

MÉTHODE DE CRIBLAGE PAR DÉCIMATION POUR L'IDENTIFICATION D'INFORMATIONS RADIOMIQUES DISCRIMINANTES

Anne-Sophie Dirand, Frédérique Frouin, Irène Buvat

IMIV (UMR1023 Inserm/CEA/Université Paris Sud et ERL 9218 CNRS), CEA SHFJ, 91401 Orsay Cedex

Introduction

En imagerie médicale, de nombreux travaux s'attachent actuellement à élaborer des modèles radiomiques, ie combinant des caractéristiques extraites d'images médicales, pour classer des patients ou des anomalies (eg, bons et mauvais répondeurs à une thérapie). Lorsqu'aucun modèle présentant des performances satisfaisantes n'est identifié, il est difficile de déterminer si cela est dû à l'absence d'informations pertinentes dans les données ou à un nombre insuffisant de patients. Nous proposons une méthode de criblage afin d'identifier la présence d'informations dans les caractéristiques radiomiques pour une tâche de classification en 2 classes et d'estimer les performances de classification atteignables si le nombre de patients n'était pas limité.

Méthode

Deux familles de méthodes statistiques (univariée et multivariées) sont considérées pour élaborer des modèles prédictifs : l'analyse Receiver Operating Characteristic (ROC) univariée et l'apprentissage automatisé (AA). Pour les études ROC et l'AA, la cohorte de patients utilisée pour construire un modèle prédictif est séparée en 2 parties, un ensemble composé de données pour l'apprentissage et pour la prédiction des performances du modèle et un ensemble pour évaluer la fiabilité du modèle. Cette séparation est répétée 50 fois, en assignant les patients de manière aléatoire à chacune des 2 parties. Les performances de classification sont caractérisées par l'index de Youden (sensibilité + spécificité - 1) et son écart-type ($YI \pm SD$). L'analyse ROC est réalisée pour chacun des paramètres extraits des images. L'AA est réalisé par validation croisée, en utilisant un modèle basé sur les séparateurs à vaste marge ou la régression logistique avec ou sans réduction de dimension et la méthode LASSO (Least Absolute Shrinkage and Selection Operator).

Le criblage est effectué en partant du nombre total de patients disponibles et en diminuant l'effectif progressivement pour créer des sous-échantillons de plus en plus petits. Les performances attendues si l'effectif était suffisant pour atteindre la convergence (YI_c) sont estimées en ajustant la courbe représentant l'évolution des YI pour chaque sous-échantillon d'effectif N_k par un modèle logarithmique :

$$YI(N_k) = a \cdot \ln(N_k) + b \quad (\text{Eq 1})$$

et en extrapolant sa valeur à convergence. La qualité de la prédiction est mesurée avec 2 index :

$$QF1 = |YI_p(N) - YI_c| \quad (\text{Eq 2})$$

$$QF2 = |YI(N) - YI_c| \quad (\text{Eq 3})$$

avec N le nombre de patients disponible, YI_p le YI prédit en extrapolant à partir de l'équation 1, et YI_c le YI effectivement observé à convergence, ie lorsque YI n'augmente plus quand l'effectif augmente. $QF1 \leq QF2$ montre l'apport de la méthode de décimation pour prédire les performances d'un modèle à partir des données disponibles de N patients.

La méthode a été évaluée sur 2 cohortes C1 (414 patients, 53 caractéristiques radiomiques) [1], C2 (569 patients, 30 caractéristiques) [2] et sur 2 cohortes C3 et C4 artificiellement créées à partir de C1 et C2 et dénuées d'information prédictive. A partir de ces 4 cohortes, 602 situations expérimentales ont été créées correspondant à des nombres de patients disponibles allant de 36 à 220 (C1 & C3) et de 19 à 73 (C2 & C4).

Résultats

Pour C1 et C2 toutes les méthodes de classification évaluées convergent lorsque le nombre total de patients est utilisé (convergence à partir 360 patients pour C1 et de 108 pour C2). Dans 183 des 216 situations expérimentales, $QF1 \leq QF2$ pour C1 (151/171 pour C2). Les performances atteignables à convergence (ie avec une cohorte plus grande que la cohorte analysée dans la situation expérimentale considérée) peuvent être prédites de manière précise (exemple dans le tableau 1) et la méthode de criblage ne conclut pas à une présence d'information lorsqu'il n'y en a pas (Youden index prédit ≤ 0.10 pour 119/120 cas pour C3 et 81/95 cas pour C4).

Nombre de patients	QF1	QF2
68	0.00	0.37
60	0.26	0.39
52	0.05	0.36
44	0.18	0.44
36	0.04	0.47

Tableau 1 : $QF1$ et $QF2$ en fonction de différents effectifs pour Recursive Feature Elimination-Support Vector Machine (C1).

Conclusion

La méthode de décimation proposée permet de déterminer, à partir d'un jeu de données, les performances de classification atteignables si l'on disposait d'un très grand nombre de patients pour créer un modèle.

Références

1. Kirienko et al, Eur J Nucl Med Mol Imaging, 45:1649-1660, 2018.
2. Wolberg et al, Human Pathology 26:792-796, 1995.