

AVIS DE SOUTENANCE DE THÈSE

DOCTORAT (Arrêté du 26 août 2022 modifiant l'arrêté du 25 mai 2016)

Madame Chiara CORDIER

candidate au diplôme de Doctorat de l'Université d'Angers, est autorisée à soutenir publiquement sa thèse

le 28/04/2025 à 14h00

**Institut de Cancérologie de l'Ouest
Amphithéâtre Noëlla Rouget
15, rue André Boquel
49055 ANGERS**

sur le sujet suivant :

Développement de modèles d'apprentissage pour le traitement de données omiques dans les cancers du sein

Directeur de thèse : **Monsieur Fabien PANLOUP**

Composition du jury :

Madame Agnès BASSEVILLE, Docteure Institut de Cancérologie de l'Ouest, Co-encadrante
Monsieur Julien CHIQUET, Directeur de Recherche INRAE Université Paris-Saclay/AgroParisTech/INRAE, Rapporteur
Monsieur Bertrand MICHEL, Professeur des Universités Ecole Centrale de Nantes, Examineur
Monsieur Raphaël MOURAD, Maître de Conférences HDR Université de Toulouse, Rapporteur
Monsieur Fabien PANLOUP, Professeur des Universités Université d'Angers, Directeur de thèse
Madame Nelle VAROQUAUX, Chargée de recherche CNRS Université Grenoble Alpes, Examinatrice

Résumé de la thèse

Le cancer du sein est le cancer le plus mortel chez la femme, principalement en raison de la résistance aux traitements. Une des stratégies pour limiter les récurrences est la médecine personnalisée. En effet, chaque tumeur étant unique, il est indispensable d'attribuer à chaque patient le traitement le plus adapté aux caractéristiques de sa tumeur. Dans ce but, il est possible de développer des modèles de prédiction de réponse au traitement en se basant sur des données dites omiques provenant de la tumeur. Cependant, le traitement de ce type de données, complexes et souffrant du fléau de la dimension, nécessite le développement de nouvelles approches plus adaptées. Dans cette thèse, trois stratégies ont été mises en place. Tout d'abord, un classifieur binaire a été construit à partir de la formulation duale de la distance de Wasserstein entre les lois des données de chaque classe, ainsi que des versions régularisées de se classer et un homologue en apprentissage profond. Ces différentes méthodes se sont avérées adaptées au traitement des données testées, particulièrement lorsque l'information prédictive était répartie sur un grand nombre de variables. La seconde stratégie a reposé sur le développement d'une nouvelle méthode de réduction de dimension. Une adaptation "labellisée" a été apportée à un autoencodeur variationnel afin de prendre en compte la classe dans l'encodage. Pour ce faire, la fonction objectif basée, entre autres, sur la divergence de Kullback-Leibler entre la loi en sortie de l'encodeur et un a priori est modifiée pour considérer un a priori différent pour chaque classe. Dans le cadre d'une analyse orientée sur la classe, cette approche offre des solutions compétitives de réduction de dimension et de génération de données synthétiques. Pour terminer, la troisième et dernière stratégie s'appuie sur l'utilisation d'un jeu de données plus grand et diversifié. Pour cela, des données omiques de tous types de cancer ont été collectées et regroupées dans une même base de données. Ensuite, un algorithme multi-tâche a été mis en place pour la classification de plusieurs éléments en simultané. L'objectif repose sur l'obtention d'un apport de connaissances complémentaires à partir de tâches similaires mais différentes – sur le principe du transfert de connaissances – et ainsi permettre de meilleures performances qu'un classifieur mono-tâche. Le développement de ses trois axes a permis la mise en place de modèles prédictifs plus performants et servira à terme à découvrir de nouveaux biomarqueurs qui serviront soit de caractéristiques aidant à la stratification des patients pour sélectionner leur traitement, soit à identifier de potentielles cibles pour l'élaboration de nouveaux traitements.