

Projet RUBY : Application d'algorithmes d'IA pour la structuration des données des comptes rendus médicaux de patientes atteintes d'un cancer du sein

R Schiappa¹, G Uzbelger², B Thamphya¹, A Fabre², S Toledano², J Gal¹, O Schneider³, C Garcia², S Benavides², J Godfroy³, J Haudebourg⁴, C Bailleux⁵, JM Ferrero⁵, O Humbert⁶, E Barranger⁵, E Chamorey¹

¹ Département d'Epidémiologie, de Biostatistiques et des Données de Santé, Centre Antoine Lacassagne, Nice

² Artificial Intelligence – Advanced Analytics Solutions – IBM France

³ Direction des Services Informatiques, Centre Antoine Lacassagne, Nice

⁴ Département d'Anatomie et Cytologie Pathologiques, Centre Antoine Lacassagne, Nice

⁵ Département d'Oncologie Sénologie, Centre Antoine Lacassagne, Nice

⁶ Département de Médecine Nucléaire, Centre Antoine Lacassagne, Nice

CONTEXTE

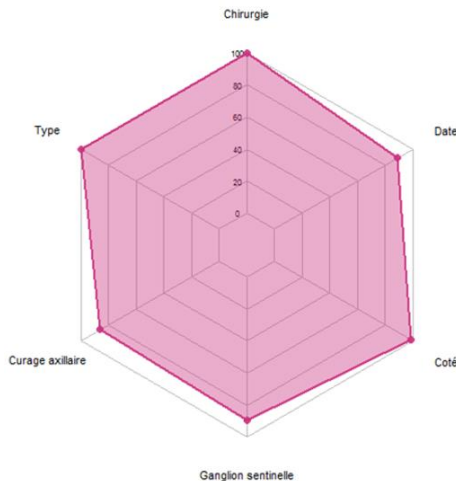
La structuration des données de santé dans les centres hospitaliers est une étape indispensable pour leur exploitation, cependant elle demeure extrêmement chronophage. Avec plus de 3 millions de comptes rendus et environ 200.000 nouveaux comptes rendus par an au Centre Antoine Lacassagne (CAL), la structuration manuelle rétrospective n'est pas une solution envisageable car elle mobiliserait beaucoup trop de ressources humaines. RUBY est un programme informatique développée au CAL en collaboration avec IBM qui permet la structuration automatique des données à l'aide d'algorithmes d'IA. L'objectif de ce travail est de tester rétrospectivement RUBY sur les patientes atteintes de cancer du sein au CAL.

OBJECTIFS

Le programme RUBY visait à développer des algorithmes d'IA (réseaux de neurones convolutifs) capables de structurer automatiquement les CR des patientes atteintes d'un cancer du sein. La première étape de ce projet était de développer des algorithmes d'IA sur une cohorte d'apprentissage puis de valider ces algorithmes sur une cohorte de test. La seconde étape, que nous présentons ici, est d'évaluer la précision et l'utilité de cet outil en comparant les résultats de la structuration des données par RUBY à une structuration manuelle.

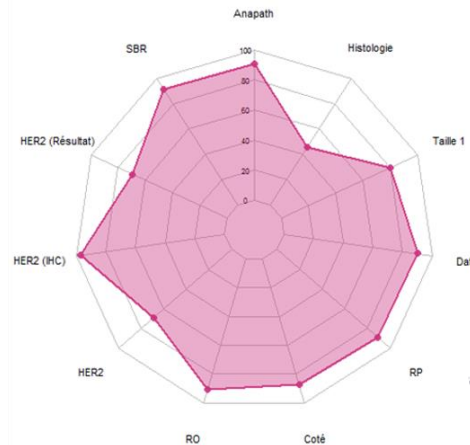
METHODES

RUBY a été appliqué aux 227 patientes de l'étude clinique promue au CAL : IODINE BREAST (IB). Les données d'IB ont été saisies par des ARC puis ont été monitorées afin de s'assurer de la qualité des données. La précision de RUBY a été évaluée en comparant les données saisies manuellement avec celles obtenues avec l'algorithme. RUBY a ensuite été appliqué sur la cohorte CHIRANA (extraite du PMSI) contenant 5702 patientes opérées d'un cancer du sein au CAL. L'objectif était de montrer la faisabilité et la rapidité de RUBY sur une base de données volumineuse.



RESULTATS

Au total, 19 variables ont été testées sur la base IB. L'algorithme RUBY (cf figure) atteint une précision >95% pour l'identification des comptes rendus de chirurgies, le type de chirurgie et le côté opéré ; une précision supérieure >90% pour les variables d'identification des comptes rendus, SBR, HER2 amplification, expression œstrogène, date du compte rendu ; une précision >85% pour les variables côté de la tumeur, expression progesterone, curage axillaire, recherche de ganglion sentinelle, date de chirurgie. La précision est <85% pour les variables taille de la tumeur et type histologique. Quatre minutes ont été nécessaires pour la structuration des 19 variables des 227 patientes d'IB. Pour CHIRANA, 25 variables ont été structurées en 1h17 pour les 5702 patientes, à partir des comptes rendus de chirurgie et d'anatomopathologie.



CONCLUSION

Les premières applications de RUBY sont très encourageantes. Mais une nouvelle phase d'apprentissage est nécessaire afin d'améliorer les performances notamment sur les tailles tumorales et les types histologiques. Les performances de RUBY sont meilleures lorsque les comptes rendus testés sont contemporains de ceux de la période d'apprentissage de l'application. RUBY permet un gain de temps considérable et cette application sera un véritable succès lorsque toutes les précisions seront >90% à >95%. Des modèles pour d'autres pathologies sont en cours de développement au CAL.