

Méthode de sélection des caractéristiques basée sur le random forest : prédiction de l'évolution du cancer de l'œsophage en imagerie TEP

Desbordes P^{1,2} Ruan S¹ Modzelewski R^{1,3} Vauclin S² Vera P^{1,3} Gardin I^{1,3}

¹ LITIS - QUANTIF - EA4108, Université de Rouen, France

² DOSISOFT, Cachan, France; ³ CENTRE HENRI BECQUEREL, Rouen, France

22 boulevard Gambetta 76000 Rouen
desbordespaul@gmail.com

Résumé

En oncologie, de nombreuses caractéristiques peuvent être extraites des images TEP afin de décrire les lésions tumorales. Afin de détecter les sous-ensembles de caractéristiques les plus pertinents, nous proposons une nouvelle stratégie de sélection, en 3 étapes, basée sur les forêts aléatoires (RF). La première étape est une analyse des corrélations de Spearman effectuée dans le but d'éliminer les caractéristiques corrélées. Puis, un classement des caractéristiques restantes par ordre d'importance est réalisé grâce à la méthode de RF. Enfin, les meilleurs sous-ensembles sont définis à l'aide d'un algorithme génétique (GA). Notre méthode est évaluée par classification RF sur une base de données de 65 patients présentant un cancer de l'œsophage. Une comparaison de la méthode est également réalisée avec d'autres méthodes de sélection de caractéristiques. Les résultats montrent que notre méthode présente de meilleures performances que celles évaluées.

Mots Clef

Sélection de caractéristiques, cancer de l'œsophage, forêts aléatoires, radiomics

Abstract

In oncology, many features can be extracted from PET images to describe tumours. We propose a new feature selection strategy based on random forests (RF) to detect the best predictive subsets of features. Our method is performed in 3 steps. First, a Spearman correlation analysis is carried out to eliminate the correlated features. Then, the achievement of importance rankings of the remaining features according to the RF is done. Finally, the best subsets of features are defined through a genetic algorithm (GA). Our method is evaluated by RF classification based on data from 65 patients with an oesophageal cancer. A comparison with other features selection strategies is performed. Our method gives more accurate results than those evaluated.

Keywords

Feature Selection, Oesophageal cancer, Random Forest, Radiomics

1 Introduction

En oncologie, l'imagerie TEP au ¹⁸FDG (tomographie par émission de positons au 18-fluorodeoxyglucose) est l'un des outils majeurs pour le suivi de l'évolution des lésions tumorales [1]. Les caractéristiques du 1^{er} ordre extraites de ces images, comme le MTV (metabolic tumor volume) ou le SUV (standardized uptake value), ont d'ores et déjà montrées leur pertinences dans des études prédictives et pronostiques de tumeurs solides [1].

Plus récemment, des indices de texture ont été étudiées dans le but de décrire l'hétérogénéité de fixation du ¹⁸FDG dans la tumeur. Cette hétérogénéité serait le reflet de l'agressivité tumorale. *El Naqa et al.* [2] ont proposé l'étude de caractéristiques extraites du SVH (histogramme SUV-volume). Ces caractéristiques se sont révélées être pertinentes dans l'étude des cancers ORL et du col de l'utérus. Cette équipe a également trouvé que des caractéristiques de texture extraites de la matrice GLCM (gray level cooccurrence matrix [3]) présentaient une valeur prédictive dans le cancer du col. Plusieurs autres matrices de texture ont également été étudiées conduisant à l'extraction d'un grand nombre d'indices. *Tixier et al.* [4] ont étudié la valeur prédictive de 38 caractéristiques extraites des images TEP au ¹⁸FDG chez des patients atteints d'un cancer de l'œsophage. Leurs résultats sont basés sur des calculs d'aires sous des courbes ROC (receiver operating characteristic). Ces auteurs ont trouvé que certaines caractéristiques issues de la matrice GLCM (second angular moment, local contrast, entropy, correlation, homogeneity, dissimilarity) et de la matrice GLSZM (gray level size zone matrix [5], zone length and gray level non uniformity) étaient prédictives de la réponse thérapeutique des patients.

Cependant, l'augmentation du nombre de caractéristiques

n'est pas nécessairement synonyme d'amélioration des performances. *Orlhac et al.* [6] ont montré, sur 3 types de cancers, que certaines caractéristiques étaient fortement corrélées avec le MTV. De même, *Tixier et al.* [4] ont montré que les caractéristiques issues de la matrice GLRLM (gray level run length matrix [7]) étaient fortement corrélées avec celles de la matrice GLSZM. Ces informations redondantes sont susceptibles de réduire les performances de prédiction, c'est pourquoi les corrélations entre caractéristiques sont à prendre en compte lors de la sélection.

Compte tenu du grand nombre de caractéristiques, les algorithmes de machine learning semblent particulièrement appropriés [2]. A notre connaissance, ces méthodes ont été peu appliquées à l'imagerie médicale TEP. Dans *Parmar et al.* [8], 14 méthodes de sélection de caractéristiques combinées avec 12 classificateurs ont été étudiées. Cette étude a été réalisée sur une base de données de 464 patients atteints d'un cancer des poumons d'où ont été extraites 440 caractéristiques des images TDM (tomodensitométrie). Les auteurs ont conclu que la méthode des forêts aléatoires (RF) est la méthode de classification présentant les meilleures performances. Plus généralement, 3 types de sélection de caractéristiques peuvent être distingués [9] : la méthode filtrante ("filter"), la méthode enveloppante ("wrapper") et la méthode intégrée ("embedded"). La méthode filtrante comme par exemple la méthode FAST (feature assessment by sliding thresholds) [10], est basée uniquement sur des propriétés générales comme la corrélation entre une caractéristique et la prédiction. La méthode enveloppante évalue les sous-ensembles de caractéristiques et détecte les interactions entre caractéristiques. La méthode SFS (sequential forward selection) [11] et HFS (hierarchical forward selection) [12] font partie de cette famille. La méthode intégrée combine plus étroitement les phases de sélection des caractéristiques et d'estimation du modèle comme par exemple la méthode RFE (recursive feature elimination) [13].

Dans cette article, nous proposons une méthode de sélection de caractéristiques basée sur le RF afin de détecter les sous-ensembles de caractéristiques prédictives et pronostiques les plus pertinents parmi un nombre important de caractéristiques multimodales (TEP et clinique) issues d'une cohorte de 65 patients présentant un cancer de l'œsophage. Cette méthode a ensuite été comparée à d'autres méthodes de sélection de caractéristiques.

2 Matériel et méthodes

Notre méthode, nommée RF-FSS, pour Random Forest Feature Selection Strategy, consiste en la détection des meilleurs sous-ensembles de caractéristiques prédictifs et pronostiques parmi celles extraites au préalable. Le calcul des caractéristiques et les analyses ont été réalisés avec le logiciel Matlab et la "Statistics Toolbox" (version 2014a,

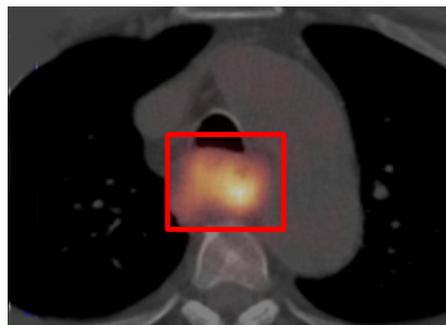


Figure 1: Coupe transverse au ^{18}F FDG-TEP/TDM d'un thorax avec une tumeur œsophagienne encadrée.

The MathWorks, Inc., Natick, Massachusetts, USA).

2.1 Cohorte de patients

Dans cette étude rétrospective, 65 patients atteints d'un cancer de l'œsophage localement avancé et éligible à une radio-chimiothérapie (RCT) ont été sélectionnés. Treize caractéristiques ont été extraites du dossier médical des patients (Tableau 1). On retrouve par exemple le poids de forme des patients ainsi que leur poids actuel, le stade de la maladie, l'évaluation de la dénutrition, etc.

Les patients ont tous bénéficié d'un examen TEP initial avant traitement au sein du même département de médecine nucléaire sur un TEP/TDM Biograph Sensation 16 (Siemens, Erlangen, Germany). L'acquisition a été faite de la base du crâne à l'extrémité proximale des cuisses après injection d'une activité de 5 MBq/kg de ^{18}F FDG, 60 minutes avant l'examen. Après reconstruction, la taille des voxels était de $4 \times 4 \times 2 \text{ mm}^3$. Un exemple de coupe TEP/TDM transversale du thorax est donné Figure 1.

La réponse thérapeutique a été évaluée un mois après la fin du traitement par un examen clinique, un examen endoscopique avec biopsies et par imagerie TEP et TDM. Les patients ont été séparés en 2 catégories : ceux présentant une réponse complète au traitement (CR) et ceux non-répondeurs ou présentant des lésions résiduelles (NCR). Dans notre cohorte, 41 patients (62%) ont été considérés comme CR, alors que 24 (38%) ont été considérés comme NCR. Ces données ont été utilisées pour l'étude prédictive de la réponse au traitement. La survie globale (OS) utilisée pour l'étude pronostique a été réalisée 3 ans après la fin de la RCT. En fin de suivi, 16 patients étaient vivants (24%) contre 40 morts (76%).

2.2 Extraction des caractéristiques de l'image

45 caractéristiques ont été extraites des images TEP au ^{18}F FDG (Tableau 1). Le MTV a d'abord été défini par une méthode de segmentation par seuillage adaptatif [14]. Le MTV moyen était de $19,6 \pm 20,5 \text{ cm}^3$ ($2,5 - 141 \text{ cm}^3$).

De ce volume ont été extraites 19 caractéristiques du 1^{er} ordre et de forme comme les dérivés du SVH [2], le COV [15] (coefficient of variation) ou la sphéricité [16].

Puis, 26 indices de texture ont été extraites de 3 matrices : 10 de la matrice GLCM (moyennées sur les 13 directions 3D [17]), 5 de la matrice GLDM (gray level difference matrix [18]) et 11 de la matrice GLSZM. Avant de calculer ces matrices, les images TEP ont été ré-échantillonnées selon l'équation 1 [19].

$$I(i) = \text{arrondi} \left(D \times \frac{\text{SUV}(i)}{\text{SUV}_{\text{max global}}} \right) \quad (1)$$

Où $I(i)$ correspond à la mesure du nouveau niveau de gris du voxel i de niveau de gris initial $\text{SUV}(i)$, $\text{SUV}_{\text{max global}}$ est le niveau de gris le plus élevé dans toute l'étude (c'est-à-dire parmi tous les patients) et D est le nombre de niveau de gris utilisé dans l'étude sachant que le pas d'intensité $D/\text{SUV}_{\text{max global}}$ est fixé à 0,5. Selon Orlhac et al. [19], cette méthode est la plus appropriée dans la comparaison des caractéristiques issues de différents patients.

Au final, les caractéristiques de l'image TEP additionnées aux 13 issues du dossier médical mènent à un nombre $F_i = 58$ caractéristiques multimodales étudiées.

2.3 Sélection des caractéristiques : RF-FSS

Notre méthode RF-FSS se déroule en 3 étapes (voir Figure 2) et peut être définie comme une méthode enveloppante. Les caractéristiques corrélées sont tout d'abord éliminées. Celles restantes sont classées selon leur coefficient d'importance calculé par la méthode RF [20]. Enfin, les meilleurs sous-ensembles sont déterminés à l'aide d'un algorithme génétique (GA) [21].

Comme 1^{ère} étape, une analyse de corrélation des rangs de Spearman est réalisée afin de conserver les caractéristiques non corrélées. Cette analyse calcule le coefficient de corrélation (sp) entre 2 caractéristiques étudiées [22], selon :

$$sp = 1 - \frac{6 \sum d^2}{N(N^2 - 1)} \quad (2)$$

où N est le nombre d'observations et d la différence de rang entre les deux caractéristiques étudiées. Le rang est la position de la variable par ordre croissant. Une corrélation a été considérée comme importante si $|sp| \geq 0,8$ avec une p-value inférieure à 5% [6]. Les caractéristiques corrélées entre elles ont été regroupées au sein de groupes vérifiant ces paramètres. Pour chaque groupe, la valeur moyenne de $|sp|$ et l'écart-type ont été calculés. Chaque groupe a été représenté par la caractéristique possédant l'indice de Spearman le plus élevé vis-à-vis des autres membres de son groupe. Le nombre de caractéristiques sélectionnées par cette étape F_{nc} est égale au nombre de caractéristiques

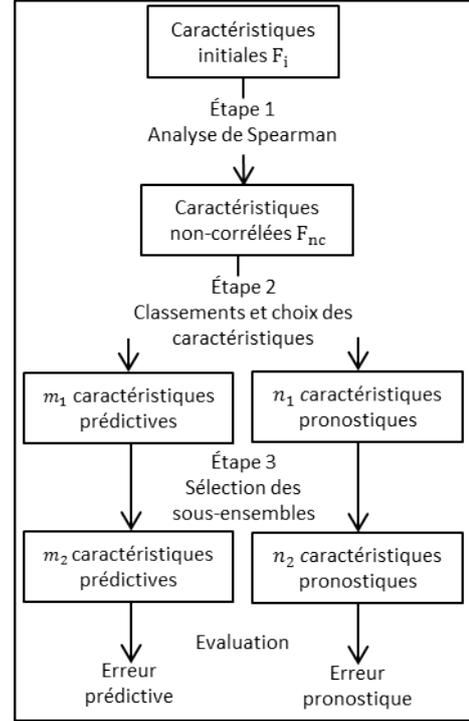


Figure 2: Stratégie de sélection des caractéristiques dans la méthode RF-FSS.

non corrélées auquel il faut ajouter le nombre de groupes de corrélation.

La 2^{ème} étape consiste à réaliser un classement des caractéristiques en fonction de leur coefficient d'importance C_i déterminé par l'algorithme RF. Dans un premier temps, il est nécessaire de construire une forêt. Cette dernière réalise la classification de données en se basant sur une multitude d'arbres de décisions T . Soit une observation x de label connu $L(x)$, le label estimé $\hat{L}_T(x)$ est calculé par vote majoritaire sur l'ensemble des arbres T , selon :

$$\hat{L}_T(x) = \text{vote majoritaire} \left(\hat{L}_t(x) \right)^T \quad (3)$$

où $\hat{L}_t(x)$ est le label estimé par le $t^{\text{ème}}$ arbre.

La construction de chaque arbre est basée sur un échantillon bootstrap de k observations tirées aléatoirement avec remise. Pour chaque nœud d'un arbre, une caractéristique est sélectionnée en 2 étapes afin de séparer le nœud étudié en 2 nœuds fils. Tout d'abord, une liste de $\sqrt{F_{nc}}$ caractéristiques est tirée aléatoirement sans remise. La caractéristique sélectionnée pour le nœud sera la plus discriminante selon le critère de Gini [23].

Le coefficient d'importance est calculé à partir de l'erreur out-of-bag (OOB_{err}) de la façon suivante. Soit t^x l'ensemble des arbres n'utilisant pas l'observation x dans

Tableau 1: Récapitulatif des caractéristiques tumorales étudiées.

Type de caractéristiques	Caractéristiques
Cliniques	Age du patient, Sexe du patient, Poids actuel du patient (kg), Poids de forme, Perte de poids (%), Albuminémie (g/l), NRI, Malnutrition*, TNM, Score de Zubrod, Localisation tumorale (tiers sup, moy, inf), Histologie (ADC or SCC), Longueur tumorale par endoscope (cm)
1 ^{er} ordre	SUV _{max} , SUV _{mean} , SUV _{peak} , somme des SUV MTV, TLG, Ecart-type (SD), COV, Sphéricité, Skewness, Kurtosis, Energie, Entropie, SUV ₁₀ , SUV ₉₀ , SUV ₁₀₋₉₀ , V ₁₀ , V ₉₀ , V ₁₀₋₉₀
Texture **	<i>GLCM</i> [3] : Variance, Énergie, Entropie, Corrélacion, Dissimilarité, Contraste, Homogénéité, Moment différentiel inverse (IDM), Cluster Shade (CS), Cluster Tendency (CT) <i>GLSZM</i> [5] : Short Zone Emphasis (SZE), Long Zone Emphasis (LZE), Low Gray level Zone Emphasis (LGZE), High Gray-level Zone Emphasis (HGZE), Short Zone Low Gray-level Emphasis (SZLGE), Long Zone Low Gray-level Emphasis (LZLGE), Short Zone High Gray-level Emphasis (SZHGE), Long Zone HighGray-level Emphasis (LZHGE), Zone Percentage (ZP), Gray Level Non Uniformity (GLNUz), Zone Length Non Uniformity (ZLNU) <i>GDLM</i> [7] : Coarseness, Contrast, Busyness, Complexity, Strength

* absence si le NRI > 97.5, moyen si 83.5 ≤ NRI ≤ 97.5 et sévère si le NRI < 83.5.
** formules mathématiques dans les données supplémentaires de [6].

leur construction et $\widehat{L}_{t\bar{x}}(x)$ leur estimation par vote majoritaire du label de x , alors l' OOB_{err} est définie de la manière suivante :

$$OOB_{err} = \frac{\sum_{x=x_1, \dots, x_N} \delta(L(x), \widehat{L}_{t\bar{x}}(x))}{N} \quad (4)$$

Où δ est une fonction égale à 1 si $L(x) \neq \widehat{L}_{t\bar{x}}(x)$ et N le nombre d'observations.

L'importance des caractéristique est alors calculée de la façon suivante. Une caractéristique X est permutée aléatoirement dans l'échantillon test. L' OOB_{err} calculée initialement est comparée à celle issue de cet échantillon perturbé, notée \widetilde{OOB}_{err} . Le coefficient d'importance C_i est obtenu en calculant la différence entre ces 2 erreurs, tel que :

$$C_i(X) = \widetilde{OOB}_{err}(X) - OOB_{err}(X) \quad (5)$$

Plus cette différence est grande, plus la caractéristique est importante. Cette étape est réalisée pour l'intégralité des caractéristiques. Seules les caractéristiques ayant un C_i supérieur à un seuil de 20% du coefficient le plus élevé ont été sélectionnées menant, à la fin de l'étape 2, à m_1 caractéristiques prédictives (liées à réponse tumorale au traitement) et n_1 caractéristiques pronostiques (liées à la survie

globale des patients).

Après avoir retenu les caractéristiques ayant les coefficients d'importance les plus élevés, la dernière étape consiste à détecter quelles sont les sous-ensembles de caractéristiques présentant les meilleures valeurs prédictives et pronostiques. A cet effet, un algorithme génétique a été utilisé. Cette méthode récursive, basée sur la sélection naturelle, consiste à tester un lot de sous-ensembles de caractéristiques (appelés chromosomes) puis de détecter le ou les meilleurs grâce à une fonction fitness. Le meilleur chromosome "père" subira alors quelques mutations menant à de nouveaux chromosomes "fils". Après plusieurs générations, l'algorithme va converger vers la meilleure solution possible. Dans cet article, nous proposons de définir la fonction fitness comme étant la précision d'une classification par RF évaluée par la méthode de K -fold cross validation avec $K = 5$. Cette étape mène à la détection du meilleur sous-ensemble prédictif de caractéristique de taille m_2 et du meilleur sous-ensemble pronostique de taille n_2 .

3 Résultats

3.1 Critères d'évaluation des classifieurs

Les sous-ensemble finaux de caractéristiques générés ont été évalués par une classification de type RF associée à une

méthode de K -fold cross validation ($K = 5$). L'évaluation a porté sur les calculs de précision de la classification (Acc), de la sensibilité (Se) et de la spécificité (Sp) par comparaison des labels estimées et des labels connus. Enfin, en utilisant une régression logistique nous avons calculé des AUC (area under the ROC curves) pour chaque sous-ensemble.

3.2 Analyse des corrélations

Les résultats de l'analyse des corrélations de Spearman des 58 caractéristiques initiales sont donnés dans le tableau 2. On note que concernant les caractéristiques cliniques, le poids de forme du patient et son poids actuel sont fortement corrélés ($|sp| > 0,96$), de même pour l'albuminémie, le NRI et la malnutrition ($|sp| > 0,84$). Seul 8 des caractéristiques de l'image TEP sont non-corrélées : COV, Skewness, Kurtosis, SUV₉₀, V₁₀, Contraste (GLDM), SZE, SZLGE et GLNUz. A la fin de cette étape, 9 groupes de corrélations ($|sp| > 0,8, p < 5\%$) ont été identifiés et une 1^{ère} sélection de $F_{nc} = 26$ caractéristiques sur les 58 initiales (10 cliniques et 16 d'images) a été réalisée.

Tableau 2: Groupes de corrélations avec la valeur $|sp|$ et l'écart-type associé. En gras, la caractéristique représentant chaque groupe.

N°	Caractéristiques corrélées	Indices de Spearman
1	Poids actuel - Poids de forme	0,96
2	NRI - Albuminémie - Malnutrition	0,88 ± 0,04
3	V₁₀₋₉₀ - V ₉₀	0,96
4	Énergie - Entropie	0,99
5	MTV - sumSUV - TLG - Corrélation (GLCM)	0,89 ± 0,03
6	HGZE - SUV ₁₀ - Variance - CT - SZHGE - SUV _{max} - SUV _{peak} - SUV _{mean} - Complexité - SD - SUV ₁₀₋₉₀ - Contraste (GLDM) - LGZE	0,91 ± 0,03
7	IDM - Homogénéité - LZE - ZP - Dissimilarité - Énergie (GLCM) - LZLGE - LZHGE - Strength - Entropie (GLCM)	0,91 ± 0,03
8	ZLNU - CS	0,81
9	Busyness - Coarseness - Sphéricité	0,94 ± 0,02

3.3 Sélection des caractéristiques

Dans la 2^{ème} étape, l'importance des caractéristiques est étudiée. Nous avons utilisé un nombre d'arbres de décision T égal à 500. Des valeurs plus élevées ont été également testées sans montrer d'amélioration significative. Chaque arbre du classifieur par RF est basé sur un échantillon bootstrap de 65 patients. $m_1 = 9$ caractéristiques prédictives et $m_2 = 6$ pronostiques sont sélectionnées par

Tableau 3: Caractéristiques prédictives et pronostiques sélectionnées par l'étape 2 de la méthode RF-FSS en fonction de la valeur de C_i . Le groupe de corrélation est indiqué.

N°	Caractéristiques prédictives	Caractéristiques pronostiques
1	MTV (grp 5)	Age
2	Busyness (grp 9)	NRI (grp 2)
3	GLNUz	Score de Zubrod
4	Énergie (grp 4)	Poids actuel (grp 1)
5	Perte de poids	MTV (grp 5)
6	IDM (GLCM) (grp 7)	Énergie (grp 4)
7	HGZE (grp 6)	-
8	Poids actuel (grp 1)	-
9	SZLGE	-

l'étape 2 de l'algorithme RF-FSS (voir Tableau 3). Ces caractéristiques sont ensuite utilisées par l'algorithme génétique pour déterminer les sous-ensembles ayant les meilleures valeurs prédictive et pronostique avec les paramètres suivants : 30 chromosomes et 30 générations.

Trois autres méthodes de sélection ont été étudiées à visée de comparaison. Il s'agit des méthodes HFS, SFS et RFE. Ces 3 méthodes sont couplées avec un SVM présentant un noyau gaussien RBF (radial basis function avec $\sigma = 1$, $C = 1$). Elles ont réalisé une classification à partir des F_{nc} caractéristiques non corrélées après la première étape, comme notre méthode.

Concernant l'étude prédictive, notre méthode RF-FSS a sélectionné comme meilleur sous-ensemble les 3 caractéristiques suivantes : le MTV (grp 5), la perte de poids et le moment différentiel inverse (IDM, grp 7) issu de la matrice GLCM. Parallèlement, la méthode SVM-HFS a sélectionné 3 caractéristiques : le MTV (grp 5), le contraste (GLDM) et le stade N de la maladie. La méthode SVM-SFS en a sélectionné 2 : le MTV (grp 5) et l'IDM (GLCM, grp 7). Enfin, la méthode SVM-RFE en a sélectionné 3 : Kurtosis, Busyness (GLDM, grp 9) et GLNUz.

Concernant l'étude pronostique, le meilleur sous-ensemble sélectionné par notre méthode RF-FSS est composé des 3 caractéristiques suivantes : l'age, le score de Zubrod, ainsi que le poids actuel du patient. La méthode SVM-HFS en a sélectionné 4 : le MTV (grp 5), le NRI, l'IDM (GLCM, grp 7) et le contraste (GLDM), la méthode SVM-SFS 3 : le MTV (grp 5), le GLNUz et le sexe, ainsi que la méthode SVM-RFE : le MTV (grp 5), le GLNUz et HGZE (grp 6).

3.4 Évaluation des méthodes de sélection

Dans un 1^{er} temps, l'influence de l'étape 1 qui concerne les corrélations entre caractéristiques, a été étudiée. Les

résultats sont présentés dans le tableau 4.

Tableau 4: Résultats de l'évaluation de l'impact de l'étape 1 dans la méthode de sélection des caractéristiques RF-FSS pour l'étude prédictive (pré) et pronostique (pro). Nbr est le nombre de caractéristiques sélectionnées.

	Méthode	Nbr	Acc	Δ Acc	Se	Sp
pré	Sans étape 1	10	80	+3%	85	71
	Avec étape 1	3	83		85	79
pro	Sans étape 1	9	75	+8%	85	60
	Avec étape 1	3	83		88	76

Dans le tableau 5 sont donnés les résultats de l'évaluation des différents sous-ensembles de caractéristiques sélectionnées par les différentes méthodes de sélection étudiées.

Tableau 5: Résultats de l'évaluation des différents sous-ensembles de caractéristiques avec la méthode RF-FSS et les 3 autres méthodes testées. Δ Acc correspond à la différence de précision de classification entre RF-FSS et les autres méthodes.

	Méthode	Nbr	Acc	Δ Acc	Se	Sp	AUC
Prédictif	RF-FSS	3	83	/	85	79	0,757
	SVM-HFS	3	68	+15%	81	46	0,751
	SVM-SFS	2	80	+3%	85	71	0,774
	SVM-RFE	3	63	+19%	73	46	0,760
Pronostique	RF-FSS	3	83	/	88	76	0,738
	SVM-HFS	4	71	+12%	80	56	0,688
	SVM-SFS	3	59	+24%	75	32	0,604
	SVM-RFE	3	61	+22%	73	44	0,561

4 Discussion

Notre méthode de sélection de caractéristiques RF-FSS a montré de très bons résultats avec une précision de classification par la méthode RF de 83% à la fois pour l'étude prédictive et pour l'étude pronostique. De même, la comparaison de notre méthode avec 3 autres méthodes de sélection de caractéristiques montre également les très bonnes performances de notre stratégie (Tableau 5).

Une analyse des corrélations de Spearman a été réalisée afin de conserver les caractéristiques non-corrélées [6]. Le fait qu'aucune des caractéristiques cliniques ne soit corrélée avec celles issues de l'image TEP (Tableau 2) montre qu'il existe une potentielle complémentarité entre ces 2 modalités d'informations. Les résultats de la classification comparant l'impact de l'étape 1 (cf. Tableau 4) montrent que l'élimination des caractéristiques corrélées améliore les résultats de la sélection de caractéristiques de 3% en prédictif et de 8% en pronostique. En effet, les caractéristiques corrélées n'apportent pas d'informations supplémentaires. Cette redondance pourrait induire en erreur la sélection de sous-ensembles de caractéristiques

complémentaires. C'est pourquoi l'élimination de ces caractéristiques ($F_{nc} = 26/58$) permet d'améliorer les résultats.

Concernant les sous-ensembles de caractéristiques sélectionnées par les différentes méthodes, notons la présence du MTV (groupe 5) dans la quasi-totalité des résultats prédictifs (seul le RFE ne l'a pas retenu). Cela confirme l'intérêt particulier de cette caractéristique dans le suivi du traitement du patient [1]. Au contraire, le SUV_{max} (groupe 6) qui est généralement considéré comme une caractéristique importante, ne ressort pas dans les études prédictives et pronostiques (excepté par l'étude pronostique avec la méthode RFE).

Enfin, notons que le sous-ensemble prédictif sélectionné par RF-FSS combine des caractéristiques cliniques et d'imagerie TEP initiales (de 1^{er} ordre et de texture). Cela montre l'intérêt prédictif des caractéristiques de la TEP initiale. Cependant, le sous-ensemble pronostique contient uniquement des caractéristiques cliniques (age, poids actuel et score de Zubrod). Nous pouvons alors supposer qu'il est difficile de prédire la survie à long terme à partir de l'examen TEP réalisé avant le début du traitement uniquement. Il peut alors être intéressant de se tourner vers l'analyse de l'évolution longitudinale de ces caractéristiques entre l'examen TEP initial et un de ceux réalisés en cours de traitement [1].

La comparaison de nos résultats avec ceux des autres méthodes étudiées montre que notre stratégie est nettement plus efficace avec un gain d'au moins 3% pour l'étude prédictive et 12% pour l'étude pronostique (Tableau 5). Seul l'étude prédictive de la méthode SVM-SFS présente des résultats similaires aux nôtres. Si l'on s'intéresse à l'étude prédictive, l'AUC est équivalent pour l'ensemble des méthodes (0,751 pour la méthode HFS à 0,774 pour la méthode SFS). Cependant, notre méthode RF-FSS présente une précision de classification (Acc) supérieure avec 83% de précision contre 80% pour SFS, 68% pour la HFS et 63% pour la RFE. RF-FSS présente également la meilleure sensibilité (85%) et la meilleure spécificité (79%).

Concernant l'étude pronostique, notre méthode montre de meilleurs résultats à tous les niveaux. L'AUC est de 0,738 contre 0,688 pour la méthode HFS qui est la 2^{ème} meilleure méthode. Les observations sont similaires pour la précision de classification (83% de bonne classification contre 80% pour la méthode HFS), la sensibilité (88% contre 80% pour la méthode HFS) et la spécificité (76% contre 56% pour la méthode HFS).

5 Conclusion

En conclusion, nous avons montré que notre méthode de sélection de caractéristiques RF-FSS améliore la précision de la classification par rapport aux méthodes concurrentes étudiées d'au moins 3% en prédictif et d'au moins 12% en pronostique. Ces bonnes performances se retrouvent

pour les autres critères d'évaluation (AUC, sensibilité et spécificité). Les techniques de machine learning, et plus particulièrement le RF, apporte une expertise nécessaire dans la détection de sous-ensembles de caractéristiques multimodales.

Cependant, il reste nécessaire de tester cette méthode sur une cohorte de patients plus importante afin d'en évaluer la robustesse. De plus, il pourrait être intéressant de réaliser cette sélection sur une cohorte de patients avec un autre type de cancer pour comparer les variations. Enfin, une piste d'étude à envisager est la modification de la fonction fitness associée au GA afin d'optimiser les résultats.

References

- [1] C. Van de Wiele, V. Kruse, P. Smeets et al. Predictive and Prognostic Value of Metabolic Tumour Volume and Total Lesion Glycolysis in Solid Tumours. *Eur J Nucl Med Mol Imaging*, 40(2):290-301, 2013.
- [2] I. El Naqa, PW. Grisby, A. Apte et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit*, 42(6):1162-1171, 2009.
- [3] R. Haralick, K. Shanmugam, I. Dinstein. Textural Features for Image Classification. 1973.
- [4] F. Tixier, C. Cheze-le Rest, M. Hatt et al. Intratumor Heterogeneity Characterized by Textural Features on Baseline 18F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer. *J Nucl Med*, 52(3):369-378, 2011.
- [5] G. Thibault, B. Fertil, C. Navarro et al. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. *Pattern. Recognit. Inf. Process*, 140-145, 2009.
- [6] F. Orlhac, M. Soussan, JA. Maisonobe et al. Tumor Texture Analysis in 18F-FDG PET: Relationships Between Texture Parameters, Histogram Indices, Standardized Uptake Values, Metabolic Volumes, and Total Lesion Glycolysis. *J Nucl Med*, 55(3):414-422, 2014.
- [7] M. Galloway. Texture Analysis Using Gray Level Run Lengths. *Comput Graph Image Process*, 4(2):172-179, 1975.
- [8] C. Parmar, P. Grossmann, J. Bussink et al. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep*, 5:13087, 2015.
- [9] G. Chandrashekar, F. Sahin. A Survey on Feature Selection Methods. *Computers & Electrical Engineering*, 40(1):16-28, 2013
- [10] W.M. Chen. FAST:A roc-based feature selection metric for small samples and imbalanced data classification problems. *ACM SIGKDD, KDD 08 New York, NY, USA*, 124-132, 2008.
- [11] A. Whitney. A direct method of nonparametric measurement selection. *IEEE Trans Comput*, 100(9):1100-1103, 1971.
- [12] H. Mi, C. Petitjean, B. Dubray et al. Robust Feature Selection to Predict Tumor Treatment Outcome, *Elsevier, Artificial Intelligence in Medicine*, 64(3):195-204, 2015.
- [13] I. Guyon, J. Weston, S. Barnhil et al. Gene Selection for Cancer Classification using Support Vector Machines. *Mach Learn*, 46:389-422, 2002.
- [14] S. Vauclin, K. Doyeux, S. Hapdey et al. Development of a generic thresholding algorithm for the delineation of 18F-FDG-PET- positive tissue: Application to the comparison of three thresholding models. *Phys Med Biol*, 54(22):6901-6916, 2009.
- [15] R. Bunduschuh, J. Dinges, L. Neumann et al. Textural Parameters of Tumor Heterogeneity in 18F-FDG PET/CT for Therapy Response Assessment and Prognosis in Patients with Locally Advanced Rectal Cancer. *J Nucl Med*, 55(6):891-897, 2014.
- [16] F. Hofheinz, A. Lougovski, K. Zöphel et al. Increased Evidence for the Prognostic Value of Primary Tumor Asphericity in Pretherapeutic FDG PET for Risk Stratification in Patients with Head and Neck Cancer. *Eur J Nucl Med Mol Imaging*, 42(3):429-437, 2014.
- [17] M. Hatt, M. Majdoub, M. Vallieres et al. 18F-FDG PET Uptake Characterization Through Texture Analysis: Investigating the Complementary Nature of Heterogeneity and Functional Tumor Volume in a multi-Cancer Site Patient Cohort. *J Nucl Med Imaging*, 56(1):38-44, 2015.
- [18] M. Amadasun, R. King. Textural Features Corresponding to Textural Properties. *IEEE Trans Syst Man Cybern*, 19(5):1264-1273, 1989.
- [19] F. Orlhac, M. Soussan, K. Martinod et al. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. *Plos One*, doi:10.1371/journal.pone.0145063, 2015.
- [20] L. Breiman. Random Forest. *Mach Learn*, 45(1):5-35, 2001.
- [21] JH. Holland. Adaptation in natural and artificial systems. *MIT Press*, Cambridge, MA, USA, 1992.
- [22] C. Spearman. The proof and measurement of association between two things. *Amer J Psychol*, 15:72-101, 1904.
- [23] C. Gini. Measurement of inequality of income. *Economic Journal*, 31:22-43, 1921.