

Prédiction du risque à deux ans d'événements cardiovasculaires indésirables majeurs (MACE) par un modèle d'apprentissage automatique utilisant des biomarqueurs non spécifiques et des variables cliniques.

Kamyar M. Hedayat^{1,2}, Théo Hennion¹, Maël Yang¹, Ahava Cohen¹, Rima Braukyliene³, Laura Zajanckauskiene³, Martynas Jurenas³, Ramunas Unikas³, Ali Aldujeli³, Osvaldas Petrokas³, Vytautas Zabiela³, Rasa Steponaviciute³, Astra Vitkauskine³, Brigita Hedayat², Sandrita Simonyte³, Vaiva Lesauskaite³, Jean Claude Lapraz², Diana Zaliaduonyte³

1-Numa Health International, La Rochelle, France 17000 - 2-Systems Biology Research Group, Chicago, IL 60626, USA. - 3-Cardiology Department, Lithuanian University of Health Sciences, LT 50161 Kaunas, Lithuania

Objectif

L'objectif de cette étude est de prédire le risque d'événements cardiovasculaires indésirables majeurs (MACE) chez les sujets ayant survécu à un premier infarctus aigu du myocarde en combinant des **facteurs biologiques non spécifiques** à des facteurs cliniques.

Matériel et Méthodes

En collaboration avec la société de bioanalyse SBRG de Chicago et l'Université Lithuanienne de Sciences Médicales, cette étude est venue compléter une analyse antérieure sur 315 sujets prédisant le risque de mortalité à deux ans après un premier infarctus du myocarde. L'étude inclut des analyses de dérivés biologiques clustérisés issus d'une NFS, développés par SBRG, indiqués par les préfixes C1, C2, etc.

Deux méthodes de cross-validation ont été employées sur un jeu d'entraînement (80% du jeu de données initial). Les variables d'intérêt ont été sélectionnées par analyse en composantes principales (PCA). Trois modèles d'apprentissage automatique ont été entraînés : **Decision Tree, Random Forest et XGBoost**, puis hyperparamétrés et enfin testés sur le jeu de test.

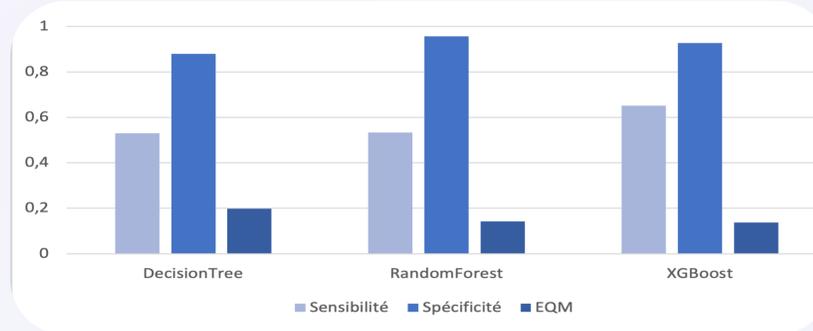


Figure 1: Métriques des modèles d'apprentissage automatique réalisés

Résultats

La population étudiée compte un âge moyen de 65 ans (64% d'hommes), dont 25,4% patients ont connu des événements cardiovasculaires indésirables majeurs en 2 ans.

La PCA a sélectionné **11 variables** contenant 90% de l'information nécessaire à la prédiction la plus performante : **âge, poids, IMC, CCMH, basophiles, leucocytes, monocytes, potassium, C3, C234, et C265**. Le modèle XGBoost s'est avéré le plus performant dans la prédiction moyenne de MACE : spécificité 92,8%, sensibilité 65,1%, erreur quadratique moyenne de 0,1373 (Fig1).

Discussion

Cette étude démontre la capacité de prédiction de MACE à 2 ans en utilisant des marqueurs non spécifiques aux maladies cardiovasculaires, et également la valeur prédictive des analyses de dérivés biologiques à composantes multiples clustérisés. Cette approche permettrait de réduire les coûts du suivi actuel en permettant d'identifier les patients à risques. Cependant, la faible sensibilité observée, ainsi que le risque de surapprentissage, indiquent qu'un échantillon de données plus important est nécessaire pour améliorer la robustesse et la généralisation du modèle.

Conclusion

Cette étude est prometteuse en tant que première étape vers l'analyse prédictive en utilisant des données non spécifiques pour une pathologie en exploitant l'approche d'apprentissage automatique de Numa Health. Les dérivés de biomarqueurs offrent la possibilité de décrypter les indicateurs subtils de la physiopathologie sous-jacente des maladies.