

Prédiction de la progression du myélome multiple par imagerie TEP : adaptation des forêts de survie aléatoires et de réseaux de neurones convolutionnels.

Contexte :

L'objectif de ces travaux est la conception de modèles par apprentissage statistique permettant la prédiction de la survie et l'identification de biomarqueurs dans le contexte du myélome multiple, à l'aide de l'imagerie TEP (Tomographie à émission de positons) et de données cliniques. Cette thèse est divisée en deux parties : la première décrit un modèle basé sur les forêts de survie aléatoires (RSF). La seconde détaille l'adaptation de réseaux de neurones convolutionnels à la survie.

Méthodes :

Les RSF, bien que devenus la méthode de référence dans l'analyse de survie, n'ont jamais été étudiés dans le contexte du myélome multiple et des images TEP. Une première méthode basée sur les RSF est donc conçue, englobant la prédiction de la progression et la sélection de variables. Elle comporte trois étapes : une étape d'optimisation automatique des paramètres des arbres, une étape de prédiction de l'ordonnancement des variables les plus prédictives par VIMP (Variable Importance), et une étape de prédiction par RSF. Ce modèle fournit la survie des patients, une classification en deux groupes de risque et une liste de biomarqueurs par ordre d'importance. En entrée du modèle sont utilisées les valeurs radiomiques et volumiques des images TEP sur la lésion focale la plus fixante et les données cliniques.

La deuxième méthode proposée, appelée M2P2, est un réseau de neurones convolutionnels 3D qui apprend à extraire les valeurs radiomiques les plus pertinentes, à l'aide d'un module d'attention sur les filtres. Pour entraîner le modèle à prédire des risques nous faisons une étude de fonctions de coût adaptées à la survie, en proposant deux nouvelles fonctions basées sur le contraste de risques entre les éléments d'un triplet. Ces nouvelles fonctions sont aussi efficaces pour apprendre un espace de caractéristiques discriminant pendant une étape de pré-entraînement du réseau. Après entraînement, M2P2 permet lui aussi de prédire la survie et un groupe de risque.

Deux bases de données prospectives multi-centriques, l'une française (IMAJEM) et l'autre italienne (EMN02/HO95), sont utilisées pour entraîner et valider nos modèles.

Résultats :

Nous avons montré que notre modèle basé sur les RSF est plus performant que les modèles classiques (erreur de prédiction de 0,36 au lieu de 0,56 pour Lasso-Cox et 0,48 pour Gradient-Boosting Cox). Nous avons aussi montré la prédictibilité et la stabilité de la sélection par VIMP comparée aux autres méthodes de sélection (0,47 et 0,43 pour "Minimal Depth" et "Variable-Hunting"). Parmi les modèles testés, notre modèle est le seul à permettre une séparation significative des patients en groupes de prédiction (p -value inférieure à 0,05). Finalement, nous avons montré l'intérêt d'utilisation des radiomiques en combinaison des données cliniques comme biomarqueurs d'intérêt dans un contexte pronostique du myélome multiple.

Concernant l'apprentissage profond, nous avons créé un modèle prenant en compte les problématiques de petites lésions de tailles variables, du nombre limité de patients et de la survie. Nous avons montré l'intérêt de l'utilisation du modèle CNN (c-index de 0,57) et plus particulièrement le CNN 3D (c-index de 0,60), comparé aux méthodes de prédiction de survie classiques (c-index de 0,51). Nous avons aussi prouvé l'intérêt de l'attention sur les filtres (c-index de 0,63 au lieu de 0,61) et des méthodes de pré-entraînement par classification binaire puis contrastif (c-index de 0,66 au lieu de 0,61), et testé d'autres comme l'attention spatiale et la SPP (Spatial Pyramid Pooling).

Contributions :

Les contributions principales de la thèse sont les suivantes : * Conception d'un modèle basé sur les RSF et les images TEP permettant la prédiction d'un groupe de risque pour les patients atteints de myélome multiple. * Détermination de biomarqueurs grâce à ce modèle. * Démonstration de l'intérêt des radiomiques TEP. * Production d'un modèle CNN 3D, appelé M2P2, permettant la prédiction de la survie et de groupes de risques des patients atteints de myélome multiple. * Extension de l'état de l'art des méthodes d'adaptation de l'apprentissage profond à un nombre limité de patients et à de petites images. * Étude des fonctions de coût utilisées en survie, et proposition de nouvelles.

De plus nous sommes, à notre connaissance, les premiers à avoir étudié l'utilisation des RSF dans le contexte du myélome multiple et des images TEP, à étudier la survie du myélome multiple grâce à des CNN, à utiliser du pré-entraînement auto-supervisé et contrastif avec des images TEP et avec une tâche de survie, à adapter la fonction de coût triplet à la survie.

Mots-clés : Myélome multiple, Réseaux de neurones convolutionnels, Analyse de survie, Forêts de survie aléatoires, Tomographie à émission de positons