Prognostic and Predictive Feature Selection for Oesophageal Cancer Using Random Forest Classifier

Desbordes, P.[1,2], Modzelewski, R.[1,3], Ruan, S.[1], Pineau, P.[2], Vauclin, S.[2], Vera, P.[1,3], and Gardin, I.[1,3]

[1] Litis - QuantIF - EA4108, University of Rouen, France
[2] Dosisoft, Cachan, France;
[3] Henri Becquerel Center, Rouen, France

Abstract. Prediction of cancer evolution is an important issue for adaptation and evaluation of a therapy. Many features can be extracted from PET images to describe cancer. However which ones are relevant for prediction? In this paper, we propose a feature selection strategy based on random forest to select features having a prognostic or predictive value among a large amount of different kinds of characteristics. Our method is performed in 3 steps. First, a Spearman rank correlation is carried out to keep uncorrelated features. Then, a random forest algorithm is applied to find the most relevant subset of features. Our method is evaluated on a PET database of 66 patients with an oesophageal cancer using two classifiers: support vector machine and random forest. Results show an improvement of the classification accuracy in a range of [4.2 to 16.5%] compared to when using all features without selection. These results are compared with a classical feature selection method (SVM-SFFS) showing that our method gives better results than SVM-SFFS.

Keywords Feature Selection · Outcome prediction · Oesophageal cancer · Random Forest · ^{18}FDG PET

1 Introduction

^{18}FDG PET (Positron Emission Tomography with 18-fluorodeoxyglucose) normalized by Standard Uptake Value (SUV), is now widely used for diagnosing, staging and monitoring response to therapy in oncology [1]. Predictive and prognostic studies using image features derived usually from first order statistics, such as MTV (Metabolic Tumour Volume) or TLG (Total Lesion Glycolysis), have also been carried out. In solid tumours, predictive and prognostic values were found for these features [2].

More recently, for describing ^{18}FDG uptake heterogeneity within the lesion other features have been proposed. For instance, Bundschuh et al. have found that the COV (Coefficient Of Variation) is an important predictive factor in patients with rectal cancer, better than conventional features [3]. El Naqa et

al. have proposed to extract features from the SVH (SUV-Volume Histogram [4]), such as SUV_x (minimum SUV of the x% highest SUV) and V_x (percentage volume having at least x% of SUV). Furthermore in this paper, El Naqa et al. found that features extracted from the GLCM (Gray-Level Cooccurrence Matrix [5]), characterizing the intensity relationships between pairs of neighbouring pixels, are some of the most important predictive characteristics in cervix cancer. Other texture matrices have also been proposed in the literature, such as the GLDM (Gray Level Difference Matrix [6]) characterizing the intensity differences between neighbours, the GLRLM (Gray Level Run Length Matrix [7]) and the GLSZM (Gray Level Size Zone Matrix [8]) characterizing the size ranges of intensities in a direction or in all the directions respectively. At the end, it is possible to extract several features by matrix leading to a large number of characteristics. As far as we know, Tixier et al. [9] are the only ones who studied the importance of a large number of PET image features on oesophageal cancer using ROC (Receiver Operating Characteristic) curves measuring associated AUC (Areas Under the ROC Curves), including texture features. Among 38 features they found that GLCM features (second angular moment, local contrast, entropy, correlation, homogeneity and dissimilarity) and GLSZM features (ZLNU, GLNUz) are relevant to predict patients' response to CRT (Chemo RadioTherapy).

Using all the extractable characteristics from images does not improve accuracy, but may be responsible for information redundancy disturbing classifier. Orlhac et al. [13] have shown that some texture characteristics are highly correlated with MTV on three types of tumour. In the same way, Tixier et al. [9] have shown that GLRLM features are highly correlated with GLSZM ones, and so, do not bring complementary information. Usually, classical statistics are used to assess the importance of each feature for the prediction [2]. Because of the high number of studied characteristics and the nonlinear pattern relationships between features these mathematical tools are not powerful enough. Another possibility is to use a classifier to identify to which set of categories (i.e. responder or non-responder) a new patient belongs. In this context, methods using machine learning could be of great interest due to their ability to manage many features and to capture nonlinear pattern relationships as well as providing a better discriminant power than classical statistics when analysing several tens of features [4].

To date, many feature selection methods have been developed. Globally, three types of feature selection methods can be distinguished: "filter", for which a model design method is not needed; "wrapper" methods, which include the prediction performance in the score calculation; and, finally, "embedded" methods which combine variable selection and model estimation more closely [10].

Although classification using Random Forest algorithm (RF) has shown its good performances [11] [12], it has not been yet used in the context of PET

imaging as far as we know. In this paper, we propose a feature selection strategy based on the RF to find prognostic and predictive values of initial PET features in patients after CRT for oesophageal cancer on a database of 66 patients. This feature selection strategy is studied using two classifiers' performances: SVM (Support Vector Machine) and RF. This paper is organized as follows: Section 2 introduces the protocol used to extract characteristics and to perform the feature selection strategy. Section 3 presents the experimental results on patients' database followed by a discussion in Section 4.

2 Method

Our approach consists of feature extraction, feature selection and determination of features having a prognostic and/or predictive value.

2.1 Feature Extraction

Forty-seven features (Table 1) were extracted from PET images according to the following workflow. First, the MTV is defined by using a contrast-based adaptive threshold algorithm [14]. Nineteen first order features are extracted from SUV, MTV, TLG, COV, SVH [4] and sphericity [15].
Secondly, twenty-eight texture parameters are extracted from 3 texture matrices - GLCM, GLDM and GLSZM, leading to 12 features for GLCM, 5 for GLDM and 11 for GLSZM. To compute these matrices a linear gray-level resampling is applied on the MTV according to [13] (Eq. 1) - leading to a tumour volume image with 64 gray-levels:

$$R(i) = 64 \times \frac{\mathrm{SUV}(i) - \mathrm{SUV}_{min}}{\mathrm{SUV}_{max} - \mathrm{SUV}_{min}} \tag{1}$$

where $R(i)$ is the new intensity of voxel i after the resampling process. To compute GLCM we used 13 matrices, one for each spatial direction, followed by averaging the values calculated separately in each matrix [16]. Finally, 13 clinical features are extracted from the medical record.

2.2 Feature Selection Strategy

The feature selection strategy is composed of three steps. In the first one, a correlation analysis is done on all 60 features using a Spearman rank correlation. Then, a ranking of the most important features is done using a RF algorithm. Finally, the best subset of features is detected thanks to the RF. Thus, the feature selection method is an embedded one because it combines variable selection and classification.

For the first step, a correlation analysis based on a Spearman rank correlation is done in order to keep uncorrelated features. Features are compared one by one and are considered as significantly correlated if the absolute value

Table 1. List of the initial tumour features

Kind of features	Characteristics
Clinical	Age, Gender, Albumin level (g/l), NRI (Nutritional Risk Index), Malnutrition*, Initial weight (kg), Normal weight (kg), Weight loss (%), Tumour location (up, mid, low), Histology (ADC or SCC), TNM stage, WHO stage, Endoscopic tumour length (cm)
First Order	SUV_{max}, SUV_{mean}, SUV_{peak}, MTV, Sum of SUV, TLG, Standard Deviation (SD), COV, Sphericity, Skewness, Kurtosys, Energy, Entropy, SUV_{10}, SUV_{90}, SUV_{10}-SUV_{90}, V_{10}, V_{90}, V_{10}-V_{90}
Texture**	*GLCM*: Energy, Entropy, Correlation, Dissimilarity, Contrast, Homogeneity, Variance, Mean, Cluster Shade, Cluster Tendency, Intensity, Inverse variance *GLSZM*: 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 *GDLM*: Busyness, Coarseness, Complexity, Contrast, Strength

* absence if NRI > 97.5, average if $83.5 \leq$ NRI ≤ 97.5 and severe if NRI < 83.5

** Mathematical expression of features in Table 1 of supplemental data from [13]

of the Spearman correlation coefficient (sp) is higher than 0.8 and $p - value$ smaller than 0.05 [13]. Correlated features are placed in a group verifying these parameters. The mean $|sp|$ value and the associated Standard Deviation (SD) were calculated for each group. In each correlation group the selected feature to represent the group is the one that is the most practical to be calculated in clinical routine. The total number of selected features is limited to the number of groups plus the other uncorrelated features.

The second step consists in a selection of the most relevant features among the remaining features using the RF algorithm. RF is a machine learning technique developed by Breiman [17] to classify data using a set of decision trees. A multitude (k) of trees is built from an initial sample corresponding to N patients with F studied features, represented by a matrix of size (N, F). For each tree node f features are randomly pulled among the F features (f is equal to rounded $\sqrt{F}$, [11]). One of RF intrinsic properties allows assessing the importance of each feature thanks to the computation of the OOB (Out Of Bag) error (%). Here, the OOB^t_{err} is the misclassification rate using a test sample OOB^t (data not

included in the boostrap sample used to construct a t tree). Then, a X feature is randomly permutated in OOB^t to get a perturbed sample denoted by ($\widetilde{OOB^t}$ to compute $\widetilde{OOB^t_{err}}$, the error of the t tree with the perturbed sample. The score of importance (C_i) of the X feature is then equal to:

$$C_i = \frac{1}{k}\sum_t (\widetilde{OOB^t_{err}} - OOB^t_{err}) \tag{2}$$

where the sum is over all t trees of the forest and k denotes the number of RF trees. Greater the error is, more important the feature is. This step is performed for all features in order to obtain a ranking of features.

Finally, classification using RF is performed recursively with an increasing F number of features, in order to define the best prognostic and predictive subset. In this paper, only results with F varying from 1 to 10 are studied. Those with $F > 10$ are not considered as relevant. For each iteration a new feature is added according to the feature ranking obtained previously.
To assess the classification error of the RF algorithm a K-fold Cross Validation protocol is used, with $K = 5$. The number and the list of features leading to the best classification (i.e. lower error) are found this way.

2.3 Classification

Two classifiers are used to evaluate the performances of the feature selection strategy: RF and SVM. SVM [18] is a supervised method which constructs a hyperplane by maximizing distance to the nearest training data points of any class. A Gaussian RBF (Radial Basis Function) is chosen as kernel function [19]. The RBF kernel has 2 parameters to be determined in order to optimize the classification performances: the kernel parameter σ and the error punishment parameter c. Different values of these parameters were tested to find the most suitable. A K-fold Cross Validation protocol with $K = 5$ is used again to assess performance of the classifiers (RF and SVM).

3 Results

3.1 Tumour Features

In this retrospective study, data from 66 patients with an oesophageal cancer eligible for CRT are used to test the feature selection strategy. All the pre-treatment FDG PET images come from the same Nuclear Medicine Department on a Biograph Sensation 16 device (Siemens, Erlangen, Germany). PET images are acquired from the base of the skull to the proximal thighs after a dose injection of 5 MBq/kg of ^{18}FDG, 60 minutes before the exam. Images are reconstructed with attenuation correction using the CT-derived data and attenuation-weighted ordered-subsets expectation maximization (AW-OSEM) algorithm. The voxel

size is 4 x 4 x 2 mm^3. To assess therapeutic response of patients the gold standard was based on the entire set of clinical, endoscopic, histological, and imaging data from the follow up. Complete metabolic response was detected in 41 patients (62%), whereas the histological proof of relapse was obtained in 25 patients (38%). These data are used for the predictive study. The overall survival, used for the prognostic study, was estimated after a follow-up of one year after the end of the CRT. Sixteen patients died (24%) and 50 were alive (76%). The mean MTV was 542.9 ± 629.7 cm^3 (76 – 4272 cm^3) and the mean SUV_{max} was 12.8 ± 9.2 (1.4 - 73).

3.2 Correlation

Results of the Spearman analysis are given in Table 2. Concerning clinical data, normal weight and current weight are correlated ($|sp| > 0.96$). Likewise for the albumin level, the NRI and malnutrition which are highly correlated ($|sp| > 0.84$). This last result can be explained by the fact that the NRI and malnutrition are derived from the albumin level. Nevertheless, none of the clinical data are correlated with the studied image characteristics. Only few image features are not correlated: sphericity, COV, cluster shade (GLCM), ZP (GLZLM) and V_{10}. At the end of this correlation study, 12 groups of significant correlated features ($|sp| \geq 0.8$, $p < 0.05$) were identified leading to a first selection of 30/60 features (in bold in Table 2).

Table 2. Groups of correlated features with the mean absolute value of the Spearman correlation coefficient per group and associated SD. The feature of each group selected for the next step is in bold

Group	Correlated indices	Spearman coefficient absolute value
1	**Current weight** - Usual weight	0.96
2	**Albumin level** - NRI - Malnutrition	0.90 ± 0.06
3	**V_{90}** - V_{10}-V_{90}	0.99
4	**SUV_{max}** - SUV_{10} - SUV_{peak} - SUV_{mean} - SD (1^{st} order) - SUV_{90} - SUV_{10}-SUV_{90}	0.94 ± 0.04
5	**MTV** - Entropy (GLCM) - Energy (GLCM)- ZLNUz - Correlation (GLCM) - GLNUz - $_{sum}$SUV - TLG - Complexity - Contrast (GLDM) - Strength	0.93 ± 0.04
6	**Skewness** - Kurtosys - Mean - Intensity	0.90 ± 0.05
7	**Variance** - Cluster Tendency	0.81
8	**SZE** - LZE	0.89
9	**HGZE** - SZHGE	0.94
10	**LGZE** - SZLGE - LZLGE	0.90 ± 0.08
11	**Energy (1^{st} order)** - Entropy (1^{st} order) - LZHGE	0.92 ± 0.06
12	**Homogeneity** - Dissimilarity - Inverse Variance - Contrast	0.95 ± 0.03

3.3 Feature Selection by Random Forest

The number of trees (k) is initialized to 500 to build the RF algorithm, higher values than 500 have been also tested and no significant difference was observed. An initial matrix with 66 rows (N, patients) and 60 columns (F, features) is built. Each tree of the forest is created based on a bootstrap of 66 patients randomly picked with replacement, and for each node of this tree the best feature is selected among 8 features picked randomly among F ($f = \sqrt{F}$). After the second step, rankings of the 10 most relevant features of the prognostic and predictive studies given by the random forest analysis are presented (Table 3).
According to the RF classification, the minimum prognostic error is obtained using the 3 best features of the ranking: cluster shade (GLCM), homogeneity (GLCM, group 12) and ZP (GLSZM) leading to an error equal to 19.8%. The minimum predictive error is equal to 23.1% and is obtained with the 4 best features of the ranking: ZP (GLSZM), MTV, weight loss and energy (1^{st} order, group 11).

Table 3. Rankings of the most important prognostic and predictive features using RF*

	Prognostic features	Predictive features
1	Cluster Shade (GLCM)	ZP (GLSZM)
2	Homogeneity (GLCM) (grp 12)	MTV (grp 5)
3	ZP (GLSZM)	Weight loss
4	HGZE (GLSZM) (grp 9)	Energy
5	N stage	Lower third location
6	MTV	Homogeneity (GLCM) (grp 12)
7	Energy (1^{st} order)	WHO stage
8	Albumin level	M stage
9	Histology	Weight
10	Weight loss	Tumour length

* correlation group is indicated if necessary

3.4 Classification Results

Results of classifications obtained with RF and SVM are given in Table 4. Performances of these classifications are assessed after each step of the feature selection: step 0 before any selection, step 1 after the correlation analysis and step 2 at the end of the selection using the RF algorithm. The values of parameters σ and c optimizing SVM algorithm are also given in Table 4. The 2 best classification results are obtained after the feature selection strategy for RF and SVM; respectively. Among these classifiers, it is SVM which gives the best final result. For the prognostic study, the corresponding minimum error is equal to 13.7% with $\sigma = 0.5$ and $c = 0.5$ (3 features: cluster shade (GLCM), homogeneity (GLCM)

and ZP (GLSZM)). The minimum predictive error is equal to 20% with $\sigma = 16$ and $c = 64$ (4 features: ZP (GLSZM), MTV, weight loss and energy (1^{st} order)).

To evaluate our feature selection strategy, our results are compared with those obtained with the wrapper method called SFFS (Sequential Forward Floating Selection [20]). This method considers the feature selection as a search problem, where the selection criterion is based on the classification performance. SVM is used as a classifier with a K-fold cross validation ($K = 10$). The SVM-SFFS method selected ZP (GLSZM), cluster shade (GLCM) and weight loss as the 3 best prognostic features. A SVM classification done with these features leads to a classification error of 25.7%. In the same way for the predictive study, the 4 features selected by the SVM-SFS are ZP (GLSZM), albumin level, MTV and weight loss, leading to a classification error of 24.2%. Compared to the SVM-SFFS method, our method gives better results.

Table 4. Prediction errors with two classifiers (SVM and RF) according to the feature selection step

Method \ Characteristics n		Step 0 $n = 60$	Step 1 $n = 30$	Step 2 $n < 10$	Improvement Step 2 - Step 0
Prognostic	RF	36.3%	27.4%	19.8% ($n = 3$)	-16.5%
	SVM*	24.2% $\sigma = 1, c = 1$	24.2% $\sigma = 1, c = 1$	13.7% ($n = 3$) $\sigma = 0.5, c = 0.5$	-10.5%
Predictive	RF	34.7%	27.3%	23.1% ($n = 4$)	-11.6%
	SVM*	24.2% $\sigma = 1, c = 1$	24.2% $\sigma = 1, c = 1$	20% ($n = 4$) $\sigma = 16, c = 64$	-4.2%

* σ and c parameters are those optimizing SVM

4 Discussion

A feature selection strategy in three steps is proposed, showing an improvement of the prediction of PET image features through 2 classifiers (SVM and RF). To our knowledge, this work is one of the first using machine learning to find predictive and prognostic features from FDG PET images. It is now well established that first order features extracted from FDG PET images (SUV, MTV, TLG) have a predictive and a prognostic value [2]. These features are clinically used to choose or to modify treatment. More recently, it has been shown that texture features could also be interesting [4] even if some issues still remain. A high number of texture characteristics can be extracted from PET images but they are not all relevant. Indeed, some of them are correlated [13] and it has been shown that relevant features depend on the type of tumour [16]. Furthermore, in oncology it is difficult to have an important patient database in order to obtain

robust knowledge. Therefore, feature selection strategy is a prerequisite to find predictive and prognostic features among several tens.

At first, a Spearman correlation analysis is done in order to keep uncorrelated features [13]. Tixier et al. [9] have shown on an oesophageal cancer database that GLRLM features are highly correlated with GLSZM ones. Therefore GLRLM features are not integrated in this study in order to limit the number of features. The results shown in Table 2 are similar with those in [13] on three other types of cancer and the created correlation groups are almost the same. These results tend to show that correlations are similar for all the type of tumour. Nevertheless, at first it has to be verified for each new type of cancer. These correlations can explain different results found in the literature. For instance, energy and entropy (1^{st} order) are correlated ($|sp| = 0.92$) leading to similar outcomes in [4]. Using ROC curves and measuring associated AUC, El Naqa et al. have found that energy and entropy have respectively an AUC of 0.50 and 0.53 for head & neck cancer, and 0.72 and 0.65 for cervical cancer. Table 4 shows that the reduction to 30/60 features after the first step improves the classification performances. Indeed, as redundant data are eliminated, the information available in the 60 initial features is denoised.

After the second step of the feature selection, we notice that a subset of 3 texture characteristics (cluster shade, homogeneity (groups 12) and ZP) give the best prognostic performances. It shows that image features bring important information for the evaluation of the disease. The same conclusions with the predictive study can also be obtained. We notice that one texture feature (ZP) and two first order features (MTV and energy) are included among the 4 features used for the classification. They are associated with one clinical characteristic (weight loss). The feature selection impact on the improvement of the classification performances is higher with RF (-16.5% for the prognostic study and -11.6% for the predictive) than with SVM (-10.5% and -4.2% respectively, see Table 4). This result could be explained by the fact that the same algorithm is used during the feature selection strategy and the RF classification while it is not the case with SVM. Nevertheless, the best performances are obtained using SVM after applying the feature selection strategy.

Only few results were published concerning prognostic features in oesophageal cancer. Classical 1^{st} order features (SUV, MTV, TLG) have been studied by Hatt et al. [21]. They have found that MTV, TLG and tumour length have a prognostic value using Kaplan-Meier survival methods and ROC curves. With our strategy, MTV is highly correlated to TLG (group 5). This feature appears at the 6^{th} place in the ranking and was not used during the classification due to the fact that best results were found using only 3 features. In [16], Hatt et al. have studied texture parameters: entropy and dissimilarity from the GLCM, ZP and LZHGE from the GLSZM. They have found that only the dissimilarity (GLCM) is relevant. From our study, this feature (from group 12 represented by

homogeneity) is also one of the most relevant (2^{nd} place in the ranking, see Table 3). On the other hand, we have found that ZP (GLSZM) is the 3^{rd} most relevant feature while it was not found as relevant in [16]. These result discrepancies can be attributed to the difference of patient populations between both studies.
Concerning best predictive features, in [2], Van de Wiele et al. have found that MTV and classical 1^{st} order features are relevant. Similar conclusions are obtained in our results where MTV ranks 2^{nd} place, but the importance of the SUV does not appear in our study. This difference could be due to the presence of more relevant features in our analysis, such as texture parameters. In [9], Tixier et al. found that among 38 texture features studied using AUC measuring on ROC curves, a high predictive value was found using the GLCM features (second angular moment, local contrast, entropy, correlation, homogeneity and dissimilarity). They found similar conclusions concerning the variability in the size (ZLNU) and the intensity (GLNUz) of identified homogeneous tumour zones from GLSZM. In our study, ZLNU and GLNUz are correlated with the MTV (group 5). This group was statistically significant to predict therapy response (2^{nd} place in the ranking), but we do not detect a particular predictive importance for GLCM features. Group 12, with homogeneity, dissimilarity and contrast, is only 6^{th} place in the predictive ranking.

The comparison of our results with those obtained by a well-recognized method (SVM-SFSS, [20]) shows that our feature selection strategy is more efficient. Indeed, if we look at SVM results our feature selection strategy gives better results than the SFFS one: prognostic error of 13.7% versus 25.7%, and 20% versus 24.2% for the predictive error. Utilization of the RF in our method, gives a resistance to noise of training data but the Spearman correlation, studying the features 2 by 2, can be tiresome. To face this drawback, a development of an automation of the 3 steps of our strategy could be a solution.

5 Conclusion

The use of machine learning is helpful in studying combinations of PET image features and clinical data. We have demonstrated that a feature selection strategy can improve classification performances in a range of [4.2 to 16.5%] compared to those using all features without selection.

One of the problems in the medical field is that usually only a small number of data are available, making classifier training difficult. The proposed feature selection strategy presents a real improvement by using RF intrinsic property. Indeed, the capacity to create a ranking of studied features in order to know the importance of each feature in the classification is a good advantage versus classical statistical methods. In the future, our method will be tested on a larger cohort and on different cancers such as lung cancer and lymphoma.

References

1. Czernin, J., Allen-Auerbach, M., Schelbert, H.R.: Improvements in Cancer Staging with PET/CT: Literature-Based Evidence as of September 2006. J. Nucl. Med. 48, Suppl 1, 78S–88S (2007)
2. Van de Wiele, C., Kruse, V., Smeets, P., 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)
3. Bunduschuh, R., Dinges, J., Neumann, L., 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)
4. El Naqa, I., Grisby, P.W. Apte, A., et al.: Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern. Recognit. 42, 6, 1162–1171 (2009)
5. Haralick, R., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. (1973)
6. Amadasun, M., King, R.: Textural Features Corresponding to Textural Properties. IEEE Trans Syst Man Cybern. 19, 5, 1264–1273 (1989)
7. Galloway, M.: Texture Analysis Using Gray Level Run Lengths. Comput. Graph. Image Process. 4, 2, 172–179 (1975)
8. Thibault, G., Fertil, B., Navarro, C., et al.: Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. Pattern. Recognit. Inf. Process. 140–145 (2009)
9. Tixier, F., Cheze-le Rest, C., Hatt, M., 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)
10. Chandrashekar, G., Sahin., F.: A Survey on Feature Selection Methods. Computers & Electrical Engineering. 40, 1, 16–28 (2013)
11. Genuer, R., Poggi, J.M., Tuleau-Malot, C.: Variable Selection Using Random Forests. Pattern. Recognit. Lett. 31, 14, 2225–2236 (2010)
12. Genuer, R., Michel, V., Eger, E. et al.: Random Forests Based Feature Selection for Decoding fMRI Data. Proceed. of the 19th COMPSTAT. 1071–1078 (2010)
13. Orlhac, F., Soussan, M., Maisonobe, J.A., 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)
14. Vauclin, S., Doyeux, K., Hapdey, S. 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. Hofheinz, F., Lougovski, A., Zphel, K., 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)
16. Hatt, M., Majdoub, M., Vallieres, M., 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)
17. Breiman, L. Random Forest. Mach. Learn. 45, 1, 5–35 (2001)

18. Cortes, C., Vapnik, V.: Support-Vector Networks. Mach. Learn. 20, 3, 273–297 (1995)
19. Gao, X., Xue, Z., Li, Y., et al.: The Method and Efficacy of Support Vector Machine Classifiers Based on Texture Features and Multi-Resolution Histogram From 18F-FDG PET-CT Images for the Evaluation of Mediastinal Lymph Nodes in Patients with Lung Cancer. Eur. J. Radiol. 84, 2, 312–317 (2015)
20. Pudil., P., Novovicova., J., Kittler., J.: Floating Search Methods in Feature Selection. Pattern Recognition Letters. 15, 11, 1119-1125 (1994)
21. Hatt, M., Visvikis, D., Pradier, O., et al. Baseline 18F-FDG PET Image-Derived Parameters for Therapy Response Prediction in Oesophageal Cancer. Eur. J. Nucl. Med. Mol. Imaging. 38, 9, 1595–1606 (2011)