

AI for Health 2020

Le challenge : développer et valider un algorithme permettant de classifier les patients atteints d'un cancer broncho-pulmonaire à partir de données multi-modales (cliniques, radiologiques, anatomopathologiques et génétiques) pour orienter les décisions thérapeutiques.

L'Institut Curie associe un Ensemble hospitalier et un Centre recherche. Les équipes de la Direction des data et de l'Institut du Thorax Curie - Montsouris (ITCM) proposent d'utiliser des données cliniques, de biologie et d'imagerie pour alimenter des algorithmes utilisant les techniques d'intelligence artificielle (IA). Le but est de permettre à une start-up d'innover pour trouver des solutions utilisables pour améliorer les diagnostics, comprendre et prédire la réponse aux traitements et formuler de nouvelles hypothèses biologiques ouvrant des perspectives de recherche.

Les cancers broncho-pulmonaires sont la première cause de mortalité par cancer en France avec plus de 46 000 cas en 2018 (10 % de tous les nouveaux cas de cancers) et près de 35 000 décès (1^{ère} cause de décès par cancer en France). **Près de 70 % des cancers broncho-pulmonaires sont diagnostiqués à un stade avancé** ou métastatique ce qui représente un défi majeur pour les oncologues. En 2019, l'Institut Curie a pris en charge plus de 1000 patients atteints d'un cancer du poumon.

Les cancers broncho-pulmonaires ne sont pas une entité unique. Ils n'ont de commun que la localisation anatomique alors que la présentation clinique, l'évolution, le pronostic peuvent très différents. Pour le traitement de ces maladies, on peut utiliser la chirurgie, la radiothérapie, la chimiothérapie, l'immunothérapie ou des thérapies ciblées qui peuvent s'associer entre elles dans différentes séquences. La cancérogénèse est un mécanisme complexe et de ce fait les cancers du poumon ne sont pas une entité unique mais « une multitude de maladies orphelines ».

Les classifications actuelles reposent sur des critères cliniques, de diffusion de la maladie, sur des images radiologiques (Scanners X, TEP, IRM), des images d'anatomopathologie, des résultats du séquençage de l'ADN tumoral. Ces différents critères, associés entre eux, devraient permettre de mieux classer les patients en cas jumeaux, voire en jumeaux numériques afin de mieux comprendre la maladie, en déterminer le pronostic et décider le meilleur traitement. L'utilisation de toutes ces données (**intégration multi-modale**) associées aux techniques faisant appel à l'apprentissage machine pourraient permettre de mieux comprendre chaque tumeur, de mieux les classer, les regrouper, pour mieux en prédire l'évolution, le pronostic et décider le meilleur traitement adapté à chaque cas : ce qu'on appelle la **médecine de précision**.

C'est un projet qui associe des médecins et des data scientist, un hôpital spécialisé dans le traitement du cancer et une startup, coopération indispensable pour relever ce défi. L'Institut Curie dispose de données numériques complètes car la numérisation des dossiers et des images remonte au début des années 2000.

Le cancer du poumon, malgré les progrès liés aux thérapies ciblées et à l'immunothérapie, reste encore un des cancers parmi les plus graves. L'utilisation des techniques d'intelligence artificielle, intégrant un maximum de données disponibles, doit nous permettre de progresser pour mieux caractériser de nouveaux biomarqueurs, mieux prédire l'évolution pour mieux comprendre la maladie : mieux classer pour mieux prédire, c'est là le grand défi où l'IA peut apporter des solutions.

La création de l'ITCM est l'illustration que l'Institut Curie a inscrit le cancer du poumon comme une priorité dans son programme médico-scientifique. Plusieurs équipes au sein de l'ITCM, du Centre de recherche de l'Institut, de la Direction des données coopèrent ou s'associent à des startup pour accélérer la recherche dans ce domaine. Plusieurs programmes sont en cours pour mieux comprendre l'action des chimiothérapies, des thérapies ciblées, de l'immunothérapie. Des bases de données sont constituées pour permettre des coopérations entre structures académiques. La coopération avec des startup existe depuis plusieurs années, comme par exemple la nécessité d'assurer l'interopérabilité, le partage et l'échange de données, l'apprentissage fédéré, l'analyse d'images ou le traitement automatique du langage. Ces coopérations sont souvent nées d'un projet précis mais peuvent se perpétuer dans un cadre contractuel plus large.

L'Institut Curie dispose de données médicales utilisées dans le cadre du soin. Les données utilisables proviennent des textes des comptes rendus médicaux, de données structurées, de lames virtuelles d'anatomopathologie, d'images radiologiques numérisées, de l'ensemble des traitements réalisés et de données évolutives sur le long terme.

Pour que ce projet d'analyse multi-modale puisse se réaliser, une partie du budget obtenu par le lauréat sera consacré à la curation, le contrôle qualité, la mise en conformité au RGPD dans un entrepôt sécurisé pour permettre au lauréat de réaliser ses travaux, en coopération avec les équipes médicales et scientifiques de l'Institut Curie.