



**AfIA**

Association française  
pour l'Intelligence Artificielle

# Bulletin N° 112

---

*Association française pour l'Intelligence Artificielle*

---

AfIA



---

## PRÉSENTATION DU BULLETIN

Le [Bulletin](#) de l'Association française pour l'Intelligence Artificielle vise à fournir un cadre de discussions et d'échanges au sein de la communauté universitaire et industrielle. Ainsi, toutes les contributions, pour peu qu'elles aient un intérêt général pour l'ensemble des lecteurs, sont les bienvenues. En particulier, les annonces, les comptes rendus de conférences, les notes de lecture et les articles de débat sont très recherchés. Le [Bulletin](#) de l'AfIA publie également des dossiers plus substantiels sur différents thèmes liés à l'IA. Le comité de rédaction se réserve le droit de ne pas publier des contributions qu'il jugerait contraire à l'esprit du bulletin ou à sa politique éditoriale. En outre, les articles signés, de même que les contributions aux débats, reflètent le point de vue de leurs auteurs et n'engagent qu'eux-mêmes.

---

### ■ Édito

Ce numéro de printemps du [Bulletin](#) de AfIA est l'occasion de saluer le départ de Claire LEFEVRE du comité de rédaction, et toute l'équipe de rédaction tenait à la remercier pour son travail. L'équipe de rédacteurs ne se réduit pas pour autant car nous saluons l'arrivée de Gaël LEJEUNE parmi nous, et le remercions à l'avance pour son travail.

Cette année est marquée également par un changement important car Yves DEMAZEAU a quitté ses fonctions de président de l'AfIA. Nous le remercions chaleureusement pour son investissement exceptionnel, et nous saluons Benoit LE BLANC qui prend sa succession. Afin que nos adhérents puissent mieux connaître le nouveau président de l'AfIA, vous trouverez en page [3](#) un mot du président, ainsi qu'une courte biographie en page [4](#). La composition actuelle du Conseil d'Administration se trouve en quatrième de couverture.

Dans la suite de ce [Bulletin](#), vous retrouverez les rubriques habituelles, à commencer par le dossier monté par Fleur MOUGIN (BPH, Université de Bordeaux) et Lina F. SOUALMIA (LIMICS / LITIS, Université de Rouen Normandie) sur « Santé & IA ». C'est un large dossier qui fait suite aux journées « Santé & IA » que l'AfIA a organisées les années passées, avec 14 contributions provenant d'autant d'équipes ou laboratoires français (voir page [6](#)).

Ce [Bulletin](#) vous présente également le compte rendu des Journées « Enseignement et Formation en Intelligence Artificielle » du 7 janvier 2021 et « Technologies du Langage Humain et Santé » du 4 février 2021 (voir pages [63](#) et [65](#)). Vous y trouverez enfin la liste des thèses et HDR soutenues lors du trimestre écoulé, page [71](#). Encore un grand merci à tous les contributeurs de ce numéro, sans oublier Gaël LEJEUNE pour sa relecture assidue.

**Bonne lecture à tous !**

Grégory BONNET, Rédacteur



---

## SOMMAIRE

### DU BULLETIN DE L'AFIA

---

Mot du Président . . . . .	3
Biographie de Benoit Le Blanc . . . . .	4
<b>5</b>	<b>Dossier « Santé &amp; IA »</b>
La Santé, un domaine d'application majeur de l'IA . . . . .	6
Équipe « Decision Aid and Information Discovery » (DECIDE) . . . . .	9
Traitement Automatique de Langues pour la simplification de documents de santé. . . . .	12
Le Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-santé : LIMICS . . . . .	15
IA au sein d'un entrepôt de données de santé à Rouen . . . . .	18
Intelligence artificielle pour la santé au LTCl : équipe IMAGES et chaire C2M . . . . .	20
Intelligence embarquée en milieu contraint . . . . .	25
L'équipe-projet HeKA . . . . .	29
Apprentissage automatique et graphes spatio-temporels pour l'analyse d'images médicales . . . . .	33
LACODAM : Science des données appliquée aux parcours de soins . . . . .	36
IA & Neuroimagerie . . . . .	41
Intelligence Artificielle pour la Chirurgie . . . . .	45
Représentations des médecins généralistes et des patients concernant l'utilisation de l'IA en Santé : une étude qualitative . . . . .	50
Le LIFAT et CETU ILIAD3 œuvrent conjointement en santé et IA . . . . .	54
Laboratoire LAMIH : Recherches en Santé & IA . . . . .	59
<b>62</b>	<b>Comptes rendus de journées, événements et conférences</b>
Journée « Enseignement et Formation en Intelligence Artificielle » . . . . .	63
Journée « Technologies du Langage Humain et Santé » . . . . .	65
<b>70</b>	<b>Thèses et HDR du trimestre</b>
Thèses de Doctorat . . . . .	71
Habilitations à Diriger les Recherches . . . . .	72



## ■ Mot du Président

### **Valeur économique et interdisciplinarité : les deux mutations actuelles de l'IA**

Depuis un peu moins de dix ans, les capacités offertes par les réseaux de neurones dans le domaine de la reconnaissance des formes sont venues rebattre les arguments de l'opposition symbolisme / connexionnisme parmi les chercheurs en intelligence artificielle. La machine doit-elle s'en tenir aux éléments de sens fournis par les programmeurs ? ou bien la machine peut-elle produire ses propres éléments logiques sur lesquels un raisonnement va se bâtir ? La question reste entière mais le paysage de l'intelligence artificielle a été durablement modifié. Les résultats bien réels de l'apprentissage profond dans des applications de reconnaissance des visages, de repérage des plaques d'immatriculation, d'analyse des tâches cutanées ou encore de l'examen de l'écriture manuscrite, ont ouvert par milliers les perspectives d'automatisation de tâches considérées jusque-là comme exclusivement humaines, telle que par exemple la conduite d'une automobile en environnement urbain.

La recherche en intelligence artificielle est passée en quelques années d'une situation de laboratoire à un sujet incontournable des discussions économiques, politiques et médiatiques. Les journalistes ont abandonné le chiffon rouge de l'intelligence artificielle forte pour passer à l'explication d'applications toutes plus novatrices les unes que les autres. Les dirigeants des pays industrialisés se sont tous lancés dans la mise en place de stratégies nationales avec l'ambition de préserver leurs capacités de recherche académique, d'attirer les investisseurs et entrepreneurs mais aussi d'anticiper les mutations sociales qui vont se manifester dans le monde du travail et dans l'édu-

cation. Les grandes entreprises se sont toutes dotées de centres de recherche en IA et de perspectives d'intégration d'algorithmes de traitement de leurs données massives dans le développement de leurs nouveaux produits et services.

En conséquence de quoi, la recherche académique en intelligence artificielle évolue maintenant dans un monde où les repères économiques et les lignes disciplinaires sont en pleine mutation. Dans ses activités et dans ses perspectives, l'AfIA se doit de tenir compte de cette double ouverture de l'IA : vers la monétisation de ses produits d'une part et vers les disciplines de la santé, de l'environnement, de la défense, de la mobilité, de l'éducation, de la justice, etc. d'autre part.

C'est de façon très lucide que la dernière Assemblée Générale de l'Association française pour l'Intelligence Artificielle a pris comme résolutions (1) de faire croître et consolider le Collège industriel mis en place avec les acteurs historiques de l'IA en entreprise, (2) de renforcer les aspects liés à la pluridisciplinarité, voir même d'ouvrir les chercheurs en IA à la pratique de l'interdisciplinarité avec leurs collègues de campus, (3) d'inscrire l'événement annuel de la Plate-Forme en IA comme le lieu permettant d'observer et de rendre tangibles ces évolutions de la recherche en IA.

Pour mener cette évolution, nous bénéficions de la vision stratégique d'Yves DEMAZEAU qui a présidé l'AfIA de 2011 à 2020. Il a su méthodiquement mettre en place et structurer l'activité associative de plus de trois cents chercheurs à travers des Groupes de Travail dédiés aux chantiers de l'association (rédaction des bulletins trimestriels, réalisation des journées thématiques mensuelles, tenue de la plate-forme hébergeant simultanément plus d'une demi-douzaine de conférences annuelles, attri-



**AfIA**

Association française  
pour l'Intelligence Artificielle

bution d'un prix de thèse, etc.) et à travers des Collèges thématiques destinés à explorer la variété des sujets de l'IA (apprentissage artificiel, interaction avec l'humain, représentation et raisonnement, science de l'ingénierie des connaissances, systèmes multi-agents et agents autonomes, technologies du langage humain, compétitions, et un collège consacré aux applications industrielles de l'IA).

Membre du regroupement des sociétés académiques européennes consacrées à l'IA (EurAI) et membre du tout nouveau Collège des Sociétés Savantes Académiques de France (CoSSAF), l'AfIA a pour volonté de mettre en partage le sujet de la recherche académique en intelligence artificielle : partage entre ses membres issus de toute la variété des thématiques de l'IA et partage avec les autres acteurs de la recherche, qu'ils soient en entreprise ou bien dans des disciplines où l'IA s'applique. A ce sujet, l'AfIA a toujours eu une attention particulière pour promouvoir la pluridisciplinarité, organisant mensuellement des rencontres scientifiques sous forme de journées thématiques, la plupart du temps co-organisées avec une autre société savante : santé et IA, jeux et IA, défense et IA, agronomie et IA, robotique et IA, éthique et IA, etc.

La question est maintenant de regarder comment renforcer, voire étendre ces ponts entre disciplines. Si la pluridisciplinarité réside dans l'association de « disciplines qui concourent à une réalisation commune, mais sans que chaque discipline ait à modifier sa propre vision des choses et ses propres méthodes », l'interdisciplinarité « a pour but d'élaborer un formalisme suffisamment général et précis pour permettre d'exprimer dans ce langage unique les concepts, les préoccupations, les contributions d'un nombre plus ou moins grand de disciplines » (voir [ici](#) et [ici](#)).

L'AfIA rassemble, de façon historique et reconnue, les chercheurs dont les activités se consacrent au cœur de l'IA (les algorithmes et les modèles pour le travail sur les connaissances) et aux plateformes intégratives de l'IA (comme par exemple pour aborder les questions du traitement de la langue, des systèmes multi-agents, des interactions avec le monde physique, etc.). L'association est maintenant aussi attendue pour travailler dans les domaines d'applications de l'IA, avec des spécialistes des disciplines ou domaines cibles.

*Benoit LE BLANC  
Président de l'AfIA  
Mars 2021*

## ■ Biographie de Benoit Le Blanc

Benoit Le Blanc est professeur à Bordeaux INP (doctorat en 1992, HDR en 2002). Ses travaux sur la formalisation des connaissances recueillies auprès des experts et sur les systèmes d'inférence pour la décision médicale ont servi de base au développement de recherches actuelles sur la cognition collective et sur la collaboration humains-IA. Chargé de mission pour l'IA auprès de la DGRI de 2014 à 2019, il est aujourd'hui directeur de l'École nationale supérieure de cognitive (ENSC-Bordeaux INP).



**AfIA**  
Association française  
pour l'Intelligence Artificielle

---

## Dossier

### « Santé & IA »

---

Dossier réalisé par

**Fleur MOUGIN**

*Bordeaux Population Health/ERIAS*

*Université de Bordeaux*

*[Fleur.Mougin@u-bordeaux.fr](mailto:Fleur.Mougin@u-bordeaux.fr)*

**Lina F. SOUALMIA**

*LIMICS et LITIS/TIBS*

*Université de Rouen Normandie*

*[Lina.Soualmia@litislab.fr](mailto:Lina.Soualmia@litislab.fr)*

## ■ La Santé, un domaine d'application majeur de l'IA

Les premiers dossiers de l'AfIA dédiés à la "Médecine" ont été publiés en 1993 [1], en 2002 [2], ainsi que dans le numéro spécial de la Revue d'Intelligence Artificielle en 2011 [3]. En 2014, le dernier dossier du [Bulletin 84 de l'AfIA](#), coordonné par Pierre Zweigenbaum et Jean Charlet [4] a évolué vers le thème plus général qui est celui de la "Santé", et présentait ainsi les différentes conférences internationales spécialisées, les équipes en France, ainsi que leurs projets phares dans ce domaine.

Le [rapport Villani](#), publié en mars 2018 [5], avait mis en avant le domaine de la santé comme l'un des quatre secteurs pour lesquels les opportunités offertes par l'Intelligence Artificielle (IA) devaient être prioritairement exploitées en France. Un des résultats escomptés était d'arriver à une médecine personnalisée et prédictive pour le suivi en temps réel du patient, mais aussi une meilleure détection d'anomalies dans les imageries, comme les électrocardiogrammes. Le [dossier de l'Inserm](#), datant de la même année, précise les champs d'application des recherches menées pour la santé (médecine prédictive, médecine de précision, prothèses intelligentes, chirurgie assistée par ordinateur, etc.) en utilisant la logique et/ou l'apprentissage automatique sur de grands volumes de données [6].

En parallèle et sous l'égide de la société savante *International Medical Informatics Association (IMIA)*, le [volume 28 du Yearbook of Medical Informatics](#) recense les meilleurs travaux internationaux publiés en 2018 en Santé & IA [7]. Ces travaux sont sélectionnés selon leur originalité, leur rupture avec l'existant, ou encore leur succès en santé dans plusieurs sous-domaines, comme les systèmes d'aide à la décision, la gestion des connaissances, l'imagerie et les capteurs, la recherche clinique, les systèmes d'information hospitaliers, la sécurité des pa-

tients, le traitement du langage clinique, etc. Le constat qui en a été tiré est une évolution de l'IA des systèmes experts vers des approches guidées par les données et l'apprentissage automatique avec des résultats probants en analyse d'images.

En France, et depuis 2014, plusieurs « journées » dédiées à ce thème ont été organisées sous la forme d'ateliers dans le cadre des Journées Francophones d'Ingénierie des Connaissances (IC). En effet, les principales recherches de la communauté se focalisent sur le traitement des données et connaissances pour leur application au soin (aide à la décision diagnostique et thérapeutique), et plus largement dans l'accès aux informations et connaissances médicales (représentation, indexation, recherche, traduction, échange dans des systèmes interopérables). Ces recherches reposent principalement sur des processus d'extraction d'informations et de connaissances à partir de supports variés, de leur validation, évaluation et mise en application au bénéfice des patients, des professionnels de santé, des étudiants en santé, des chercheurs et du grand public (premier [Atelier IA & Santé - Ingénierie des Connaissances & Santé – IC 2014](#), Clermont-Ferrand).

À partir de 2016, ces ateliers sont soutenus par l'AfIA, le Collège Science de l'Ingénierie des Connaissances (SIC) et l'Association Française d'Informatique Médicale (AIM) : [Atelier IA & Santé 2016](#) à Montpellier, [Atelier IA & Santé 2018](#) à Nancy, [Atelier IC IA & Santé 2019](#) à Toulouse.

Au regard du succès de ces différents ateliers en nombre de participants et en nombre de contributions, démontrant l'intérêt du domaine de la santé, la journée Santé & IA devient en 2020 une journée hébergée par la Plate-Forme annuelle de l'AfIA : [Santé & IA 2020](#) (virtuelle) à Angers et la prochaine [Santé & IA 2021](#) (vir-



**AfIA**

Association française  
pour l'Intelligence Artificielle

tuelle) à Bordeaux.

Ce dossier est loin de dresser un panorama exhaustif de toutes les équipes ou unités menant des activités en IA et santé en France. Certaines d'entre elles sont toujours actives et représentées dans le dossier du [Bulletin 84 de l'AfIA](#) (il en est de même pour les projets phares pilotés en France dans ce domaine). On peut y trouver des laboratoires CNRS, Inserm, universitaires, incluant des hôpitaux et des CHU dont plusieurs disposent d'Entrepôts de Données de Santé (EDS), ainsi que des Instituts.

Les travaux recensés dans ce dossier permettent cependant d'en dégager une tendance forte qui est l'imagerie médicale, quasi absente des bulletins précédents de l'AfIA. En effet, avec les succès et la dynamique de l'apprentissage automatique et de l'apprentissage profond en imagerie et en traitement du signal, c'est naturellement que l'application à la santé y a trouvé sa place.

Ainsi, outre les travaux en Traitement Automatique du Langage Naturel (D2IM, DECIDE, LIMICS, STL), en IA symbolique (D2IM, LIMICS, LTSI/MEDICIS), en analyse de données et fouille de données à partir d'EDS (D2IM, HeKA, LACODAM, LAMIH), ou encore dans le domaine des systèmes de santé apprenants et d'aide à la décision (HeKA, LIFAT, LIMICS), on peut trouver dans ce dossier plusieurs équipes qui sont spécialisées en traitement de données d'imagerie par apprentissage (classification, *clustering*, réseaux de neurones profonds).

En terme d'applications, les recherches de l'équipe CNRS DECIDE englobent le TAL et la fouille de données relatives à la santé mentale (problématique majeure de santé publique depuis la crise du COVID-19) ainsi que la classification d'images issues d'imagerie par résonance magnétique (IRM) structurale dans la maladie neuro-dégénérative d'Alzheimer.

Le LTCI, au sein des équipes IMAGES et de la chaire C2M, par l'association de modèles et de réseaux de neurones profonds, contribue dans le domaine de l'interprétation d'images et l'étude de l'exposition des personnes aux ondes électromagnétiques.

Dans l'équipe SYEL du LIP6, trois axes majeurs visent à valider des nouveaux dispositifs comme les capteurs, ou encore de nouvelles architectures embarquées : la modélisation des systèmes sur puces hétérogènes (consommation, intégrité du signal), la conception d'architectures reconfigurables et l'intégration des modèles d'inférence d'IA sur des architectures embarquées (reconfigurables et non).

L'apprentissage automatique est également une méthodologie privilégiée dans l'équipe SDC du laboratoire CNRS ICube pour l'analyse de données d'imagerie médicale, l'apprentissage profond pour l'imagerie histopathologique et la combinaison avec des approches de modélisation (graphes spatio-temporels) pour l'analyse d'images obtenues par IRM fonctionnelle.

L'institut GIN et l'équipe « Neuroimagerie Fonctionnelle et Perfusion Cérébrale » de l'Inserm se proposent d'explorer et de raffiner les nouvelles techniques d'apprentissage automatique dans le contexte de l'imagerie *in vivo* en neuroimagerie.

Les membres de l'équipe Inserm MEDICIS du LTSI combinent représentation des connaissances (ontologies) et apprentissage profond avec trois thèmes principaux : les représentations sémantiques pour le partage de données radiologiques et chirurgicales, les réseaux de neurones profonds et les modèles de connaissances pour l'analyse des procédures en chirurgie ainsi que l'apprentissage profond pour le traitement d'images et de signaux médicaux en neurochirurgie fonctionnelle.

Toujours dans le domaine de l'imagerie et de la chirurgie, le LIFAT développe des



recherches en segmentation d'images neuro-anatomiques en 3D, et en réalité augmentée en chirurgie prothétique de l'épaule. L'ordonnement et l'optimisation pour l'aide à la planification de la production de chimiothérapie, la gestion de flotte d'ambulances et la planification d'actes médicaux sont autant d'applications exploitant des méthodes issues de la recherche opérationnelle.

Les nombreux systèmes à base d'IA développés pour la santé ont généralement pour but d'aider les cliniciens à améliorer la prise en charge de leurs patients tout en réduisant les coûts. Ainsi, l'intégration de données, leur stockage et leur gestion représentent encore en 2021 de réels défis, parmi lesquels le manque d'explicabilité pour le clinicien des résultats produits par les systèmes guidés par les données, ce qui peut limiter l'utilisation des systèmes d'aide à la décision développés.

D'autres défis tout aussi importants seront à relever comme les aspects liés à l'éthique et à la protection des données personnelles, et ce malgré le *Health Insurance Portability and Accountability Act* (HIPAA) de 1996 aux États-Unis et le Règlement Général sur la Protection des Données (RGPD) voté en 2016 par le parlement européen et applicable dans les états membres de l'Union Européenne depuis le 25 mai 2018. Le projet Droit et Responsabilité des IA en Santé, mené au LACODAM, en est une illustration. Par ailleurs, la transparence dans les processus de traitement des données de santé est une condition pour obtenir la confiance des cliniciens et des patients et leur adoption des IA et des outils numériques dans leur routine [8].

Comme illustré dans ce dossier, les travaux en IA & Santé sont en constante évolution. Enfin, sans modèle robuste, nous ne ferons pas

de prédiction sur la prochaine publication d'un dossier IA & Santé dans un bulletin de l'AfIA ([le Bulletin 84 de l'AfIA](#) prédisait un dossier en 2026). Nous sommes persuadées qu'un nouveau dossier IA & Santé sera publié dans une échéance très proche, compte tenu des nombreux Instituts et des Chaires IA & Santé qui voient actuellement le jour en France.

## Références

- [1] J. Charlet. Dossier IA & Médecine. In : : *Bulletin de l'AFIA 14*, pages 22–48, 1993.
- [2] V. Corruble et J. Charlet. Dossier IA & Médecine. In : *Bulletin de l'AFIA 48*, pages 15–49, 2002.
- [3] LF. Soualmia. IA et Santé : Vers Quelles Applications en Médecine ? In *Revue d'Intelligence Artificielle*, Vol.25, N.4, 2011.
- [4] P. Zweigenbaum et J. Charlet. Dossier IA & Santé. In : *Bulletin de l'AFIA 84*, pages 4–37, 2014.
- [5] C. Villani [Donner un sens à l'Intelligence Artificielle](#) ; pages 194–203, 2018.
- [6] Dossier Inserm. [Intelligence artificielle et santé : des algorithmes au service de la médecine](#), 2018.
- [7] HK. Hollis, LF. Soualmia, B. Séroussi. Artificial Intelligence in Health Informatics : Hype or Reality ? In *Yearbook of Medical Informatics*, Volume 28, N. 1, pages 3–4, 2019.
- [8] B. Séroussi, KF. Hollis, LF. Soualmia. Transparency of Health Informatics Processes as the Condition of Healthcare Professionals' and Patients' Trust and Adoption : the Rise of Ethical Requirements. In *Yearbook of Medical Informatics*, Volume 29, N.1, pages 7–10, 2020.



## ■ Équipe « Decision Aid and Information Discovery » (DECIDE)

<sup>1</sup> CHRU de Brest / IMT Atlantique / LATIM

<sup>2</sup> IMT Atlantique / Lab-STICC / DECIDE  
<https://www.labsticc.fr>

<sup>3</sup> CHRU de Brest / IMT Atlantique /  
Lab-STICC / DECIDE

**Sofian BERROUIGUET<sup>1</sup>**

**Romain BILLOT<sup>2</sup>**

**Yannis HARALAMBOUS<sup>2</sup>**

**Christophe LEMEY<sup>3</sup>**

**Philippe LENCA<sup>2</sup>**

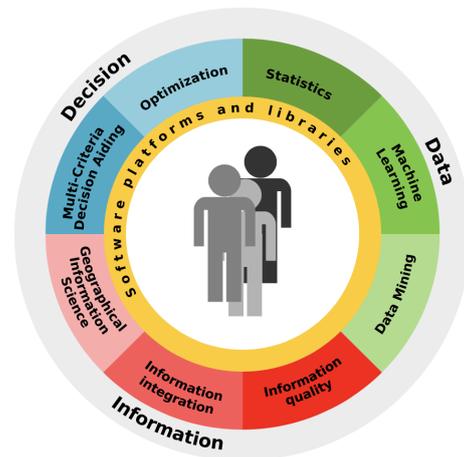
**Sorin MOGA<sup>2</sup>**

**John PUENTES<sup>2</sup>**

### Introduction

L'équipe « Decision Aid and Information Discovery » (DECIDE) de l'UMR CNRS 6285 Lab-STICC mène des travaux algorithmiques et méthodologiques afin de faciliter la prise de décision dans des contextes impliquant un ou plusieurs décideurs confrontés à des données hétérogènes et complexes. Cette aide à la décision est abordée à travers le triptyque données – information – décision. Ainsi, les recherches menées puisent leurs sources dans les disciplines de la statistique, de l'apprentissage automatique et de la fouille de données (pour le point de vue données); des sciences de l'information géographique, de l'intégration et fusion d'informations et de la qualité de l'information (pour le point de vue information); et de l'optimisation et de l'aide multi-critère à la décision (pour le point de vue décision). Cette aide à la décision est soutenue à travers différentes plate-formes et logiciels, auxquels l'équipe contribue fortement, et où l'humain garde une place centrale.

Les travaux de l'équipe sont appliqués à des domaines variés, dont celui de la santé. Nous présentons brièvement ci-dessous nos principales contributions, principalement en ce qui concerne l'axe « données », dans le cadre de la santé mentale et de la détection de la maladie d'Alzheimer.



Triptyque données – information – décision

### Données et santé mentale

En santé mentale, nous avons développé, avec les CHRU de Brest et Montpellier et l'*Instituto de Investigación Sanitaria de la Fundación Jiménez Díaz* de Madrid, des travaux sur la prévention du risque suicidaire et sur la détection précoce de la schizophrénie.

Les principaux résultats d'un processus complet de fouille de données sociologiques et cliniques sur un échantillon de suicidants de deux hôpitaux européens sont présentés dans [2]. Nous avons identifié trois groupes de patients similaires, à partir d'une analyse de correspondances multiples et d'une classification ascendante hiérarchique. Ensuite, les



facteurs de risque, par groupe et associés au nombre de tentatives sont, eux, identifiés à l'aide d'arbres de régression. Les résultats mettent en lumière l'apport de la fouille de données aux activités de description et d'explication ; ils mettent également en exergue la nécessité d'améliorer la qualité des données. Enfin, ces travaux seront enrichis par des évaluations écologiques [1], par le biais de données issues du milieu de vie habituel des patients, comme la qualité du sommeil [3], ou encore à travers de données liées à la représentation sociale du suicide [5].

Enfin, nous nous intéressons à la détection et à l'intervention précoce dans le cadre des troubles psychotiques débutants. À partir d'un (petit) corpus d'entretiens de bilan psychiatrique, en fusionnant des données provenant d'un analyseur syntaxique et des données chronométrées, et en y étudiant des phénomènes reliant la syntaxe du langage naturel, la prosodie et les pauses (vides ou pleines), nous avons obtenu des indicateurs de comorbidités psychiatriques permettant d'assister les psychiatres dans leurs diagnostics [4]. Pour mieux identifier les indicateurs de comorbidités dans les entretiens, nous avons créé un corpus témoin, suivant le même protocole que les entretiens psychiatriques mais avec des sujets sains. Dans le cadre d'une thèse, nous allons corrélérer les indicateurs obtenus avec des symptômes cliniques observés au cours des deux années de suivi.

Ces dernières années il y a eu, mondialement, pléthore d'articles sur l'application de l'apprentissage artificiel et du traitement automatique de la langue à la santé mentale. Ces articles diffèrent tant sur le plan des méthodes et des plate-formes que sur le plan du vocabulaire et du public visé. Pour donner une vision globale de ces efforts, nous avons étudié 327 articles et établi une synthèse de 58 d'entre eux dans [6].

## Données et détection de la maladie d'Alzheimer

La maladie d'Alzheimer (MA) est un type de démence affectant des millions de personnes dans le monde. Parmi les moyens de diagnostique, l'imagerie par résonance magnétique structurale (sMRI en anglais) du cerveau peut révéler des signes de neurodégénérescence. Or, l'interprétation de ces signes tant par des experts humains que par des algorithmes est difficile, principalement à cause de la complexité de la maladie, de la variabilité des images et des différences démographiques. En outre, la fiabilité de l'analyse d'images assistée par ordinateur est directement associée à la manière de représenter l'information extraite des images.

L'équipe DECIDE, en collaboration avec le *Departamento d'Ingeniería de Sistemas y Computación* de l'*Universidad de los Andes* de Bogota, explore les approches de représentation guidées par les données des images sMRI du cerveau, focalisées sur l'analyse de la neurodégénérescence et son lien potentiel avec la MA. Des résultats encourageants suggèrent la possibilité d'identifier des images 2D représentatives d'un volume 3D, comme étant caractéristiques de l'état de la neurodégénérescence d'un sujet [8]. Deux approches de classification ont obtenu des résultats performants, en appliquant une quantification par dictionnaire non supervisé pour représenter le volume total du cerveau [7] et en définissant un modèle d'auto-encodeur supervisé qui utilise le contenu d'une seule image 2D du volume 3D du cerveau par individu [9]. Finalement, l'analyse régionale basée sur la réponse locale des autoencodeurs supervisés révèle des régions anatomiques pertinentes pour l'identification de la neurodégénérescence éventuellement associée à la MA, permettant de différencier les images pathologiques des saines [9].



## Références

- [1] Sofian Berrouiguet, Romain Billot, Mark Larsen, Jorge Lopez-Castroman, Isabelle Jaussent, Michel Walter, Philippe Lenca, Enrique Baca-Garcia, and Philippe Courtet. An approach for data mining of electronic health record data for suicide risk management: Database analysis for clinical decision support. *JMIR Mental Health*, 6(5):e9766, May 2019.
- [2] Romain Billot, Sofian Berrouiguet, Mark Larsen, Michel Walter, Jorge López Castroman, Enrique Baca-Garcia, Philippe Courtet, et Philippe Lenca. Apport de la fouille de données pour la prévention du risque suicidaire. In *Conférence Internationale sur l'Extraction et la Gestion des Connaissances, Revue des Nouvelles Technologies de l'Information*, volume RNTI-E-34, pages 143–154, Paris, France, janvier 2018.
- [3] Elise Guillodo, Christophe Lemey, Mathieu Simonnet, Michel Walter, Enrique Baca-Garcia, Vincent Masetti, Sorin Moga, Mark Larsen, Hugopsy Network, Juliette Ropars, and Sofian Berrouiguet. Clinical applications of mobile health wearable-based sleep monitoring: Systematic review. *JMIR mHealth and uHealth*, 8(4), April 2020.
- [4] Yannis Haralambous, Christophe Lemey, Philippe Lenca, Romain Billot, and Deok-Hee Kim-Dufor. Using dependency syntax-based methods for automatic detection of psychiatric comorbidities. In Dimitrios Kokkinakis, Kristina Lundholm Fors, Charalambos Themistocleous, Malin Antonsson, and Marie Eckerström, editors, *Resources and processing of linguistic, para-linguistic and extra-linguistic data from people with various forms of cognitive/psychiatric/developmental impairments*, pages 142–150, Marseille, France, May 2020.
- [5] Louis Jehel, Dupont Emmanuel, Philippe Lenca, and Céline Kopp-Bigault. The ReSCU-M study: Comprehension of suicide social representation in Martinique island. Perspective for prevention and communication strategies. In *International Association for Suicide Prevention Caribbean Regional Symposium*, Trinidad, Trinidad and Tobago, May 2019.
- [6] Aziliz Le Glaz, Yannis Haralambous, Deok-Hee Kim-Dufor, Philippe Lenca, Romain Billot, Ryan Taylor, Jonathan Marsh, Jordan DeVlyder, Michel Walter, Sofian Berrouiguet, and Christophe Lemey. Machine learning and natural language processing in mental health: a systematic review. *Journal of Medical Internet Research*, 8(4), forthcoming/in press.
- [7] Ricardo Mendoza-León, Fabio A. González, Pablo Arbeláez, John Puentes, and Marcela Hernández Hoyos. Analysis of PHOW representations for Alzheimer disease classification on brain structural MRI. In *IEEE 13th International Symposium on Biomedical Imaging*, pages 24–27. IEEE, 2016.
- [8] Ricardo Mendoza-Léon, John Puentes, Fabio A. González, and Marcela Hernández Hoyos. Empirical evaluation of general-purpose image features for pathology-oriented image retrieval of Alzheimer disease cases. In *30th International Congress on Computer Assisted Radiology and Surgery, International Journal of Computer Assisted Radiology and Surgery*, volume 11, pages S39–S40. Springer, 2016.
- [9] Ricardo Mendoza-Léon, John Puentes, Luis Felipe Uriza, and Marcela Hernández Hoyos. Single-slice Alzheimer's disease classification and disease regional analysis with supervised switching autoencoders. *Computers in biology and medicine*, 116:103527, 2020.



## ■ Traitement Automatique de Langues pour la simplification de documents de santé

CNRS, UMR 8163 STL, Université de Lille  
<http://natalia.grabar.free.fr>

**Natalia GRABAR**

[natalia.grabar@univ-lille.fr](mailto:natalia.grabar@univ-lille.fr)

**Anaïs KOPTIENT**

[an.koptient@hotmail.fr](mailto:an.koptient@hotmail.fr)

**Rémi CARDON**

[remi.cardon@univ-lille.fr](mailto:remi.cardon@univ-lille.fr)

### Introduction

La simplification automatique vise à fournir une version simplifiée des textes. La simplification peut concerner le lexique, la syntaxe, la sémantique mais aussi la pragmatique et l'organisation des textes. La simplification peut être vue comme une aide fournie aux lecteurs [19, 6] ou comme un pré-traitement pour d'autres applications de TAL [5, 21].

Nous nous intéressons à la simplification de documents médicaux car ils comportent une terminologie technique (myocarde, cholécystectomie, aplasie...). De tels termes sont souvent incompréhensibles par les patients. Le travail décrit est effectué dans le cadre du projet CLEAR (*Communication, Literacy, Education, Accessibility, Readability*) financé par l'ANR sous référence ANR-17-CE19-0016-01.

### Travaux existants en simplification

Ces dernières années, la simplification automatique a connu un grand succès auprès des chercheurs. Il existe actuellement trois types de méthodes qui sont exploitées pour effectuer la simplification :

- *Méthodes guidées par les connaissances et les règles*. Ces méthodes exploitent des ressources de type WordNet [4], la fréquence des mots [8, 7] ou leur longueur [1] ;

- *Méthodes issues de la traduction automatique*, où la simplification correspond à une traduction monolingue allant d'un texte technique vers un texte simple [22, 18] ;
- *Méthodes exploitant la sémantique distributionnelle*. Entraînés sur des données convenables, les vecteurs de mots peuvent contenir des équivalents plus simples de termes, qui peuvent alors être exploités pour effectuer la simplification [9, 13].

La majorité des travaux existants sont réalisés sur des documents en anglais et exploitent des données provenant de Wikipedia, Simple Wikipedia ou encore Newsela.

### Nos contributions

Nous avons proposé plusieurs contributions, qu'elles concernent les méthodes ou les ressources, concernant différents aspects de la simplification automatique :

1. création d'un corpus avec des documents comparables en français, qui met en regard les documents médicaux en version technique et leur version simplifiée ou simple [10]. Ce corpus regroupe des informations sur les médicaments (RCP et notices de médicaments), des résumés des revues systématiques Cochrane et des articles encyclopédiques (Wikipedia et Vikidia) ;
2. méthodes pour la recherche de phrases parallèles au sein de documents comparables



- [2]. Ces méthodes effectuent la catégorisation des paires de phrases en deux catégories : phrases alignables ou non alignables. Plusieurs descripteurs sont exploités, comme des descripteurs lexicaux (nombre de mots communs, longueur des phrases), des descripteurs basés sur les chaînes d'édition, les descripteurs basés sur des similarités lexicales (cosinus, Dice et Jaccard), des descripteurs basés sur les n-grammes de mots et de caractères, et des descripteurs basés sur des plongements lexicaux. Ces travaux ont permis de constituer un ensemble de plus de 10 000 couples de phrases alignées ;
3. annotation de couples de phrases alignées selon les transformations dues à la simplification et élaboration d'une typologie des procédés de simplification [14] ;
  4. méthodes pour la détection de mots et de passages difficiles dans les documents médicaux : (1) classification de mots grâce à l'exploitation d'un lexique médical annoté en difficulté [12], (2) exploitation de méthodes d'oculométrie, où les fixations plus longues et les saccades plus courtes indiquent la présence de passages plus difficiles à comprendre [11], (3) exploitation de méthodes neuronales avec un lexique médical annoté en difficulté [20] ;
  5. création d'un lexique où les termes médicaux techniques sont associés avec leurs paraphrases grand public, comme {*myocarde*; *muscle du cœur*} ou {*cholécystectomie*; *ablation de la vésicule biliaire*}. Ce lexique contient actuellement 7 580 paraphrases pour 4 516 termes médicaux. De plus, les termes et les paraphrases du lexique sont associés à des indices de leur lisibilité [16]. À terme, ce lexique sera aligné avec les concepts d'UMLS [17] afin de rendre cette ressource plus facilement exploitable ;
  6. méthodes pour la simplification automatique de documents médicaux : une méthode à base de règles [15] et une méthode exploitant les approches neuronales issues de la traduction automatique [3]. Ce dernier travail a permis également de produire la traduction française du corpus WikiLarge [23] avec 300 000 couples de phrases.
- Ces différentes contributions ont permis de faire des avancées sur différents aspects liés à la simplification automatique. Notons que plusieurs des ressources constituées sont librement disponibles pour la recherche<sup>1</sup>. Nous espérons que ces ressources pourront motiver les travaux autour de la simplification en français.

## Références

- [1] Susana Bautista, Pablo Gervás, and R. Ignacio Madrid. Feasibility analysis for semi-automatic conversion of text to improve readability. In *Int Conf on Inform and Comm Technology and Accessibility (ICTA)*, pages 33–40, 2009.
- [2] Rémi Cardon and Natalia Grabar. Construction d'un corpus parallèle à partir de corpus comparables pour la simplification de textes médicaux en français. volume 61, pages 15–39, 2020.
- [3] Rémi Cardon and Natalia Grabar. French biomedical text simplification : When small and precise helps. In *COLING 2020*, pages 1–8, 2020.
- [4] John Carroll, Guido Minnen, Yvonne Canning, SiobhanDevlin, and John Tait. Practical simplification of English newspaper text to assist aphasic readers. In *AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10, 1998.
- [5] Raman Chandrasekar and Bangalore Srinivas. Automatic induction of rules for

1. <http://natalia.grabar.free.fr/resources.php>



- text simplification. *Knowledge Based Systems*, 10(3) :183–190, 1997.
- [6] Ping Chen, John Rochford, David N. Kennedy, Soussan Djamasbi, Peter Fay, and Will Scott. Automatic text simplification for people with intellectual disabilities. In *AIST*, pages 1–9, 2016.
- [7] Jan De Belder and Marie-Francine Moens. Text simplification for children. In *Workshop on Accessible Search Systems of SIGIR*, pages 1–8, 2010.
- [8] Siobhan Devlin and John Tait. The use of psycholinguistic database in the simplification of text for aphasic readers. In *Linguistic Database*, pages 161–173, 1998.
- [9] Goran Glavas and Sanja Stajner. Simplifying lexical simplification : Do we need simplified corpora? In *ACL-COLING*, pages 63–68, 2015.
- [10] Natalia Grabar and Rémi Cardon. Clear – simple corpus for medical French. In *Workshop on Automatic Text Adaption (ATA)*, pages 1–11, 2018.
- [11] Natalia Grabar, Emmanuel Farce, and Laurent Sparrow. Study of readability of health documents with eye-tracking approaches. In *Workshop on Automatic Text Adaption (ATA)*, pages 1–11, 2018.
- [12] Natalia Grabar and Thierry Hamon. A large rated lexicon with French medical words. In *LREC (Language Resources and Evaluation Conference)*, pages 1–12, 2016.
- [13] Yea-Seul Kim, Jessica Hullman, Matthew Burgess, and Eytan Adar. SimpleScience : Lexical simplification of scientific terminology. In *EMNLP*, pages 1–6, 2016.
- [14] Anaïs Koptient, Rémi Cardon, and Natalia Grabar. Simplification-induced transformations : typology and some characteristics. In *Proc of the 18th BioNLP Workshop and Shared Task*, pages 309–318, Florence, Italy, 2019.
- [15] Anaïs Koptient and Natalia Grabar. Fine-grained text simplification in french : steps towards a better grammaticality. In *Proc of ISHIMR 2020*, pages 1–10, 2020.
- [16] Anaïs Koptient and Natalia Grabar. Rated lexicon for the simplification of medical texts. In *Proc of HEALTHINFO 2020*, pages 1–6, 2020.
- [17] Donald A Lindberg, Betsy L Humphreys, and Alexa T McCray. The Unified Medical Language System. *Methods Inf Med*, 32(4) :281–291, 1993.
- [18] Sergiu Nisioi, Sanja Stajner, Simone Paolo Ponzetto, and Liviu P. Dinu. Exploring neural text simplification models. In *Ann Meeting of the Assoc for Comp Linguistics*, pages 85–91, 2017.
- [19] Gustavo H. Paetzold and Lucia Specia. Benchmarking lexical simplification systems. In *LREC*, pages 3074–3080, 2016.
- [20] Hanna Pylieva, Artem Chernodub, Natalia Grabar, and Thierry Hamon. Generalizability of readability models for medical terms. In *MEDINFO 2019*, pages 1–5, 2019.
- [21] Chih-Hsuan Wei, Robert Leaman, and Zhiyong Lu. SimConcept : A hybrid approach for simplifying composite named entities in biomedicine. In *BCB '14*, pages 138–146, 2014.
- [22] Sander Wubben, Antal van den Bosch, and Emiel Krahmer. Sentence simplification by monolingual machine translation. In *Annual Meeting of the Association for Computational Linguistics*, pages 1015–1024, 2012.
- [23] Xingxing Zhang and Mirella Lapata. Sentence simplification with deep reinforcement learning. In *Conference on Empirical Methods in Natural Language Processing*, pages 584–594, 2017.



## ■ Le Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-santé : LIMICS

LIMICS  
Inserm U1142

LIMICS  
Sorbonne Université & Inserm U1142

LIMICS  
APHP.Sorbonne Université & Inserm U1142

**Marie-Christine JAULENT**  
[marie-christine.jaulent@inserm.fr](mailto:marie-christine.jaulent@inserm.fr)

**Xavier TANNIER**  
[xavier.tannier@sorbonne-universite.fr](mailto:xavier.tannier@sorbonne-universite.fr)

**Ferdinand DHOMBRES**  
[ferdinand.dhombres@inserm.fr](mailto:ferdinand.dhombres@inserm.fr)

### Introduction

Le laboratoire d'informatique médicale et d'ingénierie des connaissances en e-Santé (LIMICS) est une unité de recherche interdisciplinaire en informatique et en informatique médicale. Nous développons des approches innovantes de traitement de l'information de santé sur les plans à la fois méthodologique et applicatif. Il s'agit de rendre les systèmes d'information plus autonomes sur la compréhension des données et des connaissances qu'ils manipulent afin qu'ils communiquent entre eux et s'intègrent mieux dans les activités décisionnelles de soins et de recherche.

L'unité comporte 38 membres permanents incluant des chercheurs Inserm, des enseignants-chercheurs en informatique et des médecins et pharmaciens hospitalo-universitaires qui sont proches de la pratique et permettent des collaborations multiples avec des cliniciens pour le développement et l'évaluation des produits de la recherche. Ainsi, une ambition forte du LIMICS est de se positionner comme une des principales équipes européennes en e-santé. Nous décrivons ici l'activité du LIMICS à l'exclusion de celle des chercheurs de Rouen, décrite indépendamment.

### Aperçu de l'état actuel du domaine de recherche

Le LIMICS propose des approches d'intelligence artificielle (IA) symboliques et numériques de traitement de l'information de santé sur les plans à la fois méthodologique et applicatif. Les domaines d'application sont liés à la fouille de textes et de données pour les hôpitaux (principalement en association avec l'entrepôt de données de santé de l'AP-HP), à l'aide à la décision en médecine, à la prise en compte de nouvelles sources de données (les objets connectés, les réseaux sociaux, les données environnementales mais aussi les nouvelles données médicales, telles que les données -omiques), la position centrale du patient nécessitant une communication des systèmes d'information entre ville et hôpital (par exemple en psychiatrie), ou encore le vieillissement de la population qui pousse nos gouvernants à afficher comme priorité nationale le domaine de la prévention, les parcours de soins (en particulier pour des patients poly-pathologiques et polymédiqués) et, au delà, les parcours de santé ainsi que la lutte contre les déserts médicaux en favorisant le maintien à domicile des patients.



## Détails

Pour cette section, nous mettons en avant quelques projets marquants récemment terminés ou en cours au [LIMICS](#).

- **Aide à la décision.** C'est une des thématiques principales du LIMICS. Nous présentons ici deux projets, DESIREE et SUOG. Le projet européen DESIREE s'appuie sur l'approche symbolique pour aider les cliniciens dans le traitement et le suivi des patientes atteintes de cancers du sein. Ces maladies très complexes nécessitent en effet souvent des adaptations de protocoles classiques. La plateforme intègre des recommandations de bonne pratique par la mise en œuvre d'un raisonnement fondé sur une ontologie. Le système peut également apprendre à partir des cas déjà résolus ou d'un raisonnement par expérience (réutilisation de décisions) [1].  
Le projet EIT-HEALTH *Smart Ultrasound in Obstetrics and Gynecology* (SUOG) prend la suite du projet OPPIO [3] pour développer un système d'aide à la décision en temps réel dans les systèmes d'échographie en obstétrique. Le but est d'apporter une assistance à l'acquisition d'images et à l'orientation diagnostique des anomalies fœtales. Une base de connaissances comportant l'expertise de 11 centres de référence en échographie prénatale est à la disposition des échographistes moins expérimentés. L'IA est utilisée dans ce projet : 1) pour générer des protocoles d'imagerie personnalisés par inférence sur l'ontologie au cœur du système et 2) pour présenter les bonnes images des cas de la base de connaissance par détection de similarité avec les images produites par l'échographiste lors de son examen. L'IA développée est à la fois symbolique (inférences sur une base de connaissance) et numérique (*deep learning* sur une collection d'images) [4].

- Le projet RHU [PsyCARE](#), lancé en janvier 2020 vise la prévention de la psychose précoce qui touche 15 000 adolescents par an. Pour cela, le projet met en œuvre les technologies de l'IA à deux niveaux : 1) dans le développement de l'**ontologie**, pilotée par le LIMICS, qui servira de pivot d'intégration des données, et pour un annotateur sémantique comme dans le projet PARON [2] et 2) au niveau du "calcul" d'un score avec des algorithmes d'**apprentissage** dans le but de construire un système d'aide à l'orientation des patients dans le parcours de soin.
- Plusieurs projets de **traitement automatique des langues** sont menés au LIMICS dans le cadre de partenariats avec les entrepôts de données de l'APHP et de Rouen. À l'APHP, des projets sur l'application clinique d'extraction d'information dans les comptes-rendus médicaux sont menés, comme l'identification des patients ayant subi des fractures de faible énergie (projet Epifractal), la détection des sites infectieux d'une bactériémie (projet BactHub), la caractérisation des lésions décrites à l'issue des mammographies (projet EZ Mammo), etc. D'autres projets plus fondamentaux sont également menés pour permettre la réduction de la supervision nécessaire dans les systèmes d'apprentissage profond pour l'extraction d'information [6].
- Plusieurs projets sont également en cours avec l'hôpital de la Pitié-Salpêtrière dans le domaine du diagnostic et de la lutte contre le paludisme, en utilisant d'une part les **images** de frottis sanguins (identification de globules rouges infectés) et les résultats de **spectrométrie** de masse pour l'étude des moustiques vecteurs de la maladie [5].

## Thèses en cours

A. **Abdellatif** (dir : B Séroussi-J Bouaud, Amélioration des pratiques des soignants en



EHPAD, développement et évaluation d'un assistant numérique intelligent), A. **Acherar** (dir : X Tannier-R Piarroux, analyse d'images pour le diagnostic du paludisme), N. **Bannour** (dir : A Névéol-X Tannier-B Rance, Traitement automatique des langues pour les réunions pluridisciplinaires dans le domaine du cancer), J. **Bonnard** (dir. K Bailly-F Dhombres, Intelligence artificielle statistique et symbolique pour l'aide à la pratique de la médecine fœtale), T. **Bouchez** (dir : J Charlet-M Schuers, Modèles de prise de décision interprofessionnelle en soins primaires), R. **Dandan** (dir : S Despres-J Nobécourt, Intégration de données guidée par une ontologie. Application à l'assistance personnalisée aux personnes âgées), J. **Hilbey** (dir : J Charlet, Médecine de précision : la voie ontologique. Application au domaine de la psychiatrie), P. **Karapetiantz** (dir : MC Jaulent-C Bousquet, Détection automatisée de signaux de pharmacovigilance dans les médias sociaux), A. **Koivogui** (dir : C Duclos, Calcul du taux de participation réel au dépistage organisé du cancer colorectal après détermination optimisée du taux d'exclusion par chaînage des données du programme de dépistage et des données du SNDS), N. **Mohammad** (dir : R Piarroux-X Tannier, Apprentissage automatique des résultats de spectrométrie de masse), M. **Mouazer** (dir : JB Lamy-K Sedki, Conception et évaluation d'un outil d'aide informatisée au bilan de médicaments), JS **Rech** (dir : O Steichen, Recours aux soins des patients drépanocytaires : caractérisation à partir de l'entrepôt de données de santé de l'AP-HP et du SNDS), A. **Redjal** (dir : B Séroussi-J Bouaud, Aide à la décision pour la prise en charge des patients polyopathologiques - complémentarité des approches à base de connaissances et des méthodes d'apprentissage à partir des données), A. **Saab** (dir : JB Lamy, Détection et classification d'alertes pour les événements indésirables cliniques en combinant ontologie et apprentis-

sage automatique), C. **Tahri** (dir : X Tannier, Exploration d'information dans des corpus de textes), Y. **Taillé** (dir : X Tannier-C Daniel, Extraction d'information peu supervisée dans les comptes-rendus médicaux), P. **Wajsbürt** (dir : X Tannier-C Daniel, Phénotypage et extraction d'information dans les comptes-rendus médicaux).

## Références

- [1] Jacques Bouaud, Sylvia Pelayo, Jean-Baptiste Lamy, Coralie Prebet, Charlotte Ngo, Luis Teixeira, Gilles Guézennec, and Brigitte Séroussi. Implementation of an ontological reasoning to support the guideline-based management of primary breast cancer patients in the DESIREE project. *Artificial Intelligence in Medicine*, 108 :101922, August 2020.
- [2] Sonia Cardoso, Pierre Meneton, Xavier Aimé, Vincent Meininger, David Grabli, Gilles Guezennec, and Jean Charlet. Use of a modular ontology and a semantic annotation tool to describe the care pathway of patients with amyotrophic lateral sclerosis in a coordination network. *PLoS One*, 16(1) :e0244604, 2021.
- [3] Ferdinand Dhombres, Paul Maurice, Stéphanie Friszer, Lucie Guilbaud, Nathalie Lelong, Babak Khoshnood, Jean Charlet, Nicolas Perrot, Eric Jauniaux, Davor Jurkovic, and Jean-Marie Jouannic. Developing a knowledge base to support the annotation of ultrasound images of ectopic pregnancy. *Journal of Biomedical Semantics*, 8(1) :4, January 2017.
- [4] Ferdinand Dhombres, Paul Maurice, Lucie Guilbaud, Loriane Franchinard, Barbara Dias, Jean Charlet, Eléonore Blondiaux, Babak Khoshnood, Davor Jurkovic, Eric Jauniaux, and Jean-Marie Jouannic. A novel intelligent scan assistant system for early pregnancy diagnosis by ultrasound :



Clinical decision support system evaluation study. *Journal of medical Internet research*, 21(7), July 2019.

- [5] Cécile Nabet, Aurélien Chaline, Jean-François Franetich, Jean-Yves Brossas, Noémie Shahmirian, Olivier Silvie, Xavier Tannier, and Renaud Piarroux. Prediction of malaria transmission drivers in Anopheles mosquitoes using artificial intelligence cou-

pled to MALDI-TOF mass spectrometry. *Scientific Reports*, 10(1), July 2020.

- [6] Perceval Wajsbürt, Arnaud Sarfati, and Xavier Tannier. Medical concept normalization in French using multilingual terminologies and contextual embeddings. *Journal of Biomedical Informatics*, 12(114), January 2021.

## ■ IA au sein d'un entrepôt de données de santé à Rouen

D2IM, CHU de Rouen/LIMICS  
Université de Rouen Normandie  
<http://www.cismef.org/>

**Stéfan J. DARMONI**  
[stefan.darmoni@chu-rouen.fr](mailto:stefan.darmoni@chu-rouen.fr)

### Introduction

Depuis 5 ans, le département de l'informatique et de l'information médicales (D2IM) du CHU de Rouen développe un entrepôt de santé, mis à disposition de la communauté des professionnels de santé [1].

Plus de 170 cas d'usage ont été réalisés, dépassant largement la recherche clinique, pour aller vers le calcul semi-automatique d'indicateurs de qualité, la détection de cas de pharmacovigilance ou de prescription hors autorisation de mise sur le marché par exemple, sans oublier plus récemment la détection des patients éligibles à la vaccination anti-COVID.

Le D2IM se focalise sur des travaux de recherche sur les terminologies et ontologies de santé [2, 3], l'indexation [4] et l'extraction automatique [5, 6] et la recherche d'information [8, 10].

Le D2IM est une équipe pluridisciplinaire, composée de 16 membres avec des profils différents (Kevin Billey - Ingénieur, Badisse Dahamna - Chef de projets, Stéfan Darmoni - PUPH, Mikaël Dusenne - Médecin, Julien Grosjean - Ingénieur, Gaëtan Kerdelhué - Documen-

taliste, Ivan Kergourlay - Ingénieur, Romain Le Guillon - Interne en pharmacie, Emeline Lejeune - Documentaliste, Romain Lelong - Ingénieur, Jean-Philippe Leroy - Médecin, Catherine Lertord - Documentaliste, Clément Massonnaud - Interne de santé publique, Thibaut Pressat - Doctorant, Matthieu Schuers - MCUMG, Lina Soualmia - MCF (membre associée)). Les acteurs du D2IM sont membres du LIMICS UMRS 1142, S.J. Darmoni en étant le directeur adjoint.

### Positionnement

Si de nombreuses équipes académiques ont développé des entrepôts de données de santé (EDS) aux États Unis (comme I2B2), puis en France (comme eHOP à Rennes, Dr. Warehouse à Paris Imagine ou Consore pour les Centres de Lutte Contre le Cancer), notre approche sémantique à tous les niveaux des outils employés (serveur terminologique, annotateur sémantique et moteur de recherche sémantique) est une des caractéristiques fortes de nos choix stratégiques. Un autre choix stratégique est le choix de développer notre EDS



en utilisant NoSQL pour minimiser les temps de réponse, afin que les utilisateurs soient en quasi temps réel, leur permettant d'améliorer lors d'une même session de travail leurs requêtes.

### Architecture de l'EDS

L'EDS du CHU de Rouen repose sur trois couches sémantiques : (a) HeTOP, un serveur terminologique interlingue contenant à ce jour plus de 85 terminologies et ontologies de santé en 45 langues. Nous avons entrepris un effort particulier sur la traduction de certaines terminologies étasuniennes, comme la *Human Phenotype Ontology* (HPO) ou OMIM par exemple, aboutissant à ce qu'HeTOP contient plus de 520 000 concepts avec un CUI (*Concept Unique Identifier*) de l'UMLS. Comme seulement 20 terminologies d'HeTOP sont intégrées dans l'UMLS, nous avons créé un pendant français au CUI : le HUI pour *HeTOP Unique Identifier*, avec plus de 600 000 HUI différents au sein d'HeTOP ; (b) ECMT, annotateur sémantique, capable en 6ms d'indexer un document de santé (le plus souvent un compte-rendu textuel) en se fondant sur les concepts d'HeTOP. Cet annotateur est en passe de devenir hybride, en se reposant sur l'apprentissage profond en utilisant les 19 millions de documents de santé de notre entrepôt [7] ; (c) moteur de recherche sémantique, permettant de rechercher dans la même requête des données structurées et non structurées [8, 9, 10].

### Projets en cours

À ce jour, le D2IM, en collaboration étroite avec le LIMICS, participe à quatre projets de recherche financés par l'ANR (programme Technologies pour la Santé) : ApiApps [11], PSYHAMM, REMIAMES et ABIMeD, ainsi qu'un projet européen EIT-HEalth (HDQ4HP).

### Références

- [1] R. Lelong, LF. Soualmia, J. Grosjean, M. Talbaa, SJ. Darmoni. Building a Semantic Health Data Warehouse in the Context of Clinical Trials : Development and Usability Study. In *JMIR Med Inform.* 2019 Dec 20 ;7(4) :e13917.
- [2] A. Benis, M. Crisan-Vida, L. Stoicu-Tivadar, SJ. Darmoni. A Multi-Lingual Dictionary for Health Informatics as an International Cooperation Pillar. *Stud Health Technol Inform.* 2019 Vol. 262 ; pages : 31–34. IOS press 2019.
- [3] J. Grosjean, K. Billey, J. Charlet, S.J. Darmoni. Manual Evaluation of the Automatic Mapping of International Classification of Diseases (ICD)-11 (in French). In *Stud Health Technol Inform* 2020. Vol. 270 ; pages 1335–1336. IOS Press, 2020.
- [4] C. Siefriidt, J. Grosjean J, T. Lefebvre, L. Rollin, SJ. Darmoni, M. Schuers. Evaluation of Automatic Annotation by a Multi-terminological Concepts Extractor within a Corpus of Data from Family Medicine Consultations. In *Int J Med Inform.* 2020 Jan ;133 :104009.
- [5] C. Cabot, SJ. Darmoni, LF Soualmia. Cimind : A Phonetic-based Tool for Multilingual Named Entity Recognition in Biomedical Texts. In *J Biomed Inform.* 2019 Jun ;94 :103176.
- [6] N. Tvardik, I. Kergourlay, A. Bittar, F. Segond, SJ. Darmoni, MH. Metzger. Accuracy of using natural language processing methods for identifying healthcare-associated infections. *Int J Med Inform.* 2018 Sep ; Vol. 117 ; pages 96–102.
- [7] E. Dynomant, R. Lelong, B. Dahamna, C. Massonnaud, G. Kerdelhué, J. Grosjean, S. Canu, SJ. Darmoni. Word Embedding for the French Natural Language



**Afia**

Association française  
pour l'Intelligence Artificielle

- in Health Care : Comparative Study. *JMIR Med Inform.* 2019 Jul 29;7(3) :e12310.
- [8] C.R. Massonnaud, G. Kerdelhué, J. Grosjean, R. Lelong, N. Griffon, S.J. Darmoni. Identification of the Best Semantic Expansion to Query PubMed Through Automatic Performance Assessment of Four Search Strategies on All Medical Subject Heading Descriptors : Comparative Study. In *JMIR Med Inform.* 2020 Jun 4;8(6) :e12799.
- [9] A. Gillibert, N. Griffon, M. Schuers, K. Hardy, A. Elmerini, C. Letord, P. Staccini, S.J. Darmoni, J. Benichou. Impact on medical practice of accessing pharmaceutical records. *Int J Med Inform.* 2019 Jan;121 :58-63.
- [10] M. Ndangang, J. Grosjean, R. Lelong, B. Dahamna, I. Kergourlay, N. Griffon, S.J. Darmoni. Terminology Coverage from Semantic Annotated Health Documents. *Stud Health Technol Inform.* 2018; Vol. 255; pages 20–24. IOS Press, 2018.
- [11] J.C. Dufour, J. Grosjean, S.J. Darmoni, M. Yasini, G. Marchand, C. Simon, A. Sarradon-Eck, M. Préau, D. Darmon, M. Schuers, P. Hassanaly, R. Giorgi. ApiAppS : A Project to Study and Help Practitioners in Recommending mHealth Apps and Devices to Their Patients. In *Stud Health Technol Inform.* 2019. Vol. 264; pages 1919–1920. IOS Press 2019.

## ■ Intelligence artificielle pour la santé au LTCI : équipe IMAGES et chaire C2M

LTCI et LIP6/IMAGES et LFI  
LTCI, Télécom Paris, Institut Polytechnique  
de Paris - Sorbonne Université, CNRS, LIP6  
[perso.telecom-paris.fr/bloch](mailto:perso.telecom-paris.fr/bloch)

LTCI/IMAGES

LTCI/C2M

**Isabelle BLOCH**

[isabelle.bloch@telecom-paris.fr](mailto:isabelle.bloch@telecom-paris.fr),  
[isabelle.bloch@sorbonne-universite.fr](mailto:isabelle.bloch@sorbonne-universite.fr)

**Pietro GORI**

[pietro.gori@telecom-paris.fr](mailto:pietro.gori@telecom-paris.fr)

**Joe WIART**

[joe.wiart@telecom-paris.fr](mailto:joe.wiart@telecom-paris.fr)

### Introduction

Cette section présente les activités de recherche en intelligence artificielle (IA) pour la santé menées au LTCI, d'une part en interprétation d'images médicales dans l'équipe IMAGES, et d'autre part pour l'étude de l'exposition des personnes aux ondes électromagnétiques dans le cadre de la chaire C2M. Une composante commune est l'association entre des modèles (de connaissances ou physiques) et des méthodes d'apprentissage (en particu-

lier avec des réseaux de neurones profonds).

### Aperçu de l'état actuel du domaine

Les dernières années ont connu un essor considérable des méthodes d'apprentissage profond pour l'imagerie médicale, avec souvent des résultats spectaculaires, mais laissant encore un certain nombre de problèmes ouverts. En particulier, les masses de données nécessaires pour l'apprentissage ne sont pas toujours disponibles, par exemple en imagerie pédi-



**Afia**

Association française  
pour l'Intelligence Artificielle

trique ou lorsqu'il s'agit de maladies rares, ce que ne compensent que partiellement les méthodes de transfert d'apprentissage ou d'augmentation de données. Le coût de l'annotation des données dépasse également souvent ce qui est acceptable par les experts médicaux, et des méthodes visant à le réduire (approches semi-supervisées ou auto-supervisées par exemple) émergent. Enfin, peu de méthodes actuellement exploitent l'important corpus de connaissances médicales existant. La place respective des données et des connaissances dans ce type d'approche reste une question ouverte.

### **Contributions de l'équipe IMAGES et de la chaire C2M**

L'équipe Images, Modélisation, Analyse, GÉométrie et Synthèse (IMAGES) du Laboratoire Traitement et Communication de l'Information (LTCI) de Télécom Paris regroupe les recherches sur les images et les objets à trois dimensions, et en informatique graphique. L'expertise de l'équipe IMAGES réside dans le développement des méthodologies et des outils théoriques dans ces domaines, et s'attache à la résolution globale de problèmes complexes, intégrant des techniques multiples et complémentaires, reposant sur des bases théoriques solides, permettant de conduire des données brutes à leur interprétation. Les contributions comportent donc à la fois des aspects théoriques (représentation et modélisation des données et des connaissances, traitement, analyse, interprétation et raisonnement sur l'espace, au carrefour des mathématiques appliquées et de l'informatique, en particulier en IA), des aspects algorithmiques permettant de mettre en œuvre ces modèles sur des données volumineuses et complexes, et des applications dans des domaines de pointe, en particulier en imagerie médicale.

Dans le domaine de l'IA pour l'interprétation d'images médicales, les contributions prin-

cipales concernent le raisonnement spatial dans les images et l'apprentissage, pour la segmentation et la reconnaissance de structures, l'analyse de formes et l'anatomie computationnelle. Dans le premier domaine, des contributions en représentation des connaissances et des informations extraites des images ont permis de développer des méthodes de segmentation et de reconnaissance de structures dans les images guidées par des modèles. Un des points clés de ces modèles est la définition de représentations de connaissances structurelles, via la modélisation mathématique de relations spatiales combinant morphologie mathématique et ensembles flous, et permettant le raisonnement sur ce type de connaissances. Pour le raisonnement, le formalisme de la morphologie mathématique en logique (morpho-logique) a donné lieu à de nouvelles formulations de modes de raisonnement tels que la révision et l'abduction, en particulier dans le cadre de l'ANR LOGIMA, et en collaboration avec le LAMSADE (université Paris Dauphine), l'équipe MICS de Centrale-Supélec, et l'université Los Andes, Merida, Vénézuéla [1]. Ces travaux ont été appliqués à l'interprétation d'images médicales, par exemple en exprimant l'interprétation comme un problème d'abduction, permettant d'expliquer les observations (les images et des résultats de segmentation par exemple) à partir d'une base de connaissances, intégrant les modèles de relations spatiales obtenus par morphologie mathématique floue [2]. Ces représentations constituent un moyen de combler le fossé sémantique entre des connaissances ontologiques et des observations concrètes dans l'image, et prennent en compte l'imprécision intrinsèque de ce type de relations. Du point de vue théorique, ces travaux ont contribué à étendre le champ de la morphologie mathématique, au-delà de l'analyse d'images, et à la faire connaître dans d'autres domaines scientifiques.



**Afia**

Association française  
pour l'Intelligence Artificielle

Les approches structurelles ont été exploitées pour l'interprétation d'images (en imagerie médicale en particulier), avec le LAMSADE et l'université de São Paulo au Brésil (projets FA-PESP, ANR-FAPESP), et ont montré une plus grande stabilité que des méthodes exploitant surtout les formes des objets. Les approches de reconnaissance, segmentation et interprétation développées s'appuient sur une exploration séquentielle des images, à partir de l'optimisation de chemins dans un graphe de connaissances, ou encore sur des problèmes de satisfaction de contraintes (nettes ou floues), prenant en compte desinstanciations multiples d'objets dans la scène et des relations spatiales complexes.

Nos travaux actuels portent sur l'intégration de ces modèles dans des réseaux de neurones profonds, dans une approche d'IA hybride. Il s'agit ainsi de contraintes géométriques pour la segmentation de vaisseaux sanguins, d'architectures de réseaux reproduisant des modes de raisonnement des experts médicaux (par exemple à plusieurs échelles pour des images histologiques, ou encore exploitant les connaissances sur plusieurs modalités en IRM). Les applications en imagerie médicale, développées avec des doctorants et en collaboration avec des CHU ou des équipes industrielles, visent à lever des verrous importants, tels que développer des modèles mathématiques de représentation à la fois des connaissances et des informations issues de ces images, conduisant à des algorithmes opérationnels et efficaces pour l'interprétation des images; combler le fossé sémantique entre connaissances symboliques et percepts visuels; développer des méthodes traitant les cas pathologiques s'écartant significativement des modèles ou exemples normaux; développer les outils informatiques associés, répondre aux besoins des cliniciens, radiologues, chirurgiens, chercheurs en neurosciences. Parmi les projets importants actuelle-

ment, une collaboration de long terme avec le CHU du Kremlin-Bicêtre porte sur les images de nouveaux nés et de prématurés, pour la segmentation des principaux tissus du cerveau et la détection de zones hyper-intenses dans la matière blanche, en exploitant les connaissances anatomiques. Ces problèmes ont été abordés avec des méthodes d'apprentissage, par des réseaux convolutionnels pré-entraînés sur des bases classiques d'images, et intégrant des informations 3D au moment du transfert d'apprentissage (collaboration avec l'EPITA et avec l'université de Huazhong en Chine) [13]. Toujours en imagerie pédiatrique, une collaboration avec le CHU Necker Enfants Malades (équipe IMAG2) porte sur l'apport de l'imagerie dans la planification chirurgicale (tumeurs et déformations du pelvis chez l'enfant) [9]. D'autres projets sont réalisés en collaboration avec l'ISEP et l'hôpital des Quinze-Vingts sur l'imagerie de la rétine, avec des modalités très récentes telles que l'optique adaptative, dans le cadre d'un RHU porté par l'hôpital Lariboisière portant sur les maladies des petits vaisseaux sanguins [8]. Une collaboration avec une nouvelle équipe Inserm en imagerie neurologique à l'hôpital Sainte-Anne (Brain) porte sur les gliomes du cerveau [14]. Les approches actuellement développées portent sur l'association de données multi-modales inégalement disponibles dans des réseaux de type professeur-étudiant pour la segmentation des tumeurs, et sur le recalage pour la constitution d'un atlas. Nous travaillons aussi sur la reconnaissance de faisceaux de fibres de matière blanche, à partir de la description de leurs relations spatiales par rapport à des structures anatomiques, ainsi que sur leur recalage par des méthodes de transport optimal, sur leurs représentations multi-échelles, facilitant la visualisation et la navigation (projet DIGICOSME avec le LIX) [11]. Toujours sur le cerveau, nous avons contribué à un projet de segmentation multi-modale sur



des images issues de la machine TEP-IRM installée à Orsay [4]. Avec les industriels, les applications peuvent être aussi variées que l'amélioration d'images acquises par rayons X à faible dose, la tomosynthèse en mammographie, la constitution de jumeaux numériques, le transfert d'apprentissage pour la classification et la reconnaissance de structures anatomiques, et pour l'explicabilité [5, 6, 10]. Ce tournant vers des méthodes d'apprentissage profond répond à une demande des étudiants et des industriels, et bénéficie de bases d'images disponibles de plus en plus importantes. Nos projets en cours visent à combiner ces approches avec des méthodes structurelles, en particulier des représentations par graphes permettant de modéliser des connaissances spatiales (projet ANR-FAPESP avec le Brésil).

Dans la plupart des travaux ci-dessus, les modèles jouent un rôle clé pour la représentation des connaissances et leur utilisation en interprétation d'images. Inversement, les images peuvent aussi être utilisées pour créer des modèles anatomiques réalistes. C'est le cas des travaux avec l'hôpital Necker, où de tels modèles viennent en appui de la chirurgie. Dans des projets ANR et ANSES (nationaux et internationaux), nous avons aussi exploré cette piste pour construire des modèles anatomiques numériques 3D (voxélisés ou maillés). Associés à des modèles de déformation pour simuler différents âges et différentes positions, ces modèles ont été utilisés par nos partenaires dans ces projets pour mesurer l'influence des ondes électromagnétiques sur les tissus biologiques. Certains de ces modèles sont disponibles pour la recherche<sup>2</sup>. Ces travaux ont été réalisés en collaboration avec la chaire C2M.

Toutes ces recherches sont menées dans le cadre de projets collaboratifs (qui contribuent à leur financement) : thèses en convention CIFRE avec des industriels, projets ANR

(nationaux et internationaux), ANSES, RHU, DIM RFSI, Labex, etc.

La Chaire C2M de Télécom Paris étudie l'exposition des personnes aux ondes électromagnétiques. Après des travaux, mentionnés ci-dessus, portant sur la construction de modèles réalistes du corps humain et leur utilisation pour simuler la propagation des ondes et l'exposition, les études actuelles portent sur les méthodes numériques et expérimentales permettant la quantification de ces expositions, mais également sur la construction de modèles simplifiés (krigeage, expansion polynomiale...) permettant le calcul et la représentation de ces expositions [7]. Parmi les méthodes utilisées se trouvent les réseaux de neurones artificiels, soit pour aider à la représentation spatiale et temporelle de mesures opérées par des réseaux de sondes, soit pour créer des modèles de substitution dans le cas de grandes dimensions. Dans le premier cas, les architectures des réseaux de neurones artificiels sont étudiées pour utiliser au mieux les mesures des réseaux de capteurs et les informations existantes sur les infrastructures des réseaux de communication [12]. Dans le second cas, les réseaux de neurones sont utilisés pour réduire la dimension d'un problème afin de permettre la création de modèles de substitution dans le cas de problèmes en grande dimension [3].

## Références

- [1] Marc Aiguier, Jamal Atif, Isabelle Bloch, and Ramón Pino Pérez. Explanatory relations in arbitrary logics based on satisfaction systems, cutting and retraction. *International Journal of Approximate Reasoning*, 102 :1–20, 2018.
- [2] Jamal Atif, Céline Hudelot, and Isabelle Bloch. Explanatory reasoning for image understanding using formal concept analysis and description logics. *IEEE Transac-*

2. <http://femonum.telecom-paristech.fr>



- tions on Systems, Man and Cybernetics : Systems, 44(5) :552–570, May 2014.
- [3] Xi Cheng, Clément Henry, Francesco P. Andriulli, Christian Person, and Joe Wiart. A surrogate model based on artificial neural network for RF radiation modelling with high-dimensional data. *International Journal of Environmental Research and Public Health*, 17(7), 2020.
- [4] Olivier Commowick et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Nature Scientific Reports*, 8(13650) :1–17, 2018.
- [5] Vincent Couteaux et al. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. *Diagnostic and Interventional Imaging*, 100 :235–242, 2019.
- [6] Vincent Couteaux et al. Kidney cortex segmentation in 2D CT with U-Nets ensemble aggregation. *Diagnostic and Interventional Imaging*, 100 :211–217, 2019.
- [7] Pierrick Kersaudy, Bruno Sudret, Nadège Varsier, Odile Picon, and Joe Wiart. A new surrogate modeling technique combining Kriging and polynomial chaos expansions - Application to uncertainty analysis in computational dosimetry. *Journal of Computational Physics*, 286 :103–117, 2015.
- [8] Nicolas Lermé, Florence Rossant, Isabelle Bloch, Michel Paques, Edouard Koch, and Jonathan Benesty. A fully automatic method for segmenting retinal artery walls in adaptive optics images. *Pattern Recognition Letters*, 72 :72–81, 2016.
- [9] Cécile O Muller, et al. Integrating tractography in pelvic surgery : a proof of concept. *Journal of Pediatric Surgery Case Reports*, 48, 2019.
- [10] Antoine Pirovano, Hippolyte Heuberger, Sylvain Berlemont, Saïd Ladjal, and Isabelle Bloch. Automatic feature selection for improved interpretability on whole slide imaging. *Machine Learning and Knowledge Extraction*, 2021.
- [11] Corentin Mercier et al. QFib : Fast and efficient brain tractogram compression. *Neuroinformatics*, 2020.
- [12] Shanshan Wang and Joe Wiart. Sensor-aided EMF exposure assessments in an urban environment using artificial neural networks. *International Journal of Environmental Research and Public Health*, 17(9), 2020.
- [13] Yongchao Xu et al. The challenge of cerebral magnetic resonance imaging in neonates : A new method using mathematical morphology for the segmentation of structures including diffuse excessive high signal intensities. *Medical Image Analysis*, 48 :75–94, 2018.
- [14] Marc Zanello et al. Automated neurosurgical stereotactic planning for intraoperative use : a comprehensive review of the literature and perspectives. *Neurosurgical Review*, pages 1–22, 2020.



## ■ Intelligence embarquée en milieu contraint

LIP6/SYEL  
Sorbonne Université  
[lip6.fr/recherche/team.php?id=960](http://lip6.fr/recherche/team.php?id=960)

**Adrien UGON**  
[adrien.ugon@lip6.fr](mailto:adrien.ugon@lip6.fr)

**Andrea PINNA**  
[andrea.pinna@lip6.fr](mailto:andrea.pinna@lip6.fr)

**Sylvain FERUGLIO**  
[sylvain.feruglio@lip6.fr](mailto:sylvain.feruglio@lip6.fr)

**Bertrand GRANADO**  
[bertrand.granado@lip6.fr](mailto:bertrand.granado@lip6.fr)

**Orlando CHUQUIMIA**  
[orlando.chuquimia@lip6.fr](mailto:orlando.chuquimia@lip6.fr)

### Introduction

Les membres de l'équipe SYEL étudient à la fois les aspects fondamentaux et appliqués de la conception des systèmes embarqués hétérogènes. Les projets menés au sein de l'équipe couvrent à la fois des aspects capteurs (transducteur), les processeurs génériques et dédiés pour le calcul embarqué, l'optimisation de la consommation énergétique. La conception de Système-on-Chip à base d'architectures matérielles de traitement (reconfigurables et non) et leur programmation permet l'exploration des limites et verrous qui se positionnent à la frontière entre l'électronique, le traitement du signal et l'informatique.

Trois axes majeurs sont priorisés : la modélisation des systèmes sur puce hétérogènes (consommation, intégrité du signal), la conception d'architectures reconfigurables et l'intégration des modèles d'inférence d'intelligence artificielle sur des architectures embarquées (reconfigurables et non).

L'objectif de ces activités de recherche est la validation des nouveaux capteurs, des nouvelles architectures pour les systèmes embarqués et leur évaluation dans des domaines d'application comme le transport, les dispositifs médicaux et

la e-santé.

Les méthodes d'évaluation développées au sein de l'équipe aident à mettre en exergue les limites du système pour une application envisagée en termes de contraintes non fonctionnelles, telles que le temps réel et la faible consommation.

Ces limites et la manière de les surmonter posent de nouvelles questions qui alimentent nos recherches.

### Le projet Automated Embedded Polysomnography

Le projet *Automated Embedded Polysomnography* a pour objectif de concevoir un système intelligent capable d'analyser et d'interpréter les signaux assis lors d'une polysomnographie à l'aide de techniques de fusion symbolique. Cette analyse concerne à la fois la macrostructure du sommeil (stades de sommeil), les événements physiologiques — pathologiques ou non — survenant pendant le sommeil, mais également, grâce à une collaboration avec l'Université de Pise en Italie, la microstructure du sommeil. En particulier, un algorithme de reconnaissance des *Cycling Alternating Patterns* a été mis au point. Il s'agit d'événements reliés à la qualité du sommeil,



particulièrement fastidieux à reconnaître visuellement.

Le système décisionnel commence par segmenter les courbes physiologiques acquises et extrait des paramètres numériques, transformés en symboles par des règles logiques issues de la formalisation des guides de bonne pratique clinique d'interprétation des enregistrements polysomnographiques du sommeil. Le système expert prenant la décision intègre la dimension temporelle afin de relier des événements physiologiques non concomitants.

Plusieurs systèmes de règles ont été formalisés afin de reconnaître chacun des cinq stades de sommeil définis par les guides de bonne pratique internationaux en médecine du sommeil. Les conflits entre décisions sont réglés par un système de préférences [2]. Ils ont été testés sur 131 enregistrements polysomnographiques, ce qui a abouti à un taux de reconnaissance similaire à l'accord entre experts [15, 1].

Ces règles sont basées sur des paramètres et des seuils qui peuvent être optimisés et personnalisés pour chaque individu. Plusieurs systèmes d'apprentissage de seuils ont été mis au point permettant de personnaliser l'analyse automatique [4, 3].

## La capsule Cyclope

En France, le cancer colorectal est le deuxième cancer le plus meurtrier. Son dépistage est recommandé tous les deux ans pour les personnes âgées entre 50 et 74 ans. La capsule Cyclope est une capsule endoscopique qui enregistre des images haute définition d'une résolution d'un mégapixel. Elle est dotée d'une intelligence artificielle (forêts d'arbres flous [5, 6]) permettant d'effectuer un tri des images pour ne conserver que celles sur lesquelles un polype est détecté afin de les présenter ensuite aux gastro-entérologues. Cela permet un gain de temps car les capsules actuelles obligent les gastro-entérologues à visionner l'ensemble des

images de basse résolution et mauvaise qualité. Une des difficultés est d'intégrer toute l'électronique dans une capsule de 3 centimètres sur 1,4, mais aussi de trouver un système décisionnel dont la complexité a été suffisamment faible pour pouvoir être embarqué [7, 11, 12]. Lors des tests macroscopiques dans un intestin en PVC de 20 centimètres, Cyclope a reconnu les anomalies en forme de polype avec un taux de réussite d'environ 95 %.

## Le projet SpinalCOM

SpinalCOM est un projet de bio-ingénierie, à l'interface de l'informatique, de la physique, de l'architecture de traitement et de la médecine. Ce projet vise à concevoir un système embarqué innovant et peu coûteux, utilisant des composants optoélectroniques [8, 10, 9], répondant à de réels problèmes de santé publique et complétant les systèmes d'imagerie déjà utilisés. Il soulève des questions scientifiques et techniques qui n'ont pas encore été explorées, ou pas suffisamment.

Les pathologies de la colonne vertébrale sont des causes majeures de douleur, d'invalidité et de mortalité. Comme les techniques d'imagerie habituelles fournissent des informations limitées, il est nécessaire de disposer de méthodes permettant de surveiller, d'une part, l'état fonctionnel local de la moelle épinière en temps réel et, d'autre part, d'évaluer les conséquences fonctionnelles de ses lésions, l'effet des interventions thérapeutiques (médicaments ou manipulations) et des stimulations (électriques ou autres).

Pour répondre à ces besoins, SpinalCOM étudie une nouvelle approche d'imagerie de l'activité de la moelle épinière sur un modèle animal, d'abord, avant un transfert potentiel à l'être humain, par la réalisation d'un dispositif médical intelligent multimodal et de son logiciel associé. Cette approche permettra d'établir lo-



**Afia**

Association française  
pour l'Intelligence Artificielle

calement et spécifiquement l'état fonctionnel de la moelle épinière, ce qui constituera une avancée dans le traitement des pathologies de la moelle épinière. À court terme, nous nous concentrerons sur les animaux vertébrés afin de proposer un nouvel outil d'investigation pour les chercheurs et les vétérinaires. Il permettra d'acquérir des informations fondamentales précieuses pour toute la communauté médicale, telles que le rôle de la moelle épinière en interaction avec le cerveau. En ce qui concerne l'homme, l'outil développé pourrait également contribuer, grâce notamment à l'aide de l'intelligence artificielle, à la compréhension des mécanismes de blessure chez les vertébrés, à préserver la moelle épinière pendant la chirurgie et après un traumatisme et à améliorer l'utilisation de certains dispositifs utilisés en routine. Une étude de faisabilité a été conduite [13] et des résultats expérimentaux préliminaires ont été obtenus [14].

Une thèse est en cours sur les aspects de modélisation et de prototypage virtuel.

## Références

- [1] Ugon Adrien, Philippe Carole, Kotti Amina, Dalloz Marie-Amélie, and Pinna Andrea. Automatic sleep stages classification combining semantic representation and dynamic expert system. *Studies in Health Technology and Informatics*, 264 (MEDINFO 2019 : Health and Wellbeing e-Networks for All) :848–852, 2019.
- [2] Ugon Adrien, Sedki Karima, Kotti Amina, Seroussi Brigitte, Philippe Carole, Ganascia Jean-Gabriel, Garda Patrick, Bouaud Jacques, and Pinna Andrea. Decision system integrating preferences to support sleep staging. *Studies in Health Technology and Informatics*, 228 (Exploring Complexity in Health : An Interdisciplinary Systems Approach) :514–518, 2016.
- [3] Chen Chen, Adrien Ugon, Chenglu Sun, Wei Chen, Carole Philippe, and Andrea Pinna. Towards a hybrid expert system based on sleep event's threshold dependencies for automated personalized sleep staging by combining symbolic fusion and differential evolution algorithm. *IEEE Access*, 7 :1775–1792, 2019.
- [4] Chen Chen, Adrien Ugon, Xun Zhang, Amara Amara, Patrick Garda, Jean-Gabriel Ganascia, Carole Philippe, and Andrea Pinna. Personalized sleep staging system using evolutionary algorithm and symbolic fusion. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, August 2016.
- [5] Orlando Chuquimia, Andrea Pinna, Xavier Dray, and Bertrand Granado. Polyps Recognition Using Fuzzy Trees. In *BHI-2017 International Conference on Biomedical and Health Informatics*, pages 9–12, Orlando, FL, United States, February 2017. IEEE.
- [6] Orlando Chuquimia, Andrea Pinna, Xavier Dray, and Bertrand Granado. Polyp follow-up in an intelligent wireless capsule endoscopy. In *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, October 2019.
- [7] Orlando Chuquimia, Andrea Pinna, Christophe Marsala, Xavier Dray, and Bertrand Granado. FPGA-based Real Time Embedded Hough Transform Architecture for circles detection. In *DASIP 2018 - Conference on Design and Architectures for Signal and Image Processing*, Porto, Portugal, October 2018.
- [8] Sylvain Feruglio, Thierry Courcier, Armine Karami, Annick Alexandre-Gauthier, Olivier Romain, Vincent Aimez, Paul G. Charette, Patrick Pittet, and Guo-Neng Lu. Opto-electrical modeling of CMOS buried



- quad junction photodetector. *Key Engineering Materials*, 605 :470–473, April 2014.
- [9] Sylvain Feruglio, Thierry Courcier, Olivier Tsiakaka, Armine Karami, Annick Alexandre-Gauthier, Olivier Romain, Vincent Aimez, Paul G. Charette, Patrick Pittet, and Guo-Neng Lu. A CMOS buried quad p-n junction photodetector model. *IEEE Sensors Journal*, 16(6) :1611–1620, March 2016.
- [10] Guo-Neng Lu, Thierry Courcier, Bastien Mamdy, Sylvain Feruglio, Paul G. Charette, Vincent Aimez, Olivier Romain, and Patrick Pittet. CMOS buried multi-junction (BMJ) detector for bio-chemical analysis. In Haimei Gong, Nanjian Wu, Yang Ni, Weibiao Chen, and Jin Lu, editors, *AOPC 2015 : Optical and Optoelectronic Sensing and Imaging Technology*. SPIE, October 2015.
- [11] Chuquimia Orlando, Pinna Andrea, Dray Xavier, and Granado Bertrand. A real time hough transform architecture useable inside a WCE. In *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, October 2019.
- [12] Chuquimia Orlando, Pinna Andrea, Dray Xavier, and Granado Bertrand. A low power and real-time architecture for hough transform processing integration in a full HD-wireless capsule endoscopy. *IEEE Transactions on Biomedical Circuits and Systems*, 14(4) :646–657, August 2020.
- [13] Olivier Tsiakaka and Sylvain Feruglio. Toward the monitoring of the spinal cord : A feasibility study. *Microelectronics Journal*, 88 :145–153, June 2019.
- [14] Olivier Tsiakaka, Benoit Gosselin, and Sylvain Feruglio. Source–detector spectral pairing-related inaccuracies in pulse oximetry : Evaluation of the wavelength shift. *Sensors*, 20(11) :3302, June 2020.
- [15] Adrien Ugon, Amina Kotti, Brigitte Séroussi, Karima Sedki, Jacques Bouaud, Jean-Gabriel Ganascia, Patrick Garda, Carole Philippe, and Andrea Pinna. Knowledge-based decision system for automatic sleep staging using symbolic fusion in a turing machine-like decision process formalizing the sleep medicine guidelines. *Expert Systems with Applications*, 114 :414–427, December 2018.



## ■ L'équipe-projet HeKA

*Équipe-projet HeKA/ Équipe 22  
Centre de Recherche des Cordeliers,  
Inria Paris, Inserm et Université de Paris  
[team.inria.fr/heka](http://team.inria.fr/heka)*

**Adrien COULET,**  
**Stéphanie ALLASSONNIERE,**  
**François ANGOULVANT,**  
**Anita BURGUN,**  
**Xiaoyi CHEN,**  
**David DRUMMOND,**  
**Nicolas GARCELON,**  
**Anne-Sophie JANNOT,**  
**Sandrine KATSAHIAN,**  
**Antoine NEURAZ,**  
**Bastien RANCE,**  
**Brigitte SABATIER,**  
**Rosy TSOPRA,**  
**Moreno URSINO,**  
**Sarah ZOHAR**  
[adrien.coulet@inria.fr](mailto:adrien.coulet@inria.fr),  
[anita.burgun@aphp.fr](mailto:anita.burgun@aphp.fr),  
[sarah.zohar@inserm.fr](mailto:sarah.zohar@inserm.fr)

### Introduction

**HeKA** est une équipe-projet de recherche commune à Inria, l'Inserm et l'Université de Paris. Plus précisément, HeKA dépend du Centre de Recherche des Cordeliers et du Centre Inria de Paris. En plus de deux chercheurs Inria et Inserm, HeKA est composé de chercheurs hospitalo-universitaires de l'AP-HP associés à des services de l'hôpital Européen Georges Pompidou, l'hôpital Necker et de l'Institut Imagine. Les thèmes de recherche de l'équipe sont l'**informatique médicale**, les **biostatistiques** et les **mathématiques appliquées** pour l'**aide à la décision clinique**. Le terme HeKA est à la fois une référence à la divinité égyptienne de la médecine et un acronyme pour *Health data- and model- driven Knowledge Acquisition*. L'équipe HeKA fait suite à l'équipe 22 (*Information Sciences to support Personalized Medicine*) dirigée par Anita Burgun au Centre de Recherche des Cordeliers (Inserm, Université de Paris). La responsable de HeKA est Sa-

rah Zohar, elle est secondée par Adrien Coulet.

### Vers un système de santé apprenant

L'objectif partagé au sein de l'équipe est le développement de méthodes, modèles et outils pour un **système de santé apprenant** [8]. Ce paradigme, que nous développons en particulier pour les maladies rares dans le RHU C'IL-LICO [1, 7] tire parti des données générées au cours du soin pour apprendre de nouvelles connaissances, qui sont à leur tour utilisées pour guider la pratique clinique, de façon continue. Pour atteindre ce but, HeKA s'intéresse à trois axes de recherche très liés : (1) l'extraction de connaissances à partir des données de santé et notamment le phénotypage profond ; (2) les approches stochastiques et supervisées pour l'aide à la décision ; et (3) les essais cliniques du futur et leur design, qui permettent l'évaluation des systèmes d'aide à la décision médicale. Les sections qui suivent décrivent brièvement ces axes et les illustrent avec des projets récents.



## Axe 1 : l'extraction de connaissances à partir de données de santé

Au sein de l'axe 1, nous nous attachons au développement de méthodes et d'outils pour tirer parti au mieux des données de patient·e·s, et ce malgré leur nature hétérogène et complexe (structurée vs. non-structurée, temporelle, incomplète, etc.). Nous étudions l'extraction et la transformation des données brutes, et notamment du texte clinique, en des descripteurs plus élaborés ou en des représentations apprises, qui facilitent le développement sous-jacent de systèmes d'aide à la décision ou de découverte de connaissances, comme ceux développés dans les axes 2 et 3.

**Le projet Dr Warehouse.** En collaboration avec l'institut Imagine et l'hôpital Necker-Enfants Malades, et pour faciliter la réutilisation des données hospitalières, nous avons conçu l'entrepôt de données Dr Warehouse centré sur le document clinique [6]. À travers trois cas d'usage, nous avons tenté d'adresser les problématiques inhérentes aux données textuelles : (i) le recrutement de patient·e·s à travers un moteur de recherche adapté aux données textuelles (notamment en traitant la négation et les antécédents familiaux), (ii) le phénotypage automatisé à partir des données textuelles et (iii) l'aide au diagnostic par similarité entre patient·e·s, basée sur le texte [5]. Ce projet a fait l'objet de la création de la start-up Codoc, en charge de l'installation de Dr Warehouse dans d'autres hôpitaux (licence *open source*).

**Les projets TALONCO et TALREP.** Il s'agit de deux projets d'extraction de connaissances, notamment à partir de textes en lien avec l'oncologie et le programme de recherche **CARPEN** (*Cancer Research and Personalized Medicine*) de l'AP-HP et Université de Paris. **TALONCO** vise à l'extraction d'un en-

semble d'informations (localisation, classification histologique, stade, etc.) pour la constitution automatique de "fiches de synthèses" de patient·e·s qui facilitent les réunions de concertations pluridisciplinaires en oncologie. **TALREP** vise quant à lui à l'extraction, à partir des dossiers patient·e·s informatisés (DPI), de la réponse clinique aux chimiothérapies (réponse positive, absence de réponse, effets indésirables). Cette première étape, indispensable, nous permettra ensuite de développer des modèles prédictifs de la réponse à de tels traitements.

### Les projets PractiKPharma et PyMedExt.

PractiKPharma est un projet ANR qui vise la comparaison des connaissances synthétisées dans l'état de l'art (bases de données expertes et littérature) avec celles qui peuvent être extraites des dossiers patients informatisés (DPI) [2]. Nous avons développé pour cela des méthodes permettant d'assurer la qualité et la reproductibilité de l'extraction de connaissances à partir des DPI, notamment en proposant la réutilisation de bibliothèques standards et indépendantes de l'environnement [3]. Le besoin de manipuler le texte clinique dans PractiKPharma et dans d'autres projets nous a amenés à développer la bibliothèque libre **PyMedExt** qui facilite leur transformation, échange et annotation [4].

## Axe 2 : les approches stochastiques et supervisées pour l'aide à la décision

L'axe 2 vise à proposer des méthodes originales d'apprentissage statistique ou automatique, notamment dans le contexte particulier de jeux de données de grande dimension, mais à faible effectif. En effet, même lorsque l'on a la chance d'avoir des données pour un grand nombre de patient·e·s, la sélection des cas pertinents dans un objectif d'aide à la décision diagnostique ou thérapeutique réduit



drastiquement le nombre d'individus. Dans ce contexte, nous prêtons une attention particulière à la modélisation du parcours de soin (ou trajectoire) des patient·e·s, ce qui intègre une dimension temporelle au sein de nos modèles. Enfin, la génération de données synthétiques et leur prise en compte dans nos modèles doit permettre de mieux les évaluer en augmentant artificiellement notre échantillonnage.

**Le projet AntibioHelp®.** Dans le cadre du projet RaMiPA (“Raisonnement pour Mieux Prescrire les Antibiotiques”), financé par l’ANSM, nous avons développé AntibioHelp®, un système d’aide à la décision en antibiothérapie empirique. Ce système est capable de retrouver les antibiotiques recommandés dans les guides de bonnes pratiques cliniques, à partir des propriétés pondérées des antibiotiques. AntibioHelp® fournit des recommandations de prescription, notamment pour des situations rares ou non décrites dans les guides [9]. Afin de personnaliser les recommandations, nous travaillons à la mise en place de flux de données à partir d’un entrepôt de données cliniques, dans l’idée de générer, pour un·e patient·e donné·e, un score prédictif d’efficacité pour chaque antibiotique.

**Le projet Lights.** Nous développons “Lights”, un modèle de survie qui intègre les variables longitudinales de grande dimension en petit effectif dans le cadre du projet InCa “Thérapies personnalisées en oncologie pour les cancers métastatiques dans les populations asiatiques et caucasiennes : étude de transition réutilisant les dossiers patients informatisés”. Lights permet d’identifier automatiquement les variables prédictives du pronostic à partir de données longitudinales de grande dimension, ainsi que des données expertes. Autrement dit, Lights permet d’identifier au cours d’un parcours de soin, les éléments les plus importants pour le pronostic des patient·e·s. Le développement théorique

de cette méthode a déjà fait l’objet de deux communications orales, dont l’une au *Joint Statistical Meetings (JSM) - American Statistical Association* en 2020.

**Le projet Data Augmentation.** Deux problèmes récurrents lorsque l’on travaille avec des données médicales sont le petit nombre d’exemples et leur grande dimension. En utilisant des *Variational Auto Encoders* (VAE) dont l’espace latent est muni d’une structure de variété riemannienne, nous pouvons résoudre les deux problèmes en un seul modèle génératif. Les VAE permettent de réduire la dimension des données. La structure riemannienne permet d’apprendre une loi de probabilité sur l’espace latent qui permet de gérer de nouvelles données beaucoup plus informatives que les alternatives de la littérature. Ce travail fait l’objet d’une thèse de mathématiques appliquées débutée en 2020, mêlant statistiques computationnelles et géométrie riemannienne.

### **Axe 3 : les essais cliniques du futur et leur design**

Le premier objectif de cet axe est de proposer des méthodes d’évaluation pour des outils logiciels en tant que dispositifs médicaux, et dans ce cadre de proposer des designs d’essais cliniques qui soient adaptés aux méthodes d’apprentissage continu. Son second objectif consiste à développer des modèles statistiques ou d’apprentissage automatique qui participent à la construction d’essais cliniques à partir de sources de connaissances extérieures comme des modèles de maladies, des modèles pré-cliniques, des données expertes, des DPI, des patient·e·s synthétiques. Il s’agit dans ce cas d’optimiser les futurs essais cliniques pour faciliter l’acquisition de nouvelles connaissances biomédicales.



**AfIA**

Association française  
pour l'Intelligence Artificielle

### Les projets européens ITFoC et PeCan.

Dans le cadre des projets européens Flag-Era ITFoC (*Information Technology for the Future of Cancer Treatment*) et Era PerMed Pe-Can (*Parametrisation of large scale cancer models for personalised therapy of triple negative breast*), nous développons une méthodologie de validation précoce des algorithmes d'intelligence artificielle à partir des données de vie réelle, avant leur implantation dans le processus de soins. 'A partir de cette méthodologie, nous allons mettre en place un environnement d'expérimentation pour valider les algorithmes prédictifs de réponse au traitement des patientes atteintes de cancer du sein triple négatif. Cet environnement sera construit en collaboration avec le CEA, l'institut Curie, le CHU de Rennes, et le centre Unicancer Eugène Marquis.

### Le projet collaboratif européen FAIR.

Dans le cadre du projet européen FAIR (*Flagellin Aerosol therapy as an Immunomodulatory adjunct to the antibiotic treatment of drug-Resistant bacterial pneumonia*), nous développons une plateforme de simulation translationnelle prenant en compte des modèles pré-cliniques (cellulaires et animaux) ainsi que des données expertes afin de construire de manière séquentielle un modèle physiologique. Cette approche de modélisation, par extrapolation et apprentissage par transfert, intègre différentes sources d'informations et va permettre de mettre en évidence le design optimal pour les essais cliniques évaluant les effets de la flagelline chez l'humain.

En conclusion, HeKA est une équipe interdisciplinaire, en lien étroit avec le monde hospitalier, qui vise des contributions méthodologiques en science des données, capables d'impacter la pratique clinique.

### Références

- [1] Xiaoyi Chen *et al.* Phenotypic similarity for rare disease : Ciliopathy diagnoses and subtyping. *Journal of Biomedical Informatics*, 100 :103308, 2019.
- [2] Adrien Coulet and Malika Smaïl-Tabbone. Mining Electronic Health Records to Validate Knowledge in Pharmacogenomics. *ERCIM News*, 104 :56, January 2016.
- [3] William Digan *et al.* Can reproducibility be improved in clinical natural language processing ? A study of 7 clinical NLP suites. *Journal of the American Medical Informatics Association*, December 2020.
- [4] William Digan *et al.* Pymedext - a library to process clinical text : [https://github.com/equipe22/pymedext\\_core/](https://github.com/equipe22/pymedext_core/), 2020.
- [5] Nicolas Garcelon *et al.* Finding patients using similarity measures in a rare diseases-oriented clinical data warehouse : Dr. warehouse and the needle in the needle stack. *Journal of Biomedical Informatics*, 73 :51–61, 2017.
- [6] Nicolas Garcelon *et al.* A clinician friendly data warehouse oriented toward narrative reports : Dr. warehouse. *Journal of Biomedical Informatics*, 80 :52–63, 2018.
- [7] Nicolas Garcelon *et al.* Electronic health records for the diagnosis of rare diseases. *Kidney International*, 97(4) :676–686, 2020.
- [8] Leigh Anne Olsen *et al.* The learning healthcare system : workshop summary, 2007.
- [9] Rosy Tsopra *et al.* Helping GPs to extrapolate guideline recommendations to patients for whom there are no explicit recommendations, through the visualization of drug properties. The example of AntibioHelp® in bacterial diseases. *Journal of the American Medical Informatics Association*, 26(10) :1010–1019, 05 2019.



## ■ Apprentissage automatique et graphes spatio-temporels pour l'analyse d'images médicales

*ICube/SDC*  
Université de Strasbourg  
[sdc.icube.unistra.fr](mailto:sdc.icube.unistra.fr)

**Thomas LAMPERT**

[lampert@unistra.fr](mailto:lampert@unistra.fr)

**Florence LE BER**

[florence.le-ber@unistra.fr](mailto:florence.le-ber@unistra.fr)

**Aurélien LEBORGNE**

[aurelie.leborgne@unistra.fr](mailto:aurelie.leborgne@unistra.fr)

**Stella MARC-ZWECKER**

[stella@unistra.fr](mailto:stella@unistra.fr)

**Cédric WEMMERT**

[wemmert@unistra.fr](mailto:wemmert@unistra.fr)

### Introduction

L'équipe "Science des Données et Connaissances" couvre un large spectre de recherches en informatique, et plus précisément en intelligence artificielle. Nos activités de recherche s'articulent autour de deux thèmes de recherche théoriques (l'apprentissage automatique et la modélisation de connaissances) et de plusieurs domaines d'application privilégiés dont l'analyse de données d'imagerie médicale.

Nous présentons ici une partie de nos travaux concernant deux modalités d'images médicales sur lesquelles se concentre une grande partie de l'équipe : le *deep learning* pour l'imagerie histopathologique et les graphes spatio-temporels pour l'analyse d'images IRMf.

### Deep learning pour l'imagerie histopathologique

La pathologie numérique a révolutionné le domaine de l'anatomopathologie en démocratisant l'utilisation de scanners permettant d'obtenir des images microscopiques à très haute résolution de lames histologiques. Cela permet de développer des algorithmes d'analyse de ce type d'images permettant d'aider le patholo-

giste dans sa tâche de diagnostic.

C'est dans ce cadre que notre équipe s'intéresse à toutes les problématiques liées à ces données complexes. En effet, ces images sont très hétérogènes, en termes de signal (colorations multiples, variabilité intercentre) et de sémantique (types de cellules, etc.), causant des difficultés pratiques de traitement par des algorithmes classiques. Les principales difficultés pour les analyser sont leur très grande taille, le manque de données étiquetées, les différents niveaux de résolution, la nature même des images (qui sont plus proches d'images de texture que d'images naturelles) et les variations entre les différentes colorations. De nombreux travaux et projets ont été menés dans notre équipe autour de cette thématique [1, 2, 7] et notre recherche se concentre actuellement sur la réduction du besoin d'annotations.

Une première idée consiste à développer des approches non supervisées pour la régularisation des réseaux profonds en utilisant l'augmentation des données. Par exemple, dans les images histopathologiques avec différentes colorations, nous pouvons utiliser les annotations d'une coloration pour apprendre un modèle invariant en utilisant le transfert de coloration via

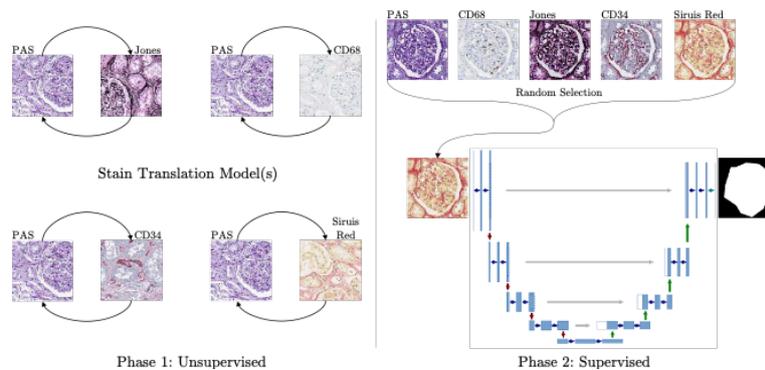


Figure 1.1 – Schéma de principe de l'utilisation de réseaux génératifs pour le transfert de domaine en analyse multi-coloration des images histopathologiques. Phase 1 : entraînement de GANs pour la génération d'images de colorations non étiquetées - Phase 2 : utilisation de ces nouvelles données pour l'entraînement supervisé d'un réseau de type U-Net pour la segmentation automatique de glomérules

des réseaux adversaires générateurs (GAN). La seconde stratégie consiste à mieux utiliser une quantité limitée de données annotées en développant des stratégies d'auto-régularisation pour parvenir à un apprentissage avec peu de données. Enfin, lorsque des étiquettes existent mais dans des modalités de données différentes, par exemple des modalités de balayage, de résolutions, etc., les approches "standard-s" ne peuvent pas être utilisées, et nous avons donc développé un cadre contradictoire grâce auquel les représentations invariantes de domaines multi-modèles peuvent être apprises de manière non supervisée [3, 8].

Ces stratégies (Figure 1.1) permettent de réutiliser les données étiquetées pour des problèmes qui n'auraient pas été possibles auparavant, d'augmenter les performances en encourageant l'apprentissage de caractéristiques plus robustes et d'utiliser l'apprentissage sur des données provenant de différents capteurs. Il s'agit donc de contribuer à prévenir leur obsolescence et de tirer un meilleur parti de la tâche coûteuse que représente l'étiquetage, tout en développant des stratégies visant à construire des modèles plus robustes et généralisables.

## Graphes spatio-temporels pour l'analyse de données IRMf

Notre équipe travaille aussi (en collaboration avec l'équipe [IMAGeS](#) dans l'objectif d'analyser les données issues de l'imagerie par Résonance Magnétique fonctionnelle (IRMf). Cette modalité d'imagerie permet d'explorer *in vivo* le fonctionnement cérébral, selon quatre dimensions, trois spatiales et une temporelle. Les approches mises en œuvre exploitent les graphes comme outils de représentation des corrélations spatiales et temporelles détectées par IRMf et les techniques de fouille de graphes pour extraire des réseaux fonctionnels caractéristiques.

D'un point de vue applicatif, nous nous intéressons à l'analyse des réseaux fonctionnels entre les différentes régions cérébrales de deux populations de souris (une malade, une témoin) : les données de chaque examen sont moyennées, puis représentées par des graphes qui sont fouillés pour extraire les réseaux caractéristiques des deux populations [4, 5]. Nous cherchons aussi à faciliter la compréhension des données d'un examen par une représentation sous forme de graphe spatio-temporel qui met

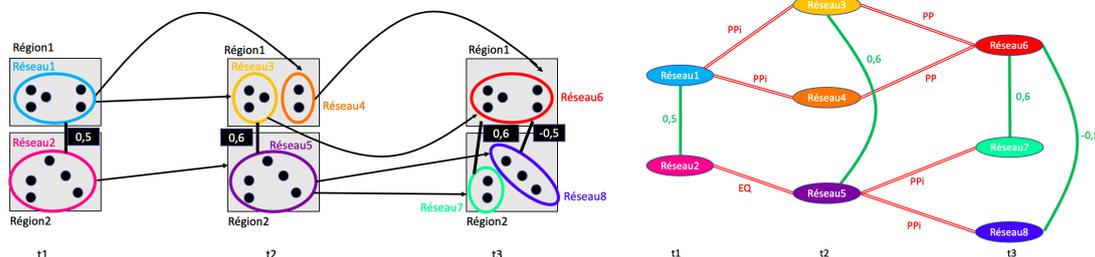


Figure 1.2 – Un exemple de graphe construit à partir de données IRMF. À gauche, les données sur les réseaux au cours du temps, à droite, le graphe correspondant.

en évidence les évolutions des corrélations cérébrales (Figure 1.2) [6].

## Références

- [1] A. Grote, N. Schaadt, G. Forestier, C. Wemmert, and F. Feuerhake. Crowd-sourcing of histological image labeling and object delineation by medical students. *IEEE Transactions on Medical Imaging*, 2018.
- [2] J. Krüger, C. Wemmert, L. Sternberger, C. Bonnas, G. Dietmann, P. Gançarski, and F. Feuerhake. Combat or surveillance? evaluation of the heterogeneous inflammatory breast cancer microenvironment. *The Journal of Pathology*, 259(4) :569–578, Nov 2013.
- [3] T. Lampert, O. Merveille, J. Schmitz, G. Forestier, F. Feuerhake, and C. Wemmert. Strategies for training stain invariant cnns. In *16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 905–909. IEE, Apr 2019.
- [4] A. Leborgne, F. Le Ber, J. Degiorgis, L. Harsan, S. Marc-Zwecker, and V. Noblet. Analysis of brain functional connectivity by frequent pattern mining in graphs. application to the characterization of murine models. In *ISBI*, April 2021.
- [5] A. Leborgne, F. Le Ber, L. Degiorgis, J. Massart, L. Harsan, S. Marc-Zwecker, and V. Noblet. Analyse de la connectivité fonctionnelle cérébrale de modèles murins par extraction de motifs fréquents dans des graphes. In *EGC*, Jan 2021.
- [6] A. Leborgne, F. Le Ber, D. Niezgoda, C. Meillier, and S. Marc-Zwecker. Utilisation des graphes pour la représentation spatio-temporelle lors d'un examen d'irm fonctionnelle cérébrale. In *Journée Santé & IA 2020 (dans le cadre PFIA)*, Jun 2020.
- [7] N. Schaadt, J. Alfonso, R. Schönmeier, A. Grote, G. Forestier, C. Wemmert, N. Krönke, M. Stoeckelhuber, H. Kreipe, H. Hatzikirou, and F. Feuerhake. Image analysis of immune cell patterns in the human mammary gland during the menstrual cycle refines lymphocytic lobulitis. *Breast Cancer Research and Treatment*, pages 1–11, Apr 2017.
- [8] J. Vasiljevic, F. Feuerhake, C. Wemmert, and T. Lampert. Self adversarial attack as an augmentation method for immunohistochemical stainings. In *IEEE International Symposium on Biomedical Imaging*, Jan 2021.



## ■ LACODAM : Science des données appliquée aux parcours de soins

IRISA-Inria/LACODAM  
Institut Agro  
<https://team.inria.fr/lacodam/>

**Thomas GUYET**  
[thomas.guyet@irisa.fr](mailto:thomas.guyet@irisa.fr)

### Introduction

L'équipe **LACODAM** (Large Scale Collaborative Data Mining) développe des travaux à la frontière entre l'analyse de données et l'intelligence artificielle (IA). Cette équipe s'intéresse en particulier à faciliter l'émergence de sens à partir de volumes importants de données. Elle développe pour cela des méthodes automatiques pour découvrir des structures dans les données, et elle s'intéresse également aux *workflows* de découverte de connaissances et à l'implication des utilisateurs dans ces *workflows*.

Les travaux appliqués à la santé de LACODAM concernent en particulier la modélisation et l'analyse de parcours de soins à partir de données médico-administratives. L'exploitation de ces données fait l'objet d'un intérêt croissant notamment avec le déploiement de la *plateforme des données de santé* (ou *Health Data Hub*) mis en place par le **SNDS**.

Les données médico-administratives sont des données collectées pour des raisons de gestion administrative des soins (e.g. données de l'assurance maladie, données hospitalières). Les travaux qui sont menés chez LACODAM visent à développer des outils pour exploiter ces informations à des fins médicales, et en particulier épidémiologiques. L'enjeu est de faciliter leur exploitation par des médecins-chercheurs. Le défi principal est de combler le saut sémantique qui existe entre les données brutes et l'information à en extraire. Ceci nous amène à développer des approches originales combinant raisonnement et science des données.

Nous nous intéressons en particulier à la di-

mension temporelle des données qui sont collectées et qui donnent une vision longitudinale d'un parcours patient pendant un séjour hospitalier, ou sur une période plus longue comme le permettent les données de l'assurance maladie.

Nous présentons dans la suite quatre travaux passés et en cours relatifs à ces défis :

1. Un *workflow* pour la conduite d'étude épidémiologique à partir des données de l'assurance maladie (projet ANSM/PEPS),
2. Le phénotypage de patients à partir des données hospitalières (chaire APHP/Inria/CS),
3. La génération de données synthétiques,
4. Les questions de responsabilité juridique liées à l'utilisation des outils d'IA en santé (projet CNRS/DRIAS).

### Épidémiologie et bases de données médico-administratives

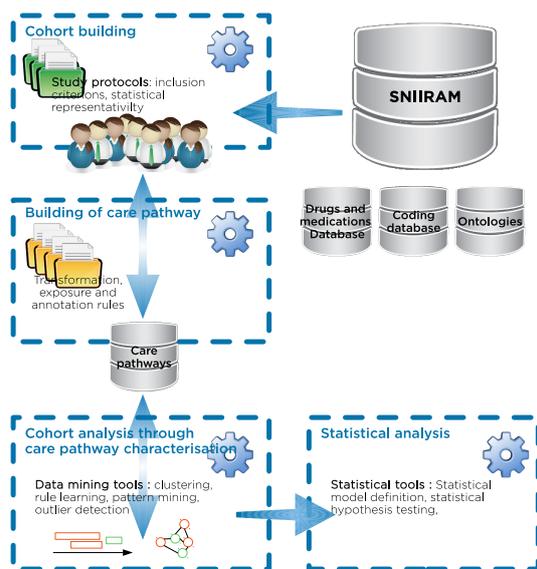
*Ce travail a été notamment mené en collaboration avec les équipes IRISA-Inria/DRUID, IRISA-Inria/DYLISS et l'unité REPERES (CHU Rennes/EHESP) dans le cadre du projet Pharmaco-Épidémiologie des Produits de Santé (PEPS) financé par l'ANSM (2015-2018).*

Une étude épidémiologique étudie la survenue d'événements médicaux en vie réelle. Par exemple, dans l'étude GenEpi [8], il s'agit d'étudier un lien éventuel entre un changement dans la prescription de médicaments anti-épileptiques et la survenue de crises.

Dans une telle étude, des personnes qui partagent des caractéristiques communes sont recrutées. Ensuite, des données significatives (ex-



positions aux médicaments, procédures médicales, survenue d'un événement médical, etc.) sont recueillies auprès de la population échantillonnée dans une période de temps définie par l'étude. Enfin, une analyse statistique met en évidence les liens (ou l'absence de lien) entre une exposition à un traitement et la survenue d'un événement d'intérêt (par exemple, les effets indésirables). Le principal inconvénient des études est le temps nécessaire à la collecte des données nécessaires à l'étude statistique : il faut en effet prendre en compte le temps de recrutement des patients d'une étude, puis le temps d'observation nécessaire à la collecte des données.



L'utilisation de bases de données médico-administratives (BDMA) se montre alors utile pour répondre rapidement aux questions épidémiologiques. Les données ont déjà été collectées (données historisées) et elles couvrent une large population. La couverture large d'une population permet d'obtenir des résultats représentatifs de la population et elle permet des études sur des phénomènes peu fréquents lorsque la population est grande. C'est le cas en particulier de la base de l'assurance maladie (nommée SNIIRAM).

La figure ci-contre présente les principales étapes de réalisation d'une étude à partir d'une BDMA [4] qui sont décrites ci-dessous :

**Construction d'une cohorte.** Un ensemble de patients doivent être inclus dans l'étude. La principale difficulté pour les cliniciens est d'exprimer des critères d'inclusion et d'exclusion (basés sur les données brutes disponibles) qui soient suffisamment précis pour identifier exhaustivement les patients d'intérêt. Mais identifier un patient avec une pathologie donnée (e.g. épilepsie) uniquement à partir de ses prestations de soin n'est pas une chose aisée.

Nous cherchons actuellement à proposer des outils pour exprimer des critères d'inclusion exploitant les dimensions temporelle et ontologique des données [1]. On ne s'intéresse plus seulement à la présence ou à l'absence d'un événement médical, mais à des motifs temporels dans la séquence des soins des patients qui définissent un parcours de soin identifiant un type de patients. La combinaison de techniques du Web Sémantique à des modèles temporisés permet d'exploiter la sémantique riche dans des temps raisonnables sur des données massives.

**Construction des parcours de soins.** Cette étape vise à donner une vue abstraite des patients afin : 1) de combler partiellement le fossé sémantique entre les données administratives et la question médicale, et 2) d'élaguer les données que le clinicien a supposé sans incidence sur l'étude. Cette transformation peut être mise en œuvre sous forme d'algorithmes ou de règles de transformation. Par exemple, les délivrances de médicaments sont transformées en expositions aux médicaments. De nouveau, les motifs temporels sémantiquement riches aident à formaliser ces transformations.

**Caractérisation des parcours de soins.** Cette étape consiste à utiliser des algo-



**AfIA**

Association française  
pour l'Intelligence Artificielle

rithmes d'exploration de données supervisés et non supervisés pour extraire des caractérisations de trajectoires de soins, et plus particulièrement sous la forme de motifs temporels. Nous avons plus particulièrement proposé une méthode d'extraction de chroniques temporelles discriminantes [3]. Une chronique représente un ensemble d'événements liés par des contraintes temporelles numériques qui sont découvertes à partir des données. Les chroniques extraites peuvent ainsi aider le clinicien à générer de nouvelles hypothèses (i.e. possibles événements adverses).

**Analyse des données** (analyse statistique, fouille de données, etc.). Celle-ci est finalement effectuée pour évaluer l'association entre le résultat et les covariables.

Une étude épidémiologique à partir de données médico-administratives devient ainsi une chaîne d'analyse de données (*workflow*). Nous proposons alors de faire le parallèle avec le domaine de l'*automatic data science* qui développe des approches et outils pour automatiser partiellement les analyses de données. De la même manière, les épidémiologistes ont besoin d'outils pour faciliter la conduite de ces études. Dans une première approche, l'outil que nous avons développé [5] représente formellement une étude épidémiologique, l'exécute et compare différents résultats obtenus pour différentes chaînes de traitements. Dans une seconde approche [7], nous avons exploité la capacité d'un formalisme de programmation déclarative (ASP – *Answer Set Programming*) pour intégrer raisonnement et extraction de motifs dans un même outil.

## Phénotypage temporel de patients

*Ce travail est mené en collaboration avec l'APHP et Central Supelec dans le cadre de la Chaire APHP/Inria/CS (2020-2024).*

L'APHP est un regroupement des hôpitaux franciliens qui constitue l'une des plus grandes structures hospitalières européennes. L'APHP dispose d'un entrepôt de données de santé (EDS) qui comporte les informations administratives et médicales des patients effectuant un séjour dans l'un de ses hôpitaux. De la même manière que pour les données du SNIIRAM, nous nous intéressons à représenter et exploiter les trajectoires de soins ou le déroulement d'un séjour hospitalier. Contrairement aux données du SNIIRAM qui sont uniquement des données structurées, les données de l'APHP comportent également des comptes-rendus médicaux, des images ou des résultats de laboratoire. Ces informations sont riches et doivent être pré-traitées pour en tirer une information structurée et datée. L'ensemble de ces informations constitue un parcours longitudinal de soins.

Dans le cadre de notre collaboration avec l'APHP, l'objectif est d'identifier des groupes homogènes de patients qui partagent les mêmes types de trajectoire, et de caractériser ces trajectoires. On parle de phénotypage temporel des patients. De nouveau, la dimension temporelle est au cœur de nos travaux. À cette fin, nous explorons des méthodes d'extraction de motifs temporels et d'apprentissage automatique.

Les travaux en cours sont menés sur la caractérisation des trajectoires de soins de patients hospitalisés lors de la première vague du COVID19 (projet CoviPreds). L'objectif est de décrire de manière globale les prises en charge pour notamment en faire une analyse rétrospective.

## Génération d'un SNIIRAM synthétique

Comme indiqué précédemment, le SNIIRAM [2] est une base de données contenant des informations riches pour répondre à de nombreuses questions épidémiologiques et

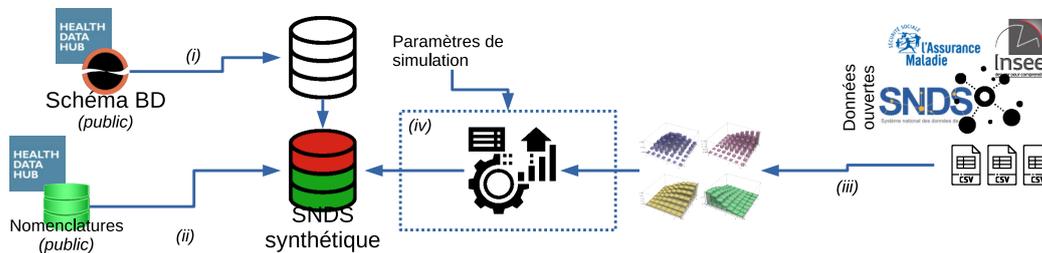


Figure 1.3 – Illustration du processus de génération d'un jeu de données synthétiques.

médico-économiques. De part son contenu médical sensible et personnel, son usage est restreint, ce qui limite les possibilités d'expérimentation de nouveaux algorithmes sur ces données.

Nous avons proposé de générer des données synthétiques pour alimenter une base de données, d'une part, respectant la structure originale du SNIIRAM et, d'autre part, reproduisant des statistiques connues sur les agrégats de variables épidémiologiques en s'appuyant pour cela sur les données ouvertes (*open data*) [6]. Les mesures de protection statistique mises en œuvre sur les données ouvertes librement accessibles assurent ainsi leur réutilisabilité dans le respect de la vie privée.

Nous avons ainsi développé des outils, disponibles en ligne<sup>3</sup> pour générer un SNIIRAM synthétique.

Le processus de génération en quatre phases principales est illustré dans la figure 1.3 : (i) création de la structure générale de la base de données à partir du schéma de la base de données, (ii) chargement des nomenclatures qui alimentent 416 tables, (iii) reconstruction de distributions des variables à partir de données ouvertes, (iv) simulation de nouvelles bases aléatoires (12 tables).

Le génération de données synthétiques est configurée en précisant la liste des départements à considérer ainsi que l'année à simuler.

Ces paramètres servent à générer une population synthétique dont on connaît, à une date donnée, la structure par sa localisation (code commune), son sexe et son âge quinquennal (i.e. par classe d'âge de 5 ans).

Le principe de génération d'un parcours de soins comprend trois dimensions : (i) déterminer si et combien de prestations d'une nature donnée (hospitalisation, visite, acte, etc.) un bénéficiaire a eu dans l'année, (ii) déterminer le code de chacune de ces prestations (code CIM, CCAM, CIP ou NABM) en utilisant des probabilités conditionnelles d'observer ces codes, (iii) lier cette prestation aux informations complémentaires (e.g. spécialité de l'exécutant).

Les distributions statistiques et des estimations de quantités de prestations sont estimées à partir des données ouvertes provenant de l'INSEE ou de la DREES et publiées sur la plateforme ouverte des données publiques françaises<sup>4</sup>. Les quantités et les probabilités sont conditionnées aux caractéristiques du patient (âge, sexe et localisation). La granularité de la localisation est par défaut celle du département. Le défi est lié à l'incomplétude des informations disponibles. Pour le résoudre, différentes techniques bayésiennes ont été utilisées pour estimer des distributions jointes à partir de distributions marginales.

L'outil permet ainsi à chacun de générer un jeu de données respectant la structure de la

3. [https://gitlab.inria.fr/tguyet/medtrajectory\\_datagen](https://gitlab.inria.fr/tguyet/medtrajectory_datagen)

4. <http://www.data.gouv.fr>



base de données et contenant des informations statistiquement réalistes.

## Responsabilité des IA en santé

*Ce travail est mené en collaboration avec l'Institut de Droit Public de l'Université de Rennes 1 (IDPSP), le centre Borelli, le CERSA, l'équipe IRISA-Inria/DRUID et l'unité REPERES (CHU Rennes/EHESP) dans le cadre du projet DRIAS (Droit et Responsabilité des IA en Santé) financé par le CNRS (2020).*

L'objet du projet DRIAS est d'explorer les questions juridiques en matière de responsabilité publique que soulève l'utilisation des méthodes d'IA dans un contexte médical. La notion de responsabilité est un concept clé qui motive des travaux de recherche récents en IA sur la transparence des décisions (interprétabilité, explicabilité), la limitation des biais (équité des traitements, robustesse aux exemples adversariaux) ou encore le respect de la vie privée (données personnelles). Derrière chacun de ces volets sont identifiés des risques potentiellement nouveaux liés à l'utilisation de méthodes d'IA, et en particulier dans un contexte médical. Alors que ces technologies sont en plein développement et que les législateurs (et en particulier la commission européenne) entreprend de mettre en phase le cadre législatif de la responsabilité civile avec les révolutions liées à l'utilisation des techniques d'IA, il s'agit de confronter ces cadres législatifs existants ou en cours de construction avec des situations pratiques.

Dans ce projet en cours, nos réflexions sont conduites selon trois axes : 1) les responsabilités liées aux questions de décision automatique ou d'aide à la décision, 2) les responsabilités liées à l'utilisation des données personnelles et sensibles (de santé) dans les chaînes de traitement d'IA, et 3) les responsabilités adverses auxquelles sont notamment confrontés les chercheurs.

Un travail collectif a été mené sur l'étude des documents sur la responsabilité des IA de la Commission Européenne publiés pendant l'année 2020 (incluant une résolution votée). Ces documents définissent un cadre très général qui sera l'objet de discussions par la Commission dans les années à venir pour définir de nouveaux régimes de responsabilité applicables aux systèmes d'IA. Nous avons ainsi travaillé à identifier quelques cas pratiques dans le champ du projet (responsabilité administrative et contexte médical) qui puissent mettre en lumière les conséquences de leur mise en place.

## Références

- [1] Johanne Bakalara, Thomas Guyet, Olivier Dameron, André Happe, and Emmanuel Oger. An extension of chronicles temporal model with taxonomies – application to epidemiological studies. In *Proceedings of HEALTHINFO*, page à paraître, 2021.
- [2] Julien Bezin, Mai Duong, Régis Lassalle, Cécile Droz, Antoine Pariente, Patrick Blin, and Nicholas Moore. The national healthcare system claims databases in France, SNIIRAM and EGB : Powerful tools for pharmacoepidemiology. *Pharmacoepidemiology and drug safety*, 26(8) :954–962, 2017.
- [3] Yann Dauxais, Thomas Guyet, David Gross-Amblard, and André Happe. Discriminant chronicles mining : Application to care pathway analytics. In *Proceedings of the Conference on Artificial Intelligence in Medicine*, 2017.
- [4] Erwan Drezen, Thomas Guyet, and André Happe. From medico-administrative databases analysis to care trajectories analytics : an example with the French SNDS. *Fundamental and Clinical Pharmacology*, 32(1) :78–80, 2018.
- [5] Thomas Guyet. *Enhancing sequential pattern mining with time and reasoning*. Habi-



litation à diriger des recherches, Université de Rennes 1, 2020.

- [6] Thomas Guyet. Génération d'un snds synthétique et réaliste à partir de données ouvertes. In *Actes de la conférence Extraction et Gestion des Connaissances (EGC)*, page à paraître, 2021.
- [7] Thomas Guyet, André Happe, and Yann Dauxais. Declarative Sequential Pattern Mining of Care Pathways. In *Proceedings of the Conference on Artificial Intelligence*

*in Medicine in Europe*, volume 24, pages 1161–266, 2017.

- [8] Elisabeth Polard, Emmanuel Nowak, André Happe, Arnaud Biraben, and Emmanuel Oger. Brand name to generic substitution of antiepileptic drugs does not lead to seizure-related hospitalization : a population-based case-crossover study. *Pharmacoepidemiology and drug safety*, 24(11) :1161–1169, 2015.

## ■ IA & Neuroimagerie

GIN

Équipe Neuroimagerie Fonctionnelle & Perfusion cérébrale, Inserm-UGA

<https://neurosciences.univ-grenoble-alpes.fr/>

**Emmanuel BARBIER**

[Emmanuel.barbier@univ-grenoble-alpes.fr](mailto:Emmanuel.barbier@univ-grenoble-alpes.fr)

**Thomas CHRISTEN**

[thomas.christen@univ-grenoble-alpes.fr](mailto:thomas.christen@univ-grenoble-alpes.fr)

**Michel DOJAT**

[Michel.dojat@inserm.fr](mailto:Michel.dojat@inserm.fr)

**Benjamin LEMASSON**

[benjamin.lemasson@univ-grenoble-alpes.fr](mailto:benjamin.lemasson@univ-grenoble-alpes.fr)

### Introduction

L'équipe "Neuroimagerie fonctionnelle et perfusion cérébrale" (Dir. E. Barbier) est labellisée par l'Inserm et l'Université de Grenoble Alpes. Elle regroupe 23 membres permanents avec des compétences en médecine, biologie, ingénierie, neuroimagerie et neuroinformatique ; et focalise ses travaux méthodologiques à l'interface préclinique et clinique. Depuis 2019, elle contribue au projet grenoblois MIAI, un des quatre projets 3AI français, sur les axes "Machine learning models (ML)" (chaire "Towards a robust and understandable neuro-morphic system") et "Health" (chaire "AI for high throughput biomedical investigations"). Notre activité en IA & santé se propose d'explorer et de raffiner les nouvelles techniques

ML dans le contexte de l'imagerie *in vivo*. Une collaboration étroite est effective avec l'équipe **Statify** (Inria, dir F. Forbes). Nous travaillons selon trois axes : acquisition des images, définition de biomarqueurs et construction d'environnements informatiques adaptés à l'apprentissage sur de larges bases de données d'imagerie.

### Axe 1 - Techniques avancées d'acquisition et de reconstruction d'IRM

Ces dernières années ont connu des améliorations majeures dans les domaines de l'acquisition et de la reconstruction d'IRM. Les nouveaux cadres de reconstruction d'images utilisent désormais des algorithmes de ML, le plus souvent d'apprentissage profond (Deep Lear-



**Afia**

Association française  
pour l'Intelligence Artificielle

ning (DL)), pour récupérer des données à partir d'acquisitions sous-échantillonnées et non linéaires; tandis que l'approche MRF (MR Fingerprinting) relie des simulations numériques avancées et des acquisitions *in vivo* pour accélérer les séquences IRM et fournir des données quantitatives [2]. Nous développons des outils les plus réalistes et flexibles possibles pour simuler le signal IRM en présence de structures biologiques, auquel nous associons des outils de ML. In fine, cela conduira à de nouvelles façons d'acquérir et d'analyser les données de perfusion ou d'oxygénation tissulaire. Avec l'approche MRF, nous pouvons créer des « séquences » IRM qui génèrent des cartes anatomiques, des cartes de paramètres biophysiques ou physiologiques, le tout en une seule acquisition. En utilisant l'accélération fournie par la technique multi-bandes et le ML, nous cherchons à fournir des solutions rapides (quelques minutes) et robustes aux cliniciens (indépendantes du scanner IRM et avec correction de mouvement) en une étape pour la prise en charge notamment de l'AVC aigu (financement ANR et UGA).

## **Axe 2 - Définition de biomarqueurs d'imagerie**

Pour extraire des informations utiles à partir de données d'imagerie nous travaillons sur deux axes : fusion d'images multiparamétriques et multimodales (IRM, TEP, Scanner X et microscopie) et connectivité cérébrale par IRM anatomique et fonctionnelle.

### *2.1. Histologie non invasive par IRM (h-IRM)*

Ce projet a pour objectif de démontrer que l'IRM multiparamétrique *in vivo*, couplée à des analyses statistiques avancées, permet d'obtenir de façon non invasive des cartes fournissant des informations équivalentes à celles obtenues par histologie. De plus, dans le cas des glioblastomes, tumeurs très hétérogènes, une biop-

sie unique ne permet pas d'obtenir un diagnostic clinique fiable. L'h-IRM est aussi en cours d'évaluation en tant que biomarqueur d'imagerie pour l'évaluation de l'efficacité d'une immunothérapie chez des patients porteurs d'un glioblastome en récurrence (projet SIG2MA avec l'Oncopôle de Toulouse).

### *2.2 Biomarqueurs d'imagerie de l'évolution lésionnelle cérébrale des traumatisés crâniens*

Le traumatisme crânien (TC) est un ensemble complexe d'entités nosologiques d'évolutions différentes dont l'identification précoce, centrale pour la prise en charge des lésions secondaires à l'agression initiale, est difficile. Il n'existe pas, à l'heure actuelle, de méthode d'imagerie utilisable en routine clinique permettant de visualiser, de quantifier et de prédire les évolutions lésionnelles. Notre objectif principal est de rechercher des signatures d'évolution scannographique cérébrale précoces associées à l'évolution clinique des TC. Nous faisons l'hypothèse que l'analyse automatique de scanners cérébraux répétés et précoces permettra de distinguer l'évolution des tissus cérébraux des patients définissant ainsi de véritables « biomarqueurs d'imagerie ». Nous souhaitons mettre en place un outil informatique utilisable en routine clinique associant ces profils scannographiques à la prise en charge thérapeutique lors du séjour en réanimation. Ce projet est soutenu par le CHU-GA et la fondation des « Gueules Cassées ».

### *2.3 Harmonisation de données d'imagerie pour les études multicentriques*

L'introduction d'études IRM multicentriques permet d'augmenter la robustesse statistique des résultats. Cependant, chaque protocole d'acquisition ou scanner utilisés introduisent un bruit qui leur est propre, source de contamination des résultats. L'harmonisation est un moyen de s'affranchir de ces biais, tout en conservant l'information biologique contenue dans les images. Nous étudions des méthodes



d'harmonisation statistiques telles que combat [1] et à base de GAN (Generative Adversarial Network). À l'aide de données synthétiques, nous évaluons leur capacité à supprimer les artefacts dans les images tout en conservant les différences utiles propres aux populations imagées dans les centres. Ce travail est une étude ancillaire à un projet multicentrique national (cohorte DefiDiag) qui vise à améliorer le diagnostic de certaines maladies génétiques rares conduisant à des déficits intellectuels. Ce projet bénéficie d'un soutien financier du projet 3IA @MIAI.

#### *2.4 Détection d'anomalies cérébrales précoces dans les maladies neurodégénératives*

Nous cherchons à extraire des signatures spécifiques, à partir des données d'imagerie, chez des patients parkinsoniens *de novo* (nouvellement diagnostiqués, sans traitement). Nous explorons des modèles de mélange à base de lois de Student généralisées qui permettent de générer des lois de distributions variées et ce, quel que soit le nombre de cartes de paramètres considérées. Un premier dictionnaire multiparamétrique de « normalité » décrit le groupe de sujets sains. Chez les patients, nous recherchons les voxels atypiques ne pouvant pas être expliqués par ce dictionnaire. Ces voxels permettent de construire un nouveau dictionnaire de clusters pathologiques et des signatures chez chaque patient. Il s'agit enfin de les relier aux connaissances d'experts pour faire émerger des biomarqueurs pertinents et proposer un sous-typage des patients. Les techniques alternatives à base de GAN sont aussi utilisées pour reconstruire fidèlement des données de sujets sains. Lorsque que ce même réseau est confronté à des données de patients, ses erreurs de reconstruction permettent de détecter les anomalies. Ces travaux sont aussi poursuivis en collaboration avec la start-up Inserm-Inria [Pixyl](#) spécialisée dans le développement d'outils d'analyse d'images médicales à base d'IA.

Ils sont financés par l'Idex NeuroCOG.

#### *2.5 Apprentissage de graphes de connectivité cérébrale*

La notion de graphe, où des nœuds sont reliés entre eux par des arêtes, est un bon modèle pour représenter à la fois l'anatomie et le fonctionnement cérébral. Des outils mathématiques sont alors disponibles pour caractériser ces graphes de connectivité : calculer le plus court chemin entre deux régions, extraire les régions à forte connectivité, la densité la centralité, la modularité... Ces graphes permettent de capturer à la fois l'information spatiale et temporelle des mécanismes cérébraux. Leur comparaison permet d'obtenir de nouvelles connaissances quant à l'impact d'une pathologie sur la structure et le fonctionnement du réseau cérébral. Nous combinons la théorie de graphes et le DL. Des caractéristiques sont extraites des graphes (*network embedding*) qui sont ensuite utilisés pour faire de l'apprentissage et définir un graphe moyen, caractéristique de la population étudiée, normale ou pathologique. Nous explorons en particulier les dysfonctions de connectivité dans les situations de coma cérébral. L'objectif à long terme est d'améliorer notre compréhension des mécanismes cérébraux et possiblement de raffiner la nosologie des pathologies cérébrales sur la base de la connectivité. Ce projet bénéficie d'un soutien financier du projet 3IA @MIAI.

#### *2.6 Apprentissage fédératif*

En médecine, des données peuvent être sensibles (e.g. génétiques, nominatives) et difficiles à exporter d'un centre à l'autre. Nous démarrons une étude en collaboration avec le Laboratoire d'Informatique de Grenoble sur l'évaluation de techniques permettant de réaliser l'apprentissage d'un modèle à partir de données distribuées sur différents sites hospitaliers. Un enjeu théorique est aussi d'estimer le nombre d'exemples annotés nécessaires pour la réalisa-

tion d'une tâche donnée avec des performances acceptables.

### **Axe 3 - Infrastructures spécifiques pour la gestion de larges masses de données et d'outils de traitement en imagerie *in vivo***

Trois actions collaboratives sont menées :

- (i) *FLI-IAM* : action nationale dans le cadre du plan d'investissement d'avenir pour le développement d'une architecture fédérative de partage d'outils de traitement, de données et de métadonnées d'imagerie *in vivo*. La première phase (2013-2020) a permis la mise en place d'un démonstrateur utilisé notamment pour la réalisation de challenges (MICCAI 2016) sur la segmentation de lésion de sclérose en plaques et en oncologie. L'exécution des pipelines de traitement d'images est réalisée sur l'infrastructure VIP (Créatis, Lyon). La seconde phase (2020-2025) vise à rendre opérationnel ce démonstrateur et à son déploiement.
- (ii) *textitShanoir-NG* : outils de gestion de données et métadonnées en imagerie clinique et préclinique. Différents projets de recherche clinique multicentriques (Ofsep, OxyTC, Braini, Radiomics...) sont gérés via Shanoir.
- (iii) *textitMP3* : outil de gestion de données et métadonnées en imagerie préclinique et clinique. MP3 permet l'importation directe des données d'imagerie multimodale des consoles d'acquisition, leur visualisation et la création ainsi que l'exécution locale de pipelines d'analyse complexes (composés d'une multitude de briques algorithmiques interconnectées).

### **Conclusion**

Le développement de l'imagerie médicale connaît actuellement un nouveau souffle avec l'apport de concepts mathématiques et informatiques et d'outils logiciels puissants qui permettent de repenser la façon dont les données sont acquises, reconstruites et préparées pour l'interprétation médicale. L'évaluation clinique de ces nouveaux outils passe par le renforcement des appuis en ingénierie sur le versant académique et par des collaborations entre les centres hospitalo-universitaires et les entreprises susceptibles de proposer ces nouveaux services. De nouveaux modes de valorisation de ces coopérations autour des données médicales sont à inventer.

### **Références**

- [1] Jean-Philippe Fortin, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, Crystal Cooper, Maurizio Fava, Patrick J. McGrath, Melvin McInnis, Mary L. Phillips, Madhukar H. Trivedi, Myrna M. Weissman, and Russell T. Shinohara. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167 :104–120, 2018.
- [2] Dan Ma, Vikas Gulani, Nicole Seiberlich, Kecheng Liu, Jeffrey L. Sunshine, Jeffrey L. Duerk, and Mark A. Griswold. Magnetic resonance fingerprinting. *Nature*, 495(7440) :187–192, March 2013.



## ■ Intelligence Artificielle pour la Chirurgie

*LTSI / MEDICIS*  
Inserm et Université de Rennes 1  
<https://medicis.univ-rennes1.fr>

**John BAXTER**

[john.baxter@univ-rennes1.fr](mailto:john.baxter@univ-rennes1.fr)

**Bernard GIBAUD**

[bernard.gibaud@univ-rennes1.fr](mailto:bernard.gibaud@univ-rennes1.fr)

**Arnaud HUAULMÉ**

[arnaud.huaulme@univ-rennes1.fr](mailto:arnaud.huaulme@univ-rennes1.fr)

**Pierre JANNIN (RESP. ÉQUIPE)**

[pierre.jannin@univ-rennes1.fr](mailto:pierre.jannin@univ-rennes1.fr)

### Introduction

Le projet de l'équipe Inserm MediCIS du LTSI vise à améliorer la qualité des interventions chirurgicales grâce à la science des données chirurgicales (*Surgical Data Science*). S'appuyant sur les récents progrès de l'intelligence artificielle (IA) et de l'apprentissage automatique profond, la science des données chirurgicales vise à s'appuyer sur des données collectées tout au long du processus chirurgical, qui sont analysées pour produire des connaissances explicites pouvant être utilisées pour l'aide à la décision, l'évaluation ou la formation chirurgicale. La science des données chirurgicales comprend l'analyse des processus humains et des effecteurs en plus des données sur les patients. Une analyse holistique de l'environnement chirurgical est ensuite utilisée pour développer la prochaine génération de systèmes chirurgicaux assistés par ordinateur et de simulateurs chirurgicaux. Sur les aspects méthodologiques, nous abordons à la fois les approches *bottom-up*, s'appuyant sur les données, et *top-down*, s'appuyant sur des modèles génériques de connaissances instanciables. L'équipe Inserm est localisée à la Faculté de Médecine de Rennes et inclut des chercheurs de disciplines différentes : mathématiques, informatique et traitement du signal. Elle inclut aussi de nombreux chercheurs hospitalo-universitaires cou-

vrant différentes spécialités chirurgicales : neurochirurgie, chirurgie orthopédique, et chirurgie gynécologique, par exemple. L'équipe est composée de 8 personnels permanents et d'environ une vingtaine de post-docs, doctorants et ingénieurs. Nous présentons trois thèmes de recherche pour illustrer comment l'IA est au cœur de nos travaux de recherches actuels.

### Thème 1. Ontologies et représentations sémantiques pour le partage de données radiologiques et chirurgicales

Les travaux menés au cours des années précédentes ont exploré différents aspects du partage de données en imagerie médicale et dans le domaine de la représentation symbolique des procédures chirurgicales. Les premiers se situent dans le contexte de la mise en place de systèmes de partage d'images appelés *biobanques d'images* (par analogie avec les biobanques de tissus), destinées à permettre l'utilisation secondaire des données d'imagerie pour différentes applications (recherche biomédicale, méta-analyse, conception et validation de biomarqueurs d'imagerie). Ils se fondent sur l'hypothèse que les technologies sémantiques peuvent apporter la précision sémantique requise en recherche biomédicale, aussi bien dans les activités visant à constituer de grands jeux de données à partir de données hétérogènes



issues de sites multiples (fédération de données), que dans les traitements appliqués à ces données dans le cadre de la recherche elle-même. Ces travaux incluent : i) la modélisation conceptuelle inhérente à la conception d'ontologies, ii) l'intégration d'extraits d'ontologies existantes pour la création d'ontologies d'application dédiées à un domaine donné, iii) la création de logiciels permettant de "sémantiser" les données, c'est-à-dire de produire des données instanciant les ontologies choisies. L'originalité de nos travaux tient notamment au choix d'une approche réaliste, inspirée des travaux de Barry Smith et de Werner Ceusters, et à l'utilisation d'une ontologie fondatrice pour l'alignement cohérent des différents extraits, en l'occurrence l'ontologie *Basic Formal Ontology* (BFO2). Ce type d'approche a été mis en jeu dans différents projets : la création d'une ontologie dédiée au domaine des biomarqueurs d'imagerie, en collaboration avec l'IRT B-COM à Rennes [1, 2] et la création d'une base de données sémantique décrivant la nature et la provenance d'images et de données dosimétriques au sein d'un système de partage de données appelé IRDBB (pour Image and Radiation Dose BioBank), développé en collaboration avec l'hôpital de Genève (Osman Ratib) et l'IRT B-COM [13]. Ce dernier système a été conçu dans le cadre du projet européen MEDIRAD (programme EURATOM), dédié à la recherche sur les risques associés à l'exposition à des faibles doses de rayonnement en médecine au cours d'explorations diagnostiques ou liées à des traitements (chirurgicaux, ou associés à des radiothérapies internes). Les travaux réalisés ont consisté à définir une ontologie d'application, appelée OntoMEDIRAD, couvrant le domaine des projets de recherche clinique du projet MEDIRAD, puis à créer les logiciels capables de peupler cette base de données sous la forme d'instances des classes de l'ontologie, représentées en langage RDF, et obtenues

par traduction des métadonnées (standard DICOM) associées aux images et de descriptions des calculs dosimétriques estimant les doses absorbées par les organes du fait de leur exposition aux faibles doses de rayonnement. Un second domaine d'application concerne la représentation symbolique des activités chirurgicales. Les travaux ont permis de proposer une ontologie dédiée, appelée OntoSPM, fournissant un socle commun de représentation des processus chirurgicaux (à différents niveaux de granularité). Cette ontologie a été utilisée dans différents projets, notamment, le projet *S3PM* pour l'analyse et création de scénarios de simulation utilisant la réalité virtuelle (financé dans le cadre du Labex CominLabs), le projet *CONDOR* (sur la salle d'opération du futur) et le projet LapOntoSPM sur la reconnaissance des phases en chirurgie laparoscopique [17]. OntoSPM a donné lieu en 2016 à la création d'une collaboration internationale appelée *OntoSPM Collaborative Action* chargée de maintenir et étendre l'ontologie OntoSPM et de promouvoir son adoption et son utilisation dans les applications relevant de la science des données chirurgicales, et la conception de nouveaux systèmes informatisés (systèmes de gestion de *workflow*, robots chirurgicaux, systèmes automatisés de *reporting*, etc.) [12].

## **Thème 2. Réseaux profonds et modèles de connaissances pour l'analyse procédurale en chirurgie**

La modélisation des processus chirurgicaux (SPM pour *Surgical Process Models*) est essentielle pour pouvoir comprendre le déroulement des opérations chirurgicales et, in fine, pour concevoir des outils d'aide à la prise de décision ou de formation pour le chirurgien et l'ensemble de l'équipe chirurgicale.

Ces modèles décomposent les procédures chirurgicales en différents concepts, nommés niveaux de granularité. Ceux-ci, comme les



phases par exemple, peuvent décrire les temps chirurgicaux, des actions physiques réalisées sur une structure anatomique par un instrument chirurgical (activités), ou encore des représentations numériques des mouvements (dexèmes) [19, 10]. Ces niveaux de granularité ont été formalisés dans l'ontologie OntoSPM [12]. De nombreuses applications peuvent être déduites de cette description des processus chirurgicaux, comme l'évaluation du niveau d'expertise [11], l'assistance robotique [18] et l'évaluation de la qualité [6]. Afin d'arriver à de telles applications dans un contexte clinique, un enjeu majeur est de développer des méthodes capables de reconnaître automatiquement et en temps réel les différents niveaux de granularité à partir de données issues de capteurs disponibles en salle d'opération. Ces capteurs peuvent être la vidéo chirurgicale ou la cinématique générée par des robots chirurgicaux. Au cours des dernières années, l'équipe MediCIS a proposé différentes méthodes innovantes basées machine et *deep learning* dans ce domaine. Par exemple, elle a proposé des méthodes de reconnaissance d'activités basées sur le flux optique [23], de reconnaissance d'étapes avec des informations vidéos et la présence d'outils [8]. Nous avons aussi développé d'autres méthodes comme l'aide à l'annotation à partir de vidéos [20], ou l'annotation automatique à partir des informations fournies par un simulateur chirurgical [15]. Dans le cadre de MICCAI 2020, nous avons proposé un challenge nommé MISAW (*Micro-Surgical Anastomose Workflow recognition challenge on training sessions*). L'objectif de ce challenge était de développer un modèle capable de reconnaître plusieurs niveaux de granularité à partir de données vidéos et cinématiques de sessions d'entraînement de sutures chirurgicales. Les 27 séquences utilisées pour ce challenge seront prochainement disponibles publiquement.

Ces dernières contributions s'axaient sur

la création ou la reconnaissance des SPM, conduisant à différentes applications, comme la détection de déviation d'un processus chirurgical en fonction des spécificités des patients ou de l'apparition d'événements indésirables [16], et l'identification de signatures chirurgicales spécifiques du niveau d'expertise [14].

L'équipe mène de nombreux projets axés sur cette thématique, comme l'étude de l'impact de l'utilisation régulière d'un simulateur dans la formation initiale de jeunes internes en arthroscopie, l'étude de la courbe d'apprentissage d'un système d'assistance robotique, l'analyse de données physiologiques comme la fréquence cardiaque ou des informations de postures, en utilisant les méthodes d'IA pilotées par les données. Nous avons aussi soumis un challenge pour MICCAI 2021 afin d'évaluer l'apport de différentes modalités dans les méthodes de reconnaissance automatique de SPM.

### **Thème 3. Apprentissage profond pour le traitement d'images et de signaux médicaux en neurochirurgie fonctionnelle**

La Stimulation Cérébrale Profonde (SCP) est un traitement symptomatique applicable à un grand nombre de désordres neurologiques, particulièrement la maladie de Parkinson. Le but de cette intervention est de remédier à des dysfonctionnements de l'activité neuronale. Sa mise en œuvre fait intervenir plusieurs étapes cliniques, notamment :

- Le "planning" pré-opératoire, c'est-à-dire la détermination des cibles pour la stimulation et le choix des meilleures trajectoires chirurgicales pour la mise en place des électrodes ;
- La mise en place des électrodes selon la trajectoire choisie en fonction de l'anatomie sous-jacente ;
- La détermination exacte post-opératoire de la position des électrodes et l'ajustement des paramètres électriques pour atténuer au



mieux les symptômes du patient.

La complexité de la procédure et ses nombreuses étapes cliniques nécessitent des données d'imagerie de différentes modalités, en particulier des IRM et des enregistrements électrophysiologiques réalisés pendant l'intervention chirurgicale. Le but de nos travaux est d'optimiser les processus de traitement du signal et de l'image, en associant à la fois les méthodes traditionnelles et l'apprentissage automatique, pour assister les clinicien(ne)s qui réalisent ces interventions.

Pour le planning pré-opératoire et l'analyse post-opératoire de positionnement des électrodes, notre équipe a développé le système PyDBS [7] qui stocke les images pré- et post-opératoires et effectue leur mise en correspondance spatiale (recalage). Ce système a permis d'améliorer le planning en estimant la probabilité des effets secondaires en fonction du positionnement de l'électrode [3] et en suggérant les trajectoires optimales à partir des images [9]. Actuellement, nous cherchons à optimiser la segmentation des structures anatomiques sous-corticales à partir des images pré-opératoires en utilisant des méthodes d'apprentissage automatique avec lesquelles les cliniciens peuvent interagir pour apporter des corrections [4, 5]. Nous utilisons aussi les images pour mieux comprendre et reconnaître les différents sous-types de la maladie de Parkinson, là aussi à partir d'apprentissage automatique [22]. Nous avons aussi proposé une approche innovante pour la prédiction du résultat clinique par apprentissage automatique à partir des données et images cliniques pré- et post-opératoires. Cette approche a été adaptée aux différents moments du *workflow* chirurgical pour l'aide à la décision : prédiction du résultat clinique avant même l'implantation, prédiction des effets secondaires pour optimiser l'implantation et prédiction des résultats cliniques pour optimiser les paramètres électriques de stimu-

lation [21].

Concernant la mise en place intra-opératoire des électrodes SCP, notre équipe travaille sur la reconnaissance des signaux électro-physiologiques. Nous avons créé un réseau de neurones convolutif qui distingue les signaux provenant du noyau sous-thalamique (une cible privilégiée de la SCP) des signaux provenant des structures voisines en utilisant seulement une seconde de signal [21]. Ce réseau rendra plus efficace les interventions SCP en réduisant le temps nécessaire pour collecter et interpréter ces signaux, actuellement utilisés par les neurophysiologistes pour identifier "à l'oreille" la structure anatomique dans laquelle l'électrode est implantée.

## Références

- [1] Emna Amdouni and Bernard Gibaud. Conception d'une ontologie générique pour la représentation sémantique des biomarqueurs d'imagerie. In *Journées Françaises sur les Ontologies*, Bordeaux, France, October 2016.
- [2] Emna Amdouni and Bernard Gibaud. Imaging Biomarker Ontology (IBO) : A Biomedical Ontology to Annotate and Share Imaging Biomarker Data. *Journal on Data Semantics*, 7(4) :223, September 2018.
- [3] Clement Baumgarten et al. Improvement of Pyramidal Tract Side Effect Prediction Using a Data-Driven Method in Subthalamic Stimulation. *IEEE Transactions on Biomedical Engineering*, 64(9) :2134–2141, September 2017.
- [4] John S. H. Baxter et al. Localisation of the subthalamic nucleus in MRI via convolutional neural networks for deep brain stimulation planning. In *Medical Imaging 2020 : Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11315, pages 155 – 160. International



- Society for Optics and Photonics, SPIE, 2020.
- [5] John S. H. Baxter et al. Segmentation of the subthalamic nucleus in MRI via Convolutional Neural Networks for deep brain stimulation planning. In *Medical Imaging 2021 : Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11598, pages 371 – 376. International Society for Optics and Photonics, SPIE, 2021.
- [6] Loubna Bouarfa and Jenny Dankelman. Workflow mining and outlier detection from clinical activity logs. *Journal of Biomedical Informatics*, 45(6) :1185–1190, 2012.
- [7] Tiziano D’Albis et al. PyDBS : an automated image processing workflow for deep brain stimulation surgery. *International Journal of Computer Assisted Radiology and Surgery*, 10(2) :117–128, February 2015.
- [8] Olga Dergachyova et al. Automatic data-driven real-time segmentation and recognition of surgical workflow. *International Journal of Computer Assisted Radiology and Surgery*, 11(6) :1081–1089, June 2016.
- [9] Olga Dergachyova et al. Automatic preoperative planning of DBS electrode placement using anatomo-clinical atlases and volume of tissue activated. *International Journal of Computer Assisted Radiology and Surgery*, 13(7) :1117–1128, July 2018.
- [10] Fabien Despinoy et al. Unsupervised trajectory segmentation for surgical gesture recognition in robotic training. *IEEE Transactions on Biomedical Engineering*, 63(6) :1280–1291, 2016.
- [11] Germain Forestier et al. Multi-site study of surgical practice in neurosurgery based on surgical process models. *Journal of Biomedical Informatics*, 46(5) :822–829, 2013.
- [12] Bernard Gibaud et al. Toward a standard ontology of surgical process models. *International Journal of Computer Assisted Radiology and Surgery*, 13(9) :1397–1408, September 2018.
- [13] Bernard Gibaud et al. A semantic database for integrated management of image and dosimetric data in low radiation dose research in medical imaging. In *AMIA Annual Symposium Proceedings*, 2020.
- [14] Arnaud Hualmé et al. Sequential surgical signatures in micro-suturing task. *International Journal of Computer Assisted Radiology and Surgery*, 13(9) :1419–1428, September 2018.
- [15] Arnaud Hualmé et al. Automatic annotation of surgical activities using virtual reality environments. *International Journal of Computer Assisted Radiology and Surgery*, 14(10) :1663–1671, October 2019.
- [16] Arnaud Hualmé et al. Offline identification of surgical deviations in laparoscopic rectopexy. *Artificial Intelligence in Medicine*, 104 :101837, 2020.
- [17] Darko Katić et al. LapOntoSPM : an ontology for laparoscopic surgeries and its application to surgical phase recognition. *International Journal of Computer Assisted Radiology and Surgery*, 10(9) :1427–1434, September 2015.
- [18] Seong-Young Ko et al. Surgery task model for intelligent interaction between surgeon and laparoscopic assistant robot. *Int. J. Assitive Robot. Mechatronics*, 8(1) :38–46, 2007.
- [19] Florent Lalys and Pierre Jannin. Surgical process modelling : a review. *International Journal of Computer Assisted Radiology and Surgery*, 9(3) :495–511, May 2014.



- [20] Gurvan Lecuyer et al. Assisted phase and step annotation for surgical videos. *International Journal of Computer Assisted Radiology and Surgery*, 15(4) :673–680, April 2020.
- [21] Maxime Peralta et al. SepaConvNet for Localizing the Subthalamic Nucleus using One Second Micro-Electrode Recordings. In *42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society in conjunction with the 43rd, Annual Conference of the Canadian Medical and Biological Engineering Society*, Montreal, Canada, July 2020.
- [22] Maxime Peralta et al. Striatal shape alteration as a staging biomarker for parkinson's disease. *NeuroImage : Clinical*, 27 :102272, 2020.
- [23] Duygu Sarikaya and Pierre Jannin. Surgical gesture recognition with optical flow only, 2020.

## ■ Représentations des médecins généralistes et des patients concernant l'utilisation de l'IA en Santé : une étude qualitative

Département Universitaire Médecine Générale  
Université Sorbonne Paris Nord  
<https://dumg.univ-paris13.fr>

**Alexis ASTRUC**

[alexis.astruc@sorbonne-paris-nord.fr](mailto:alexis.astruc@sorbonne-paris-nord.fr)

**Ammar ALSHEIKHLY**

**Mehdi OULMOUDDANE**

### Introduction

L'entrée dans le XXI<sup>ème</sup> siècle a été marquée par une révolution numérique avec le développement, en seulement deux décennies de nouvelles technologies qui se sont intégrées progressivement dans notre quotidien : applications mobiles sur smartphone, objets connectés, réseaux sociaux, messageries instantanées. Parmi elles, l'intelligence artificielle (IA) reste celle dont les tenants et les aboutissants ne sont pas encore bien maîtrisés. L'IA est née dans les années 1950, et à ce jour, il n'existe toujours pas de définition précise. Le Larousse décrit celle-ci comme "l'ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine". En pratique, il peut s'agir d'un simple programme informatique (algorithme) ou un ensemble de techniques plus complexes (*deep learning*, *machine learning*, réseaux de neu-

rones) dont le but est d'accomplir des tâches qui sont, pour le moment, purement accomplies par l'intelligence humaine [12, 6]. Dans le domaine de la santé, d'après la Ministre de la Santé, Agnès Buzyn, l'apparition de cette nouvelle technologie pourrait améliorer notre système de santé, tant au niveau individuel que collectif [2]. D'après un article de l'Institut national de la santé et de la recherche médicale datant de juillet 2018 : "L'IA est au cœur de la médecine du futur, avec les opérations assistées, le suivi des patients à distances, les traitements personnalisés grâce au recoupement d'un nombre croissant de données (*big data*), etc." [7]. Les experts s'accordent actuellement à dire que l'IA ne vise pas à se substituer à l'expertise du médecin, mais qu'elle pourrait aider le médecin au quotidien dans certaines prises en charge diagnostique et thérapeutique [9]. Le Conseil national de l'Ordre souhaite "soutenir



le développement d'une société numérique au service des soignants et des patients et faire de l'IA un apport essentiel pour l'aide à la décision et à la stratégie thérapeutique" [10]. L'académie nationale de médecine souhaite que la formation des futurs médecins à l'usage des nouvelles technologies devienne indispensable [8]. C'est dans cet esprit que, depuis septembre 2019, un nouveau diplôme universitaire "Intelligence artificielle appliquée en santé" a vu le jour, inauguré par l'Université Paris Descartes. L'objectif de ce DU est de comprendre les enjeux de l'IA en santé, ses diverses applications, ses potentialités comme ses limites [3]. En 2020, les applications possibles de l'IA en médecine sont nombreuses et concrètes. Des outils sont désormais capables de diagnostiquer un mélanome mais aussi de dépister une rétinopathie diabétique avec la même précision que les spécialistes [14, 13]. Le développement de l'IA aura donc probablement à l'avenir un impact important sur de nombreuses spécialités médicales, y compris en médecine générale [11]. Plusieurs pistes sont en cours de réflexions, par exemple pour épargner du temps administratif et se concentrer sur la relation médecin-malade l'idée a germé d'une IA "assistante médicale" avec un questionnement personnalisé du patient en amont de la consultation toujours dans le but d'aider le médecin et non le remplacer. Plus récemment, dans le contexte actuel de crise sanitaire, des moyens utilisant l'IA ont été mis à profit dans la lutte contre le coronavirus notamment dans la prédiction de la propagation du virus à l'aide de l'analyse des données récoltées [15]. En France, deux outils numériques ont été mis à disposition des professionnels de santé SI-DEP et Contact COVID [1]. L'un permet de récolter les données d'identification du patient atteint et l'autre réunit automatiquement l'ensemble des résultats des tests réalisés par plus 600 laboratoires d'analyse publics et privés. La récolte

de ces données permettra d'assurer un suivi épidémiologique fiable et de prévenir la naissance de nouveau cluster. Le médecin traitant est l'interlocuteur des patients pour faire le test et identifier les "personnes contacts" et l'enregistrer dans la base de données "Contact COVID". Cette tâche est automatisée dans certains pays grâce à l'IA, par exemple, en Italie, une entreprise a développé une application permettant de reconstituer l'itinéraire d'un individu atteint d'un virus et d'avertir les personnes ayant eu un contact avec elle avec d'après le constructeur "la garantie de la vie privée" [4]. En France, le gouvernement a travaillé à l'élaboration d'une application "TousAntiCovid" de suivi de contact qui permettrait d'alerter et d'isoler les derniers contacts d'un malade [5]. Les médecins sont-ils au courant de ces expérimentations? Sont-elles source de fascination et d'espoirs ou bien au contraire source de craintes? Qu'en est-il des patients? D'après nos recherches récentes y compris dans la littérature grise (thèse, congrès), aucune étude n'a exploré ce sujet auprès de médecins exerçant en soins primaires en France.

### **Quelles sont les représentations des médecins généralistes et patients en 2020, dans le département de la Seine Saint Denis (93), concernant l'utilisation de l'IA en santé ?**

L'objectif est d'explorer les représentations des médecins généralistes et des patients concernant l'utilisation de IA en santé.

#### **Population étudiée**

Deux populations seront incluses dans l'étude : patients et médecins généralistes. Concernant les patients, est éligible tout patient majeur habitant en Seine-Saint-Denis. Concernant les médecins, est éligible tout médecin généraliste exerçant en Seine-Saint-



Denis. Notre étude recrute uniquement des participants majeurs, et ne prévoit pas par ailleurs l'inclusion de personnes vulnérables.

### **Méthode de recrutement des participants à l'étude**

Le *recrutement des médecins* se fera avec pour objectif d'obtenir un échantillon varié sur les critères de diversité suivants (choix empirique) : âge, sexe, ancienneté d'installation, zone d'exercice urbaine ou rurale, et si possible, niveau de connaissance informatique. Dans un premier temps, nous recruterons des médecins généralistes ayant une activité libérale ou salariée sélectionnés au hasard sur les annuaires Pages Jaunes ou Ameli à partir des critères : âge, sexe, zone d'exercice. A noter que ces critères de diversités pourront être modifiés/complétés au fur à mesure de la réalisation et de l'analyse des entretiens. Devant la difficulté d'obtenir une variété sur les certains critères uniquement à partir d'annuaire, le recrutement sera complété par de praticiens proposés par des médecins déjà inclus correspondant aux "profils manquants". Le lieu de recrutement sera principalement le département de la Seine-Saint-Denis (93). Un questionnaire de caractérisation sera soumis pour recueillir les critères de diversité : âge, sexe, degré de connaissance en nouvelles technologies, zone d'exercice, type d'exercice.

Le *recrutement des patients* se fera sur proposition de leur médecin généraliste. Nous demanderons aux médecins généralistes de nous proposer une liste de patients potentiellement interrogeables. Le recrutement pourra être complété en allant solliciter des associations de patients afin d'obtenir un échantillon varié sur les critères de diversité définis dans le questionnaire de caractérisation. Le lieu de domicile est le département de la Seine Saint-Denis (93). Un questionnaire de caractérisation sera soumis pour recueillir les critères de

diversité : âge, sexe, degré de connaissance en nouvelles technologies, géographique/territoire de vie, niveau socio-économique, maladie chronique ou non.

La *date de début du recrutement prévisible* est septembre 2020, sous réserve de l'obtention de l'ensemble des accords réglementaires et universitaires.

### **Méthode de recueil des données**

Nous avons choisi une méthode de recueil des données par entretiens individuels semi-structurés. L'entretien individuel permet une expression plus libre de l'individu qu'un questionnaire structuré auto-administré. L'entretien permettra ainsi l'émergence d'idées nouvelles que nous n'aurions pas envisagées en début de l'étude. Nous avons donc établi un guide d'entretien avec 5 questions ouvertes et 2 questions de relance. Les entretiens seront enregistrés avec un téléphone mobile personnel grâce à l'aide d'un dictaphone après accord oral du participant, puis sauvegardés dans un fichier audio une clé USB chiffrée. Les enregistrements seront effacés après retranscription. Un pseudonyme sera attribué à chaque participant (type M1 ou P1) pour faire le lien entre la retranscription de l'entretien et la vignette sociodémographique.

### **Méthode d'analyse des données**

Les entretiens seront retranscrits mot à mot sur le logiciel Word par l'investigateur sous forme d'un fichier texte. Toutes les données permettant d'identifier des personnes seront supprimées. Les retranscriptions sous forme de verbatim constitueront les données de base pour l'analyse dans notre étude. Le logiciel NVivo (QSR international, Melbourne, Australie) sera utilisé pour aider l'analyse des données. Chaque retranscription d'entretien sera découpée en unités de sens puis codée sous formes



d'étiquettes (analyse descriptive) et regroupée ensuite selon des propriétés théoriques caractérisant le phénomène étudié (analyse thématique). Un double codage sera réalisé indépendamment par deux internes. Les résultats des deux codages seront ensuite comparés et mis en commun. Ce double codage permettra la triangulation des chercheurs et ainsi un renforcement de la validité interne de l'étude. La méthode d'analyse sera qualitative et se fera sur la base d'une analyse thématique selon une matrice SWOT. L'intérêt dans notre étude était de pouvoir visualiser et clarifier les idées de manière logique et de pouvoir en dégager des axes stratégiques [4]. L'analyse sera réalisée après chaque entretien, au fur et à mesure, afin d'une part de pouvoir adapter le guide d'entretien et afin d'autre part de pouvoir détecter la suffisance des données. Ils seront arrêtés lorsque la suffisance des données sera atteinte, c'est-à-dire, à partir du moment où aucune nouvelle thématique ne semblera plus émerger de l'analyse des données, pendant au moins deux entretiens.

## Références

- [1] "Contact Covid" : arrêter les chaînes de transmission du virus pour stopper l'épidémie. [Internet] Disponible [ici](#).
- [2] Discours de madame Agnès Buzyn - colloque « données de santé et intelligence collective » - lundi 18 novembre 2019. [Internet] Disponible [ici](#).
- [3] DU intelligence artificielle IA appliquée en santé. [Internet] Disponible [ici](#).
- [4] IA et lutte contre le coronavirus Covid-19. [Internet] Disponible [ici](#).
- [5] Info coronavirus COVID-19 - application TousAntiCovid. [Internet] Disponible [ici](#).
- [6] Intelligence artificielle - LAROUSSE. [Internet] Disponible [ici](#).
- [7] Intelligence artificielle et santé | Inserm - la science pour la santé. [Internet] Disponible [ici](#).
- [8] L'académie nationale de médecine s'engage pour le numérique et l'intelligence artificielle en santé. [Internet] Disponible [ici](#).
- [9] Rapport de Cédric Villani : donner un sens à l'intelligence artificielle (IA). [Internet] Disponible [ici](#).
- [10] Recommandations sur les data et l'intelligence artificielle. [Internet] Disponible [ici](#).
- [11] Mireille Blais and Stéphane Martineau. L'analyse inductive générale : Description d'une démarche visant à donner un sens à des données brutes. *Recherches Qualitatives*, 26(2) :1-18.
- [12] Camille Bourdairé-Mignot and Tatiana Gründler. Intelligence artificielle et robotisation : la performance de l'IA au prix de la relation humaine? Publisher : Centre de recherches et d'études sur les droits fondamentaux (CREDOF).
- [13] Jeffrey De Fauw and al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9) :1342-1350.
- [14] Michael Phillips, Helen Marsden, Wayne Jaffe, Rubeta N. Matin, Gorav N. Wali, Jack Greenhalgh, Emily McGrath, Rob James, Evmorfia Ladoyanni, Anthony Bewley, Giuseppe Argenziano, and Ioulios Palamaras. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Network Open*, 2(10).
- [15] Raju Vaishya, Mohd Javaid, Ibrahim Haleem Khan, and Abid Haleem. Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome : Clinical Research & Reviews*, 14(4) :337-339.



## ■ Le LIFAT et CETU ILIAD3 œuvrent conjointement en santé et IA

LIFAT / CETU ILIAD 3

Université de Tours

[www.lifat.univ-tours.fr](http://www.lifat.univ-tours.fr) - [www.liad3.univ-tours.fr](http://www.liad3.univ-tours.fr)

**Jean-Yves RAMEL**

[ramel@univ-tours.fr](mailto:ramel@univ-tours.fr)

**Barthélémy SERRES**

[serres@univ-tours.fr](mailto:serres@univ-tours.fr)

Les préoccupations scientifiques du LIFAT concernent la conception de modèles et de méthodes pour l'extraction d'informations et de connaissances dans des ensembles de données, la résolution de problèmes d'apprentissage automatique et d'optimisation combinatoire avec la volonté d'obtenir de bons résultats dans un temps acceptable et en intégrant des considérations d'interaction homme-machine. La santé constitue un domaine d'application privilégié du laboratoire qui peut s'adosser sur le Centre d'Expertise et de Transfert Universitaire (CETU) ILIAD3 pour répondre à de nombreux projets de recherche pluridisciplinaires menés avec les organismes régionaux de recherche en santé humaine et animale.

### **Dans le domaine de la segmentation d'images 3D neuro-anatomique avec l'équipe IBrain de l'Inserm et l'INRAE de Nouzilly**

Plusieurs projets menés au LIFAT, en collaborations avec l'équipe PRC de l'INRAE et l'équipe IBrain de l'Inserm, ont pour objectifs de décrire et de comparer les structures ou connexions de l'encéphale de différents modèles animaux ou humains. La plateforme SILA 3D est développée dans ce cadre et propose une nouvelle méthode de segmentation interactive et incrémentale d'images médicales 3D. Dans ce travail, une nouvelle manière, plus locale, de modéliser les caractéristiques visuelles et anatomiques des structures ou organes à segmenter a été proposée (thèse de G. Galisot [1]). Cette modélisation s'inspire des mé-

thodes se basant sur des atlas pour guider la segmentation, mais propose de l'instancier de manière plus locale en permettant l'apprentissage de modèles locaux dédiés chacun uniquement à une des structures d'intérêt composant un organe donné. Les modèles sont définis localement sur une zone autour de chaque région d'intérêt. Il est ainsi possible d'imaginer utiliser des modalités différentes pour apprendre le modèle de chaque région, voire même d'utiliser des types d'algorithmes de segmentation différents (CNN, Atlas...). Les modèles locaux appris sont liés structurellement les uns avec les autres dans une représentation nommée "Graphe de Modèles Anatomiques". Les nœuds du graphe servent à stocker l'ensemble des modèles locaux tandis que les arcs du graphe servent à stocker les relations spatiales existantes entre les différentes régions. Les informations de tailles relatives et de distances entre les bords des boîtes englobantes de chaque région sont exploitées pour cela. Cette représentation "Graphe de Modèles Anatomiques" peut être apprise à partir de très peu d'images et renferme à la fois les informations structurales apprises (relations spatiales) et les informations visuelles (modèles locaux) associées à chaque région de manière répartie et indépendante. Cette modélisation est générique et peut être appliquée à différents organes ou types d'images médicales 3D [2]. Lors de la segmentation, le graphe est utilisé suivant un processus incrémental et il devient possible de réaliser des segmentations partielles, rapides en faisant varier ou en choisissant l'ordre de segmentation des différentes régions représentées dans le



graphe. Outre le choix de l'ordre d'extraction des régions, l'utilisateur peut intervenir avant et après chaque segmentation locale afin de corriger le positionnement et la qualité de la segmentation obtenus/proposés automatiquement (Fig. 1.4).

Dans un contexte similaire, le LIFAT et le CETU ILIAD3 participent au projet Fibratlas et à ses extensions visant à reconstruire les principaux faisceaux de fibres blanches pour valider la technique de tractographie (détection de fibres par imagerie IRM). Il s'agit ici de développer des outils permettant aux spécialistes du cerveau de travailler sur des acquisitions 3D successives produites lors du suivi d'une dissection. L'outil développé permet aux anatomistes une navigation interactive et immersive dans les données 3D ainsi que la documentation et l'extraction de connaissances à partir des données de dissection [3]. Après avoir scanné la surface de dissection au moyen d'un scanner laser 3D, ces différentes surfaces sont recalées et texturées avec une photographie haute-résolution avant d'être mises à disposition de l'expert. L'expert neuro-anatomiste peut ainsi naviguer entre les étapes destructrices de la dissection qu'il a effectuée. À l'issue d'une étape d'étiquetage manuel, les différents segments 3D de faisceaux identifiés sont assemblés pour fournir une reconstruction du faisceau d'intérêt. Les algorithmes de tractographie (détection de fibres) reposent sur de nombreux paramètres et produisent de ce fait des résultats de faisceaux de fibres très différents selon leur paramétrage initial. Un second objectif de ce travail est d'étudier ces variations et de proposer une méthode semi-automatique de détermination du paramétrage initial optimal. Une méthode se basant sur des algorithmes génétiques a été proposée et des outils de comparaison de faisceaux de fibres ont également été développés pour quantifier les écarts entre des résultats issus de l'imagerie (IRM de diffusion,

tractographie) et les vérités terrain (thèse de M. Sta [4]). La tractographie est une technique utilisée aujourd'hui en routine clinique pour extraire les faisceaux de fibres à partir d'images IRM de patients. Une connaissance de l'organisation de ces faisceaux chez le patient est nécessaire aux neurochirurgiens en phases pré- et per-opératoire pour minimiser les dommages lors de la résection de tumeurs (gliomes).

Des travaux sur l'étude de l'arborescence vasculaire sont également menés sur des images issues de l'imagerie IRM très hauts champs, en collaboration avec l'équipe de Neurospin (CEA-Saclay). Un pipeline de traitement automatisé a été mis en place pour extraire les vaisseaux visibles dans les différentes modalités à partir d'un volume haute-résolution constitué lors de découpes progressives du spécimen (*block face imaging*) afin de permettre une comparaison de l'imagerie à la vérité terrain. La difficulté réside dans la phase de segmentation des vaisseaux qui sont de petites structures à faible contraste. L'ensemble de ces travaux académiques sont menés en lien étroit avec le pôle de neurochirurgie et le laboratoire d'anatomie du CHRU de Tours. La connaissance de l'organisation vasculaire dans les zones corticales et sub-corticales est une donnée importante pour expliquer le développement des fibres lors de la maturation cérébrale. Ces travaux participent également à une meilleure compréhension de la survenue de pathologies sur les vaisseaux cérébraux.

### **Dans la réalité augmentée en chirurgie prothétique de l'épaule avec le CHRU de Tours**

De plus en plus, les techniques d'imagerie 3D contribuent à solutionner différentes difficultés opératoires et ainsi aident à la préparation de l'intervention. La réalité augmentée constitue désormais la dernière technologie d'assistance opératoire chirurgicale en dé-

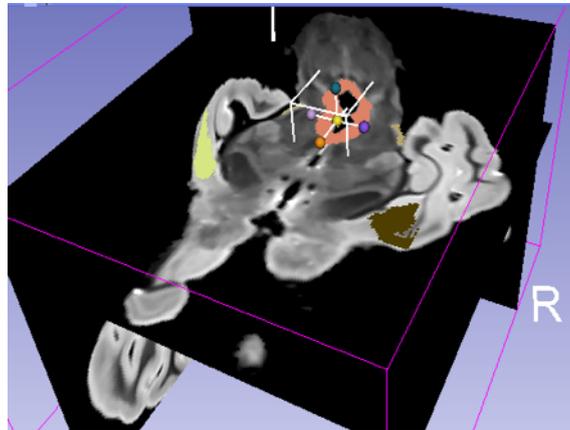


Figure 1.4 – Segmentation interactive et incrémentale.

veloppement. Elle a pour challenge de délivrer des informations virtuelles pertinentes aux chirurgiens, adaptées à la situation chirurgicale en temps réel. Ces informations peuvent être un support à la réalisation du geste chirurgical lors des différentes étapes de l'opération. L'objectif final est d'augmenter la précision et la reproductibilité de l'implantation prothétique, quels que soient l'expérience des chirurgiens et le niveau de difficulté technique du cas à opérer. Dans ce cadre, le LIFAT collabore depuis de nombreuses années avec le Pr Julien Berhouet du Centre de Chirurgie Orthopédique et Traumatologique de Tours – Universitaire (CCOTT-U Tours). Ce travail s'est notamment traduit par une thèse CIFRE soutenue en 2020 [5] (menée avec la startup IMAS-CAP, pionnière dans le développement de solutions de réalité mixte appliquée à l'orthopédie). Tout d'abord, la problématique de calibrage d'un couple de caméras RGB-D très précis, nécessaire pour l'affichage en réalité augmentée en chirurgie, a été étudié. Une nouvelle méthode de calibrage RGB-D complète, utilisant des sphères comme objets de référence, a été produite. La seconde problématique étudiée est celle du suivi de pose d'objets, nécessaire pour obtenir un affichage cohérent des objets

virtuels (les guides) ajoutés dans le champ de vision des chirurgiens et ce, malgré les mouvements de tête des chirurgiens et de l'épaule des patients. Une nouvelle approche d'application du suivi, par l'apprentissage d'un modèle 3D reconstruit de l'objet à suivre, a été proposée. Des informations contextuelles de voisinage sont intégrées au modèle pour améliorer les performances de suivi. Une nouvelle stratégie d'apprentissage est utilisée, permettant d'offrir des résultats similaires à d'autres approches de l'état de l'art pour un temps d'apprentissage significativement réduit. Ce résultat permet d'envisager son exécution sur un matériel de puissance modeste, et tout juste en amont de la phase de suivi. Pour chacune de ces contributions, des ensembles de données de test spécifiques ont été créés, et de nombreuses expérimentations ont permis de mettre en avant les apports des approches proposées par rapport à l'état de l'art. Bien que ces travaux se concentrent sur le contexte applicatif d'une chirurgie de l'épaule, les méthodes proposées sont suffisamment génériques pour envisager leur usage dans d'autres contextes.



**AfIA**

Association française  
pour l'Intelligence Artificielle

## Dans le domaine de l'ordonnancement et de l'optimisation avec le CHRU de Tours

Parmi les multiples facettes du "diamant AfIA", une facette concerne les outils et les langages pour l'IA dans laquelle se reflète la Recherche Opérationnelle (RO). Plusieurs membres de l'équipe ROOT (ERL CNRS 7002) du LIFAT travaillent sur la thématique de la RO avec comme domaine d'application la santé. Leurs recherches visent à développer des outils d'aide à la décision pour résoudre des problèmes d'optimisation combinatoire auxquels les gestionnaires et coordinateurs en santé font face. Ces problématiques décisionnelles et/ou organisationnelles sont complexes en raison de la particularité et des nombreuses contraintes du milieu de la santé, de leurs impacts sur les patients et les coûts de gestion, du caractère incertain des activités de santé, des ressources limitées et spécifiques au milieu, ou encore de la nature dynamique des problèmes. Dans le passé, plusieurs travaux de recherche ont porté sur la logistique hospitalière, le transport de patients, l'analyse comparative de la métabolomique ciblée ou encore la planification de soins à domicile. Actuellement, plusieurs projets de recherche sont en cours, dont les trois suivants :

- Un projet portant sur la production et la livraison de chimiothérapie. Ce projet a commencé il y a plusieurs années par le développement d'outil d'aide à la planification de la production de chimiothérapie. Puis il s'est poursuivi en intégrant les aspects de livraison dans le problème d'optimisation [6]. Actuellement, les travaux portent toujours sur ces aspects de problème intégré production-distribution [7] mais aussi sur la gestion des reliquats des molécules cytotoxiques lors de la production. Un projet financé par la région Centre Val de Loire et une thèse sont actuellement en cours sur ce sujet ;

- Un tout autre projet concerne la gestion de flotte d'ambulances pour répondre à des demandes de transports urgentes de patients et non urgentes. Ce projet s'inscrit dans le cadre d'une collaboration avec des chercheurs de Montréal et a pour objectif de développer et d'évaluer des outils de gestion d'une flotte dans un service médical d'urgence (EMS) de type "nord-américain" [8] ;
- Enfin, un dernier projet s'intéresse aux plateformes ambulatoires et à la planification d'actes médicaux afin de diagnostiquer ou soigner des patients en hôpital de jour. Cette série d'actes, ou parcours patients, est généralement définie et nécessite différentes ressources humaines et matérielles nécessitant un pilotage et une coordination en temps réel. Des travaux préliminaires ont commencé pour le développement d'outils intelligents d'aide à la décision pour la conception, l'optimisation et la gestion de ces structures médicales ambulatoires. Une thèse sur ce sujet devrait bientôt démarrer.

## Références

- [1] G. Galisot . Segmentation incrémentale et interactive d'images médicales 3D. Thèse de doctorat en informatique, Université de Tours, mars 2018. <https://lifat.univ-tours.fr/phd-defense-gaetan-galisot>
- [2] X. Zhuang, G. Galisot, et al. Evaluation of algorithms for Multi-Modality Whole Heart Segmentation : An open-access grand challenge. Medical Image Analysis. Vol 58. 2019
- [3] R. Zemmoura, B. Serres, F. Andersson, L. Barantin, C. Tauber, et al. FIBRAS-CAN : A novel method for 3D white matter tract reconstruction in MR space from cadaveric dissection. NeuroImage, Vol 103, pp.106-118. 2014. Elsevier.
- [4] M. Sta. Comparaison tractographie IRM - tissu cérébral et optimisa-



- tion de la reconstruction tractographique par algorithme génétique, Université de Tours, sept. 2017. <http://www.theses.fr/2017TOUR3305>
- [5] D. Boas. Réalité Augmentée pour la chirurgie de l'épaule. Thèse de doctorat en informatique. Université de Tours, juil. 2020. <http://www.theses.fr/s238667>
- [6] J.C. Billaut, V. André, Y. Kergosien, and J.F. Tournamille. Optimization Issues in Chemotherapy Delivery. *Health Efficiency*, pages 91–118. Elsevier, 2018.
- [7] A. Robbes, Y. Kergosien, and J.C. Billaut. Multi-level heuristic to optimize the chemotherapy production and delivery. In *International Conference on Human-Centred Software Engineering*, pages 263–273, May 2019. Springer, Cham.
- [8] V. Bélanger, Y. Kergosien, A. Ruiz, and P. Soriano. An empirical comparison of relocation strategies in real-time ambulance fleet management. *Computers & Industrial Engineering*, 94, pages 216–229. 2016.



## ■ Laboratoire LAMIH : Recherches en Santé & IA

LAMIH CNRS UMR 8201  
Université Polytechnique Hauts-de-France  
<https://www.uphf.fr/LAMIH>

**Maxence BIGERELLE**

**Patrice CAULIER**

**Sondes CHAABANE**

**Véronique DELCROIX**

**Emanuelle GRISLIN**

**Sylvain PIECHOWIACK**

[prenom.nom@uphf.fr](mailto:prenom.nom@uphf.fr)

### Introduction

Le LAMIH (Laboratoire d'Automatique, de Mécanique et d'Informatique industrielles et Humaines) est une Unité Mixte de Recherche (UMR 8201) entre l'UPHF (Université Polytechnique Hauts-de-France) et le CNRS. Il compte près de 250 enseignants-chercheurs, ingénieurs, post-doctorants et doctorants répartis dans quatre départements : Automatique, Mécanique, Informatique et Sciences de l'Homme et du Vivant. Nous décrivons ici les recherches actuelles, mais aussi passées, menées au sein du LAMIH, dans le domaine de l'intelligence artificielle appliquée au secteur de la santé. Ces recherches concernent la pratique médicale mais aussi son organisation.

### Prédiction de la durée de séjours hospitaliers

La durée de séjour est un indicateur de la performance des services hospitaliers. Sa maîtrise contribue à l'optimisation des ressources, à l'amélioration de l'organisation des soins et à une meilleure planification des activités. Dans ce cadre et en partenariat avec la société [Alicante](#), nous travaillons au développement d'un outil de prédiction de la durée de séjour [1]. Après avoir modélisée la notion de « séjour hospitalier » et analysé différents algorithmes de

prédiction de sa durée, nous avons défini une approche générique de prédiction de la durée de séjour. Fondée sur la régression, cette approche conjugue des traitements à la fois hors ligne et en ligne, depuis l'admission puis au gré des actes médicaux. L'approche s'applique à des séjours internes à une unité médicale mais aussi à des séjours inter-unités médicales (thèse en cours).

### Prévision de la demande dans les services d'urgence

La prévision de la demande dans les services d'urgences constitue un enjeu majeur de la gestion de ces services. Elle permet d'anticiper des situations de tension qui engendrent une surcharge de travail laquelle impacte la qualité des soins et engendre des situations de stress chez les soignants et les patients. Pour répondre à cette problématique, nous avons proposé un modèle de prévision des arrivées journalières des patients basée sur la méthode ARIMA (*AutoRegressive Integrated Moving Average*) [2]. Ce modèle a été appliqué à des données réelles issues du Service des Urgences Pédiatriques (SUP) du CHRU de Lille. Les résultats obtenus ont été acceptables. En effet, pour la série totale, l'erreur moyenne de la prévision était de 2,66 (3 patients) avec un taux d'erreur de prévision de 3,79%. Le taux d'erreur maximum



**Afia**

Association française  
pour l'Intelligence Artificielle

était de 9,03% pour un horizon de cinq jours. Dans une autre étude, nous nous sommes intéressés à la prévision du nombre d'arrivée par heure au SUP. Lors de cette étude, nous avons développé un modèle basé sur la méthode SARMA (*Seasonal autoregressive moving average*). La valeur élevée du coefficient de détermination et la faible valeur de l'erreur absolue moyenne en pourcentage du modèle de référence montrent que ce dernier est très représentatif de la série chronologique observée.

### **Interprétation clinique de données**

L'objectif du travail de recherche consistait à, d'une part, assister le chirurgien orthopédiste dans l'évaluation objective des résultats de l'arthroplastie gléno-humérale et, d'autre part, améliorer cette pratique [3]. Une technique fiable d'évaluation de la cinématique du membre supérieur a d'abord été établie. Ensuite, une aide à l'interprétation clinique des données de la cinématique a été réalisée. Elle repose sur l'induction de règles floues à partir de forêts d'arbres de décision flous. La capacité d'explication des règles induites est évaluée par une mesure de satisfaisabilité. L'aide a été éprouvée dans le cadre de la mise en correspondance du mouvement avec l'état fonctionnel des patients, déterminé subjectivement. Les résultats ont montré une bonne capacité d'explication de l'état fonctionnel des patients n'ayant pas participé à l'apprentissage.

### **Aide à la décision pré-opératoire**

L'aide à la décision pré-opératoire porte sur la chirurgie orthopédique de la scoliose. Cette thérapie nécessite une analyse pré-opératoire et approfondie des données cliniques et géométriques du patient afin de déterminer les niveaux vertébraux de fixation de l'arthrodèse [4]. Pour ce faire, une des méthodes appliquée par le chirurgien expérimenté consiste à réutiliser des cas antérieurs opérés. Dans ce sens, nous

avons développé un système d'assistance fondé sur les étapes du cycle de raisonnement à partir de cas. L'exploitation du système d'assistance montre que l'étape de récupération des cas les plus similaires dans la base conditionne la performance de résolution de problèmes du système. Ainsi, nous proposons trois types d'organisation de la mémoire.

### **Aide à la prévention des chutes**

La prévention des chutes des personnes âgées est un enjeu majeur de santé publique. En France, la chute est la première cause de mortalité accidentelle chez les personnes âgées de plus de 65 ans. La prévention des chutes est un problème multifactoriel qui implique d'évaluer les facteurs de risques de chutes de la personne. Cela nécessite du temps, de l'expertise et des équipements. L'objectif du projet est de proposer un outil d'aide à la prévention des chutes, destiné en premier lieu aux médecins traitants. Il doit permettre de : (1) cibler le recueil d'informations utiles, (2) capitaliser les informations au fil du temps, (3) évaluer les facteurs de risque de chute et (4) fournir des recommandations adaptées à la personne. Dans un premier temps, nous avons défini une ontologie des facteurs de risques de chutes avec le Pr. Puisieux, responsable du pôle gériatrie du CHRU de Lille, et du service sur la chute [5]. L'étape suivante est la construction d'un modèle de raisonnement intégrant les observations incomplètes pour évaluer les facteurs de risque de chute. En effet, les observations disponibles sur une personne sont à la fois incomplètes et variables d'une personne à l'autre. Pour atteindre cet objectif, nous utilisons des modèles graphiques probabilistes. Un réseau bayésien a été élaboré sur la base des données du CHU de Lille. Il a été comparé avec d'autres classificateurs pour évaluer les facteurs de risque de chute à partir d'observations partielles sur la personne [6]. En parallèle, nous travaillons sur



la gestion d'une base d'informations personnelles, datées, incomplètes et incertaines, afin de caractériser la pertinence et la confiance à accorder aux informations en fonction de leur ancienneté (thèse en cours).

### **Texturation intelligente des implants endo-osseux**

Notre projet "MUST<sup>implant</sup>" regroupe quatre laboratoires de recherche académique qui mettent en commun leurs compétences pour innover un nouveau concept dans le domaine de la science des surfaces : le *Multiscale Smart Texturing* [7]. Prenant en compte quantitativement un vaste ensemble de données rencontrées dans la bibliographie, le concept consiste à concevoir une surface fonctionnelle optimale grâce à une méthodologie originale basée sur un nouvel outil de gestion des connaissances. Cette approche générique de texturation de surface sera appliquée dans le domaine de la texturation intelligente des implants endo-osseux pour augmenter leur performance et leur sécurité. Après la fabrication d'implants présentant ces topographies optimales grâce à des techniques de fabrication innovantes basées sur le Femtolaser, ils seront testés mécaniquement, chimiquement et biologiquement pour fournir des preuves de concept et des topographies ostéoconductrices (recherche en cours).

### **Références**

- [1] R.N. Mekhaldi, P. Caulier, S. Chaabane, S. Piéchowiak, J. Taillard, et A. Hansske. Apprentissage automatique dans la prédiction des durées de séjours hospitaliers. In *Actes de la 10<sup>ème</sup> Conférence Francophone en Gestion et Ingénierie des Systèmes Hospitaliers*, GISEH 2020, Valenciennes, France, 26-29 octobre, 2020.
- [2] F. Kadri, H. Fouzi, S. Chaabane, et C. Tahon. Time Series Modelling and Forecasting of Emergency Department Overcrowding. *Journal of Medical Systems*, 38(9), 107, 2014.
- [3] E. Roux, A-P. Maquinghen, P. Caulier, et D. Boutens. A support method for the contextual interpretation of biomechanical data. *IEEE Transactions on Information Technology in Biomedicine*, 10(1), pages 109–118, 2006.
- [4] S. Chusseau, D. Chopin, A-P. Maquinghen, et P. Caulier. An assistance system to determine fusion levels in idiopathic scoliosis surgical treatment using case-based reasoning. *Scoliosis European Society*, Innsbruck, Austria, june 1998.
- [5] F. Essghaier, V. Delcroix, K. Oliveira, F. Puisieux, C. Gaxatte, et P. Pudlo. Towards a fall prevention system design by using ontology. In *5<sup>ème</sup> Atelier IN-OVIVE, INTégration de sources/masses de données hétérogènes et Ontologies, dans le domaine des sciences du VIVant et de l'Environnement*, Plate-Forme PFIA, IC, Toulouse, France, juillet 2019.
- [6] G. Sihag, V. Delcroix, E. Grislin, X. Siebert, S. Piechowiak, et F. Puisieux. Prediction of Risk Factors for Fall using Bayesian Networks with Partial Health Information. In *IEEE GLOBECOM 2020 Workshop on AI-driven Smart Healthcare*, Taipei, Taiwan, december, 2020.
- [7] L. Pieuchot, J. Marteau, A. Guignandon, T. Dos Santos, I. Brigaud, P.F. Chauvy, M. Bigerelle, et K. Anselme. Curvotaxis directs cell migration through cell-scale curvature landscapes *Nature communications*, 9(1), pages 1–13, 2018.



**AfIA**  
Association française  
pour l'Intelligence Artificielle

---

## Comptes rendus de journées, événements et conférences

---



## ■ Journée « Enseignement et Formation en Intelligence Artificielle »

Par

**Sandra BRINGAY**  
LIRMM/ADVANCE  
Université Paul-Valéry Montpellier  
[Sandra.Bringay@lirmm.fr](mailto:Sandra.Bringay@lirmm.fr)

**Catherine FARON**  
I3S/WIMMICS  
Université Côte d'Azur  
[faron@unice.fr](mailto:faron@unice.fr)

**Gauthier PICARD**  
DTIS  
ONERA  
Université de Toulouse  
[gauthier.picard@onera.fr](mailto:gauthier.picard@onera.fr)

L'AfIA a organisé sa deuxième journée « Enseignement et Formation en IA » le 7 Janvier 2021. Prévue à l'Université Paris Nanterre, amphithéâtre Weber, cette journée a finalement été virtualisée du fait des contraintes sanitaires.

La journée avait deux objectifs : d'une part, réaliser un tour d'horizon des besoins des entreprises et de la société civile. D'autre part, dresser un état de l'art des formations académiques en IA qui existent en France et qui participent à la démocratisation de l'utilisation d'IA.

### Programme de la journée

**09h00.** « Ouverture et introduction », par Benoit LE BLANC, Sandra BRINGAY, Catherine FARON et Gauthier PICARD (AfIA)

#### Session 1 : Cartographie de l'enseignement en IA

**09h15.** « Stratégie nationale sur la formation en IA », par Laurent VERCOUTER (responsable formation PNRIA INRIA) et Anne BOYER (conseillère scientifique IA DGESIP MESRI)

**09h45.** « Vers l'ajout de compétences IA dans la définition de la littératie numérique à l'Unesco », par Colin DE LA HIGUERA (U. Nantes et UNESCO)

**10h30.** Pause

#### Session 2 : L'introduction de l'IA dans les contenus pédagogiques

**10h30.** « Le De Vinci Innovation Center : un laboratoire d'enseignement supérieur expérimental et trans-disciplinaire », par Clément DUHART (ESILV La Défense et MIT Media Lab Boston)

**11h00.** « IA centrée sur l'humain dans les industries et la création », par Sotiris MANITSARIS (Mines Paristech)

**11h30.** « IA et informatique visuelle à l'X », par Erwan SCORNET (École Polytechnique Palaiseau)

**12h00.** « Ouverture du département 2IA Informatique et IA », par Sylvie RANWEZ (IMT Mines Alès)

**12h30.** Pause

#### Session 3 : Les besoins des entreprises

**14h00.** « L'IA au cœur des outils de gestion ! Comment passer du buzz à une vision opérationnelle, les défis de formations et acculturation interne », par Youssef MILOUDI (Berger-Levrault Lyon)

**14h30.** « Stratégie de développement des compétences en IA pour les ingénieurs », par Bruno CARRON (Airbus Defence and Space Elancourt)

**15h00.** Pause

#### Session 4 : Les nouvelles formes de l'enseignement de l'IA

**15h30.** « 42 : le *peer-learning* au service de l'IA », par Olivier CROUZET (École 42 Pa-



ris)

**16h00.** « Quand les collégiens et les lycéens font du machine learning », par Romain LIBLAU (Magic Makers Paris)

**16h30.** Clôture

## Résumé des interventions

Cette deuxième édition a réuni entre cinquante et soixante quinze participants tout au long de la journée, autour de l'enseignement et de la formation en intelligence artificielle. Après un discours de bienvenue et de présentation de l'association par Benoit LEBLANC, président de l'Afia, la thématique et le déroulement de la journée ont été brièvement introduits par Catherine FARON.

Lors de la première session, animée par Gaël DIAS, dédiée à la cartographie de l'enseignement de l'IA, la stratégie nationale française sur la formation en IA a tout d'abord été présentée par Laurent VERCOUTER, professeur à l'INSA Rouen Normandie et responsable en charge de la formation du Programme National de Recherche en IA (PNRIA, INRIA). Le point de vue de l'UNESCO sur l'enseignement de l'IA en tant qu'élément de la littératie numérique a été également abordé par Colin DE LA HIGUERA, professeur à l'Université de Nantes et titulaire de la Chaire Unesco en technologie pour la formation des enseignants par les ressources éducatives libres.

La deuxième session, animée par Gauthier PICARD, a été dédiée à l'introduction de l'IA dans les contenus pédagogiques avec la présentation de quatre formations de niveau universitaire. Clément DUHART, créateur et responsable du De Vinci Innovation Center, rattaché à l'école d'ingénieurs du Pôle Léonard de Vinci nous a présenté le principe innovant sur lequel repose la formation à savoir l'antidisciplinarité qui permet un travail exploratoire des étudiants avec un haut niveau de liberté. Sotiris MANITSARIS, chercheur senior, a présenté

comment l'IA centrée sur l'humain et les industries de la création sont abordées à l'école des Mines Paritech. Erwan SCORNET, maître de conférences au Centre de Mathématiques Appliquées (CMAP) de l'École Polytechnique a présenté comment l'IA et l'informatique visuelle sont abordés à l'X. Sylvie RANWEZ, professeur, a finalement décrit l'ouverture du département 2IA Informatique et IA à l'école des Mines d'Alès. Les compétences à acquérir par les étudiants sont évidemment techniques mais également liées à des réflexions éthiques, en lien avec le *green IT*.

La troisième session, animée par Sandra BRINGAY, a été dédiée aux besoins des entreprises en terme de compétences, de recrutements et de formations des personnels. Youssef MILOUDI, chercheur chez Berger-Levrault, a présenté la place de l'IA au cœur des outils de gestion développés par l'entreprise et en particulier comment passer du buzz à une vision opérationnelle. Bruno CARRON, chercheur chez Airbus Defence and Space, a présenté la stratégie de cette dernière pour le développement des compétences en IA de ses ingénieurs. Ces présentations ont permis de mieux comprendre les défis en termes de formations et d'acculturation interne à relever au sein de ces entreprises.

La quatrième session, animée par Catherine FARON, a porté sur les nouvelles formes de l'enseignement de l'IA. La session a commencé par une présentation de personne Olivier Crouzet, directeur de l'école 42 à Paris, qui a montré comment est mise en œuvre dans cette école la pédagogie innovante de *peer-learning* au service de l'IA. Romain LIBLAU, responsable pédagogie et innovation chez Magic Makers Paris, a ensuite présenté un programme qui permet de faire comprendre les principes et utiliser des algorithmes de *machine learning* à des collégiens et lycéens sans pré-requis mathématiques et sans entrer dans la complexité des



algorithmes et de la programmation. Ces deux présentations ont mis en lumière l'importance de savoir apprendre à apprendre et l'efficacité de l'approche par projet.

Les participants ont été très actifs tout au long de la journée, posant de nombreuses questions ouvrant à la discussion et la journée s'est conclue notamment sur la nécessité de définir 1) une typologie des compétences en IA attendues selon le public, des enfants aux étudiants et aux professionnels, des spécialistes aux non spécialistes, et 2) des approches pédagogiques

innovantes adaptées à l'enseignement de l'IA.

Les présentations des intervenants de cette journée sont accessibles sur le site de l'[AfIA](#).

### La suite...

Cette journée, organisée par le groupe de travail « Enseignement » de l'AfIA, sera suivie cette année par des sessions tutorielles à PFIA, du 28 Juin - 2 Juillet 2021, à Bordeaux. Face à son succès et aux besoins réels et aux intérêts exprimés, une troisième édition d'EFIA sera organisée en janvier 2021.

## ■ Journée « Technologies du Langage Humain et Santé »

Par

**Corinne FREDOUILLE**

LIA  
Université d'Avignon

**José G. MORENO**

IRIT  
Université de Toulouse

**Aurélie NÉVÉOL**

LISN  
Université Paris Saclay, CNRS

**Christophe SEVRAN**

QWANT

l'intersection du traitement automatique des langues, de la recherche d'information, de la communication parlée, de l'informatique médicale et de la santé publique.

Cette journée fait suite à d'autres journées communes AfIA/ ATALA organisées entre les années 2012 et 2018 sur les thèmes suivants :

- [Intelligence artificielle et traitement automatique des langues se retrouvent](#) (Mars 2012 - Paris INALCO)
- Langue, apprentissage automatique et fouille de données (Mars 2014 - Paris INALCO)
- [Représentation](#) (Mars 2016 - Paris IHP/UPMC)
- [Apprentissage profond pour le traitement automatique des langues](#) (Juillet 2018 - Nancy)

### Introduction

La Journée « Technologies du Langage Humain (TLH) et Santé » organisée conjointement par le Collège TLH de l'AfIA et par l'Association pour le Traitement Automatique des Langues ([ATALA](#)) a eu lieu en distanciel le 4 février 2021. Elle avait pour objectif de proposer un panorama des recherches réalisées par les laboratoires francophones sur la thématique du **traitement du langage humain et de ses applications dans le domaine de la santé**. Ainsi, nous avons souhaité réunir des collègues issus d'instituts de recherche menant des travaux à

On peut également noter que la conférence [TALN 2019](#) a été organisée à Toulouse dans le cadre de la plateforme AfIA (1-5 juillet 2019), ce qui montre l'importance des liens entre les deux associations.



## Présentation de la journée

Suite à un appel à participation communiqué sur les listes de diffusion françaises des domaines de recherche en TLH et santé fin 2020, nous avons reçu 12 contributions, issues de laboratoires académiques en Belgique, Italie, France et Suisse. La diversité des recherches présentées ainsi que la qualité et la quantité des contributions reçues démontrent à la fois une dynamique importante des TLH dans la communauté francophone. Ainsi, nous avons pu articuler l'organisation de la journée autour de quatre sessions thématiques, d'une présentation invitée et d'une table ronde. Les sessions thématiques ont porté sur la traduction de la parole médicale, le traitement automatique de la langue clinique, le traitement automatique de la langue pathologique et la recherche d'information en santé.

Du fait de l'évolution de la situation sanitaire, la journée s'est déroulée en distanciel uniquement. Nous remercions l'université d'Avignon pour avoir fourni l'infrastructure BBB de la journée, ainsi que pour son soutien technique. Nous avons reçu plus de 120 demandes d'inscription à la journée. En terme de participation effective, cela s'est traduit par la présence de nombreux participants répartis sur la journée, avec un pic de présence à plus de 70 participants en fin de matinée. Par ailleurs, une dizaine de participants ont profité de l'environnement [Gathertown](#) proposé pour offrir un moment de convivialité pendant la pause déjeuner. La journée a en outre bénéficié d'un public varié issu de laboratoires d'informatique, de linguistique et d'informatique médicale. Nous sommes ravis d'avoir pu réunir à cette occasion la communauté scientifique intéressée par le traitement de la langue biomédicale dans toute sa diversité. Nous remercions à ce titre l'ATALA, l'AfIA, le comité scientifique, les intervenants et le public.

## Programme

**9h15.** « Accueil virtuel et ouverture de la journée », par Corinne FREDOUILLE et Aurélie NÉVÉOL

**9h30.** « Mots des présidents des associations AfIA et ATALA », par Domitile LOURDEAUX représentant Benoit LE BLANC (AfIA) et Christophe SERVAN (ATALA)

### Session « Transcription et traduction de la parole médicale » animée par Christophe SERVAN

**9h40.** « Présentation du projet BabelDr », par Pierrette BOUILLON

**9h50.** « [BabelDr : un système de traduction médicale avec des pictogrammes pour les patients allophones aux urgences et dans un secteur de dépistage COVID-19](#) », par Magali NORRÉ, Pierrette BOUILLON, Johanna GERLACH et Hervé SPECHBACH

**10h08.** « [Vers une communication médicale adaptée aux personnes sourdes en période de confinement](#) », par Bastien DAVID, Pierrette BOUILLON et Hervé SPECHBACH

**10h26.** « [Reconnaissance vocale du discours spontané pour le domaine médical](#) », par Lucía ORMAECHEA GRIJALBA, Pierrette BOUILLON, Johanna GERLACH, Benjamin LECOUTEUX, Didier SCHWAB et Hervé SPECHBACH

**10h45.** Pause

### Session « Traitement de la langue humaine dans la pratique clinique » animée par Aurélie NÉVÉOL

**11h00.** « [Vers un système de dialogue oral pour la saisie de prescriptions médicales](#) », par Ali CAN KOCABIYIKOGLU, François PORTET, Jean-Marc BABOCHKINE et Hervé BLANCHON

**11h20.** « [PyMedExt, un couteau suisse pour](#)



le traitement des textes médicaux », par William DIGAN, Alice ROGIER, David BAUDOIN, Bastien RANCE et Antoine NEURAZ

**11h40.** « Traitement de la langue naturelle pour une réponse rapide aux maladies émergentes: COVID-19 », par Antoine NEURAZ, Ivan LERNER, William DIGAN, Nicolas GARCELON, Rosy TSOPRA, Alice ROGIER, David BAUDOIN, Anita BURGUN et Bastien RANCE

**12h00.** « Traitement Automatique de la Langue et Intégration de Données pour les Réunions de Concertations Pluridisciplinaires en Oncologie », par Nesrine BANNOUR, Aurélie NÉVÉOL, Xavier TANNIER et Bastien RANCE

**12h30.** Pause déjeuner – Gathertown

**Session « Présentation invitée » animée par Mathieu ROCHE**

**13h30.** « Traitement automatique des langues pour la santé : travaux récents au LISN », par Pierre ZWEIGENBAUM

**Session « Traitement de la Langue Humaine appliqué aux pathologies langagières » animée par Corinne FREDOUILLE**

**14h40.** « Investigation des marqueurs langagiers non-lexicaux et spécifiques des personnes souffrant de schizophrénie dans des conversations spontanées », par Chuyuan Li, Maxime Amblard, Chloé Braud, Caroline Demily, Nicolas Franck et Michel Musiol

**15h00.** « Measurements of turn-taking and linguistic behaviors in clinical settings », par Rachid Riad, Lucas Gautheron, Emmanuel Dupoux, Anne-Catherine Bachoud-Lévi et Alejandrina Cristia

**15h20.** « Exploration de la temporalité dans la désignation des pathologies du lan-

gage en orthophonie : aspects cliniques et termino-ontologiques », par Frédérique BRIN-HENRY

**Session « Traitement de la Langue Humaine appliqué à l'information de santé » animée par José G. MORENO**

**15h45.** « Fouille de la littérature médicale à l'aide de graphes », par Elise BIGEARD, Aman SINHA, Marianne CLAUSEL et Mathieu CONSTANT

**16h05.** « La communication en santé publique au temps du Covid-19 dans les contextes français, québécois et tunisien », par Silvia CALVI et Klara DANKOVA

**16h30.** Table ronde animée par Aurélie NÉVÉOL et Christophe SERVAN

**18h00.** Clôture de la journée

---

## Compte-rendu de la table ronde

La table ronde a vu la participation de différents intervenants permettant d'avoir un regard croisé pluridisciplinaire sur la problématique spécifique des TLH sur les données de santé. Cette table ronde a mis en relief les nouveaux défis scientifiques en considérant les dimensions juridiques et éthiques associées. En effet, la prise en compte de ces problématiques constitue un verrou à prendre en compte dans les travaux de recherche en TLH sur lesquels les intervenants ci-dessous ont pu échanger. Les points saillants évoqués ont été : l'accès aux données de santé textuelles dans des langues autres que l'anglais, le biais des données collectées et des modèles qui en découlent, le coût des modèles de langue modernes.

- Sandra BRINGAY a obtenu un doctorat en informatique de l'Université de Picardie Jules Verne. Maître de conférences à l'Université Paul Valéry Montpellier de 2007 à



**Afia**

Association française  
pour l'Intelligence Artificielle

- 2016, elle est professeur dans cet établissement depuis septembre 2016. Elle réalise ses recherches dans le Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM-CNRS-UM). Ses intérêts de recherche actuels incluent la science des données et l'intelligence artificielle appliquées au domaine de la santé.
- Eric BRUNET-GOUET est praticien hospitalier en psychiatrie adulte au Centre Hospitalier de Versailles. Il a conduit des recherches essentiellement en neurosciences cognitives portant sur les troubles de la cognition sociale dans la schizophrénie et leurs répercussions fonctionnelles : troubles de la communication, difficultés dans la vie courante. Dans le cadre du réseau de centres experts FondaMental, il a travaillé à l'exploitation des bases de données de recherche et aussi cliniques portant sur des cohortes de patients souffrant de troubles bipolaires et schizophréniques. Ce travail, dans le cadre de la supervision d'une thèse en santé publique, a conduit à utiliser des modèles statistiques appropriés à des nombre de sujets plus élevés (équations structurales), et donc à s'intéresser aux possibilités d'accroître massivement le recueil d'informations en employant des techniques innovantes comme le traitement du langage naturel appliqué aux données cliniques. Rattaché au Centre Epidémiologie et Santé des Populations (Inserm CESP) et à l'équipe PsyDev (Pr Speranza & Passerieux), il travaille notamment à l'application de modèles pré-entraînés type BERT à des corpus de dossiers cliniques ou de textes, dans la perspective de concevoir des modèles applicables et pertinents en recherche clinique.
  - Hélène GUIMIOT-BRÉAUD est docteur en droit. Elle a exercé des fonctions de juriste dans le secteur de l'industrie pharmaceutique et au sein d'un Centre hospitalier universitaire. Elle a rejoint la CNIL en 2015 et est chef du service de la santé depuis février 2018.
  - Antoine NEURAZ est praticien hospitalier universitaire à l'Université de Paris et à l'hôpital Necker – Enfants Malades depuis septembre 2020 (après 4 ans comme assistant hospitalier universitaire). Médecin de santé publique depuis 2015, il a obtenu un doctorat en informatique médicale de l'Université de Paris en 2020 et est membre de l'équipe de recherche INSERM « Information Sciences to support Personalized Medicine » (Pr. A. Burgun). Ses travaux portent sur la réutilisation des données de soins pour la recherche et le *machine learning*, notamment via l'utilisation du traitement automatique de la langue pour faciliter la recherche d'information et l'accès à l'information dans les dossiers patients informatisés. Un autre axe de travail porte sur les méthodes de générations d'hypothèse (fouille de données) via des études d'association à large spectre dans les dossiers patients.
  - Pierre ZWEIGENBAUM est directeur de recherche au CNRS depuis 2006 au LIMSI, maintenant Laboratoire Interdisciplinaire des Sciences du Numérique (LISN). Il a passé vingt ans à l'Assistance publique – Hôpitaux de Paris et à l'Inserm, et a été dix ans Professeur associé à l'Inalco. Il mène des recherches sur l'extraction d'information à partir de textes avec des applications dans le domaine médical. Il est auteur ou co-auteur de méthodes et d'outils pour la détection de divers types d'entités médicales, l'expansion d'abréviations, la résolution de coréférences, la détection de relations, et la normalisation d'entités, notamment pour le codage CIM-10. Il a aussi conçu des méthodes pour l'acquisition automatique de connaissances linguistiques à partir de corpus et de thésaurus, pour aider à étendre des lexiques et des



**AfIA**  
Association française  
pour l'Intelligence Artificielle

terminologies, y compris bilingues. Ses travaux lui ont valu la reconnaissance de l'American College of Medical Informatics (Fel-

low ACMI, 2014) et de l'International Academy of Health Sciences Informatics (Fellow IAHSI, 2019).



**Afia**  
Association française  
pour l'Intelligence Artificielle

---

## Thèses et HDR du trimestre

---

Si vous êtes au courant de la programmation de soutenances de thèses ou HDR en Intelligence Artificielle cette année, vous pouvez nous les signaler en écrivant à [redacteur@afia.asso.fr](mailto:redacteur@afia.asso.fr).



## ■ Thèses de Doctorat

### Iago BONNICI

« Vers un apprentissage protéiforme : accommoder les changements de signature des agents artificiels »

Supervision : *Fabien MICHEL*

Le 29/03/2021, à l'Université de Montpellier

### Silvia PAGLIARINI

« Modélisation du réseau neuronal responsable de l'apprentissage du chant chez l'oiseau chanteur »

Supervision : *Xavier HINAUT*  
*Arthur LEBLOIS*

Le 25/03/2021, à l'Université de Bordeaux

### Mustafizur Rahman SHAHID

« Deep learning for Internet of Things (IoT) network security »

Supervision : *Hervé DEBAR*  
*Grégory BLANC*

Le 22/03/2021, à l'Institut Polytechnique de Paris

### Audren Boulic BOUADJIO

« Génération multi-agents de réseaux sociaux »

Supervision : *Frédéric AMBLARD*

Le 18/03/2021, à l'Université de Toulouse 1

### Sara MEFTAH

« Apprentissage par transfert neuronal pour l'adaptation aux domaines en traitement automatique de la langue »

Supervision : *Nasredine SEMMAR*

Le 10/03/2021, à l'Université Paris-Saclay

### Tatiana BABICHEVA

« Machine learning pour la gestion distribuée et dynamique d'une flotte de taxis et navettes autonomes »

Supervision : *Leila KLOUL*

*Dominique BARTH*

Le 10/03/2021, à l'Université Paris-Saclay

### Karim ADERGHAL

« Classification of multimodal MRI images using deep learning: application to the diagnosis of Alzheimers disease »

Supervision : *Jenny Benois PINEAU*  
*Karim AFDEL*

Le 26/02/2021, à l'Université de Bordeaux

### Adrien BARDET

« Architectures neuronales multilingues pour le traitement automatique des langues naturelles »

Supervision : *Sylvain MEIGNIER*  
*Loïc BARRAULT*

Le 22/02/2021, à l'Université de Le Mans

### Daniel EI OURAoui

« Méthodes pour le raisonnement d'ordre supérieur dans SMT »

Supervision : *Stephan MERZ*

*Pascal FONTAINE*

Le 11/02/2021, à l'Université de Lorraine

### Anh Khoa Ngo HO

« Modèles d'alignement probabilistes génératifs pour les mots et sous-mots: une exploration systématique des limites et potentialités des paramétrisations neuronales »

Supervision : *Francois YVON*

Le 09/02/2021, à l'Université Paris-Saclay



**Afia**  
Association française  
pour l'Intelligence Artificielle

**Syrielle MONTARIOL**

« [Models of diachronic semantic change using word embeddings](#) »

Supervision : *Alexandre ALLAUZEN*  
*Jean Baptiste JANVIER*

Le 08/02/2021, à l'Université Paris-Saclay

**Leo HEMAMOU**

« [Analyse automatique des comportements multimodaux lors d'entretiens vidéo différés pour le recrutement](#) »

Supervision : *Jean-Claude MARTIN*

Le 04/02/2021, à l'Université Paris-Saclay

## ■ **Habilitations à Diriger les Recherches**

Nous n'avons malheureusement pas eu connaissance ce trimestre d'HDR dans le domaine de l'IA. N'hésitez pas à nous envoyer les informations concernant celles dont vous avez entendu parler à [redaction@afia.asso.fr](mailto:redaction@afia.asso.fr).



**AfIA**

Association française  
pour l'Intelligence Artificielle

---

## À PROPOS DE L'AfIA

---

L'objet de l'AfIA, Association Loi 1901 sans but lucratif, est de promouvoir et de favoriser le développement de l'Intelligence Artificielle (IA) sous ses différentes formes, de regrouper et de faire croître la communauté française en IA et, à la hauteur des forces de ses membres, d'en assurer la visibilité.

L'AfIA anime la communauté par l'organisation de grands rendez-vous. Se tient ainsi chaque été une semaine de l'IA, la « Plate-forme IA » (Pfia 2019 Toulouse, Pfia 2020 Angers, Pfia 2021 Bordeaux) au sein de laquelle se tiennent la Conférence Nationale d'Intelligence Artificielle (CNIA), les Rencontres des Jeunes Chercheurs en IA (RJCIA) et la Conférence sur les Applications Pratiques de l'IA (APIA) ainsi que des conférences thématiques invitées qui évoluent d'une année à l'autre, sans récurrence obligée.

Ainsi, Pfia 2021 héberge du 28 juin au 2 juillet 2021 à Bordeaux, outre la 24<sup>e</sup> CNIA, les 19<sup>es</sup> RJCIA et la 7<sup>e</sup> APIA : les 32<sup>es</sup> IC, les 16<sup>es</sup> JFPDA, les 29<sup>es</sup> JFSMA et les 15<sup>es</sup> JIAF, trois journées thématiques (Défense & IA, Jeux & IA, Santé & IA) et une section tutorielle accessible à tous les inscrits à la plateforme.

Forte du soutien de ses 340 adhérents à jour de leur cotisation en 2021, l'AfIA assure :

- le maintien d'un [site web](#) dédié à l'IA reproduisant également les [Brèves](#) de l'IA ;
- une *journée recherche* « Perspectives et Défis en IA » (PDIA 2021) ;
- une *journée enseignement* « Enseignement et Formation en IA » (EFIA 2021) ;
- une *journée industrielle* « Forum Industriel en IA » (FIIA 2020) ;
- la remise annuelle d'un [prix de thèse](#) en IA ;
- le soutien à plusieurs collèges, actuellement au nombre de 8, ayant leur propre activité :
  - Collège [Industriel](#) (depuis janvier 2016) ;
  - Collège [Apprentissage Artificiel](#) (depuis janvier 2020) ;
  - Collège [Interaction avec l'Humain](#) (depuis juillet 2020) ;

- Collège [Représentation et Raisonnement](#) (depuis avril 2017) ;
  - Collège [Science de l'Ingénierie des Connaissances](#) (depuis avril 2016) ;
  - Collège [Systèmes Multi-Agents et Agents Autonomes](#) (depuis octobre 2016) ;
  - Collège [Technologies du Langage Humain](#) (depuis juillet 2019) ;
  - Collège [Compétitions](#) (depuis octobre 2018).
- la parution trimestrielle des [Bulletins](#) de l'AfIA ;
  - un lien entre ses membres et sympathisants sur les réseaux sociaux [LinkedIn](#), [Facebook](#) et [Twitter](#) ;
  - le [parrainage](#) scientifique, mais aussi éventuellement financier, d'événements en IA ;
  - la diffusion mensuelle de [Brèves](#) sur les actualités de l'IA en France ([abonnement](#) ou [envoi](#) à la liste) ;
  - la réponse aux consultations officielles ou officieuses (Ministères, Missions, Organismes) ;
  - la réponse aux questions de la presse, écrite ou orale, également sur internet ;
  - la divulgation d'offres de [collaborations](#), de [formations](#), d'[emploi](#), de [thèses](#) et de [stages](#).

L'AfIA organise aussi mensuellement des [journées communes](#) avec d'autres associations. Pour 2021 : [TAL & IA](#) avec [ATALA](#) ; [IHM & IA](#) avec [AFIHM](#) ; [EGC & IA](#) avec [EGC](#).

Enfin, l'AfIA encourage la participation de ses membres aux grands événements de l'IA, dont Pfia. Ainsi, les membres de l'AfIA, pour leur inscription à Pfia, bénéficient d'une réduction équivalente à deux fois le coût de leur adhésion, leur permettant d'assister à [Pfia 2021](#) sur 5 jours au tarif de 0€ !

Rejoignez-vous vous aussi et [adhérez](#) à l'AfIA pour contribuer au développement de l'IA en France. L'adhésion peut être individuelle ou, à partir de cinq adhérents, être réalisée au titre de personne morale. Parmi les personnes morales, certaines peuvent vouloir rejoindre notre Collège [Industriel](#), au plus près de l'IA !

Merci également de susciter de telles adhésions en diffusant ce document autour de vous !



## CONSEIL D'ADMINISTRATION

Benoit LE BLANC, *président*  
Domitile LOURDEAUX, *vice-présidente*  
Isabelle SESÉ, *trésorière*  
Grégory BONNET, *secrétaire*  
Dominique LONGIN, *rédacteur*  
Emmanuel ADAM, *webmestre*

### Autres membres :

Sandra BRINGAY, Yves DEMAZEAU, Gaël DIAS,  
Catherine FARON-ZUCKER, Pierre FEILLET,  
Thomas GUYET, Marie LEFEVRE, Engelbert  
Mephu NGUIFO, Gauthier PICARD, Valérie REI-  
NER, Céline ROUVEIROL, Laurent SIMON,  
Charlotte TRUCHET

## COMITÉ DE RÉDACTION

redaction@afia.asso.fr

Emmanuel ADAM  
*Rédacteur*

Grégory BONNET  
*Rédacteur en chef adjoint*  
resp-gt-redaction@afia.asso.fr

Gaël LEJEUNE  
*Rédacteur*

Dominique LONGIN  
*Rédacteur en chef*  
resp-gt-redaction@afia.asso.fr

Laurent SIMON  
*Rédacteur*

## LABORATOIRES ET SOCIÉTÉS ADHÉRANT COMME PERSONNES MORALES

.....  
Ardans, Berger Levrault, CRIL, CRIStAL, Dassault Aviation, ENIB, EURODECISION, GRETTIA,  
GREYC, Huawei, I3S, IBM, INRIA Sophia Antipolis Méditerranée, IRIT, ISAE-SUPAERO, Lab-  
STICC, LAMSADE, LERIA, LGI2P, LHC, LIG, LIMICS, LIMSI, LIP6, LIPADE, LIRIS, LIRMM,  
LITIS, MaIAGE, Naver Labs, Renault, Thales, Université Paris-Saclay, Veolia.

## ■ Pour contacter l'Afia

### Président

Benoit LE BLANC  
École Nationale Supérieure de Cognitique  
Bordeaux-INP  
109 avenue Roul, 33400 Talence  
Tél. : +33 (0) 5 57 00 67 00  
[president@afia.asso.fr](mailto:president@afia.asso.fr)

### Serveur WEB

<http://www.afia.asso.fr>

### Adhésions, liens avec les adhérents

Isabelle SESÉ  
[tresorier@afia.asso.fr](mailto:tresorier@afia.asso.fr)

## ■ Calendrier de parution du Bulletin de l'Afia

	Hiver	Printemps	Été	Automne
Réception des contributions	15/12	15/03	15/06	15/09
Sortie	31/01	30/04	31/07	31/10