8):621-5.
55. National Multiple Sclerosis Society [Internet]. [cité 25 août 2024]. Empowering people affected by MS to live their best lives. Disponible sur: <https://www.nationalmssociety.org>
56. Demir S. Expanded Disability Status Scale (EDSS) in Multiple Sclerosis. *csmedj*. 28 déc 2022;2(3):82-9.
57. MS Progression Chart: Stages of MS, Disability Scale, and More [Internet]. [cité 28 août 2024]. Disponible sur: <https://www.healthline.com/health/progressing-ms/ms-progression-chart>
58. Australian Journal of General Practice [Internet]. [cité 25 août 2024]. Multiple sclerosis diagnosis therapy and prognosis. Disponible sur: <https://www1.racgp.org.au/ajgp/2022/april/multiple-sclerosis-diagnosis-therapy-and-prognosis>
59. Sormani MP, Rio J, Tintorè M, Signori A, Li D, Cornelisse P, et al. Scoring treatment response in patients with relapsing multiple sclerosis. *Mult Scler*. avr 2013;19(5):605-12.

60. Briganti G, Le Moine O. Artificial Intelligence in Medicine: Today and Tomorrow. *Front Med (Lausanne)*. 5 févr 2020;7:27.
61. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>
62. Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR Data Science Institute Artificial Intelligence Survey. *J Am Coll Radiol*. août 2021;18(8):1153-9.
63. European Society of Radiology (ESR). Current practical experience with artificial intelligence in clinical radiology: a survey of the European Society of Radiology. *Insights Imaging*. 21 juin 2022;13(1):107.
64. Mello-Thoms C, Mello CAB. Clinical applications of artificial intelligence in radiology. *Br J Radiol*. oct 2023;96(1150):20221031.
65. Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health*. mars 2020;2(3):e138-48.
66. Ahn JS, Ebrahimian S, McDermott S, Lee S, Naccarato L, Di Capua JF, et al. Association of Artificial Intelligence-Aided Chest Radiograph Interpretation With Reader Performance and Efficiency. *JAMA Netw Open*. 1 août 2022;5(8):e2229289.
67. Regnard NE, Lanseur B, Ventre J, Ducarouge A, Clovis L, Lassalle L, et al. Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays. *Eur J Radiol*. sept 2022;154:110447.
68. Fasen BACM, Berendsen RCM, Kwee RM. Artificial intelligence software for diagnosing intracranial arterial occlusion in patients with acute ischemic stroke. *Neuroradiology*. août 2022;64(8):1579-83.
69. Hardy M, Harvey H. Artificial intelligence in diagnostic imaging: impact on the radiography profession. *Br J Radiol*. mars 2020;93(1108):20190840.
70. Vasey B, Ursprung S, Beddoe B, Taylor EH, Marlow N, Bilbro N, et al. Association of Clinician Diagnostic Performance With Machine Learning-Based Decision Support Systems: A Systematic Review. *JAMA Netw Open*. 1 mars 2021;4(3):e211276.
71. TURING AM. I.—Computing Machinery and Intelligence. *Mind*. 1 oct 1950;LIX(236):433-60.
72. Neri E, de Souza N, Brady A, Bayarri AA, Becker CD, Coppola F, et al. What the radiologist should know about artificial intelligence – an ESR white paper. *Insights into Imaging*. 4 avr 2019;10(1):44.
73. A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE [Internet]. [cité 1 sept 2024]. Disponible sur: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

74. Sorantin E, Grasser MG, Hemmelmayer A, Tschauner S, Hrzic F, Weiss V, et al. The augmented radiologist: artificial intelligence in the practice of radiology. *Pediatr Radiol*. 2022;52(11):2074-86.
75. Bonacchi R, Filippi M, Rocca MA. Role of artificial intelligence in MS clinical practice. *Neuroimage Clin*. 28 mai 2022;35:103065.
76. Data Analytics Post [Internet]. [cité 31 août 2024]. SVM. Disponible sur : <https://dataanalyticspost.com/Lexique/svm/>
77. Zhang S, Li X, Zong M, Zhu X, Wang R. Efficient kNN Classification With Different Numbers of Nearest Neighbors. *IEEE Transactions on Neural Networks and Learning Systems*. mai 2018;29(5):1774-85.
78. Bush WS, Thornton-Wells TA, Ritchie MD. Association Rule Discovery Has the Ability to Model Complex Genetic Effects. *IEEE Symp Comput Intell Data Min*. 1 mars 2007;2007:624-9.
79. Built In [Internet]. [cité 1 sept 2024]. C-Means Clustering Explained. Disponible sur: <https://builtin.com/data-science/c-means>
80. Kimes PK, Liu Y, Hayes DN, Marron JS. Statistical Significance for Hierarchical Clustering. *Biometrics*. sept 2017;73(3):811-21.
81. Markov Random Field Optimisation [Internet]. [cité 1 sept 2024]. Disponible sur: https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0809/ORCHARD/
82. Yuan C, Yang H. Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*. juin 2019;2(2):226-35.
83. Moraru L, Moldovanu S, Dimitrievici LT, Dey N, Ashour AS, Shi F, et al. Gaussian mixture model for texture characterization with application to brain DTI images. *J Adv Res*. 4 janv 2019;16:15-23.
84. Singular value decomposition. In: Wikipedia [Internet]. 2024 [cité 1 sept 2024]. Disponible sur: https://en.wikipedia.org/w/index.php?title=Singular_value_decomposition&oldid=1238419013
85. Zhu X (Jerry). Semi-Supervised Learning Literature Survey [Internet]. University of Wisconsin-Madison Department of Computer Sciences; 2005 [cité 31 août 2024]. Disponible sur: <https://minds.wisconsin.edu/handle/1793/60444>
86. Xu L, Zhu S, Wen N. Deep reinforcement learning and its applications in medical imaging and radiation therapy: a survey. *Phys Med Biol*. 11 nov 2022;67(22).
87. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput*. juill 2006;18(7):1527-54.
88. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. août 2018;9(4):611-29.

89. Datascientest. Convolutional Neural Networks. Datascientest. [date de publication non spécifiée]. Disponible à : <https://datascientest.com/convolutional-neural-network>. Consulté le : 23 septembre 2024.
90. Gautherot M, Yepremian S, Bretzner M, Jacques T, Hutt A, Pruvo JP, et al. 15 minutes pour comprendre et évaluer un logiciel d'intelligence artificielle appliquée à l'imagerie médicale. Journal d'imagerie diagnostique et interventionnelle. 1 juin 2021;4(3):167-71.
91. Kassel R. Hyperparamètres : Qu'est-ce que c'est ? À quoi ça sert ? [Internet]. Formation Data Science | DataScientest.com. 2023 [cité 23 sept 2024]. Disponible sur: <https://datascientest.com/hyperparametres-tout-savoir>
92. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. Sci Rep. 13 mars 2024;14:6086.
93. Nabizadeh F, Ramezannezhad E, Kargar A, Sharafi AM, Ghaderi A. Diagnostic performance of artificial intelligence in multiple sclerosis: a systematic review and meta-analysis. Neurol Sci. févr 2023;44(2):499-517.
94. Duong MT, Rauschecker AM, Mohan S. Diverse Applications of Artificial Intelligence in Neuroradiology. Neuroimaging Clin N Am. nov 2020;30(4):505-16.
95. Amin M, Martínez-Heras E, Ontaneda D, Prados Carrasco F. Artificial Intelligence and Multiple Sclerosis. Curr Neurol Neurosci Rep. 2024;24(8):233-43.
96. Zeng C, Gu L, Liu Z, Zhao S. Review of Deep Learning Approaches for the Segmentation of Multiple Sclerosis Lesions on Brain MRI. Front Neuroinform [Internet]. 20 nov 2020 [cité 26 sept 2024];14. Disponible sur: <https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2020.610967/full>
97. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik. 1 mai 2019;29(2):102-27.
98. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, éditeurs. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing; 2015. p. 234-41.
99. Yin XX, Sun L, Fu Y, Lu R, Zhang Y. U-Net-Based Medical Image Segmentation. J Healthc Eng. 15 avr 2022;2022:4189781.
100. La Rosa F, Abdulkadir A, Fartaria MJ, Rahmanzadeh R, Lu PJ, Galbusera R, et al. Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE. Neuroimage Clin. 30 juin 2020;27:102335.
101. Wahlig SG, Nedelec P, Weiss DA, Rudie JD, Sugrue LP, Rauschecker AM. 3D U-Net for automated detection of multiple sclerosis lesions: utility of transfer learning from other pathologies. Front Neurosci. 27 oct 2023;17:1188336.

102. Gros C, De Leener B, Badji A, Maranzano J, Eden D, Dupont SM, et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage*. 1 janv 2019;184:901-15.
103. Commowick O, Istace A, Kain M, Laurent B, Leray F, Simon M, et al. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. *Sci Rep*. 12 sept 2018;8:13650.
104. Moraal B, Meier DS, Poppe PA, Geurts JJG, Vrenken H, Jonker WMA, et al. Subtraction MR Images in a Multiple Sclerosis Multicenter Clinical Trial Setting. *Radiology*. févr 2009;250(2):506-14
105. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. févr 2021;18(2):203-11.
106. Kamraoui RA, Mansencal B, Manjon JV, Coupé P. Longitudinal detection of new MS lesions using deep learning. *Front Neuroimaging* [Internet]. 25 août 2022 [cité 26 sept 2024];1. Disponible sur: <https://www.frontiersin.org/journals/neuroimaging/articles/10.3389/fnimg.2022.948235/full>
107. Eshaghi A, Wottschel V, Cortese R, Calabrese M, Sahraian MA, Thompson AJ, et al. Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology*. 6 déc 2016;87(23):2463-70.
108. Kim H, Lee Y, Kim YH, Lim YM, Lee JS, Woo J, et al. Deep Learning-Based Method to Differentiate Neuromyelitis Optica Spectrum Disorder From Multiple Sclerosis. *Front Neurol*. 2020;11:599042.
109. Maggi P, Fartaria MJ, Jorge J, La Rosa F, Absinta M, Sati P, et al. CVSnet: A machine learning approach for automated central vein sign assessment in multiple sclerosis. *NMR Biomed*. mai 2020;33(5):e4283.
110. Cavaliere C, Vilades E, Alonso-Rodríguez MC, Rodrigo MJ, Pablo LE, Miguel JM, et al. Computer-Aided Diagnosis of Multiple Sclerosis Using a Support Vector Machine and Optical Coherence Tomography Features. *Sensors (Basel)*. 3 déc 2019;19(23):5323.
111. Goyal M, Khanna D, Rana PS, Khaibullin T, Martynova E, Rizvanov AA, et al. Computational Intelligence Technique for Prediction of Multiple Sclerosis Based on Serum Cytokines. *Front Neurol*. 18 juill 2019;10:781.
112. Bendfeldt K, Taschler B, Gaetano L, Madoerin P, Kuster P, Mueller-Lenke N, et al. MRI-based prediction of conversion from clinically isolated syndrome to clinically definite multiple sclerosis using SVM and lesion geometry. *Brain Imaging Behav*. 28 août 2018;13(5):1361-74.
113. Pinto MF, Oliveira H, Batista S, Cruz L, Pinto M, Correia I, et al. Prediction of disease progression and outcomes in multiple sclerosis with machine learning. *Sci Rep*. 3 déc 2020;10:21038.
114. Seccia R, Gammelli D, Dominici F, Romano S, Landi AC, Salvetti M, et al. Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis. *PLoS One*. 2020;15(3):e0230219.

115. Roca P, Attys A, Colas L, Tucholka A, Rubini P, Cackowski S, et al. Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. *Diagn Interv Imaging*. déc 2020;101(12):795-802.
116. High-dimensional detection of imaging response to treatment in multiple sclerosis - PMC [Internet]. [cité 4 sept 2024]. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6556513/>
117. Pixyl : quand l'IA décrypte l'imagerie médicale | Inria [Internet]. 2021 [cité 26 sept 2024]. Disponible sur: <https://www.inria.fr/fr/intelligence-artificielle-medecine-imagerie-pixyl>
118. Pixyl.Neuro.MS - Pixyl [Internet]. [cité 4 sept 2024]. Disponible sur: <https://pixyl.ai/multiple-sclerosis/>
119. Bosc M, Heitz F, Armspach JP, Namer I, Gounot D, Rumbach L. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. *Neuroimage*. oct 2003;20(2):643-56.
120. Data [Internet]. [cité 30 sept 2024]. Disponible sur: <https://portal.fli-iam.irisa.fr/msseg-2/data/>
121. Commowick O, Combès B, Cervenansky F, Dojat M. Editorial: Automatic methods for multiple sclerosis new lesions detection and segmentation. *Front Neurosci* [Internet]. 14 mars 2023 [cité 24 sept 2024];17. Disponible sur: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2023.1176625/full>
122. Basaran BD, Matthews PM, Bai W. New lesion segmentation for multiple sclerosis brain images with imaging and lesion-aware augmentation. *Front Neurosci* [Internet]. 21 oct 2022 [cité 30 sept 2024];16. Disponible sur: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2022.1007453/full>
123. Rival M. Évaluation de l'apport d'un outil de segmentation automatique pour le suivi longitudinal des lésions de sclérose en plaques en IRM cérébrale. 10 oct 2022;121.
124. Méreaux JL. Comparaison d'une analyse automatique d'IRM cérébrale par PIXYL versus imagerie de fusion dans le suivi de la sclérose en plaques. 18 oct 2021;96.
125. Altay EE, Fisher E, Jones SE, Hara-Cleaver C, Lee JC, Rudick RA. Reliability of Classifying Multiple Sclerosis Disease Activity Using Magnetic Resonance Imaging in a Multiple Sclerosis Clinic. *JAMA Neurol*. 1 mars 2013;70(3):338-44.
126. McKinley R, Wepfer R, Grunder L, Aschwanden F, Fischer T, Friedli C, et al. Automatic detection of lesion load change in Multiple Sclerosis using convolutional neural networks with segmentation confidence. *Neuroimage Clin*. 9 déc 2019;25:102104.
127. Valencia L, Clérigues A, Valverde S, Salem M, Oliver A, Rovira À, et al. Evaluating the use of synthetic T1-w images in new T2 lesion detection in multiple sclerosis. *Front Neurosci* [Internet]. 29 sept 2022 [cité 1 oct 2024];16. Disponible sur: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2022.954662/full>
128. Bruschi N, Boffa G, Inglese M. Ultra-high-field 7-T MRI in multiple sclerosis and other demyelinating diseases: from pathology to clinical practice. *Eur Radiol Exp*. 22 oct 2020;4:59.

AUTEURE : Nom : KOUEMO

Prénom : Sheena-Chloé

Date de soutenance : 15 octobre 2024

Titre de la thèse : Évaluation de la performance d'un outil de segmentation automatisée dans le suivi des lésions de sclérose en plaques par IRM au CHU de Lille

Thèse - Médecine – Lille 2024

Cadre de classement : Radiologie et Imagerie Médicale

DES + FST/option : Radiologie et Imagerie Médicale

Mots-clés : sclérose en plaques, IRM, activité inflammatoire radiologique, intelligence artificielle, apprentissage profond, algorithme, segmentation

Résumé :

Contexte : Suivre l'évolution des lésions de SEP par IRM cérébrale est primordial pour détecter la présence d'une activité inflammatoire constituant un facteur pronostic et thérapeutique de la maladie. Le suivi par le radiologue peut être chronophage et peu reproductible. C'est pour ces raisons qu'il existe de plus en plus d'outils de segmentation automatisée utilisant la technique du *deep learning* visant à pallier ces problèmes. L'objectif principal de cette étude était d'étudier le logiciel Pixyl© récemment utilisé au CHU de Lille afin d'évaluer sa performance à détecter une activité inflammatoire radiologique de la SEP en comparaison à un radiologue-expert. Les objectifs secondaires étaient de déterminer la concordance sur la détection et le nombre des nouvelles lésions et lésions croissantes puis d'analyser les étiologies pour chaque discordance de ces valeurs en étudiant les masques de segmentation du logiciel.

Méthode : Il s'agit d'une étude rétrospective descriptive et analytique d'une cohorte de 106 patients (2016 – 2024). Un test de concordance entre le radiologue expert et le logiciel est réalisé (coefficient Kappa) sur la détection d'une activité inflammatoire radiologique et le nombre de lésions croissantes et nouvelles. Puis une description analytique des étiologies pour chaque cas de discordance avec la recherche d'un lien entre la variabilité des paramètres techniques IRM et la discordance du nombre de nouvelles lésions et croissantes via respectivement un test de Fisher exact et Khi-deux.

Résultats : La concordance expert/logiciel pour la détection d'une activité inflammatoire est faible avec un kappa à 0,24 [0,12 ; 0,37] due à une faible sensibilité pour la détection et le décompte des lésions croissantes avec des concordances proche de l'aléatoire : kappa tous 2 à 0,06 (respectivement [-0,04 ; 0,16] et [-0,01 ; 0,13]). En revanche, la concordance expert/logiciel pour la détection et le nombre de nouvelles lésions est forte avec un kappa tous deux à 0,81 (respectivement [0,69 ; 0,94] et 0,81 [0,72 ; 0,91]). Les trois étiologies les plus fréquentes expliquant les 25 discordances entre le logiciel et l'expert pour le décompte des nouvelles lésions sont des lésions étiquetées à tort comme nouvelles alors qu'elles sont croissantes (lésions de petites tailles et passées inaperçues par le logiciel sur une IRM antérieure) (8 cas) ; un oubli du logiciel à cause d'une charge lésionnelle importante rendant difficile la visualisation d'une nouvelle lésion (8 cas) et à cause de la variabilité des paramètres techniques (8 cas). L'étiologie la plus fréquente expliquant les 74 discordances pour le nombre de lésions croissantes est une erreur de segmentation pour les lésions de topographie majoritairement périventriculaire étiquetées comme croissantes alors qu'elles sont stables (66 cas). Enfin la variabilité des paramètres techniques entre 2 IRM ne sont pas des facteurs significatifs de discordance sur ces nombres ($p > 0,05$)

Conclusion : Le logiciel Pixyl© est robuste pour la détection et le décompte des nouvelles lésions en raison d'une concordance satisfaisante avec le radiologue-expert. Concernant la détection des lésions croissantes, le logiciel semble peu performant. La nature purement inflammatoire de ces lésions croissantes reste toutefois à préciser. Pour la détection des nouvelles lésions, le logiciel Pixyl© peut être utile aux radiologues qui ne disposent pas d'un logiciel de fusion performant.

Composition du Jury :

Président : Monsieur Le Professeur Grégory KUCHCINSKI

Assesseurs : Madame La Professeure Hélène ZÉPHIR

Monsieur Le Docteur Arnaud BEAUGUITTE

Directeur de thèse : Monsieur Le Docteur Olivier OUTTERYCK