

Projet doctoral :
Développement de méthodes pour la création de modèles d'IA performants et robustes en imagerie médicale

Contact: irene.buvat@universite-paris-saclay.fr

Sujet

Contexte

Le développement de méthodes relevant de l'Intelligence Artificielle (IA) pour faciliter l'interprétation des images médicales et mieux exploiter leur contenu est en pleine expansion. Ces développements se heurtent cependant à des obstacles récurrents. Par exemple, la quantité d'images médicales composant les bases de données d'apprentissage est bien souvent limitée, notamment en comparaison aux volumes d'images « naturelles » accessibles pour entraîner des algorithmes de reconnaissance d'objets, d'animaux, de plantes, etc. Ce volume limité d'images disponibles pour l'apprentissage est d'autant plus fréquent dans le contexte de la médecine de précision où les maladies sont distinguées par un nombre croissant de critères, qui rend difficile la collecte de bases d'images homogènes vis-à-vis de l'ensemble de ces critères. Au-delà de la quantité d'images composant les bases d'apprentissage, la plupart des méthodes d'IA développées en imagerie médicale relèvent actuellement de l'apprentissage supervisé. Elles nécessitent un travail laborieux d'annotation des images par des experts. Les annotations peuvent s'avérer de qualité variable, bruitées et biaisées. Ces limites sont omniprésentes et appellent de nouvelles solutions.

L'hypothèse du travail de thèse est que les méthodes d'IA développées pour des applications différentes de celles de l'imagerie médicale pourraient être revisitées et adaptées au contexte spécifique des images médicales, dans le but de dépasser les performances actuellement atteintes au moyen d'algorithmes développés à partir de bases de données limitées en quantité et qualité.

Objectifs

L'objectif du travail de thèse est d'exploiter les avancées récentes du domaine de l'IA, notamment concernant l'apprentissage faiblement supervisé et non-supervisé ou l'apprentissage par transfert, pour résoudre les obstacles auxquels se heurtent systématiquement le développement d'applications d'IA en imagerie médicale, à savoir le faible volume de données imparfaitement annotées. Il s'agira également de démontrer la pertinence des méthodes développées sur des cas d'usage relevant de l'imagerie TEP/TDM et IRM.

Méthodes

Plusieurs bases de données d'images TEP/TDM (lymphomes, cancer du poumon et cancer du sein) et IRM (cancer du sein, tumeurs cérébrales) annotées sont disponibles au sein du laboratoire, auxquelles sont associées des questions d'intérêt médical. Ces bases, contenant chacune les données de 100 à 500 patients, ont été exploitées pour construire des modèles d'IA par apprentissage supervisé dont les performances ont été précisément caractérisées (1-4). Le travail de thèse s'appuiera sur cet existant pour apporter des solutions aux questions suivantes :

- Peut-on enrichir artificiellement ces bases d'apprentissage pour construire des modèles plus performants et plus robustes ? Pour cela, des méthodes originales d'augmentation de données ou de génération d'images synthétiques pourront par exemple être développées.
- Peut-on caractériser la qualité des annotations et mieux exploiter les annotations imparfaites pour améliorer les performances des modèles construits ? A cet effet, certaines bases de données ont été partiellement annotées par plusieurs experts, ce qui permet d'évaluer la variabilité de l'annotation et d'étudier son impact.
- Peut-on présélectionner les images à annoter de façon préférentielle pour assurer une variabilité suffisante dans les données annotées permettant de construire des modèles

robustes ? Pour faciliter cette étude, nous disposons de plusieurs bases de données indépendantes pour lesquelles la même question est posée.

- Comment mieux tirer parti de données supplémentaires (nouveaux examens acquis chez le même patient, patients supplémentaires) pour améliorer les performances et la robustesse d'un modèle existant ? Pour plusieurs bases de données, nous disposons notamment d'images de suivi longitudinal de patients (imageries réalisées avant puis au cours du traitement).
- Pouvons-nous prédire si un modèle peut être appliqué avec succès à une nouvelle cohorte de patients ? Nous disposons de plusieurs ensembles de données pour tester la capacité du modèle à se généraliser à différents échantillons.

Résultats attendus

Le travail de thèse devrait conduire à de nouvelles méthodes pour faciliter la conception de modèles d'IA performants et robustes en imagerie médicale, et à la démonstration de la pertinence de ces méthodes sur des cas d'usage très concrets, répondant à des problématiques cliniques de grande actualité dans le contexte de la médecine de précision.

Références

1. Saint Martin MJ, Frouin F, Malhaire C, Orlhac F. Decrypting the information captured by MRI-radiomic features in predicting the response to neoadjuvant chemotherapy in breast cancer. 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 3227-3230, 2022. DOI: 10.1109/EMBC48229.2022.9871844.
2. Rahimpour M, Saint Martin MJ, Frouin F, Akl P, Orlhac F, Koole M, Malhaire C. Visual ensemble selection of deep convolutional neural networks for 3D segmentation of breast tumors on dynamic contrast enhanced MRI. Eur Radiol, in press, 2022. DOI: 10.1007/s00330-022-09113-7
3. Girum KB, Rebaud L, Cottureau AS, Meignan M, Clerc J, Vercellino L, Casasnovas O, Morschhauser F, Thieblemont C, Buvat I. 18F-FDG PET maximum intensity projections and artificial intelligence: a win-win combination to easily measure prognostic biomarkers in DLBCL patients. J Nucl Med. 63:1925-1932, 2022. DOI: 10.2967/jnumed.121.263501
4. Wallis D, Soussan M, Lacroix M, Akl P, Dubouchier C, Buvat I. An FDG-PET/CT deep learning method for fully automated detection of pathological mediastinal lymph nodes in lung cancer patients. Eur J Nucl Med Mol Imaging. 49:881-888, 2022. DOI: 10.1007/s00259-021-05513-x
5. Bradshaw TJ, Boellaard R, Dutta J, Jha AK, Jacobs P, Li Q, Liu C, Sitek A, Saboury B, Scott PJH, Slomka PJ, Sunderland JJ, Wahl RL, Yousefirizi F, Zuehlsdorff S, Rahmim A, Buvat I. Nuclear Medicine and Artificial Intelligence: Best Practices for Algorithm Development. J Nucl Med. 63:500-510, 2022. DOI: <https://doi.org/10.2967/jnumed.121.262567>

Modalités d'encadrement et d'accueil dans le laboratoire

Encadrement

L'encadrement sera assuré par Irène Buvat (DR CNRS) et Frédérique Frouin (CRHC Inserm). Le projet se déroulera dans le contexte du programme de recherche AI.DReAM, piloté par General Electric et financé par BPI France. Le doctorant participera aux réunions d'équipe toutes les semaines pour suivre l'avancée des travaux réalisés dans l'unité et devra régulièrement présenter ses résultats à cette occasion. Il participera aux réunions bimestrielles avec General Electric. Il sera également suivi par un tuteur, membre permanent interne au laboratoire mais non impliqué dans le projet doctoral pour prévenir d'éventuelles difficultés pouvant survenir pendant la thèse.

Ouverture nationale et internationale

Depuis 10 ans, l'unité s'investit dans le domaine de l'intelligence artificielle pour l'analyse des images médicales et collabore avec d'autres équipes, à l'Institut Curie, à l'École des Mines, à Neurospin, expertes en bioinformatique et en intelligence artificielle. Elle entretient également des collaborations

étroites avec les équipes de l'Université Médicale de Vienne (Autriche), de l'Université d'Amsterdam (Pays-Bas) et des laboratoires internationaux participant à la Task Force Intelligence Artificielle de la Society of Nuclear Medicine and Molecular Imaging (USA).

Objectifs de valorisation des travaux de recherche du doctorant

Le doctorant aura à présenter ses travaux via des communications dans des congrès nationaux et internationaux dans le domaine de l'imagerie et de la cancérologie. Il devra également publier ses résultats dans des revues internationales.

Modalités de financement de la thèse

Le doctorat est financé via le contrat AI.DReAM, soutenu par BPI France (2021-2026), et géré par le Centre de Recherche de l'Institut Curie. Le doctorant sera ainsi recruté par le Centre de Recherche de l'Institut Curie. Le salaire est fonction du niveau de formation et de l'expérience du candidat.

Le doctorant aura accès aux ressources informatiques de l'unité moyennant le respect de la charte informatique. Il aura également accès à des données d'imagerie et des données cliniques anonymisées respectant le règlement RGPD grâce à la collaboration établie entre le LITO et l'Institut Curie. Outre les formations proposées par l'Ecole Doctorale, le doctorant participera à des conférences (une par an) sous réserve d'acceptation d'une communication en premier auteur. Ces participations seront financées par le contrat AI.DReAM. Le doctorant se conformera au règlement intérieur de l'unité. Il ne sera pas exposé à des risques spécifiques (chimiques, biologiques ou radiologiques).

Le doctorat se déroulera dans le cadre de l'Ecole Doctorale EOBE (Electrical, Optical, Bio-physics and Engineering) de l'Université Paris Saclay. Le sujet de thèse y est recensé sur :

https://www.adum.fr/as/ed/voirproposition.pl?site=PSaclay&matricule_prop=45977

Lieu de travail

La thèse se déroulera dans le Laboratoire d'Imagerie Translationnelle en Oncologie (lito-web.fr) actuellement situé sur dans les locaux du Centre de Recherche de l'Institut Curie, Bâtiment 101B, sur le Campus Universitaire de Paris Saclay, à 3 min à pied de la station de RER B Orsay-Ville.

Le laboratoire a vocation à déménager en 2025, pour rejoindre le Centre de Recherche de l'Institut Curie à Saint-Cloud, à 2 min à pied de la gare de Saint Cloud (Transiliens L et U), Tramway T2, Métro ligne 10 Boulogne – Porte de Saint Cloud.

Le télétravail est autorisé à raison de 2 jours maximum par semaine.