



Université Claude Bernard  Lyon 1

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de :

l'Université Claude Bernard Lyon 1

Ecole Doctorale ED205

Interdisciplinaire Science-Santé

Spécialité de doctorat : Traitement du signal et des images

Discipline : Ingénierie biomédicale, biotechnologie

Soutenue publiquement le 13/10/2020, par :

Noëlie Debs

Prédiction de l'issue lésionnelle et clinique de l'accident vasculaire cérébral par approche d'intelligence artificielle

Devant le jury composé de :

Stéphanie Allassonnière

Professeure, Université Paris-Descartes

Caroline Petitjean

Maître de Conférences, Université de Rouen

Emmanuel Barbier

Directeur de recherche, INSERM

Yves Berthezène

PU-PH, Hospices Civils de Lyon

Mikael Mazighi

PU-PH, Université Paris-Diderot

Florence Zara

Maître de Conférences, Université Claude Bernard Lyon 1

Tae-Hee Cho

PU-PH, Hospices Civils de Lyon

Carole Frindel

Maître de Conférences, INSA Lyon

David Rousseau

Professeur, Université d'Angers

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

Examinatrice

Directeur de thèse

Co-directrice de thèse

Co-directeur de thèse

UNIVERSITÉ CLAUDE BERNARD - LYON 1

Président de l'université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-président du Conseil d'Administration	M. Didier REVEL
Vice-président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Damien VERHAEGHE

COMPOSANTE SANTÉ

Faculté de Médecine Lyon Est – Claude Bernard	Doyen : M. Gilles RODE
Faculté de Médecine et Maïeutique Lyon Sud Charles Mérieux	Doyenne : Mme Carole BURILLON
UFR d'Odontologie	Doyenne : Mme Dominique SEUX
Institut des Sciences Pharmaceutiques et Biologiques	Directrice : Mme Christine VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. Xavier PERROT
Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT

COMPOSANTE ET DÉPARTEMENTS DE SCIENCES ET TECHNOLOGIE

UFR Biosciences	Directrice : Mme Kathrin GIESELER
Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur : M. Marc BUFFAT
UFR - Faculté des Sciences	Admin. provisoire : M. Bruno ANDRIOLETTI
UFR (STAPS)	Directeur : M. Yannick VANPOULLE
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Ecole Polytechnique Universitaire Lyon 1	Directeur : Emmanuel PERRIN
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Institut de Science Financière et d'Assurances ESPE	Directeur : M. Nicolas LEBOISNE
	Admin. Provisoire : M. Pierre CHAREYRON

Résumé

L'accident vasculaire cérébral (AVC) ischémique constitue la première cause de handicap acquis et la troisième cause de décès dans les pays industrialisés. La prédiction de l'évolution des lésions observées en phase aiguë est un challenge clinique. Actuellement, cette question est abordée au moyen de l'imagerie par résonance magnétique (IRM). Deux modalités sont principalement utilisées : l'imagerie de diffusion qui montre l'étendue de l'œdème cytotoxique et l'imagerie dynamique de perfusion qui donne accès à l'état hémodynamique des tissus. L'imagerie de diffusion est connue pour être l'imagerie qui porte le potentiel prédictif le plus important pour détecter la zone d'infarctus irréversible. Une question ouverte est l'apport de l'imagerie de perfusion en phase aiguë pour réaliser cette prédiction. Plusieurs approches d'apprentissage automatique ont été testées et se sont avérées efficaces, mais les résultats témoignent que le problème est loin d'être résolu : la tâche de prédiction demeure complexe, dans la mesure où la lésion visible en phase aiguë évolue jusqu'à un mois après.

La thèse s'articule autour des deux questions suivantes : (1) quel est la valeur de l'imagerie de perfusion pour la prédiction de l'AVC ischémique? (2) comment fusionner l'imagerie de perfusion et de diffusion au sein de modèle d'apprentissage?

La première partie de la thèse répond à la question (1) en deux sous-parties : dans une première sous-partie, on démontrera l'intérêt d'apprendre directement sur l'imagerie de perfusion brute plutôt que sur des cartes paramétriques extraites. On discutera alors de différentes méthodes d'encodage de la perfusion brute, et de différents modèles capables d'en extraire les informations pertinentes pour prédire l'infarctus final. Ces premières expériences nous permettront aussi de comprendre quels sont les facteurs physiologiques déterminant qui semblent impacter l'apprentissage dans les différents modèles proposés. Dans une deuxième sous-partie, on proposera d'apprendre des modèles sur des données de perfusion brute simulées, et de les tester sur des données réelles. Ces données simulées sont issues d'un simulateur physique déjà publié, auquel nous proposerons des améliorations.

Dans un deuxième temps, pour répondre à la question (2), on proposera des modèles présentant différentes stratégies de fusion de données. Dans un premier temps, on proposera pour la tâche de prédiction un modèle bayésien naïf fusionnant tardivement les différentes modalités d'entrée. Dans un second temps, on explorera de manière plus approfondie un modèle de réseau de neurones fusionnant tardivement les modalités d'entrée. Au travers de ce réseau de neurones, on discutera de l'impact de l'homogénéité de jeux d'apprentissage et notamment des paramètres physiologiques permettant cette stratification ; enfin on visera une explicabilité du modèle en détaillant ses échecs de prédictions lié à la complexité de la course naturelle de la maladie. Plus spécifiquement, on s'attardera sur l'étude de la réversion de la lésion aiguë de diffusion, une cause classique et récurrente d'échec de prédiction.

Comme ouverture, on s'intéressera à la possibilité de prédire l'état clinique des patients après un AVC en intégrant des données en lien avec la phase « chronique » de la maladie. Avec l'émergence de l'internet des objets (IoT), de plus en plus de personnes sont équipées d'un smartphone (60% de la population française) et d'objets connectés (bracelet de suivi de l'activité, montre intelligente,...) remontant des informations sur l'activité de la personne voir sur sa physiologie. On étudiera de manière exploratoire comment ces nouveaux capteurs peuvent assurer un meilleur suivi et accompagnement des patients après un AVC, pendant leur rééducation. Plus spécifiquement, nous présenterons des travaux préliminaires dans lesquels nous explorons des méthodes capables, à partir de signaux actimétriques issus de capteurs, de reconnaître l'activité d'individus tout en masquant leur identité.

Mots clés

imagerie médicale, IRM, accident vasculaire cérébral, prédiction, apprentissage automatique, deep learning, fusion de données, explicabilité, descripteurs texture, simulation physique

Abstract

Ischemic stroke is the leading cause of acquired disability and the third leading cause of death in industrialized countries. Predicting the lesion evolution observed in the acute phase is a clinical challenge. This question is currently being addressed using magnetic resonance imaging (MRI). Two modalities are mainly used : diffusion imaging which shows the extent of cytotoxic edema and dynamic perfusion imaging which shows the hemodynamic state of the tissues. Diffusion MRI is known to be the image with the greatest predictive potential for detecting the area of irreversible infarct. An open question is the contribution of acute perfusion MRI to achieve this prediction. Several machine learning approaches have been tested and turned out to be effective, but results show that the problem is far from being resolved : the prediction task remains complex, since visible lesions in the acute phase may change up to a month later.

The thesis is organized around the following two questions : (1) what is the value of perfusion imaging for ischemic stroke prediction? (2) how to merge perfusion and diffusion MRI within a machine learning model?

The first part of the thesis answers question (1) in two sub-parts : in a first sub-part, we will demonstrate the advantage of learning directly from raw perfusion imaging rather than on extracted parametric maps. We will then discuss different methods for encoding raw perfusion, and different models capable of extracting its relevant information to predict the final infarct. These first experiences will also allow us to understand which determining physiological factors impact the learning process. In a second sub-section, we will propose to train models on simulated raw perfusion data, and to test them on real data. This simulated data comes from a published physical simulator, to which we will suggest improvements.

In a second step, to answer question (2), we will propose models with different data fusion strategies. Firstly, we will propose for the prediction task a naive Bayes model that merges lately the different input modalities. In a second step, we will explore in more depth a neural network model that merges lately the input modalities. Through this neural network, we will discuss the impact of the homogeneity in the training dataset : in particular, physiological parameters allowing its stratification ; Finally, we will aim for a model explicability by detailing its prediction failures associated with the complex course of the disease. More specifically, we will focus on acute diffusion lesion reversal, a recurrent cause of prediction failure.

As an opening, we will study the possibility of predicting patient clinical state after stroke by integrating data related to the “chronic” phase of the disease. With the emergence of the Internet of Things (IoT), more and more people have a smartphone (60% of the French population) and connected objects (activity tracking bracelet, intelligent watch,...) relaying information on someone activity and physiology. We will explore how these new sensors can ensure better follow-up and support for patients after stroke, during their rehabilitation. More specifically, we will present preliminary work in which we explore methods capable, from actimetric signals from sensors, of recognizing the activity of individuals while masking their identity.

Keywords

medical imaging, MRI, stroke, prediction, machine learning, deep learning, data fusion, explainability, texture descriptors, physical simulation

Remerciements

J'aimerais tout d'abord remercier Stéphanie Allassionnière et Caroline Petitjean qui m'ont fait l'honneur d'accepter de lire et juger mes travaux en qualité de rapporteuses. J'aimerais aussi remercier vivement Emmanuel Barbier, Yves Berthezène, Mikael Mazighi et Florence Zara d'avoir accepté d'évaluer mes travaux en tant qu'examineurs.

Je voudrai ensuite remercier mes trois encadrants de thèse Tae-Hee Cho, Carole Frindel et David Rousseau, de m'avoir soutenue et encouragée tout au long de ma thèse. Je leur suis extrêmement reconnaissante pour leur bienveillance, pour le temps qu'ils ont consacré à me transmettre leur savoir, ainsi que pour la confiance qu'ils ont accordé à mon travail. J'ai eu la chance, aussi bien pendant ma thèse que pendant mes stages précédents, d'avoir été encadré par des gens compétents et motivants. Ce sont eux qui m'ont conforté dans mon choix de me tourner vers la recherche aujourd'hui.

Je salue tous les membres du laboratoire de m'avoir aussi bien accueilli à CREATIS. Je tiens à remercier en particulier mes camarades et amis doctorants qui m'ont accompagné pendant mes trois ans de thèse : Anchen Chai, Fei Ge, Sarah Leclerc, Pierre-Antoine Ganaye, William Romero, Thu NGuyen, Yunyun Sun, Méghane Decroocq, Léon Victor, Maxime DiFolco, Pilar Sango, Arthur Gautheron, Paul Nobre, Charly Caredda, Benoit Vernier, Habib Rebbah, Ludmilla Penarrubia, Ali Ahmad et Clément Douarre.

Je voudrai également remercier chaleureusement mes proches compagnons insaliens et amis lyonnais, avec qui j'ai tant partagé ces dernières années à Lyon : Théo Jourdan, Audrey Duran, Samuel Laheux, Marianne Lèbre, Thomas Grégoris, Théophile Sanchez, Rosa Huaman, Vincent Nieto, Lucas Michon, Thomas Chaverondier, Alexis Piat, Zineb Lahjouji, Marianne Borderès, Baptiste Chassagne, Alexandre Labas et Clarisse Peyraud.

Merci à mes fidèles amis de licence Pitchounes, véritables compagnons contre vents et marées, d'avoir été présents ces dernières années.

Enfin, je remercie profondément ma famille, pour leur soutien inconditionnel.

Table des matières

Table des matières	1
Liste des figures	3
Liste des tableaux	8
I Introduction	1
I.1 Diagnostic et pronostic de l'accident vasculaire cérébral ischémique en clinique	2
I.1.1 L'accident vasculaire cérébral ischémique	2
I.1.2 Place de l'imagerie par résonance magnétique multimodale dans la prise en charge de l'accident vasculaire cérébral ischémique	4
I.1.3 Modélisations actuelles de l'évolution de l'infarctus final ischémique en clinique	8
I.2 Champs d'investigation et contributions	13
I.2.1 Intérêt de l'imagerie de perfusion brute pour la prédiction de l'infarctus final	13
I.2.2 Enjeux de la fusion des données multimodales pour la prédiction de l'infarctus final	14
I.2.3 Nouvelles prises en charge de la phase chronique de l'accident vasculaire cérébral	16
II Valeur de l'imagerie de perfusion brute pour la prédiction de l'infarctus final	18
II.3 L'encodage spatio-temporel de l'image de perfusion	19
II.3.1 Encodage de la signature spatio-temporelle locale de chaque voxel	19
II.3.2 Extraction de descripteurs de motifs binaires locaux	21
II.3.3 Évaluation des performances	22
II.3.4 Résultats	23
II.3.5 Discussion et questions ouvertes	25
II.4 Nouvelles approches de description de l'information spatio-temporelle	27
II.4.1 Les coefficients d'Haralick	27
II.4.2 Les coefficients de transformées en ondelettes	28
II.4.3 Évaluation des deux approches proposées	30
II.4.4 Résultats	31
II.4.5 Discussion	33
II.5 Valeur ajoutée de la simulation patient-spécifique	35
II.5.1 Le simulateur IRM utilisé	35
II.5.2 Ensemble de données d'entraînement et de test	37
II.5.3 Évaluation des performances	38
II.5.4 Résultats	44

II.5.5	Discussion	45
II.6	Apport d'un nouveau degré de réalisme relatif à l'arbre vasculaire au simulateur	48
II.6.1	Intégration de la variabilité du flux artériel	48
II.6.2	Intégration du temps de transport de l'agent de contraste	49
II.6.3	Évaluation des améliorations proposées	50
II.6.4	Résultats et Discussion	52
III	Couplage des imageries de diffusion et perfusion pour la prédiction de l'infarctus final	54
III.7	Classification bayésienne naïve à l'échelle du voxel	55
III.7.1	Modélisation et fusion des imageries multimodales	55
III.7.2	Évaluation des performances	56
III.7.3	Résultats	57
III.7.4	Discussion	59
III.8	Segmentation sémantique par réseau de neurones profond à l'échelle de l'image	61
III.8.1	Stratification des données d'entraînement selon la variable de reperfusion précoce	61
III.8.2	Stratégies de fusion proposées pour l'apprentissage par réseau de neurones	63
III.8.3	Évaluation des performances	65
III.8.4	Résultats	66
III.8.5	Discussion	72
III.9	Classification du statut de réversion des voxels de la lésion de diffusion aiguë par forêt aléatoire	75
III.9.1	Facteurs cliniques associés à la réversion, à l'échelle du patient	75
III.9.2	Marqueurs d'images associés à la réversion, à l'échelle du voxel	77
III.9.3	Résultats	78
III.9.4	Discussion	82
IV	Ouverture, conclusions et perspectives	84
IV.10	Potentiel des capteurs portables pour le suivi de l'accident vasculaire cérébral	85
IV.10.1	Description des technologies	85
IV.10.2	Application dans le domaine de la rééducation	87
IV.10.3	Nécessité de valider les capteurs portables grand public	88
IV.10.4	Compromis entre utilité des données et vie privée	88
IV.11	Anonymisation de données actimétriques issus de capteurs portables par filtrage temps-fréquence	90
IV.11.1	Matériel	90
IV.11.2	Domaine temps-fréquence	90
IV.11.3	Filtrage des images temps-fréquences pour anonymiser les signaux	92
IV.11.4	Évaluation de l'approche proposée	93
IV.11.5	Résultats et discussion	95
IV.11.6	Conclusion	96
IV.12	Conclusions sur la tâche de prédiction de l'accident vasculaire cérébral en phase aiguë et chronique	98
IV.12.1	Bilan des contributions	98

IV.12.2	Discussion et perspectives de recherche	99
V	Annexes	117
A	Présentation des cohortes analysées et des pré-traitements associés	118
B	Matériel supplémentaire pour l'apprentissage de réseaux de neurones sur données de perfusion brutes simulées	120

Liste des figures

I.1.1	Les deux types d'accident vasculaire cérébral : l'AVC hémorragique (à gauche) et l'AVC ischémique (à droite). <i>Illustration reproduite de www.coeuretavc.ca/avc/questce-quun-avc</i>	3
I.1.2	Progression dans le temps (gauche à droite) du cœur ischémique (en rouge), avec des dommages irréversibles aux dépens de la pénombre ischémique (en vert). <i>Illustration adaptée de l'article the New England Journal of Medecine 2007; 357 :572-579</i>	3
I.1.3	Illustration des techniques interventionnelles de thrombolyse (A) et de thrombectomie (B). La thrombolyse consiste à dissoudre le caillot, tandis que la thrombectomie consiste à le retirer mécaniquement à l'aide d'un cathéter. <i>Illustrations adaptée de https://www.chumontreal.qc.ca</i>	4
I.1.4	Imagerie par résonance magnétique pondérée T2* de deux patients présentant une hémorragie intra-cérébrale, visible en hypo-signal.	5
I.1.5	Imagerie par résonance magnétique FLAIR aiguë d'un patient atteint d'AVC. Les deux coupes font apparaître en hyper-signal une leucaraïose étendue et un ancien accident vasculaire cérébral lacunaire.	6
I.1.6	Angiographie IRM par temps de vol d'un patient atteint d'AVC présentant une occlusion de l'artère cérébrale moyenne gauche. L'image de gauche correspond à une reconstruction axiale de coupe 2D, et l'image de droite correspond une reconstruction 3D du cercle de Willis.	6
I.1.7	Imagerie par résonance magnétique pondérée de diffusion de trois patients atteints d'AVC. Sur chaque coupe, la lésion ischémique est visible en hyper-signal.	7
I.1.8	Imagerie par résonance magnétique pondérée de perfusion d'un patient atteint d'AVC. Le passage de l'agent de contraste dans les tissus est illustré pour une coupe donnée (z=12). Au moment du pic de passage de l'agent de contraste (t=24), au niveau de la flèche rouge, on voit que le produit d'agent de contraste en hypo-signal circule difficilement par rapport aux tissus environnants : il s'agit de la zone de pénombre.	7
I.1.9	Imagerie de suivi FLAIR de trois patients atteints d'AVC. La lésion ischémique est visible en hyper-intensité.	8

I.1.10	Étude de cas afin d'illustrer les IRM intervenant dans l'examen clinique de l'AVC. Depuis l'angiographie (A), on localise l'occlusion au niveau de l'artère cérébrale moyenne droite (segment M1). Aucun tissu hémorragique n'est détectable sur la séquence T2* (B). Pas de trace de lésion parenchymateuse sur la séquence FLAIR (C). L'infarctus est visible en hyper-intensité sur la séquence de diffusion (D) et en hypo-intensité sur la séquence ADC (E). La perfusion à un temps donné (F) rend compte d'une zone hypo-perfusée plus étendue que la lésion aiguë de diffusion. L'évolution finale de la lésion est visible en hyper-intensité sur l'IRM FLAIR de suivi (G).	8
I.1.11	Estimation de la courbe moyenne de concentration d'agent de contraste depuis le signal de perfusion.	9
I.1.12	Courbe tissulaire de concentration d'agent de contraste avant et après déconvolution. Plusieurs cartes de paramètres de perfusion sont présentées (temps de pic du TTP, temps de transit moyen du MTT, débit sanguin cérébral du CBF, volume sanguin cérébral du CBV). <i>Illustration adaptée de l'article Stroke. 2014;16:131-45.</i>	10
I.1.13	Exemples cliniques de deux patients traités par thrombectomie présentant des évolutions différentes. La colonne 1 correspond à l'IRM aiguë de diffusion, la colonne 2 à l'IRM aiguë de perfusion, et la colonne 3 à l'IRM de suivi FLAIR. Le patient A présente une discordance significative entre la lésion de diffusion (en jaune) et la pénombre (en bleu). Sur l'IRM de suivi, on voit que le patient a bénéficié d'une reperfusion : la pénombre a été sauvée et la lésion finale (en rouge) s'apparente à la lésion de diffusion aiguë. Le patient B ne présente pas de discordance significative entre la lésion de diffusion et la pénombre, et sa lésion finale s'apparente à l'union de la lésion de diffusion aiguë et de la pénombre ischémique (en bleu). Ce patient n'a pas bénéficié d'une reperfusion.	11
I.2.1	Nécessité de prendre en charge les patients victimes d'accident vasculaire cérébral : dès l'admission à l'hôpital, dans la phase aiguë (A); après la sortie de l'hôpital, dans phase chronique (B).	13
I.2.2	Les trois niveaux classiques de fusion d'images. <i>Illustration inspirée de l'ouvrage CRC (2018). Multisensor data fusion and machine learning for environmental remote sensing.</i>	16
II.3.1	Illustration de la signature spatio-temporelle des signaux IRM bruts en IRM de perfusion d'un patient donné. La séquence présentée a été acquise sur 20 coupes spatiales pendant 60 pas de temps consécutifs.	20
II.3.2	Encodage de la signature spatio-temporelle des signaux d'IRM de perfusion sous forme de patch.	21
II.3.3	Illustration des signatures typiques obtenues pour les patches de voisinage $N_p = 1$, $N_p = 3$ et $N_p = 5$ (de haut en bas) pour les voxels sains (à gauche) et les voxels pathologiques (à droite).	21
II.3.4	Pipeline de l'approche de classification proposée.	22
II.3.5	Visualisation de la répartition des 4 patients choisis (encerclés en rouge) parmi les 77 patients I-Know à disposition sur un plan de projection 2D obtenu après ACP sur 3 descripteurs de forme (la boîte englobante de la lésion, le logarithme de la lésion et le nombre de composantes connexes de la lésion). Les deux axes sont les deux premiers vecteurs propres de l'ACP et constituent 95% de la variance totale. Chaque patient a été assigné à un groupe en utilisant un clustering K-Means. Chaque groupe est représenté par un nombre (1, 2, 3, 4) et une couleur (respectivement bleu, vert, jaune, rouge). Les ellipses représentent les enveloppes convexes qui rassemblent les patients d'un même groupe.	26

II.4.1	Un propagateur scattering \tilde{W} appliqué à une image x calcule la première couche du module des coefficients d'ondelette $U[\lambda_1]_x = x * \psi_1 $ et donne en sortie sa moyenne locale $S[0]x = x * \phi_{2j}$ (flèche noire). De manière générale, en appliquant \tilde{W} sur chaque signal propagé $U[p]x$, on obtient en sortie $S[p]x = U[p]x * \phi_{2j}$ (flèche noire). La variable m correspond à l'indice de couche. <i>Figure reproduite de l'article IEEE transactions on pattern analysis and machine intelligence (2013), 35(8), 1872-1886.</i>	29
II.4.2	Images de sortie pour chaque classe (patch interpolé sain à gauche et patch interpolé lésionnel à droite) et pour trois couches m de la transformée en scattering, avec $r=2$ et $j=3$	30
II.4.3	Images de sortie pour chaque classe (patch interpolé sain à gauche et patch interpolé lésionnel à droite) et pour trois couches m de la transformée en scattering, avec $r=2$ et $j=5$	30
II.4.4	Prédictions obtenues pour deux patients de validation à partir des coefficients scattering (colonne 1) et des coefficients d'Haralick (colonne 2). Les prédictions sont à comparer avec la vérité terrain (colonne 3).	33
II.5.1	Courbes AIF moyennes extraites sur chacun des 76 patients de la cohorte. Les courbes bleues et rouges correspondent aux AIF avec des distributions respectivement étroites et larges. Les lignes grises correspondent aux AIF restantes. On peut constater la grande variabilité et l'inhomogénéité de forme des AIF. . .	37
II.5.2	Courbes AIF des 8 patients testés (en couleur) parmi toutes les AIF de la cohorte (en gris). Afin de mieux les visualiser, nous n'avons pas montré le délai d . Les patients choisis présentent de grandes différences de forme d'AIF.	38
II.5.3	Vue d'ensemble du pipeline de prédiction proposé. (a) Les images initiales sont des images de concentration d'agent de contraste. Dans l'expérience 1, l'ensemble de données d'entraînement consiste en des patchs provenant d'images de concentration réelles, tandis que dans les expériences 2 et 3, l'ensemble de données d'entraînement consiste en des patchs synthétiques obtenus à partir du simulateur. Dans l'expérience 2, le paramètre d'entrée AIF est réglé sur une valeur par défaut, et dans l'expérience 3, les paramètres d'entrée AIF sont ceux du patient testé. (b) Les images de concentration sont codées en patchs spatio-temporels. (c) Un modèle de réseau neuronal convolutif (CNN) est entraîné à partir de patchs d'images de concentration. (d) Chaque voxel des images de concentration testées est classé comme sain ou infarci.	42
II.5.4	Illustration d'un patch sain (à gauche) et d'un patch pathologique (à droite) issus de l'image de concentration pour les patients test 1 et 8. Les voxels sains présentent un segment hyperintense étroit résultant du passage rapide du bolus d'agent de contraste. Les voxels pathologiques présentent un segment hyperintense étalé, bruité et peu contrasté, résultant du passage difficile du bolus d'agent de contraste. Les deux patients présentent des caractéristiques hémodynamiques différentes : le temps de transit de l'agent de contraste, représenté par l'espace entre les deux flèches, est plus rapide pour le patient 8 que pour le patient 2.	43
II.5.5	Métrique DSC en fonction du nombre de patchs d'entraînement lors de l'expérience 3. Chaque point correspond au DSC moyenné sur l'ensemble des patients test et sur les 10 expériences répétées. La ligne pointillée rouge indique le nombre optimal de patchs d'entraînement.	45

II.5.6	Prédictions obtenues lors de l'expérience 1 (1ère colonne), de l'expérience 2 (2ème colonne) et de l'expérience 3 (3ème colonne) pour les patients testés. La barre de couleur à gauche présente la probabilité pour chaque voxel d'être sain. Les voxels dans les tons bleus ont été prédits sains et les voxels dans les tons rouges ont été prédits infarcis. Le biomarqueur classique T_{max} , dont le seuil est fixé à 6 secondes, est indiqué (4e colonne). Les colonnes 1 à 4 doivent être comparées au masque lésionnel du FLAIR final (5e colonne).	47
II.6.1	Pipeline de la nouvelle version du simulateur d'IRM de perfusion. Les nouvelles sources de variabilité introduites sont affichées en bleu foncé.	48
II.6.2	Ensemble des AIF extraites manuellement (noir) et le signal médian associé (rouge) pour 2 patients d'étude. À titre d'exemple, la variabilité des AIF intra-patient est plus importante pour le patient à droite que le patient de gauche. . .	49
II.6.3	Carte de délai pour une coupe d'un patient donné, calculée avec des seuils de délai successifs et croissants. Cette représentation donne des informations sur la propagation de l'agent de contraste depuis les artères principales vers les tissus.	50
II.6.4	Pipeline général de l'approche proposée.	50
II.6.5	Les différentes bases de données prises en compte dans nos expériences. Chacune correspond à un degré de réalisme différent, qui augmente de la base de données A à la base de données D.	51
II.6.6	Prédictions obtenues après entraînement de la base de données C (2ème colonne) et de la base de données D (3ème colonne) pour deux patients testés. L'image DSC-PWI des patients testés est donnée en 1ère colonne. Les colonnes 2 et 3 doivent être comparées à la vérité de terrain (4ème colonne). Les voxels dans les tons bleus ont été prédits sains et les voxels dans les tons rouges ont été prédits infarcis. Le patient de la première ligne a un cercle de Willis (CW) incomplet alors que le patient de la deuxième ligne a un CW complet. Le délai dans la base de données D semble améliorer la prédiction de la lésion lorsque le cercle de Willis est complet (voir les cercles en gras).	53
III.7.1	Vraisemblances des 6 cartes binaires pour la classe lésion, pour toutes les combinaisons de patients possibles dans l'ensemble d'apprentissage.	57
III.7.2	DSC de la classe "lésion" en fonction du nombre de patients dans l'ensemble d'apprentissage, pour $N = 0$. Les lignes vertes correspondent à un résultat DSC d'apprentissage individuel, tandis que les aires violettes représentent la distribution DSC pour plusieurs combinaisons de patients. Toutes les combinaisons possibles de patients sont représentées.	58
III.7.3	Résultats du modèle bayésien naïf pour une coupe d'un patient testé. Deux cartes hémodynamiques X_i (DWI et CBF) des six utilisées pour l'apprentissage ainsi que l'imagerie FLAIR final Y sont représentées binarisées. En rose sont représentés les voxels considérés "sain" et en violet les voxels considérés "lésion".	58
III.8.1	Aperçu de l'architecture CNN proposée. En haut à gauche : Le réseau prend en entrée cinq images IRM (coupes 2D des volumes DWI, ADC, CBV, CBF, T_{max}). En bas : Chaque image d'entrée est traitée indépendamment sur 5 branches distinctes. Les cartes de caractéristiques roses, violettes, jaunes, rouges et vertes résultent de convolutions 2D et d'opérations de maxpooling. Les sorties des 5 branches sont ensuite concaténées, et sur-échantillonnées par des couches de déconvolution 2D. Le réseau produit en sortie une carte avec 3 classes (lésion, tissu sain et fond). En haut à droite : La lésion prédite doit être comparée à la lésion réelle du FLAIR final.	64
III.8.2	Organigramme d'inclusion des patients.	66

III.8.3	Prédictions de l'infarctus final de deux patients, obtenues à partir des CNN re-perfusé et non re-perfusé : le patient 1 ne présente pas de reperfusion (TICI=2a); le patient 2 est re-perfusé (TICI=2b).	69
III.8.4	Prédictions de l'infarctus final, obtenues à partir des modèles CNN et des modèles mismatch perfusion-diffusion. Cinq patients testés sont présentés : deux cas de réussite où le modèle CNN réalise de meilleures performances que le modèle mismatch correspondant (patient A avec TICI=2a et patient B avec TICI=3) et trois patients difficiles à prédire, tant pour le modèle CNN que pour le modèle mismatch (patient C avec TICI=2a, patient D avec TICI=3 et patient E avec TICI=2b).	70
III.9.1	Organigramme d'inclusion des patients.	76
III.9.2	Patients d'étude en fonction du volume de lésion DWI aiguë réversé (axe x) et du ratio de volume de lésion DWI réversé (axe y). Les patients réverseurs (marqueurs triangulaires rouges) ont réversé d'au moins 10 ml et 25% de leur volume de lésion DWI aiguë.	76
III.9.3	Les voxels d'étude inclus dans la lésion DWI aiguë sont classés réversibles ou non réversibles.	77
III.9.4	Boxplots des performances du modèle proposé en fonction de la taille des lésions de DWI (lésions peu volumineuses : volume < 50mL, 70 patients; lésion volumineuses : volume > 50mL, 21 patients). Les différences entre les deux groupes ont été évaluées à l'aide d'un test de Wilcoxon non apparié.	80
III.9.5	Prédictions données par le modèle proposé, comparée avec la vérité terrain. Les tons bleus désignent les voxels réversibles, tandis que les tons rouges désignent les voxels non réversibles. Cinq patients de validation sont montrés : trois patients bien prédits - avec une AUC > 0.7 (patient 1 avec TICI=3, patient 2 avec TICI=3 et patient 3 avec TICI=2b), et deux patients difficiles à prédire - avec une AUC < 0.5 (patient 4 avec TICI=2b et patient 5 avec TICI=3). DSC_r and DSC_{nr} désignent les DSC des voxels réversibles et non réversibles, respectivement.	81
III.9.6	Graphique récapitulatif mesurant à quel point les caractéristiques d'entrée (DWI, T_{max} , ADC, CBV, CBF) contribuent (positivement ou négativement) à discriminer les voxels d'entraînement. Chaque point correspond à un voxel d'entraînement. Les points à droite de l'axe vertical gris correspondent aux voxels non réversibles, tandis que les points à gauche correspondent aux voxels réversibles. Les points dans les tons chauds représentent des valeurs de voxels élevées pour une caractéristique donnée, et les couleurs plus froides indiquent des valeurs plus faibles. Pour une caractéristique donnée, plus les points le long de l'axe x sont nombreux et éloignés de l'axe vertical gris, plus cette caractéristique contribue à la classification.	82
IV.10.1	Illustration d'un système de surveillance à distance de la santé basé sur des capteurs portables. <i>Figure reproduite de "A review of wearable sensors and systems with application in rehabilitation." Journal of neuroengineering and rehabilitation 9.1 (2012) : 1-17.</i>	86
IV.11.1	Représentation de la transformation de Stockwell optimisée pour 2 utilisateurs différents (#8 et #15) pour deux activités différentes (marche et course).	92
IV.11.2	Aperçu de l'architecture CNN proposée. Le réseau prend trois images TF (TF_x , TF_y , TF_z) en entrée. Chaque image d'entrée est traitée indépendamment sur 3 branches distinctes. Les cartes de caractéristiques roses, jaunes et vertes résultent de convolutions 2D et d'opérations <i>maxpooling</i> . Les sorties des 3 branches sont ensuite concaténées, et passent à travers une couche cachée de N nœuds, avec $N = 4$ pour le CNN activité et $N = 24$ pour le CNN identité.	93

IV.11.3	Vue d'ensemble du pipeline proposé, divisé en 4 étapes : A. Transformation du signal en une image temps-fréquence (TF), B. Méthode d'anonymisation basée sur le filtrage des images, C. Reconnaissance d'activité et D. Identification de l'utilisateur.	94
IV.11.4	Accuracy de l'activité en fonction de l'accuracy de l'identité pour différentes représentations : la transformée de Fourier (marqueurs en croix en vert), la STFT (marqueurs ronds en bleu), la transformée de Stockwell (marqueurs carrés en orange) et la transformée de Stockwell optimisée (marqueurs triangulaires en rouge). Chaque point correspond à un résultat de classification moyen sur 10 expériences. Le coin supérieur gauche représente le compromis idéal entre l'utilité et la vie privée. Pour chaque courbe, les points les plus performants en activité et en identité correspondent à des cas sans filtrage alors que les autres (les points qui tendent vers la gauche du graphique) correspondent à des cas avec filtrage avec un pas de 10%.	96
A.1	Exemple de recalage non-linéaire effectué sur deux patients de la cohorte HIBISCUS-STROKE. L'IRM de diffusion (colonne A) et l'IRM de suivi FLAIR (colonne B/D) sont représentées dans leur espace initial. Après recalage, l'IRM de diffusion (colonne C) est dans l'espace initial de l'IRM FLAIR.	119

Liste des tableaux

II.3.1	Précision (%), Sensibilité (%), Spécificité (%) obtenues par validation croisée 100-Folds pour des patches de type I_1 (échelle d'observation $N_p = 1$). Différentes tailles N_h (largeur du segment) sont évaluées. N_f correspond au nombre de descripteurs sélectionnés après réduction de dimension par ACP.	23
II.3.2	Précision (%), Sensibilité (%), Spécificité (%) obtenues par validation croisée 100-Folds pour $N_h = 3$ et différents types de patch I_{N_p} sont évalués. N_f correspond au nombre de descripteurs sélectionnés après réduction de dimension par ACP.	24
II.3.3	Précision (%), sensibilité (%), spécificité (%) obtenues par validation croisée 100-Folds pour $N_h = 3$ et $N_p = 9$. N_f correspond au nombre de descripteurs sélectionnés après réduction de dimension par ACP. Ces résultats comparent la qualité de la prédiction obtenue en utilisant : des signaux de perfusion bruts PWI en entrée du modèle (1ère ligne), l'image de concentration de l'agent de contraste CAC (2ème ligne), ou l'image de concentration déconvoluée avec l'algorithme de Tikhonov IRF [1] (3ème ligne).	24
II.4.1	Coefficient Dice (DSC) et distance de Hausdorff (HD) de la lésion prédite obtenus à partir des deux descripteurs manuels proposés : les coefficients d'Haralick et les coefficients de la transformée en scattering. Les métriques sont moyennées sur l'ensemble de validation (moyenne \pm écart-type). Le temps d'extraction des descripteurs est précisé pour un patient donné (d'environ 32 500 patches). Pour les coefficients scattering, le calcul a été parallélisé sur 4CPU d'une station standard.	32

II.5.1	Paramètres par défaut du simulateur pour la variabilité tissulaire des paramètres hémodynamiques de débit sanguin cérébral (CBF, en mL/100g/s) et temps de transit moyen (MTT, en s). La distribution de chaque paramètre est modélisée par une distribution gaussienne de moyenne μ , écart-type σ . Trois classes de tissus sont considérés : tissus gris sains (TGS), tissus blancs sains (TBS) et tissus lésionnels (TL). Le fond (F) présente une distribution nulle.	36
II.5.2	Valeurs des paramètres d'AIF (y_{max} , t_{max} , d , α) des patients testés obtenues à partir de leur courbe gamma moyenne d'agent de contraste.	38
II.5.3	Architecture du réseau neuronal convolutif proposé. Les deux premières couches sont de type Convolutionnel 2D (Conv2D) et les deux dernières couches sont des couches entièrement connectées (FC pour <i>Fully Connected</i>).	40
II.5.4	Distance de Hausdorff (HD) et coefficient de similarité (DSC) après avoir réalisé 10 fois l'expérience 1 (1ère colonne), l'expérience 2 (2ème colonne) et l'expérience 3 (3ème colonne). La moyenne de toutes les mesures est calculée sur les 10 réalisations, et indiquée pour chaque patient testé (moyenne \pm écart type). Les métriques DSC et HD entre le $T_{max} \geq 6s$ et la lésion finale sont reportées (4ème colonne). Nous avons indiqué en gras les valeurs pour lesquelles l'expérience 3 a donné les meilleures performances à l'échelle du patient. Dans les trois expériences, l'écart-type de la métrique HD est faible, voire nul, car certains voxels aberrants ont été systématiquement mal prédits. Les métriques moyennes sur l'ensemble des données test ont également été calculées (dernière ligne).	44
II.6.1	Métriques DSC et HD entre les voxels prédits infarctés et la lésion finale de la vérité de terrain. La moyenne de toutes les mesures est calculée sur l'ensemble des données de test (moyenne \pm écart type).	52
III.7.1	Encodage en 6 valeurs pour les cartes binaires prenant en compte, pour chaque voxel, l'état des voxels de son voisinage $N \times N$	56
III.7.2	Coefficient de similarité DSC des tissus infarctés et des tissus sains en fonction de l'échelle d'observation N de chaque voxel. Les mesures sont moyennées sur toutes les combinaisons de patients possibles (moyenne \pm écart-type).	59
III.8.1	Partie contractante du CNN proposé, composée de 5 blocs de convolution (Conv Block), d'opérations de maxpooling (2D MaxPooling) et de <i>dropout</i> . Chaque Conv Block est composé de : Convolution 2D (3*3)+ <i>batch normalisation</i> + convolution 2D (3*3)+ <i>batch normalisation</i>	63
III.8.2	Caractéristiques cliniques des patients sélectionnés (médiane avec écart interquartile, sauf indication contraire). NIHSS : National Institutes of Health Stroke Scale; DWI : imagerie pondérée de diffusion; ICA : artère carotide interne. . . .	67
III.8.3	Mesure des performances des modèles généraux, reperfusés et non-reperfusés chez les patients (a) reperfusés et (c) non-reperfusés (valeurs moyennes \pm écart type). Les valeurs en gras correspondent à la meilleure valeur de métrique en colonne. Les P-values des tests bilatéraux de Wilcoxon comparent les modèles généraux, reperfusés et non-reperfusés chez les patients (b) reperfusé et (d) non-reperfusé . Les P-values en gras correspondent à des différences significatives, (*) indiquant $P < 0.05$, (**) indiquant $P < 0.01$ et (***) indiquant $P < 0.001$	68

III.8.4	Comparaison entre les modèles CNN et les modèles mismatch perfusion-diffusion. Parmi les patients reperfusés (deux premières lignes), le modèle reperfusé CNN a été comparé à la lésion de diffusion aiguë. Parmi les patients non reperfusés (deux dernières lignes), le modèle non reperfusé CNN a été comparé à l'union de la lésion de diffusion aiguë et de la pénombre ischémique. Les valeurs en gras correspondent à la meilleure valeur de métrique en colonne. Un test bilatéral de wilcoxon a été effectué entre les modèles CNN proposés et les modèles cliniques mismatch, avec (*) indiquant $P < 0.05$, (**) indiquant $P < 0.01$ et (***) indiquant $P < 0.001$	68
III.8.5	Métriques d'évaluation des modèles reperfusés et non reperfusés après ablation successive des entrées IRM, testés parmi les patients (a) reperfusés et (b) non reperfusés , respectivement (valeurs moyennes \pm écart type). Les valeurs en gras correspondent à la meilleure valeur de métrique en colonne. Un test bilatéral de Wilcoxon a été effectué entre les modèles complets (5 IRM en entrées) et les modèles après ablation, avec (.) indiquant $P < 0.10$, (*) indiquant $P < 0.05$, (**) indiquant $P < 0.01$ et (***) indiquant $P < 0.001$	71
III.8.6	Métriques d'évaluation des modèles généraux, reperfusés et non reperfusés avec différentes stratégies de fusion (précoce et tardive) sur des patients reperfusés (a) et des patients non reperfusés (b) (valeurs moyennes \pm écart type). Les valeurs en gras correspondent à la meilleure valeur de la métrique en colonne. Un test bilatéral de Wilcoxon a été effectué entre le modèle global et les deux autres modèles (reperfusé et non-reperfusé) pour une stratégie de fusion donnée, avec (*) indiquant $P < 0.05$, (**) indiquant $P < 0.01$ et (***) indiquant $P < 0.001$	74
III.9.1	Caractéristiques cliniques (médiane avec écart inter-quartile, sauf indication contraire) pour les patients réverseurs et non réverseurs. NIHSS : National Institutes of Health Stroke Scale; DWI : diffusion-weighted imaging; TIV : thrombolyse intra-veineuse; d_1 : délai entre les symptômes et l'IRM; d_2 : délai entre les symptômes et la reperfusion.	79
III.9.2	Régression logistique univariée. NIHSS : National Institutes of Health Stroke Scale; TIV : thrombolyse intra-veineuse; d_1 : délai entre les symptômes et l'IRM; d_2 : délai entre les symptômes et la reperfusion. IC : Intervalle de Confiance.	79
III.9.3	Performances du modèle proposé pour la prédiction de la réversion de la lésion DWI. Les résultats ont été obtenus par une approche de validation croisée 5-Fold. L'AUC, la spécificité, la sensibilité et les DSC donnés correspondent aux moyennes sur les 5 validations (moyenne \pm écart-type). Les valeurs minimales (Min) et maximales (Max) sur les 5 validations ont également été données.	80
III.9.4	Contribution des modalités d'entrée (en %) à la tâche de prédiction (moyenne \pm écart-type).	82
IV.11.1	Aire sous la courbe utilité-vie privée normalisée (AUC) pour chaque représentation	96
IV.11.2	Filtre optimal en % pour chaque représentation, et les performances associées en % (activité <i>acc</i> /identité <i>acc</i>)	96
B.1	Coefficient de corrélation moyen de Matthews (MCC), aire moyenne sous la courbe (AUC), rappel moyen et précision moyenne après avoir effectué les expériences 1, 2 et 3 en utilisant le réseau neuronal convolutif du Tableau II.5.4. Les métriques sont indiquées pour chaque patient testé et correspondent à la moyenne sur 10 répétitions (moyenne \pm écart-type).	121
B.2	Fully Convolutional Network 1 (FCN1) proposé.	122
B.3	Fully Convolutional Network 2 (FCN2) proposé.	122

B.4	Fully Convolutional Network 3 (FCN3) proposé.	122
B.5	Dice (DSC) et distance de Hausdorff (HD) après avoir effectué 10 répétitions des expériences 1, 2 et 3 en utilisant trois différents réseaux entièrement convolutifs : FCN1 (1ère colonne), FCN2 (2ème colonne) et FCN3 (3ème colonne). Les métriques sont indiquées pour chaque patient testé et correspondent à la moyenne sur 10 répétitions (moyenne \pm écart-type). Nous avons indiqué en gras quand l'expérience 3 a donné les meilleures performances à l'échelle du patient.	123

Partie I

Introduction

Chapitre I.1

Diagnostic et pronostic de l'accident vasculaire cérébral ischémique en clinique

I.1.1 L'accident vasculaire cérébral ischémique

I.1.1.1 L'accident vasculaire cérébral : une priorité de santé publique

L'accident vasculaire cérébral (AVC) ischémique est une pathologie grave provoquée par un arrêt brutal de la circulation sanguine dans une partie du cerveau. L'arrêt de la circulation du sang entraîne la mort des cellules cérébrales au niveau de la zone du cerveau touchée en les privant d'un apport vital en oxygène et en nutriment. L'AVC se traduit par un déficit neurologique d'installation soudaine, qui peut concerner la motricité, la sensibilité, le langage, ou la vision, pour ne citer que les manifestations les plus fréquentes, selon le siège de la lésion cérébrale.

En France, l'AVC est la troisième cause de décès, première cause d'handicap acquis et la deuxième cause de démence. L'incidence est de 100 000 à 145 000 AVC par an, et le risque est particulièrement élevé chez les personnes âgées : 75% des patients ont plus de 65 ans. En moyenne, une personne sur 6 aura un AVC dans sa vie [2]. À l'échelle internationale, l'AVC est la deuxième cause de mortalité et troisième cause d'invalidité, avec un AVC survenant toutes les 5 secondes dans le monde, résultant en près de 15 millions de cas d'AVC chaque année [3]. Les risques vitaux et fonctionnels associés à cette pathologie en font une priorité de santé publique mondiale.

I.1.1.2 Origines de l'accident vasculaire cérébral ischémique

On distingue deux formes d'accident vasculaire cérébral : l'AVC hémorragique (hématome intracérébral) et l'AVC ischémique (infarctus cérébral). L'AVC ischémique est la forme la plus courante (80% des cas d'AVC) et survient lorsqu'un caillot bloque le flux d'une artère cérébrale. Le plus souvent, le mécanisme de l'occlusion artérielle est de nature embolique (c.a.d. caillot provenant d'une autre artère ou des cavités cardiaques, et véhiculé par le sang jusqu'au cerveau).

I.1.1.3 La zone de pénombre ischémique : cible de toute intervention thérapeutique

Au niveau de la zone cérébrale touchée, l'oxygène vient à manquer. Le tissu dans lequel le débit sanguin est inférieur à 22 mL/min/100g est défini comme hypoperfusé. L'activité des neurones qui le constitue est fortement altérée, voire nulle. Les neurones soumis à un débit inférieur à 10 mL/min/100g pendant plus de quelques minutes meurent, provoquant une nécrose irréversible : ils constituent ce qu'on appelle le cœur ischémique. Le tissu hypoperfusé non nécrosé (débit sanguin maintenu entre 10 et 22 mL/min/100g) est classiquement appelé zone de pénombre ischémique. Cependant cette définition stricte - fondée sur un seuil - est difficilement applicable à des outils d'imagerie non quantitatifs. La zone de pénombre est donc avant tout un concept, pour définir un

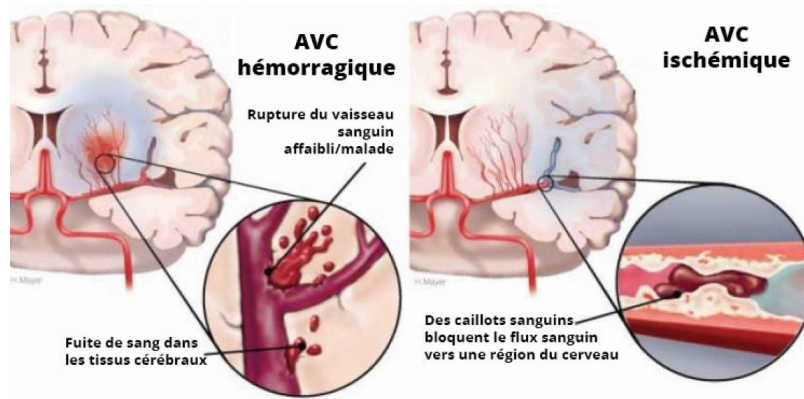


FIGURE I.1.1 – Les deux types d'accident vasculaire cérébral : l'AVC hémorragique (à gauche) et l'AVC ischémique (à droite). *Illustration reproduite de www.couretavc.ca/avc/questce-quin-avc*

zone ischémique à risque de nécrose, potentiellement viable, dont la définition dépend de l'outil pour l'évaluer.

On estime qu'environ 2 millions de neurones meurent chaque minute après la survenue d'une occlusion d'une artère intracrânienne proximale. Il est donc nécessaire de prendre en charge les patients dès les premiers symptômes le plus rapidement possible pour désobstruer l'artère touchée afin de rétablir le débit sanguin : la pénombre ischémique peut ainsi avoir une chance de retrouver une activité normale. Sans intervention, la pénombre ischémique risque de complètement nécroser avec le temps. Avec les traitements actuels, le délai pour intervenir est de quelques heures. La formule utilisée par les neurologues vasculaires pour décrire cette urgence thérapeutique est "Time is brain" ("Le temps, c'est du cerveau"), comme le montre la Figure I.1.2.

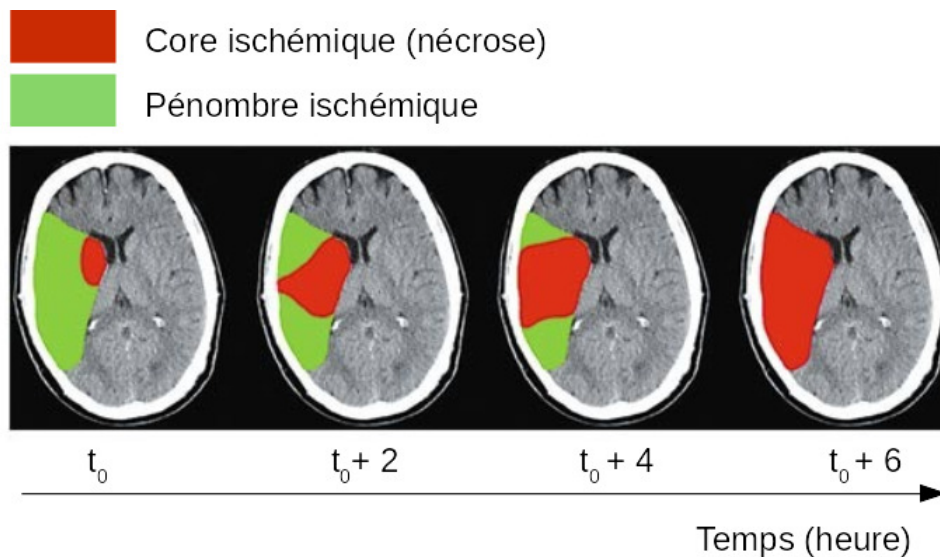


FIGURE I.1.2 – Progression dans le temps (gauche à droite) du cœur ischémique (en rouge), avec des dommages irréversibles aux dépens de la pénombre ischémique (en vert). *Illustration adaptée de l'article the New England Journal of Medicine 2007; 357 :572-579*

I.1.2 Place de l'imagerie par résonance magnétique multimodale dans la prise en charge de l'accident vasculaire cérébral ischémique

I.1.2.1 Traitements de l'accident vasculaire cérébral aigu

Dans les premières heures après le début des symptômes, un traitement de revascularisation est possible. Les deux approches thérapeutiques qui existent à l'heure actuelle sont la thrombolyse et la thrombectomie.

La thrombolyse intraveineuse

Le premier traitement de revascularisation urgente validé pour les AVC ischémiques fut la thrombolyse, une intervention qui consiste à injecter par voie veineuse un fibrinolytique (le *rtPA* : recombinaut tissulaire Plasminogène Activateur) pour dissoudre le caillot. Le bénéfice attendu est la reperfusion tissulaire afin d'améliorer le pronostic fonctionnel et vital. Sa principale limite est sa faible efficacité pour les occlusions des gros troncs artériels intracrâniens (moins de 30% de recanalisation du segment M1 de l'artère cérébrale moyenne). La thrombolyse comporte aussi un risque d'hémorragie intracérébrale (~2% d'hématomes fatals). La thrombolyse présente aussi l'inconvénient de devoir être réalisée moins de 4h30 après l'apparition des premiers symptômes pour être efficace. Le nombre de patients pouvant bénéficier de ce traitement est donc relativement faible (environ 10 à 15% des patients admis en unité neuro-vasculaire [2]).

La thrombectomie

Depuis 2015, un nouveau traitement a révolutionné la prise en charge aiguë des AVC ischémiques : la thrombectomie mécanique. Cette intervention consiste à retirer le caillot par voie endovasculaire. La thrombectomie peut être réalisée en cas d'occlusion d'une artère de gros calibre (1,5 à 4 mm de diamètre) jusqu'à 6h après les premiers symptômes, voire même 24h dans certaines situations. Elle peut être pratiquée en complément de la thrombolyse intraveineuse ou d'emblée si cette dernière est contre-indiquée. De nombreuses études ont montré le bénéfice de cette intervention, en offrant de meilleures chances de récupération aux patients. En revanche, cet acte radiologique nécessite une bonne expertise et un plateau technique adapté; conditions qu'aujourd'hui peu d'hôpitaux en France peuvent satisfaire.

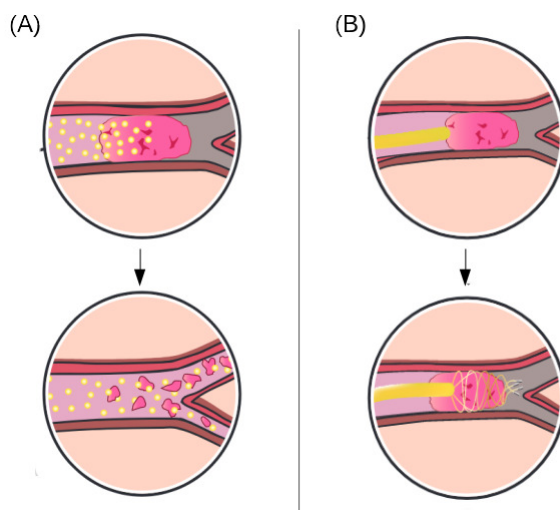


FIGURE I.1.3 – Illustration des techniques interventionnelles de thrombolyse (A) et de thrombectomie (B). La thrombolyse consiste à dissoudre le caillot, tandis que la thrombectomie consiste à le retirer mécaniquement à l'aide d'un cathéter. Illustrations adaptées de <https://www.chumontreal.qc.ca>

La reperfusion précoce par thrombolyse intraveineuse et/ou thrombectomie est le principal objectif thérapeutique de l'accident ischémique cérébral aigu : elle consiste à revasculariser les

tissus de la pénombre ischémique afin qu'ils retrouvent des conditions préalables à leur fonctionnement actif.

La neuroimagerie joue un rôle central dans la décision thérapeutique : une évaluation le plus souvent qualitative est effectuée afin d'évaluer le rapport bénéfices/risques d'un potentiel traitement. Malgré l'expérience et l'expertise, des éventuels détails peuvent échapper à l'œil humain. Dans cette optique, la vision par ordinateur peut potentiellement capturer des descripteurs non perceptibles, et les combiner pour guider des décisions. Nos travaux s'inscrivent donc dans une démarche d'automatisation, afin d'investiguer des méthodes innovantes capables d'accompagner l'expertise neurologique dans la prise de décision.

I.1.2.2 Intégration de l'imagerie par résonance magnétique dans la prise de décision thérapeutique

Aujourd'hui, les principales technologies pour diagnostiquer et pronostiquer l'AVC ischémique sont la tomodensitométrie à rayons X (CT pour *computed tomography*, ou scanner) et l'imagerie par résonance magnétique (IRM). Le scanner est simple, rapide et économique, mais manque de sensibilité (la région de l'AVC ne semble pas anormale pendant les premières heures suivant le début de l'AVC, et peut être trop petite pour être vue sur le scanner). Par opposition, l'IRM est beaucoup plus précise et moins invasive (pas de rayon X), mais le temps d'acquisition est plus long et la technologie reste plus onéreuse et moins disponible. La majorité des patients faisant l'objet d'un traitement de revascularisation en France sont évalués par IRM, ce qui n'est pas le cas dans d'autres pays (e.g. Etats-Unis). L'IRM reste l'examen idéal de phase aiguë à privilégier dans la mesure des disponibilités pour la prise en charge de l'AVC.

En routine clinique, les séquences pondérées en T2*, FLAIR, diffusion et l'angiographie IRM par temps de vol sont systématiques dans le protocole IRM. L'imagerie de perfusion n'est pas systématique, mais peut aisément être effectuée (durée d'acquisition d'environ 1 minute). Nous présentons succinctement le rôle de chacune de ces séquences selon leur place dans le protocole IRM.

IRM à l'admission

Les séquences acquises à l'admission du patient à l'hôpital permettent de dresser un bilan rapide en phase aiguë afin de fournir un traitement et une intervention optimale.

— IRM pondérée T2* : évaluer de potentiels hématomes

La séquence T2* est sensible à la présence de saignements intra-cérébraux, en les faisant apparaître en hypo-signal. La présence de saignement compromet le traitement par thrombolyse.

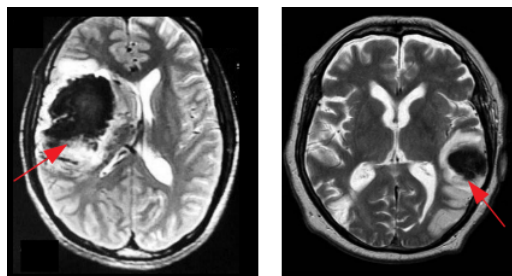


FIGURE I.1.4 – Imagerie par résonance magnétique pondérée T2* de deux patients présentant une hémorragie intra-cérébrale, visible en hypo-signal.

— IRM FLAIR : évaluer de potentiels lésions du parenchyme

La séquence FLAIR (pour *Fluid Attenuated Inversion Recovery*) est une séquence T2 dont le signal provenant du liquide céphalo-rachidien est atténué. Elle fait apparaître les lésions parenchymateuses en hyper-signal. Dans le cas des AVC au réveil (pour lequel le début des

symptômes n'est pas connu), le différentiel de lésion observé entre les imageries FLAIR et diffusion permet d'estimer le temps écoulé, et de guider les indications de traitement.

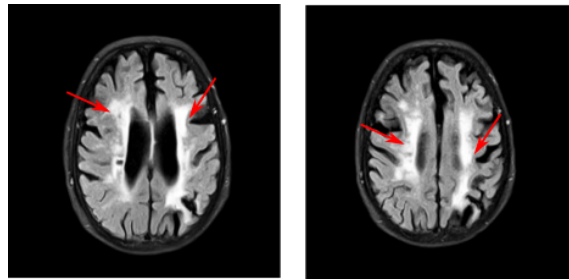


FIGURE I.1.5 – Imagerie par résonance magnétique FLAIR aiguë d'un patient atteint d'AVC. Les deux coupes font apparaître en hyper-signal une leucaraïose étendue et un ancien accident vasculaire cérébral lacunaire.

— **L'angiographie IRM par temps de vol : détecter l'occlusion**

L'angiographie (MRA pour *Magnetic resonance angiography*) par temps de vol est une séquence qui permet de voir, sans injecter de produit de contraste, le flux sanguin circulant dans les artères intracrâniennes. L'importance du signal vasculaire dépend notamment de la vitesse et du type de flux, permettant ainsi de déterminer la localisation de l'occlusion artérielle.

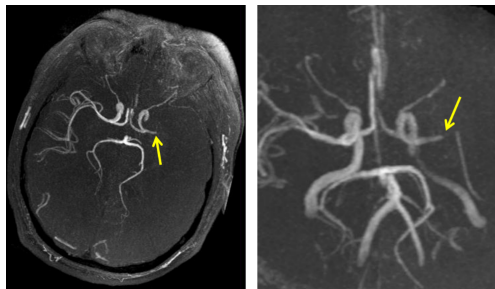


FIGURE I.1.6 – Angiographie IRM par temps de vol d'un patient atteint d'AVC présentant une occlusion de l'artère cérébrale moyenne gauche. L'image de gauche correspond à une reconstruction axiale de coupe 2D, et l'image de droite correspond une reconstruction 3D du cercle de Willis.

— **IRM pondérée de diffusion : évaluer l'infarctus**

La séquence de diffusion (DWI pour *Diffusion Weighted Imaging*) est une imagerie dont le contraste est influencé par les différences de mobilité de molécule d'eau. Les molécules d'eau diffusent rapidement dans les tissus sains, en apparaissant en hypointensité, tandis que leur mouvement est restreint au niveau de l'infarctus en apparaissant en hyperintensité. La lésion visible sur l'imagerie de diffusion est communément assimilée au coeur ischémique, à haut risque d'évolution vers la nécrose. Pendant l'acquisition, il est nécessaire d'appliquer les gradients de diffusion dans au moins 3 directions de l'espace. La magnitude de la diffusion calculée à partir des trois images de diffusion ainsi obtenues donne l'image pondérée en diffusion globale (image *trace*). À partir de deux séquences de diffusion avec des degrés de pondération différents, on peut calculer le coefficient de diffusion apparent (ADC pour *Apparent Diffusion Coefficient*), une cartographie quantitative du degré de mobilité des molécules d'eau (en mm^2/s), faisant apparaître l'infarctus en hypointensité.

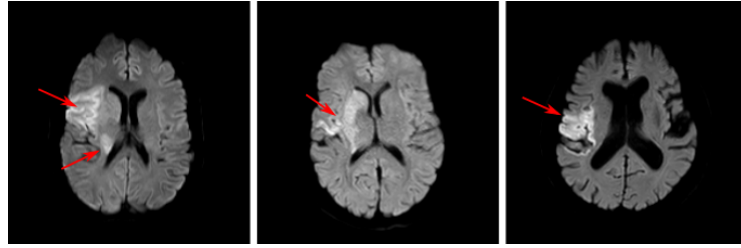


FIGURE I.1.7 – Imagerie par résonance magnétique pondérée de diffusion de trois patients atteints d'AVC. Sur chaque coupe, la lésion ischémique est visible en hyper-signal.

— **IRM de perfusion : évaluer la zone de pénombre**

La séquence de perfusion (PWI pour *Perfusion Weighted Imaging*) est une imagerie T2* dynamique (3D+temps) qui rend compte de la microcirculation capillaire des tissus (on parle d'état hémodynamique des tissus). La différenciation des tissus perfusés repose sur l'injection d'un bolus d'agent de contraste au gadolinium dans la circulation, dont on peut suivre le signal en fonction du temps (DSC : dynamic susceptibility contrast). La chute du signal lors du passage de l'agent de contraste permet d'extraire, après post-traitement, les paramètres quantitatifs de perfusion, qui sont classiquement : le temps au pic (TTP pour *Time To Peak*, en s), le débit sanguin cérébral (CBF pour *Cerebral Blood Flow*, en ml/100 g de tissu/min), le volume sanguin cérébral (CBV pour *Cerebral Blood Volume*, en ml/100 g de tissu), le temps de transit moyen (MTT pour *Mean Transit Time*, en s), le délai entre l'arrivée de l'agent contraste au niveau de l'entrée artérielle et au niveau des tissus (T_{max} , en s). L'estimation de ces paramètres est détaillé en section I.1.3.1.

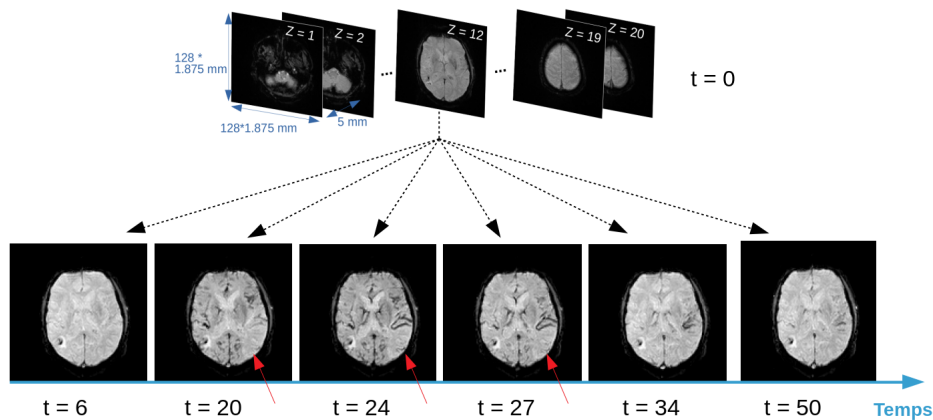


FIGURE I.1.8 – Imagerie par résonance magnétique pondérée de perfusion d'un patient atteint d'AVC. Le passage de l'agent de contraste dans les tissus est illustré pour une coupe donnée ($z=12$). Au moment du pic de passage de l'agent de contraste ($t=24$), au niveau de la flèche rouge, on voit que le produit d'agent de contraste en hypo-signal circule difficilement par rapport aux tissus environnants : il s'agit de la zone de pénombre.

IRM de suivi FLAIR : évaluer l'infarctus final

L'étendue de la lésion évolue durant les jours suivant le début de l'ischémie cérébrale. Une séquence FLAIR effectuée entre 24h et plusieurs semaines après l'admission peut rendre compte de l'étendue de l'infarctus final (lésion en hypersignal). Très schématiquement, on distingue deux scénarios possibles d'évolution de la lésion finale. Dans le scénario favorable, la lésion finale se limite au cœur ischémique observé à l'admission (la pénombre ischémique a pu être sauvée). Dans un scénario moins favorable, la zone de pénombre n'a pas pu être sauvée et l'infarctus s'est étendu.

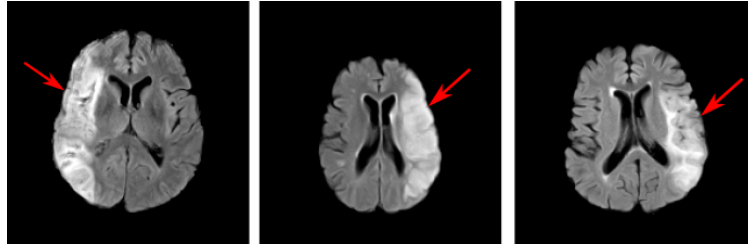
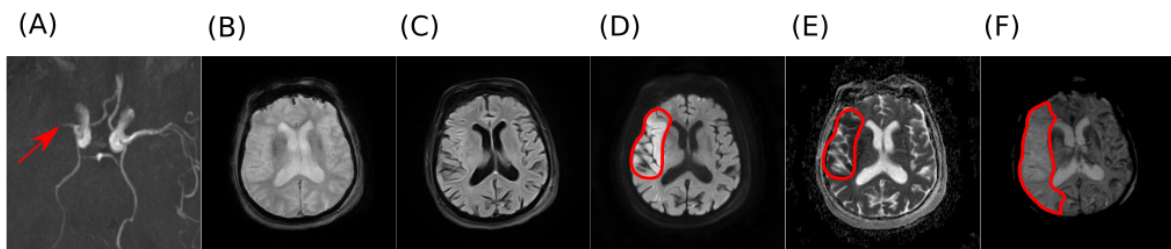


FIGURE I.1.9 – Imagerie de suivi FLAIR de trois patients atteints d'AVC. La lésion ischémique est visible en hyper-intensité.

IRM à l'admission



IRM de suivi

(G)

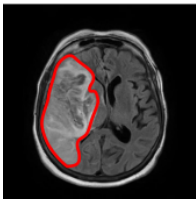


FIGURE I.1.10 – Étude de cas afin d'illustrer les IRM intervenant dans l'examen clinique de l'AVC. Depuis l'angiographie (A), on localise l'occlusion au niveau de l'artère cérébrale moyenne droite (segment M1). Aucun tissu hémorragique n'est détectable sur la séquence T2* (B). Pas de trace de lésion parenchymateuse sur la séquence FLAIR (C). L'infarctus est visible en hyper-intensité sur la séquence de diffusion (D) et en hypo-intensité sur la séquence ADC (E). La perfusion à un temps donné (F) rend compte d'une zone hypoperfusée plus étendue que la lésion aiguë de diffusion. L'évolution finale de la lésion est visible en hyper-intensité sur l'IRM FLAIR de suivi (G).

I.1.3 Modélisations actuelles de l'évolution de l'infarctus final ischémique en clinique

I.1.3.1 Déconvolution de l'imagerie de perfusion en clinique

Dans la pratique clinique actuelle, l'imagerie de diffusion est connue pour être l'imagerie qui porte le potentiel prédictif le plus important pour détecter le cœur ischémique. L'imagerie de perfusion DSC-MRI quant à elle permet de distinguer la pénombre ischémique. Bien qu'elle soit d'une utilité majeure pour la prédiction de l'infarctus final, l'imagerie de perfusion demeure sous-utilisée en clinique du fait de sa complexité et des verrous qui la caractérise.

L'imagerie de perfusion correspond à un ensemble de données en 4 dimensions (un volume en 3 dimensions dont on fait l'acquisition au fil du temps). Cela peut être visualisé comme un film dans lequel le tissu en aval d'une occlusion montre généralement une arrivée retardée du bolus d'agent de contraste. Lors du passage du bolus dans les vaisseaux, le signal enregistré diminue proportionnellement à la concentration de l'agent de contraste (cf courbe A de la Figure I.1.11). Ainsi, l'imagerie de perfusion associe à chaque voxel une courbe temporelle d'intensité du signal

en résonance magnétique. Le profil de la courbe dépend non seulement du profil du bolus à son arrivée dans le cerveau (au niveau de l'entrée artérielle) mais aussi de l'état de perfusion propre au tissu du voxel considéré.

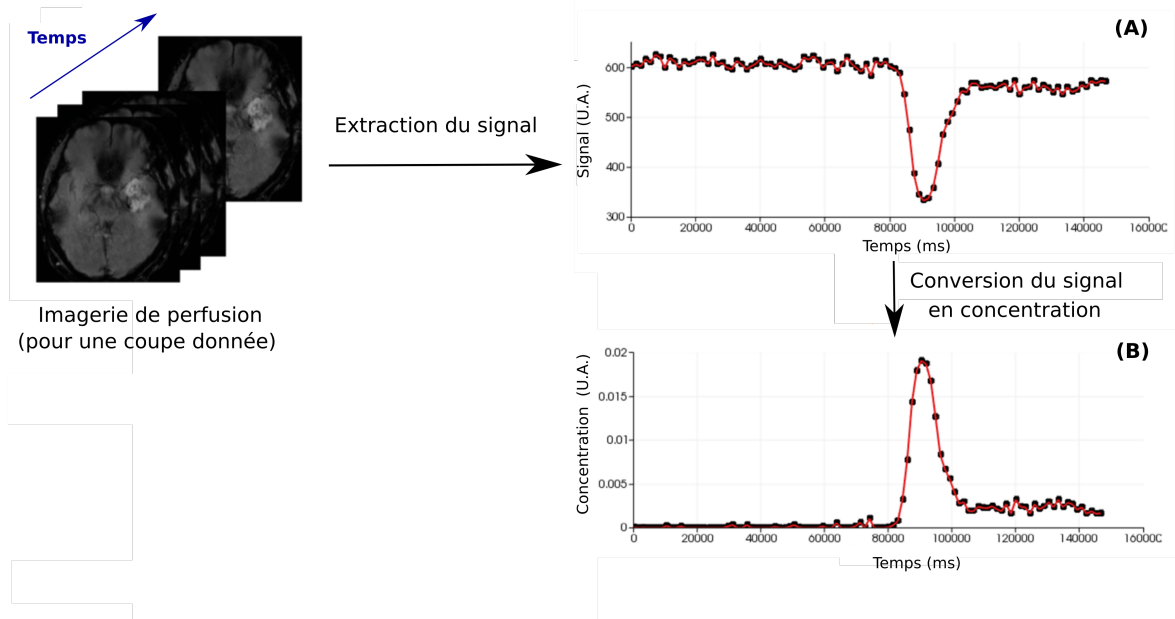


FIGURE I.1.11 – Estimation de la courbe moyenne de concentration d'agent de contraste depuis le signal de perfusion.

La courbe signal peut être convertie en courbe de concentration tissulaire en agent de contraste, (*cf* courbe B de la Figure I.1.11) sous l'hypothèse que la concentration en agent de contraste est proportionnelle au taux de relaxation transverse et que l'intensité du signal pour une séquence pondérée en T2* a une relaxation mono-exponentielle. Formellement, la concentration en agent de contraste tissulaire C peut être estimée par la relation logarithmique suivante :

$$C(t) = -\frac{1}{TE} \ln \frac{I(t)}{I(t_0)} \quad (\text{I.1.1})$$

avec TE le temps d'écho de l'IRM, $I(t)$ l'intensité de signal IRM d'un pixel en fonction du temps t , et t_0 le premier temps d'acquisition avant l'arrivée du bolus d'agent de contraste.

La courbe en agent de contraste produit des paramètres fortement influencés par la forme de la fonction d'entrée artérielle (AIF pour *Arterial Input Function*). L'AIF présente une grande variabilité inter-patients (selon le protocole d'injection de contraste ou l'état cardiaque du sujet), et à moindre mesure une variabilité intra-patient (selon l'emplacement choisi, l'AIF mesurée sur l'imagerie de perfusion d'un même patient peut plus ou moins varier). La dépendance de la courbe de concentration C à l'AIF rend alors difficile toute comparaison intra et inter-patients. Afin de pallier ce problème, la courbe de concentration peut être déconvoluée par l'AIF selon la formule suivante :

$$C_a(t) = \kappa \text{CBF} \int_0^1 \text{AIF}(\tau) R(t - \tau) d\tau \quad (\text{I.1.2})$$

avec κ une constante proportionnelle à la valeur de l'hématocrite dans les capillaires et à la densité locale du tissu cérébral, CBF le flux sanguin cérébral, AIF la fonction d'entrée artérielle et $R(t)$ la réponse impulsionnelle caractéristique de la micro-circulation des tissus. Contrairement à la courbe de concentration tissulaire C, la courbe de concentration déconvoluée C_a n'est pas dépendante de l'AIF. Pour chaque voxel, des biomarqueurs quantitatifs peuvent être extraits de la courbe déconvoluée C_a qui renseignent sur les propriétés hémodynamiques des tissus, comme illustrés en Figure I.1.12. Le CBF peut être directement estimé à partir de l'Équation I.1.2 en prenant le maximum de la réponse impulsionnelle :

$$CBF = \max(R(t)). \quad (I.1.3)$$

Le CBV est calculé comme le ratio entre le volume total d'agent de contraste entrant et le volume total d'agent de contraste mobile au travers du voxel d'intérêt :

$$CBV = \frac{\int_0^{\infty} C_a(t) dt}{\int_0^{\infty} AIF(t) dt}. \quad (I.1.4)$$

Le MTT est le temps moyen que le sang passe au sein d'un voxel d'intérêt :

$$MTT = \frac{CBV}{CBF}. \quad (I.1.5)$$

Le T_{max} est défini comme le moment où $R(t)$ atteint son maximum tel que :

$$T_{max} = \arg \max_t (R(t)). \quad (I.1.6)$$

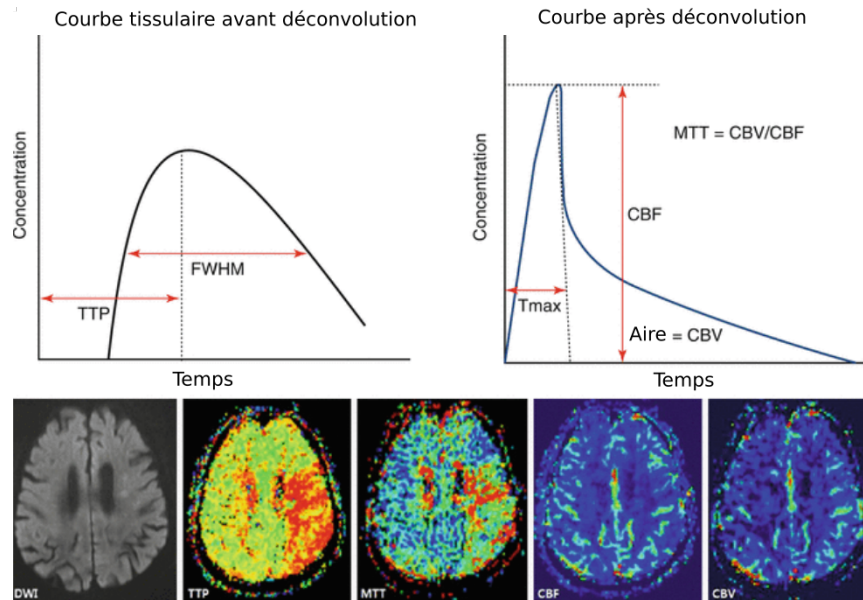


FIGURE I.1.12 – Courbe tissulaire de concentration d'agent de contraste avant et après déconvolution. Plusieurs cartes de paramètres de perfusion sont présentées (temps de pic du TTP, temps de transit moyen du MTT, débit sanguin cérébral du CBF, volume sanguin cérébral du CBV). *Illustration adaptée de l'article Stroke. 2014;16:131-45.*

Ces biomarqueurs (CBF, CBV, MTT, T_{max}) sont considérés prédictifs de la santé du tissu et sont actuellement utilisés dans la pratique clinique pour prédire l'infarctus final. Cependant, un certain nombre de difficultés se posent lors de la résolution de l'Équation I.1.2. En raison du bruit expérimental, la déconvolution est connue pour être un problème inverse mal posé, ce qui signifie que des AIF de formes différentes peuvent aboutir à des courbes de concentration tissulaire similaires [4]. Plusieurs approches de déconvolution ont été proposées dans la littérature ces dernières années pour tenter de résoudre l'Équation I.1.2 [5, 6, 7], mais aucune méthode spécifique n'a abouti sans présenter un minimum de lacunes techniques.

I.1.3.2 Le mismatch perfusion-diffusion : méthode clinique admise pour guider les procédures thérapeutiques

Actuellement, l'approche du *mismatch* perfusion-diffusion [8] est la méthode admise en clinique pour prédire l'infarctus final. Cette méthode est basée sur des seuils pour délimiter le cer-

veau encore récupérable (c'est-à-dire la pénombre ischémique) du tissu déjà perdu (cœur ischémique). Les critères pour délimiter le cœur ischémique et la pénombre sont le coefficient de diffusion apparent (ADC, extrait de l'imagerie pondérée par la diffusion) $< 600 \sim 620 \times 10^{-6} \text{ mm}^2/\text{s}$, et la cartographie $T_{max} > 6$ secondes, respectivement. En seuillant les imageries de diffusion et de perfusion, deux profils de patients se dégagent :

- les patients avec un **profil "target mismatch"**, qui présentent une grande pénombre et un cœur ischémique limité. Ils ont une forte probabilité de bénéficier d'une reperfusion, même dans des fenêtres temporelles tardives. Ils sont d'excellents candidats à une thérapie endovasculaire et présentent un bon pronostic clinique en cas de reperfusion. Un exemple de patient au profil "target mismatch" est donné en Figure I.1.13 A.
- les patients avec un **profil radiologique malin**, qui ne présentent pas de discordance significative entre la pénombre et le cœur ischémique, et dont la lésion de diffusion est volumineuse. Ces patients ont un mauvais pronostic et ont moins de chance de bénéficier d'une reperfusion cérébrale. Le rapport bénéfique/risque d'un traitement pour ces patients reste à l'heure actuelle incertain [9, 10]. Un exemple de patient au profil malin est donné en Figure I.1.13 B.

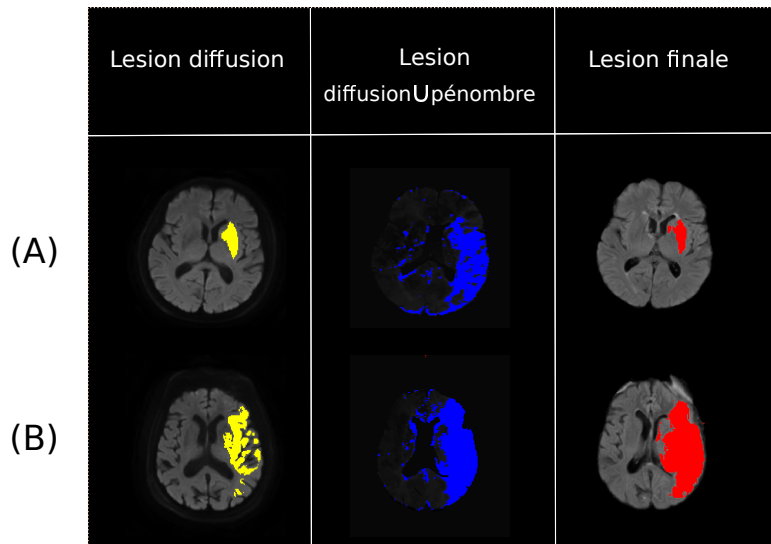


FIGURE I.1.13 – Exemples cliniques de deux patients traités par thrombectomie présentant des évolutions différentes. La colonne 1 correspond à l'IRM aiguë de diffusion, la colonne 2 à l'IRM aiguë de perfusion, et la colonne 3 à l'IRM de suivi FLAIR. Le patient A présente une discordance significative entre la lésion de diffusion (en jaune) et la pénombre (en bleu). Sur l'IRM de suivi, on voit que le patient a bénéficié d'une reperfusion : la pénombre a été sauvée et la lésion finale (en rouge) s'apparente à la lésion de diffusion aiguë. Le patient B ne présente pas de discordance significative entre la lésion de diffusion et la pénombre, et sa lésion finale s'apparente à l'union de la lésion aiguë de diffusion et de la pénombre ischémique (en bleu). Ce patient n'a pas bénéficié d'une reperfusion.

Comme en témoigne la Figure I.1.13, l'information de reperfusion est la métadonnée clinique la plus importante connue pour influencer la progression des lésions ischémiques depuis les imageries aiguës jusqu'à l'infarctus final [11]. Dans ce contexte, un modèle très simplifié, dit modèle mismatch perfusion-diffusion, est admis en clinique. Selon ce modèle, l'infarctus final inféré peut être défini comme suit : chez les patients reperfusés, l'infarctus final est représenté par la lésion de diffusion aiguë (*c.f.* patient A de la Figure I.1.13) ; chez les patients non reperfusés, l'infarctus final est défini comme l'union de la lésion aiguë de diffusion et de la pénombre ischémique (*c.f.* patient B de la Figure I.1.13).

Cependant, l'évolution de la lésion finale est beaucoup plus complexe et variable que le modèle clinique mismatch largement simplifié. On sait par exemple que les lésions aiguës de diffusion ne correspondent pas nécessairement à la nécrose. Des études IRM longitudinales ont mon-

tré que près de 30 à 50% du volume lésionnel initial observé en imagerie de diffusion était réversible, même sans reperfusion. De nombreux cofacteurs interviennent également dans l'évolution de la pénombre ischémique, contribuant à la complexité du pronostic des malades. Le facteur le plus important influençant les différences d'évolution inter-individus est probablement la qualité de la circulation sanguine collatérale autour de la zone ischémisée. Parmi les autres facteurs, on compte aussi le site d'occlusion artériel, la tolérance à l'ischémie selon le tissu considéré, les facteurs génétiques, l'âge. On compte également des facteurs qui ne dépendent pas du dossier clinique des patients, mais qui sont relatifs à leur prise en charge : plus le délai du traitement est long (temps entre les symptômes et l'admission à l'hôpital, ainsi que le temps entre l'admission et les premiers examens IRM), moins bonnes seront les chances de récupération.

Ces éléments témoignent de la complexité de la maladie et de la difficulté de la tâche de prédiction, que de simples modèles linéaires peuvent difficilement modéliser. Nos travaux explorent en profondeur plusieurs modèles d'apprentissage automatique, incluant notamment des approches d'apprentissage classique et réseaux de neurones.

Chapitre I.2

Champs d'investigation et contributions

Nous détaillons dans cette section les trois grands fronts de recherche autour desquels s'articule nos contributions. Dans la première partie du manuscrit, nous proposons des travaux fondamentaux sur la valeur prédictive de l'imagerie de perfusion. Puis dans une seconde partie, nous proposons des travaux plus appliqués où nous investiguons différentes méthodes capables de coupler les imageries de perfusion et de diffusion pour la prédiction de l'infarctus final. Ces deux parties tentent de répondre à la tâche de prédiction de l'AVC ischémique dans l'urgence (phase A sur la Figure I.2.1). Comme ouverture, nous nous intéressons à la phase chronique de la maladie (phase B sur la Figure I.2.1). Nous étudions de manière exploratoire comment assurer un meilleur suivi et accompagnement des patients après un AVC, pendant leur rééducation.

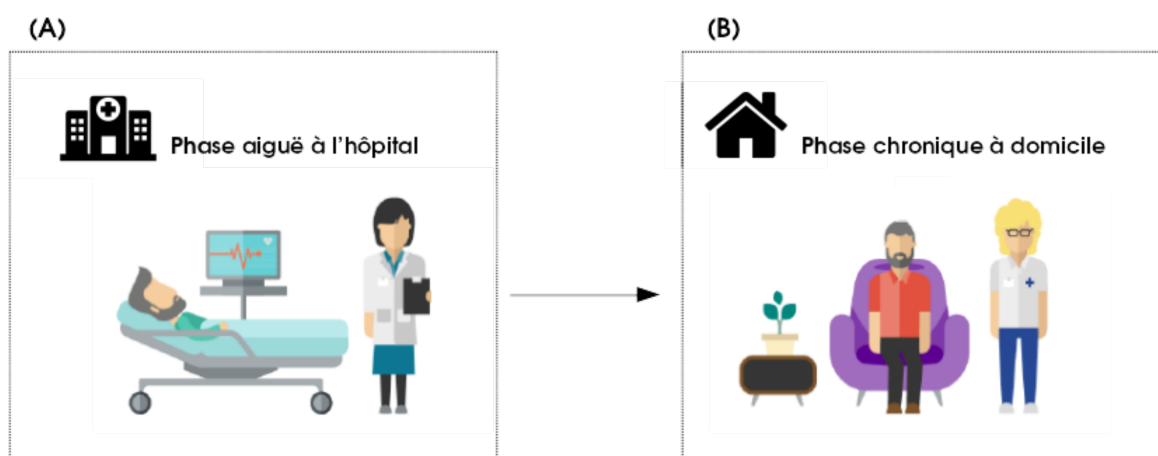


FIGURE I.2.1 – Nécessité de prendre en charge les patients victimes d'accident vasculaire cérébral : dès l'admission à l'hôpital, dans la phase aiguë (A) ; après la sortie de l'hôpital, dans phase chronique (B).

I.2.1 Intérêt de l'imagerie de perfusion brute pour la prédiction de l'infarctus final

Dans la pratique courante clinique, l'IRM de perfusion n'est jamais étudiée sous sa forme brute, mais est systématiquement analysée après avoir été déconvoluée au préalable. Comme nous l'avons introduit précédemment, la déconvolution est connue pour être un problème mathématique mal posé, ce qui est problématique étant donné le faible rapport signal/bruit des images de perfusion. Par conséquent, le calcul des biomarqueurs hémodynamiques extraits du signal déconvolué est instable. Dans le domaine de l'imagerie des accidents vasculaires cérébraux aigus, il existe depuis plusieurs années une certaine controverse concernant les avantages de l'utilisation de paramètres basés sur la déconvolution par rapport à la quantification à l'aide de paramètres de la courbe non déconvoluée, suggérant que la prédiction de l'infarctus basée sur des paramètres

non déconvolués est aussi bonne (et, dans certains cas, même meilleure) que les prédictions basées sur les paramètres déconvolués [12, 13, 14]. Récemment, plusieurs travaux ont proposé d'utiliser des techniques d'apprentissage automatique afin de prédire l'infarctus final à partir des IRM d'admission [15, 16, 17]. Cependant, la majorité des approches utilisent en entrée de leur modèle d'apprentissage les biomarqueurs standards de perfusion. Une nouvelle problématique a alors émergé : est-il possible de prédire la lésion finale en utilisant les données de perfusion brutes en entrée des modèles d'apprentissage machine, contournant ainsi l'étape de pré-traitement standard de déconvolution? Quelques études y ont répondu très récemment en montrant le bénéfice d'apprendre sur des données brutes de perfusion [18, 19, 20]. Dans ce contexte, nous proposons de répondre aux problématiques suivantes : compte tenu de la complexité dimensionnelle de la perfusion, quels nouveaux encodages et descripteurs peuvent être proposés pour caractériser la perfusion brute? Ces descripteurs sont-ils en mesure de capturer l'information pertinente pour la prédiction de la lésion ischémique?

La faible quantité de données annotées constitue aujourd'hui une limite à l'utilisation d'algorithmes d'apprentissage machine supervisé. Cela peut être considéré comme un goulot d'étranglement, en particulier pour des algorithmes d'apprentissage dépendant d'un grand nombre de paramètres comme les réseaux de neurones. Un moyen de remédier à ce problème est de générer davantage de données grâce à la simulation. Dans la première partie de ce manuscrit, nous proposons de prédire la lésion ischémique à partir d'IRM de perfusion brute simulées. Ces IRM synthétiques ont été produites à partir d'un simulateur physique récemment publié [21], capable de fournir un nombre de données annotées illimité. Parmi les nombreux paramètres physiologiques paramétrables du simulateur, nous nous focalisons sur celui de l'entrée artérielle. En ajustant les données d'entraînement au plus près des données test selon ce critère, nous proposons un apprentissage patient-spécifique. Dans ce contexte, nous proposons de répondre aux questions générales suivantes : peut-on recourir à la simulation physique de l'imagerie de perfusion afin de constituer une large base de données d'entraînement? Injecter des a priori physiques dans le jeu de données d'entraînement peut-il conduire à contraindre un réseau de neurones purement stochastique? Ces contraintes peuvent-elles simplifier la convergence du réseau? Comment évaluer la validité de la simulation physique? La réponse à ces questions fait l'objet de la première partie du manuscrit.

1.2.2 Enjeux de la fusion des données multimodales pour la prédiction de l'infarctus final

Dans la seconde partie du manuscrit, nous proposons de coupler l'imagerie de perfusion à l'imagerie de diffusion pour répondre à la tâche de prédiction de l'infarctus final. Comment fusionner une imagerie volumique (l'imagerie de diffusion) avec une imagerie dynamique (l'imagerie de perfusion) au sein d'un modèle d'apprentissage? Compte tenu de l'hétérogénéité dimensionnelle des deux imageries aiguës, nous avons pris le parti de ne plus considérer l'imagerie de perfusion sous sa forme brute mais de la réduire à ses biomarqueurs déconvolués. Déconvoluer l'imagerie de perfusion présente certes des inconvénients - comme ceux soulevés dans la section précédente - mais cette étape reste un pré-traitement qui fait consensus dans la littérature clinique. Pour notre application, le choix de déconvolution nous permettait avant tout de rapporter l'imagerie de perfusion à des cartes paramétriques 3D, nous ouvrant ainsi à davantage de possibilités méthodologiques de couplage avec l'imagerie de diffusion.

Dans le domaine de l'imagerie médicale, l'utilisation de méthodes de fusion d'images multimodales permet d'offrir une plus grande diversité de caractéristiques (*features*) pour les applications d'analyse [22]. Dans le cadre de notre application, les informations portées par les imageries de diffusion et perfusion sont complémentaires. Fusionner les deux modalités constituait donc un apport évident pour prédire l'infarctus final. Dans cette section, nous introduisons les concepts fondamentaux caractérisant la fusion de données.

La fusion de données est une opération qui combine une ou plusieurs sources de données

provenant de différentes positions et à différents moments dans le but d'obtenir une meilleure fiabilité et meilleure précision pour des problèmes de classification, détection, ou prédiction [23, 24]. Dans la littérature, il est courant de répertorier la fusion de données en trois niveaux d'abstraction [25] : au niveau de l'image, au niveau des caractéristiques (*features*) et au niveau de la décision. Nous détaillons les principales caractéristiques de ces trois catégories de fusion. Une illustration est donnée en Figure 1.2.2.

Fusion au niveau de l'image : La fusion au niveau de l'image (également appelé fusion au niveau du pixel) est le plus bas niveau de couplage possible. Ce type de fusion fait référence à la fusion pixel à pixel d'images pré-traitées et recalées. Ainsi, pour chaque pixel, une nouvelle valeur de donnée est calculée en utilisant plusieurs opérateurs mathématiques combinant les informations d'autres images au pixel. Dans le domaine de l'imagerie médicale, ce type de fusion a été proposé pour coupler des données de nature différente : par exemple pour fusionner des données scanner avec des données IRM [26, 27], des données TEP avec des données IRM [28], ou encore des données de vibro-acoustographie avec des données rayons X [29].

Fusion au niveau des caractéristiques : La fusion au niveau des caractéristiques combine des caractéristiques cibles extraites de chaque image individuellement pour produire une image (ou information) fusionnée plus informative. Autrement dit, ce type de fusion cible les caractéristiques extraites de chaque image d'entrée plutôt que leur valeur d'origine au niveau des pixels. En fusionnant ces caractéristiques, un espace de description commun est créé qui est potentiellement plus informatif et capable de cibler les variables d'intérêt de l'environnement hétérogène et complexe de départ. Ce type de fusion a été appliqué avec succès dans le domaine de l'imagerie médicale pour la fusion d'images multimodales, et en particulier depuis l'avènement des réseaux de neurones convolutionnels (CNN). Bien que la plupart des techniques de segmentation multimodale existantes de CNN opèrent une stratégie de fusion précoce au niveau du pixel (c.a.d. les images sont concaténées en tant que canaux dès l'entrée du réseau), un certain nombre d'études proposent de fusionner les caractéristiques à un plus haut niveau du réseau [30, 31].

Fusion au niveau de la décision : La fusion au niveau de la décision s'effectue à un niveau encore plus élevé que les deux autres catégories présentées. Elle consiste à extraire des caractéristiques des images d'entrée en utilisant différentes règles de décision externes, pour ensuite fusionner directement les différentes décisions. En observant individuellement les caractéristiques via différents outils de décision, on les analyse sous de multiples aspects avant de les fusionner. La fusion de décision permet ainsi une interprétation commune des différentes règles de décision. Ce type de fusion est dit de haut niveau, dans le sens où la prise de décision finale dépend fortement de la qualité des caractéristiques extraites en amont par chaque règle de décision. Dans le domaine de l'imagerie médicale, la plupart des études qui opèrent ce type de fusion proposent des méthodes dites ensemblistes, combinant les décisions de différents modèles pour obtenir une décision finale robuste et précise [32, 33, 34, 35].

Dans la seconde partie du manuscrit, nous avons investigué dans un premier temps une approche par classification bayésienne naïve basée voxel fusionnant les modalités au niveau de la décision. Cette étude pilote sur un nombre restreint de patient discute de l'influence de l'échelle d'observation de chaque voxel et de la taille de l'ensemble d'entraînement. Dans un second temps, nous avons travaillé de façon plus approfondie une approche réseau de neurones fusionnant tardivement les modalités d'entrée (fusion au niveau des caractéristiques). Au travers de cette seconde étude, nous discutons de l'impact de l'homogénéité de jeux d'entraînement, et notamment des paramètres physiologiques permettant sa stratification.

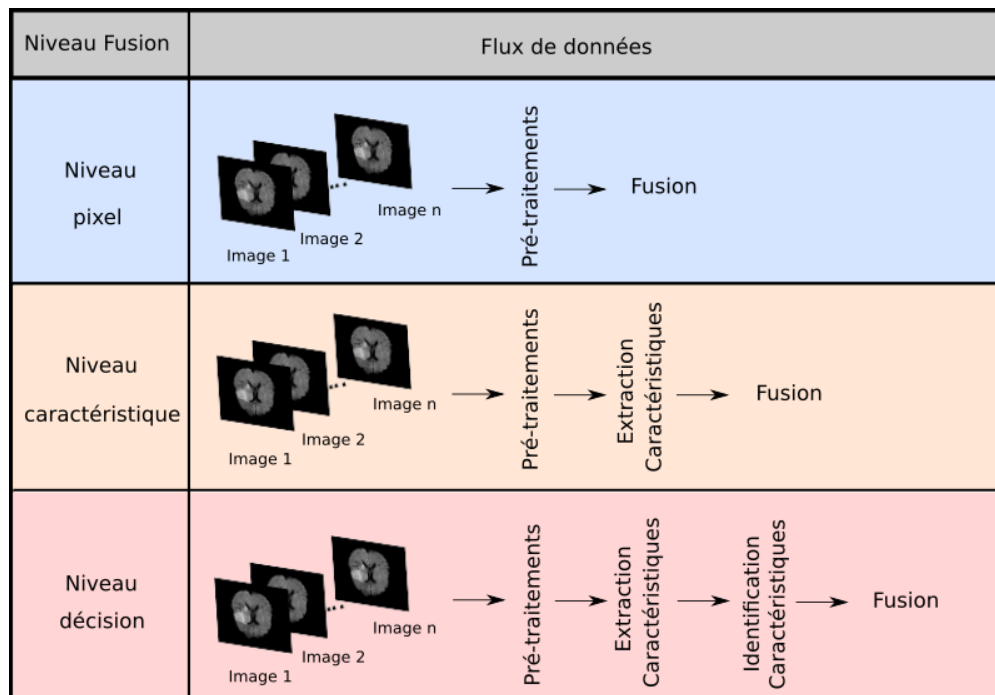


FIGURE I.2.2 – Les trois niveaux classiques de fusion d’images. *Illustration inspirée de l’ouvrage CRC (2018). Multisensor data fusion and machine learning for environmental remote sensing.*

I.2.3 Nouvelles prises en charge de la phase chronique de l’accident vasculaire cérébral

Avec l’amélioration de la prise en charge des AVC au cours des dernières décennies dans les pays industrialisés, les patients survivent maintenant plus longtemps, mais avec des états de santé compliqués : un AVC entraîne des incapacités motrices permanentes dans 80% des cas [36]. Bien que l’apparition de la thrombectomie depuis 2015 a permis d’améliorer l’état clinique des patients [37], la grande majorité nécessite des thérapies de rééducation. Aussi, d’un point de vue épidémiologique, les "baby-boomers" atteignent aujourd’hui un âge autour duquel le risque d’AVC est particulièrement accentué (>65 ans). Par conséquent, on peut prévoir que de plus en plus de patients seront amenés à suivre des thérapies de rééducation dans les prochaines années. Ces données soulèvent des questions fondamentales, généralisables même à d’autres pathologies chroniques : comment prendre soin d’un nombre croissant de personnes souffrant d’handicap moteur et nécessitant un suivi médical ? Comment assurer un tel suivi dans les déserts médicaux ?

En complément de ces questions relatives à la gestion de la prise en charge, on peut s’interroger sur l’efficacité des thérapies actuelles de rééducation motrice. Pendant les stades aigus et subaigus (<6 mois après un AVC), les patients reçoivent des thérapies de réadaptation dans des centres de santé spécialisés, consistant en un ensemble d’exercices dont l’objectif est de gagner au maximum en autonomie. Après avoir quitté le centre de réadaptation (c’est-à-dire après être entré dans le stade chronique, environ 6 mois après un AVC), seuls quelques rendez-vous médicaux ponctuels sont prévus afin de suivre l’évolution du patient. Pendant cette phase chronique, l’absence de suivi continu entrave une récupération optimale : une fois de retour à domicile, environ 65% des patients sont incapables d’intégrer à nouveau les membres affectés dans les activités de la vie quotidienne [38]. Cela montre la nécessité de nouvelles options thérapeutiques permettant aux patients de s’entraîner de manière intensive et approfondie après avoir quitté le centre spécialisé, tout en garantissant la qualité, l’efficacité et la sécurité du traitement.

Avec l’émergence de l’internet des objets (IoT), de plus en plus de personnes sont équipées d’un smartphone (60% de la population française) et d’objets connectés (bracelet de suivi d’activité, montre intelligente) [39] remontant des informations sur l’activité de la personne voire sur

sa physiologie. Les études récentes tendent à montrer que les mesures de ces capteurs sont suffisamment précises pour retourner des informations fiables concernant les utilisateurs [40]. Ces technologies font donc l'objet de nombreuses recherches ces dernières années dans le domaine médical, en particulier de la rééducation en phase chronique de l'AVC, dans le but de les intégrer dans le processus de surveillance d'activité à domicile. Utilisés de manière régulière, les capteurs portables peuvent réduire les temps d'évaluation et fournir des données objectives et quantifiables sur les capacités physiques et physiologiques des patients, complétant le jugement expert des spécialistes de santé rééducateurs. Les capteurs ont également le potentiel de fournir des informations continues relatives aux activités de la vie quotidienne. Ces nombreuses informations peuvent ainsi être utilisées pour analyser finement le dossier médical du patient, ce qui peut ensuite conduire à des soins spécifiques mieux ciblés (vers une médecine plus personnalisée).

À l'heure où les capteurs sont assez précis pour fournir des services personnalisés, les utilisateurs peuvent raisonnablement craindre que le respect de leur vie privée soit compromis [41]. L'environnement complexe dans lequel les données sont collectées multiplie les risques de sécurité et de confidentialité, au moment de la collecte et de la transmission des données [42], ainsi que du traitement et du stockage [43]. Quelles méthodes peuvent être mises en place pour permettre de supprimer des données capteurs les informations relatives à l'identité, tout en préservant les informations relatives à l'activité? Quel est l'équilibre optimal entre l'utilité des données (c'est-à-dire la reconnaissance de l'activité) et la confidentialité des données (c'est-à-dire éviter la ré-identification des utilisateurs)? Nous proposons une méthode capable de répondre à ces questions en ouverture du manuscrit.

Partie II

Valeur de l'imagerie de perfusion brute pour la prédiction de l'infarctus final

Chapitre II.3

L'encodage spatio-temporel de l'image de perfusion

Dans ce chapitre nous introduisons une approche d'encodage de l'IRM de perfusion brute, qui décrit l'information spatio-temporelle de chaque voxel en un patch 2D de petite dimension. Des descripteurs de texture sont ensuite extraits de ces patches, et une machine à vecteurs de support est entraînée pour réaliser une prédiction voxel-à-voxel de l'état final des tissus. Ce travail récemment publié [19] a constitué la base de recherche de la première partie de ce manuscrit. À la fin de ce chapitre, nous soulevons différentes questions relatives à la pertinence de l'encodage proposé et des descripteurs associés, auxquelles nous répondons tout au long de la première partie de ce manuscrit.

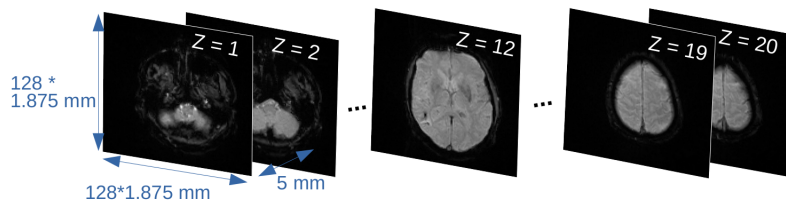
II.3.1 Encodage de la signature spatio-temporelle locale de chaque voxel

Un nouvel encodage de l'IRM de perfusion a été proposé par Giacalone et al., motivé par le fait que la signature spatio-temporelle de chaque voxel est difficile à observer en raison de la nature 3D+temps de l'imagerie (*cf.* Figure II.3.1). Cet encodage propose de traduire la signature spatio-temporelle en une texture discriminante facilement identifiable par l'œil humain et des outils de vision par ordinateur [19]. Cette texture témoigne de l'état des tissus dans chaque voxel, en encodant ses informations locales contenues dans le voisinage de Moore d'ordre 1.

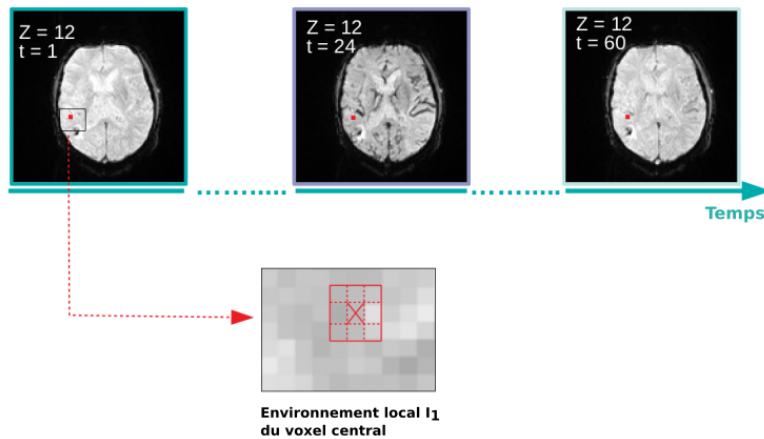
L'encodage spatio-temporel se caractérise par une imagerie 2D traduisant une information locale 3D (deux dimensions géométriques, c.a.d. largeur et hauteur, et une dimension temporelle). L'information contenue dans chaque voxel est encodée en déroulant les signaux des voxels temporels le long d'une dimension spatiale, et en empilant les signaux des voxels du voisinage (3×3) du voxel d'intérêt les uns à la suite des autres, créant ainsi une imagerie de taille $(9 \times N_t)$, où N_t est le nombre de points d'acquisition temporels dans la séquence d'imagerie de perfusion. Cette méthode d'encodage est illustrée dans Figure II.3.2, pour $N_t = 60$.

Giacalone et al. propose également de prendre en compte des informations provenant d'un plus large voisinage dans son encodage, qui s'affranchit de potentiels effets de bords. Il s'agit d'un pré-traitement consistant à lisser l'imagerie de perfusion par un filtre moyen de taille $N_p \times N_p$, avec N_p l'échelle d'observation du voisinage [19]. Les patches résultants de ce pré-traitement sont illustrés en Figure II.3.3. Cette étape préliminaire de lissage permet également de réduire le bruit contenu dans le patch qui perturberait l'efficacité de classification. Toutefois, si l'échelle d'observation N_p est trop grande, les informations locales pertinentes sont perdues, ce qui peut nuire aussi à la pertinence de classification.

Les signatures spatio-temporelles caractéristiques des voxels sains sont bien distinctes de celles des voxels pathologiques [19], en particulier en termes de texture. Comme le montre la Figure II.3.3, une tendance générale se dégage pour caractériser les voxels sains des voxels pathologiques. Les patches typiques obtenus à partir de voxels sains présentent une texture en hypo-intensité relativement étroite, qui traduit le passage rapide et local du bolus d'agent de contraste dans le tissu. Les



(a) Visualisation de la structure spatiale de l'IRM de perfusion. Seules 5 coupes ($z=1$, $z=2$, $z=12$, $z=19$ et $z=20$) pour un pas de temps donné ($t=1$) sont affichées.



(b) Visualisation de la structure temporelle de l'IRM de perfusion. Seuls 3 pas de temps ($t=1$, $t=24$ et $t=60$) pour une coupe donnée ($z=12$) sont affichés. Un voxel d'intérêt et son voisinage local d'ordre 1 sont indiqués en rouge.

FIGURE II.3.1 – Illustration de la signature spatio-temporelle des signaux IRM bruts en IRM de perfusion d'un patient donné. La séquence présentée a été acquise sur 20 coupes spatiales pendant 60 pas de temps consécutifs.

patches typiques obtenus à partir de voxels pathologiques présentent une texture en hypo-intensité relativement étalée et peu contrastée. Cela traduit un comportement localement plus erratique du bolus d'agent de contraste dans les tissus pathologiques.

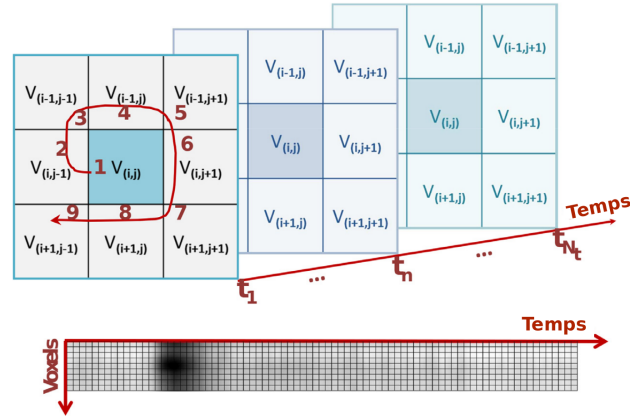


FIGURE II.3.2 – Encodage de la signature spatio-temporelle des signaux d’IRM de perfusion sous forme de patch.

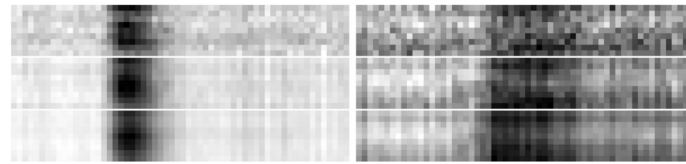


FIGURE II.3.3 – Illustration des signatures typiques obtenues pour les patches de voisinage $N_p = 1$, $N_p = 3$ et $N_p = 5$ (de haut en bas) pour les voxels sains (à gauche) et les voxels pathologiques (à droite).

II.3.2 Extraction de descripteurs de motifs binaires locaux

Giacalone et al. propose des descripteurs de motifs binaires (LBP pour *Local Binary Patterns*) introduits par Ojala et al. [44]. Les LBP ont été largement appliqués pour différentes tâches de classification dans le domaine de l’imagerie médicale, en particulier en tant que descripteurs de texture [45, 46, 47]. Une propriété importante du LBP est notamment sa robustesse aux variations d’intensité moyenne au sein d’une image. Une autre est leur simplicité de calcul qui rend possible leur utilisation dans des applications d’analyse d’images en temps réel.

Giacalone et al. propose ainsi d’utiliser les LBP pour décrire les signatures observées sur les patches. Les LBP correspondent à des descripteurs de texture qui décrivent l’environnement local de chaque voxel par un label calculé par seuillage simple des valeurs en niveau de gris de ses voxels environnants. L’idée de base des LBP est qu’une image est composée de micro-modèles et que l’histogramme de ces micro-modèles contient des informations sur la distribution des bords et d’autres caractéristiques locales de l’image. L’opérateur LBP conventionnel avec voisinage (L, R) [44] est calculé à chaque emplacement de voxel en considérant à la fois la valeur d’un voxel central (q_c) et les valeurs de ses voxels voisins L dans un périmètre de rayon R (q_l avec $l \in \{1, \dots, L\}$). Formellement, l’opérateur LBP de voisinage (L,R) est défini comme :

$$\text{LBP}_{(L,R)} = \sum_{l=1}^L s(q_l - q_c) 2^{l-1} \quad (\text{II.3.1})$$

où $s(x) = 1$ si $x \geq 0$ et 0 sinon. Il y a donc 2^L labels distincts résultant des différents motifs circulaires possibles autour de chaque voxel. On peut distinguer deux types de motifs dans les LBP : les motifs uniformes, qui présentent au maximum deux transitions de $s(q_l - q_c) = 0$ à $s(q_{l+1} - q_c) = 1$ (ou inversement), et les motifs non-uniformes. Ojala et al. [44] ont observé que les motifs uniformes constituent la majorité des textures de leurs images (les $\text{LBP}_{(8,1)}$ constituent un peu moins de 90% des motifs dans leurs images texturées). Giacalone et al. propose d’utiliser les LBP avec un voisinage (8,1) en se limitant aux descripteurs uniformes. Le nombre de labels résultants pour un patch est alors de 59 motifs (58 labels pour les motifs uniformes et 1 label pour les motifs non-uniformes)

fusionnés comme un même label).

Une fois que l'opérateur LBP a été appliqué sur un patch, il peut être décrit en calculant les histogrammes de ses sous-patches adjacents de largeur N_h . Soit N_t la largeur totale du patch, le nombre de sous-patches contigus est donc de N_t/N_h . Les histogrammes concaténés des sous-patches peuvent être alors utilisés comme vecteurs caractéristiques pour décrire la texture du patch, de longueur totale $N_t/N_h \times 59$. Les vecteurs caractéristiques peuvent ainsi servir de descripteurs pour classer chaque voxel en fonction de son état final (classe 1 pour les voxels pathologiques appartenant à la lésion finale ou classe 0 pour les voxels sains).

II.3.3 Évaluation des performances

II.3.3.1 Classification par machine à vecteur de support

L'apprentissage par machine à vecteur de support (SVM pour *Support Vector Machine*) consiste à trouver l'hyperplan séparateur permettant de séparer au mieux, dans l'espace des variables d'entrée (dit espace de description), des points d'entraînement associés à des classes (dans notre cas d'application, la classe est 0 ou 1). Les points d'entraînement les plus proches de la frontière de séparation forment les vecteurs supports, et définissent la marge (c'est-à-dire la distance entre l'hyperplan séparateur et les vecteurs supports). Avec les SVM, la frontière de séparation est choisie comme celle qui maximise la marge. Un nouveau point à tester est alors classé comme appartenant à la classe 0 ou 1 selon sa position par rapport à l'hyperplan de séparation. Plus le point est éloigné de l'hyperplan séparateur, plus le niveau de confiance accordé à sa classification est élevé.

Trouver un hyperplan séparateur suppose que les points des deux différentes classes sont linéairement séparables. Cependant dans de nombreuses situations, les deux classes ne sont pas séparables linéairement dans l'espace de description. Dans ce cas, il est possible de projeter les points d'entraînement dans un nouvel espace de variables (dit espace de redescription), de dimension plus élevée que l'espace de départ, dans lequel il est possible de trouver un séparateur linéaire. En pratique, le calcul de projection de données est réalisé par l'intermédiaire d'une fonction noyau (ou fonction *kernel*), qui permet de simplifier le calcul de la transformation.

Ici, après des tests empiriques, un classifieur SVM avec un noyau polynomial de degré 3, a été choisi en prenant comme variables d'entrée les éléments du vecteur caractéristique LBP de chaque voxel. Afin de réduire la dimension du vecteur d'entrée, une analyse en composantes principales (ACP) a été réalisée au préalable pour conserver les vecteurs propres principaux contenant 90% de la variance totale. Un diagramme résumant l'approche de classification proposée est donné à la Figure II.3.4.

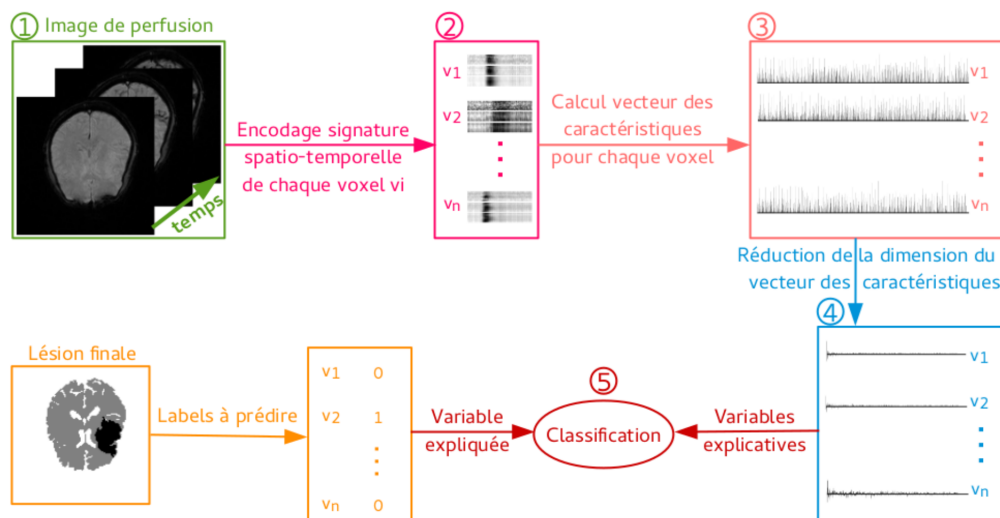


FIGURE II.3.4 – Pipeline de l'approche de classification proposée.

II.3.3.2 Matériel

Des données IRM ont été extraites d'une cohorte de patients acquise aux Hospices Civils de Lyon [48]. Le travail a été effectué sur les données longitudinales de quatre patients victimes d'un accident ischémique cérébral de la circulation antérieure. Ces quatre patients n'ont pas reçu de traitement et n'ont pas reperfusé d'eux-mêmes. Tous les patients ont reçu le même protocole IRM à l'admission : une séquence d'imagerie pondérée par diffusion (DWI) ; une séquence FLAIR ; une séquence pondérée T2* ; une angiographie par résonance magnétique et une séquence de perfusion (DSC-PWI). Une IRM FLAIR de suivi a été effectuée à 1 mois. Une correction de mouvement a été appliquée sur l'IRM de perfusion brute. Coupe par coupe, tous les points de temps ont été recalés sur le premier point de temps par une approche d'information mutuelle. Cela a été fait en recalant chaque point (n+1) sur son point temporel précédent (n) en appliquant récursivement les matrices de transformation obtenues jusqu'à ce que tous les points soient alignés sur le premier point temporel. La segmentation du masque de la lésion finale a été réalisée pour chaque patient sur l'IRM de suivi FLAIR par des experts. L'IRM FLAIR a été recalée sur l'imagerie de perfusion. Après recalage, les lésions finales ont été binarisées en appliquant un seuil de 50%.

II.3.4 Résultats

II.3.4.1 Optimisation de la largeur des segments

Le Tableau II.3.1 montre l'influence de la largeur de segment N_h utilisée pour le calcul des vecteurs de caractéristiques sur la performance de la classification. De plus, N_f , le nombre de descripteurs conservés après ACP, est fourni comme mesure de la complexité du modèle. La largeur de segment N_h influence grandement la qualité de la classification obtenue. Les performances générales augmentent lorsque N_h diminue, ce qui se traduit par une différence de précision pouvant atteindre 15% entre une largeur de segment de $N_h = 60$ et $N_h = 3$. Le bolus d'agent de contraste observé sur nos données dans les tissus sains a une largeur d'environ 10 voxels. La largeur de segment optimale trouvée ici, $N_h = 3$, est donc approximativement un tiers de la largeur du bolus de l'agent de contraste, ce qui semble raisonnable. Les motifs encodés par les LBP peut être considérés comme la concaténation de trois sous-motifs : un sous-motif sombre correspondant au passage du bolus, au milieu de deux sous-motifs plus clairs qui correspondent principalement à du bruit. Une valeur N_h trop élevée mélangerait les variations locales des sous-motifs produisant, comme le montre le Tableau II.3.1, une diminution des performances de prédiction.

N_h	Accuracy (%)				Sensibilité (%)				Spécificité (%)				N_f
	Min	Moy	Max	σ	Min	Moy	Max	σ	Min	Moy	Max	σ	
60	58.32	66.46	75.11	3.34	49.53	61.59	71.79	4.79	59.83	71.34	82.61	4.59	22
30	61.09	68.81	76.47	3.27	55.00	68.04	78.38	4.79	59.26	69.55	79.65	4.45	50
20	58.37	69.51	78.73	3.39	55.14	68.71	80.91	5.22	60.87	70.33	79.63	4.41	79
15	65.61	72.07	78.73	2.83	60.95	71.48	80.85	4.28	60.55	72.69	81.42	3.91	111
12	68.33	74.40	81.00	2.84	64.22	74.20	84.11	4.52	65.77	74.59	81.90	3.54	143
10	68.78	74.78	82.35	3.14	65.09	74.17	84.68	4.58	63.54	75.43	84.91	4.12	176
6	71.49	78.09	85.97	2.72	66.38	77.85	88.79	4.25	69.37	78.33	87.61	3.70	309
5	71.49	79.05	85.52	2.59	68.97	79.07	87.07	3.95	69.72	79.01	88.50	3.74	379
4	72.40	80.21	85.97	2.73	70.69	80.03	87.27	4.07	70.87	80.37	88.50	3.64	473
3	75.11	80.21	88.24	2.71	72.41	81.43	89.81	3.71	73.33	82.16	91.15	3.56	659

TABLEAU II.3.1 – Précision (%), Sensibilité (%), Spécificité (%) obtenues par validation croisée 100-Folds pour des patches de type I_1 (échelle d'observation $N_p = 1$). Différentes tailles N_h (largeur du segment) sont évaluées. N_f correspond au nombre de descripteurs sélectionnés après réduction de dimension par ACP.

II.3.4.2 Observation de l'échelle optimale

La largeur du segment est fixée à la valeur optimale trouvée dans la section précédente, à savoir $N_h = 3$, afin d'étudier l'influence de l'échelle d'observation N_p utilisée pour le calcul du patch. Les résultats sont présentés dans le Tableau II.3.2. Il apparaît que la performance moyenne de classification du modèle augmente et que sa variabilité diminue à mesure que l'échelle d'observation N_p augmente jusqu'à $N_p = 9$. En observant des échelles supérieures à 9, la performance semble atteindre un plateau et son augmentation semble négligeable par rapport à la variabilité des résultats obtenus. L'optimisation de l'échelle d'observation de $N_p = 1$ à $N_p = 9$ permet d'améliorer la précision de 13%.

Type de patch	Accuracy (%)				Sensibilité (%)				Spécificité (%)				Nf
	Min	Moy	Max	σ	Min	Moy	Max	σ	Min	Moy	Max	σ	
I ₁	75.11	81.81	88.24	2.71	72.41	81.43	89.81	3.71	90.18	82.16	91.15	3.56	659
I ₃	85.07	89.85	95.02	2.11	81.65	89.62	96.58	3.26	81.82	90.07	95.10	2.85	607
I ₅	89.14	93.43	96.38	1.63	85.71	93.65	98.28	2.41	88.46	93.20	97.64	2.09	545
I ₇	90.05	94.73	98.19	1.46	88.99	95.23	99.13	1.84	85.96	94.22	99.12	2.21	494
I ₉	92.31	95.04	98.19	1.28	90.76	95.38	100.00	1.80	89.81	94.70	100.00	1.92	454
I ₁₁	91.40	95.26	98.64	1.50	89.92	95.77	100.00	2.07	87.85	94.72	99.09	2.32	420
I ₁₃	91.86	95.37	98.19	1.39	88.24	96.02	100.00	1.83	90.18	94.70	98.28	2.00	391

TABLEAU II.3.2 – Précision (%), Sensibilité (%), Spécificité (%) obtenues par validation croisée 100-Folds pour $N_h = 3$ et différents types de patch I_{N_p} sont évalués. N_f correspond au nombre de descripteurs sélectionnés après réduction de dimension par ACP.

II.3.4.3 Impact du post-traitement de l'IRM de perfusion

Les résultats présentés jusqu'à présent ont été obtenus à partir de la dynamique brute de l'imagerie de perfusion sans autre post-traitement que le recalage décrit dans la section Matériel. Le pipeline proposé s'affranchit donc de l'étape habituelle de déconvolution par la fonction d'entrée artérielle. Dans une dernière expérience, Giacalone et al. compare les résultats de classification obtenus à partir d'imagerie de perfusion brutes à celles obtenues en utilisant les cartographies déconvoluées conventionnelles. Pour cette expérience, une déconvolution de Tikhonov temporelle standard [1] a été utilisée pour la comparaison avec les signaux de perfusion bruts. Les résultats présentés dans le Tableau II.3.3 ont été obtenus en utilisant les valeurs N_h et N_p optimales trouvées dans les expériences précédentes. Il apparaît que les performances en termes de précision, de spécificité et de sensibilité sont très proches les unes des autres. Le gain potentiel lié à l'utilisation d'images post-traitées se situe peut-être dans la taille du vecteur caractéristique associé qui est plus petite (350 variables caractéristiques pour les données déconvoluées contre 454 pour les données de perfusion brutes).

	Accuracy (%)				Sensibilité (%)				Spécificité (%)				Nf
	Min	Moy	Max	σ	Min	Moy	Max	σ	Min	Moy	Max	σ	
PWI	92.31	95.04	98.19	1.28	90.76	95.38	100.00	1.80	89.81	94.70	100.00	1.92	454
CAC	90.50	94.59	99.10	1.49	89.29	95.18	99.10	1.99	88.89	94.03	99.17	2.02	365
IRF	90.05	93.94	97.74	1.56	89.62	94.05	98.21	1.87	87.85	93.87	99.15	2.38	350

TABLEAU II.3.3 – Précision (%), sensibilité (%), spécificité (%) obtenues par validation croisée 100-Folds pour $N_h = 3$ et $N_p = 9$. N_f correspond au nombre de descripteurs sélectionnés après réduction de dimension par ACP. Ces résultats comparent la qualité de la prédiction obtenue en utilisant : des signaux de perfusion bruts PWI en entrée du modèle (1ère ligne), l'image de concentration de l'agent de contraste CAC (2ème ligne), ou l'image de concentration déconvoluée avec l'algorithme de Tikhonov IRF [1] (3ème ligne).

II.3.5 Discussion et questions ouvertes

II.3.5.1 Indépendance entre les données d’entraînement et les données de validation

La validation du modèle proposé se fait par validation croisée 100-Folds sans séparer distinctement des patients d’entraînement de patients de validation. L’absence d’indépendance entre les données d’entraînement et les données de validation induit un risque très élevé de sur-apprentissage, d’autant plus que le nombre de patients d’étude est restreint (4 patients). Ce détail pourrait expliquer que les performances observées soient particulièrement élevées, à savoir une métrique accuracy entre 75 et 90 %. Est-ce que l’encodage proposé serait toujours aussi adapté en assurant une indépendance entre les données d’entraînement et les données de validation? Dans la suite de nos travaux, nous assurons cette indépendance, en veillant à ce que nos modèles soient systématiquement entraînés avec des patients différents des patients de validation et de test. D’autre part, nous évaluerons préférentiellement nos modèles à l’aide du coefficient de similarité Dice (DSC) [49] plutôt que l’accuracy, en raison du déséquilibre de classes de l’ensemble test (beaucoup plus de voxels sains à évaluer que de voxels pathologiques).

II.3.5.2 Détermination de l’environnement local optimal

Giocalone et al. suggère une échelle d’observation N_p optimale d’ordre 9 sur la base de quatre patients d’étude [19]. Notons que ces quatre patients présentent tous une lésion volumineuse et compacte, et un niveau de bruit similaire dans leur séquence d’acquisition. Cependant à l’échelle d’une cohorte plus large, il apparaît plus difficile d’établir une échelle d’observation optimale, dans la mesure où toutes les séquences d’acquisitions ne présentent pas le même niveau de bruit, et les volumes et formes de lésions sont très différents d’un patient à l’autre. Afin d’appuyer notre propos, nous avons illustré la diversité de forme de lésion de la base I-KNOW sur la Figure II.3.5. Les patients sont représentés selon trois descripteurs de forme (le logarithme du volume total de la lésion en mm^3 , le rapport entre le volume de la lésion et le volume de sa boîte englobante et le nombre de composantes connexes constituant la lésion). L’axe des abscisses est corrélé au volume, tandis que l’axe des ordonnées est corrélé au nombre de composantes connexes, tels que les clusters 1 et 2 contiennent des patients dont la lésion est volumineuse et compacte, tandis que le cluster 3 contient des patients avec une lésion éclatée et peu volumineuse. Les quatre patients d’études, entourés en rouge, présentent bien des lésions relativement volumineuses et compactes au regard des autres patients de la cohorte.

Notons également qu’avec l’encodage proposé, deux patches voisins d’un patient donné ne sont pas indépendants dans la mesure où ils sont chevauchants : nous évoquons là une autre forme d’indépendance que celle mentionnée dans paragraphe II.3.5.1, du fait que deux patches voisins partagent des voxels en commun. Dans l’ensemble de nos travaux, nous avons fait le choix de rester sur un voisinage de Moore d’ordre 1 afin de conserver le signal le plus brut possible et de permettre une meilleure indépendance entre les patches, au risque d’être sensible au bruit lors de l’étape de classification. On peut alors se demander : quels types d’erreurs de classification obtient-on lorsqu’on limite le contexte spatial de chaque voxel à son voisinage d’ordre 1 ?

II.3.5.3 Choix du descripteur

En tant que méthode non paramétrique, l’opérateur LBP résume efficacement les structures locales des images en comparant chaque pixel avec ses voisins. Son inconvénient majeur est qu’il génère un vecteur caractéristique de grande taille. Afin d’obtenir un plus petit ensemble de descripteurs discriminants, il est possible de réduire la dimension du vecteur caractéristique LBP par des techniques de réduction de dimension tel que l’ACP. Notons que les LBP sont des descripteurs invariants en translation. Dans le cadre de l’application présentée, ils se révèlent particulièrement adaptés car ils sont insensibles à l’information non discriminante de délai contenue dans les patches spatio-temporels (c.a.d. le temps entre l’administration du produit d’agent de contraste et l’arrivée dans les tissus). Cependant est-il possible d’utiliser des opérateurs plus parcimonieux

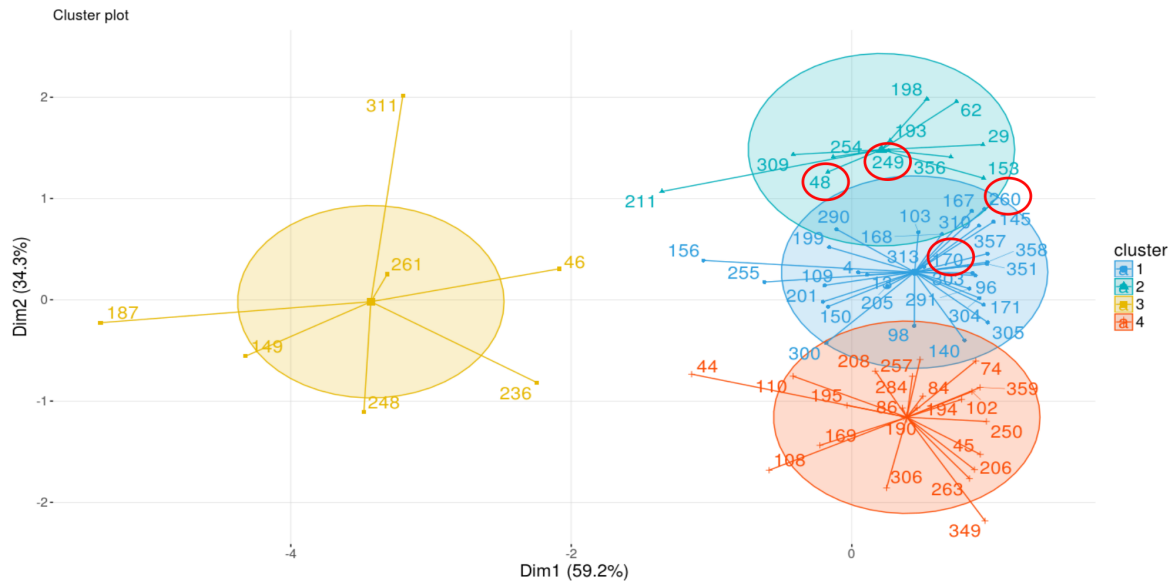


FIGURE II.3.5 – Visualisation de la répartition des 4 patients choisis (encerclés en rouge) parmi les 77 patients I-Know à disposition sur un plan de projection 2D obtenu après ACP sur 3 descripteurs de forme (la boîte englobante de la lésion, le logarithme de la lésion et le nombre de composantes connexes de la lésion). Les deux axes sont les deux premiers vecteurs propres de l'ACP et constituent 95% de la variance totale. Chaque patient a été assigné à un groupe en utilisant un clustering K-Means. Chaque groupe est représenté par un nombre (1, 2, 3, 4) et une couleur (respectivement bleu, vert, jaune, rouge). Les ellipses représentent les enveloppes convexes qui rassemblent les patients d'un même groupe.

et translation invariants qui s'affranchissent d'une étape préalable de réduction de dimension? Nous nous sommes alors intéressés à la matrice de co-occurrence : un opérateur mono-échelle simple capable de résumer efficacement des informations de texture en un nombre restreint de descripteurs au travers des coefficients d'Haralick [50]. D'autre part, nous nous sommes également interrogés sur la valeur des descripteurs mono-échelles que sont les LBP et les matrice de co-occurrence. Est-ce-que des opérateurs multi-échelles plus élaborés seraient capables d'améliorer notre tâche de classification? Pour y répondre, nous avons investigué l'opérateur multi-échelle de la transformée en scattering [51, 52], potentiellement capable d'analyser aussi bien l'information transitoire que les textures des patches spatio-temporels.

Chapitre II.4

Nouvelles approches de description de l'information spatio-temporelle

Dans ce chapitre, nous explorons d'autres descripteurs manuels capables d'encoder les patches spatio-temporels de l'imagerie de perfusion (introduits dans le chapitre précédent) [19]. Nous proposons d'abord d'analyser les patches à l'aide de la matrice de co-occurrence [50], une matrice créée sur la base des corrélations entre les voxels. Tout comme les LBP, la matrice de co-occurrence est une méthode structurale, qui décrit l'organisation spatiale des niveaux de gris d'une image. Mais à la différence des LBP, les coefficients retenus correspondent aux propriétés statistiques texturales de l'image. Aussi, le vecteur caractéristique résultant de la matrice de co-occurrence est de dimension moins grande que les LBP. Dans un second temps, nous décrivons les patches spatio-temporels par une approche fréquentielle au travers de la transformée en scattering multi-échelle récemment introduite pour la classification des textures [51, 52]. Cette toute autre approche consiste à analyser les patches comme des signaux transitoires, qu'on perturbe sur plusieurs échelles au moyen d'une famille d'ondelette. Les coefficients issus des combinaisons de transformation sont alors utilisés comme descripteurs pour l'étape de classification.

II.4.1 Les coefficients d'Haralick

Haralick et al. [50] ont proposé d'utiliser une matrice de co-occurrence de niveau de gris (GLCM, pour *Grey Level Co-occurrence Matrix*) comme méthode de quantification de la relation spatiale de pixels voisins dans une image. Les caractéristiques de texture de Haralick, calculées à partir du GLCM, sont largement utilisées en raison de leur simplicité et de leurs interprétations intuitives, et ont été appliquées avec succès dans de nombreuses tâches de classification, en particulier pour l'analyse des images ultra-sonores et IRM du foie [53], des images IRM pour l'étude du cancer du sein [54, 55], du cancer de la prostate [56] et de tumeurs cérébrales [57].

Soit I une image donnée de taille $M \times K$ avec N niveaux de gris. La matrice GLCM notée X , avec $X \in \mathbb{N}^{N \times N}$, est calculée en comptant le nombre de fois où chaque paire de niveaux de gris se produit comme voisins de l'image I . Le voisin d'un pixel est défini par un vecteur de déplacement $\theta = (d_x, d_y)$ en deux dimensions, avec $d_x, d_y \in \mathbb{Z}$ qui représentent le déplacement en x et y en unités de pixels. Chaque élément $X(i, j)$ dans la matrice GLCM est calculé par :

$$X(i, j) = \sum_{m=0}^M \sum_{k=0}^K \begin{cases} 1 & \text{si } I(m, k) = i \text{ and } I(m + d_x, k + d_y) = j \\ 0 & \text{sinon.} \end{cases} \quad (\text{II.4.1})$$

avec $I(m, k)$ la valeur de l'image I en (m, k) . Il est possible de créer une matrice GLCM avec plusieurs vecteurs de déplacement $\delta(d_x, d_y)$. Par exemple, si l'on considère les 8 voisins d'un pixel en deux dimensions, il existe 8 déplacements possibles :

$$\begin{aligned}
\delta_0^\circ &= (1, 0) & \delta_{180}^\circ &= (-1, 0) \\
\delta_{45}^\circ &= (1, 1) & \delta_{225}^\circ &= (-1, -1) \\
\delta_{90}^\circ &= (0, 1) & \delta_{270}^\circ &= (0, -1) \\
\delta_{135}^\circ &= (-1, 1) & \delta_{315}^\circ &= (1, -1).
\end{aligned}$$

Haralick et al. [50] définit 14 attributs statistiques calculables à partir de la matrice GLCM, tels que l'entropie, la somme de variance, le second moment angulaire. Ces mesures sont utilisées dans de nombreuses applications radiomiques d'imagerie médicale en tant que descripteurs généraux de textures.

Dans le cadre de notre application, nous considérons l'image I comme un patch spatio-temporel. Nous souhaitons privilégier des déplacements dans deux directions particulières : un déplacement horizontal $\delta_0^\circ = (1, 0)$ afin de mesurer la relation des voxels voisins selon la dimension spatiale du patch, ainsi qu'un déplacement vertical $\delta_{90}^\circ = (0, 1)$ afin de mesurer la relation des voxels voisins selon la dimension temporelle du patch. En combinant les statistiques issus des deux matrices de co-occurrence, 28 attributs statistiques sont obtenus (1 matrice par direction, et 14 statistiques par matrice).

II.4.2 Les coefficients de transformées en ondelettes

Une ondelette est une fonction oscillante, de moyenne nulle, avec un certain degré de régularité et dont le support est fini. Une ondelette mère ψ peut être dilatée par un paramètre d'échelle s (ou contractée si $s < 1$) et translatée d'un paramètre de temps u , générant ainsi une famille d'ondelettes

$$\begin{aligned}
(u, s) &\in \mathbf{R} \times \mathbf{R}_0^+, \\
\psi_{u,s}(t) &= \frac{1}{\sqrt{s}\psi\left(\frac{t-u}{s}\right)}.
\end{aligned} \tag{II.4.2}$$

En faisant varier les paramètres d'échelles s et de temps u , tout signal x peut être analysé par ses coefficients d'ondelettes $g_{(u,s)}$ où :

$$\begin{aligned}
g(u, s) &= x * \psi_{u,s} \\
&= \int_{-\infty}^{\infty} x(t) \psi_{u,s}(t) dt
\end{aligned} \tag{II.4.3}$$

qui correspondent à la convolution du signal avec la famille d'ondelettes. Les coefficients d'ondelettes mesurent une information locale du signal x en temps et en fréquence. L'opération associant les coefficients d'ondelettes au signal est appelée transformée en ondelettes.

Dans des travaux récents, S. Mallat propose de caractériser des images présentant des textures par des coefficients issus d'une série de transformations d'ondelettes particulières. Pour une telle application, la transformée d'ondelettes est alors discrétisée en limitant les valeurs des paramètres (u, s) à un ensemble de rotations r d'angle entre $[0, \pi)$ et un ensemble de dilatations 2^{-j} (où $j \in \mathbb{Z}$) :

$$\psi_\lambda(t) = 2^{-2j} \psi(2^{-j} r^{-1} t) \text{ avec } \lambda = 2^{-j} r. \tag{II.4.4}$$

Le signal x est également discrétisé, et ne correspond plus à un scalaire (avec $\psi : \mathbf{R} \mapsto \mathbf{R}$) mais un signal vectoriel (avec $\psi : \mathbf{R} \mapsto \mathbf{R}^{N \times M}$, $M \times N$ étant la dimension d'une image deux dimensions). Pour notre application, les patches spatio-temporels étant de faible dimension, nous les avons interpolés linéairement afin d'éviter des effets de bords, passant d'une dimension (9×60) à (90×600) . Le signal d'entrée était donc tel que : $\psi : \mathbf{R} \mapsto \mathbf{R}^{90 \times 600}$.

La transformée en scattering proposée [51, 52] définit un propagateur de module d'ondelettes \tilde{W} qui calcule des coefficients d'ondelettes à partir d'un opérateur de filtrage gaussien ϕ à échelle 2^J et d'une ondelette de Morlet ψ opérant une dilatation de 2^{-j} et une rotation d'angle θ (avec $j = \{0, 1, \dots, J\}$ et $\theta = \{0, \pi/L, \dots, \pi(L-1)/L\}$, où $(J, L) \in \mathbb{Z} \times \mathbb{Z}$) tel que :

$$\tilde{W}x(t) = \{x * \phi_{2^j}(t), |x * \psi_{j,\theta}(t)|\}. \quad (\text{II.4.5})$$

En itérant sur \tilde{W} , on définit un réseau convolutionnel profond particulier que nous présentons schématiquement en Figure II.4.1.

BRUNA AND MALLAT: INVARIANT SCATTERING CONVOLUTION NETWORKS

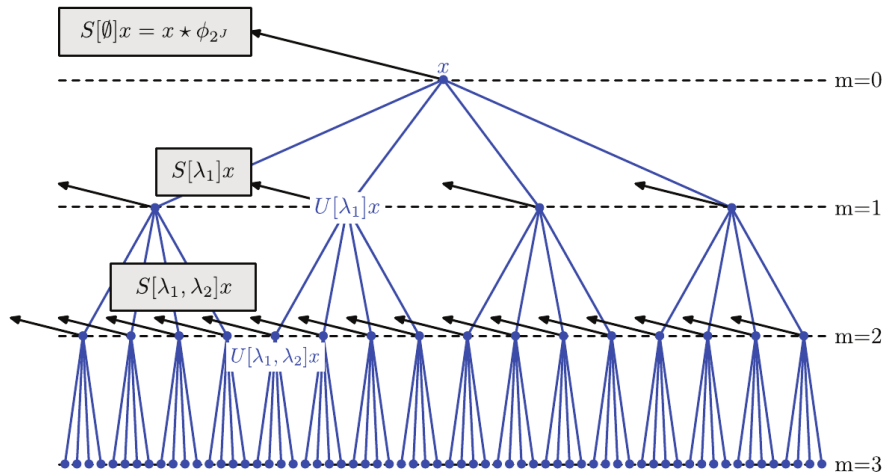


FIGURE II.4.1 – Un propagateur scattering \tilde{W} appliqué à une image x calcule la première couche du module des coefficients d'ondelette $U[\lambda_1]_x = |x * \psi_{\lambda_1}|$ et donne en sortie sa moyenne locale $S[0]x = x * \phi_{2^J}$ (flèche noire). De manière générale, en appliquant \tilde{W} sur chaque signal propagé $U[p]x$, on obtient en sortie $S[p]x = U[p]x * \phi_{2^J}$ (flèche noire). La variable m correspond à l'indice de couche. *Figure reproduite de l'article IEEE transactions on pattern analysis and machine intelligence (2013), 35(8), 1872-1886.*

Les coefficients en sortie du réseau (c.f. Figure II.4.1) représentent des descripteurs aux propriétés intéressantes, offrant notamment une représentation invariante en translation tout en permettant de préserver des informations haute-fréquences utiles à la classification. D'après l'Équation II.4.5, les coefficients S_m à la sortie des trois premières couches convolutives ($m = 0, 1, 2$) pour une image d'entrée x sont tels que :

$$\begin{aligned} S_0 &= x * \phi_{2^J} \\ S_1 &= |x * \psi(j_1, \theta_1)| * \phi_{2^J} \\ S_2 &= ||x * \psi(j_1, \theta_1)| * \psi(j_2, \theta_2)| * \phi_{2^J}. \end{aligned}$$

La transformée proposée diffère d'une décomposition en ondelettes pure en raison de l'opérateur de module non linéaire. Avec cette non-linéarité, la décomposition de l'image ne se fait pas sur une base orthonormale, ce qui offre un potentiel avantage lorsqu'on opère différentes décompositions d'ondelettes à différentes échelles et sur plusieurs couches. Il est intéressant de noter que ces propriétés spécifiques s'accordent avec la nature texturale multi-échelle intrinsèque des patches spatio-temporels, ce qui confirme l'utilisation appropriée de la transformée en scattering pour notre problème de classification. Dans la mesure où l'on cherche à extraire l'information spatiale distinctement de l'information temporelle des patches, nous avons fixé le nombre de rotations r à 2. Une visualisation des images de sortie est donnée : pour $r = 2$ et $j = 3$ en Figure II.4.2 ; pour $r = 2$ et $j = 5$ en Figure II.4.3. Il apparaît clairement que les textures discriminantes peuvent

être saisies avec ces deux facteurs d'échelle. Cependant on peut constater que plus le paramètre d'échelle augmente, plus la dimension des images diminue et plus l'information est moyennée. Pour cette étude, nous avons donc limité les valeurs prises par le paramètre de dilatation j , en le faisant varier de 3 à 5. D'après les recommandations de Bruna et al. [58], le nombre de couches nécessaires n'a pas besoin d'excéder $m = 2$, car au-delà, l'énergie contenue dans les couches est négligeable (la variabilité résiduelle stochastique, soit la variance des coefficients S_x , devient faible pour $m > 2$). Nous avons donc fixé le nombre de couches à $m = 2$.

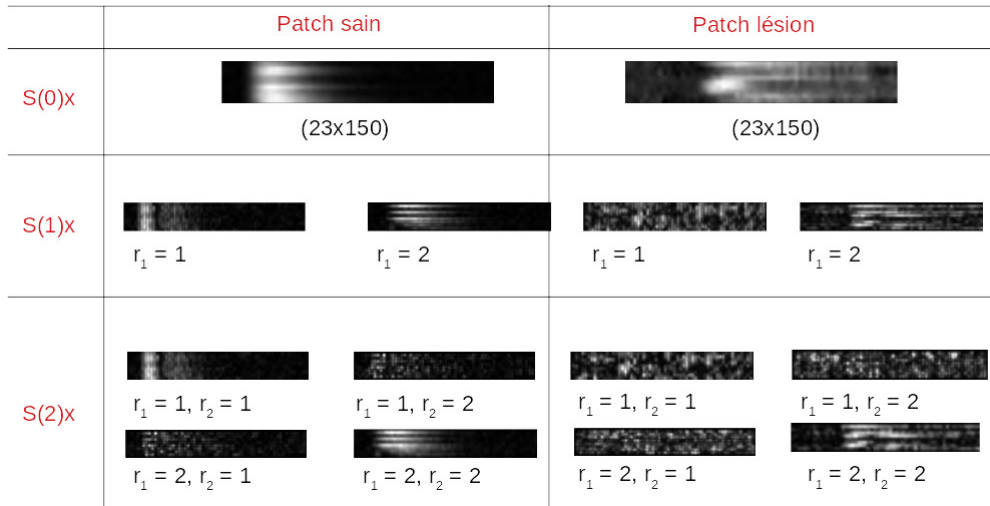


FIGURE II.4.2 – Images de sortie pour chaque classe (patch interpolé sain à gauche et patch interpolé lésionnel à droite) et pour trois couches m de la transformée en scattering, avec $r=2$ et $j=3$.

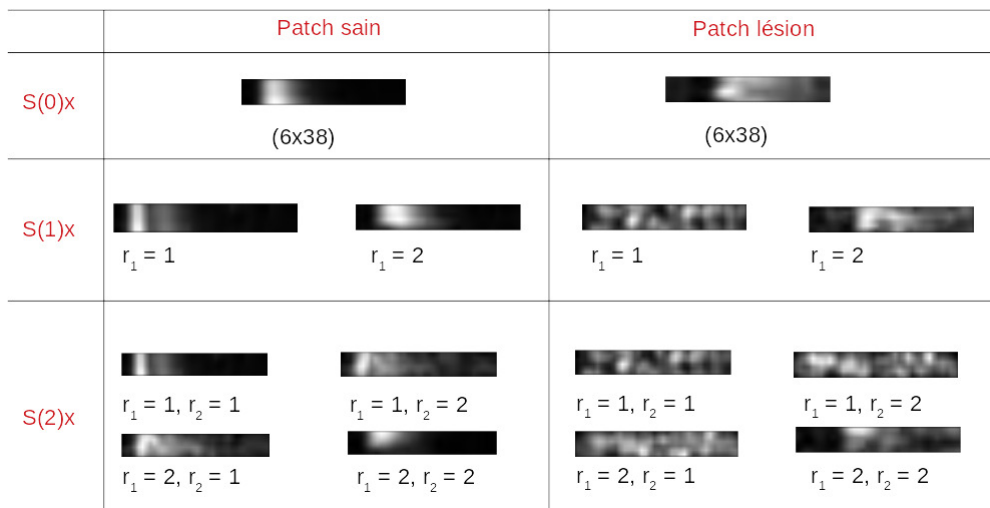


FIGURE II.4.3 – Images de sortie pour chaque classe (patch interpolé sain à gauche et patch interpolé lésionnel à droite) et pour trois couches m de la transformée en scattering, avec $r=2$ et $j=5$.

II.4.3 Évaluation des deux approches proposées

II.4.3.1 Matériel

Des données IRM ont été extraites d'une cohorte de patients acquise aux Hospices Civils de Lyon [48]. Nous avons travaillé sur les données longitudinales de dix patients victimes d'un accident ischémique cérébral de la circulation antérieure. Ces dix patients n'ont pas reperfusé d'eux-mêmes, et présentaient des lésions particulièrement compactes et volumineuses. Tous les patients

ont reçu le même protocole IRM à l'admission : une séquence d'imagerie pondérée par diffusion (DWI); une séquence FLAIR; une séquence pondérée T2*; une angiographie par résonance magnétique et une séquence de perfusion (DSC-PWD). Une IRM FLAIR de suivi a été effectuée à 1 mois. Une correction de mouvement a été appliquée sur l'IRM de perfusion brute. Nous avons recalé, coupe par coupe, tous les points de temps sur le premier point de temps par une approche d'information mutuelle. Cela a été fait en recalant chaque point (n+1) sur son point temporel précédent (n) en appliquant récursivement les matrices de transformation obtenues jusqu'à ce que tous les points soient alignés sur le premier point temporel. La segmentation du masque de la lésion finale a été réalisée pour chaque patient sur l'IRM de suivi FLAIR par des experts. L'IRM FLAIR a été recalée sur l'imagerie de perfusion. Après recalage, les lésions finales ont été binarisées en appliquant un seuil de 50%.

II.4.3.2 Encodage de l'imagerie de perfusion

Nous encodons l'imagerie de perfusion à l'aide de l'encodage patch spatio-temporels proposé par Giacalone et al. [19]. À la différence de Giacalone et al. qui génère les patches à partir de l'imagerie de perfusion brute, nous générons les patches à partir de l'imagerie de concentration de l'agent de contraste. Cette étape de conversion permet de s'affranchir de la baseline de l'AIF, et donc de garantir une meilleure normalisation inter-patients.

II.4.3.3 Classification par machine à vecteur de support et évaluation

Après plusieurs tests empiriques, un classifieur SVM à un noyau gaussien a été utilisé, en prenant comme variables d'entrée les coefficients propres à chaque patch. Aucun pré-traitement de réduction de dimension n'a été nécessaire. L'évaluation a été réalisée par validation croisée, à l'aide de l'approche "Leave-One-Out", c'est-à-dire que le modèle a été paramétré sur la base de N-1 patients et a ensuite été évalué sur la performance du patient exclu. Le processus a été répété N fois, et la performance finale est donnée par la moyenne de performance des N itérations. Cette approche maximise le nombre de données d'entraînement à chaque itération tout en garantissant l'indépendance entre les patients d'entraînement et les patients de validation. Pour constituer la base d'entraînement, nous avons systématiquement sélectionné au hasard pour chaque patient le même nombre de voxels sains que de voxels infarctés. Dans l'ensemble de validation en revanche, tous les voxels ont été évalués. Nous avons évalué nos résultats en utilisant le coefficient de similitude Dice (DSC) [49] et la distance de Hausdorff (HD) [59]. Ces métriques ont été calculées entre les voxels prédits comme infarctés et le masque de la lésion finale fourni par l'IRM FLAIR.

II.4.4 Résultats

Les performances obtenues à partir des coefficients d'Haralick sont indiquées sur les trois premières lignes du Tableau II.4.1. À titre de référence et de comparaison, il convient de noter que les modèles réseaux de neurone pour la prédiction de lésion ischémique dans la littérature présentent des DSC entre 0.3 et 0.5 [17, 60]. Les résultats obtenus à partir des coefficients d'Haralick peuvent être considérés satisfaisants, si l'on garde à l'esprit que le travail est réalisé sur des données de perfusion brutes et avec des descripteurs de texture très simples basés uniquement sur les 14 coefficients d'Haralick. Les performances obtenues à partir d'un déplacement horizontal (dimension temporelle) sont supérieures aux performances obtenues avec un déplacement vertical (dimension spatiale). Cela indique que la redondance spatiale locale comporte un degré de prédictibilité inférieur à l'évolution temporelle du signal de perfusion. Un incrément de performance est observé lorsque les deux déplacements sont considérés, ce qui indique que certaines informations issues de chaque déplacement sont non-redondantes et qu'il y a un intérêt à considérer les deux dimensions du patch pour la prédiction de la lésion.

Les performances obtenues à partir des coefficients de la transformée en scattering sont indiquées sur les trois dernières lignes du Tableau II.4.1. On observe un incrément de performance

lorsqu'on augmente le nombre de dilatations j . Ce résultat était attendu : en augmentant le nombre de déformations, l'espace des descripteurs est plus grand et peut potentiellement capturer plus d'informations. Cependant le nombre de déformations que nous pouvons tester est limité pour notre application : au-delà de $j = 5$, la dimension et la résolution des patchs deviennent trop faibles (c.f. Figure II.4.2 et II.4.3). De manière plus générale, nous sommes également limités par le temps de calcul : augmenter le nombre de déformations fait augmenter la combinatoire dans le réseau, et par extension le temps de calcul des descripteurs, comme en témoigne le Tableau II.4.1. À l'échelle d'une cohorte, il apparaît nécessaire de paralléliser le code sur d'avantages de CPU pour étendre l'ensemble de déformations.

Descripteur	Paramètres	# coefficients	DSC lésion	HD lésion	Temps calcul/patient (en min)
Haralick	$\delta = (0,1)$	14	0.31 ± 0.16	45.02 ± 3.55	<1
	$\delta = (1,0)$	14	0.34 ± 0.15	45.08 ± 3.56	<1
	$\delta = (0,1) + \delta = (1,0)$	28	0.35 ± 0.15	45.09 ± 3.56	<1
Scattering	$j=3, r=2$	19	0.31 ± 0.14	45.07 ± 3.55	11
	$j=4, r=2$	33	0.32 ± 0.14	45.07 ± 3.55	13
	$j=5, r=2$	51	0.33 ± 0.16	45.07 ± 3.58	18

TABLEAU II.4.1 – Coefficient Dice (DSC) et distance de Hausdorff (HD) de la lésion prédite obtenus à partir des deux descripteurs manuels proposés : les coefficients d'Haralick et les coefficients de la transformée en scattering. Les métriques sont moyennées sur l'ensemble de validation (moyenne \pm écart-type). Le temps d'extraction des descripteurs est précisé pour un patient donné (d'environ 32 500 patchs). Pour les coefficients scattering, le calcul a été parallélisé sur 4CPU d'une station standard.

On peut visualiser les prédictions obtenues pour deux patients de validation sur la Figure II.4.4. Quelle que soit l'approche utilisée, on distingue les deux mêmes types d'erreurs de classification. Le premier type d'erreur, facilement identifiable sans même vérité terrain, concerne les voxels faux positifs qui apparaissent isolés, ou éloignés de la lésion (typiquement les voxels en bordure du tissu cérébral, ou sur l'hémisphère contralatéral). L'agent de contraste passe de façon plus chaotique dans ces zones, et les voxels y présentent un signal particulièrement bruité. Le modèle détecte alors une perturbation du flux sanguin cérébral et la relie directement à l'accident ischémique cérébral final. Ce type d'erreur témoigne du manque de contexte spatial dans l'apprentissage : en augmentant la taille des patchs, nous serions moins sensibles à ces phénomènes locaux. Le deuxième type d'erreur, plus difficile à identifier sans vérité terrain, concerne des faux positifs proches de la lésion. On constate que la forme prédite ne correspond par tout à fait à la celle de la lésion finale. Cette différence témoigne de l'évolution de la lésion entre l'imagerie d'admission et l'imagerie de suivi (vérité terrain).

On constate également que les coefficients de la transformée en scattering semblent beaucoup plus sensibles au bruit : les voxels localisés en bordure du cerveau et dans les vaisseaux sont classifiés plus systématiquement lésionnels. On peut supposer que les descripteurs scattering portent donc une information plus fine que celle des coefficients d'Haralick, qui nuit à la classification. Ceci expliquerait pourquoi les performances générales sont moins bonnes en utilisant les coefficients de la transformée en scattering.

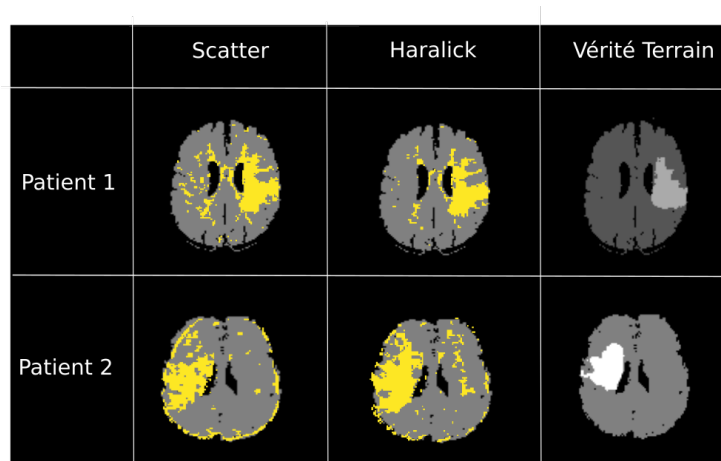


FIGURE II.4.4 – Prédictions obtenues pour deux patients de validation à partir des coefficients scattering (colonne 1) et des coefficients d’Haralick (colonne 2). Les prédictions sont à comparer avec la vérité terrain (colonne 3).

II.4.5 Discussion

Dans ce chapitre, nous avons proposé de résumer l’information spatio-temporelle contenue dans un patch à l’aide de descripteurs manuels que nous jugeons pertinents. Nous les avons choisis le plus judicieusement possible afin de représenter l’information utile pour répondre à la tâche de prédiction. Dans notre cas, la taille du vecteur caractéristique des descripteurs choisis est petite : le patch de taille (9x60) est réduit en un vecteur caractéristique optimal de taille 28 pour les descripteurs d’Haralick et 51 pour les coefficients scattering. Choisir un nombre réduit de descripteurs permet de mieux cibler le processus d’apprentissage. Si les descripteurs sont adaptés et pertinents, ils ont le potentiel de guider l’optimisation du modèle vers un optimum global.

En revanche, trouver des descripteurs adaptés constitue une étape supplémentaire dans la chaîne de traitement. Cela constitue non seulement un certain temps de recherche pour trouver le descripteur le plus pertinent au vu des données, mais aussi un biais de subjectivité dans l’apprentissage : on injecte dans le modèle un a priori fort vis-à-vis de la compétence du descripteur pour répondre à la tâche d’apprentissage. Le risque est donc de négliger à tort des informations qu’on ne jugeait pas utiles.

Enfin notons que le temps pour générer ces descripteurs doit être considéré dans le cadre de notre application, à savoir diagnostiquer et traiter le plus rapidement possible les patients dès leur admission à l’hôpital. Pour un patient donné (d’environ 32 500 patches), les descripteurs de la matrice de co-occurrence sont générés en moins d’une minute, contre 18 minutes avec le réseau scattering ($j = 5, r = 2$). Vis-à-vis de ce critère, les descripteurs Haralick sont mieux adaptés.

Ces éléments nous ont motivé à tester des approches supervisées qui s’affranchissent de l’étape de redescription manuelle des données. Dans le chapitre suivant, notre choix s’est porté naturellement vers les réseaux de neurones convolutionnels. Le réseau de neurones est un système probabiliste composé d’une succession de couches dont chacune prend ses entrées sur les sorties de la précédente. Chaque couche (i) est composée de N_i unités, prenant leurs entrées sur les N_{i-1} unités de la couche précédente. Entre chaque couche interviennent des transformations non-linéaires. Avec ce type d’apprentissage, l’extraction des descripteurs est automatique et est pilotée directement à partir des données.

Enfin, on peut noter que dans ce chapitre l’apprentissage a été réalisé sur une cohorte de patients d’entraînement. Dans la mesure où nous apprenons à partir des données de perfusion brutes (c.a.d. non déconvoluées), on peut alors s’interroger de l’impact de la variabilité d’entrée artérielle dans l’apprentissage. En effet, nous avons remarqué que certains patients dans la cohorte présentaient une fonction d’entrée artérielle particulièrement peu représentée. Comment assurer un ensemble d’entraînement suffisamment représentatif des données test, en particulier

vis-à-vis du critère d'entrée artérielle? Nous proposons une solution dans le chapitre suivant via une approche de simulation patient-spécifique.

Chapitre II.5

Valeur ajoutée de la simulation patient-spécifique

Une limitation à l'utilisation de l'apprentissage automatique supervisé dans le domaine de l'imagerie médicale est la quantité limitée de données annotées. Cela peut être considéré comme un goulot d'étranglement, en particulier lors de l'utilisation d'algorithmes hautement discriminants grâce à un grand nombre de paramètres. L'augmentation des données ("data augmentation") est un moyen d'améliorer la régularisation et de réduire le sur-apprentissage, en injectant davantage de variabilité dans la base de données d'entraînement. Dans ce chapitre, nous évaluons l'apport de la simulation de l'IRM de perfusion brute pour la prédiction du devenir de la lésion ischémique. Cette prédiction est effectuée ici avec un apprentissage par réseaux neuronaux convolutifs (CNN), car ils sont connus pour avoir un grand nombre de paramètres à ajuster et sont donc susceptibles de bénéficier de l'augmentation de données.

En particulier, nous proposons d'entraîner le réseau à partir d'images IRM de perfusion synthétiques produites par un simulateur physiquement et physiologiquement réaliste récemment publié [21]. Nous paramétrons cette simulation en nous concentrant sur la fonction d'entrée artérielle qui a été soulignée comme le facteur limitant du simulateur [21]. De plus, contrairement aux autres études en apprentissage profond appliquées à l'accident vasculaire cérébral [61, 15, 62, 63, 64] dans lesquelles l'ensemble des données d'entraînement est basé sur une cohorte de patients, nous démontrons la valeur ajoutée de l'apprentissage patient-spécifique.

II.5.1 Le simulateur IRM utilisé

Les données simulées ont été générées avec un simulateur de perfusion DSC-MRI [21] capable de simuler des images de concentration d'agent de contraste. Ce simulateur peut prendre en entrée des paramètres de classes distinctes de tissus (sain, infarctus, matière grise et blanche) et leurs paramètres hémodynamiques associés ainsi que la fonction d'entrée artérielle (AIF). À l'origine, le simulateur a été conçu pour tester la robustesse d'un algorithme de déconvolution de l'AIF [7]. La sensibilité du simulateur à tous les paramètres a été systématiquement testée, et il a été démontré que l'incertitude liée à l'AIF était le facteur limitant, comme le montre également Calamante et al. [65]. Dans cette étude, nous souhaitons étendre l'application du simulateur [21] en le proposant comme générateur de données pour l'apprentissage supervisé. En particulier, nous avons décidé d'investiguer l'importance du paramétrage de l'AIF sur les performances d'apprentissage. Les autres paramètres hémodynamiques ont été fixés à leurs valeurs par défaut rapportées de la littérature [21], que nous rappelons brièvement dans le Tableau II.5.1. Les paramètres d'acquisition ont été fixés à 200 u.a. pour la valeur de base, 60 s pour le temps d'acquisition, 0.030 s pour le temps d'écho et 21 dB pour le SNR.

L'AIF est modélisée par distribution gamma qui peut être exprimée à l'aide de la formulation simplifiée proposée par Madsen et al. [66] :

Paramètres Hémodynamiques	F	TGS	TBS	TL
CBF ($\mu \pm \sigma$)	0 \pm 0	60 \pm 9	25 \pm 2.1	10 \pm 4.3
MTT ($\mu \pm \sigma$)	0 \pm 0	4 \pm 2.2	4.8 \pm 3.2	10 \pm 5

TABLEAU II.5.1 – Paramètres par défaut du simulateur pour la variabilité tissulaire des paramètres hémodynamiques de débit sanguin cérébral (CBF, en mL/100g/s) et temps de transit moyen (MTT, en s). La distribution de chaque paramètre est modélisée par une distribution gaussienne de moyenne μ , écart-type σ . Trois classes de tissus sont considérés : tissus gris sains (TGS), tissus blancs sains (TBS) et tissus lésionnels (TL). Le fond (F) présente une distribution nulle.

$$f(t) = \begin{cases} 0, & \text{if } t \leq d \\ y_{max} \cdot \left(\frac{t-d}{t_{max}}\right)^\alpha \cdot \exp\left(\alpha\left(1 - \frac{t-d}{t_{max}}\right)\right), & \text{if } t \geq d \end{cases}, \quad (\text{II.5.1})$$

où y_{max} et t_{max} correspondent respectivement à la magnitude et à la position du maximum de la fonction d'entrée artérielle, d est le temps d'arrivée de l'agent de contraste et α correspond au paramètre de forme de la fonction gamma.

Extraction de l'AIF Nous avons caractérisé les AIF de chaque patient à l'aide d'une méthode de sélection multiple d'AIF. Pour ce faire, les AIF ont été extraites pour chaque patient à partir de voxels situés dans les artères cérébrales moyennes visibles sur l'imagerie de perfusion [67]. La sélection de voxels a été effectuée selon une région d'intérêt délimitée manuellement par trois opérateurs différents. Ensuite, pour chaque patient, les signaux de perfusion bruts des voxels sélectionnés ont été moyennés pour produire un signal de concentration moyen de l'agent de contraste. Les courbes de l'agent de contraste ont ensuite été ajustées à une fonction gamma définie dans l'Équation II.5.1.

Caractérisation des AIF L'estimation des paramètres d'AIF est connue pour être critique dans la tâche de prédiction des lésions ischémiques [65, 21]. Il existe différentes origines possibles de la variabilité de l'AIF. Tout d'abord, dans une étude multicentrique, la durée du protocole de perfusion peut être différente. Dans notre cas, la durée a été normalisée à 1 minute pour tous les centres. De plus, le délai entre le début de l'acquisition et l'injection du bolus d'agent de contraste peut être différent d'un patient à l'autre, ce qui produit un éventuel décalage temporel. Cependant, les CNN étant invariants en translation, ils sont supposés ne pas y être sensibles. La variabilité intra-patient s'explique par la variation de l'AIF selon l'emplacement du voxel sélectionné. La variabilité inter-patients est attribuée à la quantité de sang arrivant au cerveau qui peut varier d'un patient à l'autre. Ces deux sources de variabilité (intra- et inter-patients) étaient présentes dans notre ensemble de données. La variabilité intra-patient a été réduite dans notre étude en calculant la moyenne de l'AIF pour chaque patient après sélection artérielle par trois experts distincts. En ce qui concerne la variabilité inter-patients, comme observé dans par Meijs et al. [14] et dans notre ensemble de données, la plupart des patients présentaient une AIF haute et étroite représentés en bleu sur la Figure II.5.1 (le temps de transit de l'agent de contraste est compris entre 10 et 15 secondes). Les patients présentant une AIF plus large et aplatie (représentés en rouge dans la Figure II.5.1) étaient peu nombreux, et étaient par conséquent sous-représentés dans la cohorte. Ces observations ont motivé notre choix d'étudier différentes approches pour simuler l'AIF dans cette étude, selon les différents ensembles de données décrits ci-dessous.

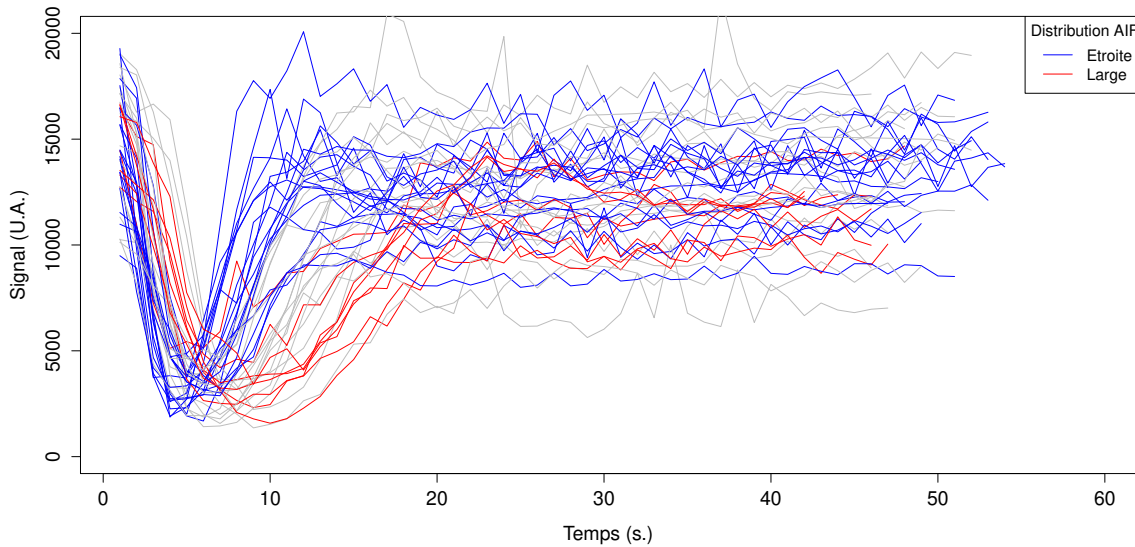


FIGURE II.5.1 – Courbes AIF moyennes extraites sur chacun des 76 patients de la cohorte. Les courbes bleues et rouges correspondent aux AIF avec des distributions respectivement étroites et larges. Les lignes grises correspondent aux AIF restantes. On peut constater la grande variabilité et l’inhomogénéité de forme des AIF.

II.5.2 Ensemble de données d’entraînement et de test

Ensemble d’entraînement constitué d’une sélection de patients réels Tout d’abord, un ensemble de données d’entraînement de 6 patients de la cohorte de la section II.5.3.1 a été créé : tous les patients présentaient des AIF étroites comme indiqué en bleu dans la Figure II.5.1. Cette approche permet de tester la valeur prédictive d’un ensemble d’entraînement biaisé, constitué de patients réels présentant une très faible variabilité d’AIF par rapport aux patients testés. Cet ensemble de données est évidemment très faible, mais sert de référence afin d’évaluer les prédictions obtenues à partir de patients simulés.

Ensemble d’entraînement constitué de simulations obtenues à partir d’une AIF théorique de la littérature (ensemble de données A) Un ensemble de données simulées, noté A, a été créé avec des images de concentration générées avec les paramètres AIF par défaut du simulateur, qui correspondent à des paramètres AIF moyens issus de la littérature [68]. Dans cette configuration, les paramètres de simulation d’AIF ont été fixés à une valeur unique, c’est-à-dire : $y_{max} = 0,61$, $t_{max} = 4,5$, $d = 3$, $\alpha = 3$. Cette approche permet de tester la valeur prédictive des images de perfusion synthétiques, simulées à partir d’une AIF théorique non ajustée aux valeurs de nos patients testés.

Ensemble d’entraînement constitué de simulations obtenues à partir d’une AIF spécifique au patient testé (ensemble de données B) Un deuxième ensemble de données simulées, noté B, a été créé en ajustant les paramètres d’entrée d’AIF sur celles des AIF extraites de chaque patient testé (voir les valeurs choisies dans le tableau II.5.2). Cette approche permet de tester la valeur ajoutée prédictive des images de perfusion synthétiques, simulées à partir d’une AIF ajustée aux valeurs de nos patients testés.

Ensemble test Nous avons choisi 8 patients de la cohorte pour construire l’ensemble de données test. Ils ont été sélectionnés pour couvrir la diversité des formes d’AIF observée dans la cohorte. En plus de devoir présenter une large diversité de forme d’AIF, les 8 patients sélectionnés présentaient une lésion finale volumineuse et compacte. En effet, lorsque les lésions finales sont très

petites (généralement inférieures à quelques mL), les métriques d'évaluation sont calculées sur un nombre très restreint de voxels, et chaque voxel mal prédit pénalise très rapidement les métriques représentatives du patient. Les lésions éclatées sont également plus difficiles à prédire, comme le montre Frindel et al. [69]. Notre sélection de patient permettait de garantir une évolution contrôlée de la lésion afin de se concentrer sur l'AIF et la variabilité résultant de ce paramètre. Comme le montre la Figure II.5.2, les patients 1, 2, 3 et 6 présentent une AIF large, tandis que les autres patients présentent des formes d'AIF relativement étroites. L'AIF de chaque patient individuel test a été extrait. Les paramètres d'ajustement à la loi gamma sont présentés dans le Tableau II.5.2 et illustrés dans la Figure II.5.2. L'ensemble B contient des données simulées ajustées sur les paramètres d'AIF du Tableau II.5.2.

Patient	y_{max}	t_{max}	d	α
1	0.90	6.32	13.14	4.23
2	0.86	10.53	11.99	8.54
3	0.60	3.41	18.67	0.99
4	0.92	5.26	19.64	0.94
5	0.74	2.35	11.62	1.01
6	0.60	4.58	14.01	2.30
7	0.77	3.45	9.40	2.48
8	0.85	7.08	8.79	8.03

TABLEAU II.5.2 – Valeurs des paramètres d'AIF (y_{max} , t_{max} , d , α) des patients testés obtenues à partir de leur courbe gamma moyenne d'agent de contraste.

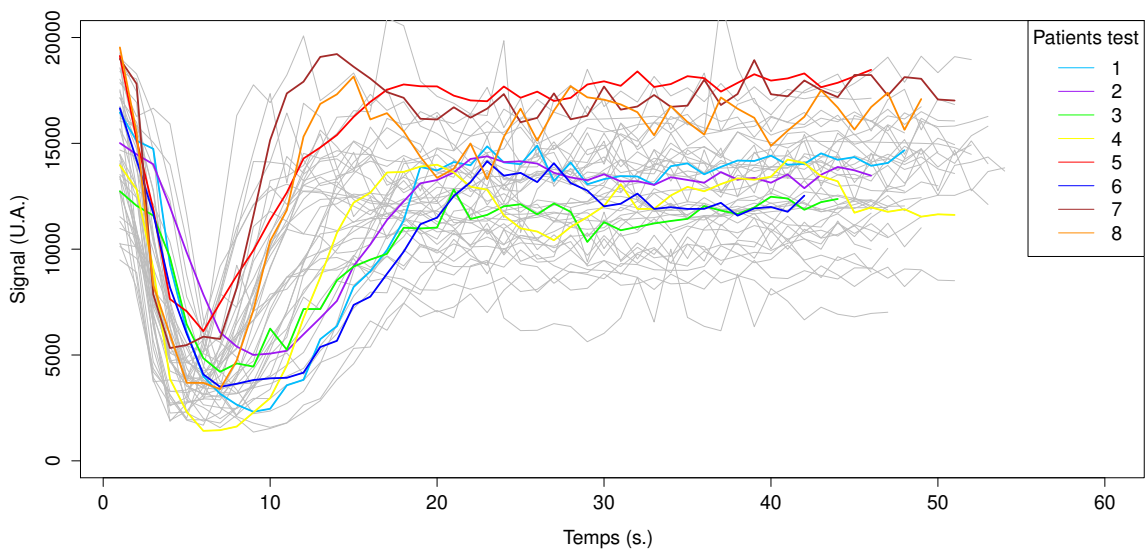


FIGURE II.5.2 – Courbes AIF des 8 patients testés (en couleur) parmi toutes les AIF de la cohorte (en gris). Afin de mieux les visualiser, nous n'avons pas montré le délai d . Les patients choisis présentent de grandes différences de forme d'AIF.

II.5.3 Évaluation des performances

La prédiction du devenir du tissu cérébral a été faite en encodant les images de perfusion en patches spatio-temporels [19]. Ces patches sont les images d'entrée de réseaux neuronaux convolutifs (CNN). La Figure II.5.3 montre le pipeline d'apprentissage proposé.

II.5.3.1 Données cliniques IRM

Les données d'IRM ont été extraites d'une cohorte de patients acquise aux Hospices Civils de Lyon [48]. Nous avons travaillé sur les données longitudinales de huit patients victimes d'un accident ischémique cérébral de la circulation antérieure. Seuls 2 de ces 8 patients ont reçu un traitement thrombolytique, mais aucun d'entre eux n'a reperfusé de lui-même. Tous les patients ont reçu le même protocole IRM à l'admission : une séquence d'imagerie pondérée par diffusion (DWI); une séquence FLAIR; une séquence pondérée T2*; une angiographie par résonance magnétique et une séquence de perfusion (DSC-PWI). Une IRM FLAIR de suivi a été effectuée à 1 mois. Une correction de mouvement a été appliquée sur l'IRM de perfusion brute. Nous avons recalé, coupe par coupe, tous les points de temps sur le premier point de temps par une approche d'information mutuelle. Cela a été fait en recalant chaque point (n+1) sur son point temporel précédent (n) en appliquant récursivement les matrices de transformation obtenues jusqu'à ce que tous les points soient alignés sur le premier point temporel. La segmentation du masque de la lésion finale a été réalisée pour chaque patient sur l'IRM de suivi FLAIR par des experts. L'IRM FLAIR a été recalée sur l'imagerie de perfusion. Après recalage, les lésions finales ont été binarisées en appliquant un seuil de 50%.

II.5.3.2 Encodage de l'imagerie de perfusion

Les images de perfusion ont été converties en images de concentration d'agent de contraste après une transformation logarithmique, en supposant une relation linéaire entre la concentration d'agent de contraste et la variation du taux de relaxation transverse [70, 71]. Cette transformation permet de normaliser les images entre les patients puisqu'on s'affranchit de la baseline. Ensuite, les images de concentration ont été encodées en patchs spatio-temporels [19] : le signal temporel d'un voxel d'intérêt est alors déployé horizontalement, et ses 8 voxels dans le voisinage de Moore d'ordre 1 sont également déployés horizontalement, empilant la signature temporelle de chaque voxel voisin les uns sous les autres. Ainsi, un patch de taille 9 par Nt est créé pour chaque voxel, où Nt est le nombre de points d'acquisition dans le temps de la séquence d'imagerie de perfusion. Afin d'obtenir des patchs indépendants les uns des autres, nous n'avons pas considéré un voisinage de Moore d'ordre supérieur.

Il a été démontré que les patchs lésionnels et les patchs sains peuvent être discriminés en termes de texture [19]. Pour aller plus loin que le Giacalone et al., on peut remarquer sur la Figure II.5.4 que cette discrimination dépend fortement de l'AIF : le temps de transit de l'agent de contraste est beaucoup plus court pour le patient 8 (présentant une AIF de forme étroite) que le patient 2 (présentant une AIF de forme plus étalée). Cela confirme la nécessité d'un apprentissage patient-spécifique, en tenant compte de l'AIF de chaque patient pour mieux prédire ses voxels pathologiques.

II.5.3.3 Classification par réseau de neurones convolutif

Un réseau de neurones convolutif (CNN) a été conçu pour prendre en entrée les patchs spatio-temporels de dimension (9,60), afin de fournir une prédiction à l'échelle du voxel. En sortie du CNN, on associe à chaque patch la probabilité d'appartenir à deux classes (tissu sain ou tissu infarci). Nous avons choisi un CNN car ce type de modèle est invariant en translation [72]. Cette propriété est un point clé de l'apprentissage, car nous souhaitons uniquement discriminer les patchs sur le transit de l'agent de contraste dans les tissus et non pas sur son délai d'arrivée d . Ainsi, le réseau ne serait pas sensible au délai d , mais seulement au motif blanc de la signature spatio-temporelle. Une architecture parcimonieuse décrite dans le Tableau II.5.3 a été conçue. Cette architecture présente un nombre limité de couches convolutives afin d'éviter la réduction de la taille des patchs et le sur-apprentissage. Les poids du réseau ont été initialisés de manière aléatoire au début de l'apprentissage. La fonction ReLU a été utilisée comme fonction d'activation, connue pour être plus performante et plus rapide que les fonctions tangentes sigmoïdes ou hyperboliques [73, 74]. Dans la dernière couche entièrement connectée (*fully connected layer*),

nous avons utilisé une fonction d'activation softmax, avec 2 unités de sortie car notre tâche s'apparente à un problème de classification binaire. Dans la mesure où les patchs en entrée sont de petite dimension (9,60) et que la convolution tend à réduire la dimension de l'image en sortie, nous avons décidé de ne pas utiliser de fonction *max-pooling* pour éviter de réduire la taille des patchs. Nous avons utilisé du *dropout* [75, 76] dans les couches entièrement connectées afin d'éviter le sur-apprentissage. Nous avons utilisé la fonction entropie croisée catégorique (*categorical cross-entropy*) comme fonction de perte (*loss function*) et une descente de gradient stochastique pour optimiser le modèle. Pour toutes les expériences, le nombre total de poids à entraîner était de 197 087, le *dropout* a été fixé à 0.5, le nombre d'époques a été fixé à 30 et la taille des *batches* à 32.

Couche	Type	Taille Filtre	Stride	# filtres	unités FC	Taille Sortie
1	Conv2D	2*2	1*1	16	-	(8, 59, 16)
2	Conv2D	2*2	1*1	32	-	(7, 58, 32)
3	FC	-	-	-	15	(15)
4	FC	-	-	-	2	(2)

TABLEAU II.5.3 – Architecture du réseau neuronal convolutif proposé. Les deux premières couches sont de type Convolutionnel 2D (Conv2D) et les deux dernières couches sont des couches entièrement connectées (FC pour *Fully Connected*).

Chaque modèle a été entraîné 10 fois afin d'avoir des performances globales robustes et pas seulement le meilleur apprentissage possible. Pour obtenir un ensemble d'entraînement équilibré, nous avons veillé à ce que la moitié des patchs soient sains, l'autre moitié lésionnels. L'entraînement des réseaux durait un peu moins de 15 minutes sur une station de travail équipée d'une carte GPU NVIDIA GeForce GTX 1080 et 128 Go de mémoire.

II.5.3.4 Métriques d'évaluation

Nous avons évalué nos résultats en utilisant le coefficient de similitude Dice (DSC) [49] et la distance de Hausdorff (HD) [59], dans la mesure où ils ont été utilisés pour le challenge international de segmentation des lésions de l'accident ischémique cérébral de MICCAI [17]. Ces métriques ont été calculées entre les voxels prédits comme infarctés et le masque de la lésion finale fourni par l'IRM FLAIR. À titre de comparaison, la prédiction de la lésion de perfusion a également été calculée à partir d'une carte de perfusion T_{max} seuillée à 6 secondes. L'approche du T_{max} seuillé reste la référence standard en recherche clinique [77]. Nous avons calculé les métriques DSC et HD entre les voxels au-dessus de ce seuil et le masque de la lésion finale. L'utilisation de données simulées pour l'entraînement permet de produire des données à la demande. Nous avons évalué le nombre minimum de patchs simulés nécessaires pour obtenir un apprentissage stable en rapport avec l'architecture CNN proposée. Pour chaque patient, nous avons 125 000 patchs initiaux que nous avons divisés en 8 sous-ensembles de taille différente par tirage aléatoire avec remise : respectivement 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% du nombre initial de patchs. Pour chaque patient, chaque ensemble de patchs a été entraîné 10 fois.

II.5.3.5 Détails expérimentaux

Nous avons classifié les patchs test par le biais de trois expériences, chacune utilisant un ensemble de données d'entraînement différent.

Expérience 1 : entraînement à partir d'une sélection de patients réels Dans cette expérience, l'ensemble d'entraînement était constitué de patchs provenant de 6 patients de la cohorte présentant une AIF étroite. Sur ces 6 patients, nous avons pu obtenir 21 914 patchs dont la moitié

étaient sains et l'autre moitié infarcis. L'ensemble de validation comprenait 17 766 patchs provenant de 3 patients indépendants présentant une forme d'AIF différente. Dans cette expérience, le taux d'apprentissage (*learning rate*) a été fixé à 0.01.

Expérience 2 : entraînement à partir de simulations issues d'une AIF théorique Dans cette expérience, l'ensemble d'entraînement était constitué de 125 000 patchs provenant d'images simulées de l'ensemble de donnée A, dont la moitié des patchs étaient sains et l'autre moitié infarcis. L'ensemble de validation comprenait 125 000 patchs provenant d'autres images obtenues avec les mêmes paramètres de simulation. Dans cette expérience, le taux d'apprentissage (*learning rate*) a été fixé à 0.0001.

Expérience 3 : entraînement à partir de simulations issues d'une AIF patient-spécifique Dans cette expérience, nous avons appris d'un patient synthétique spécifique. L'ensemble de données B a été séparé en plusieurs sous-ensembles, chaque sous-ensemble représentant des images simulées d'un patient spécifique testé. Par conséquent, 8 entraînements différents ont été effectués. Pour chaque entraînement, l'ensemble d'entraînement comprenait 125 000 patchs dont la moitié étaient sains et l'autre moitié infarcis, et l'ensemble de validation comprenait un ensemble de 125 000 patchs provenant d'autres images obtenues avec les mêmes paramètres de simulation spécifiques au patient. Dans cette expérience, le taux d'apprentissage (*learning rate*) a été fixé à 0.0001.

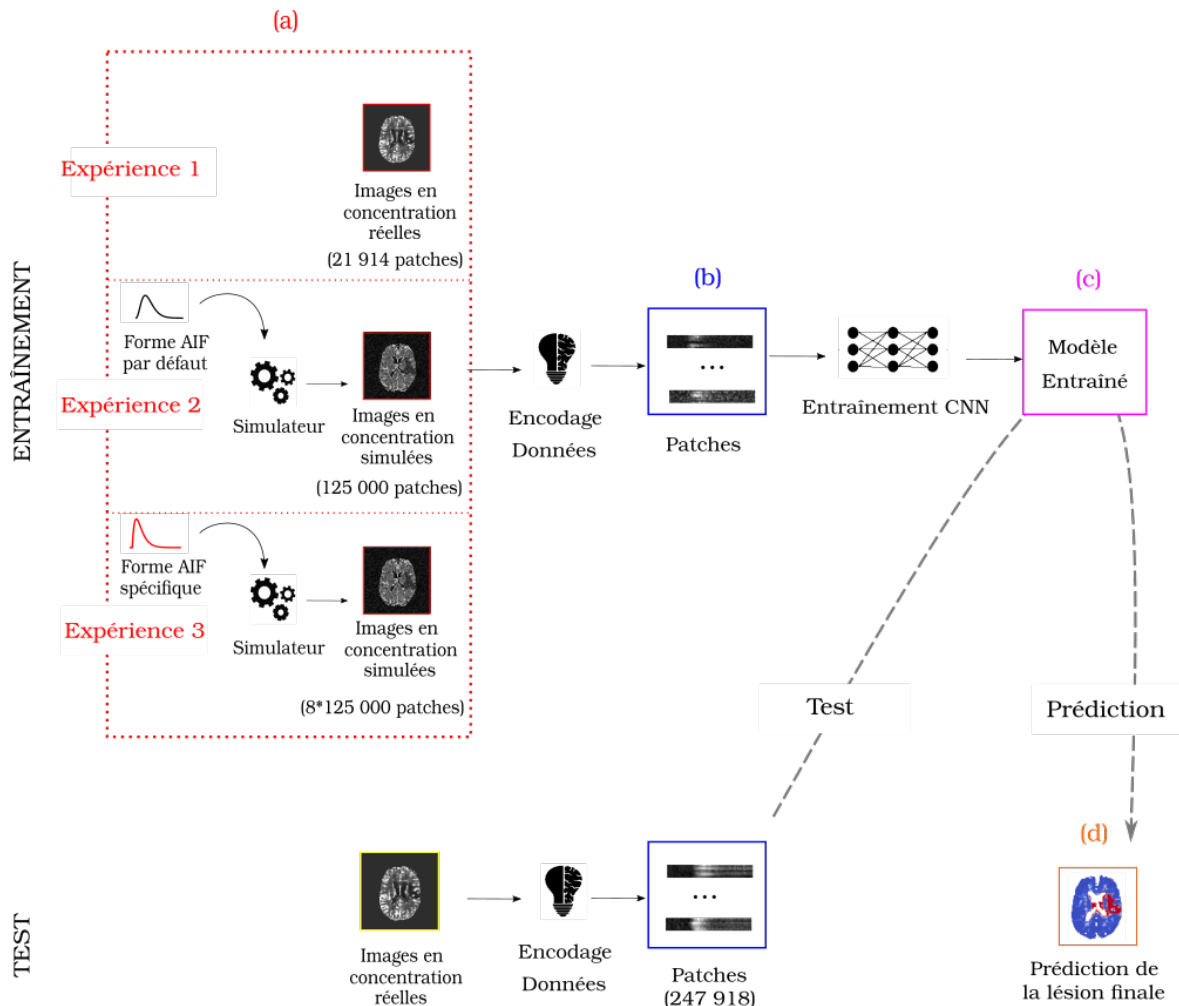


FIGURE II.5.3 – Vue d’ensemble du pipeline de prédiction proposé. (a) Les images initiales sont des images de concentration d’agent de contraste. Dans l’expérience 1, l’ensemble de données d’entraînement consiste en des patchs provenant d’images de concentration réelles, tandis que dans les expériences 2 et 3, l’ensemble de données d’entraînement consiste en des patchs synthétiques obtenus à partir du simulateur. Dans l’expérience 2, le paramètre d’entrée AIF est réglé sur une valeur par défaut, et dans l’expérience 3, les paramètres d’entrée AIF sont ceux du patient testé. (b) Les images de concentration sont codées en patchs spatio-temporels. (c) Un modèle de réseau neuronal convolutif (CNN) est entraîné à partir de patchs d’images de concentration. (d) Chaque voxel des images de concentration testées est classé comme sain ou infarci.

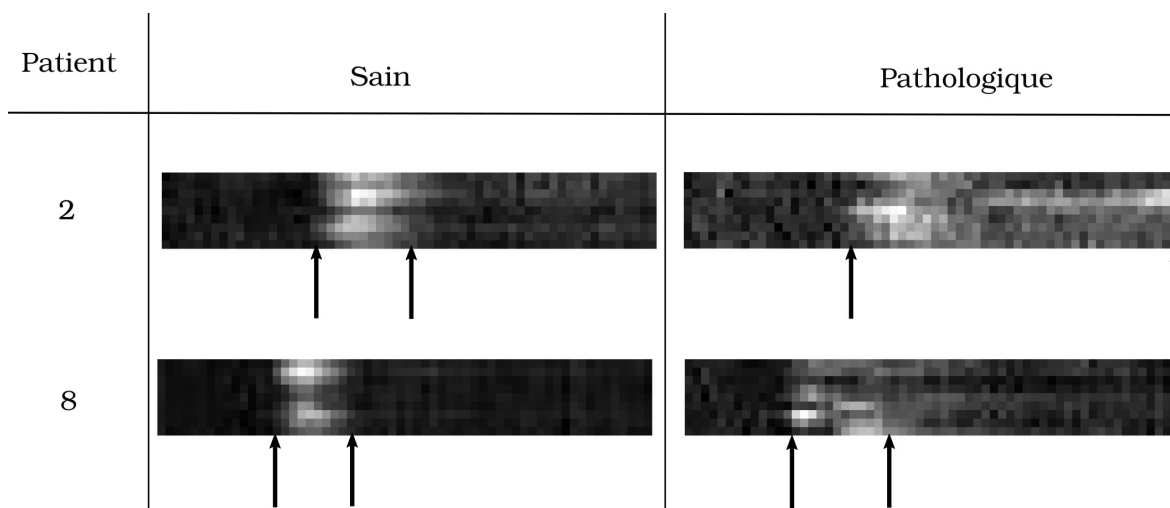


FIGURE II.5.4 – Illustration d'un patch sain (à gauche) et d'un patch pathologique (à droite) issus de l'image de concentration pour les patients test 1 et 8. Les voxels sains présentent un segment hyperintense étroit résultant du passage rapide du bolus d'agent de contraste. Les voxels pathologiques présentent un segment hyperintense étalé, bruité et peu contrasté, résultant du passage difficile du bolus d'agent de contraste. Les deux patients présentent des caractéristiques hémodynamiques différentes : le temps de transit de l'agent de contraste, représenté par l'espace entre les deux flèches, est plus rapide pour le patient 8 que pour le patient 2.

II.5.4 Résultats

Le tableau II.5.4 présente les résultats issus des trois expériences menées : les valeurs moyennes et écart-types des métriques DSC et HD sont reportées pour chaque patient testé. Les valeurs de DSC et HD entre le T_{max} seuillé et la lésion finale sont également indiquées, à titre de référence clinique.

Patient	Expérience 1		Expérience 2		Expérience 3		TMAX thr	
	HD	DSC	HD	DSC	HD	DSC	HD	DSC
1	50.4 ± 0.0	0.17 ± 0.031	50.2 ± 0.0	0.12 ± 0.00	50.2 ± 0.0	0.086 ± 0.003	49.7	0.11
2	51.1 ± 0.0	0.18 ± 0.002	51.1 ± 0.0	0.18 ± 0.00	48.8 ± 0.829	0.62 ± 0.008	43.2	0.51
3	44.1 ± 0.0	0.054 ± 0.001	43.8 ± 0.0	0.053 ± 0.00	43.8 ± 0.0	0.23 ± 0.010	43.4	0.14
4	44.4 ± 0.0	0.47 ± 0.041	44.9 ± 0.0	0.17 ± 0.00	44.8 ± 0.085	0.48 ± 0.008	42.2	0.47
5	43.8 ± 1.08	0.36 ± 0.069	45.4 ± 0.0	0.11 ± 0.00	44.4 ± 0.0	0.47 ± 0.005	46.1	0.31
6	43.0 ± 0.0	0.11 ± 10 ⁻⁵	43.0 ± 0.0	0.11 ± 0.00	42.5 ± 0.0	0.20 ± 0.006	44.7	0.28
7	44.0 ± 0.36	0.47 ± 0.031	45.0 ± 0.0	0.26 ± 0.00	45.0 ± 0.0	0.53 ± 0.004	46.3	0.36
8	40.1 ± 3.07	0.64 ± 0.026	47.1 ± 0.0	0.17 ± 0.00	44.0 ± 10 ⁻⁶	0.48 ± 0.017	45.8	0.40
Moyenne Totale	45.1 ± 3.73	0.30 ± 0.22	46.3 ± 2.94	0.14 ± 0.074	45.4 ± 2.65	0.40 ± 0.19	45.18 ± 2.36	0.32 ± 0.14

TABLEAU II.5.4 – Distance de Hausdorff (HD) et coefficient de similarité (DSC) après avoir réalisé 10 fois l’expérience 1 (1ère colonne), l’expérience 2 (2ème colonne) et l’expérience 3 (3ème colonne). La moyenne de toutes les mesures est calculée sur les 10 réalisations, et indiquée pour chaque patient testé (moyenne ± écart type). Les métriques DSC et HD entre le $T_{max} \geq 6s$ et la lésion finale sont reportées (4ème colonne). Nous avons indiqué en gras les valeurs pour lesquelles l’expérience 3 a donné les meilleures performances à l’échelle du patient. Dans les trois expériences, l’écart-type de la métrique HD est faible, voire nul, car certains voxels aberrants ont été systématiquement mal prédits. Les métriques moyennes sur l’ensemble des données test ont également été calculées (dernière ligne).

Les résultats de l’expérience 1 montrent l’impact d’une disparité d’AIF entre les ensembles d’entraînement et de test. Pendant la phase d’entraînement, la discrimination entre les voxels sains et pathologiques n’a été faite qu’à partir de patients présentant une AIF étroite : il s’avère alors impossible de prédire correctement les voxels des patients testés présentant des formes d’AIF différentes, tels que les patients 1, 2, 3 et 6 qui ont un DSC moyen inférieur à 0.13. En revanche, les patients testés ayant des AIF proches de ceux des patients d’entraînement, comme les patients 4, 5, 7 et 8, présente un DSC moyen supérieure à 0.49.

Dans l’expérience 2, l’entraînement de CNN avec des données simulées obtenues à partir d’une AIF théorique (correspondant à des paramètres AIF moyens tirés de la littérature sans aucun ajustement patient-spécifique) donne de très mauvais résultats. Presque tous les voxels sont prédits infarcis, de sorte que le DSC moyen est très faible et la distance de Hausdorff moyenne élevée. Cela montre qu’une AIF théorique n’est pas en mesure de saisir la variabilité qui existe entre les patients. L’AIF diffère d’un patient à l’autre, et comme nous travaillons sur des données brutes d’imagerie de perfusion, il est important d’intégrer l’AIF spécifique à chaque patient test.

La variabilité apportée par l’AIF d’un patient à l’autre est bien connue [14]. C’est la raison pour laquelle la plupart des approches classiques utilisent la déconvolution afin de compenser cette variabilité. Or ce problème inverse est mal posé, et les données déconvoluées sont de ce fait biaisées. Dans l’expérience 3, pour éviter de recourir à la déconvolution, nous avons plutôt simulé des signaux de perfusion à partir de fonctions d’entrée artérielle spécifiques. L’ensemble d’entraînement contient alors que des images simulées, où l’AIF est ajustée sur celle du patient à prédire (ensemble de données B). Il apparaît clairement que l’ajustement des paramètres liés à l’AIF dans le simulateur a un impact considérable sur les performances d’apprentissage. Le DSC moyen dans l’expérience 3 est de 0.40 (± 0.19), contre 0.14 (± 0.074) dans l’expérience 2 et 0.30 (± 0.22) dans l’expérience 1. L’apprentissage à partir d’une AIF spécifique dans l’expérience 3 a clairement amélioré le DSC médian, qui est de 0.48 dans l’expérience 3, contre 0.27 dans l’expérience 1 et 0.15 dans l’expérience 2. Il apparaît également que l’apprentissage à partir de données non déconvoluées dans l’expérience 3 donne de meilleurs résultats que le seuillage du T_{max} déconvolué qui présente un DSC moyen de 0.32 (± 0.14). À titre de référence, il est intéressant de noter que les meilleurs modèles utilisés jusqu’à présent dans le cadre du challenge de prédiction ISLES 2017

présentaient un DSC moyen de $0.38 (\pm 0.22)$ et une distance de Hausdorff moyenne de $29.21 (\pm 15.04)$. Les résultats moyens de l'expérience 3 sont du même ordre de grandeur. Toutefois, une comparaison absolue n'est pas strictement possible car les deux ensembles de données sont différents. Pour tester la transférabilité de la méthode que nous proposons, nous avons réalisé 10 fois l'expérience 3 pour les patients 7 et 20 de la base de données ISLES, qui présentaient des lésions particulièrement larges. Le DSC moyen était alors de $0.39 (\pm 0.30)$ et la distance moyenne de Hausdorff était de $46.9 (\pm 1.80)$.

En ce qui concerne l'architecture CNN décrite dans l'expérience 3, nous avons également investigué le nombre de patches requis pour un apprentissage stable. Les résultats sont présentés dans la Figure II.5.5. Il apparaît qu'entre 25 000 et 100 000 patches d'entraînement, le DSC augmente de manière presque linéaire. Au-delà de 100 000 patches, la courbe semble atteindre un plateau : toute la diversité des informations fournies par les images simulées a été apprise, et l'apport de nouvelles images est redondant. De plus, l'écart type des valeurs DSC est plus faible au-delà de 100 000 patches. Ces observations indiquent qu'étant donné notre architecture CNN et notre ensemble de données, le nombre minimum de patches d'entraînement pour obtenir des résultats stables et optimaux est d'environ 100 000.

Comme expériences additionnelles, nous avons comparé les expériences 1, 2 et 3 avec d'autres architectures de réseaux de neurones et d'autres métriques. Les résultats de cette comparaison sont fournis en Annexe B. Quelle que soit l'architecture ou la métrique utilisée, l'expérience 3 donne systématiquement de meilleures performances que les expériences 1 et 2. La valeur ajoutée de la simulation utilisée dans l'expérience 3 est confirmée de manière robuste.

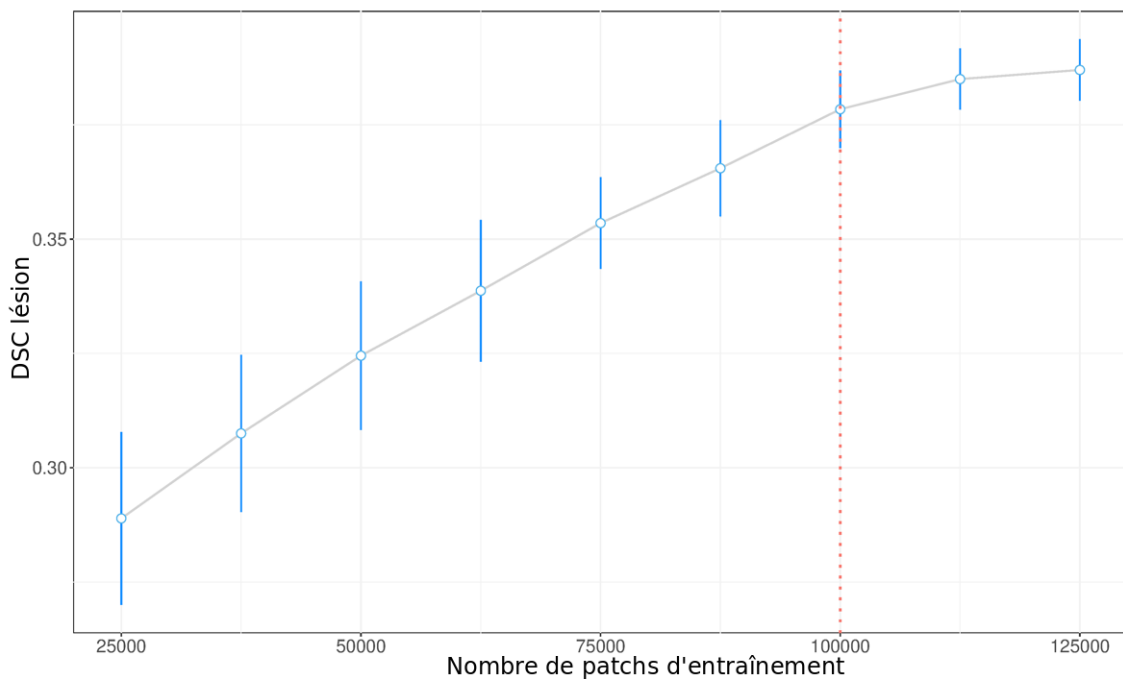


FIGURE II.5.5 – Métrique DSC en fonction du nombre de patches d'entraînement lors de l'expérience 3. Chaque point correspond au DSC moyenné sur l'ensemble des patients test et sur les 10 expériences répétées. La ligne pointillée rouge indique le nombre optimal de patches d'entraînement.

II.5.5 Discussion

Les résultats précédents ont démontré l'intérêt de la simulation de l'imagerie de perfusion basée sur l'AIF pour la prédiction de la lésion ischémique. Dans cette section, nous allons au-delà de la seule observation de nos performances et discutons des limites de nos expériences.

Face au problème de la représentativité des AIF dans l'ensemble d'entraînement (souligné

dans les expériences 1 et 2), nous avons proposé via l'expérience 3 d'apprendre directement à partir de l'AIF du patient à tester. Grâce à ce type d'apprentissage, chaque patient a été correctement prédit individuellement, même pour les patients dont l'AIF était sous-représentée dans la cohorte globale. Il convient de noter que pour l'expérience 3, nous avons simulé des images à partir d'une AIF unique et spécifique au patient testé. Cependant, certaines études montrent qu'il peut être bénéfique de prendre en compte la variabilité de l'AIF intra-patient [78]. En effet, dans l'expérience 3, l'AIF extraite semble pertinente pour les patients 2, 3, 4, 5, 6 et 7, mais pas assez pour les patients 1 et 8 car ils montrent de meilleures performances dans l'expérience 1. Ces deux patients ne peuvent probablement pas être résumés en une seule AIF car ils pourraient présenter une grande variabilité intra-patient. Cela nous encourage à poursuivre nos travaux pour simuler des images avec plusieurs AIF en fonction de la variabilité intra-AIF du patient testé, et donc potentiellement mieux représenter ses caractéristiques hémodynamiques.

La Figure II.5.6 montre la position des voxels mal classés résultant des expériences 1, 2 et 3. Nous pouvons voir que beaucoup de nos erreurs se situent au niveau des ventricules. Comme l'agent de contraste ne passe pas dans ces zones, les voxels y présentent un signal particulièrement bruité. Le modèle détecte donc une perturbation du flux sanguin cérébral et la relie directement à l'accident ischémique cérébral final. Ces erreurs pourraient être facilement filtrées en limitant les erreurs de segmentation des masques du cerveau.

Actuellement en recherche clinique, la prédiction de la lésion finale se fait classiquement à partir des cartes de paramètres déconvolués (CBV, CBE, T_{max} , TTP et MTT) apprises sur des cohortes de patients. Avec notre approche, il n'est en fait pas possible d'apprendre directement à partir de ces cartes puisque notre encodage est basé sur la signature spatio-temporelle des données de perfusion. Une comparaison possible serait de comparer l'apprentissage à partir de signaux de perfusion bruts, encodés avec notre approche spatio-temporelle, avec un apprentissage à partir de signaux déconvolués (celui utilisé pour produire les cartes paramétriques standard CBV, CBE, T_{max} , TTP et MTT). Une telle comparaison a été présentée dans [19] et a démontré l'intérêt de l'encodage spatio-temporel du signal de perfusion brut.

Dans ce travail, l'architecture CNN proposée est constituée d'un faible nombre de couches. Le nombre minimum de patches requis pour l'apprentissage d'un patient spécifique a donc été limité à 100 000 (ce qui représente environ 25 simulations nécessaires pour représenter un patient avec 2 000 voxels pathologiques, et un temps de calcul de 3 minutes). Cependant, avec d'autres méthodes d'encodage et des modèles plus complexes, le nombre de données nécessaires à l'apprentissage pourrait être plus élevé. Le simulateur que nous avons utilisé dans cet article [21] serait capable de surmonter ce problème, en permettant de produire une quantité théoriquement illimitée de données annotées.

Enfin, il convient de noter que cette approche, bien que non instantanée (étapes de simulation et d'apprentissage), semble pleinement compatible avec la prise en charge en temps réel des patients victimes d'un accident vasculaire cérébral dans la routine clinique. En effet, pour des performances comparables à celles de l'état de l'art, le temps de calcul global (simulation et apprentissage) pour une approche patient-spécifique telle que développée dans l'expérience 3 est d'environ 20 minutes. Ce temps reste du même ordre de grandeur que le temps annoncé pour une approche de déconvolution [7].

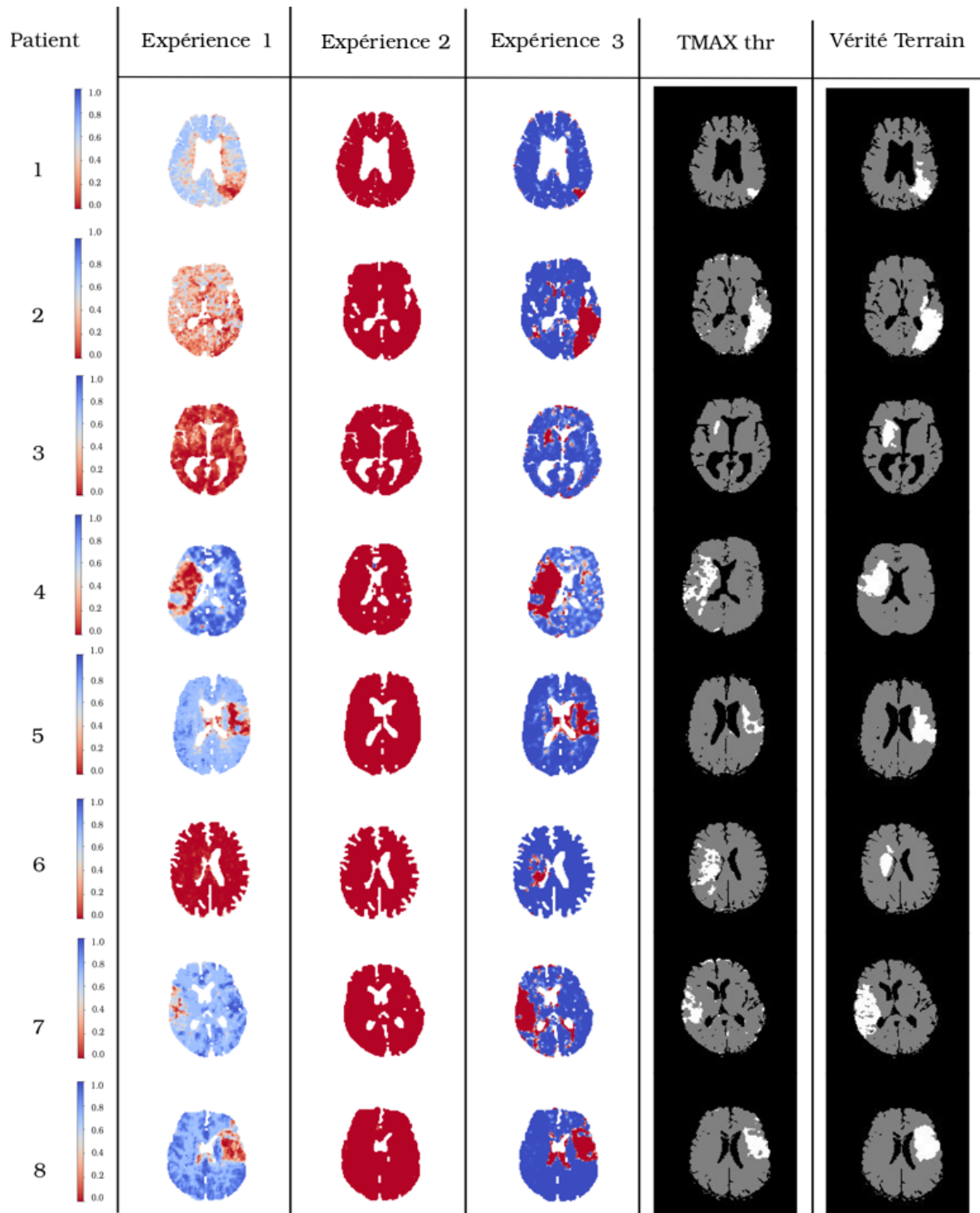


FIGURE II.5.6 – Prédictions obtenues lors de l'expérience 1 (1ère colonne), de l'expérience 2 (2ème colonne) et de l'expérience 3 (3ème colonne) pour les patients testés. La barre de couleur à gauche présente la probabilité pour chaque voxel d'être sain. Les voxels dans les tons bleus ont été prédits sains et les voxels dans les tons rouges ont été prédits infarcis. Le biomarqueur classique T_{max} , dont le seuil est fixé à 6 secondes, est indiqué (4e colonne). Les colonnes 1 à 4 doivent être comparées au masque lésionnel du FLAIR final (5e colonne).

Chapitre II.6

Apport d'un nouveau degré de réalisme relatif à l'arbre vasculaire au simulateur

Dans ce chapitre, nous proposons d'améliorer le réalisme physiologique du simulateur présenté dans le chapitre II.5. Nous suggérons une modélisation patient-spécifique de l'anatomie de l'arbre vasculaire, en intégrant les variations du flux artériel et du temps de transport de l'agent de contraste au sein de cerveau. L'intégration de cette nouvelle étape de modélisation au sein du simulateur est résumée en Figure II.6.1.

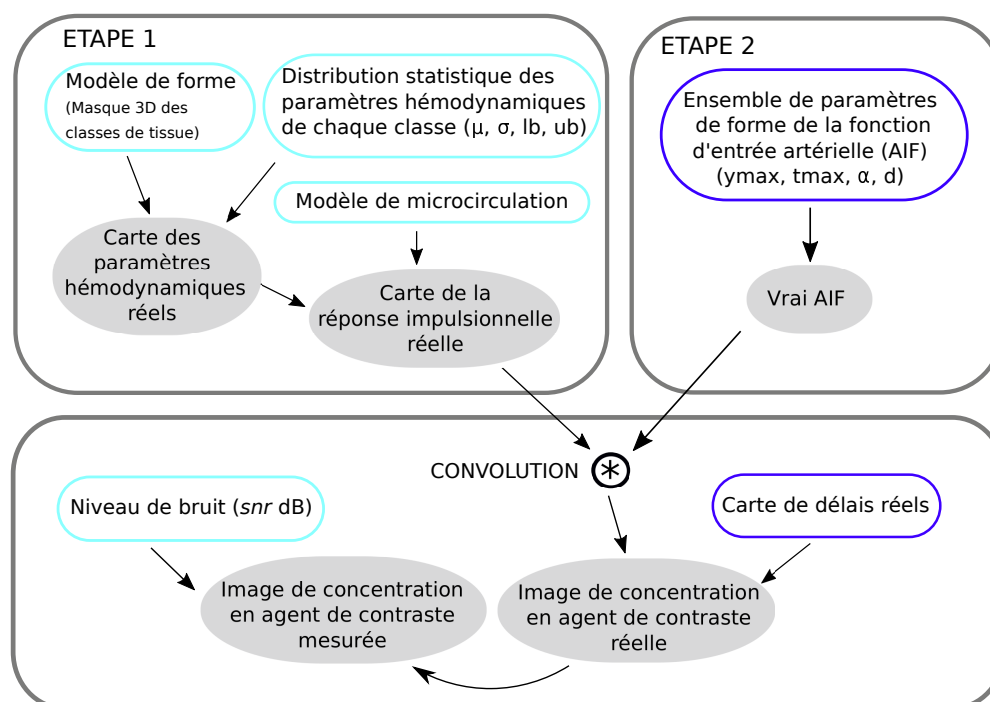


FIGURE II.6.1 – Pipeline de la nouvelle version du simulateur d'IRM de perfusion. Les nouvelles sources de variabilité introduites sont affichées en bleu foncé.

II.6.1 Intégration de la variabilité du flux artériel

L'hypothèse d'une AIF globale et unique a récemment été remise en question [78, 79] car pour un même patient, une importante variabilité de l'AIF peut être observée lors de l'extraction des signaux de concentration d'une artère principale contralatérale sur l'imagerie de perfusion (voir Figure II.6.2). Ces variations peuvent être causées par des artefacts d'imagerie (tels qu'un effet de volume partiel, une faible résolution temporelle) [80] ou même la présence d'une sténose entraî-

nant un pic plus bas et une forme de bolus plus large dans le territoire alimenté par l'artère sténosée [81]. Cette variabilité de l'AIF intra-patient a été intégrée dans le simulateur de Giacalone et al. [21]. Dans la nouvelle version du simulateur proposée, un ensemble de multiples AIF spécifiques peuvent être donné en paramètre d'entrée (contre une seule AIF globale dans l'ancienne version).

À cette fin, nous avons d'abord caractérisé les AIF des patients test en modélisant chaque AIF à une fonction gamma définie par l'expression simplifiée de Madsen et al. [66] :

$$\Gamma(t) = \begin{cases} 0, & \text{if } t \leq d \\ y_{max} \cdot \left(\frac{t-d}{t_{max}}\right)^\alpha \cdot \exp\left(\alpha\left(1 - \frac{t-d}{t_{max}}\right)\right), & \text{if } t \geq d \end{cases} \quad (\text{II.6.1})$$

où y_{max} et t_{max} correspondent respectivement à la magnitude et à la position du maximum de l'AIF, d est le temps d'arrivée de l'agent de contraste (ou le délai) et α correspond au paramètre de forme de la fonction gamma. L'AIF de chaque patient test est alors décrite, non plus par quatre paramètres uniques (une valeur unique pour y_{max} , t_{max} , d et α), mais par quatre ensembles de paramètres (plusieurs valeurs pour y_{max} , t_{max} , d et α). Les quatre ensembles de paramètres sont ensuite donnés en entrée du simulateur. Le simulateur ne sélectionne alors que les AIF les plus pertinentes, contenues dans le polygone de Delaunay 4D formé par les quatre ensembles de paramètres.

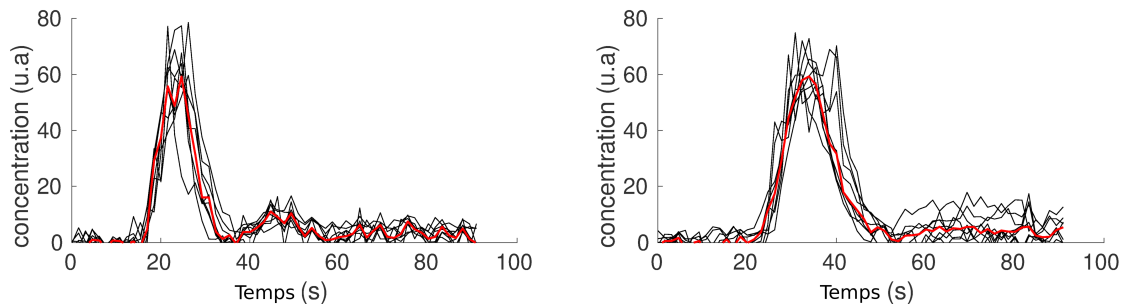


FIGURE II.6.2 – Ensemble des AIF extraites manuellement (noir) et le signal médian associé (rouge) pour 2 patients d'étude. À titre d'exemple, la variabilité des AIF intra-patient est plus importante pour le patient à droite que le patient de gauche.

II.6.2 Intégration du temps de transport de l'agent de contraste

Le transport de l'agent de contraste le long de l'arbre vasculaire entraîne un retard croissant (c'est-à-dire un décalage temporel) de l'AIF dans la périphérie du cerveau [82]. La prise en compte de ce phénomène dans la simulation permettrait d'atteindre un nouveau degré de réalisme spatial, et pourrait fournir des informations importantes sur l'approvisionnement vasculaire de la lésion et sur d'éventuels flux collatéraux (circulation de suppléance autour d'une artère bloquée). Ces informations additionnelles pourrait avoir des conséquences sur la géométrie de la lésion finale prédite.

Afin de construire la carte de délai spécifique au patient, la fonction gamma (défini dans l'Équation II.6.1) a d'abord été ajustée à la courbe de concentration temporelle de chaque voxel dans l'image de perfusion originale en utilisant l'algorithme de Levenberg-Marquart. Le délai a ensuite été estimé à l'aide de la méthode de Rose et al. [83]. La carte résultante était particulièrement bruitée : les voxels situés aux bords du cerveau, proches des ventricules ou dans le cœur de la lésion donnaient de mauvais résultats d'ajustement. Dans ces zones, l'agent de contraste ne passe pas, ce qui se traduit par un très faible rapport signal/bruit et donc des valeurs de délai artificielles. Les performances d'ajustement ont été évaluées pour chaque voxel en utilisant la somme des erreurs au carré (SCE) après normalisation de la hauteur du signal temporel.

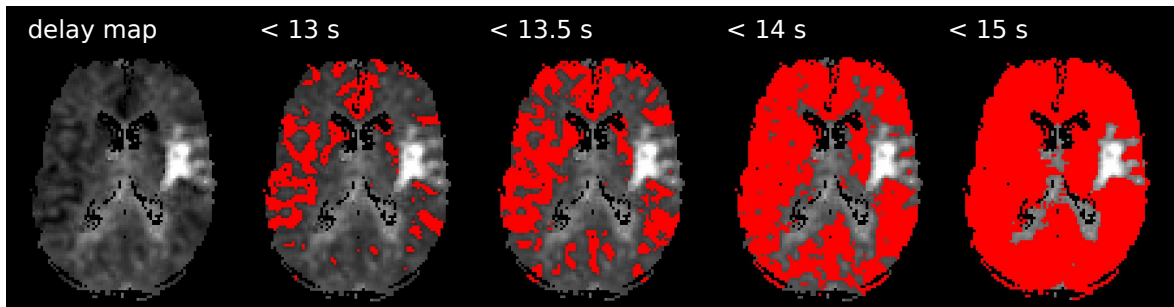


FIGURE II.6.3 – Carte de délai pour une coupe d’un patient donné, calculée avec des seuils de délai successifs et croissants. Cette représentation donne des informations sur la propagation de l’agent de contraste depuis les artères principales vers les tissus.

Les voxels ayant un indice SCE supérieur à 0.20 ont été supprimés. Les valeurs manquantes ont ensuite été complétées à l’aide d’un filtre moyenneur adaptatif. Comme les images de perfusion sont anisotropes, le filtre utilisé était en 2 dimensions. La taille optimale du voisinage de chaque voxel a été choisie de manière à inclure au moins 9 voxels bien ajustés. Pour la visualisation, nous avons utilisé la représentation de la propagation du front d’onde proposée par Christensen et al. [82] (Figure II.6.3). La carte de délai résultante a été définie comme une nouvelle entrée du simulateur.

II.6.3 Évaluation des améliorations proposées

Afin d’évaluer le réalisme de la nouvelle version du simulateur, nous avons prédit la lésion finale de 8 patients réels à partir d’images simulées associées à l’aide de réseaux de neurones convolutifs. Nous avons adopté une approche d’apprentissage patient-spécifique : pour chaque patient réel testé, un modèle a été entraîné uniquement à partir de ses images simulées spécifiques. Au travers de différentes expériences, nous montrons l’évolution des performances de prédiction à mesure que nous augmentons le degré de réalisme du simulateur. Le pipeline général est illustré dans la Figure II.6.4.

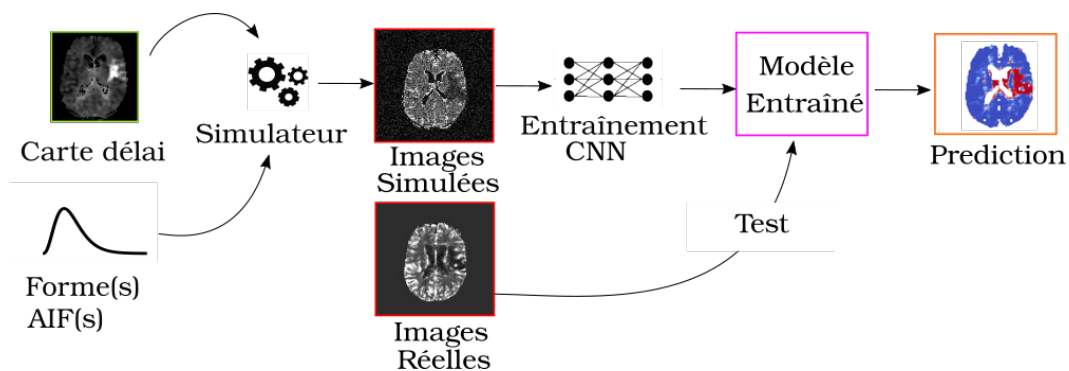


FIGURE II.6.4 – Pipeline général de l’approche proposée.

II.6.3.1 Données cliniques IRM

Nous considérons les 8 mêmes patients test de la section précédente (section II.5). Pour rappel, ces patients ont reçu à l’admission une imagerie pondérée de diffusion (DWI), une imagerie FLAIR, une imagerie pondérée en T2*, une angiographie par temps de vol et une imagerie pondérée de perfusion (DSC-PWI). Une IRM FLAIR de suivi a été effectuée un mois après l’admission. Les IRM de perfusion ont été recalées, pour chaque coupe, en utilisant le premier point temporel comme référence pour tous les autres points temporels, avec une approche d’information mu-

tuelle maximale. La lésion finale a été segmentée pour chaque patient à partir de l'IRM FLAIR de suivi par 3 experts. Les volumes FLAIR-MRI ont recalés sur les volumes DSC-PWI. La matrice de transformation obtenue a ensuite été utilisée pour recaler le masque de la lésion finale dans l'espace du DSC-PWI.

Pour ce travail, nous procurons des informations complémentaires concernant les 8 patients sélectionnés. Ces patients sont représentatifs de différents niveaux d'occlusion dans l'artère cérébrale moyenne : 2 patients présente une occlusion du segment M1, 3 patients du segment M2, 2 patients du segment M3 et 1 patient dont l'occlusion n'est pas référencé). Ces patients présentent également différentes géométries du cercle de Willis avec un nombre d'artères communicantes différent : 3 patients avec 3 communications, 2 patients avec 2 communications, 2 patients avec 1 communication et 1 patient sans communication.

II.6.3.2 Données d'entraînement simulées

Pour chacun des 8 patients, 4 bases de données d'entraînement différentes ont été produites, chacune avec un degré de réalisme croissant. Dans la **base de données A**, les images ont été générées en utilisant une AIF global unique théorique, similaire quel que soit le patient testé. Les paramètres de l'AIF ont alors été réglés sur les paramètres par défaut du simulateur. Dans la **base de données B**, une AIF unique a également été utilisée pour toutes les simulations, mais variant d'un patient à l'autre. Cette valeur a été définie en fonction du signal médian de tous les AIF spécifiques extraites pour un patient donné (voir Figure II.6.2). Dans la **base de données C**, une AIF différente a été utilisée pour chaque simulation. Chaque AIF a été générée à partir de la représentation de Delaunay des AIF extraits pour un patient donné. Enfin, dans la **base de données D**, les simulations ont été réalisées à l'aide d'un ensemble de formes multiples d'AIF (comme dans la base de données C), mais en intégrant en supplément une carte de délai spécifique au patient testé.

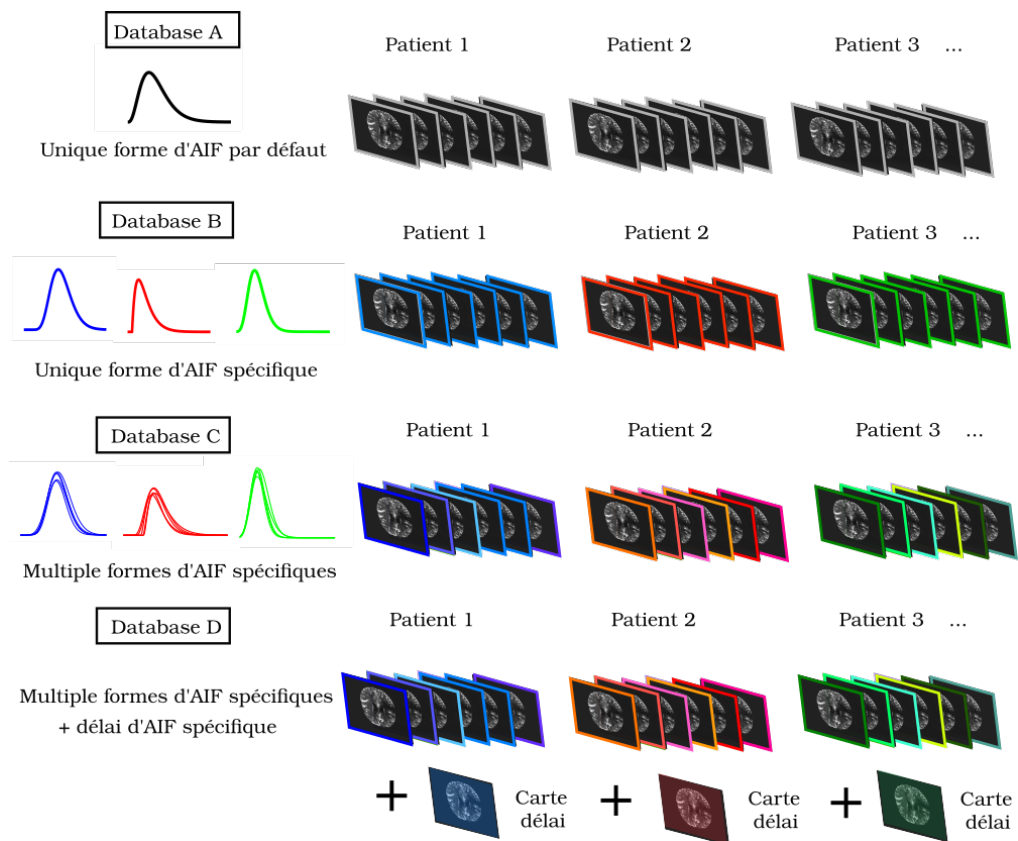


FIGURE II.6.5 – Les différentes bases de données prises en compte dans nos expériences. Chacune correspond à un degré de réalisme différent, qui augmente de la base de données A à la base de données D.

II.6.3.3 Classification par réseau de neurones

Un réseau de neurones parcimonieux, similaire à celui décrit dans la section II.5, a été construit afin de fournir une prédiction à l'échelle du voxel. Nous détaillons brièvement les caractéristiques générales de son architecture, mais sa description intégrale est fournie en section II.5. Le réseau prend directement en entrée des patchs spatio-temporels [19], et présente quatre couches au total : deux couches convolutionnelles 2D, suivies de deux couches entièrement connectées. Pour chaque patch, le réseau fournit en sortie une classification binaire (sain ou lésion). L'entraînement des réseaux dure globalement moins de 15 minutes sur une station de travail équipée d'une carte GPU NVIDIA GeForce GTX 1080 avec 128 Go de mémoire.

II.6.3.4 Détails expérimentaux

Une simulation est un volume de 25 coupes, obtenue en configurant le simulateur avec une ou plusieurs AIF et éventuellement avec une carte de délai. Chaque ensemble d'entraînement était constitué de 150 000 patchs tirés aléatoirement sans remise parmi 100 simulations (1 500 patchs tirés par simulation). Pour garantir un ensemble d'entraînement équilibré, la moitié des patchs d'entraînement correspondaient à des voxels lésionnels. L'ensemble de données de validation correspondant consistait en 150 000 patchs indépendants simulés. Pour tous les modèles d'entraînement, le nombre de poids à entraîner était de 197 087, le *dropout* était fixé à 0.5, le nombre d'époques à 30, la taille des *batches* à 32 et le taux d'apprentissage (*learning rate*) à 0.0001. Chaque modèle a été entraîné 10 fois, et le modèle plus performant a été donné dans les résultats. Chaque ensemble test était constitué de patchs d'un des 8 patients réels. Pour chaque base de données, 8 modèles ont été entraînés indépendamment (un modèle par patient testé) et nous avons évalué nos résultats en utilisant le coefficient de similarité Dice (DSC) et la distance de Hausdorff (HD). Ces métriques ont été calculées entre les voxels infarcis prédits et le masque de la lésion finale fourni par l'IRM FLAIR de suivi.

II.6.4 Résultats et Discussion

Les métriques DSC et HD associées à la prédiction des lésions à partir des 4 bases d'entraînement simulées sont indiquées dans le Tableau II.6.1. Chaque degré de réalisme supplémentaire apporté au simulateur a amélioré la performance globale de la classification sur les données réelles. Comme observé dans la section II.5, l'ajustement de l'AIF des données simulées aux données réelles est crucial : sans aucun ajustement (base de données A), le DSC moyen est de 0.15 contre 0.39 après ajustement avec une seule AIF spécifique (base de données B). L'intégration de la variabilité de flux artériel semble bénéfique également dans les performances générales : le DSC moyen est de 0.44 après ajustement avec plusieurs AIF spécifiques (base de données C). Les meilleures performances de classification sont obtenues avec la base de données D (c.a.d. en tenant compte des AIF multiples et du délai dans les données simulées d'entraînement), bien que l'amélioration par rapport à la base de données C soit modérée. Notons que la base A présente l'écart-type le plus faible parmi toutes les bases de données : comme tous les patients sont mal prédits, tous les scores sont faibles et proches les uns des autres.

Base de données d'entraînement	Réalisme	DSC	Hausdorff
Base A	Unique AIF par défaut	0.15 ± 0.062	46.31 ± 2.9
Base B	Unique AIF spécifique	0.39 ± 0.19	45.30 ± 2.5
Base C	Multiple AIF spécifiques	0.44 ± 0.18	44.76 ± 2.5
Base D	Multiple AIF spécifiques + délais	0.45 ± 0.18	44.66 ± 2.6

TABLEAU II.6.1 – Métriques DSC et HD entre les voxels prédits infarcis et la lésion finale de la vérité de terrain. La moyenne de toutes les mesures est calculée sur l'ensemble des données de test (moyenne ± écart type).

Il convient de souligner que tous les patients semblent bénéficier de degrés de réalisme de base (comme l'intégration de formes d'AIF spécifiques), tandis que seuls quelques patients bénéficient d'améliorations plus fines (comme la modélisation de l'arbre vasculaire). En particulier, l'intégration du délai dans la base de données d'entraînement synthétique améliore la prédiction des lésions uniquement pour les patients présentant un cercle de Willis complet, comme le montre la Figure II.6.6. Ces patients présentent une circulation alternative dans l'arbre vasculaire, et sont donc plus susceptibles de voir leur taille de lésion régresser.

À titre de comparaison, le meilleur modèle jusqu'à présent du challenge de prédiction des lésions ischémiques ISLES [17] a présenté un score DSC moyen de $0.38 (\pm 0.22)$. Nos mesures de prédiction sont du même ordre de grandeur, bien que nos modèles soient entraînés uniquement à partir de données simulées. Ces valeurs de DSC sont faibles et témoignent de la complexité de la tâche de prédiction. En effet contrairement à la tâche de segmentation classique, un délai d'un mois sépare l'entrée PWI-DSC de la vérité de terrain FLAIR : toutes les informations prédictives contenues dans PWI-DSC ne sont pas toujours suffisantes pour prédire précisément la lésion finale.

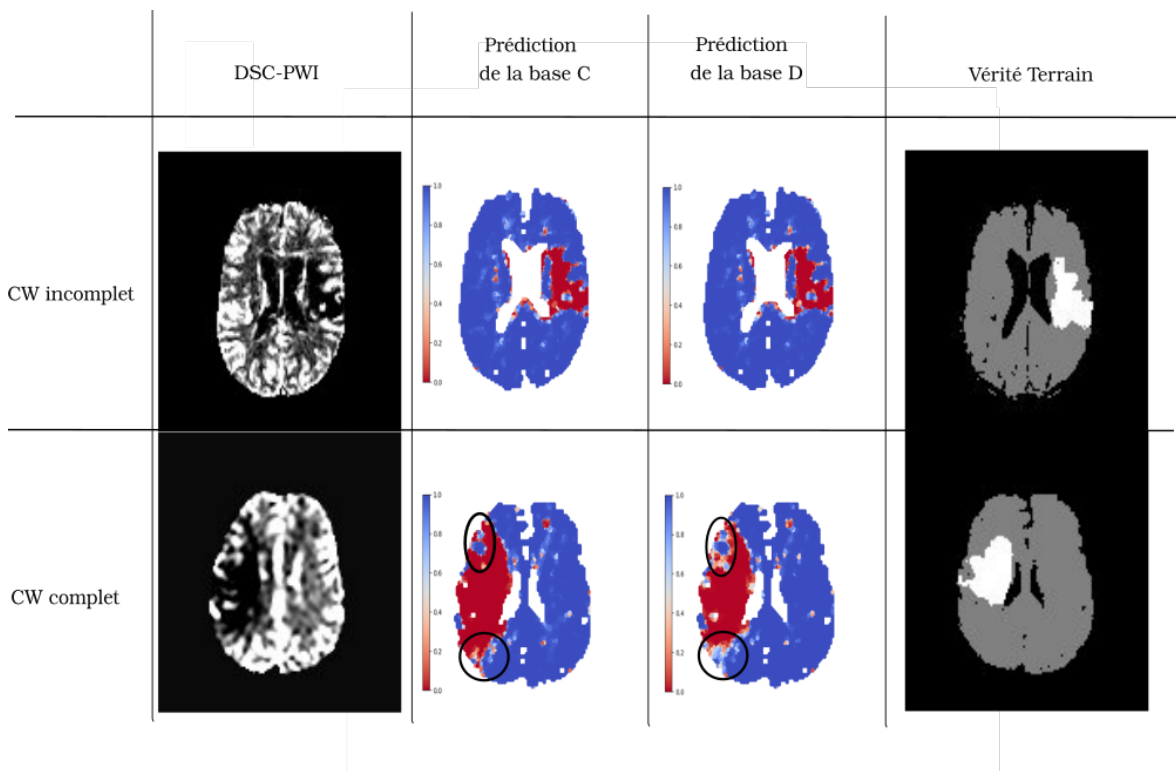


FIGURE II.6.6 – Prédictions obtenues après entraînement de la base de données C (2ème colonne) et de la base de données D (3ème colonne) pour deux patients testés. L'image DSC-PWI des patients testés est donnée en 1ère colonne. Les colonnes 2 et 3 doivent être comparées à la vérité de terrain (4ème colonne). Les voxels dans les tons bleus ont été prédits sains et les voxels dans les tons rouges ont été prédits infarcis. Le patient de la première ligne a un cercle de Willis (CW) incomplet alors que le patient de la deuxième ligne a un CW complet. Le délai dans la base de données D semble améliorer la prédiction de la lésion lorsque le cercle de Willis est complet (voir les cercles en gras).

Partie III

Couplage des imageries de diffusion et perfusion pour la prédiction de l'infarctus final

Chapitre III.7

Classification bayésienne naïve à l'échelle du voxel

Dans ce chapitre, nous proposons une approche bayésienne pour réaliser un apprentissage voxel-à-voxel de l'infarctus final à partir des imageries de perfusion et de diffusion. Au lieu de traiter chaque voxel indépendamment sur les imageries aiguës, nous proposons d'inclure les éléments du contexte local dans l'encodage. Nous montrons l'utilité d'une telle approche de modélisation pour aborder des questions pratiques. Nous discutons de l'influence de l'échelle d'observation de chaque voxel, de l'intérêt de l'encodage proposé pour l'estimation des probabilités conditionnelles du modèle bayésien, et du compromis entre le nombre de paramètres et la taille de l'ensemble d'entraînement.

III.7.1 Modélisation et fusion des imageries multimodales

III.7.1.1 Intégration et encodage des images

L'imagerie DWI donne une image 3D de la lésion initiale. L'imagerie PWI produit une image dynamique (3D + temps), à partir de laquelle nous extrayons, suite à une étape de déconvolution, plusieurs cartes paramétriques 3D. Nous avons généré cinq cartes parmi les plus communément utilisées : le flux cérébral sanguin (CBF), le volume cérébral sanguin (CBV), le temps de transit moyen (MTT), le temps au maximum (TMAX) et le temps au pic (TTP). L'imagerie DWI et les 5 cartes hémodynamiques (CBF, CBV, MTT, TMAX, TTP) sont seuillées pour ne travailler qu'avec des masques binaires respectivement notés X_1 et $(X_2, X_3, X_4, X_5, X_6)$. Ainsi, nous avons décrit chaque patient grâce à ces 6 cartes 3D binaires de la phase aiguë. L'IRM-FLAIR acquise 1 mois après l'admission du patient à l'hôpital rend compte de la lésion finale. Elle a été segmentée manuellement par des experts pour en extraire une image 3D binaire de l'infarctus final noté Y .

III.7.1.2 Échelle d'observation optimale pour la prédiction de l'infarctus final

Nous proposons d'inclure des informations sur l'état des voxels voisins des 6 cartes binaires de la phase aiguë, tout en conservant un petit nombre de valeurs possibles. Ainsi chaque voxel ne dépend plus seulement de sa propre valeur, mais aussi de la valeur de son voisinage $N \times N$ (en ignorant les voxels voisins en dehors de la région du cerveau) avec N l'échelle d'observation. Chaque voxel de chaque carte étant initialement binaire et donc encodé par deux valeurs, nous proposons désormais de l'encoder par six valeurs comme décrit dans la Table [III.7.1](#).

III.7.1.3 Classification bayésienne naïve

La prédiction de l'infarctus final est vu comme un problème de classification binaire où, sur la base de 6 cartes binaires de la phase aiguë, chaque voxel appartient à une des deux classes suivantes : sain ou lésion.

Valeur	Voxel central supérieur au seuil	% de voxels voisins supérieur au seuil
0	non	[0,25]
1	non]25,75[
2	non	[75,100]
3	oui	[0,25]
4	oui]25,75[
5	oui	[75,100]

TABLEAU III.7.1 – Encodage en 6 valeurs pour les cartes binaires prenant en compte, pour chaque voxel, l'état des voxels de son voisinage $N \times N$.

Nous avons fait le choix d'un classifieur bayésien naïf, dont le modèle sous-jacent estime les probabilités conditionnelles qu'une observation appartienne à une classe donnée compte tenu des valeurs de variables caractéristiques supposées indépendantes entre elles. Dans notre travail, une observation correspond à un voxel qui peut appartenir à 2 classes distinctes. Les variables caractéristiques correspondent aux 6 cartes binaires et sont supposées statistiquement indépendantes. Ainsi, la probabilité a posteriori qu'un voxel Y appartienne à la classe k étant donné n variables caractéristiques X_1, X_2, \dots, X_n peut être modélisée selon le théorème de Bayes :

$$P(Y = k | X_1, X_2, \dots, X_n) = \frac{P(Y = k) \prod_{i=1}^n P(X_i | Y = k)}{P(X_1, X_2, \dots, X_n)} \quad (\text{III.7.1})$$

où $P(Y = k)$ est la probabilité a priori que Y soit de classe k , $P(X_1, \dots, X_n | Y = k)$ est la vraisemblance, et $P(X_1, \dots, X_n)$ est l'évidence qui correspond à une constante (avec $n = 6$ et $k \in \{0, 1\}$). Les probabilités a priori sont estimées à partir d'un ensemble de voxels d'entraînement, où $P(Y = k)$ correspond au ratio du nombre de voxels observés de classe k sur le nombre de voxels observés au total. Nous présumons également la loi suivie par chacune des variables caractéristiques X_1, X_2, \dots, X_n :

- en encodant les 6 cartes X_1, X_2, \dots, X_6 comme des images binaires ($N = 0$), on fait l'hypothèse qu'elles suivent une loi de Bernoulli multivariée (qui correspond à un produit de lois de Bernoulli univariées)
- en encodant les 6 cartes X_1, X_2, \dots, X_6 comme des images multi-labels ($N > 0$, c.a.d. chaque voxel peut prendre 6 valeurs possibles : $\{0, 1, 2, 3, 4, 5\}$), on fait l'hypothèse qu'elles suivent une loi multinomiale.

Les paramètres des lois de probabilité présumées sont calculés sur la base de l'ensemble de voxels d'entraînement. À partir du modèle probabiliste décrit, le classifieur bayésien naïf peut être modélisé par une fonction C_{BN} qui classe les voxels par maximum a posteriori selon la règle de décision

$$C_{BN}(k, x_1, \dots, x_n) = \underset{k \in \{0,1\}}{\text{arg max}} P(Y = k) \prod_{i=1}^n P(X_i = x_i | Y = k) \quad (\text{III.7.2})$$

où x_1, x_2, \dots, x_n sont les valeurs prises par les variables caractéristiques X_1, X_2, \dots, X_n d'un voxel à classer.

III.7.2 Évaluation des performances

III.7.2.1 Métriques d'évaluation

Le degré de similarité spatiale entre la lésion prédite par le modèle bayésien et la lésion apparente sur l'imagerie FLAIR (vérité terrain) a été quantifié à l'aide du score Dice (DSC), calculé selon

$$\text{DSC} = \frac{2VP}{2VP + FN + FP} \quad (\text{III.7.3})$$

où TP correspond aux voxels vrais positifs, FN aux voxels faux négatifs et FP aux voxels faux positifs. Les performances de classification ont été étudiées pour plusieurs combinaisons de patients possibles : nous avons considéré i patients d'apprentissage parmi un total de 9 patients, pour $(9 - i)$ patients test ($i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$).

III.7.2.2 Matériel

Nous avons utilisé des données cliniques du projet européen I-KNOW [48]. Tous les patients présentaient un AVC de la circulation antérieure. À l'admission à l'hôpital, des images PWI et DWI ont été réalisées sur chaque patient. Une imagerie de suivi FLAIR a été également réalisée un mois après l'admission. Pour cette étude pilote, nous avons sélectionné un nombre restreint de 9 patients admis à l'hôpital dans les 3 heures suivant l'apparition des premiers symptômes. Les artefacts de mouvement des images PWI ont été corrigés. Ces mêmes images PWI ont également été déconvoluées par décomposition singulière circulaire afin d'obtenir les cartes paramétriques hémodynamiques. Ces cartes ont ensuite été normalisées en définissant manuellement une région de référence. Enfin, les images DWI et PWI ont été recalées sur le FLAIR par déformation rigide pour chacun des patients. Les masques binaires de la lésion finale à partir de l'imagerie FLAIR (Y) et de l'œdème cytotoxique à partir de l'image DWI (X_1) ont été segmentés manuellement par des experts. Les masques binaires issus des cartes hémodynamiques (X_2, X_3, X_4, X_5, X_6) ont été obtenus par seuillage. Le seuil a été défini comme la valeur associée au point le plus proche du point idéal (1,1) de la courbe ROC au regard de la lésion finale (FLAIR). La courbe ROC a été calculée à partir de voxels d'une région d'intérêt, avec un ratio de voxels sains et de voxels infarctés de 0.5.

III.7.3 Résultats

III.7.3.1 Information prédictive des modalités d'imagerie d'apprentissage

Nous avons représenté la vraisemblance de chacune des cartes binaires sur la Figure III.7.1. Toutes les cartes semblent contribuer à la prédiction de l'infarctus final, en présentant des probabilités supérieures à 0.2. Cependant les images DWI, CBF et CBV semblent les plus prédictives en présentant des probabilités respectives de 0.59, 0.65 et 0.54. Ces probabilités de vraisemblance semblent dépendantes du choix des patients dans l'ensemble d'apprentissage. Cependant quelle que soit la combinaison de patients, la tendance reste la même.

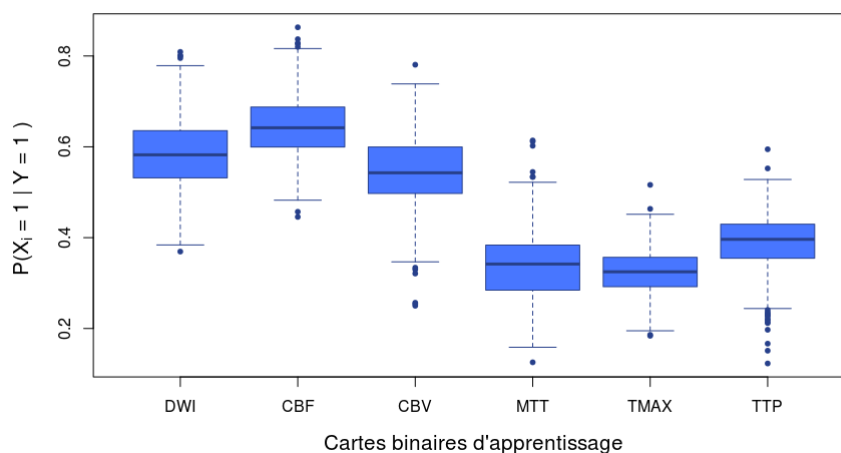


FIGURE III.7.1 – Vraisemblances des 6 cartes binaires pour la classe lésion, pour toutes les combinaisons de patients possibles dans l'ensemble d'apprentissage.

III.7.3.2 Impact du choix des patients dans l'apprentissage

Nous avons comparé les performances de classification du modèle bayésien naïf selon différentes combinaisons de patients dans l'ensemble d'apprentissage. La Figure III.7.2 présente les statistiques descriptives du DSC de la classe "lésion", pour un voisinage $N = 0$. Il apparaît qu'en augmentant le nombre de patients dans l'ensemble d'apprentissage, le DSC moyen varie peu : le DSC moyen est représenté par un trait horizontal noir sur chaque distribution violette de la Fig III.7.2 et a une valeur moyenne stabilisée à 0.6. Les résultats de classification apparaissent plutôt cohérents : les voxels classés comme "lésion" se situent bien au niveau de lésion finale, avec quelques erreurs près des ventricules comme le montre la Figure III.7.3 pour un patient testé. En revanche, tous les patients testés ne sont pas tous aussi bien classés. La Figure III.7.2 représente les résultats de DSC pour toutes les combinaisons de patients d'apprentissage possibles par un trait horizontal vert. Lorsqu'on apprend sur 8 patients et qu'on teste individuellement un 9ème patient, les résultats de DSC varie entre 0.32 et 0.74. Ceci nous indique que tous les patients testés ne présentent pas tous la même évolution de lésion, et que le modèle parvient à prédire la majorité des patients mais certains restent difficiles à prédire.

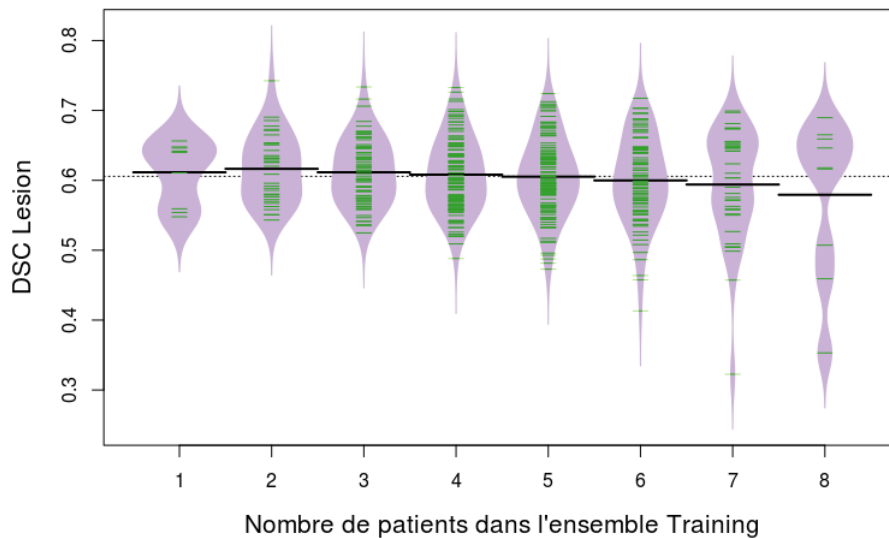


FIGURE III.7.2 – DSC de la classe "lésion" en fonction du nombre de patients dans l'ensemble d'apprentissage, pour $N = 0$. Les lignes vertes correspondent à un résultat DSC d'apprentissage individuel, tandis que les aires violettes représentent la distribution DSC pour plusieurs combinaisons de patients. Toutes les combinaisons possibles de patients sont représentées.

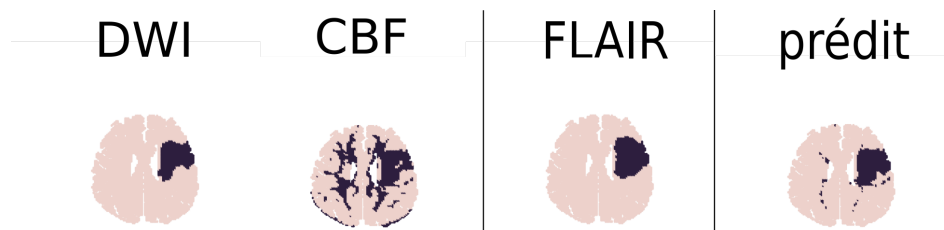


FIGURE III.7.3 – Résultats du modèle bayésien naïf pour une coupe d'un patient testé. Deux cartes hémodynamiques X_i (DWI et CBF) des six utilisées pour l'apprentissage ainsi que l'imagerie FLAIR final Y sont représentées binarisées. En rose sont représentés les voxels considérés "sain" et en violet les voxels considérés "lésion".

III.7.3.3 Échelle d'observation optimale pour la prédiction de l'infarctus final

La Table III.7.2 présente l'évolution du DSC moyen des différentes combinaisons de patients en fonction de l'échelle d'observation N de l'environnement local de chaque voxel. On constate que plus l'échelle est grande, meilleures sont les performances de classification. Cependant l'augmentation semble atteindre un palier à mesure que l'échelle est grande : le DSC de la lésion augmente de 0.7 entre un voisinage 0 et un voisinage 1, tandis qu'il augmente moins de 0.1 au-delà d'un voisinage 3. Le DSC des voxels sains présente la même tendance bien que l'augmentation soit plus modérée. L'échelle d'observation N = 3 semble ici optimale.

TABLEAU III.7.2 – Coefficient de similarité DSC des tissus infarctés et des tissus sains en fonction de l'échelle d'observation N de chaque voxel. Les mesures sont moyennées sur toutes les combinaisons de patients possibles (moyenne \pm écart-type).

N	DSC lésion	DSC sain
0	0.606 \pm 0.0564	0.938 \pm 0.0187
1	0.666 \pm 0.0435	0.948 \pm 0.0111
2	0.680 \pm 0.0436	0.952 \pm 0.0108
3	0.693 \pm 0.0432	0.954 \pm 0.0104
4	0.701 \pm 0.0427	0.956 \pm 0.0102
5	0.710 \pm 0.0422	0.957 \pm 0.0100
6	0.715 \pm 0.0420	0.958 \pm 0.00997
7	0.718 \pm 0.0414	0.959 \pm 0.00991

III.7.4 Discussion

Le classifieur naïf bayésien présente un nombre de paramètres NP particulièrement faible, estimé par

$$NP = (k - 1) + nrk, \quad (\text{III.7.4})$$

avec k le nombre de classes de la variable Y , n le nombre de variables caractéristiques X_i et r le nombre de valeurs possibles qu'elles prennent. Dans notre étude, le modèle présente donc 13 paramètres en encodant de façon binaire les voxels des différentes cartes, ou 73 paramètres en encodant les voxels des différentes cartes avec 6 valeurs différentes. Le faible nombre de paramètres permet d'apprendre à partir d'un nombre relativement faible d'observations, comme nous avons pu le constater dans nos résultats. Cette particularité du modèle bayésien naïf est d'autant plus intéressante dans le domaine de l'imagerie médicale où le nombre de données reste souvent le facteur limitant des analyses. Outre le faible nombre de paramètres, le modèle bayésien naïf a l'avantage également d'être très peu coûteux en temps de calcul : une fois les données pré-traitées, la prédiction pour une combinaison de patients donnée est de l'ordre de la seconde. Ces propriétés présentent des avantages considérables au regard des réseaux de neurones convolutionnels largement utilisés dans la littérature pour ce type de problème [15].

Dans nos résultats, nous avons pu observer l'importance de considérer chaque voxel avec son environnement local : les performances de classification augmentent à mesure qu'on considère un voisinage de plus en plus étendu. Cependant, il faut souligner que nous avons travaillé avec des patients présentant des lésions volumineuses et compactes (en moyenne : 579 voxel², soit le double du volume moyen de la cohorte). L'environnement local optimal que nous avons identifié s'avère donc spécifique à des patients présentant de grosses lésions. En intégrant des patients avec des lésions plus petites ou plus éclatées, l'environnement local optimal serait sûrement plus faible, au risque de sous-estimer l'infarctus final. Il serait alors nécessaire de stratifier les patients selon leur forme de lésion pour modéliser finement les règles de décisions et les échelles d'observations de façon optimales.

Notons enfin que pour ce travail, nous avons tiré profit de l'encodage simple des imageries aiguës, dans la mesure où il a permis d'estimer précisément les lois de probabilité suivies par les variables caractéristiques. Sans l'étape de binarisation des imageries aiguës, les variables caractéristiques correspondraient à des niveaux de gris dont le comportement aurait été difficilement modélisable, et les coûts computationnels du modèle auraient été bien plus grands. La binarisation participe énormément à contraindre l'apprentissage bayésien. En ce sens, la fusion que nous avons réalisé peut être assimilée à une fusion de décision, puisque la règle de décision du modèle bayésien intervient en aval de décisions prises sur l'encodage des données d'entrée. On comprend ainsi que la performance finale du modèle bayésien repose grandement sur la pertinence de seuillage des imageries aiguës.

Chapitre III.8

Segmentation sémantique par réseau de neurones profond à l'échelle de l'image

Dans ce chapitre, nous proposons de répondre à la tâche de prédiction de l'infarctus final en couplant l'imagerie de perfusion et de diffusion au travers d'une approche réseau de neurones profond. Contrairement au chapitre précédent, aucun pré-traitement ni seuillage n'est réalisé en amont de l'apprentissage : les images brutes en niveau de gris sont données directement au réseau. L'apprentissage profond est un ensemble de méthodes d'apprentissage flexibles, basées sur les données, capables d'extraire automatiquement des caractéristiques non linéaires, avec des résultats prometteurs en segmentation de lésions cérébrales [84]. En contrepartie, l'apprentissage profond présente un très grand nombre d'hyperparamètres à ajuster lors de la phase de l'apprentissage, nécessitant ainsi une grande quantité de données d'entraînement et de validation.

Seul un nombre limité d'études, avec des paradigmes de traitement et des mesures d'évaluation hétérogènes, ont évalué les réseaux de neurones profonds pour la prédiction de la lésion cérébrale finale à partir d'IRM [17, 18, 15, 85] ou CT aigus [20]. Dans ce chapitre, nous avons évalué l'impact de l'intégration du statut de reperfusion sur la performance des réseaux de neurones pour prédire l'infarctus final. La reperfusion est la métadonnée clinique la plus importante connue pour influencer la progression des lésions ischémiques depuis l'imagerie aiguë jusqu'à l'infarctus final [11]. Nous avons émis l'hypothèse qu'entraîner des réseaux à partir de sous-ensembles de patients ayant un statut de reperfusion homogène pourrait améliorer leurs performances. Aussi, nous avons confronté la valeur prédictive de nos réseaux avec celle des modèles cliniques actuels de mismatch perfusion-diffusion. Enfin, nous avons étudié l'importance prédictive relative des données de l'IRM au travers d'une étude d'ablation.

III.8.1 Stratification des données d'entraînement selon la variable de reperfusion précoce

III.8.1.1 Présentation des données IRM cliniques

Patients et protocole d'imagerie Les patients inclus dans cette étude proviennent des cohortes HIBISCUS-STROKE et I-KNOW. HIBISCUS-STROKE est une cohorte d'observation monocentrique en cours qui regroupe des patients présentant une occlusion artérielle intracrânienne importante et traités par thrombectomie, après une IRM de diffusion et de perfusion aiguës. I-KNOW (2007-2011) était une étude observationnelle multicentrique prospective de patients victimes d'un accident vasculaire cérébral. Une IRM de suivi de perfusion aiguë (3 heures à partir de l'IRM de base) a été acquise sur un sous-ensemble des patients I-KNOW afin d'évaluer leur reperfusion précoce [86]. Les patients des deux cohortes ont donné leur consentement éclairé et le protocole d'imagerie a été approuvé par des comités régionaux d'éthique.

Les critères d'inclusion pour HIBISCUS-STROKE étaient : (1) les patients présentant un AVC de la circulation antérieure lié à une occlusion intracrânienne proximale (artère carotide interne,

occlusion M1 ou M2); (2) les patients présentant une IRM de diffusion et de perfusion aiguës; (3) les patients traités par thrombectomie avec ou sans thrombolyse intraveineuse. Les critères d'inclusion et d'exclusion pour I-KNOW étaient : (1) NIHSS ≥ 4 ; (2) IRM de diffusion et de perfusion correspondant à un accident ischémique cérébral aigu de la circulation antérieure; et (3) IRM à l'admission effectuée dans les 6 heures pour les patients traités par thrombolyse intraveineuse, ou dans les 12 heures pour ceux traités sans thrombolyse. Les patients ayant subi un accident vasculaire cérébral de circulation lacunaire ou postérieur, ou une hémorragie intracérébrale ont été exclus. Aucun patient dans HIBISCUS-STROKE n'a reçu de thrombectomie.

Dans les deux cohortes, au moment de leur admission, les patients ont tous été soumis au protocole d'IRM suivant : une IRM pondérée de diffusion (DWI), une IRM FLAIR, une IRM pondérée T2*, une angiographie RM et IRM pondérée de perfusion (DSC-PWI) à contraste de susceptibilité (DSC-PWI). Une IRM FLAIR de suivi a été effectuée plusieurs jours après l'admission (plus précisément, 6 et 30 jours pour HIBISCUS et I-KNOW, respectivement).

Post-traitements des images Des cartes paramétriques ont été extraites de l'IRM de perfusion par décomposition circulaire en valeurs singulières des courbes de concentration tissulaire : le débit sanguin cérébral (CBF), le volume sanguin cérébral (CBV), le temps de transit moyen (MTT), le temps au maximum (T_{max}) et le temps au pic (TTP). Les lésions de l'IRM DWI aiguë et de l'IRM FLAIR de suivi ont été segmentées par un expert (THC) à l'insu des données cliniques à l'aide d'une méthode semi-automatique (3D Slicer, <https://www.slicer.org/>). Plus précisément, un seuillage de la région d'intérêt a été utilisé avec des corrections manuelles si nécessaire (pour la lésion de diffusion, un seuil supérieur à $620 \times 10^{-6} \text{ mm}^2/\text{s}$ a été utilisé).

Pour chaque patient d'HIBISCUS-STROKE, les images ont été recalées sur l'IRM de diffusion aiguë en utilisant un recalage non linéaire Ants [87]. Pour chaque patient de I-KNOW, les imageries ont été recalées sur l'IRM de perfusion (matrice 128x128) en utilisant un recalage affine (Statistical Parametric Mapping 8). Davantage de détails de recalage sont donnés en Annexe A. Une fois le recalage effectué, les patients HIBISCUS-STROKE présentaient des coupes d'IRM de taille 192x192 contre 128x128 pour les patients I-KNOW. Comme les patients I-KNOW étaient largement minoritaires (17 patients sur un total de 109), nous avons rééchantillonné leurs images à 192x192. La boîte crânienne de tous les patients a été retirée en utilisant FSL [88]. Enfin, les images ont été normalisées entre 0 et 1 afin d'assurer une standardisation entre les patients.

III.8.1.2 Intégration de la reperfusion précoce dans les ensembles d'entraînement

Évaluation de la reperfusion précoce Dans HIBISCUS-STROKE, la reperfusion précoce a été évaluée à la fin de la procédure endovasculaire avec le score mTICI (pour *modified Thrombolysis in Cerebral Infarction*, avec grade 0 : pas de reperfusion; grade 1 : reperfusion antérograde au-delà de l'occlusion initiale, mais remplissage sanguin limité de la branche distale avec peu ou pas de reperfusion distale; grade 2a : reperfusion antérograde de moins de la moitié du territoire ischémique de l'artère cible précédemment occluse; grade 2b : reperfusion antérograde de plus de la moitié du territoire ischémique de l'artère cible précédemment occluse; grade 2c : reperfusion presque complète, c.a.d. $>90\%$ mais moins que mTICI 3; grade 3 : reperfusion antérograde complète) [89]. La reperfusion angiographique a été définie comme l'ensemble des scores mTICI de 2b à 3, alors que la non reperfusion comme l'ensemble des scores mTICI de 0 à 2a.

Dans I-KNOW, aucun patient n'a été traité par des procédures endovasculaires. La reperfusion précoce a été évaluée 3 heures après la première IRM (H3) et a été définie comme les voxels dont $T_{max} \geq 6 \text{ s}$ à l'admission (H0) et $T_{max} < 6 \text{ s}$ à H3. La reperfusion aiguë a été définie par un ratio de reperfusion (volume des voxels reperfusés à H3/volume de la lésion de reperfusion à H0) $\geq 50\%$.

Ensembles d'entraînement Trois ensembles d'entraînement distincts et les trois modèles associés ont été construits pour évaluer l'impact de la reperfusion sur la précision de la prédiction de l'infarctus final : un modèle "général", entraîné sur la cohorte entière indépendamment du statut de reperfusion; un modèle "reperfusé", entraîné uniquement avec les patients reperfusés; un

modèle "non reperfusé", entraîné uniquement avec les patients non reperfusés. Étant donné le taux élevé de succès angiographique chez les patients traités par thrombectomie (score mTICI de 2b à 3 pour plus de 70% des patients) [90], nous nous attendions à une proportion limitée de patients non reperfusés dans HIBISCUS-STROKE. Nous avons donc inclus les patients sans re-perfusion précoce de I-KNOW (identifiés par l'IRM de perfusion de suivi H3) afin de contrebalancer ce déséquilibre.

III.8.2 Stratégies de fusion proposées pour l'apprentissage par réseau de neurones

III.8.2.1 Architecture CNN proposée

Nous avons utilisé une architecture U-Net [91] : un réseau multi-échelle qui a déjà été utilisé avec succès pour répondre à la tâche de prédiction de l'infarctus final [17, 85]. L'IRM de perfusion et de diffusion ont été donnés en entrée du réseau, car les deux modalités sont complémentaires pour évaluer le risque d'infarctus [8]. Des études récentes sur d'autres applications médicales ont évalué plusieurs méthodes permettant de combiner les images données en entrée du CNN, montrant le bénéfice des stratégies de fusion de caractéristique tardive [30, 92, 31, 93]. Contrairement à la fusion précoce qui concatène les entrées à un faible niveau de représentation (c'est-à-dire avant tout processus de convolution) [17, 15, 85], la fusion tardive incorpore les images d'entrée indépendamment dans des branches convolutionnelles distinctes, fusionnant ensuite les caractéristiques à des niveaux supérieurs. Nous avons fait le choix d'une fusion de caractéristiques tardive pour notre tâche de prédiction, car cette stratégie a le potentiel de mieux intégrer chaque IRM d'entrée, et par extension l'état de reperfusion de l'ensemble d'entraînement. La comparaison des stratégies de fusion précoce et tardive est donnée dans les résultats.

Cinq images (DWI, ADC, T_{max} , CBV, CBF) ont été données en entrée du réseau, chacune destinée à être convoluée sur une branche distincte. L'architecture de fusion tardive proposée est représentée en Figure III.8.1, et ses couches contractantes sont détaillées dans le Tableau III.8.1. Chaque entrée était constituée de coupes IRM 2D entières (192x192). Aucun patch n'a été utilisé afin de garantir un contexte spatial large pour la prédiction des lésions. Le réseau a produit en sortie des cartes de probabilité avec 3 classes : lésion, tissu sain, fond. La carte de probabilité de lésion a été seuillée à 0.5 pour définir l'infarctus final.

Couche (type)	Taille Sortie
Conv Block 1	192*192*8
2D MaxPooling	96*96*8
Conv Block 2	96*96*16
2D MaxPooling	48*48*16
Conv Block 3	48*48*32
2D MaxPooling	24*24*32
Conv Block 4	24*24*64
Dropout + 2D Maxpooling	12*12*64
Conv Block 5 + Dropout	12*12*128
Concatenation	12*12*640

TABLEAU III.8.1 – Partie contractante du CNN proposé, composée de 5 blocs de convolution (Conv Block), d'opérations de maxpooling (2D MaxPooling) et de *dropout*. Chaque Conv Block est composé de : Convolution 2D (3*3) + *batch normalisation* + convolution 2D (3*3) + *batch normalisation*.

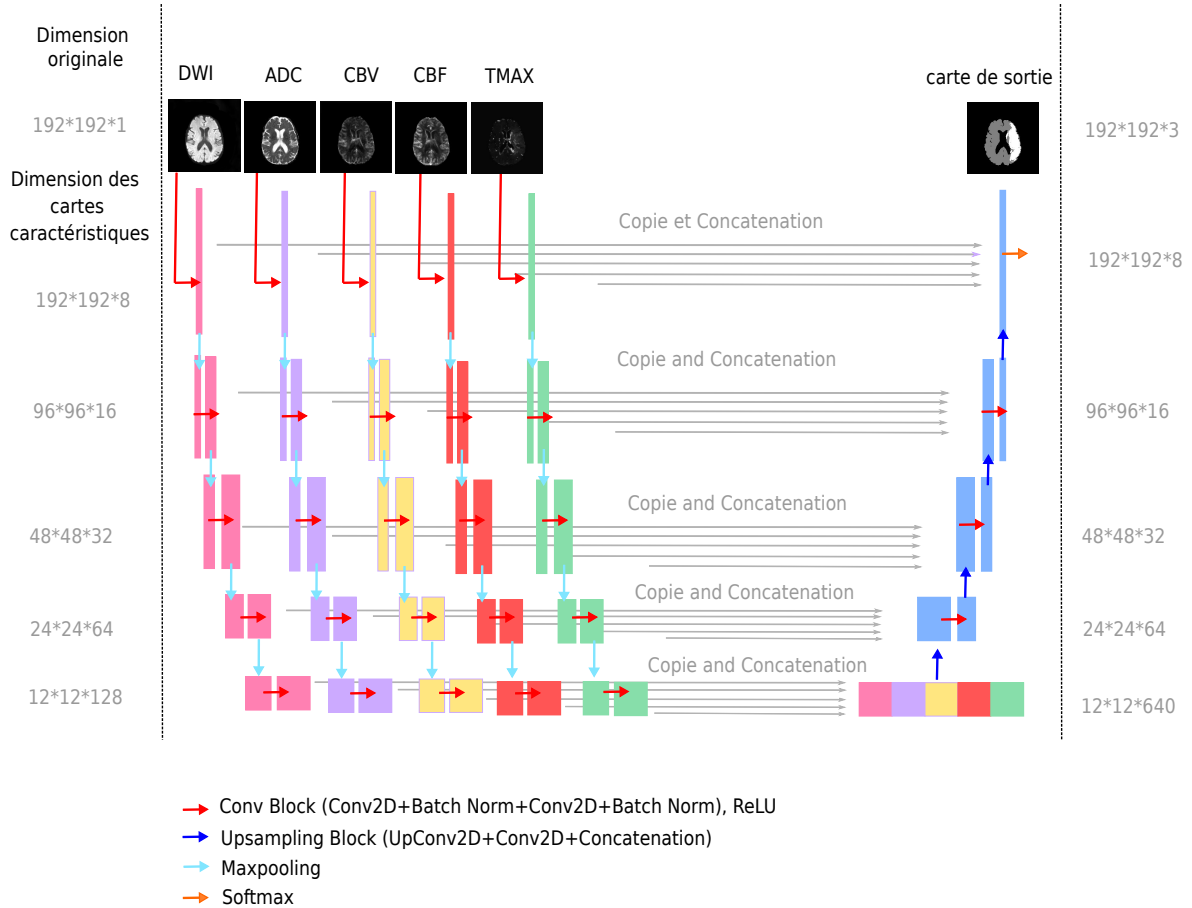


FIGURE III.8.1 – Aperçu de l’architecture CNN proposée. En haut à gauche : Le réseau prend en entrée cinq images IRM (coupes 2D des volumes DWI, ADC, CBV, CBF, T_{max}). En bas : Chaque image d’entrée est traitée indépendamment sur 5 branches distinctes. Les cartes de caractéristiques roses, violettes, jaunes, rouges et vertes résultent de convolutions 2D et d’opérations de maxpooling. Les sorties des 5 branches sont ensuite concaténées, et sur-échantillonnées par des couches de déconvolution 2D. Le réseau produit en sortie une carte avec 3 classes (lésion, tissu sain et fond). En haut à droite : La lésion prédite doit être comparée à la lésion réelle du FLAIR final.

III.8.2.2 Entraînement et paramétrage du réseau

Le réseau a été entraîné uniquement à partir de coupes présentant une lésion, et aucune augmentation de données n’a été utilisée. Nous avons utilisé une fonction Dice multi-classes \mathcal{L}_{MD} comme fonction de perte (*loss*) [94], pour laquelle nous avons attribué à la classe lésion un poids 8 fois supérieur à celui des classes saines et fond, définie par :

$$\mathcal{L}_{MD} = 1 - \frac{1}{10} (\mathcal{L}_{DSCfond} + \mathcal{L}_{DSCsain} + 8\mathcal{L}_{DSClesion}) \quad (\text{III.8.1})$$

avec \mathcal{L}_{DSC} une fonction Dice continue et dérivable d’une classe donnée (lésion, sain ou fond), telle que :

$$\mathcal{L}_{DSC} = \frac{2 \sum_i^N s_i r_i}{\sum_i^N s_i + \sum_i^N r_i} \quad (\text{III.8.2})$$

où s_i et r_i représentent respectivement les valeurs continues de la carte de probabilité $\in [0, \dots, 1]$ et les valeurs de la vérité terrain pour chaque voxel i , et N le nombre de voxel de la classe en question.

Nous avons utilisé l’optimiseur Adam ($lr = 1 \times 10^{-4}$ et $decay = 5 \times 10^{-4}$) et une taille de *batch* de 12. Pour éviter le sur-apprentissage, nous avons appliqué du *dropout* (fixé à 0.5), utilisé un régularisateur L2 *reg* à chaque couche de convolution ($reg = 2 \times 10^{-4}$) et le nombre d’époques

(fixé à 500) a été régulé par *early stopping* (c.a.d. l'entraînement a été arrêté une fois que la fonction Dice multi-classes de validation n'a pas augmenté de plus de 0.005 sur 100 époques). L'évaluation de chaque modèle a été réalisée à l'aide d'une validation croisée 5-Folds.

Le nombre de paramètres est proportionnel au nombre de branches du U-Net : ainsi, le nombre de paramètres pouvant être entraînés est de 1 997 851 pour une architecture U-Net avec 5 images d'entrée, 1 242 603 pour 3 images d'entrée, et 487 355 en utilisant une seule image d'entrée. Plus le nombre de branches est élevé, moins les informations sont compressées et plus l'architecture offre la possibilité d'apprendre des informations différentes sur chaque donnée d'entrée. Ainsi, nous avons choisi de ne pas équilibrer le nombre de paramètres entre chaque architecture. Toutefois, pour assurer une comparaison équitable, les hyperparamètres de chaque réseau ont été ajustés indépendamment sur un espace de recherche fixe. Les meilleurs paramètres se sont avérés être les mêmes pour toutes les architectures testées. Nous avons utilisé la bibliothèque Keras 2.1.3 avec une interface Python 3.6.3. La phase d'entraînement a duré environ 1 heure sur un poste de travail équipé d'un GPU NVIDIA GeForce GTX 1080 avec 128 Go de mémoire.

III.8.3 Évaluation des performances

III.8.3.1 Métriques

Nous avons utilisé les métriques d'évaluation standards pour les tâches de segmentation et prédiction, à savoir : le coefficient de similarité Dice (DSC), la précision, le rappel, la similarité volumétrique (VS) et la distance de Hausdorff (HD). Le DSC mesure le chevauchement relatif de la prédiction avec la vérité de terrain (VP, FN et FP sont respectivement les voxels vrais positifs, faux négatifs et faux positifs) :

$$\text{DSC} = \frac{2 \cdot \text{VP}}{\text{FN} + \text{FP} + 2 \cdot \text{VP}} \quad (\text{III.8.3})$$

La précision (également appelée valeur prédictive positive) mesure le pourcentage de voxels identifiés comme lésion qui ont été classés correctement, tandis que le rappel (également appelé sensibilité) mesure le pourcentage de voxels lésionnels réels correctement classés :

$$\text{Précision} = \frac{\text{VP}}{\text{VP} + \text{FP}}, \quad (\text{III.8.4})$$

$$\text{Rappel} = \frac{\text{VP}}{\text{TP} + \text{FN}}. \quad (\text{III.8.5})$$

Le VS donne un rapport relatif entre les volumes de prédiction et de vérité de terrain, sans tenir compte d'un quelconque chevauchement entre les deux volumes :

$$\text{VS} = 1 - \frac{|\text{FN} - \text{FP}|}{2 \cdot \text{VP} + \text{FP} + \text{FN}}. \quad (\text{III.8.6})$$

Le HD est une mesure la distance de la plus grande erreur entre la prédiction (A) et la vérité terrain (B) :

$$\text{HD}(A, B) = \max(h(A, B), h(B, A)) \quad \text{avec} \quad h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|. \quad (\text{III.8.7})$$

L'aire sous la courbe (AUC) est une métrique d'évaluation largement utilisée dans le domaine médical. Basée sur la courbe ROC [95], elle fournit une mesure agrégée des performances pour tous les seuils de classification possibles d'une modalité d'image ou d'une carte paramétrique. Cependant, étant donné que les voxels sains sont en très large majorité par rapport aux voxels lésionnels, cela peut entraîner des valeurs élevées d'AUC sans pour autant que l'étendue et la localisation de l'infarctus soient correctement prédits [96]. Plusieurs études recommandent donc la métrique DSC, qui est plus spécifique pour la prédiction des lésions [97, 85]. En fin de chapitre, nous donnons quand même les valeurs d'AUC associées aux performances de nos modèles proposés afin de faciliter les comparaisons avec certaines études de l'état de l'art [15, 85].

III.8.3.2 Modèle du mismatch perfusion-diffusion

Nos modèles prédictifs basés CNN ont été comparés au modèle mismatch perfusion-diffusion de référence actuellement utilisé dans la pratique clinique. Selon cette méthode, l'infarctus final peut être prédit comme suit : (1) chez les patients reperfusés, l'infarctus final est représenté par la lésion de diffusion aiguë; (2) chez les patients non reperfusés, l'infarctus final est défini comme l'union de la lésion de diffusion aiguë et de la pénombre ischémique (voxels avec une $T_{max} > 6$ secondes et une DWI normale)[77]. L'AUC du modèle mismatch perfusion-diffusion pour prédire l'infarctus final a été évaluée pour les patients avec et sans reperfusion. Les voxels non infarctés étaient ceux qui n'étaient pas inclus dans la lésion de diffusion chez les patients reperfusés, et ceux qui n'étaient pas inclus dans la pénombre de diffusion chez les patients non reperfusés. Les voxels complémentaires correspondaient aux voxels infarctés. L'AUC a été calculée selon les recommandations de Jonsdottir et al [96].

III.8.3.3 Analyses statistiques

Un test bilatéral de Wilcoxon non paramétrique a été effectué afin de comparer les performances des modèles (1) reperfusé *vs* général, non-reperfusé *vs* général et reperfusé *vs* non-reperfusé; (2) modèles avec toutes les entrées IRM *vs* modèles avec ablation d'une ou plusieurs entrées IRM; (3) modèle reperfusé *vs* modèle de lésion de diffusion; (4) modèle non-reperfusé *vs* modèle de lésion de diffusion \cup pénombre. Les analyses statistiques ont été effectuées à l'aide de la version 3.5.1 de R.

III.8.4 Résultats

Nous avons analysé 109 patients, comme le montre l'organigramme en Figure III.8.2. Parmi les 109 patients analysés, 74 présentaient une reperfusion précoce, tandis que 35 n'avaient pas de reperfusion (17 de I-KNOW et 18 d'HIBISCUS-STROKE). Les caractéristiques des patients sélectionnés sont résumées dans Tableau III.8.2.

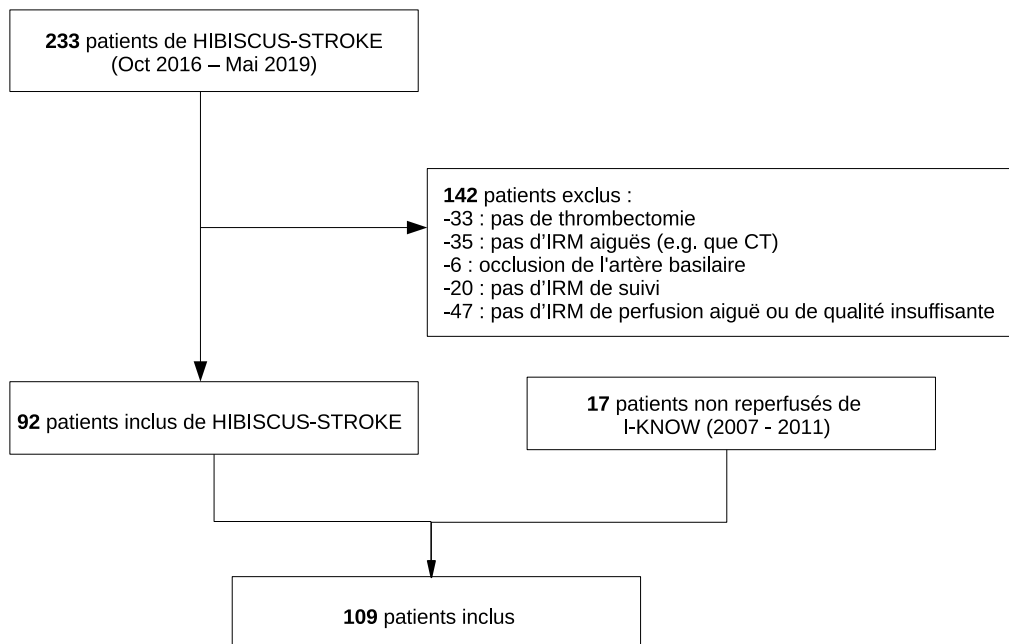


FIGURE III.8.2 – Organigramme d'inclusion des patients.

Variabes cliniques	
Femmes, n (pourcentage)	45 (41.3)
Âge	70 (57 - 79)
NIHSS score	15 (10 - 19)
Délai entre les symptômes et IRM	105 (78 - 154)
tPA intraveineuse, n (pourcentage)	59 (54.1)
Site d'occlusion, n (pourcentage) :	
ICA intracrânienne + M1	27 (24.8)
M1	54 (49.5)
ICA intracrânienne + M2	23 (21.1)
M2	5 (4.6)
ICA cervicale, n (pourcentage)	19 (17.4)
Taille lésion DWI, mL	24.9 (7.4 - 50.9)

TABLEAU III.8.2 – Caractéristiques cliniques des patients sélectionnés (médiane avec écart interquartile, sauf indication contraire). NIHSS : National Institutes of Health Stroke Scale; DWI : imagerie pondérée de diffusion; ICA : artère carotide interne.

III.8.4.1 Performances des CNN généraux, reperfusés et non reperfusés

Les performances et les comparaisons des modèles généraux, reperfusés et non reperfusés testés sur des patients reperfusés et non reperfusés sont présentées dans le Tableau III.8.3.

Parmi les patients reperfusés, les métriques du modèle non reperfusé étaient inférieures à celles du modèle reperfusé et du modèle général, à l'exception de la précision (Tableaux III.8.3-a et III.8.3-b). Inversement, aucune différence de performance nette n'a été constatée entre le modèle reperfusé et le modèle général.

Parmi les patients non reperfusés, le modèle non reperfusé présentait des performances supérieures ou similaires à celles du modèle reperfusé pour toutes les métriques, à l'exception du rappel (Tableaux III.8.3-c et III.8.3-d). Aucune différence globale nette n'a été observée entre les modèles non reperfusé et général, ou entre les modèles reperfusé et général.

Les volumes d'infarctus prédits étaient significativement plus importants avec le modèle non reperfusé qu'avec le modèle reperfusé (39,7 ml (61,3-20) contre 17,5 ml (28-5,1), $p = 4,5e-16$ pour les modèles non reperfusés et reperfusés, respectivement; médiane avec l'intervalle interquartile). En conséquence, des différences significatives de VS entre ces deux modèles ont été observées (Tableaux III.8.3-b et -d). La Figure III.8.3 illustre les prédictions des CNN reperfusé et non-reperfusé sur deux patients ayant un statut de reperfusion distinct.

III.8.4.2 Comparaison entre les modèles CNN et le modèle mismatch perfusion-diffusion

Chez les patients reperfusés et non reperfusés, le DSC, le VS et le rappel des modèles CNN étaient supérieurs à ceux des modèles mismatch perfusion-diffusion (Tableau III.8.4). À l'échelle du patient, des valeurs DSC supérieures ont été obtenues avec les modèles CNN chez 68% et 89% des patients reperfusés et non reperfusés, respectivement. À l'inverse, la précision des modèles mismatch était plus élevée que celle du CNN, ce qui suggère un plus grand nombre de voxels faux positifs avec ces dernières méthodes.

Les modèles CNN ont obtenu des valeurs d'AUC plus élevées que celles des modèles mismatch perfusion-diffusion (patients reperfusés : 0.87 ± 0.13 contre 0.79 ± 0.17 , $P < 0,001$; patients non reperfusés : 0.81 ± 0.13 vs 0.73 ± 0.14 , $P < 0.01$, respectivement pour modèles CNN vs modèles mismatch). Des cas de succès et d'échec de prédictions sont présentés en Figure III.8.4.

(a) Mesure des performances chez les patients reperfusés

Modèle	DSC	VS	Précision	Rappel	HD
Général	0.43 ± 0.24	0.69 ± 0.27	0.55 ± 0.28	0.43 ± 0.25	33.23 ± 15.6
Reperfusé	0.44 ± 0.25	0.70 ± 0.27	0.50 ± 0.27	0.50 ± 0.26	38.58 ± 18.1
Non-Reperfusé	0.35 ± 0.21	0.57 ± 0.28	0.60 ± 0.25	0.31 ± 0.24	40.05 ± 15.6

(b) Comparaisons de modèles chez les patients reperfusés

Test bilatéral	DSC P-value	VS P-value	Précision P-value	Rappel P-value	HD P-value
Général vs Reperfusé	0.43	0.53	3.7e-6 (***)	1.4e-6 (***)	0.048 (*)
Général vs Non-Reperfusé	1.4e-8 (***)	4.3e-6 (***)	0.0069 (**)	1.0e-10 (***)	0.0041 (**)
Reperfusé vs Non-Reperfusé	2.3e-8 (***)	1.6e-5 (***)	2.9e-7 (***)	2.7e-11 (***)	0.65

(c) Mesure des performances chez les patients non reperfusés

Modèle	DSC	VS	Précision	Rappel	HD
Général	0.44 ± 0.21	0.66 ± 0.26	0.39 ± 0.25	0.63 ± 0.21	30.61 ± 16.1
Reperfusé	0.44 ± 0.22	0.63 ± 0.25	0.36 ± 0.23	0.69 ± 0.22	44.53 ± 16.7
Non-Reperfusé	0.47 ± 0.17	0.74 ± 0.13	0.49 ± 0.22	0.52 ± 0.21	37.70 ± 17.7

(d) Comparaisons de modèles chez les patients non reperfusés

Test bilatéral	DSC P-value	VS P-value	Précision P-value	Rappel P-value	HD P-value
Général vs Reperfusé	0.93	0.55	0.13	0.021 (*)	0.0023 (**)
Général vs Non-Reperfusé	0.17	0.21	0.0016 (**)	0.00084 (***)	0.11
Reperfusé vs Non-Reperfusé	0.13	0.034 (*)	0.00067 (***)	5.3e-5 (***)	0.12

TABLEAU III.8.3 – Mesure des performances des modèles généraux, reperfusés et non-reperfusés chez les patients (a) **reperfusés** et (c) **non-reperfusés** (valeurs moyennes ± écart type). Les valeurs en gras correspondent à la meilleure valeur de métrique en colonne. Les P-values des tests bilatéraux de Wilcoxon comparent les modèles généraux, reperfusés et non-reperfusés chez les patients (b) **reperfusé** et (d) **non-reperfusé**. Les P-values en gras correspondent à des différences significatives, (*) indiquant $P < 0.05$, (**) indiquant $P < 0.01$ et (***) indiquant $P < 0.001$.

Patients reperfusés

Model	DSC	VS	Précision	Rappel	HD
CNN	0.44 ± 0.21 (*)	0.66 ± 0.26 (***)	0.39 ± 0.25	0.63 ± 0.21 (***)	30.61 ± 16.1
Mismatch perfusion-diffusion	0.41 ± 0.23	0.56 ± 0.27	0.71 ± 0.31 (***)	0.33 ± 0.20	19.34 ± 10.3 (***)

Patients non-reperfusés

Modèle	DSC	VS	Précision	Rappel	HD
CNN	0.47 ± 0.17 (***)	0.74 ± 0.13 (***)	0.49 ± 0.22	0.52 ± 0.21 (***)	37.70 ± 17.7 (***)
Mismatch perfusion-diffusion	0.26 ± 0.17	0.31 ± 0.21	0.84 ± 0.16 (***)	0.17 ± 0.13	69.15 ± 7.7

TABLEAU III.8.4 – Comparaison entre les modèles CNN et les modèles mismatch perfusion-diffusion. Parmi les patients reperfusés (deux premières lignes), le modèle reperfusé CNN a été comparé à la lésion de diffusion aiguë. Parmi les patients non reperfusés (deux dernières lignes), le modèle non reperfusé CNN a été comparé à l'union de la lésion de diffusion aiguë et de la pénombre ischémique. Les valeurs en gras correspondent à la meilleure valeur de métrique en colonne. Un test bilatéral de wilcoxon a été effectué entre les modèles CNN proposés et les modèles cliniques mismatch, avec (*) indiquant $P < 0.05$, (**) indiquant $P < 0.01$ et (***) indiquant $P < 0.001$.

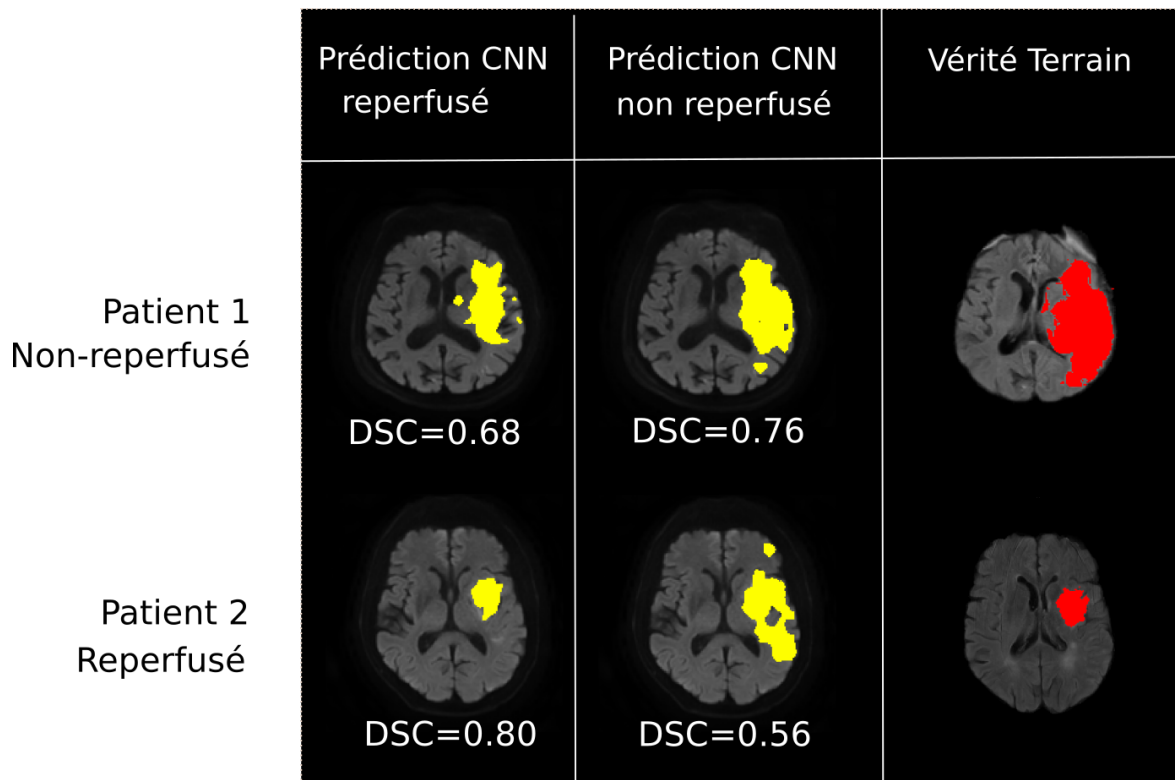


FIGURE III.8.3 – Prédications de l'infarctus final de deux patients, obtenues à partir des CNN reperfusé et non reperfusé : le patient 1 ne présente pas de reperfusion (TICI=2a) ; le patient 2 est reperfusé (TICI=2b).

III.8.4.3 Valeur des données d'entrée IRM pour prédire l'infarctus final

Une étude d'ablation a été réalisée avec les modèles reperfusés et non reperfusés (testés uniquement chez les patients reperfusés et non reperfusés, respectivement) afin d'évaluer la contribution relative des images d'entrée pour la prédiction de l'infarctus final. Chez les patients reperfusés et non reperfusés, les modèles CNN complets (c'est-à-dire incluant les images DWI, ADC, T_{max} , CBF et CBV) présentaient des performances similaires par rapport aux modèles sans CBF ni CBV, ce qui suggère que ces deux derniers biomarqueurs ont une valeur prédictive limitée (lignes 1 et 2 des tableaux III.8.5-a et III.8.5-b). Inversement, l'ajout des données de diffusion (DWI et ADC) à la carte T_{max} a considérablement augmenté le DSC moyen des modèles CNN. Cette augmentation de performance est plus prononcée chez les patients reperfusés que chez les patients non reperfusés.

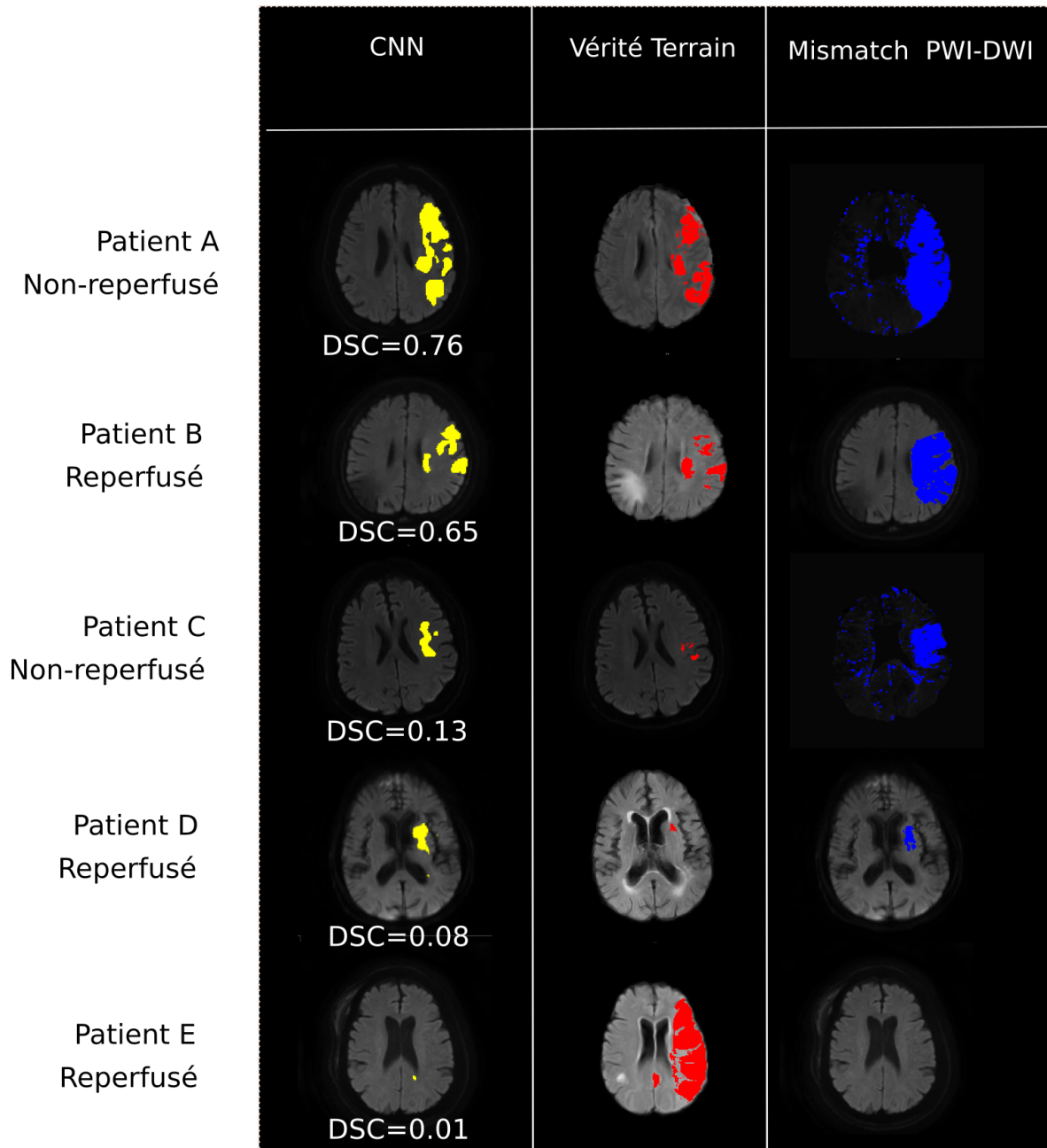


FIGURE III.8.4 – Prédications de l'infarctus final, obtenues à partir des modèles CNN et des modèles mismatch perfusion-diffusion. Cinq patients testés sont présentés : deux cas de réussite où le modèle CNN réalise de meilleures performances que le modèle mismatch correspondant (patient A avec TICI=2a et patient B avec TICI=3) et trois patients difficiles à prédire, tant pour le modèle CNN que pour le modèle mismatch (patient C avec TICI=2a, patient D avec TICI=3 et patient E avec TICI=2b).

(a) Modèle reperfusé : étude d'ablation parmi les patients reperfusés

IRM d'entrée	DSC	VS	Précision	Rappel	HD
DWI+ADC+T _{max} +CBF+CBV	0.44 ± 0.21	0.66 ± 0.26	0.39 ± 0.25	0.63 ± 0.21	30.61 ± 16.1
DWI+ADC+T _{max}	0.44 ± 0.25	0.70 ± 0.26	0.54 ± 0.28 (***)	0.46 ± 0.27 (**)	35.13 ± 15.6 (.)
DWI	0.42 ± 0.24 (*)	0.70 ± 0.26	0.51 ± 0.28	0.44 ± 0.27 (***)	31.28 ± 16.1 (**)
ADC	0.40 ± 0.24 (***)	0.67 ± 0.28 (.)	0.47 ± 0.27 (.)	0.43 ± 0.27 (***)	34.35 ± 20.4 (*)
T _{max}	0.32 ± 0.20 (***)	0.63 ± 0.30 (*)	0.44 ± 0.25 (*)	0.35 ± 0.25 (***)	29.99 ± 13.7 (**)

(b) Modèle non reperfusé : étude d'ablation parmi les patients non reperfusés

IRM d'entrée	DSC	VS	Précision	Rappel	HD
DWI+ADC+T _{max} +CBF+CBV	0.47 ± 0.17	0.74 ± 0.13	0.49 ± 0.22	0.52 ± 0.21	37.70 ± 17.7
DWI+ADC+T _{max}	0.47 ± 0.18	0.74 ± 0.16	0.52 ± 0.22	0.50 ± 0.22	35.77 ± 20.2
DWI	0.45 ± 0.17	0.71 ± 0.17	0.50 ± 0.22	0.50 ± 0.25	33.20 ± 17.2
ADC	0.42 ± 0.15 (.)	0.73 ± 0.23	0.47 ± 0.18	0.46 ± 0.21 (.)	28.35 ± 12.9 (**)
T _{max}	0.40 ± 0.19 (*)	0.65 ± 0.21	0.50 ± 0.29	0.46 ± 0.24 (**)	26.86 ± 13.3 (.)

TABLEAU III.8.5 – Métriques d'évaluation des modèles reperfusés et non reperfusés après ablation successive des entrées IRM, testés parmi les patients (a) reperfusés et (b) non reperfusés, respectivement (valeurs moyennes ± écart type). Les valeurs en gras correspondent à la meilleure valeur de métrique en colonne. Un test bilatéral de Wilcoxon a été effectué entre les modèles complets (5 IRM en entrées) et les modèles après ablation, avec (.) indiquant P < 0.10, (*) indiquant P < 0.05, (**) indiquant P < 0.01 et (***) indiquant P < 0.001.

III.8.5 Discussion

III.8.5.1 Impact du statut de reperfusion sur les performances des CNN

Les imageries aiguës ont une valeur prédictive importante pour prédire la lésion finale. Des études récentes ont déjà montré que les CNN ont le potentiel de prédire avec succès la lésion finale sans aucune information sur le statut de reperfusion des patients [85, 15]. Notre étude montre que les performances des CNN peuvent s'améliorer en intégrant dans l'ensemble d'entraînement le statut de reperfusion précoce. Cependant, il est important de noter qu'aussi bien dans notre étude que les autres études de la littérature [85, 15, 20], les patients sélectionnés sont majoritairement traités et bénéficient d'une reperfusion précoce (dans notre étude, 70% des 109 patients d'étude étaient reperfusés). La disproportion entre patients reperfusés et non reperfusés explique probablement l'absence de différence significative entre le modèle général et le modèle reperfusé.

Néanmoins, la motivation physiopathologique de l'intégration du statut de reperfusion dans les modèles prédictifs est importante. Une reperfusion rapide est étroitement associée à une récupération accrue de la pénombre ischémique et à une réduction de la taille de l'infarctus final [86]. De plus, un double ensemble de cartes prédictives en sortie des CNN est cliniquement pertinent, dans la mesure où cela permettrait aux médecins de comparer l'infarctus prédit avec et sans reperfusion, et donc d'estimer l'effet du traitement. Avec nos données, l'infarctus final prédit était sensiblement plus important avec le CNN non reperfusé chez 53 patients (soit 50% des patients de notre étude) si l'on considère les critères suivants : DSC entre les deux CNN < 0.5 et le volume lésionnel du CNN non reperfusé est 20% plus élevé que le volume lésionnel du CNN reperfusé. À l'inverse, l'absence de différence nette entre les deux modèles laisserait supposer un bénéfice limité des thérapies de reperfusion. Des prédictions fiables de l'infarctus final peuvent également aider à évaluer de nouvelles stratégies de neuroprotection, en comparant la taille projetée de l'infarctus observé chez les patients atteints d'ischémie-reperfusion [98]. Cette approche pourrait faciliter le dépistage d'un plus grand nombre de neuroprotecteurs présumés à un coût moindre que celui des essais cliniques actuels.

Nos résultats indiquent que la reperfusion est une caractéristique d'intervention clinique que le CNN peut intégrer avec succès aux données d'imagerie aiguë pour obtenir des prédictions plus robustes. La validation complète de cette approche nécessiterait une collaboration multicentrique afin de collecter des données longitudinales de haute qualité, comprenant en particulier davantage de patients sans reperfusion.

III.8.5.2 Comparaison avec les modèles cliniques actuels de prédiction

Nos modèles CNN présentent des métriques DSC et AUC plus élevées que les modèles mismatch perfusion-diffusion actuellement utilisés dans la pratique clinique (comme en témoigne les patients A et B de la Figure III.8.4). Nos résultats sont du même ordre que ceux des études récemment publiés utilisant des approches d'apprentissage profond : le meilleur modèle du challenge ISLES a obtenu un DSC moyen de 0.38 [17] ; Nielsen et al présentait une AUC moyenne de 0.88 [15], tandis que Yu et al a proposé un modèle avec un DSC et une AUC moyens de 0.53 et 0.89, respectivement [85].

Nous avons montré que la prédiction de l'infarctus final reste une tâche difficile. Les DSC moyens restent modestes (0.44 et 0.47 pour le modèle avec et sans reperfusion, respectivement), et correspondent à un ensemble hétérogène de prédictions précises ($DSC > 0.7$) et de prédictions échouées aussi bien par nos modèles CNN proposés que les modèles mismatch perfusion-diffusion (les patients C, D et E sont des exemples d'échec sur la Figure III.8.4). Une réversion partielle et parfois importante de la lésion de diffusion peut être observée (patients C et D de la Figure III.8.4), notamment en cas de reperfusion précoce [99]. De plus, l'imagerie aiguë ne peut pas prendre en compte des événements ultérieurs qui peuvent modifier la progression des lésions ischémiques (par exemple, le patient E de la Figure III.8.4 est un cas possible de ré-occlusion après une reperfusion réussie). Ces patients illustrent l'hétérogénéité et la complexité de la progression des lésions

ischémiques. L'apprentissage par renforcement pourrait contribuer à améliorer les performances du CNN, en s'entraînant plus spécifiquement sur ces cas de patients difficiles et sous-représentés [100].

III.8.5.3 Valeur prédictive des données IRM et impact de leur configuration de fusion

L'étude d'ablation a montré que le CBF et le CBV avaient un impact limité sur les performances de nos CNN. Ce résultat est conforme à l'observation qualitative courante selon laquelle la lésion de perfusion est moins visible sur les cartes CBF ou CBV que sur les cartes T_{max} . Une précédente étude basée sur des seuils avait également observé que ces deux biomarqueurs étaient de mauvais prédicteurs de l'infarctus final [101].

Ainsi, les images ADC, DWI et T_{max} pourraient constituer les principales entrées du réseau de prédiction de l'infarctus final. Nos résultats suggèrent en outre que les cartes T_{max} pourraient avoir une plus grande valeur prédictive chez les patients non reperfusés, ce qui serait cohérent avec des études précédentes : Wheeler et al [102] avait montré une forte corrélation entre la lésion de diffusion aiguë et le volume final de l'infarctus chez les patients reperfusés, et une corrélation élevée entre la lésion $T_{max} > 6$ secondes et le volume final de l'infarctus chez les patients non reperfusés.

Ces observations appuient notre choix d'architecture d'apprentissage profond. La configuration de fusion tardive de caractéristique permet une meilleure intégration des informations distinctes contenues dans l'image de perfusion et de diffusion. L'entraînement de modèles spécifiques au statut de la reperfusion implique l'attribution de poids distincts à chaque entrée IRM.

Nous avons comparé notre proposition d'architecture d'apprentissage profond par fusion tardive à une architecture de fusion précoce, où toutes les images sont combinées à l'entrée du CNN, avant toute étape convolutive. Cette stratégie de fusion réduit à la fois la complexité de calcul et le nombre d'hyperparamètres à ajuster pendant l'entraînement [103]. Chaque patient étant représenté par DWI, ADC, T_{max} , CBV, CBF, l'architecture de fusion précoce concatène les 5 IRM sur 5 canaux, et ne les traite pas indépendamment. Les résultats sont présentés dans le Tableau III.8.6.

La fusion précoce montre des performances globalement moins bonnes que la fusion tardive : les valeurs moyennes de DSC, VS, précision et rappel sont plus élevées quel que soit l'ensemble d'entraînement. Cependant, les valeurs les plus basses pour la métrique HD sont obtenues lors de la fusion précoce. La fusion précoce semble offrir une meilleure délimitation spatiale de la lésion finale : moins de valeurs aberrantes semblent être prédites, ce qui diminue considérablement les valeurs de HD. Nous avons également observé moins de différences de performance entre les modèles généraux, reperfusés et non reperfusés, ce qui suggère que la fusion précoce semble moins adapté pour prendre en compte l'état de la reperfusion.

III.8.5.4 Limites de l'étude

Notre étude présente plusieurs limites. Les patients inclus proviennent de deux cohortes ayant des protocoles de traitement différents : HIBISCUS-STROKE concernait des patients traités par thrombectomie, alors que I-KNOW était une étude observationnelle multicentrique de patients traités par thrombolyse intraveineuse sans aucune procédure endovasculaire (voire pour certains patients sans aucun traitement). Cependant, seuls les patients I-KNOW ayant une occlusion proximale sans reperfusion ont contribué à l'étude : ils peuvent donc être considérés comme des cas d'échec de thrombectomie. Bien que les méthodes d'évaluation de la reperfusion précoce diffèrent, le score mTICI et le DSC-PWI évaluent tous deux la perfusion tissulaire. D'autres paramètres cliniques tels que l'âge et le délai entre l'apparition des symptômes, l'imagerie et la reperfusion sont des facteurs pronostiques reconnus. Leur intégration dans le CNN pourrait améliorer les performances du modèle et faire l'objet d'une étude plus approfondie.

(a) Évaluation chez les patients reperfusés

Fusion	Modèle	DSC	VS	Précision	Rappel	HD
Précoce	général	0.39 ± 0.25	0.59 ± 0.30	0.56 ± 0.31	0.40 ± 0.26	29.51 ± 16.26
Précoce	reperfusé	0.41 ± 0.25	0.64 ± 0.30	0.46 ± 0.29 (***)	0.49 ± 0.30 (***)	31.24 ± 15.61
Précoce	non-reperfusé	0.36 ± 0.22 (*)	0.63 ± 0.27	0.54 ± 0.26	0.33 ± 0.24 (***)	26.64 ± 11.16
Tardive	général	0.43 ± 0.24	0.69 ± 0.27	0.55 ± 0.28	0.43 ± 0.25	33.23 ± 15.64
Tardive	reperfusé	0.44 ± 0.25	0.70 ± 0.27	0.50 ± 0.27	0.50 ± 0.26 (***)	38.58 ± 18.15
Tardive	non-reperfusé	0.35 ± 0.21 (***)	0.57 ± 0.28 (***)	0.60 ± 0.25 (***)	0.31 ± 0.24 (***)	40.05 ± 15.66 (**)

(b) Évaluation chez les patients non reperfusés

Fusion	Modèle	DSC	VS	Précision	Rappel	HD
Précoce	général	0.42 ± 0.24	0.62 ± 0.27	0.42 ± 0.28	0.55 ± 0.29	30.98 ± 18.23
Précoce	reperfusé	0.41 ± 0.26	0.51 ± 0.31	0.36 ± 0.29	0.69 ± 0.24	30.94 ± 16.30
Précoce	non-reperfusé	0.42 ± 0.18	0.66 ± 0.17	0.42 ± 0.24	0.55 ± 0.22	28.48 ± 13.63
Tardive	général	0.44 ± 0.21	0.66 ± 0.26	0.39 ± 0.25	0.63 ± 0.21	30.61 ± 16.15
Tardive	reperfusé	0.44 ± 0.22	0.63 ± 0.25	0.36 ± 0.23	0.69 ± 0.22 (*)	44.53 ± 16.79 (**)
Tardive	non-reperfusé	0.47 ± 0.17	0.74 ± 0.13	0.49 ± 0.22 (**)	0.52 ± 0.21 (***)	37.70 ± 17.74

TABLEAU III.8.6 – Métriques d'évaluation des modèles généraux, reperfusés et non reperfusés avec différentes stratégies de fusion (précoce et tardive) sur des patients **reperfusés** (a) et des patients **non reperfusés** (b) (valeurs moyennes ± écart type). Les valeurs en gras correspondent à la meilleure valeur de la métrique en colonne. Un test bilatéral de Wilcoxon a été effectué entre le modèle global et les deux autres modèles (reperfusé et non-reperfusé) pour une stratégie de fusion donnée, avec (*) indiquant $P < 0.05$, (**) indiquant $P < 0.01$ et (***) indiquant $P < 0.001$.

Chapitre III.9

Classification du statut de réversion des voxels de la lésion de diffusion aiguë par forêt aléatoire

Dans ce chapitre, nous nous focalisons sur les voxels inclus dans la lésion de diffusion aiguë, afin de prédire leur statut de réversion. Nous proposons un modèle par forêt aléatoire, relativement parcimonieux et rapide d'exécution, et nous proposons d'interpréter précisément l'impact de chaque modalité d'entrée sur la prédiction. Nous discutons de la pertinence de nos résultats au regard de l'état de l'art ainsi que des besoins cliniques actuels.

III.9.1 Facteurs cliniques associés à la réversion, à l'échelle du patient

III.9.1.1 Matériel

Pour cette étude, 91 patients ont été inclus de l'étude HIBISCUS-STROKE comme le montre la Figure III.9.1. HIBISCUS-STROKE est une cohorte d'observation monocentrique en cours qui regroupe des patients présentant une occlusion artérielle intracrânienne importante et traités par thrombectomie, après des IRM de diffusion-perfusion aiguës. Tous les patients ont donné leur consentement éclairé et le protocole d'imagerie a été approuvé par le comité d'éthique régional. À leur admission, tous les patients ont été soumis au protocole IRM suivant : imagerie pondérée par diffusion (DWI), FLAIR (fluid-attenuated-inversion-recovery) pondéré en T2, écho de gradient T2, angiographie par RM et imagerie dynamique de perfusion à contraste de susceptibilité (DSC-PWI). Un FLAIR de suivi a été effectué six jours après l'admission.

Les lésions de DWI (en aiguë) et de FLAIR (en suivi) ont été segmentées par un expert à l'aide d'une méthode semi-manuelle (3D Slicer, <https://www.slicer.org/>). Des cartes paramétriques ont été extraites de la perfusion DSC-PWI par décomposition circulaire en valeurs singulières des courbes de concentration tissulaire (Olea Sphere, Olea Medical, La Ciotat, France) : débit sanguin cérébral (CBF), volume sanguin cérébral (CBV), temps de transit moyen (MTT), temps au maximum (T_{max}) et temps au pic (TTP). Les images ont été recalées dans l'espace de la DWI aiguë par recalage non linéaire Ants [87]. Les IRM aiguës ont été normalisées : (1) en soustrayant la valeur moyenne de l'hémisphère contralatéral à chaque voxel de l'hémisphère ipsilatéral pour le DWI, l'ADC, et le T_{max} ; (2) en divisant la valeur moyenne de l'hémisphère contralatéral à chaque voxel de l'hémisphère ipsilatéral pour le CBV et le CBF.

III.9.1.2 Définition d'un patient réverseur

Il n'existe pas de consensus clair dans la littérature pour définir un patient réverseur. La plupart des travaux en rapport avec la réversion se contentent de définir un patient comme réverseur lorsque le volume de lésion de DWI a régressé entre l'admission et le suivi [104]. Selon ce critère,

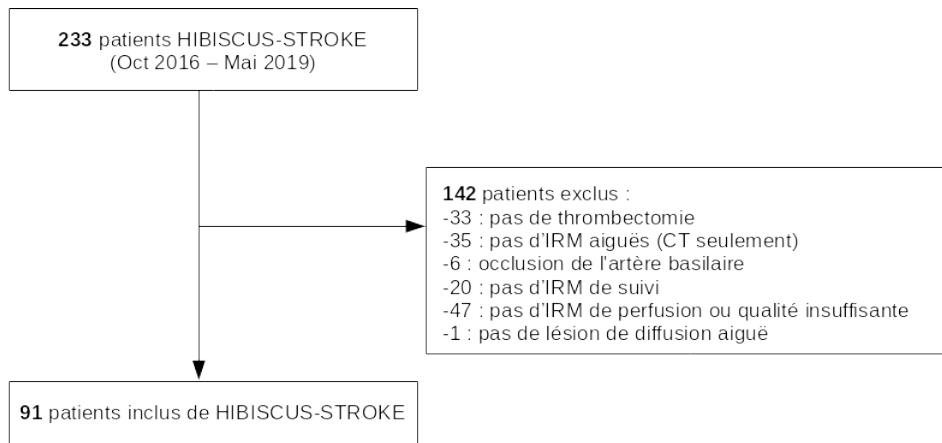


FIGURE III.9.1 – Organigramme d’inclusion des patients.

environ 1 patient sur 4 semble présenter une réversion de lésion de DWI, bien que la réversion complète soit très rare [105, 106, 107, 108, 109]. Seulement cette définition est biaisée, car elle implique qu’un patient peut être défini réverseur bien que son volume de lésion de DWI ait diminué de moins de 5mL (ce qui est très peu). Dans notre étude, nous avons choisi des seuils plus stricts : un patient a été défini comme réverseur si : (1) le volume de la lésion de DWI de départ avait diminué d’au moins 10 ml; et (2) le volume de la lésion de DWI avait diminué d’au moins 25%. Selon ces deux critères combinés, les 91 patients de l’étude ont pu être divisés en deux groupes : 29 patients réverseurs contre 62 patients non réverseurs, comme le montre la Figure III.9.2.

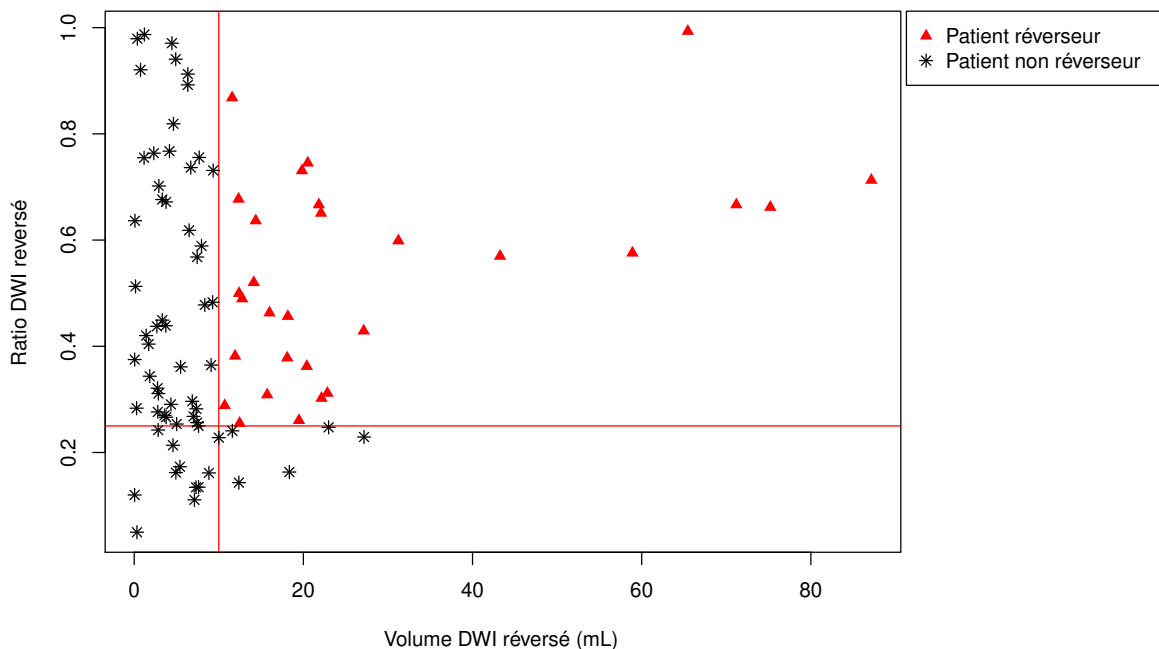


FIGURE III.9.2 – Patients d’étude en fonction du volume de lésion DWI aiguë réversé (axe x) et du ratio de volume de lésion DWI réversé (axe y). Les patients réverseurs (marqueurs triangulaires rouges) ont réversé d’au moins 10 ml et 25% de leur volume de lésion DWI aiguë.

III.9.1.3 Analyses statistiques

Afin d'évaluer les facteurs cliniques associés à la réversion, nous avons effectué plusieurs analyses statistiques descriptives. Dans un premier temps, nous avons évalué les différences entre les patients réverseurs et les patients non réverseurs à l'aide d'un test de Mann-Whitney non apparié pour les variables continues et d'un test de Fisher pour les variables catégorielles. Les variables continues évaluées étaient les suivantes : le volume de lésion DWI perdu, le ratio de lésion DWI perdu, l'âge, le NIHSS, le délai entre le début des symptômes et l'IRM (noté d_1), le délai entre le début des symptômes et la reperfusion (noté d_2). Les variables catégorielles évaluées étaient les suivantes : le sexe (femme : oui ou non), le traitement par thrombolyse (oui ou non) et le TICI (binarisé, en opposant : $TICI=\{0,1,2a\}$ vs $TICI=\{2b,2c,3\}$).

Dans un second temps, nous avons effectué une régression logistique univariée afin d'estimer la force de l'association entre chaque variable clinique et la variable explicative binaire (le fait qu'un patient soit réverseur ou non). Comme les variables d_1 et d_2 n'étaient disponibles que pour 52 patients, nous avons effectué deux régressions logistiques distinctes : (1) une qui considérait les 91 patients d'étude sans les variables d_1 et d_2 (âge, sexe, traitement par thrombolyse, NIHSS, TICI); (2) une qui ne considérait seulement que 52 patients pour qui les variables d_1 et d_2 étaient disponibles (âge, sexe, traitement par thrombolyse, NIHSS, TICI, d_1 , d_2); les deux régressions ont été effectuées avec une P-value significative seuil de 0.05.

III.9.2 Marqueurs d'images associés à la réversion, à l'échelle du voxel

III.9.2.1 Forêt aléatoire pour la prédiction du statut de réversion

Une classification binaire à l'échelle du voxel a été réalisée par un modèle de forêt aléatoire supervisé. Comme le montre la Figure III.9.3, nous avons travaillé uniquement sur les voxels inclus dans la lésion DWI aiguë. Sur la base des modalités d'imagerie aiguës, chaque voxel a été classé comme réversible ou non réversible. Les forêts aléatoires sont des classifieurs performants qui construisent un grand ensemble d'arbres de décision indépendants, où chaque arbre est construit sur un échantillon bootstrapé des voxels d'entraînement, et combine la sortie de tous les arbres avec un processus de vote. Ils peuvent ainsi obtenir une meilleure généralisation que les arbres de décision seuls. Dans ce travail, nous avons paramétré la forêt aléatoire avec 100 arbres d'une profondeur maximale de 5.

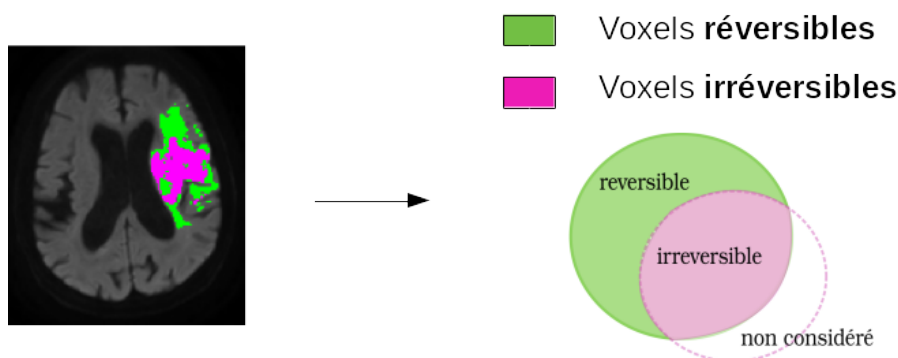


FIGURE III.9.3 – Les voxels d'étude inclus dans la lésion DWI aiguë sont classés réversibles ou non réversibles.

III.9.2.2 Validation croisée et évaluation du modèle

La validation croisée permet d'estimer la fiabilité d'un modèle en se basant sur une technique d'échantillonnage. Nous avons effectué une validation croisée 5-Folds, où le modèle a été entraîné sur les voxels de 80% des patients choisis au hasard dans la cohorte, et validé sur les voxels des

20% de patients restants. Ce processus est ensuite répété 5 fois, chacun des 5 sous-échantillons de voxels étant utilisé exactement une fois comme validation. La moyenne des résultats de ces 5 validations a ensuite été calculée pour produire une seule estimation.

La performance du modèle a été évaluée en utilisant l'aire sous la courbe ROC (AUC), la sensibilité, la spécificité et le coefficient de similarité de Dice (DSC). La sensibilité a été définie comme la proportion de voxels réels réversibles correctement identifiés comme tels, et la spécificité comme la proportion de voxels réels non réversibles correctement identifiés comme tels. Le DSC a été donné pour chaque classe, à savoir les voxels non réversibles et réversibles.

III.9.2.3 Contribution des modalités d'entrée dans la tâche de classification

Les modèles de forêt aléatoire permettent d'évaluer la contribution de chaque modalité d'entrée dans la tâche de classification. L'importance de la modalité est calculée comme la diminution de l'impureté d'un nœud, pondérée par la probabilité d'atteindre ce nœud. La probabilité de nœud peut être calculée par le nombre d'arbres qui atteignent le nœud, divisé par le nombre total d'arbres. La contribution est donnée en pourcentage, et la somme des contributions des cinq modalités d'entrée est de 100%. Plus la valeur est élevée, plus la modalité est importante.

Pour aller plus loin dans l'explicabilité du modèle, nous proposons des mesures supplémentaires basées sur l'indice de Shapley de la théorie des jeux [110]. Nous proposons un graphique récapitulatif qui combine l'importance et l'effet de chaque caractéristique (ou modalité) sur la prise de décision du modèle. Chaque point du graphique correspond à la valeur d'indice de Shapley d'un voxel d'entraînement pour une caractéristique donnée. La position sur l'axe y est déterminée par les caractéristiques et sur l'axe des x par la valeur de Shapley. La couleur représente la valeur d'une caractéristique donnée : les valeurs faibles sont dans les tons bleus tandis que les valeurs élevées sont dans les tons rouges. Lorsque plusieurs points d'une caractéristique donnée présentent la même valeur de Shapley, ils s'empilent les uns sur les autres sur le même axe x.

III.9.3 Résultats

III.9.3.1 Facteurs cliniques

Sur les 91 patients d'étude, 76 présentaient une occlusion M1. L'âge médian était de 69.2 ans (IQR 52-80) et le NIHSS médian était de 15 ans (IQR 11-19). Le délai entre le début des symptômes et l'IRM était disponible pour 61 patients et la valeur médiane était de 100 (IQR 75-120) minutes. Le délai entre le début des symptômes et la reperfusion était disponible pour 52 patients et la valeur médiane était de 212 (IQR 179-262) minutes. Le volume médian de la lésion DWI était de 21,47 (IQR 7.03-48.01) mL. Le volume et le ratio de volume médians de lésion DWI perdu étaient respectivement de 7,43 (IQR 3.7-14.3) mL et de 0.438 (IQR 0.27-0.67). Parmi toutes les caractéristiques cliniques figurant dans le Tableau III.9.1, les patients réverseurs étaient statistiquement différents des patients non réverseurs en termes de volume de lésion DWI et de volume de lésion DWI perdu en raison de la définition même des patients réverseurs. D'autres variables telles que le NIHSS et le délai entre les symptômes et l'IRM (d_1) semblaient être sensiblement différents entre les deux classes de patients (P -value <0.1).

D'après les régressions logistiques univariées réalisées, seul le NIHSS était associé de manière significative aux patients réverseurs. L'Odds ratio de cette variable étant positif, il semble que les patients réverseurs présentent des valeurs élevées de NIHSS (Tableau III.9.2). Cette observation est finalement triviale puisque les valeurs élevées de NIHSS sont attribuées aux patients présentant de grandes lésions.

Variables cliniques	Patients réverseurs (n=29)	Patients non réverseurs (n=62)	Test P-Value
Volume DWI, mL	47.8 (27.5 - 52.5)	10.3 (5.2 - 25.7)	7.9 e-8
Volume DWI perdu, mL	19.8 (14.1 - 27.1)	4.8 (2.8 - 7.4)	1.0 e-12
Âge	66 (50 - 81)	70 (56 - 79)	0.82
Femme, n (pourcentage)	12 (41.4)	28 (45.2)	0.74
Traitement par TIV, n (pourcentage)	13 (44.8)	34 (54.8)	1
NIHSS	16 (13 - 21)	15 (8 - 18)	0.087
TICI={2b, 2c, 3}, n (pourcentage)	26 (89.7)	47 (75.8)	0.16
d ₁ , min	80 (74 - 99.5) (n= 18/29)	107 (78.5 - 187.5) (n=43/62)	0.056
d ₂ , min	199 (165 - 235) (n=17/29)	217 (183 - 284) (n=35/62)	0.24

TABLEAU III.9.1 – Caractéristiques cliniques (médiane avec écart inter-quartile, sauf indication contraire) pour les patients réverseurs et non réverseurs. NIHSS : National Institutes of Health Stroke Scale; DWI : diffusion-weighted imaging; TIV : thrombolyse intra-veineuse; d₁ : délai entre les symptômes et l'IRM; d₂ : délai entre les symptômes et la reperfusion.

Variables cliniques	Odds Ratio (95% IC)	P-Value
Âge	0.99 (0.97 - 1.02)	0.646
Genre	1.17 (0.48 - 2.89)	0.735
Traitement par TIV	0.67 (0.27 - 1.62)	0.374
NIHSS	1.09 (1.00 - 1.19)	0.0375
TICI	2.77 (0.82 - 12.73)	0.134
d ₁	0.99 (0.98 - 1.00)	0.0691
d ₂	0.99 (0.98 - 1.00)	0.123

TABLEAU III.9.2 – Régression logistique univariée. NIHSS : National Institutes of Health Stroke Scale; TIV : thrombolyse intra-veineuse; d₁ : délai entre les symptômes et l'IRM; d₂ : délai entre les symptômes et la reperfusion. IC : Intervalle de Confiance.

III.9.3.2 Marqueurs images

Performances à l'échelle du voxel de la forêt aléatoire

Les performances de classification du modèle de forêt aléatoire sont données dans le Tableau III.9.3. L'AUC apparaît toujours supérieure à 0.7, ce qui indique une bonne classification générale, quel que soit le sous-échantillon de voxels d'entraînement. Le modèle présente une bonne spécificité : quels que soient les voxels d'entraînement, les voxels non réversibles sont majoritairement bien prédits. Cependant, la sensibilité est beaucoup plus variable : selon les voxels d'entraînement, le modèle n'est pas toujours capable de prédire correctement les voxels réversibles. Ces observations sont confirmées par les valeurs DSC : le DSC des voxels non réversibles est nettement plus élevé que le DSC des voxels réversibles.

La Figure III.9.4 montre que les patients présentant des lésions étendues (volume > 50 mL) sont mieux prédits que les patients présentant des lésions moins volumineuses. La différence entre les deux groupes en termes d'AUC à l'échelle du patient est significative (0.64 ± 0.08 pour les lésions étendues contre 0.60 ± 0.09 pour les lésions plus petites, P<0.05).

Les prédictions du modèle sont présentées en Figure III.9.5 pour 5 patients de validation. Nous avons montré des cas où le modèle parvenait bien à prédire la réversion de lésion de DWI : c'est le cas des patients 1, 2, et 3, qui présentent tous des valeurs de métrique d'évaluation élevées. Parmi les cas difficiles à prédire, les patients peuvent présenter des lésions de DWI particulièrement peu volumineuses : c'est le cas du patient 5. Pour d'autres patients, notre modèle tend à sur-estimer les voxels non réversibles : c'est le cas du patient 4.

	Mean \pm ET	Min - Max
AUC	0.73 \pm 0.021	0.70 - 0.75
Spécificité	0.77 \pm 0.12	0.67 - 0.91
Sensibilité	0.52 \pm 0.21	0.26 - 0.71
DSC non réversible	0.71 \pm 0.025	0.68 - 0.74
DSC réversible	0.53 \pm 0.078	0.45 - 0.62

TABLEAU III.9.3 – Performances du modèle proposé pour la prédiction de la réversion de la lésion DWI. Les résultats ont été obtenus par une approche de validation croisée 5-Fold. L'AUC, la spécificité, la sensibilité et les DSC donnés correspondent aux moyennes sur les 5 validations (moyenne \pm écart-type). Les valeurs minimales (Min) et maximales (Max) sur les 5 validations ont également été données.

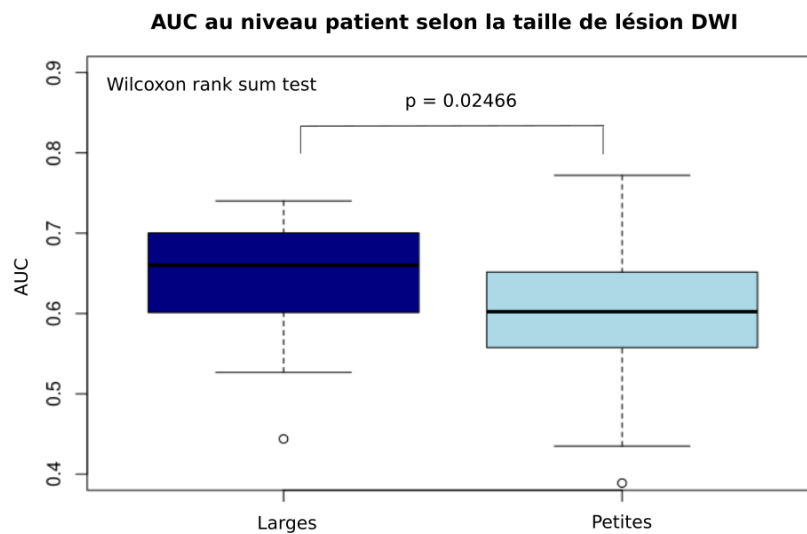


FIGURE III.9.4 – Boxplots des performances du modèle proposé en fonction de la taille des lésions de DWI (lésions peu volumineuses : volume < 50mL, 70 patients; lésion volumineuses : volume > 50mL, 21 patients). Les différences entre les deux groupes ont été évaluées à l'aide d'un test de Wilcoxon non apparié.

Contribution des modalités d'imagerie d'entrée

D'après le Tableau III.9.4, le DWI est l'image d'entrée qui contribue le plus à notre tâche de prédiction. Le T_{max} et l'ADC semblent également être impliqués mais à moindre mesure. En revanche, les images CBF et CBV semblent être peu pertinentes dans les prises de décision de la forêt.

La Figure III.9.6 montre, par le biais des valeurs d'indice Shapley (SHAP), dans quelle mesure chaque modalité d'imagerie contribue, positivement ou négativement, à la classification. Le DWI présente une grande amplitude de valeurs SHAP négatives et positives, ce qui signifie qu'il s'agit de la caractéristique la plus discriminante de notre modèle. Plus précisément, des valeurs de DWI élevées contribuent fortement à prédire qu'un voxel n'est pas réversible. Dans une moindre mesure, des valeurs T_{max} élevées, des valeurs ADC faibles et des valeurs CBV faibles contribuent aussi à prédire qu'un voxel est non réversible. D'autre part, les faibles valeurs de DWI conduisent le modèle à prédire la classe réversible. Dans une moindre mesure, des valeurs T_{max} faibles, des valeurs CBV élevées et des valeurs ADC élevées contribuent à prédire qu'un voxel est réversible. Notons que les valeurs SHAP du CBF sont proches de 0 : le CBF semble très peu contribuer aux décisions prises par le modèle prédictif. Nous pouvons également noter que les valeurs SHAP de la classe réversible (à gauche de l'axe vertical gris) ont une magnitude inférieure aux valeurs SHAP de la classe non réversible. Cela révèle que les valeurs des caractéristiques de la classe réversible sont moins

	Prédiction	Vérité Terrain
Patient 1 AUC = 0.77 DSC_nr = 0.86 DSC_r = 0.53		
Patient 2 AUC = 0.74 DSC_nr = 0.66 DSC_r = 0.74		
Patient 3 AUC = 0.75 DSC_nr = 0.81 DSC_r = 0.50		
Patient 4 AUC = 0.45 DSC_nr = 0.34 DSC_r = 0.45		
Patient 5 AUC = 0.13 DSC_nr = 0.00 DSC_r = 0.40		

FIGURE III.9.5 – Prédications données par le modèle proposé, comparée avec la vérité terrain. Les tons bleus désignent les voxels réversibles, tandis que les tons rouges désignent les voxels non réversibles. Cinq patients de validation sont montrés : trois patients bien prédits - avec une AUC > 0.7 (patient 1 avec TICI=3, patient 2 avec TICI=3 et patient 3 avec TICI=2b), et deux patients difficiles à prédire - avec une AUC < 0.5 (patient 4 avec TICI=2b et patient 5 avec TICI=3). DSC_r and DSC_{nr} désignent les DSC des voxels réversibles et non réversibles, respectivement.

discriminantes que celles de la classe non réversible. Cela explique pourquoi les performances de la classe réversible sont moins bonnes que celle de la classe non réversible.

	Moyenne \pm ET (in %)	Min - Max (in %)
DWI	68.1 \pm 2.2	65.6 - 70.3
T_{max}	15.0 \pm 3.4	10.6 - 20.2
ADC	8.8 \pm 3.4	7.1 - 12.2
CBV	4.3 \pm 2.1	2.1 - 7.3
CBF	3.8 \pm 1.4	1.8 - 5.8

TABLEAU III.9.4 – Contribution des modalités d’entrée (en %) à la tâche de prédiction (moyenne \pm écart-type).

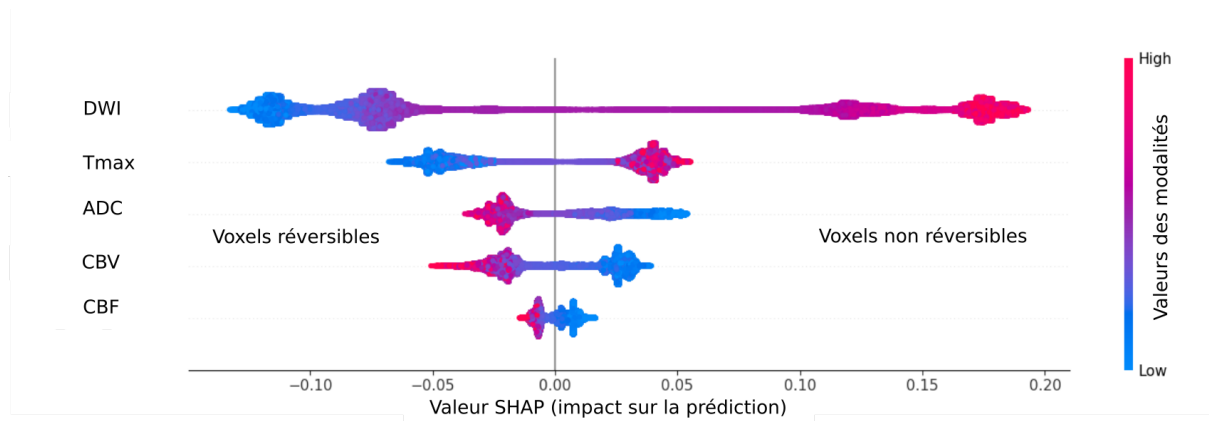


FIGURE III.9.6 – Graphique récapitulatif mesurant à quel point les caractéristiques d’entrée (DWI, T_{max} , ADC, CBV, CBF) contribuent (positivement ou négativement) à discriminer les voxels d’entraînement. Chaque point correspond à un voxel d’entraînement. Les points à droite de l’axe vertical gris correspondent aux voxels non réversibles, tandis que les points à gauche correspondent aux voxels réversibles. Les points dans les tons chauds représentent des valeurs de voxels élevées pour une caractéristique donnée, et les couleurs plus froides indiquent des valeurs plus faibles. Pour une caractéristique donnée, plus les points le long de l’axe x sont nombreux et éloignés de l’axe vertical gris, plus cette caractéristique contribue à la classification.

III.9.4 Discussion

III.9.4.1 Confrontation des résultats avec l’état de l’art

Dans la littérature, les travaux en rapport avec la réversion de la lésion DWI sont rares, et se contentent uniquement d’analyses descriptives à l’échelle du patient. Les quelques études existantes semblent converger pour affirmer que la réversion de la lésion DWI est associée à : (1) la recanalisation et/ou la reperfusion précoce rapide [111, 112, 113, 106, 114, 108, 115], (2) à un traitement thérapeutique [108], et (3) à une amélioration neurologique précoce et une amélioration de l’état clinique du patient à long terme [116, 106, 107, 114, 108]. Notons que l’association (3) est probablement plutôt due à une reperfusion précoce plutôt que la réversion DWI en elle-même [104], puisque l’amélioration de l’état clinique est déjà une conséquence reconnue de la reperfusion précoce. Notons également que la plupart des études considèrent un patient non réverseur dès que sa lésion DWI aiguë a diminué en volume entre l’admission et le suivi, sans aucun seuil critique. Ces corrélations peuvent donc être biaisées. Dans notre étude, aucune de ces associations n’a été observée de manière significative. La différence entre les patients réverseurs et les patients non réverseurs en termes de reperfusion (TICI) était notable, mais pas significative (P-value = 0.16). En revanche dans notre étude, nous avons constaté que le délai entre les symptômes et l’IRM (variable d_1) était différent entre les deux groupes (patients réverseurs vs non réverseurs), à la limite de la significativité (P-value=0.056). Il convient de souligner que nous n’avons accès à l’information de délai d_1 seulement pour 52 patients, soit un échantillon relativement petit. L’association entre le délai d_1 et la réversion de lésion DWI reste à être confirmée sur une cohorte plus large.

III.9.4.2 Choix du modèle prédictif

Nous avons choisi d'utiliser un modèle de forêt aléatoire pour prédire la réversion de lésion DWI pour plusieurs raisons. Premièrement, parce qu'aucune étude n'avait jusqu'à présent proposé un modèle d'apprentissage supervisé sur cette question. Nous avons donc voulu proposer un premier modèle prédictif non linéaire avec un nombre limité de paramètres comme première référence. Bien que les modèles d'apprentissage profond CNN se sont avérés très efficaces dans de nombreuses applications [117], ils comportent de nombreux paramètres à ajuster et nécessitent beaucoup de temps de calcul. L'idée était de proposer un premier modèle afin d'obtenir des performances de référence. Pour des travaux futurs, il serait intéressant de voir dans quelle mesure l'apprentissage profond CNN pourrait améliorer les performances globales présentées dans cette étude. Deuxièmement, nous voulions être en mesure d'expliquer précisément les prédictions obtenues avec le modèle proposé, et éviter l'effet boîte noire. Grâce à des travaux récents sur l'explicabilité des modèles basés sur les arbres de décision [110], nous avons pu comprendre et mesurer comment le modèle utilise les caractéristiques d'entrée pour faire des prédictions.

III.9.4.3 Aide à la décision thérapeutique

Le modèle de forêt aléatoire à l'échelle du voxel a permis de mettre en évidence des marqueurs d'imagerie capable de discriminer les voxels de la lésion DWI en deux classes : réversible ou non réversible. Lorsque les voxels sont en hyper-intensité sur le DWI (comparativement aux voxels contralatéraux), le modèle semble indiquer de manière robuste qu'ils ont très peu de chance de réverser. Le DWI apparaît ainsi comme un marqueur important pour la prise de décision thérapeutique en urgence. Aussi, notons que le modèle proposé est significativement plus performant pour les lésions DWI aiguës volumineuses (volume > 50mL). Ce résultat est important à souligner, car la prise de décision thérapeutique pour les lésions volumineuses est toujours plus compliquée que pour les petites lésions. En effet, en pratique, les petites lésions sont systématiquement traitées, tandis que les lésions volumineuses sont sujettes à plus d'interrogations (par rapport au compromis bénéfice/risque d'une intervention). Le modèle que nous proposons promet d'être particulièrement robuste pour les lésions DWI volumineuses, et pourrait ainsi accompagner les neuroradiologues interventionnels dans leur prise de décision thérapeutique.

Partie IV

Ouverture, conclusions et perspectives

Chapitre IV.10

Potentiel des capteurs portables pour le suivi de l'accident vasculaire cérébral

Les capteurs portables, intégrés dans un système de suivi d'activité, ont le potentiel d'aider au diagnostic des risques et de suivre de manière continue les patients dans la phase chronique de l'AVC. Une nouvelle ère s'ouvre potentiellement pour le domaine du traitement du signal image appliqué à l'étude de l'AVC. Dans ce contexte, de nombreux défis demeurent à relever pour utiliser les capteurs portables en routine clinique. On peut citer le fait que les capteurs embarqués actuels ne sont pas forcément reconnus comme des dispositifs médicaux, que leur cadre d'utilisation clinique est complexe (besoin de constituer des cohortes de patients en dehors des hôpitaux), ou encore qu'ils n'assurent pas une cybersécurité optimale vis-à-vis des données qu'ils produisent. Dans ce chapitre nous abordons en guise d'ouverture de cette thèse, certains de ces nouveaux challenges. Nous résumons les récents développements dans le domaine des capteurs portables proposés pour la rééducation de maladies chroniques. Après avoir donné une brève description de ces technologies portables, nous exposons leur intégration dans le contexte de surveillance à domicile. Nous détaillons ensuite plusieurs applications cliniques possibles. Enfin, nous discutons de la potentielle menace de la vie privée que les capteurs actuels peuvent constituer, notamment concernant la collecte, l'exploitation et le transfert de données. Ces données qui contiennent des informations sensibles peuvent en effet être exploitées sans le consentement des patients.

IV.10.1 Description des technologies

Définition Les capteurs portables sont des appareils électroniques dits "intelligents", qui sont portés à proximité et/ou à la surface de la peau, où ils détectent, analysent et transmettent des informations concernant des signaux corporels (signal physiologique, signal d'activité) et/ou des signaux de données ambiantes. Ils permettent dans certains cas un retour (*biofeedback*) immédiat à l'utilisateur [118, 119, 120]. Ces capteurs échangent des données avec d'autres utilisateurs ou appareils connectés, sans intervention humaine et via des réseaux sans fils. Cette technologie s'intègre aujourd'hui dans les systèmes de surveillance à distance de personnes. Les capteurs portables se composent de trois éléments principaux : 1) le matériel de détection et de collecte de données physiologiques ou de mouvement, 2) le matériel et le logiciel de communication pour relayer les données vers un centre distant, et 3) les techniques d'analyse de données pour extraire des informations cliniquement pertinentes à partir des données physiologiques et de mouvement [121].

Matériel de détection La microélectronique a joué un rôle clé dans le développement des systèmes portables actuels. Pour que le capteur inertiel soit adapté aux applications de surveillance à long terme, il faut avant tout qu'il soit de taille suffisamment petite, portatif, léger et peu encombrant. Si les capteurs disposent de telles propriétés aujourd'hui, c'est grâce aux progrès récents dans le domaine de la microélectronique, qui ont permis de mettre au point des circuits

miniatures flexibles, intégrant des composants tels que des microprocesseurs [122]. L'essor des systèmes microélectromécaniques (MEMS) ont également permis des avancées technologiques considérables dans la fabrication des capteurs portables, en offrant des nouvelles capacités de détection, de traitement et de communication. Ils ont également permis de réduire les coûts des composants et de limiter la consommation énergétique des capteurs [123].

Collecte et traitement de données Les systèmes portables récemment développés intègrent des capteurs qui utilisent une technologie de communication sans fils, à faible coût énergétique, peu onéreuse et à haut débit de données. Parmi les technologies les plus courantes, on compte les liaisons ZigBee, Bluetooth, Wi-Fi et radio à impulsions à bande ultra-large (UWB) [124]. La plupart des applications de surveillance exigent que les données recueillies soient transmises à un site distant (le *cloud*, c'est-à-dire un serveur dont la gestion est interne ou externe au projet de recherche). Cela peut se faire en transmettant les données à une passerelle d'information telle qu'un téléphone portable ou un ordinateur personnel.

Les téléphones portables (les smartphones) constituent des passerelles d'information incontournables. En plus d'être très accessibles sur le marché [125], ils sont très prisés car ils fournissent une plateforme "prête à l'emploi" pour enregistrer et transmettre les données. Le stockage et les calculs étant réalisés sur le cloud, le dispositif est simple d'utilisation, accessible et peu coûteux [126, 127]. Une fois recueillies, les données de patient sont traitées pour en extraire des variables cliniques pertinentes via des techniques de traitement du signal, de reconnaissance des formes, de fouille de données et d'autres méthodes basées sur l'intelligence artificielle [128]. Cette étape de traitement peut désormais se faire sur le téléphone lui-même grâce aux nouveaux développements de puces tout en un (puces SoC) [129].

Architecture complète du système de surveillance Une représentation conceptuelle d'un système de surveillance à distance est représentée en Figure IV.10.1. Les capteurs portables récoltent des données physiologiques (A) et/ou de mouvement (B) permettant de surveiller l'état du patient. La communication sans fil est utilisée pour transmettre les données du patient à un téléphone mobile (C) ou à un point d'accès et relayer les informations vers un centre distant (D) via Internet. En cas d'urgence, (par exemple, chutes détectées) un message d'alarme peut être envoyé à un centre de service d'urgence (G) pour fournir une assistance immédiate aux patients. Les membres de la famille (F) et les soignants (E) sont alertés en cas d'urgence mais peuvent également communiquer régulièrement avec le patient tout en surveillant à distance son état clinique.

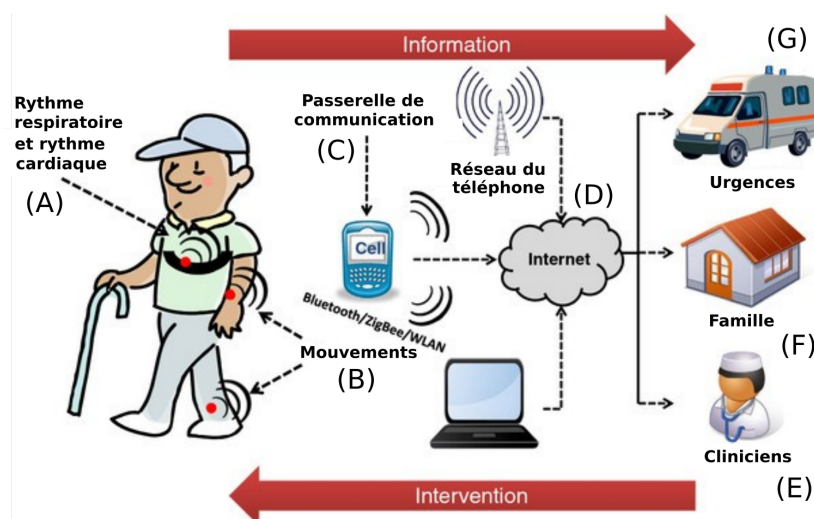


FIGURE IV.10.1 – Illustration d'un système de surveillance à distance de la santé basé sur des capteurs portables. Figure reproduite de "A review of wearable sensors and systems with application in rehabilitation." *Journal of neuroengineering and rehabilitation* 9.1 (2012) : 1-17.

Capteurs portables et capteurs ambiants Les capteurs portables (portés sur le corps) peuvent être combiné à des capteurs ambiants (placés à des endroits fixes à domicile) afin d'assurer une surveillance optimale des patients. Typiquement, les capteurs ambiants permettent de détecter des chutes lorsque l'utilisateur ne porte pas son capteur portable. Parmi les capteurs portables, les plus courants sont ceux qui mesurent des données physiologiques (la fréquence cardiaque [130], la fréquence respiratoire [131], la pression artérielle, la saturation en oxygène du sang [132], l'activité musculaire, etc...). Ces capteurs donnent des indicateurs importants de l'état de santé des utilisateurs de manière continue. Les capteurs de mouvement (tels que les gyroscopes, les accéléromètres, les capteurs de pression, les magnétomètres) et les centrales inertielles sont également très utiles pour suivre l'activité des patients sur le long terme (marche, course, montée d'escalier, repos, assis...). Parmi les capteurs ambiants et environnementaux, on peut citer les capteurs et les détecteurs de mouvement à domicile (placés sur une porte d'entrée, une armoire, etc..). La technologie de "maison intelligente", consistant à équiper son domicile de capteurs ambiants, a été adoptée dans diverses applications liées à la réhabilitation [133, 134]. Elle couvre des concepts et des services qui relient les nouvelles technologies et l'environnement social.

IV.10.2 Application dans le domaine de la rééducation

Suivi de la santé et du bien-être Des efforts de recherche considérables ont été déployés ces dernières années pour évaluer la précision des capteurs portables dans la classification des activités de la vie quotidienne (AVQ) pour la surveillance à domicile des personnes âgées et des personnes souffrant de maladies chroniques [135]. Des études ont déjà montré qu'il était possible de correctement identifier les AVQ chez les personnes âgées, que ce soit avec des accéléromètres [136], ou des capteurs de pression dans la chaussure [137]. Des études cliniques ont montré que l'utilisation de capteurs portables lors de thérapie de rééducation à domicile est motivant : en proposant des interfaces ludiques, les capteurs encouragent les patients à poursuivre régulièrement leurs exercices, ce qui s'avère bénéfique sur le long terme [138, 139, 140]. Cependant la plupart des études ont été réalisées en laboratoire, sur un échantillon de personne limité. Il serait nécessaire de valider ces méthodes de suivi à domicile, dans des conditions expérimentales moins contrôlées, et sur une cohorte représentative plus large.

Surveillance et sécurité De très nombreux dispositifs ont été développés pour des applications de surveillance de la sécurité ces dernières années. Des dispositifs de détection de chutes ont été proposés, sous la forme de capteurs portables [141, 142], ou de capteurs embarqués dans le téléphone [143, 144]. Les recherches en cours sont axées sur la prévention des blessures liées aux chutes. Des dispositifs ont été également proposés pour surveiller des déséquilibres dans la démarche de patients atteints de Parkinson [145] et de les patients atteint de démence [146], ou encore pour surveiller les crises potentielles des patients épileptiques [147]. Les applications de surveillance nécessitent généralement la détection d'événements d'urgence. Par conséquent, la technologie de détection utilisée pour de telles applications doit être extrêmement robuste et fiable.

Rééducation à domicile Les capteurs portables peuvent également être utilisés pour faciliter la rééducation à domicile. Ils permettent non seulement de détecter et surveiller les patients au travers de leurs activités quotidiennes, mais aussi de les encourager à poursuivre régulièrement leur thérapie de rééducation, au travers de défis motivants et d'exercices préventifs. Les exercices tentent d'améliorer la motricité des patients, afin de remédier aux déficits fonctionnels. Les dispositifs portables visent à motiver les patients à réaliser régulièrement leurs exercices de motricité, entraînant ainsi une meilleure récupération à long terme. De tels dispositifs ont déjà été proposés dans le cadre de nombreuses pathologies chroniques, notamment pour l'AVC [148, 149].

Évaluation de l'efficacité d'un traitement Les capteurs constituent également un outil pour quantifier l'efficacité d'un traitement dans la gestion de maladies. En fournissant des données en continu

(en particulier entre chaque consultation), les traitements peuvent être ajustés au plus près du patient afin de fournir une médecine personnalisée [150]. Par exemple dans le cadre de la maladie de Parkinson, il est essentiel de bien identifier les dysfonctionnements moteurs suffisamment en avance pour déterminer la posologie de médicaments de manière optimale. L'utilisation d'un système basé sur des capteurs apparaît donc comme une approche prometteuse pour améliorer la prise en charge clinique des patients [151, 152]. De même qu'après un AVC, la rééducation intensive est très importante pour récupérer des fonctions motrices. La surveillance de l'activité motrice peut être utilisée comme un outil de rétroaction pour guider le processus de réadaptation, mais aussi recueillir des données cliniques pertinentes pour le personnel de rééducation concernant l'état moteur du patient [153, 149].

IV.10.3 Nécessité de valider les capteurs portables grand public

Malgré leur essor, l'utilisation des capteurs portables grand public reste contestée dans la communauté médicale : les objections concernent la qualité des données collectées ainsi que la fiabilité des technologies dans un contexte clinique où les pathologies sont très variées voir combinées [154]. Il apparaît encore nécessaire de valider les capteurs afin de lever toute réticence et qu'ils soient intégrés dans un cadre clinique dédié à la réadaptation de patients. Une première étape consisterait à définir des protocoles de validation précis - en concertation avec le corps médical - adaptés à l'étude des pathologies chroniques. En effet, beaucoup d'études se contentent de valider des capteurs pour une application médicale donnée sans même les avoir testés en dehors du laboratoire, sur un nombre très restreint de patients, et sur une fenêtre de temps relativement courte (de l'ordre de quelques heures). Le protocole à définir devrait donc imposer des contraintes d'expérimentation plus proches de la vie quotidienne des patients, à savoir : que les données soient acquises à domicile, mais aussi sur suffisamment de patients, et sur une période d'acquisition suffisamment longue (plusieurs semaines, voire des mois). Il serait aussi nécessaire de définir au sein du protocole quels types de capteurs (actimétriques, physiologiques) seraient plus adaptés, combien de capteurs seraient nécessaires, où les placer sur le patient, et quels indicateurs seraient à extraire préférentiellement. D'autre part, il faudrait garantir que les relevés des capteurs soient suffisamment précis et sensibles pour la collecte de données médicales. Cela est crucial non seulement pour garantir la généralisabilité d'un capteur au sein d'une population cible, mais aussi sa capacité à mesurer la variabilité quotidienne des mesures qui peuvent être corroborées avec les symptômes de la maladie. Dans cette optique, les données acquises par les capteurs devraient systématiquement être comparées aux données acquises par des dispositifs médicaux de référence (c.a.d. des systèmes de pointe dit *Gold Standard*, fiables, souvent non embarquables).

IV.10.4 Compromis entre utilité des données et vie privée

Un autre enjeu de recherche dans le domaine des capteurs portables consiste à mieux sécuriser les données collectées. L'architecture traditionnelle du système portable de surveillance couramment utilisée aujourd'hui, illustrée en Figure IV.10.1, est centralisée et présente peu de protection contre de potentielles cyberattaques. La centralisation des données brutes sur un serveur présente un certain risque, en particulier si les données sont envoyées sur un serveur externe au projet de recherche, car elle facilite l'accès aux personnes malveillantes. Si le serveur est compromis ou si certaines données sont volées, des données brutes sont révélées entraînant la possibilité de faire de nombreuses inférences sensibles, y compris ré-identifier les utilisateurs [155]. Sur cette architecture traditionnelle, les nœuds du système (c'est-à-dire l'application client et le smartphone, ainsi que le serveur d'application) sont considérés fiables et sécurisés. En revanche les communications entre les nœuds le sont beaucoup moins, et sont la cible de potentielles attaques. Bien que chaque information soit transmise par des batch de données annotés indépendamment (en utilisant des pseudonymes aléatoires [156]), la transmission des données reste peu sécurisée et peu fiable. Face à ce problème, une solution consiste à transmettre vers le serveur, non

plus les données brutes de l'utilisateur, mais des données pré-traitées et normalisées qui empêche la ré-identification [157, 158].

Dans le contexte des capteurs portables et appareils mobiles, le défi est donc d'obfusquer les données qui peuvent compromettre la vie privée des individus tout en restant suffisamment pertinentes pour le suivi clinique [159]. Ce défi soulève deux questions importantes : 1) Les données collectées sont-elles suffisamment protégées pour que personne ne puisse en abuser pour inférer des informations sensibles ou pour ré-identifier l'utilisateur? 2) Comment évaluer si les données protégées sont encore suffisamment fiables et précises pour évaluer au mieux l'état de santé du patient? L'équilibre entre utilité des données et vie privée de l'utilisateur est un enjeu primordial pour assurer la précision, la fiabilité, et la sécurité des données envoyées, ainsi que pour renforcer la confiance des utilisateurs vis-à-vis du dispositif. Nous proposons une solution dans le chapitre suivant.

Chapitre IV.11

Anonymisation de données actimétriques issus de capteurs portables par filtrage temps-fréquence

Avec l'émergence de l'internet des objets (IoT), de plus en plus de personnes sont équipées d'un smartphone et d'objets connectés remontant des informations sur leur activité voire sur leur physiologie. Ces nouveaux capteurs ont le potentiel d'assurer un meilleur suivi et accompagnement des patients souffrant d'une maladie chronique. Dans ce chapitre, nous présentons des travaux préliminaires dans lesquels nous explorons des méthodes capables, à partir de signaux actimétriques issus de capteurs, de reconnaître l'activité d'individus tout en masquant leur identité.

IV.11.1 Matériel

Nous avons utilisé l'ensemble de données publiques Motion-Sense pour évaluer la performance de notre approche. Ces données proviennent de capteurs de mouvement à 3 axes, collectés à une fréquence constante de 50 Hz avec un iPhone 6S placé sur la poche avant des participants [160]. Au total, 24 participants ont réalisé six activités au cours de 15 expérimentations dans le même environnement et dans les mêmes conditions d'utilisation. Différentes activités ont été mesurées, tel que descendre et monter des escaliers, marcher, courir, s'asseoir et se lever. Pour permettre deux types de classification continue au cours du temps (l'une pour l'activité, l'autre pour l'identité), les signaux bruts ont été découpés sur des fenêtres glissantes. Sachant que la cadence de marche est en moyenne d'au moins 1.5 pas/seconde [161], la longueur choisie de la fenêtre était de 2.5 secondes avec un chevauchement de 50%. Pour cette étude, nous nous sommes focalisés sur les quatre activités dynamiques (monter, descendre, marcher et courir), dans la mesure où ces activités sont les plus difficiles à analyser et elles contiennent des fréquences complexes adaptés à la représentation temps-fréquence (TF). En outre, nous avons uniquement analysé le signal d'accélération des utilisateurs, qui est adapté pour la détection d'activité [162].

IV.11.2 Domaine temps-fréquence

Le domaine temps-fréquence s'efforce de surmonter les limites de la transformée de Fourier classique, qui fournit uniquement des informations de fréquence (f) sans aucune information temporelle (t). En effet, dans le contexte de signaux non stationnaires, il est nécessaire de connaître l'évolution de la fréquence des composantes du signal en tant que fonction du temps. Il existe plusieurs approches pour projeter le signal dans le domaine TF. Dans nos travaux, nous nous focalisons sur 3 différentes transformées linéaires : la transformée de Fourier à court terme [163], la transformée de Stockwell [164] et la transformée de Stockwell optimisée [165].

IV.11.2.1 La transformée de Fourier à court terme

Étant donné un signal $x(t)$, sa transformée de Fourier à court terme (STFT, pour *Short-Time Frequency Time*) Sf_x peut être exprimée par :

$$Sf_x(t, f) = \int_{-\infty}^{+\infty} x(\tau) w^*(\tau - t) e^{-2j\pi f\tau} d\tau, \quad (\text{IV.11.1})$$

où $w(t)$ est la fenêtre d'analyse de largeur fixe (analyse mono-résolution). Dans cette étude, la fenêtre $w(t)$ a été choisie comme une fonction gaussienne d'écart-type $\sigma = 0.05$.

IV.11.2.2 La transformée de Stockwell

La transformée de Stockwell peut être considérée comme une transformée hybride entre la STFT et la transformée continue en ondelettes (CWT) [166] : tout comme la STFT, elle préserve une relation directe avec le noyau de Fourier, et tout comme la CWT elle effectue une analyse multi-résolution. La transformée de Stockwell S_x du signal $x(t)$ peut être exprimée comme suit :

$$S_x(t, f) = \int_{-\infty}^{+\infty} x(\tau) w^*(\tau - t, f) e^{-2\pi j f\tau} d\tau, \quad (\text{IV.11.2})$$

où $w(t, f)$ est la fenêtre d'analyse représentée par une fonction gaussienne à deux variables : le temps t et la fréquence f . Elle peut être donnée comme :

$$w(t, f) = \frac{1}{\sigma(f)\sqrt{2\pi}} e^{\frac{-t^2}{2\sigma(f)^2}} \quad (\text{IV.11.3})$$

L'écart-type $\sigma(f)$ est inversement proportionnel à la fréquence tel que :

$$\sigma(f) = \frac{1}{|f|}, \quad (\text{IV.11.4})$$

pour faire apparaître la résolution temporelle dans les basses fréquences et la résolution fréquentielle dans les hautes fréquences.

IV.11.2.3 La transformée de Stockwell optimisée

Pour mieux adapter la fenêtre glissante à la nature du signal analysé, de nombreuses études ont tenté d'optimiser la représentation de la transformée de Stockwell en introduisant de nouveaux paramètres à la fenêtre gaussienne [167, 168]. Parmi les propositions, une fenêtre gaussienne généralisée dépendant d'un ensemble p_i de quatre paramètres $\{r, m, p, k\}$ été proposée dans [165] comme suit :

$$w^{p_i}(t, f) = \frac{|f|^r}{(mf^p + k)\sqrt{2\pi}} e^{\frac{-(t-t)^2 f^{2r}}{2(mf^p + k)^2}}. \quad (\text{IV.11.5})$$

L'idée est de choisir l'ensemble de paramètre p_i qui maximise une fonction de concentration d'énergie. Cette concentration d'énergie peut être mesurée par plusieurs approches. Les mesures de concentration CM utilisées dans cette étude sont données comme suit [169] :

$$\text{CM}(p_i) = \frac{1}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |S_x^{p_i}(t, f)| dt df}, \quad (\text{IV.11.6})$$

avec $\overline{S_x^{p_i}(t, f)}$ une normalisation de $S_x^{p_i}(t, f)$ [168] tel que :

$$\overline{S_x^{p_i}(t, f)} = \frac{S_x^{p_i}(t, f)}{\sqrt{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |S_x^{p_i}(t, f)|^2 dt df}}. \quad (\text{IV.11.7})$$

Les paramètres p_i qui maximisent $\text{CM}(p_i)$ sont choisis pour calculer la transformée de Stockwell. Dans nos travaux, l'optimisation a été effectuée sur l'ensemble du signal et les paramètres optimaux ont été calculés pour un échantillon de signaux de l'ensemble de données. À partir de ces critères, les paramètres optimaux trouvés étaient $r = 0, 7$, $m = 0$, $p = 0$ et $k = 0, 4$.

IV.11.3 Filtrage des images temps-fréquences pour anonymiser les signaux

Les images TF générées à partir des différentes transformées TF ont une dimension de 62x128 (62 pixels dans le domaine spectral, et 168 pixels dans le domaine temporel). Comme le montre la Figure IV.11.1, les images TF pour la marche présentent des motifs différents de ceux pour la course, et peuvent être discriminées en termes de texture. Plus particulièrement, le nombre de pics saillants verticaux apparaît comme un critère important de discrimination. D'autre part, des différences de contraste peuvent être observées entre les images TF d'utilisateurs différents : comme le montre Figure IV.11.1, les pics de l'utilisateur n°8 sont plus contrastés que ceux de l'utilisateur n°15. Ces observations soulignent l'intérêt de filtrer les coefficients élevés des images TF pour supprimer les informations relatives à l'identité des utilisateurs.

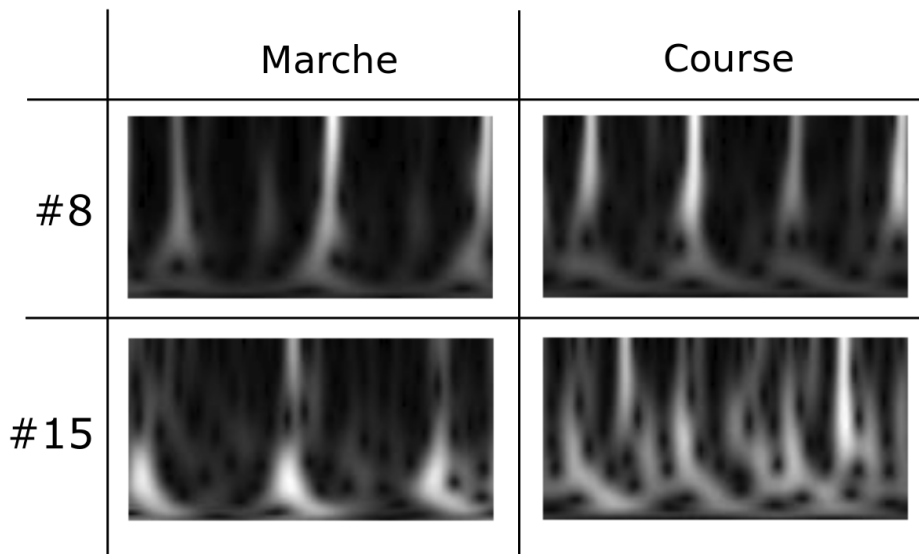


FIGURE IV.11.1 – Représentation de la transformation de Stockwell optimisée pour 2 utilisateurs différents (#8 et #15) pour deux activités différentes (marche et course).

Le filtrage d'identité que nous proposons consiste à fixer à zéro un pourcentage x des coefficients de l'image TF, ces coefficients étant triés au préalable par ordre décroissant. Dans nos travaux, nous avons testé différents pourcentages de filtrage, x allant de 10% à 90% avec un pas de 10%. Cette méthode nous permet, en accord avec [157], de nous assurer que les informations pertinentes pour la ré-identification soient supprimées.

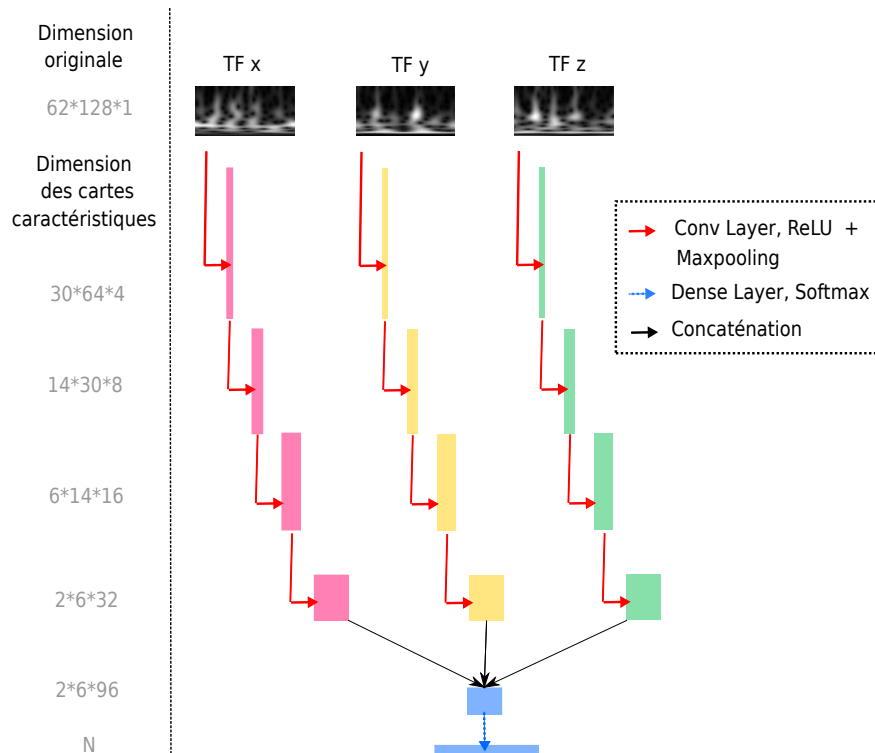


FIGURE IV.11.2 – Aperçu de l’architecture CNN proposée. Le réseau prend trois images TF (TF_x , TF_y , TF_z) en entrée. Chaque image d’entrée est traitée indépendamment sur 3 branches distinctes. Les cartes de caractéristiques roses, jaunes et vertes résultent de convolutions 2D et d’opérations *maxpooling*. Les sorties des 3 branches sont ensuite concaténées, et passent à travers une couche cachée de N nœuds, avec $N = 4$ pour le CNN activité et $N = 24$ pour le CNN identité.

IV.11.4 Évaluation de l’approche proposée

IV.11.4.1 Classification par réseau de neurones

Nous proposons deux réseaux neuronaux convolutifs (CNN) distincts pour nos deux tâches de classification : un réseau à 4 classes pour classifier les signaux en activité appelé CNN activité, et un autre réseau à 24 classes pour classifier les mêmes signaux en identité appelé CNN identité. Ces deux modèles présentent la même architecture, mais sont entraînés séparément. Le pipeline complet proposé est illustré en Figure IV.11.3.

Architecture La classification CNN de multiples images en entrée peut être abordée en utilisant la stratégie de fusion précoce, où toutes les images d’entrée sont concaténées à l’entrée du réseau. Cette stratégie de fusion présente une faible complexité de calcul et est facile à mettre en œuvre [170]. Cependant, il a été démontré dans d’autres contextes que la fusion de caractéristique tardive tient mieux compte de la complexité de chaque entrée et donne de meilleures performances que la fusion précoce [75, 30]. La stratégie de fusion tardive consiste à traiter chaque image d’entrée indépendamment sur des branches convolutionnelles distinctes, et à fusionner les caractéristiques à un niveau supérieur.

Dans notre cas, chaque fenêtre de signal a été définie par trois images TF différentes, représentant l’accélération le long des axes x , y et z (respectivement TF_x , TF_y , TF_z). L’architecture détaillée est représentée sur la Figure IV.11.2.

Détails expérimentaux Tant pour le CNN activité que pour le CNN identité, les signaux du jeu de données ont été répartis selon les expérimentations : 90% des signaux des expérimentations 1, 2, 3, 4, 7, 8, 9 ont été utilisés comme ensemble d’entraînement, et les 10% restants comme ensemble de

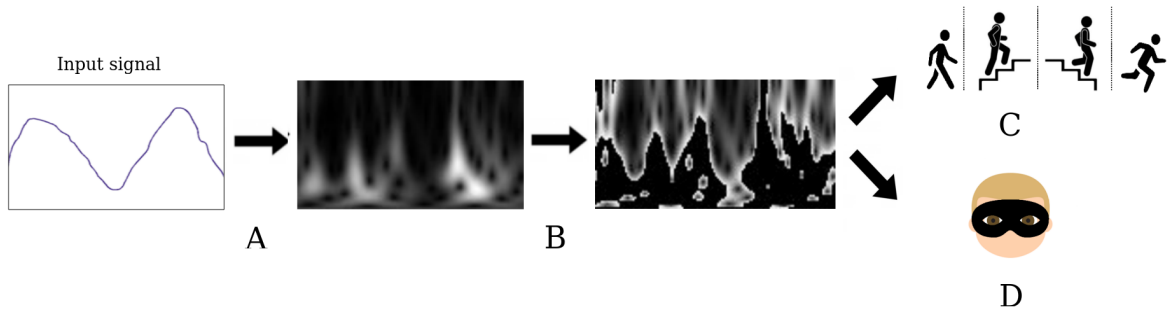


FIGURE IV.11.3 – Vue d'ensemble du pipeline proposé, divisé en 4 étapes : A. Transformation du signal en une image temps-fréquence (TF), B. Méthode d'anonymisation basée sur le filtrage des images, C. Reconnaissance d'activité et D. Identification de l'utilisateur.

validation. Les signaux des expérimentations 11,12,15,16 ont été utilisés comme ensemble de test. Pour les deux CNN, nous avons utilisé une fonction de perte (*loss*) d'entropie croisée catégorique (*categorical cross-entropy*) qui a produit des poids pour pénaliser de manière égale les classes sous- ou sur-représentées dans l'ensemble d'entraînement. L'optimisation a été faite à l'aide de l'algorithme Adam, la taille du *batch* a été fixée à 128, et le nombre d'époques a été fixé à 150 mais a été régulé par du *early stopping*. Le nombre total de poids à entraîner était de 23 044 pour le CNN activité et de 46 104 pour le CNN identité.

IV.11.4.2 Métriques de classification

Pour évaluer les performances de classification des deux CNN (activité et identité), nous avons calculé pour chaque classe la métrique d'accuracy *acc*, définie par :

$$acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (IV.11.8)$$

où TP représente les vrais positifs, TN les vrais négatifs, FN les faux négatifs et FP les faux positifs.

La reconnaissance d'activité était donc mesurée par la variable Activité *acc* (c.a.d. l'accuracy relative à l'utilité des données), et la reconnaissance d'identité était mesurée par la variable Identité *acc* (c.a.d. l'accuracy relative à la vie privée). Les valeurs d'accuracy données correspondent systématiquement aux résultats moyennés sur dix expériences.

IV.11.4.3 Comparaison des performances à celles de l'état de l'art

Pour comparer les performances obtenues à partir des différentes représentations TF, nous nous sommes comparés à une méthode plus classique basée sur la transformée de Fourier du signal d'accélération [157]. Dans cette approche de référence, tout comme dans le pipeline proposé, nous avons alors filtré différents pourcentages *x* des coefficients de la transformée par ordre décroissant (*x* allant de 10% à 90% avec un pas de 10%). Une fois la transformée de Fourier filtrée, tout comme Jourdan et al. [157], le signal a été classé en activité et en identité en fonction des caractéristiques du domaine fréquentiel à l'aide d'un classifieur de forêt aléatoire.

IV.11.4.4 Évaluation de la représentation temps-fréquence optimale

Nous avons défini la meilleure représentation TF comme celle qui maximise la surface sous la courbe utilité/vie privée (AUC) en utilisant la règle du trapèze. L'AUC est une mesure combinée efficace de l'utilité et de la protection de la vie privée, qui décrit la validité inhérente de notre approche d'anonymisation. L'AUC utilisée ici était limitée en abscisse entre le minimum et le maximum de la performance en identité. Comme ces extremums change d'une représentation TF à l'autre, nous avons normalisé l'AUC par la taille du rectangle relative aux extremums.

IV.11.4.5 Évaluation du filtre optimal

Le filtre optimal a été défini comme le filtre minimisant la distance euclidienne entre un point de la courbe utilité-vie privée et le coin supérieur gauche de la zone normalisée. Notons que le coin supérieur gauche correspond à l'intersection du bord supérieur de la Figure IV.11.4 (performance maximale de 100% en activité) et la limite minimale de performance pour l'identité. Ce filtre optimal garantit une bonne reconnaissance de l'activité tout en limitant la reconnaissance d'identité.

IV.11.5 Résultats et discussion

IV.11.5.1 Intérêt de la représentation temps-fréquence

La Figure IV.11.4 montre l'effet du filtrage sur la reconnaissance en activité et en identité pour les trois différentes représentations TF ainsi que pour la méthode de référence. Il semble que travailler uniquement dans le domaine fréquentiel (points verts sur la Figure IV.11.4) entraîne une perte rapide et importante de reconnaissance d'activité. Cette perte d'utilité des données s'explique par le fait que la transformée de Fourier n'est pas en mesure d'analyser correctement la nature non stationnaire des signaux, où le contenu fréquentiel change au fil du temps en particulier pour l'activité. Par opposition, les représentations TF semblent être capables de traiter la non-stationnarité des signaux et donc de permettre un meilleur compromis entre la reconnaissance d'activité et l'identification de l'utilisateur. Plus les coefficients sont filtrés dans les images TF, plus la performance d'identification est mauvaise. En revanche, le filtrage a beaucoup moins d'impact sur les performances de reconnaissance d'activité : qu'aucun filtrage ne soit appliqué, ou que 70% de l'image soit filtrée, l'accuracy de l'activité semble être stable entre 80% et 90%. Cette tendance s'observe quelle que soit la représentation TF (STFT, transformée de Stockwell ou transformée de Stockwell optimisée, figurant respectivement en bleu, en orange et en rouge sur la Figure IV.11.4). Ces résultats démontrent que les coefficients élevés des images TF portent spécifiquement les informations relatives à l'identité des utilisateurs. L'activité, en revanche, semble moins spécifique à une gamme de coefficients et correspond davantage à la texture générale observée dans les images TF.

IV.11.5.2 Représentation optimale

Le Tableau IV.11.1 reporte les AUC normalisées pour chaque courbe de la Figure IV.11.4. Parmi les trois représentations TF, la STFT présente l'AUC la plus faible ($AUC = 0.83$), offrant ainsi un des plus mauvais compromis entre utilité et vie privée. Cette observation peut être liée au fait que cette transformée est mono-résolution et offre donc une résolution temps-fréquence du signal associé plus faible. D'autre part, la transformée de Stockwell est multi-résolution, ce qui permet un meilleur encodage du signal analysé et donc une meilleure reconnaissance de l'activité ($AUC = 0.84$). La reconnaissance de l'activité est encore améliorée lorsque la transformée de Stockwell du signal d'accélération est optimisée selon un critère de concentration d'énergie ($AUC = 0.85$). Les différences entre la transformée de Stockwell et la transformée de Stockwell optimisée pourraient être encore plus marquées si l'optimisation de la transformée de Stockwell avait été effectuée sur des fenêtres glissantes individuelles plutôt que sur l'ensemble du signal. Comme on pouvait s'y attendre, on peut noter également que plus la concentration d'énergie dans le plan temps-fréquence est bonne – ce qui signifie une meilleure résolution tonale de l'image TF – plus le réseau neuronal converge rapidement.

IV.11.5.3 Filtrage optimal

Le Tableau IV.11.2 montre que le filtre optimal qui garantit un bon compromis utilité-vie privée est de 60% pour les représentations de Fourier, STFT et transformée de Stockwell et de 70% pour la représentation optimisée de la transformée de Stockwell. Ces observations suggèrent qu'étant

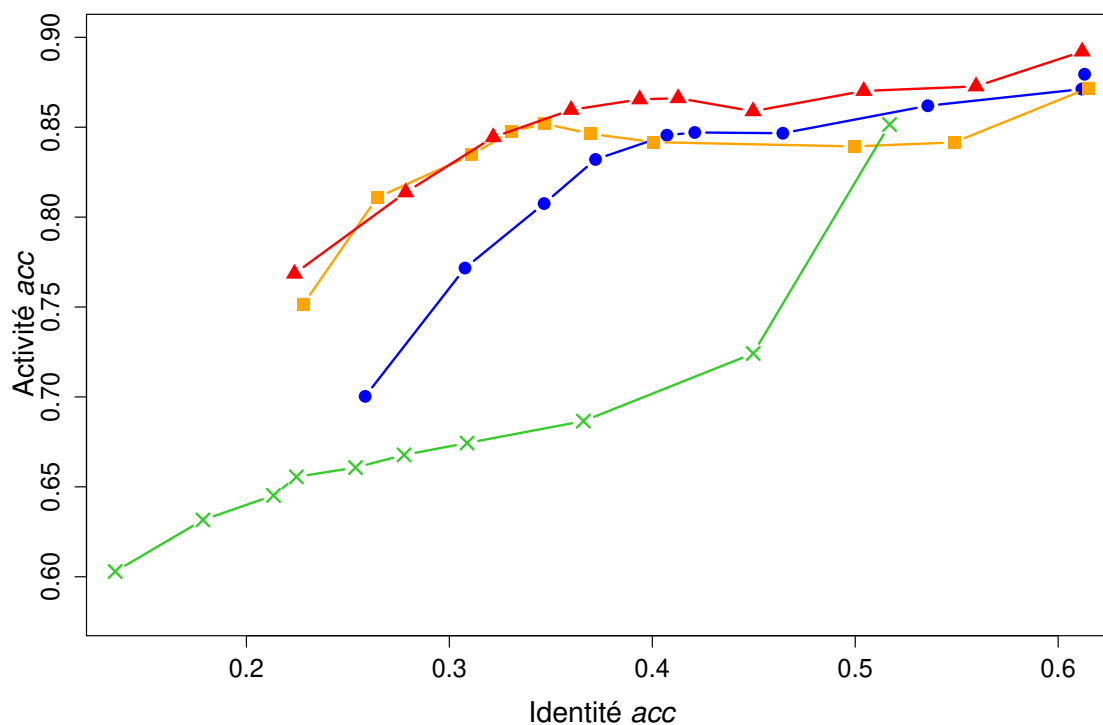


FIGURE IV.11.4 – Accuracy de l’activité en fonction de l’accuracy de l’identité pour différentes représentations : la transformée de Fourier (marqueurs en croix en vert), la STFT (marqueurs ronds en bleu), la transformée de Stockwell (marqueurs carrés en orange) et la transformée de Stockwell optimisée (marqueurs triangulaires en rouge). Chaque point correspond à un résultat de classification moyen sur 10 expériences. Le coin supérieur gauche représente le compromis idéal entre l’utilité et la vie privée. Pour chaque courbe, les points les plus performants en activité et en identité correspondent à des cas sans filtrage alors que les autres (les points qui tendent vers la gauche du graphique) correspondent à des cas avec filtrage avec un pas de 10%.

Fourier	STFT	Transformée Stockwell	Transformée Stockwell optimisée
0.69	0.83	0.84	0.85

TABLEAU IV.11.1 – Aire sous la courbe utilité-vie privée normalisée (AUC) pour chaque représentation

Fourier	STFT	Transformée Stockwell	Transformée Stockwell optimisée
60 (66/22)	60 (83/37)	60 (85/33)	70 (85/32)

TABLEAU IV.11.2 – Filtre optimal en % pour chaque représentation, et les performances associées en % (activité *acc*/identité *acc*)

donné la meilleure résolution tonale de la représentation optimisée de la transformée de Stockwell, il est possible de filtrer davantage de coefficients de l’image TF sans perdre de manière trop significative en performance d’activité (comme observé pour les autres représentations TF de la Figure IV.11.4).

IV.11.6 Conclusion

Dans ce travail, nous avons présenté une méthode preuve de concept pour préserver la vie privée d’utilisateurs de capteurs de mouvement. Cette méthode utilise une représentation temps-

fréquence des signaux d'accélération et filtre les images TF résultantes en fixant les coefficients les plus élevés à zéro avant une étape d'apprentissage machine. Les évaluations ont démontré que notre méthode permettait d'anonymiser et préserver un taux de reconnaissance d'activité élevé, car l'encodage TF que nous proposons est adapté à la nature non stationnaire des signaux (contrairement à la transformée de Fourier classique). Plus précisément, nous avons déterminé que la transformée de Stockwell optimisée offrait le meilleur compromis utilité-vie privée en filtrant les coefficients TF à 70%. Le mécanisme de filtrage proposé pour préserver la vie privée était intentionnellement simple, mais a donné des résultats prometteurs. Des méthodes de filtrage plus avancées [171] pourraient être envisagées pour améliorer les performances, ce qui fera l'objet de recherches futures. En outre, d'autres transformations temps-fréquence pourront être appliquées et comparées aux résultats obtenus.

Chapitre IV.12

Conclusions sur la tâche de prédiction de l'accident vasculaire cérébral en phase aiguë et chronique

Ce chapitre synthétise les résultats principaux développés durant ces trois années de thèse, et met l'accent sur les perspectives de recherche qui apparaissent dans la continuité des contributions présentées dans ce manuscrit. Nous présentons plusieurs pistes de recherche méthodologiques appliquées aux domaines de l'apprentissage automatique et de l'AVC.

IV.12.1 Bilan des contributions

La première partie du manuscrit proposait des travaux fondamentaux sur la valeur prédictive de l'IRM de perfusion brute.

Dans les **chapitres II.3 et II.4**, nous avons exploré différents descripteurs (les LBP, les coefficients d'Haralick et les coefficients de la transformée en scattering) pour résumer efficacement l'information spatio-temporelle locale des voxels de l'imagerie de perfusion. Ces descripteurs ont été choisis car ils présentaient des propriétés intéressantes pour notre tâche de prédiction : ils étaient translation invariants (permettant ainsi de s'affranchir de la variabilité inter-patients de délai d'arrivée de l'agent contraste), et ils balayaient différentes approches d'analyse d'images (des approches structurelles et texturales aussi bien avec les LBP que les coefficients d'Haralick, des approches statistiques avec les coefficients d'Haralick, et fréquentielles avec les coefficients de la transformée en scattering). Les descripteurs proposés semblaient tous en mesure de capturer des informations pertinentes pour la tâche de prédiction donnée. Sur la base de la métrique Dice (DSC), les coefficients d'Haralick se sont démarqués des autres car ils étaient particulièrement efficaces, parcimonieux et rapides d'exécution.

Dans le **chapitre II.5**, nous nous sommes intéressés à la possibilité de prédire l'infarctus final à partir de données de perfusion brutes simulées. Ces données synthétiques ont été obtenues à partir d'un simulateur physique de perfusion, paramétrable selon différentes variables physiologiques. Parmi toutes les paramètres ajustables, nous nous sommes focalisés sur celui de la fonction d'entrée artérielle (AIF). En effet, à travers différentes expériences, nous avons montré que la forme de l'AIF propre aux patients d'entraînement avait une grande influence sur les performances du modèle proposé. Dans ce contexte, nous avons montré l'intérêt de l'apprentissage patient-spécifique : en ajustant l'AIF de l'ensemble d'entraînement simulé à celle d'un patient test, on injectait un priori physiologique fort dans les données d'apprentissage, ce qui permettait de contraindre et faire converger le modèle d'apprentissage vers un optimum global.

Dans le **chapitre II.6**, nous avons proposé des améliorations au réalisme physiologique du simulateur, en proposant d'intégrer de nouvelles informations relatives à l'anatomie de l'arbre vasculaire (les variations du flux et du temps de transport de l'agent de contraste). Tout comme pour le chapitre précédent, les modèles proposés ont été appris sur les données de perfusion simulées,

et testés sur des données de perfusion réelles, via une approche patient-spécifique. L'intégration des nouveaux degrés de réalisme dans le simulateur a permis d'améliorer les performances générales de nos modèles prédictifs.

La seconde partie du manuscrit proposait des méthodes d'apprentissage supervisé capables de coupler l'imagerie de perfusion déconvoluée et l'imagerie de diffusion.

Dans le **chapitre III.7**, à travers une étude pilote, nous avons proposé d'encoder simplement les images d'entrée, et de prédire l'état final de chaque voxel à l'aide d'un classifieur bayésien naïf. Le modèle s'est révélé performant, peu coûteux en temps de calcul, et présentait peu de paramètres. Pour ce travail, il est important de souligner que les performances obtenues reposaient sur la pertinence de l'encodage des imageries aiguës, qui se basait sur un seuillage multi-échelle. En travaillant sur un ensemble de patients homogènes, nous avons facilement pu déterminer l'encodage optimal pour tous les patients de l'étude. À l'échelle d'une cohorte plus large, trouver un encodage optimal pour tous les patients serait probablement moins évident.

Dans le **chapitre III.8**, nous avons exploré plus en profondeur un modèle d'apprentissage profond par réseau de neurones, présentant beaucoup de paramètres mais capable de traiter les données sans aucun pré-traitement. À travers nos expériences, nous avons montré que la fusion tardive des caractéristiques semblaient améliorer les performances de prédiction, en intégrant potentiellement mieux chaque modalité au sein du réseau. Nous avons également constaté que les réseaux de neurones pouvaient être très sensibles au statut de reperfusion des données d'entraînement, et que cela avait un impact considérable sur la lésion finale prédite en sortie. Nous avons proposé de stratifier l'ensemble d'entraînement selon cette variable de reperfusion. En séparant les patients reperfusés des patients non reperfusés, les performances de prédiction étaient meilleures que celles obtenues en apprenant sur la cohorte globale. Proposer deux modèles spécifiques au statut de reperfusion faisait aussi sens d'un point de vue clinique : la comparaison entre la lésion prédite en cas de reperfusion réussie vs échouée peut aider les cliniciens à estimer l'effet d'un traitement et les guider dans leur décisions thérapeutiques.

Dans le **chapitre III.9**, nous nous sommes focalisé sur la réversion de la lésion de diffusion aiguë : un phénomène de plus en plus fréquent depuis l'introduction du traitement par thrombectomie, mais pour autant encore très peu documenté dans la littérature. Notre objectif était donc de proposer une première étude capable d'identifier des marqueurs d'imagerie importants pour prédire la réversion. Nous avons proposé un premier modèle d'apprentissage non linéaire capable de prédire la réversion de diffusion à l'échelle du voxel. Notre choix s'est porté sur les forêts aléatoires, un type de modèle basé sur les arbres de décisions reconnus dans la littérature pour être performants, parcimonieux et rapides à optimiser. Nous avons cherché des indices d'explicabilité du modèle, afin de mesurer directement les effets des variables sur le statut de réversion. Nous avons ainsi montré que les images DWI et T_{max} sont les variables les plus importantes dans le processus de décision des arbres de la forêt : lorsqu'un voxel présente des valeurs d'intensité DWI et T_{max} élevées, la probabilité qu'il ne reverse pas est particulièrement élevée.

IV.12.2 Discussion et perspectives de recherche

Au travers des différents travaux et modèles associés, présentés tout au long de ce document, nous avons montré à plusieurs reprises l'intérêt de stratifier les données d'entraînement. Autrement dit, homogénéiser les données d'entraînement permet aux modèles de mieux converger vers leur optimum global. Cette stratification a été réalisée à différents niveaux, et au regard de différents paramètres physiologiques : la fonction d'entrée artérielle (chapitre II.5), les formes de lésions (chapitre II.3, II.4, II.5, et III.7), le statut de reperfusion (chapitre III.8), ou encore la thérapeutique (chapitre III.8 et III.9). On pourrait également imaginer d'autres groupes possibles de patients. On sait par exemple que le délai entre les premiers symptômes et l'IRM est un facteur déterminant d'évolution de la lésion [11] (plus le délai est long, moins bonnes seront les chances du

patient de reperfuser). On sait aussi que lorsque le statut du réseau de collatérales est un facteur déterminant d'évolution de la lésion : après un traitement aigu, de bonnes circulations collatérales sont associées à un devenir clinique favorable tandis que de mauvaises circulations collatérales sont liées à des complications hémorragiques et à de faibles taux de recanalisation [172]. Certains marqueurs d'inflammation sériques [173], encore à l'étude, pourraient se révéler être des indices de stratification pertinents. Stratifier permet de créer des groupes de patients homogènes, qui suivent la même évolution de lésion ischémique [174]. Dans la mesure où la tâche de prédiction est complexe et les scénarios d'évolution très variés, stratifier les données peut être un moyen de réduire l'espace de recherche à partir d'a priori, pour contraindre efficacement l'optimisation d'un modèle. En réduisant l'espace de recherche, on réduit également la complexité de calcul ainsi que le nombre de paramètres nécessaires au processus d'optimisation. Bien entendu, cette approche a ses limites : en stratifiant selon trop de paramètres, on finit par créer un grand nombre de clusters avec trop peu de patients, qui ne brassent pas suffisamment de diversité.

Une autre possibilité, que nous n'avons pas exploré mais qui mérite d'être approfondie, consisterait à fusionner des variables cliniques catégorielles (par exemple le statut de reperfusion, le traitement, l'âge, le délai de traitement, le grade de collatérales) aux données images. Cette approche est loin d'être triviale, puisqu'elle consiste à fusionner à l'échelle d'un patient ses caractéristiques cliniques (des scalaires) avec ses données images (matrices 2D ou 3D). Au-delà du fait que les données proviennent de sources différentes, les dimensions et les structures diffèrent. La difficulté majeure consiste à trouver un espace de représentation commun optimal dans lequel combiner ces données de nature différente. Dans cette démarche, Pinto et al. ont proposé, dans le cadre du challenge ISLES 2017 [17], un modèle d'apprentissage profond U-Net dans lequel ils intègrent un paramètre relatif au statut de reperfusion dans la fonction de perte [175]. Bien que l'idée soit novatrice dans le domaine de la prédiction de lésion ischémique, leur modèle ne s'est pas très montré compétitif par rapport aux autres challengers. Dans le cadre du challenge ISLES 2016, Choi et al. ont proposé un modèle d'apprentissage profond FCN (*Fully Convolutional Network*) pour classifier des patches des données IRM en sain ou pathologique [176]. Les patches images ont été combinés à des données cliniques (telles que le délai entre les premiers symptômes et l'IRM, le délai entre l'IRM et le traitement, et le TIC1) au niveau de l'avant dernière couche convolutive. Contrairement au modèle U-Net, les modèles FCN ont la capacité d'opérer sur des données de tailles variables, et de faciliter leur fusion au niveau des caractéristiques. Comme le soulèvent les auteurs dans leur étude [176], l'inconvénient de leur modèle est que la prédiction est particulièrement bruitée, probablement parce que l'encodage patch des données images n'assure pas un contexte spatial optimal. Une perspective de recherche consisterait donc à proposer un FCN qui prend en entrée les images complètes, qu'il fusionnerait tardivement (c.a.d. au niveau des caractéristiques) avec des données cliniques. Une autre possibilité de recherche consisterait à s'inspirer de travaux en dehors du domaine de l'imagerie médicale qui sont en rapport avec la fusion de données hétérogènes. Par exemple, Srivastava et al. ont récemment proposé un modèle Markovien (plus précisément une machine de Boltzmann) capable de combiner des données textes avec des données images [177]. Un état de l'art approfondi serait nécessaire afin d'évaluer la pertinence et les conditions d'application de ce type de modèle pour notre problématique.

Un autre moyen de guider un modèle, en particulier les modèles d'apprentissage profond, consiste à imposer des a priori (contraintes anatomiques et/ou physiologiques) lors de la phase d'apprentissage [178, 179]. Dans le cadre de notre application, nous avons déjà imposé que l'hémisphère ipsilatéral se situe toujours du même côté pour les patients d'étude (chapitre III.7). Dans la même optique, on pourrait penser à recalculer les données sur un atlas (http://nist.mni.mcgill.ca/?page_id=714), afin d'effacer la variabilité inter-individuelle non essentielle à la tâche de prédiction. Pour aller plus loin, nous pourrions recalculer les données sur un atlas physiologique afin d'identifier de façon uniforme les différents territoires vasculaires de chaque patient. Ainsi, selon la localisation de l'occlusion, on pourrait guider le modèle vers le territoire ischémique approprié.

Un autre axe de recherche consisterait à améliorer l'interprétabilité des modèles non linéaires d'apprentissage automatique. Nous avons déjà fait un premier pas d'explicabilité dans nos travaux actuels : dans le chapitre III.8 nous avons proposé des expériences d'ablation pour mesurer l'impact de contribution de chaque modalité d'entrée sur les performances générales du modèle, et dans le chapitre III.9, nous avons utilisé un indice issu de la théorie des jeux pour représenter et mesurer les effets de chaque variable du modèle sur la prédiction. Nous souhaiterions aller plus loin dans cette démarche d'interprétabilité, en particulier concernant l'apprentissage profond par réseau de neurones (CNN). Une possibilité consisterait à mieux tenir compte des cartes de caractéristiques générées par le réseau. Ces cartes, résultantes de l'application successive de filtres aux images d'entrée, pourraient fournir un aperçu de la représentation interne que le modèle possède d'une entrée spécifique à un point donné du modèle. Au niveau des premières couches du modèle, les cartes de caractéristiques montrent souvent des détails fins, qui sont facilement interprétables par l'œil humain. Les cartes de caractéristiques des couches plus profondes sont quant à elles moins bien résolues et montrent beaucoup moins de détails. Les cartes de caractéristiques permettent donc de résumer les données d'entrée en concepts plus généraux, sur lesquels nous pourrions nous appuyer pour s'assurer que le réseau converge vers la solution souhaitée [180]. Récemment, une méthode (*DeepLIFT*) a été proposée permettant de décomposer la prédiction de sortie d'un réseau CNN en rétropropageant les contributions de tous les neurones du réseau sur les données d'entrée du réseau [181]. Concrètement, le modèle offre la possibilité pour chaque donnée test de savoir quels modalités d'entrée ont contribué (positivement ou négativement) à la prédiction, et dans quelle mesure. Cette méthode serait intéressante à tester pour mieux comprendre nos cas d'échec de prédiction. Toujours dans une optique d'interprétabilité des CNN, une autre possibilité consisterait à visualiser les caractéristiques internes apprises par les CNN à l'aide d'outils de projection de données multi-dimensionnelles, tels que t-SNE (*t-distributed Stochastic Neighbor Embedding*) [182], afin d'observer les distributions des différents clusters appris [183].

Enfin, nous proposons une dernière piste de recherche dans la continuité de nos travaux sur les capteurs adaptés au suivi médicalisé de l'AVC lors de la phase chronique. Dans l'ouverture du manuscrit, nous avons proposé de représenter les signaux actimétriques sous forme d'images temps-fréquences (TF), que nous avons seuillé de façon à optimiser le compromis utilité-vie privée. Ce seuillage était intentionnellement simple, et consistait à annuler les plus hauts coefficients sur l'image TF. Nous avons ainsi pu constater que ces hauts coefficients portaient l'information d'identité, mais pas spécifiquement celle d'activité. Pour nos prochains travaux, nous souhaiterions proposer un encodage plus sophistiqué afin de filtrer plus efficacement les informations d'identité. Une représentation simplifiée de la structure TF d'un signal peut être donnée par la triangulation de Delaunay définie sur les zéros du spectrogramme. Il s'avère que chaque région TF associée à une composante donnée peut se voir comme l'union de triangles de Delaunay adjacents dont au moins un côté est anormalement long en comparaison avec la distribution [171]. Ainsi, nous pourrions distinguer précisément les composantes utilité et identité sur l'image TF, et filtrer les régions associées à l'identité.

Contributions de l'auteur

Articles

Giacalone, M., Rasti, P., Debs, N., Frindel, C., Cho, T. H., Grenier, E., & Rousseau, D. (2018). Local spatio-temporal encoding of raw perfusion MRI for the prediction of final lesion in stroke. *Medical Image Analysis*, 50, 117-126. (c.f. **chapitre II.3**)

Debs, N., Rasti, P., Victor, L., Cho, T. H., Frindel, C., & Rousseau, D. (2020). Simulated perfusion MRI data to boost training of convolutional neural networks for lesion fate prediction in acute stroke. *Computers in Biology and Medicine*, 116, 103579. (c.f. **chapitre II.5**)

Debs, N., Cho T.H., Rousseau, D., Berthezène, Y., Buisson, M., Eker, O., Nighoghossian, N., Ovize, M. & Frindel, C. (2020) Impact of the reperfusion status for predicting the final stroke infarct using deep learning. Submitted in *NeuroImage : Clinical*. (c.f. **chapitre III.8**)

Debs, N., Frindel, C., Mechtouff L., Buisson, M., M. Berthezène, Y., Nighoghossian, N., Rousseau, D., Eker, O., & Cho T.H. (2020) Voxel-based prediction of diffusion lesion reversal in ischemic stroke patients treated with thrombectomy. To be submitted in *Stroke*. (c.f. **chapitre III.9**)

Conférence internationales

Debs, N., Giacalone, M., Rasti, P., Cho, T. H., Frindel, C., & Rousseau, D. (2018). Perfusion MRI in stroke as a regional spatio-temporal texture. In *Joint Annual Meeting ISMRM-ESMRMB*, Paris, France, June 2018. (c.f. **chapitre II.4**)

Debs, N., Decroocq, M., Cho, T. H., Rousseau, D., & Frindel, C. (2019, October). Evaluation of the Realism of an MRI Simulator for Stroke Lesion Prediction Using Convolutional Neural Network. In *International Workshop on Simulation and Synthesis in Medical Imaging* (pp. 151-160). Springer, Cham., Shenzhen, China, October 2019. (c.f. **chapitre II.6**)

Debs, N., Jourdan, T., Moukadem, A., Boutet, A., & Frindel, C., (2020) Motion sensor data anonymization by time-frequency filtering. In *28th European Signal Processing Conference (EUSIPCO 2020)*, Amsterdam, Netherlands, August 2020. (c.f. **chapitre IV.11**)

Conférence nationales

Debs, N., Frindel, C., Cho, T. H., & Rousseau, D. (2019). Classification bayésienne naïve appliquée à la prédiction de l'évolution de lésion dans l'accident vasculaire cérébrale ischémique. In *XXVII^{ème} GRETSI*, Lille, France, August 2019. (c.f. **chapitre III.7**)

Colloque d'enseignement

Debs, N., Peignier, S., Douarre, C., Jourdan, T., Rigotti, C., Frindel, C., (2020) Apprendre l'apprentissage automatique : un retour d'expérience. Submitted in *CETSI*.

Bibliographie

- [1] Fernando Calamante, David G Gadian, and Alan Connelly. Quantification of bolus-tracking mri : improved characterization of the tissue residue function using tikhonov regularization. *Magnetic Resonance in Medicine : An Official Journal of the International Society for Magnetic Resonance in Medicine*, 50(6) :1237–1247, 2003. [8](#), [24](#)
- [2] Accident vasculaire cérébral (AVC). <https://www.inserm.fr/information-en-sante/dossiers-information/accident-vasculaire-cerebral-avc>. Accessed : 2020-04-08. [2](#), [4](#)
- [3] Walter Johnson, Oyere Onuma, Mayowa Owolabi, and Sonal Sachdev. Stroke : a global response is needed. *Bulletin of the World Health Organization*, 94(9) :634, 2016. [2](#)
- [4] Leif Østergaard. Principles of cerebral perfusion imaging by bolus tracking. *Journal of Magnetic Resonance Imaging : An Official Journal of the International Society for Magnetic Resonance in Medicine*, 22(6) :710–717, 2005. [10](#)
- [5] Leif Østergaard, Robert M Weisskoff, David A Chesler, Carsten Gyldensted, and Bruce R Rosen. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. part i : Mathematical approach and statistical analysis. *Magnetic resonance in medicine*, 36(5) :715–725, 1996. [10](#)
- [6] Leif Østergaard, Alma Gregory Sorensen, Kenneth K Kwong, Robert M Weisskoff, Carsten Gyldensted, and Bruce R Rosen. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. part ii : Experimental comparison and preliminary results. *Magnetic resonance in medicine*, 36(5) :726–736, 1996. [10](#)
- [7] Carole Frindel, Marc C Robini, and David Rousseau. A 3-D spatio-temporal deconvolution approach for MR perfusion in the brain. *Medical Image Analysis*, 18(1) :144–160, 2014. [10](#), [35](#), [46](#)
- [8] PA Barber, DG Darby, PM Desmond, Q Yang, RP Gerraty, D Jolley, GA Donnan, BM Tress, and Stephen M Davis. Prediction of stroke outcome with echoplanar perfusion-and diffusion-weighted mri. *Neurology*, 51(2) :418–426, 1998. [10](#), [63](#)
- [9] Manabu Inoue, Michael Mlynash, Matus Straka, Maarten G Lansberg, Greg Zaharchuk, Roland Bammer, and Gregory W Albers. Patients with the malignant profile within 3 hours of symptom onset have very poor outcomes after intravenous tissue-type plasminogen activator therapy. *Stroke*, 43(9) :2494–2496, 2012. [11](#)
- [10] Jeffrey L Saver, Mayank Goyal, Alain Bonafe, Hans-Christoph Diener, Elad I Levy, Vitor M Pereira, Gregory W Albers, Christophe Cognard, David J Cohen, Werner Hacke, et al. Stent-retriever thrombectomy after intravenous t-pa vs. t-pa alone in stroke. *New England Journal of Medicine*, 372(24) :2285–2295, 2015. [11](#)
- [11] David S Liebeskind, Dezhi Liu, Nerses Sanossian, Sunil A Sheth, Conrad Liang, Mark S Johnson, Latisha K Ali, Doojin Kim, Jason D Hinman, Neal M Rao, et al. Time is brain on the col-

lateral clock! collaterals and reperfusion determine tissue injury. *Stroke*, 46(suppl_1) :A182–A182, 2015. [11](#), [61](#), [99](#)

- [12] Søren Christensen, Kim Mouridsen, Ona Wu, Niels Hjort, Henrik Karstoft, Gotz Thomalla, Joachim Rother, Jens Fiehler, Thomas Kucinski, and Leif Østergaard. Comparison of 10 perfusion mri parameters in 97 sub-6-hour stroke patients using voxel-based receiver operating characteristics analysis. *Stroke*, 40(6) :2055–2061, 2009. [14](#)
- [13] Lisa Willats, Alan Connelly, Soren Christensen, Geoffrey A Donnan, Stephen M Davis, and Fernando Calamante. The role of bolus delay and dispersion in predictor models for stroke. *Stroke*, 43(4) :1025–1031, 2012. [14](#)
- [14] Midas Meijs, Soren Christensen, Maarten G Lansberg, Gregory W Albers, and Fernando Calamante. Analysis of perfusion MRI in stroke : To deconvolve, or not to deconvolve. *Magnetic Resonance in Medicine*, 76(4) :1282–1290, 2016. [14](#), [36](#), [44](#)
- [15] Anne Nielsen, Mikkel Bo Hansen, Anna Tietze, and Kim Mouridsen. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. *Stroke*, pages STROKEAHA–117, 2018. [14](#), [35](#), [59](#), [61](#), [63](#), [65](#), [72](#)
- [16] Michelle Livne, Jens K Boldsen, Irene K Mikkelsen, Jochen B Fiebach, Jan Sobesky, and Kim Mouridsen. Boosted tree model reforms multimodal magnetic resonance imaging infarct prediction in acute stroke. *Stroke*, 49(4) :912–918, 2018. [14](#)
- [17] Stefan Winzeck, Arsany Hakim, Richard McKinley, José AADSR Pinto, Victor Alves, Carlos Silva, Maxim Pisov, Egor Krivov, Mikhail Belyaev, Miguel Monteiro, et al. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. *Frontiers in Neurology*, 9 :679, 2018. [14](#), [31](#), [40](#), [53](#), [61](#), [63](#), [72](#), [100](#)
- [18] Adriano Pinto, Sérgio Pereira, Raphael Meier, Victor Alves, Roland Wiest, Carlos A Silva, and Mauricio Reyes. Enhancing clinical mri perfusion maps with data-driven maps of complementary nature for lesion outcome prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 107–115. Springer, 2018. [14](#), [61](#)
- [19] Mathilde Giacalone, Pejman Rasti, Noëlie Debs, Carole Frindel, Tae-Hee Cho, Emmanuel Grenier, and David Rousseau. Local spatio-temporal encoding of raw perfusion mri for the prediction of final lesion in stroke. *Medical image analysis*, 50 :117–126, 2018. [14](#), [19](#), [25](#), [27](#), [31](#), [38](#), [39](#), [46](#), [52](#)
- [20] David Robben, Anna MM Boers, Henk A Marquering, Lucianne LCM Langezaal, Yvo BWEM Roos, Robert J van Oostenbrugge, Wim H van Zwam, Diederik WJ Dippel, Charles BLM Majoie, Aad van der Lugt, et al. Prediction of final infarct volume from native ct perfusion and treatment parameters using deep learning. *Medical image analysis*, 59 :101589, 2020. [14](#), [61](#), [72](#)
- [21] Mathilde Giacalone, Carole Frindel, Marc Robini, Frédéric Cervenansky, Emmanuel Grenier, and David Rousseau. Robustness of spatio-temporal regularization in perfusion MRI deconvolution : An application to acute ischemic stroke. *Magnetic Resonance in Medicine*, 78(5) :1981–1990, 2017. [14](#), [35](#), [36](#), [46](#), [49](#)
- [22] Alex Pappachen James and Belur V Dasarathy. Medical image fusion : A survey of the state of the art. *Information fusion*, 19 :4–19, 2014. [14](#)
- [23] Tong Meng, Xuyang Jing, Zheng Yan, and Witold Pedrycz. A survey on machine learning for data fusion. *Information Fusion*, 57 :115–129, 2020. [15](#)

- [24] Silvia Beddar-Wiesing and Maarten Bieshaar. Multi-sensor data and knowledge fusion—a proposal for a terminology definition. *arXiv preprint arXiv :2001.04171*, 2020. [15](#)
- [25] Ni-Bin Chang and Kaixu Bai. *Multisensor data fusion and machine learning for environmental remote sensing*. CRC Press, 2018. [15](#)
- [26] L Yang, BL Guo, and W Ni. Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform. *Neurocomputing*, 72(1-3) :203–211, 2008. [15](#)
- [27] Gaurav Bhatnagar, QM Jonathan Wu, and Zheng Liu. Directive contrast based multimodal medical image fusion in nsct domain. *IEEE transactions on multimedia*, 15(5) :1014–1024, 2013. [15](#)
- [28] Sabalan Daneshvar and Hassan Ghassemian. Mri and pet image fusion by combining ihs and retina-inspired models. *Information fusion*, 11(2) :114–123, 2010. [15](#)
- [29] H GholamHosseini, Azra Alizad, and Mostafa Fatemi. Fusion of vibro-acoustography images and x-ray mammography. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2803–2806. IEEE, 2006. [15](#)
- [30] Dong Nie, Li Wang, Yaozong Gao, and Dinggang Shen. Fully convolutional networks for multi-modality isointense infant brain image segmentation. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, pages 1342–1345. IEEE, 2016. [15](#), [63](#), [93](#)
- [31] Jose Dolz, Ismail Ben Ayed, and Christian Desrosiers. Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities. In *International MICCAI Brainlesion Workshop*, pages 271–282. Springer, 2018. [15](#), [63](#)
- [32] Shang-Hong Lai and Ming Fang. Adaptive medical image visualization based on hierarchical neural networks and intelligent decision fusion. In *Neural Networks for Signal Processing VIII. Proceedings of the 1998 IEEE Signal Processing Society Workshop (Cat. No. 98TH8378)*, pages 438–447. IEEE, 1998. [15](#)
- [33] Jonathan L Jesneck, Sayan Mukherjee, Loren W Nolte, Anna E Lokshin, Jeffrey R Marks, and Joseph Lo. Decision fusion of circulating markers for breast cancer detection in premenopausal women. In *2007 IEEE 7th International Symposium on BioInformatics and BioEngineering*, pages 1434–1438. IEEE, 2007. [15](#)
- [34] Hyunseok Kook, Lalit Gupta, Srinivas Kota, and Dennis Molfese. A dynamic multi-channel decision-fusion strategy to classify differential brain activity. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3212–3215. IEEE, 2007. [15](#)
- [35] Yunfeng Wu, Jinming Zhang, Cong Wang, and Sin Chun Ng. Linear decision fusions in multilayer perceptrons for breast cancer diagnosis. In *17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05)*, pages 2–pp. IEEE, 2005. [15](#)
- [36] Gert Kwakkel, Boudewijn J Kollen, Jeroen van der Grond, and Arie JH Prevo. Probability of regaining dexterity in the flaccid upper limb : impact of severity of paresis and time since onset in acute stroke. *Stroke*, 34(9) :2181–2186, 2003. [16](#)
- [37] Lancet The. 21st century management and prevention of stroke. *Lancet (London, England)*, 392(10154) :1167, 2018. [16](#)
- [38] Shu-Ya Chen and Carolee J Winstein. A systematic review of voluntary arm recovery in hemiparetic stroke : critical predictors for meaningful outcomes using the international classification of functioning, disability, and health. *Journal of Neurologic Physical Therapy*, 33(1) :2–13, 2009. [16](#)

- [39] A Ballivian, J Azevedo, W Durbin, J Rios, J Godoy, and C Borisova. *Mobile Research Methods : Opportunities and challenges of mobile research methodologies*. Ubiquity Press, 2015. [16](#)
- [40] Kelly R Evenson, Michelle M Goto, and Robert D Furberg. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1) :159, 2015. [17](#)
- [41] Natalia Neverova, Christian Wolf, Griffin Lacey, Lex Fridman, Deepak Chandra, Brandon Barbello, and Graham Taylor. Learning human identity from motion patterns. *IEEE Access*, 4 :1810–1820, 2016. [17](#)
- [42] Daniel Wood, Noah Apthorpe, and Nick Feamster. Cleartext data transmissions in consumer iot medical devices. In *Proceedings of the 2017 Workshop on Internet of Things Security and Privacy*, pages 7–12, 2017. [17](#)
- [43] Michael Rushanan, Aviel D Rubin, Denis Foo Kune, and Colleen M Swanson. Sok : Security and privacy in implantable medical devices and body area networks. In *2014 IEEE Symposium on Security and Privacy*, pages 524–539. IEEE, 2014. [17](#)
- [44] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7) :971–987, 2002. [21](#)
- [45] Faouzi Adjed, Ibrahima Faye, Fakhreddine Ababsa, Syed Jamal Gardezi, and Sarat Chandra Dass. Classification of skin cancer images using local binary pattern and svm classifier. In *AIP Conference Proceedings*, page 080006. AIP Publishing LLC, 2016. [21](#)
- [46] Yu-Ying Liu, Mei Chen, Hiroshi Ishikawa, Gadi Wollstein, Joel S Schuman, and James M Rehg. Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Medical image analysis*, 15(5) :748–759, 2011. [21](#)
- [47] Sunhua Wan, Hsiang-Chieh Lee, Xiaolei Huang, Ting Xu, Tao Xu, Xianxu Zeng, Zhan Zhang, Yuri Sheikine, James L Connolly, James G Fujimoto, et al. Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy. *Medical image analysis*, 38 :104–116, 2017. [21](#)
- [48] Laure Hermitte, Tae-Hee Cho, Brice Ozenne, Norbert Nighoghossian, Irene Klærke Mikkelsen, Lars Ribe, Jean-Claude Baron, Leif Østergaard, Laurent Derex, Niels Hjort, et al. Very low cerebral blood volume predicts parenchymal hematoma in acute ischemic stroke. *Stroke*, 44(8) :2318–2320, 2013. [23](#), [30](#), [39](#), [57](#), [118](#)
- [49] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3) :297–302, 1945. [25](#), [31](#), [40](#)
- [50] Robert M Haralick, Karthikeyan Shanmugam, and Its’ Hak Dinstein. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6) :610–621, 1973. [26](#), [27](#), [28](#)
- [51] Stéphane Mallat. Recursive interferometric representation. In *Proc. of EUSIPCO conference, Denmark*, 2010. [26](#), [27](#), [29](#)
- [52] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10) :1331–1398, 2012. [26](#), [27](#), [29](#)
- [53] Marius E Mayerhoefer, Wolfgang Schima, Siegfried Trattnig, Katja Pinker, Vanessa Berger-Kulemann, and Ahmed Ba-Ssalamah. Texture-based classification of focal liver lesions on mri at 3.0 tesla : A feasibility study in cysts and hemangiomas. *Journal of Magnetic Resonance Imaging*, 32(2) :352–359, 2010. [27](#)

- [54] Weijie Chen, Maryellen L Giger, Hui Li, Ulrich Bick, and Gillian M Newstead. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magnetic Resonance in Medicine : An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(3) :562–571, 2007. [27](#)
- [55] Ke Nie, Jeon-Hor Chen, J Yu Hon, Yong Chu, Orhan Nalcioglu, and Min-Ying Su. Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast mri. *Academic radiology*, 15(12) :1513–1525, 2008. [27](#)
- [56] Andreas Wibmer, Hedvig Hricak, Tatsuo Gondo, Kazuhiro Matsumoto, Harini Veeraraghavan, Duc Fehr, Junting Zheng, Debra Goldman, Chaya Moskowitz, Samson W Fine, et al. Haralick texture analysis of prostate mri : utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different gleason scores. *European radiology*, 25(10) :2840–2850, 2015. [27](#)
- [57] Patrik Brynolfsson, David Nilsson, Roger Henriksson, Jón Hauksson, Mikael Karlsson, Anders Garpebring, Richard Birgander, Johan Trygg, Tufve Nyholm, and Thomas Asklund. ADC texture—an imaging biomarker for high-grade glioma? *Medical physics*, 41(10) :101903, 2014. [27](#)
- [58] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8) :1872–1886, 2013. [30](#)
- [59] Jeff Henrikson. Completeness and total boundedness of the Hausdorff metric. *MIT Undergraduate Journal of Mathematics*, 1 :69–80, 1999. [31](#), [40](#)
- [60] Yannan Yu, Bhargav Parsi, William Speier, Corey Arnold, Min Lou, and Fabien Scalzo. Lstm network for prediction of hemorrhagic transformation in acute stroke. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 177–185. Springer, 2019. [31](#)
- [61] Noah Stier, Nicholas Vincent, David Liebeskind, and Fabien Scalzo. Deep learning of tissue fate features in acute ischemic stroke. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1316–1321. IEEE, 2015. [35](#)
- [62] Rongzhao Zhang, Lei Zhao, Wutao Lou, Jill M Abrigo, Vincent CT Mok, Winnie CW Chu, Defeng Wang, and Lin Shi. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional densenets. *IEEE Transactions on Medical Imaging*, 37(9) :2149–2160, 2018. [35](#)
- [63] Christian Lucas, André Kemmling, Amir Madany Mamlouk, and Mattias P Heinrich. Multi-scale neural network for automatic segmentation of ischemic strokes on acute perfusion images. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1118–1121. IEEE, 2018. [35](#)
- [64] Stefano Pedemonte, Bernardo Bizzo, Stuart Pomerantz, Neil Tenenholtz, Bradley Wright, Mark Walters, Sean Doyle, Adam McCarthy, Renata Rocha De Almeida, Katherine Andriole, et al. Detection and delineation of acute cerebral infarct on DWI using weakly supervised machine learning. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 81–88. Springer, 2018. [35](#)
- [65] Fernando Calamante, David G Gadian, and Alan Connelly. Delay and dispersion effects in dynamic susceptibility contrast MRI : Simulations using singular value decomposition. *Magnetic Resonance in Medicine*, 44(3) :466–473, 2000. [35](#), [36](#)
- [66] Mark T Madsen. A simplified formulation of the gamma variate function. *Physics in Medicine & Biology*, 37(7) :1597, 1992. [35](#), [49](#)

- [67] A Waaijer, IC Van der Schaaf, BK Velthuis, M Quist, MJP Van Osch, EPA Vonken, MS Van Leeuwen, and M Prokop. Reproducibility of quantitative CT brain perfusion measurements in patients with symptomatic unilateral carotid artery stenosis. *American journal of neuroradiology*, 28(5) :927–932, 2007. [36](#)
- [68] Elias Kellner, Irina Mader, Michael Mix, Daniel Nico Splitthoff, Marco Reisert, Katharina Foerster, Thao Nguyen-Thanh, Peter Gall, and Valerij G Kiselev. Arterial input function measurements for bolus tracking perfusion imaging in the brain. *Magnetic resonance in medicine*, 69(3) :771–780, 2013. [37](#)
- [69] Carole Frindel, Anaïs Rouanet, Mathilde Giacalone, Tae-Hee Cho, Leif Østergaard, Jens Fiehler, Salvador Pedraza, Jean-Claude Baron, Marlène Wiart, Yves Berthezène, et al. Validity of shape as a predictive biomarker of final infarct volume in acute ischemic stroke. *Stroke*, pages STROKEAHA–114, 2015. [38](#)
- [70] Leif Østergaard, Robert M Weisskoff, David A Chesler, Carsten Gyldensted, and Bruce R Rosen. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I : Mathematical approach and statistical analysis. *Magnetic Resonance in Medicine*, 36(5) :715–725, 1996. [39](#)
- [71] Arno Villringer, Bruce R Rosen, John W Belliveau, Jerome L Ackerman, Randall B Lauffer, Richard B Buxton, Yong-Sheng Chao, Van J Wedeenand, and Thomas J Brady. Dynamic imaging with lanthanide chelates in normal brain : Contrast due to magnetic susceptibility effects. *Magnetic Resonance in Medicine*, 6(2) :164–174, 1988. [39](#)
- [72] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision – ECCV 2014*, pages 818–833. Springer, 2014. [39](#)
- [73] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, pages 1097–1105, 2012. [39](#)
- [74] Kevin Jarrett, Koray Kavukcuoglu, Yann LeCun, et al. What is the best multi-stage architecture for object recognition? In *2009 IEEE 12th International Conference on Computer Vision*, pages 2146–2153. IEEE, 2009. [39](#)
- [75] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout : A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1) :1929–1958, 2014. [40](#), [93](#)
- [76] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv :1207.0580*, 2012. [40](#)
- [77] Jean-Marc Olivot, Michael Mlynash, Vincent N Thijs, Stephanie Kemp, Maarten G Lansberg, Lawrence Wechsler, Roland Bammer, Michael P Marks, and Gregory W Albers. Optimal tmax threshold for predicting penumbral tissue in acute stroke. *Stroke*, 40(2) :469–475, 2009. [40](#), [66](#)
- [78] Michelle Livne, Vince I Madai, Peter Brunecker, Olivier Zaro-Weber, Walter Moeller-Hartmann, Wolf-Dieter Heiss, Kim Mouridsen, and Jan Sobesky. A PET-guided framework supports a multiple arterial input functions approach in DSC-MRI in acute stroke. *Journal of Neuroimaging*, 27(5) :486–492, 2017. [46](#), [48](#)
- [79] Noëlie Debs, Pejman Rasti, Léon Victor, Tae-Hee Cho, Carole Frindel, and David Rousseau. Simulated perfusion mri data to boost training of convolutional neural networks for lesion fate prediction in acute stroke. *Computers in Biology and Medicine*, 116 :103579, 2020. [48](#)

- [80] Leonidas Georgiou, Daniel J Wilson, Nisha Sharma, Timothy J Perren, and David L Buckley. A functional form for a representative individual arterial input function measured from a population using high temporal resolution DCE MRI. *Magn Reson Med*, 81(3) :1955–1963, 2019. [48](#)
- [81] Fernando Calamante, Peter J Yim, and Juan R Cebal. Estimation of bolus dispersion effects in perfusion MRI using image-based computational fluid dynamics. *Neuroimage*, 19(2) :341–353, 2003. [49](#)
- [82] Søren Christensen, Fernando Calamante, Niels Hjort, Ona Wu, Anne Dorte Blankholm, Patricia Desmond, Stephen Davis, and Leif Østergaard. Inferring origin of vascular supply from tracer arrival timing patterns using bolus tracking MRI. *JMRI*, 27(6) :1371–1381, 2008. [49](#), [50](#)
- [83] Stephen E Rose, Andrew L Janke, Mark Griffin, Simon Finnigan, and Jonathan B Chalk. Improved Prediction of Final Infarct Volume Using Bolus Delay–Corrected Perfusion-Weighted MRI : Implications for the Ischemic Penumbra. *Stroke*, 35(11) :2466–2471, 2004. [49](#)
- [84] Wu Qiu, Hulin Kuang, Ericka Teleg, Johanna M. Ospel, Sung Il Sohn, Mohammed Almekhlafi, Mayank Goyal, Michael D. Hill, Andrew M. Demchuk, and Bijoy K. Menon. Machine learning for detecting early infarction in acute stroke with non-contrast-enhanced CT. *Radiology*, page 191193, 2020. [61](#)
- [85] Yannan Yu, Yuan Xie, Thoralf Thamm, Enhao Gong, Jiahong Ouyang, Charles Huang, Soren Christensen, Michael P Marks, Maarten G Lansberg, Gregory W Albers, et al. Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. *JAMA Network Open*, 3(3) :e200772–e200772, 2020. [61](#), [63](#), [65](#), [72](#)
- [86] Tae-Hee Cho, Norbert Nighoghossian, Irene Klærke Mikkelsen, Laurent Derex, Marc Hermier, Salvador Pedraza, Jens Fiehler, Leif Østergaard, Yves Berthezène, and Jean-Claude Baron. Reperfusion within 6 hours outperforms recanalization in predicting penumbra salvage, lesion growth, final infarct, and clinical outcome. *Stroke*, 46(6) :1582–1589, 2015. [61](#), [72](#)
- [87] Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *NeuroImage*, 54(3) :2033–2044, 2011. [62](#), [75](#), [118](#)
- [88] Stephen Smith, Peter R Bannister, Christian Beckmann, Mike Brady, Stuart Clare, David Flitney, Peter Hansen, Mark Jenkinson, Didier Lebovici, Brian Ripley, et al. FSL : New tools for functional and structural brain image analysis. *NeuroImage*, 13(6) :249, 2001. [62](#)
- [89] Osama O Zaidat, Albert J Yoo, Pooja Khatri, Thomas A Tomsick, Rüdiger Von Kummer, Jeffrey L Saver, Michael P Marks, Shyam Prabhakaran, David F Kallmes, Brian-Fred M Fitzsimmons, et al. Recommendations on angiographic revascularization grading standards for acute ischemic stroke : a consensus statement. *Stroke*, 44(9) :2650–2663, 2013. [62](#)
- [90] Mayank Goyal, Bijoy K Menon, Wim H van Zwam, Diederik WJ Dippel, Peter J Mitchell, Andrew M Demchuk, Antoni Dávalos, Charles BLM Majoie, Aad van der Lugt, Maria A De Miquel, et al. Endovascular thrombectomy after large-vessel ischaemic stroke : a meta-analysis of individual patient data from five randomised trials. *The Lancet*, 387(10029) :1723–1731, 2016. [63](#)
- [91] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. [63](#)

- [92] Mehmet Aygün, Yusuf Hüseyin Şahin, and Gözde Ünal. Multi modal convolutional neural networks for brain tumor segmentation. *arXiv preprint arXiv :1809.06191*, 2018. [63](#)
- [93] Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. IVD-Net : Intervertebral disc localization and segmentation in MRI with a multi-modal UNet. In *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*, pages 130–143. Springer, 2018. [63](#)
- [94] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net : Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016. [64](#)
- [95] Karimollah Hajian-Tilaki. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine*, 4(2) :627, 2013. [65](#)
- [96] Kristjana Yr Jonsdottir, Leif Østergaard, and Kim Mouridsen. Predicting tissue outcome from acute stroke magnetic resonance imaging : improving model performance by optimal sampling of training data. *Stroke*, 40(9) :3006–3011, 2009. [65](#), [66](#)
- [97] Anthony J Winder, Susanne Siemonsen, Fabian Flottmann, Götz Thomalla, Jens Fiehler, and Nils D Forkert. Technical considerations of multi-parametric tissue outcome prediction methods in acute ischemic stroke patients. *Scientific reports*, 9(1) :1–12, 2019. [65](#)
- [98] Kristina Dupont Hougaard, Niels Hjort, Dora Zeidler, Leif Sørensen, Anne Nørgaard, Troels Martin Hansen, Paul von Weitzel-Mudersbach, Claus Z Simonsen, Dorte Damgaard, Hanne Gottrup, Kristina Svendsen, Peter Vestergaard Rasmussen, Lars R Ribe, Irene K Mikkelsen, Kartheban Nagenthiraja, Tae-Hee Cho, Andrew N Redington, Hans Erik Bøtker, Leif Østergaard, Kim Mouridsen, and Grethe Andersen. Remote ischemic preconditioning as an adjunct therapy to thrombolysis in patients with acute ischemic stroke : a randomized trial. *Stroke*, 45(1) :159–167, 2013. [72](#)
- [99] Joonsang Yoo, Jin Wook Choi, Seong-Joon Lee, Ji Man Hong, Jeong-Ho Hong, Chang-Hyun Kim, Yong-Won Kim, Dong-Hun Kang, Yong-Sun Kim, Yang-Ha Hwang, Bruce Ovbiagele, Andrew M. Demchuk, Jin Soo Lee, and Sung-Il Sohn. Ischemic diffusion lesion reversal after endovascular treatment. *Stroke*, 50(6) :1504–1509, 2019. [72](#)
- [100] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning : a brief survey. *IEEE Signal Processing Magazine*, 34(6) :26–38, 2017. [73](#)
- [101] Søren Christensen, Kim Mouridsen, Ona Wu, Niels Hjort, Henrik Karstoft, Götz Thomalla, Joachim Röther, Jens Fiehler, Thomas Kucinski, and Leif Østergaard. Comparison of 10 perfusion MRI parameters in 97 sub-6-hour stroke patients using voxel-based receiver operating characteristics analysis. *Stroke*, 40(6) :2055–2061, 2009. [73](#)
- [102] Hayley M Wheeler, Michael Mlynash, Manabu Inoue, Aaryani Tipirneni, John Liggins, Greg Zaharchuk, Matus Straka, Stephanie Kemp, Roland Bammer, Maarten G Lansberg, et al. Early diffusion-weighted imaging and perfusion-weighted imaging lesion volumes forecast final infarct size in DEFUSE 2. *Stroke*, 44(3) :681–685, 2013. [73](#)
- [103] Yilong Chen, Kai Wang, Xiangyun Liao, Yinling Qian, Qiong Wang, Zhiyong Yuan, and Pheng-Ann Heng. Channel-UNet : a spatial channel-wise convolutional neural network for liver and tumors segmentation. *Frontiers in Genetics*, 10, 2019. [73](#)
- [104] Nandakumar Nagaraja, John R Forder, Steven Warach, and José G Merino. Reversible diffusion-weighted imaging lesions in acute ischemic stroke : A systematic review. *Neurology*, 94(13) :571–587, 2020. [75](#), [82](#)

- [105] Fredrik N Albach, Peter Brunecker, Tatiana Usnich, Kersten Villringer, Martin Ebinger, Jochen B Fiebach, and Christian H Nolte. Complete early reversal of diffusion-weighted imaging hyperintensities after ischemic stroke is mainly limited to small embolic lesions. *Stroke*, 44(4) :1043–1048, 2013. [76](#)
- [106] Joonsang Yoo, Jin Wook Choi, Seong-Joon Lee, Ji Man Hong, Jeong-Ho Hong, Chang-Hyun Kim, Yong-Won Kim, Dong-Hun Kang, Yong-Sun Kim, Yang-Ha Hwang, et al. Ischemic diffusion lesion reversal after endovascular treatment : Prevalence, prognosis, and predictors. *Stroke*, 50(6) :1504–1509, 2019. [76](#), [82](#)
- [107] Sebastien Soize, Marie Tisserand, Sylvain Charron, Guillaume Turc, Wagih Ben Hassen, Marc-Antoine Labeyrie, Laurence Legrand, Jean-Louis Mas, Laurent Pierot, Jean-François Meder, et al. How sustained is 24-hour diffusion-weighted imaging lesion reversal? serial magnetic resonance imaging in a patient cohort thrombolized within 4.5 hours of stroke onset. *Stroke*, 46(3) :704–710, 2015. [76](#), [82](#)
- [108] Marc-Antoine Labeyrie, Guillaume Turc, Agathe Hess, Patrice Hervo, Jean-Louis Mas, Jean-François Meder, Jean-Claude Baron, Emmanuel Touzé, and Catherine Oppenheim. Diffusion lesion reversal after thrombolysis : a mr correlate of early neurological improvement. *Stroke*, 43(11) :2986–2991, 2012. [76](#), [82](#)
- [109] Kyung-Hee Cho, Sun U Kwon, Deok Hee Lee, WooHyun Shim, ChoongGon Choi, Sang Joon Kim, Dae-Chul Suh, Jong S Kim, and Dong-Wha Kang. Early infarct growth predicts long-term clinical outcome after thrombolysis. *Journal of the neurological sciences*, 316(1-2) :99–103, 2012. [76](#)
- [110] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1) :2522–5839, 2020. [78](#), [83](#)
- [111] Bruce CV Campbell, Archana Purushotham, Soren Christensen, Patricia M Desmond, Yoshinari Nagakane, Mark W Parsons, Maarten G Lansberg, Michael Mlynash, Matus Straka, Deidre A De Silva, et al. The infarct core is well represented by the acute diffusion lesion : sustained reversal is infrequent. *Journal of Cerebral Blood Flow & Metabolism*, 32(1) :50–56, 2012. [82](#)
- [112] Thomas Chemmanam, Bruce CV Campbell, S Christensen, Y Nagakane, Patricia M Desmond, Christopher F Bladin, Mark W Parsons, Christopher R Levi, Peter Alan Barber, GA Donnan, et al. Ischemic diffusion lesion reversal is uncommon and rarely alters perfusion-diffusion mismatch. *Neurology*, 75(12) :1040–1047, 2010. [82](#)
- [113] Jean-Marc Olivot, Michael Mlynash, Vincent N Thijs, Archana Purushotham, Stephanie Kemp, Maarten G Lansberg, Lawrence Wechsler, Roland Bammer, Michael P Marks, and Gregory W Albers. Relationships between cerebral perfusion and reversibility of acute diffusion lesions in DEFUSE : insights from RADAR. *Stroke*, 40(5) :1692–1697, 2009. [82](#)
- [114] Yuki Sakamoto, Kazumi Kimura, Kensaku Shibazaki, Takeshi Inoue, Jyunichi Uemura, Junya Aoki, Kenichiro Sakai, and Yasuyuki Iguchi. Early ischaemic diffusion lesion reduction in patients treated with intravenous tissue plasminogen activator : infrequent, but significantly associated with recanalization. *International Journal of Stroke*, 8(5) :321–326, 2013. [82](#)
- [115] Jens Fiehler, Karina Knudsen, Thomas Kucinski, Chelsea S Kidwell, Jeffrey R Alger, Gotz Thomalla, Bernd Eckert, Oliver Wittkugel, Cornelius Weiller, Hermann Zeumer, et al. Predictors of apparent diffusion coefficient normalization in stroke patients. *Stroke*, 35(2) :514–519, 2004. [82](#)

- [116] Alexis Nétis Simpkins, Christian Dias, Gina Norato, Eunhee Kim, Richard Leigh, NIH Natural History of Stroke Investigators, et al. Early change in stroke size performs best in predicting response to therapy. *Cerebrovascular Diseases*, 44(3-4) :141–149, 2017. [82](#)
- [117] Kenji Suzuki. Overview of deep learning in medical imaging. *Radiological physics and technology*, 10(3) :257–273, 2017. [83](#)
- [118] John O’donoghue and John Herbert. Data management within mhealth environments : Patient sensors, mobile devices, and databases. *Journal of Data and Information Quality (JDIQ)*, 4(1) :1–20, 2012. [85](#)
- [119] Peter Düking, Andreas Hotho, Hans-Christer Holmberg, Franz Konstantin Fuss, and Billy Sperlich. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. *Frontiers in physiology*, 7 :71, 2016. [85](#)
- [120] Peter Düking, Silvia Achtzehn, Hans-Christer Holmberg, and Billy Sperlich. Integrated framework of load monitoring by a combination of smartphone applications, wearables and point-of-care testing provides feedback that allows individual responsive adjustments to activities of daily living. *Sensors*, 18(5) :1632, 2018. [85](#)
- [121] Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, and Mary Rodgers. A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation*, 9(1) :1–17, 2012. [85](#)
- [122] Oliver Brand. Microsensor integration into systems-on-chip. *Proceedings of the IEEE*, 94(6) :1160–1176, 2006. [86](#)
- [123] Bojan Milosevic and Elisabetta Farella. Wireless mems for wearable sensor networks. In *Wireless MEMS Networks and Applications*, pages 101–127. Elsevier, 2017. [86](#)
- [124] Jinyun Zhang, Philip V Orlik, Zafer Sahinoglu, Andreas F Molisch, and Patrick Kinney. Uwb systems for wireless sensor networks. *Proceedings of the IEEE*, 97(2) :313–331, 2009. [86](#)
- [125] Roy Want. iphone : smarter than the average phone. *IEEE Pervasive Computing*, 9(3) :6–9, 2010. [86](#)
- [126] Nathan Botts, Brian Thoms, Aisha Noamani, and Thomas A Horan. Cloud computing architectures for the underserved : Public health cyberinfrastructures through a network of healthatms. In *2010 43rd Hawaii international conference on system sciences*, pages 1–10. IEEE, 2010. [86](#)
- [127] Henry H Chang, Paul B Chou, and Sreeram Ramakrishnan. An ecosystem approach for healthcare services cloud. In *2009 IEEE international conference on e-business engineering*, pages 608–612. IEEE, 2009. [86](#)
- [128] Moeen Hassanali, Alex Page, Tolga Soyata, Gaurav Sharma, Mehmet Aktas, Gonzalo Mateos, Burak Kantarci, and Silvana Andreescu. Health monitoring and management using internet-of-things (iot) sensing with cloud-based processing : Opportunities and challenges. In *2015 IEEE International Conference on Services Computing*, pages 285–292. IEEE, 2015. [86](#)
- [129] Andrey Ignatov, Radu Timofte, William Chou, Ke Wang, Max Wu, Tim Hartley, and Luc Van Gool. Ai benchmark : Running deep neural networks on android smartphones. In *Proceedings of the European conference on computer vision (ECCV)*, pages 0–0, 2018. [86](#)
- [130] James AC Patterson, Douglas C McIlwraith, and Guang-Zhong Yang. A flexible, low noise reflective ppg sensor platform for ear-worn heart rate monitoring. In *2009 sixth international workshop on wearable and implantable body sensor networks*, pages 286–291. IEEE, 2009. [87](#)

- [131] Phil Corbishley and Esther Rodriguez-Villegas. Breathing detection : towards a miniaturized, wearable, battery-operated monitoring system. *IEEE Transactions on Biomedical Engineering*, 55(1) :196–204, 2007. [87](#)
- [132] H Harry Asada, Phillip Shaltis, Andrew Reisner, Sokwoo Rhee, and Reginald C Hutchinson. Mobile monitoring with wearable photoplethysmographic biosensors. *IEEE engineering in medicine and biology magazine*, 22(3) :28–40, 2003. [87](#)
- [133] Reiner Wichert, David V Keyson, Panos Markopoulos, Norbert Streitz, Monica Divitini, Nikolaos Georgantas, and Antonio Mana Gomez. *Ambient intelligence*. Springer, 2010. [87](#)
- [134] Dimitar H Stefanov, Zeungnam Bien, and Won-Chul Bang. The smart house for older persons and persons with physical disabilities : structure, technology arrangements, and perspectives. *IEEE transactions on neural systems and rehabilitation engineering*, 12(2) :228–250, 2004. [87](#)
- [135] Carole Frindel and David Rousseau. How accurate are smartphone accelerometers to identify intermittent claudication? In *International Conference on IoT Technologies for Health-Care*, pages 19–25. Springer, 2017. [87](#)
- [136] Merryn J Mathie, Adelle CF Coster, Nigel H Lovell, Branko G Celler, Stephen R Lord, and Anne Tiedemann. A pilot study of long-term monitoring of human movements in the home using accelerometry. *Journal of telemedicine and telecare*, 10(3) :144–151, 2004. [87](#)
- [137] Edward S Sazonov, George Fulk, Nadezhda Sazonova, and Stephanie Schuckers. Automatic recognition of postures and activities in stroke patients. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2200–2203. IEEE, 2009. [87](#)
- [138] Ramon Vilallonga, Albert Lecube, José Manuel Fort, Maria Angeles Boleko, Marta Hidalgo, and Manel Armengol. Internet of things and bariatric surgery follow-up : Comparative study of standard and iot follow-up. *Minimally Invasive Therapy & Allied Technologies*, 22(5) :304–311, 2013. [87](#)
- [139] Aung Pyae, Mika Luimula, and Jouni Smed. Understanding stroke patients’ motivation for motivation-driven rehabilitative game design. In *International Internet of Things Summit*, pages 99–111. Springer, 2014. [87](#)
- [140] Ali Al-Mahmood and Michael Opoku Agyeman. On wearable devices for motivating patients with upper limb disability via gaming and home rehabilitation. In *2018 Third International Conference on Fog and Mobile Edge Computing (FMEC)*, pages 155–162. IEEE, 2018. [87](#)
- [141] Tuan Nguyen Gia, Igor Tcareno, Victor K Sarker, Amir M Rahmani, Tomi Westerlund, Pasi Liljeberg, and Hannu Tenhunen. Iot-based fall detection system with energy efficient sensor nodes. In *2016 IEEE Nordic Circuits and Systems Conference (NORCAS)*, pages 1–6. IEEE, 2016. [87](#)
- [142] Tuan Nguyen Gia, Victor Kathan Sarker, Igor Tcareno, Amir M Rahmani, Tomi Westerlund, Pasi Liljeberg, and Hannu Tenhunen. Energy efficient wearable sensor node for iot-based fall detection systems. *Microprocessors and Microsystems*, 56 :34–46, 2018. [87](#)
- [143] Stefano Abbate, Marco Avvenuti, Francesco Bonatesta, Guglielmo Cola, Paolo Corsini, and Alessio Vecchio. A smartphone-based fall detection system. *Pervasive and Mobile Computing*, 8(6) :883–899, 2012. [87](#)
- [144] Bruno Aguiar, Tiago Rocha, Joana Silva, and Ines Sousa. Accelerometer-based fall detection for smartphones. In *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 2014. [87](#)

- [145] Marc Bachlin, Meir Plotnik, Daniel Roggen, Inbal Maidan, Jeffrey M Hausdorff, Nir Giladi, and Gerhard Troster. Wearable assistant for parkinson's disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine*, 14(2) :436–446, 2009. [87](#)
- [146] Frank Sposaro, Justin Danielson, and Gary Tyson. iwander : An android application for dementia patients. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 3875–3878. IEEE, 2010. [87](#)
- [147] Paula M Vergara, Enrique de la Cal, José R Villar, Víctor M González, and Javier Sedano. An iot platform for epilepsy monitoring and supervising. *Journal of Sensors*, 2017, 2017. [87](#)
- [148] Richard D Willmann, Gerd Lanfermann, Privender Saini, Annick Timmermans, Jurgen te Vrugt, and Stefan Winter. Home stroke rehabilitation for the upper limbs. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4015–4018. IEEE, 2007. [87](#)
- [149] Geng Yang, Jia Deng, Gaoyang Pang, Hao Zhang, Jiayi Li, Bin Deng, Zhibo Pang, Juan Xu, Mingzhe Jiang, Pasi Liljeberg, et al. An iot-enabled stroke rehabilitation system based on smart wearable armband and machine learning. *IEEE journal of translational engineering in health and medicine*, 6 :1–10, 2018. [87](#), [88](#)
- [150] Claude Vincent, Isabelle Deaudelin, Line Robichaud, Jacqueline Rousseau, Chantal Viscoqliosi, Lise R Talbot, and Johanne Desrosiers. Rehabilitation needs for older adults with stroke living at home : perceptions of four populations. *BMC geriatrics*, 7(1) :20, 2007. [88](#)
- [151] N Kostikis, Dimitris Hristu-Varsakelis, M Arnaoutoglou, and C Kotsavasiloglou. A smartphone-based tool for assessing parkinsonian hand tremor. *IEEE journal of biomedical and health informatics*, 19(6) :1835–1842, 2015. [88](#)
- [152] Florian Lipsmeier, Kirsten I Taylor, Timothy Kilchenmann, Detlef Wolf, Alf Scotland, Jens Schjodt-Eriksen, Wei-Yi Cheng, Ignacio Fernandez-Garcia, Juliane Siebourg-Polster, Liping Jin, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 parkinson's disease clinical trial. *Movement Disorders*, 33(8) :1287–1297, 2018. [88](#)
- [153] Carlos Ferreira, Vânia Guimarães, António Santos, and Inês Sousa. Gamification of stroke rehabilitation exercises using a smartphone. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 282–285, 2014. [88](#)
- [154] Paolo Bonato. Advances in wearable technology and its medical applications. In *2010 annual international conference of the IEEE engineering in medicine and biology*, pages 2021–2024. IEEE, 2010. [88](#)
- [155] Lynne Lamberg. Confidentiality and privacy of electronic medical records. *Jama*, 285(24) :3075–3076, 2001. [88](#)
- [156] Pierre Gard, Lucie Lalanne, Alexandre Ambourg, David Rousseau, François Lesueur, and Carole Frindel. A secured smartphone-based architecture for prolonged monitoring of neurological gait. In *International Conference on IoT Technologies for HealthCare*, pages 3–9. Springer, 2017. [88](#)
- [157] Théo Jourdan, Antoine Boutet, and Carole Frindel. Toward privacy in iot mobile devices for activity recognition. In *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems : Computing, Networking and Services*, pages 155–165, 2018. [89](#), [92](#), [94](#)

- [158] Antoine Boutet, Carole Frindel, Sébastien Gambs, Théo Jourdan, and Claude Rosin Ngueveu. Dysan : Dynamically sanitizing motion sensor data against sensitive inferences through adversarial networks. *arXiv preprint arXiv :2003.10325*, 2020. [89](#)
- [159] Berna Seref and Erkan Bostanci. Opportunities, threats and future directions in big data for medical wearables. In *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, pages 1–5, 2016. [89](#)
- [160] Jorge Luis Reyes Ortiz. *Smartphone-based human activity recognition*. Springer, 2015. [90](#)
- [161] Chiraz BenAbdelkader, Ross Cutler, and Larry Davis. Stride and cadence as a biometric in automatic person identification and verification. In *Proceedings of Fifth IEEE international conference on automatic face gesture recognition*, pages 372–377. IEEE, 2002. [90](#)
- [162] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity recognition from accelerometer data. In *Aaai*, volume 5, pages 1541–1546, 2005. [90](#)
- [163] Dennis Gabor. Theory of communication. part 1 : The analysis of information. *Journal of the Institution of Electrical Engineers-Part III : Radio and Communication Engineering*, 93(26) :429–441, 1946. [90](#)
- [164] Robert Glenn Stockwell, Lalu Mansinha, and RP Lowe. Localization of the complex spectrum : the s transform. *IEEE transactions on signal processing*, 44(4) :998–1001, 1996. [90](#)
- [165] Ali Moukadem, Zied Bouguila, Djaffar Ould Abdeslam, and Alain Dieterlen. A new optimized stockwell transform applied on synthetic and real non-stationary signals. *Digital Signal Processing*, 46 :226–238, 2015. [90](#), [91](#)
- [166] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999. [91](#)
- [167] Said Assous and Boualem Boashash. Evaluation of the modified s-transform for time-frequency synchrony analysis and source localisation. *EURASIP Journal on Advances in Signal Processing*, 2012(1) :49, 2012. [91](#)
- [168] Ervin Sejdić, Igor Djurović, and Jin Jiang. A window width optimized s-transform. *EURASIP Journal on Advances in Signal Processing*, 2008(1) :672941, 2007. [91](#)
- [169] Ljubiša Stanković. A measure of some time–frequency distributions concentration. *Signal Processing*, 81(3) :621–631, 2001. [91](#)
- [170] Pim Moeskops, Max A Viergever, Adriënne M Mendrik, Linda S De Vries, Manon JNL Benders, and Ivana Išgum. Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5) :1252–1261, 2016. [93](#)
- [171] Patrick Flandrin. Time–frequency filtering based on spectrogram zeros. *IEEE Signal Processing Letters*, 22(11) :2137–2141, 2015. [97](#), [101](#)
- [172] Suk Jae Kim, Jeong Pyo Son, Sookyung Ryoo, Mi-Ji Lee, Jihoon Cha, Keon Ha Kim, Gyeong-Moon Kim, Chin-Sang Chung, Kwang Ho Lee, Pyoung Jeon, et al. A novel magnetic resonance imaging approach to collateral flow imaging in ischemic stroke. *Annals of neurology*, 76(3) :356–369, 2014. [100](#)
- [173] Keith W Muir, Pippa Tyrrell, Naveed Sattar, and Elizabeth Warburton. Inflammation and ischaemic stroke. *Current opinion in neurology*, 20(3) :334–342, 2007. [100](#)
- [174] David S Liebeskind. Artificial intelligence in stroke care : Deep learning or superficial insight? *EBioMedicine*, 35 :14–15, 2018. [100](#)

- [175] Adriano Pinto, Richard Mckinley, Victor Alves, Roland Wiest, Carlos A Silva, and Mauricio Reyes. Stroke lesion outcome prediction based on mri imaging combined with clinical information. *Frontiers in neurology*, 9 :1060, 2018. [100](#)
- [176] Youngwon Choi, Yongchan Kwon, Hanbyul Lee, Beom Joon Kim, Myunghee Cho Paik, and Joong-Ho Won. Ensemble of deep convolutional neural networks for prognosis of ischemic stroke. In *International Workshop on Brainlesion : Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 231–243. Springer, 2016. [100](#)
- [177] Nitish Srivastava and Russ R Salakhutdinov. Multimodal learning with deep boltzmann machines. In *Advances in neural information processing systems*, pages 2222–2230, 2012. [100](#)
- [178] Hoel Kervadec, Jose Dolz, Meng Tang, Eric Granger, Yuri Boykov, and Ismail Ben Ayed. Constrained-cnn losses for weakly supervised segmentation. *Medical image analysis*, 54 :88–99, 2019. [100](#)
- [179] Alexander de Brebisson and Giovanni Montana. Deep neural networks for anatomical brain segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 20–28, 2015. [100](#)
- [180] Clinton J Wang, Charlie A Hamm, Lynn J Savic, Marc Ferrante, Isabel Schobert, Todd Schlachter, MingDe Lin, Jeffrey C Weinreb, James S Duncan, Julius Chapiro, et al. Deep learning for liver tumor diagnosis part ii : convolutional neural network interpretation using radiologic imaging features. *European radiology*, 29(7) :3348–3357, 2019. [101](#)
- [181] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through propagating activation differences. *arXiv preprint arXiv :1704.02685*, 2017. [101](#)
- [182] Laurens Van Der Maaten. Accelerating t-sne using tree-based algorithms. *The Journal of Machine Learning Research*, 15(1) :3221–3245, 2014. [101](#)
- [183] Shi-hui Zhen, Ming Cheng, Yu-bo Tao, Yi-fan Wang, Sarun Juengpanich, Zhi-yu Jiang, Yan-kai Jiang, Yu-yu Yan, Wei Lu, Jie-min Lue, et al. Deep learning for accurate diagnosis of liver tumor based on magnetic resonance imaging and clinical data. *Frontiers in oncology*, 10 :680, 2020. [101](#)

Partie V
Annexes

Annexe A

Présentation des cohortes analysées et des pré-traitements associés

Dans ces travaux, nous avons eu accès à des patients aux bases de données inclus dans les cohortes I-KNOW et HIBISCUS-STROKE. I-KNOW (2007-2011) était une étude observationnelle multicentrique prospective de 117 patients ayant subi un AVC de taille très variables [48]. Sur l'ensemble des patients, 76 ont été traités par thrombolyse et les 46 restant n'ont reçu aucun traitement. HIBISCUS-STROKE est une cohorte d'observation monocentrique en cours incluant 135 patients présentant de grandes occlusions de l'artère intracrânienne, tous traités par thrombectomie. La base de donnée I-KNOW est donc bien plus hétérogène que HIBISCUS-STROKE en terme de type d'occlusion, de forme de lésions aiguës, de traitement et par conséquent d'évolution de l'infarctus final. L'inhomogénéité des variables cliniques dans les bases d'apprentissage (train/test) a un impact considérable sur les modèles d'apprentissage automatique. Ce point est l'axe central de discussion du manuscrit.

Dans les deux cohortes, tous les patients ont subi le protocole IRM suivant à leur admission : imagerie par diffusion, imagerie FLAIR, imagerie T2*, angiographie et imagerie dynamique de perfusion. Une IRM FLAIR de suivi a été effectuée plusieurs jours après l'admission (spécifiquement, 6 et 30 jours à HIBISCUS-STROKE et I-KNOW, respectivement), et a été segmentée semi-manuellement par un ou plusieurs experts (spécifiquement, un et trois experts pour HIBISCUS-STROKE et I-KNOW, respectivement).

Concernant la cohorte I-KNOW, aucun pré-traitement spécifique n'a été nécessaire. Un recalage affine avait été préalablement réalisé afin de recalibrer toutes les modalités de chaque patient sur l'imagerie de perfusion. En revanche, la cohorte HIBISCUS-STROKE n'était pas pré-traitée, ce qui nous imposait un effort de recalage. Dans le protocole I-KNOW, le FLAIR de suivi est réalisé 30 jours après le premier examen d'admission : la lésion finale est alors stabilisée et aucun oedème n'est visible. En revanche, dans le protocole HIBISCUS-STROKE le FLAIR de suivi est réalisé à seulement 6 jours de l'admission : des oedèmes cérébraux importants y sont encore visibles, se caractérisant par un gonflement de la lésion dans le système ventriculaire, comme le montre la Figure A.1. Un travail de recalage non-linéaire [87] pour la cohorte HIBISCUS-STROKE s'est donc imposé afin de tenir compte de telles déformations.

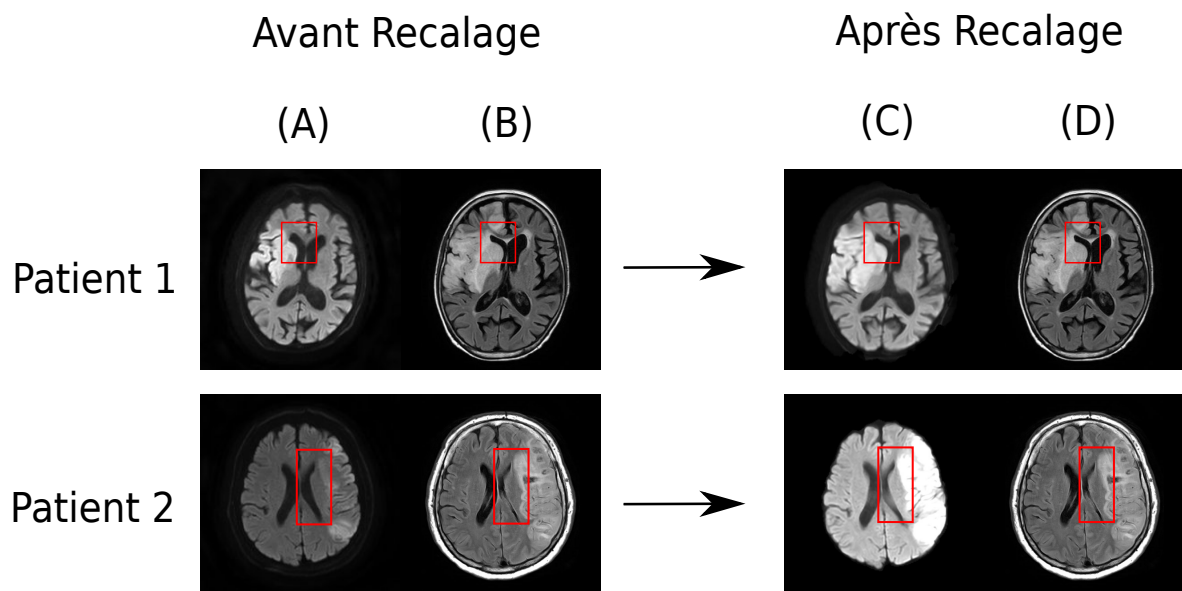


FIGURE A.1 – Exemple de recalage non-linéaire effectué sur deux patients de la cohorte HIBISCUS-STROKE. L'IRM de diffusion (colonne A) et l'IRM de suivi FLAIR (colonne B/D) sont représentées dans leur espace initial. Après recalage, l'IRM de diffusion (colonne C) est dans l'espace initial de l'IRM FLAIR.

Annexe B

Matériel supplémentaire pour l'apprentissage de réseaux de neurones sur données de perfusion brutes simulées

B.0.1 Métriques supplémentaires

Concernant les paramètres présentés dans le Tableau II.5.4 de la section II.5, nous présentons dans le Tableau B.1 davantage de métriques afin de démontrer la valeur du CNN testé dans les expériences 1, 2 et 3. Nous pouvons noter que quelle que soit la métrique, l'expérience 3 (*i.e.* entraînement sur des données de perfusion simulées) présente systématiquement les meilleures performances moyennes de classification.

B.0.2 Architectures supplémentaires

Concernant notre proposition d'architecture présentée dans II.5.3 de la section II.5, nous présentons des architectures de réseau convolutif supplémentaires (FCN1, FCN2 et FCN3) décrites dans les Tableaux B.2, B.3, B.4. Les expériences 1, 2 et 3 ont été réalisées en utilisant ces trois architectures. Pour toutes les expériences, nous avons utilisé la fonction d'entropie croisée catégorique comme fonction de perte et une descente de gradient stochastique pour optimiser le modèle. Le nombre d'époques a été fixé à 30, le taux d'apprentissage à 0.001 et la taille du batch à 32. Les performances de classification correspondantes sont indiquées dans le Tableau B.5.

Nous pouvons noter que pour toutes les architectures testées, l'expérience 3 présente systématiquement la meilleure performance de classification. Les valeurs de distance DSC et Hausdorff obtenues avec le réseau neuronal convolutif (CNN) proposé ne sont pas significativement différentes des réseaux entièrement convolutifs présentés (architectures FCN1, FCN2 et FCN3) (valeur $p > 0,05$ avec le test de Mann-Whitney pour données appariées).

Patient	Expérience 1				Expérience 2				Expérience 3			
	MCC	AUC	Recall	Precision	MCC	AUC	Recall	Precision	MCC	AUC	Recall	Precision
1	0.12 ± 0.079	0.77 ± 0.055	0.86 ± 0.061	0.10 ± 0.060	0.0 ± 0.0	0.55 ± 0.076	1.0 ± 0.0	0.12 ± 0.0	0.040 ± 0.0039	0.57 ± 0.16	0.15 ± 0.17	0.11 ± 0.0096
2	-0.039 ± 0.043	0.77 ± 0.028	0.98 ± 0.0050	0.10 ± 0.0008	0.0 ± 0.0	0.51 ± 0.031	1.0 ± 0.0	0.10 ± 0.0	0.58 ± 0.0056	0.89 ± 0.0037	0.65 ± 0.21	0.48 ± 0.15
3	-0.013 ± 0.018	0.35 ± 0.020	0.98 ± 0.041	0.037 ± 0.026	0.0 ± 0.0	0.50 ± 0.0027	1.0 ± 0.0	0.022 ± 0.0	0.31 ± 0.0076	0.72 ± 0.0075	0.26 ± 0.017	0.20 ± 0.0065
4	0.46 ± 0.030	0.94 ± 0.0025	0.94 ± 0.021	0.30 ± 0.034	0.016 ± 0.00042	0.62 ± 0.036	1.0 ± 0.0	0.091 ± 1.29e-05	0.47 ± 0.0080	0.93 ± 0.0039	0.90 ± 0.0024	0.25 ± 0.0089
5	0.35 ± 0.043	0.87 ± 0.17	0.91 ± 0.022	0.19 ± 0.037	0.0019 ± 0.0054	0.57 ± 0.048	1.0 ± 0.0	0.057 ± 0.0	0.47 ± 0.0044	0.91 ± 0.0080	0.78 ± 0.005	0.34 ± 0.0031
6	0.050 ± 0.13	0.45 ± 0.054	0.99 ± 0.0003	0.060 ± 1.0 e-05	0.0 ± 0.0	0.50 ± 0.00046	1.0 ± 0.0	0.060 ± 0.0	0.24 ± 0.0044	0.72 ± 0.014	0.22 ± 0.0054	0.28 ± 0.0039
7	0.35 ± 0.030	0.80 ± 0.011	0.62 ± 0.11	0.37 ± 0.072	0.0034 ± 0.00048	0.62 ± 0.039	1.0 ± 0.0	0.15 ± 2.71e-06	0.46 ± 0.0093	0.86 ± 0.0014	0.78 ± 0.0023	0.41 ± 0.010
8	0.60 ± 0.042	0.88 ± 0.021	0.62 ± 0.050	0.62 ± 0.17	0.0042 ± 0.00048	0.70 ± 0.0398	1.0 ± 0.0	0.092 ± 3.47e-06	0.43 ± 0.028	0.71 ± 0.012	0.46 ± 0.0035	0.50 ± 0.050

TABLEAU B.1 – Coefficient de corrélation moyen de Matthews (MCC), aire moyenne sous la courbe (AUC), rappel moyen et précision moyenne après avoir effectué les expériences 1, 2 et 3 en utilisant le réseau neuronal convolutif du Tableau II.5.4. Les métriques sont indiquées pour chaque patient testé et correspondent à la moyenne sur 10 répétitions (moyenne ± écart-type).

Couche	Type	Taille Filtre	Taille Pooling	Stride	# filtres	Taille Sortie	fonction activation
1	Conv2D	2*2	-	1*1	8	(8, 59, 8)	relu
2	maxpooling	-	1*2	1*1	-	(8, 29, 8)	relu
3	Conv2D	2*2	-	1*1	8	(7, 28, 8)	relu
4	maxpooling	-	1*2	1*1	-	(7, 14, 8)	relu
5	Conv2D	2*2	-	1*1	16	(6, 13, 16)	relu
6	maxpooling	-	1*2	1*1	-	(6, 6, 16)	relu
7	Conv2D	2*2	-	1*1	16	(5, 5, 16)	relu
8	maxpooling	-	2*2	1*1	-	(2, 2, 16)	relu
9	Conv2D	2*2	-	1*1	32	(1, 1, 32)	relu
10	Conv2D	1*1	-	1*1	2	(1, 1, 2)	softmax

TABLEAU B.2 – Fully Convolutional Network 1 (FCN1) proposé.

Couche	Type	Taille Filtre	Taille Pooling	Stride	# filtres	Taille Sortie	fonction activation
1	Conv2D	2*2	-	1*1	8	(8, 59, 8)	relu
2	maxpooling	-	1*2	1*1	-	(8, 29, 8)	relu
3	Conv2D	2*2	-	1*1	8	(7, 28, 8)	relu
4	maxpooling	-	1*2	1*1	-	(7, 14, 8)	relu
5	Conv2D	2*2	-	1*1	16	(6, 13, 16)	relu
6	maxpooling	-	1*2	1*1	-	(6, 6, 16)	relu
7	Conv2D	2*2	-	1*1	32	(5, 5, 32)	relu
8	Conv2D	2*2	-	1*1	32	(4, 4, 32)	relu
9	Conv2D	2*2	-	1*1	32	(3, 3, 32)	relu
10	Conv2D	2*2	-	1*1	64	(2, 2, 64)	relu
11	Conv2D	2*2	-	1*1	2	(1, 1, 2)	softmax

TABLEAU B.3 – Fully Convolutional Network 2 (FCN2) proposé.

Couche	Type	Taille Filtre	Taille Pooling	Stride	# filtres	Taille Sortie	fonction activation
1	Conv2D	2*2	-	1*1	8	(8, 59, 8)	relu
2	maxpooling	-	1*2	1*1	-	(8, 29, 8)	relu
3	Conv2D	2*2	-	1*1	8	(7, 28, 8)	relu
4	maxpooling	-	1*2	1*1	-	(7, 14, 8)	relu
5	Conv2D	2*2	-	1*1	16	(6, 13, 16)	relu
6	Conv2D	2*2	-	1*1	16	(5, 12, 16)	relu
7	Conv2D	2*2	-	1*1	32	(4, 11, 32)	relu
8	Conv2D	2*2	-	1*1	32	(3, 10, 32)	relu
9	Conv2D	2*2	-	1*1	32	(2, 9, 32)	relu
10	Conv2D	2*2	-	1*1	64	(1, 8, 64)	relu
11	Conv2D	1*2	-	1*1	64	(1, 7, 64)	relu
12	Conv2D	1*2	-	1*1	64	(1, 6, 64)	relu
13	Conv2D	1*2	-	1*1	64	(1, 5, 64)	relu
14	Conv2D	1*2	-	1*1	64	(1, 4, 64)	relu
15	Conv2D	1*2	-	1*1	64	(1, 3, 64)	relu
16	Conv2D	1*2	-	1*1	64	(1, 2, 64)	relu
18	Conv2D	1*2	-	1*1	2	(1, 1, 2)	softmax

TABLEAU B.4 – Fully Convolutional Network 3 (FCN3) proposé.

Patient	FCN1			FCN2			FCN3		
	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3
1	0.13 / 45.7	0.12 / 50.2	0.14 / 48.3	0.16 / 49.7	0.12 / 50.2	0.16 / 46.9	0.21 / 49.7	0.12 / 50.2	0.13 / 46.4
2	0.17 / 49.7	0.18 / 51.1	0.61 / 46.8	0.21 / 51.1	0.18 / 51.1	0.65 / 46.9	0.18 / 51.1	0.18 / 51.1	0.52 / 48.1
3	0.055 / 49.6	0.053 / 43.8	0.22 / 44.0	0.055 / 44.1	0.053 / 43.8	0.23 / 44.0	0.055 / 44.1	0.053 / 43.8	0.23 / 44.0
4	0.48 / 51.1	0.17 / 44.9	0.54 / 43.9	0.48 / 44.4	0.17 / 44.9	0.55 / 40.5	0.42 / 44.4	0.17 / 44.9	0.52 / 38.8
5	0.37 / 44.1	0.11 / 45.4	0.44 / 43.5	0.40 / 43.5	0.11 / 45.4	0.49 / 41.3	0.39 / 43.8	0.11 / 45.4	0.47 / 42.0
6	0.11 / 44.4	0.11 / 43.0	0.27 / 41.3	0.11 / 43.0	0.11 / 43.0	0.18 / 41.3	0.11 / 43.0	0.11 / 43.0	0.20 / 42.6
7	0.57 / 44.4	0.26 / 45.0	0.54 / 43.2	0.61 / 42.6	0.26 / 45.0	0.59 / 42.3	0.55 / 43.1	0.26 / 45.0	0.55 / 43.3
8	0.65 / 43.0	0.17 / 47.1	0.48 / 34.9	0.66 / 40.7	0.17 / 47.1	0.55 / 33.9	0.65 / 40.9	0.17 / 47.1	0.51 / 38.2

TABLEAU B.5 – Dice (DSC) et distance de Hausdorff (HD) après avoir effectué 10 répétitions des expériences 1, 2 et 3 en utilisant trois différents réseaux entièrement convolutifs : FCN1 (1ère colonne), FCN2 (2ème colonne) et FCN3 (3ème colonne). Les métriques sont indiquées pour chaque patient testé et correspondent à la moyenne sur 10 répétitions (moyenne ± écart-type). Nous avons indiqué en gras quand l'expérience 3 a donné les meilleures performances à l'échelle du patient.