

Offre de post-doctorat/ingénieur de recherche :
Génération d'une représentation latente des images TEP par IA
pour la classification des cancers - ARAMIS
U1288 Inserm, Institut Curie www.lito-web.fr

Contexte

Supposant que le contenu des images médicales est sous-exploité par l'analyse visuelle complétée parfois par la mesure de quelques grandeurs sur les images, la Radiomique [Gillies-2016] propose d'extraire automatiquement un grand nombre d'indices d'imagerie quantitatifs. L'approche la plus couramment utilisée consiste à calculer des indices issus d'équations mathématiques connues, reflétant par exemple la distribution des niveaux de gris ou la morphologie des lésions, mais qui ne sont pas spécialement conçus pour refléter des caractéristiques biologiques. Une seconde approche consiste à estimer ces indices grâce à des techniques d'apprentissage profond ("deep learning", DL) [Artesiani-2024]. Ces techniques ont l'avantage de ne pas nécessiter de segmenter les lésions avant l'extraction des biomarqueurs, étape chronophage et opérateur-dépendante, et favorise une analyse plus exhaustive du contenu de l'image. En effet, l'approche par DL permet d'explorer non seulement les lésions, mais aussi le tissu environnant ainsi que d'autres organes ou tissus. De plus, le DL peut identifier des "motifs" complexes, tels que l'hétérogénéité intra et inter-tumorale ou la dispersion des lésions, qui peuvent être mal ou insuffisamment décrits par l'expert humain et donc difficilement quantifiables à l'aide d'indices issus d'équations mathématiques établies.

Certaines approches de DL non supervisées montrent des résultats prometteurs en bioinformatique ou en imagerie histologique [Unger-2024] pour modéliser des processus biologiques complexes ou stratifier les lésions cancéreuses. En particulier, les auto-encodeurs variationnels (VAE) peuvent révéler des informations complexes contenues dans les données analysées via l'espace latent. Ainsi, en utilisant la représentation de l'espace latent d'un VAE construit à partir de données de séquençage ARN collectées pour 39 types de cancer et 55 tissus normaux, [Vibert-2021] ont élaboré un classifieur d'identification du site primitif pour des cancers de primitif inconnu (CPI). Le classifieur avait une précision de 96% pour prédire le site primitif sur une cohorte de test de 48 patients.

Objectifs

L'objectif du projet est développer un algorithme d'Intelligence Artificielle capable de classer automatiquement les images de tomographie par émission de positons (TEP scans) en oncologie. En analysant les images sans a priori, l'algorithme apprendra à extraire des informations concernant aussi bien les lésions tumorales, par exemple relatives à leur localisation, à leur activité métabolique, leur nombre ou leur volume, mais également des informations issues des organes/tissus non-tumoraux telles que le métabolisme de la rate par exemple, qui peut refléter la capacité du système immunitaire à influencer l'efficacité des traitements. D'autres informations, totalement inédites et jamais rapportées dans la littérature, pourraient être identifiées par l'algorithme. Cette potentielle découverte d'éléments phénotypiques actuellement inconnus et informant sur la maladie et son évolution est précisément la principale motivation du projet.

Méthodes

Le projet exploitera une base de données rétrospective comprenant plus de 4000 TEP scans acquis avant traitement dans le parcours de soin du patient. Les images sont classées en 7 catégories distinctes : cancer pulmonaire, cancer du sein, cancer ORL, lymphome, mélanome, sarcome des tissus mous ou absence de cancer diagnostiqué. Cette base de données devrait permettre à un algorithme d'apprentissage non supervisé d'identifier un espace de représentation (espace latent) faisant ressortir la variété des TEP scans, tout en structurant les informations communes ou différenciantes

des examens entre eux. Cet espace sera scruté afin de comprendre les éléments exploités par l'algorithme pour caractériser les images. Il s'agira ensuite d'exploiter cet espace latent pour concevoir un classifieur capable d'identifier automatiquement l'organe d'origine du cancer.

Dans un premier temps, le projet se concentrera sur l'identification des cancers dont le site primitif est connu, démontrant ainsi la pertinence de l'espace latent pour résoudre des tâches de classification relativement simples. Le classifieur développé sera soumis à une évaluation approfondie, grâce à des cohortes de validation interne et externe (>700 patients), à partir d'images totalement indépendantes qui n'auront pas contribué à la conception initiale du classifieur. Une fois cette phase achevée, nous explorerons les possibilités offertes par l'espace latent pour résoudre d'autres tâches complexes, telles que la prédiction de la réponse au traitement et/ou de la survie des patients.

Grâce à l'environnement de recherche propice au sein de l'Institut Curie et des collaborations avec les oncologues, il sera envisageable de généraliser notre approche en enrichissant le modèle avec de nouveaux types de cancers. Nous pourrions alors tester le modèle résultant sur des images TEP scans provenant de patients présentant un CPI. Ces cancers représentent un défi important pour les oncologues, car le choix du traitement dépend souvent de l'organe d'origine du cancer. La localisation des patients dans cet espace latent pourrait également mettre en évidence des liens avec le pronostic, et donc permettre une meilleure sélection des traitements, pour un bénéfice significatif en termes de survie et de qualité de vie pour les patients atteints de cancer.

Environnement scientifique

Le Laboratoire d'Imagerie Translationnelle en Oncologie (LITO) est spécialisé dans la production et l'analyse d'images médicales [Nioche-2018], en particulier en imagerie métabolique et fonctionnelle (TEP et IRM). Largement multidisciplinaire, il est constitué d'une quarantaine de physiciens, ingénieurs, radiologues, médecins nucléaires, biologistes, oncologues, doctorants et post-doctorants. Implanté au Centre de Recherche de l'Institut Curie et travaillant étroitement avec l'Etablissement Hospitalier de l'Institut Curie, il bénéficie d'un environnement particulièrement propice aux projets concernant l'analyse avancée des images médicales par les techniques les plus modernes de radiomique et d'intelligence artificielle.

Conditions financières

Le financement sera assuré par le contrat ARAMIS – Cancéropôle Ile-de-France, l'employeur sera le Centre de Recherche de l'Institut Curie, pour un contrat CDD de 12 mois. Le montant du salaire dépendra de l'expérience du candidat. Ce contrat pourra être prolongé.

Objectifs de valorisation des travaux de recherche : diffusion, publication et confidentialité, droit à la propriété intellectuelle,...

Le/la post-doctorant(e) aura la possibilité de présenter ses travaux via des communications dans des congrès nationaux et internationaux dans le domaine de l'imagerie médicale (EANM, SNMMI), de la cancérologie (ASCO, ESMO), et de la science des données (eg, MICCAI, ICML). Il ou elle publiera ses résultats dans des revues internationales.

Compétences requises

- Apprentissage profond
- Programmation
- Statistiques, méthodes d'apprentissage
- Imagerie médicale, radiomique, analyse d'images

Contact

Fanny Orhac : fanny.orhac@curie.fr

Références bibliographiques

[Artesani-2024] Empowering PET: harnessing deep learning for improved clinical insight. Artesani et al. *Eur Radiol Exp*. 2024;8:17. <https://doi.org/10.1186/s41747-023-00413-1>

[Gillies-2016] Radiomics: images are more than pictures, they are data. Gillies et al. *Radiology*. 2016;278:563-577. <https://doi.org/10.1148/radiol.2015151169>

[Nioche-2018] LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity . Nioche et al. *Cancer Res*. 2018;78:4786-4789. <https://doi.org/10.1158/0008-5472.CAN-18-0125>

[Unger-2024] A systematic analysis of deep learning in genomics and histopathology for precision oncology. Unger et al. *BMC Med Genomics*. 2024;17:48. <https://doi.org/10.1186/s12920-024-01796-9>

[Vibert-2021] Identification of tissue of origin and guided therapeutic applications in cancers of unknown primary using deep learning and RNA sequencing (TransCUPtomics). Vibert et al. *J Mol Diagn*. 2021;23:1380-1392. <https://doi.org/10.1016/j.jmoldx.2021.07.009>