

## DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **06 décembre 2023**

Nom de famille et prénom de l'auteur : **Monsieur RANVIER Thomas**

Titre de la thèse : « *Apprentissage profond à partir de données incomplètes et multi-sources : application à l'immunothérapie contre le cancer* »



### Résumé

Dans ce travail de thèse, nous nous intéressons à l'apprentissage automatique, et plus particulièrement au domaine de l'apprentissage profond, à partir de données incomplètes et multi-sources. Nous visons à proposer des approches innovantes pour le traitement des données incomplètes en apprentissage automatique, et plus spécifiquement lors de l'entraînement de réseaux de neurones. Nous nous concentrons également sur l'optimisation des performances d'apprentissage sur des données multi-sources, en concevant une nouvelle approche avancée d'adaptation de domaines supervisée. L'application de notre travail fait partie du projet de recherche européen QUALITOP, visant à améliorer la qualité de vie de patients souffrant de cancer et suivant un traitement par immunothérapie. Nous visons à concevoir des modèles prédictifs, qui seront intégrés au sein d'une plateforme numérique intelligente ouverte, afin d'offrir des recommandations en temps réel à des experts médicaux. Nos modèles prédictifs fourniront une aide précieuse aux experts, les aidant à personnaliser et à optimiser les stratégies de traitement pour chaque patient, ainsi qu'à identifier les facteurs influençant leur état de santé. En fin de compte, ces modèles prédictifs peuvent conduire à la conception et à la prescription de traitements par immunothérapie plus efficaces et plus sûrs, bénéficiant ainsi aux patients ainsi qu'aux médecins.

Dans ce travail, nous nous intéressons à maximiser la performance prédictive dans le contexte spécifique des données incomplètes et multi-sources. Pour atteindre cet objectif, nous avons orienté notre attention vers la résolution de problèmes de recherche critiques et pertinents pour notre contexte d'apprentissage et pour le projet QUALITOP. Nous introduisons d'abord une nouvelle méthode innovante de correction de bruit d'attributs, data Denoising and Imputation in One Step (DIOS). DIOS est la première approche dans la littérature de l'Apprentissage Automatique capable d'imputer les valeurs manquantes et de corriger les valeurs erronées dans un jeu de données tabulaires en une unique étape de prétraitement. Nous nous intéressons ensuite à maximiser la performance d'apprentissage des réseaux de neurones lorsqu'ils sont entraînés sur des données complétées. Dans ce contexte, nous proposons deux frameworks pouvant être utilisés pour tenir compte de l'incertitude d'imputation lors de l'entraînement des réseaux de neurones, conduisant à de meilleures prédictions, Single-Hotpatching (S-HOT) et Multiple-Hotpatching (M-HOT). Ce travail est une première étape vers la recherche de meilleures manières de traiter l'imputation de valeurs manquantes pour l'entraînement de modèles prédictifs, dans l'espoir de susciter l'intérêt d'autres chercheurs autour du monde sur cette question. Par la suite, nous proposons une nouvelle méthode avancée d'adaptation de domaine, Weighted Multi-Source Supervised Domain Adaptation (WMSSDA). WMSSDA est capable d'extraire des connaissances précieuses de plusieurs jeux de données labellisés, afin d'améliorer la performance d'apprentissage sur un jeu de données similaire. Notre approche proposée inclut un nouveau composant aidant à limiter le transfert négatif grâce à une pondération adaptative de l'impact de chaque source sur l'entraînement du modèle. Enfin, nous présentons l'application de

notre travail complet, unifié au sein d'un pipeline prédictif, que nous appliquons avec grand succès dans un scénario médical réel, soulignant la pertinence de notre travail pour la prédiction médicale.

**Mots Clés :**

- Bruit d'Attributs
- Incertitude d'Imputation
- Adaptation de Domaine
- Apprentissage Automatique
- Apprentissage Profond
- Prédiction Médicale