Demande n°23052 - ARED
Acronyme : MAPPIS

Identification du projet

Identification du projet

N°	23052
Statut	
Date de création	2020-01-27
Date de modification	
Date de soumission	
Acronyme	MAPPIS
Intitulé du projet	Mapping the fMRI pipeline-space towards more robust pipelines
Dispositif	ARED
Sexe	F
Nom	Fromont
Prénom	Elisa
Domaine d'innovation stratégique	Economie numérique sécurisée et responsable
Sous-domaine d'innovation stratégique	données et intelligence
DIS Secondaire	Economie de la santé et du bien-être pour une meilleure qualité de vie
Sous DIS secondaire	Technologies pour la santé

Organisme de tutelle

Organisme de tutelle

Etablissement	Université de Rennes 1
Commune	Rennes
Nom du contact	Alis
Prénom du contact	David
Adresse électronique	david.alis@univ-rennes1.fr
Tél.	02 23 23 36 36

Unité de recherche

Unité de recherche

Nom de l'unité	UMR IRISA 6074

Civilité	M.
Nom du responsable	Gravier
Prénom du responsable	Guillaume
Tél.	
Courriel	guig@irisa.fr

Précisions sur l'activité de l'équipe (HDR/doctorants/post-doc...)

Libellé (pour LACODAM)	Valeur
Nombre d'HDR dans l'équipe	4
Nombre de thèses en cours dans l'équipe	15
Nombre de post-doc en cours dans l'équipe	2

Précisions sur situation Equipe/Laboratoire porteurs

Libellé	Valeur
Statut et numéro de l'unité de recherche (UMR, USR, EA)	UMR 6074
Le cas échéant, nom de l'équipe de recherche	LACODAM/EMPENN
Le cas échéant, prénom et nom du responsable de l'équipe de recherche	Alexandre Termier / Pierre Maurel
Activités de recherche et points forts de l'équipe	Science des données / Imagerie Médicale
Si implication dans un/plusieurs programmes européens ou internationaux, préciser	H2020 ICT48-TAILOR

Co-directeur de thèse

Co-directeur de thèse

Sexe	F
Nom	Maumet
Prénom	Camille
Laboratoire	IRISA
Etablissement de tutelle	Inria

Projet

Description du projet

N°	Libellé	Valeur
	Résumé du projet (4000 caractères maximum)	Dans une étude récente, une collaboration de 200 chercheurs a démontré comment le choix d'une chaîne de traitement donnée en imagerie cérébrale peut avoir un impact important sur les conclusions d'une étude (Botvinik-Nezer et al., 2020). Dans cette collaboration historique, les chercheurs - répartis en 70 équipes - ont analysé le *même* jeu de données afin de répondre aux *mêmes* questions de recherche prédéfinies, la seule variable étant la chaîne de traitement choisie par chaque équipe en fonction de son expertise. Au total, les 70 équipes ont choisi 70 pipelines différents, menant, dans de nombreux cas, à des résultats contradictoires. En neuroimagerie, l'espace des pipelines est particulièrement vaste. De nombreuses étapes sont nécessaires pour préparer et analyser les données,

		notamment : la correction du mouvement, le recalage dans un espace standardisé, le lissage spatial, etc. Et pour chaque étape, différents algorithmes sont disponibles, implémentés dans divers logiciels de neuroimagerie. Bien qu'il existe des bonnes pratiques, les chercheurs doivent encore faire de nombreux choix. **Cette thèse s'attachera à mieux comprendre les relations qui existent entre les différents pipelines afin de pouvoir mieux guider les choix que les chercheurs lorsqu'ils choisissent une pipeline de traitement**. L'objectif de cette thèse est de construire une "carte" de l'espace du pipeline de neuroimagerie qui sera utilisée pour répondre à un ou plusieurs des défis suivants : 1) Débogage des pipelines ; 2) Identifier les cas où les pipelines souffrent de violations de leurs hypothèses sous-jacentes ; 3) Éliminer les pipelines inadaptés en fonction du jeu de données d'entrée. Les développements méthodologiques seront centrés sur l'utilisation de méthodes d'apprentissage profond qui permettront de représenter les résultats des chaînes de traitement dans un espace dimensionnel inférieur. Cet espace sera utilisé pour estimer la proximité ou la distance entre différentes chaînes de traitements. En utilisant cette carte des distances entre les pipelines dans différents contextes, nous trouverons une solution aux problèmes proposés ci-dessus.
	Hypothèses, questions posées, identification des points de blocage (4000 caractères maximum)	L'objectif de cette thèse est de construire une "carte" de l'espace des chaînes de traitements en neuroimagerie qui sera utilisée pour étudier un ou plusieurs des défis suivants : 1) Débogage des pipelines ; 2) Identifier les cas où les pipelines souffrent de violations de leurs hypothèses sous-jacentes ; 3) Éliminer les pipelines inadaptés en fonction du jeu de données d'entrée. **Débogage**. Il a été démontré que certains pipelines sont instables en présence de petites perturbations dans les données d'entrée, ce qui remet en question la robustesse de leur mise en œuvre (Kiar et al., 2020). L'objectif du débogage des pipelines est d'identifier les pipelines qui ne se comportent pas comme prévu en raison de la présence d'un bogue. **Violation d'une hypothèse**. Chaque étape d'un pipeline analytique est la mise en œuvre d'une méthode qui s'accompagne de certaines hypothèses. Si ces hypothèses ne sont pas respectées, cela peut conduire à un comportement incorrect du pipeline. **Inadéquation d'un ensemble de données spécifique**. Dans d'autres cas, un pipeline échouera pour un ensemble de données donné en raison des propriétés particulières de ces données (par exemple, propriétés spécifiques de bruit, etc.). Dans tous les cas ci-dessus, une exploration systématique de l'espace des chaînes de traitements est impossible. En outre, il n'existe généralement pas de vérité de terrain pouvant être utilisée pour comparer les résultats des pipelines dans le cadre de la neuroimagerie. Dans cette thèse, nous proposons de fournir une représentation des statistiques de l'IRMf dans un espace dimensionnel inférieur et d'utiliser cet espace pour estimer la proximité ou la distance des différents pipelines. En

		utilisant cette carte des distances entre les pipelines dans différents contextes, nous trouverons une solution aux problèmes exposés ci-dessus.
	Approches méthodologique et technique envisagées (4000 caractères maximum)	Les approches d'apprentissage profond apprennent des structures inconnues à partir des données. Certaines variantes de ces réseaux, comme les auto-codeurs ou les réseaux adversaires générateurs, ont été conçues pour apprendre la représentation latente des données dans un cadre non supervisé. Pour apprendre une représentation invariante au pipeline de l'IRMf, nous pensons que les techniques d'apprentissage métrique par paires ou par triplés pourraient également être utilisées conjointement avec ces approches d'apprentissage de représentation non supervisée (voir (Schroff, Kalenichenko, et Philbin 2015 ; Mohammadi et Kain 2017 ; Bhattarai, Sharma, et Jurie 2016)). Les applications dans le domaine de l'imagerie cérébrale (par exemple (Richard et al. 2018)) sont encore limitées en partie en raison de la petite taille des échantillons (Poldrack et al. 2017). Ce travail utilisera la plus grande base de données de cartes statistiques d'imagerie par résonance magnétique fonctionnelle (IRMf) : NeuroVault (https://neurovault.org/) (Gorgolewski et al., 2015) qui contient actuellement plus de 150 000 images. Cette base de données sera mise à profit pour construire un encodeur (GAN, autoencodeur, etc.) qui permettra d'obtenir une représentation dans un espace en dimension réduite. Cette partie du travail s'appuiera sur les travaux préliminaires réalisés en 2019 : https://github.com/gkiar/analytical-stability.
	Environnement scientifique et positionnement dans le contexte régional, national et international	Ce travail de thèse vient tout d'abord s'intégrer dans deux équipes de l'IRISA UMR 6074 : LACODAM spécialiste en "data mining" et Empenn spécialisée dans l'analyse d'images cérébrales. D'autre part, ce travail de thèse bénéficiera de deux réseaux de collaborations internationales. Tout d'abord la COST Action européenne GLIMR dont l'objectif est de développer un réseau de collaboration européen sur les gliomes. Camille Maumet dirige le work package "Multi-site Data integration" dont l'objectif est de faciliter la réutilisation de données, notamment en prenant en compte les chaînes de traitements utilisées. Ce travail bénéficiera aussi d'une collaboration internationale avec le groupe de travail "Neuroimaging Data Sharing" porté par l'International Coordinating Neuroinformatics Facility (INCF) qui vise à définir des standards et outils pour développer le partage de données dans la communauté de l'imagerie cérébrale. Camille Maumet (co-encadrante) fera le lien avec ce groupe de travail dont elle est une collaboratrice depuis 8 ans. Ce travail bénéficiera aussi de collaborations avec le projet H2020 ICT48-TAILOR "Foundations of Trustworthy AI".
	Contexte scientifique et partenarial : éléments généraux (4000 caractères maximum)	Dans une étude récente, une collaboration de 200 chercheurs a démontré comment le choix d'une chaîne de traitement donnée en imagerie cérébrale peut avoir un impact important sur les conclusions d'une étude (Botvinik-Nezer et al., 2020). Dans cette collaboration historique, les chercheurs - répartis en 70 équipes - ont analysé le *même* jeu de données afin de répondre aux *mêmes* questions de recherche prédéfinies, la seule variable étant la chaîne de traitement choisie par chaque équipe en fonction de son expertise. Au total, les 70 équipes ont choisi 70 pipelines différents, menant, dans de nombreux cas, à des résultats contradictoires.

		En neuroimagerie, l'espace des pipelines est particulièrement vaste. De nombreuses étapes sont nécessaires pour préparer et analyser les données de la neuroimagerie, notamment : la correction du mouvement, le recalage dans un espace standardisé, le lissage spatial, etc. Et pour chaque étape, différents algorithmes sont disponibles, implémentés dans divers logiciels de neuroimagerie. Bien qu'il existe des bonnes pratiques, les chercheurs doivent encore faire de nombreux choix. **Cette thèse s'attachera à mieux comprendre les relations qui existent entre les différents pipelines**.
	Adéquation du projet avec le DIS de rattachement (4000 caractères maximum)	En s'intéressant aux données d'imagerie cérébrale et à leurs chaînes de traitements, ce projet s'intègre dans le DIS "Economie numérique sécurisée et responsable", en particulier au levier "données et intelligence" avec pour objectif une meilleure extraction des connaissances des données disponibles. De par son application au domaine de l'imagerie cérébrale, ce projet participe aussi à l'amélioration des technologies pour la santé, et contribue donc de façon secondaire au DIS "Economie de la santé et du bien-être pour une meilleure qualité de vie".
	Si projet blanc, préciser les raisons de ce choix (4000 caractères maximum)	Non
	Inscription éventuelle du projet dans les objectifs de la Breizh COP (4000 caractères maximum)	Non
	Si lien avec projet ERC, préciser lequel (4000 caractères maximum)	
	Lieu(x) de déroulement du projet	UMR IRISA, Rennes
	Ecole doctorale de rattachement	MathSTIC
	Date d'obtention de l'HDR par le porteur	Decembre 2015
	Financements acquis par le porteur au cours des 3 dernières années (titre, montant) (4000 car. max.)	
	Cotutelle internationale (préciser "oui" ou "non"; si oui, préciser le pays et l'établissement)	non
	Si autre type de collaboration internationale, préciser (nature, partenaire, pays)	non
	Nom(s) du(des) cofinanceur(s) envisagé(s) ou confirmé(s)	Demande bourse IA CDO Université de Rennes 1
	Si cofinancement industriel, justifier le choix du dispositif ARED au lieu du dispositif CIFRE	

	État du cofinancement (préciser si "acquis" ou "demandé")	demandé
	Si demandé, date prévisionnelle de réponse du cofinanceur	
	Si cofinancement refusé, autre(s) source(s) de cofinancement identifiée(s)	ressources propres openAIRE H2020 et bourse IUF
	Si reversement prévu, préciser (partenaire, montant)	non
	Cofinancement international (préciser "oui" ou "non")	non
	Information(s) complémentaire(s) sur le projet	

Candidat

Candidat

Profil souhaité pour le candidat	• Some knowledge about computer vision • Interest in medical imaging • Interest (even better knowledge) in machine learning in general and deep learning particular • Good programming skills in Python (for Keras) • Very good understanding of English
Sexe (si candidat déjà identifié)	
Nationalité du candidat (si déjà indentifié)	

Cofinancement international

Cofinancement international

Pays	
Etablissement de tutelle à l'étranger	
Calendrier des périodes de séjour	

Pièces à joindre

Pièces à joindre

Type de pièce	Commentaire	Document à déposer	V	Commentaire
ARED - Dépôt - CV du porteur	Obligatoire - Le CV doit préciser si le porteur est titulaire de l'HDR et, le cas échéant, depuis quand, et mentionner la liste des doctorant·e·s déjà encadré·e·s et des publications associées.			
ARED - Dépôt - Avis directeur·rice d'unité	Facultatif - 10 lignes maximum			
ARED - Dépôt - Justificatif cofinancement	Facultatif - Dépôt de demande, justificatif d'obtention, courrier du cofinanceur pressenti...	non		

ARED - Dépôt - Lettre partenaire étranger	Facultatif - Lettre d'engagement de l'établissement partenaire étranger en cas de cotutelle internationale	non		
ARED - Dépôt - Pièce complémentaire	Facultatif	non		
ARED - Dépôt - Pièce complémentaire	Facultatif	non		