Published in Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montréal, Canada, 20-24 July 2020, which should be cited to refer to this work.
DOI: 10.1109/EMBC44109.2020.9176724

Comparison of feature selection in radiomics for the prediction of overall survival after radiotherapy for hepatocellular carcinoma

Pierre Fontaine*†, François-Georges Riet*, Joel Castelli*, Khemara Gnep*,
Adrien Depeursinge†‡, Renaud De Crevoisier*, Oscar Acosta*
*Univ Rennes, CLCC Eugne Marquis, INSERM, LTSI - UMR 1099,F-35000 Rennes, France
†University of Applied Sciences Western Switzerland (HES-SO), TechnoArk 3, CH-3960 Sierre, Switzerland
‡Nuclear Medicine and Molecular Imaging Department, Lausanne University Hospital (CHUV), Lausanne, Switzerland

***Abstract*—Hepatocellular carcinoma (HCC) is the sixth more frequent cancer worldwide. This type of cancer has a poor overall survival rate mainly due to underlying cirrhosis and risk of recurrence outside the treated lesion. Quantitative imaging within a radiomics workflow may help assessing the probability of survival and potentially may allow tailoring personalized treatments. In radiomics a large amount of features can be extracted, which may be correlated across a population and very often can be surrogates of the same physiopathology. This issues are more pronounced and difficult to tackle with imbalanced data. Feature selection strategies are therefore required to extract the most informative with the increased predictive capabilities. In this paper, we compared different unsupervised and supervised strategies for feature selection in presence of imbalanced data and optimize them within a machine learning framework. Multi-parametric Magnetic Resonance Images from 81 individuals (19 deceased) treated with stereotactic body radiation therapy (SBRT) for inoperable (HCC) were analyzed. Pre-selection of a reduced set of features based on Affinity Propagation clustering (non supervised) achieved a significant improvement in AUC compared to other approaches with and without feature pre-selection. By including the synthetic minority over-sampling technique (SMOTE) for imbalanced data and Random Forest classification this workflow emerges as an appealing feature selection strategy for survival prediction within radiomics studies.**

***Index Terms*—Radiomics, Feature selection, Liver cancer, survival prediction, SMOTE, Multi-parametric MRI**

I. Introduction

Hepatocellular carcinoma (HCC) is a major health problem worldwide. This is the sixth more frequent cancer with 749 000 new diagnosed individuals each year. With 695 000 dead patients, HCC represents 7% of all cancers [1]. This cancer has a very low overall survival rate. Although Stereotactic body radiation therapy (SBRT) is not clearly defined in the last recommendations [1], [2] its use in clinical practice is increasing [2], [3]. Despite the high control rates, patients with HCC have a poor overall survival (OS) rate mainly due to underlying cirrhosis and risk of recurrence outside the treated lesion. In order to stratify patients at risk it is of upmost importance to find markers of recurrence and survival after SBRT.

Radiomics allows qualitative and quantitative performance analyses from images to obtain predictive or prognostic information from patients treated for cancer [4]. These approaches require high computation power for processing a large amount of data coming from different centers and imaging protocols. Several issues arise in radiomics, shared by different machine learning methods. These include imbalanced data, which refers to the low number of predicted events with respect to the population but also the large amount of features compared to the reduced sample size [5]. Very often, some of the features are not thoroughly informative and can be redundant across a population. Reduction of dimensionality is therefore an important step before assessing the predictive capabilities of a given model. Supervised methods such as Minimum redundancy and maximum relevancy (mRMR), largely used in radiomics works [6], allow the selection of the most performant features. mRMR requires however a prior classification of patients. However, reducing redundancy by selecting the most contributive features without supervision is still a challenge in radiomics. Different methods such as Principal Componenent Analysis (PCA) is an example of unsupervised method [7]. PCA may be not suited in many cases as it produces a linear combination of features, but not selecting the most informative. Other unsupervised methods include clustering, such as k-means, Hierarchical clustering [8] or Affinity propagation [9], [10].

In this paper we compared different feature selection strategies within a radiomics workflow for the prediction of free disease survival in case of hepatocellular carcinoma from Multi-Parametric Magnetic Resonance Images (MRI). We implemented three unsupervised clustering methods, namely k-means, Hierarchical clustering [8] and affinity propagation [9], [10] and compared with the supervised minimum redundancy and maximum relevancy (mRMR) [11]. Random Forest was used for assessing the prediction capabilities of each one of the methods in a cross-validation scheme [6]. The synthetic minority over-sampling technique (SMOTE) [12] was used to generate synthetic samples in the minority class during the training.

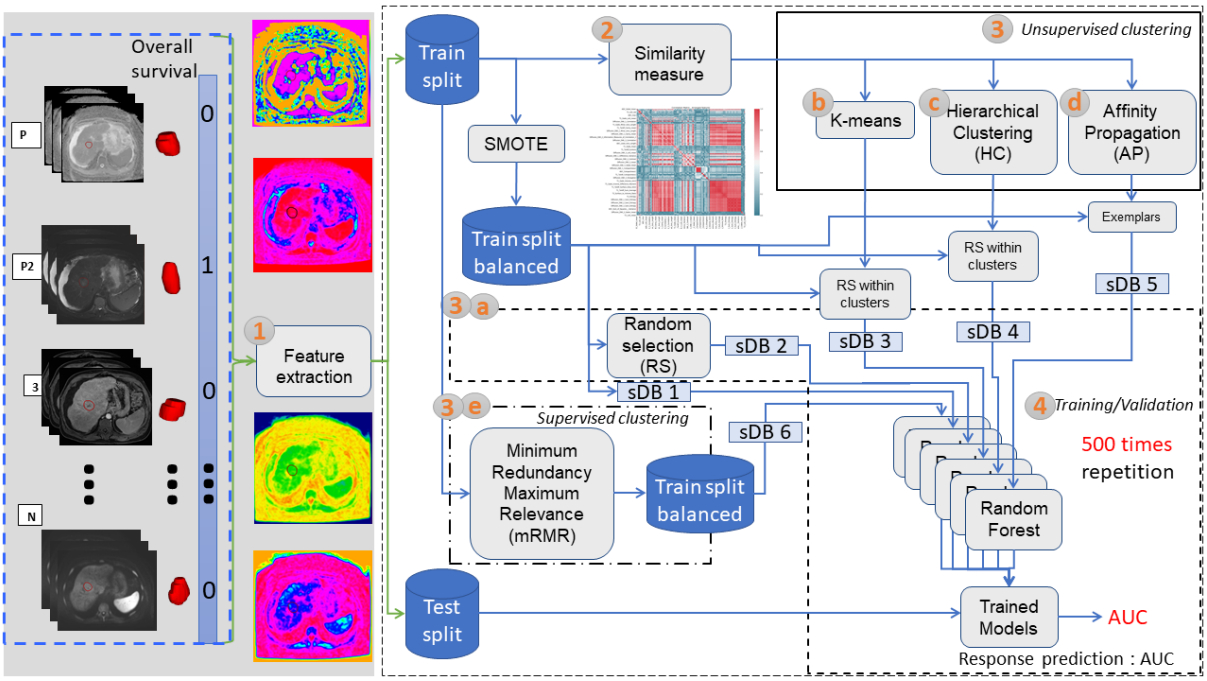

Fig. 1. Workflow of the study. Data were split in training and test. From multi-modal MR Images 273 tumor features were extracted (1). A similarity matrix across the features was computed (2). Different feature selection methods were assessed namely, unsupervised (3b,c,d) and supervised (3e) clustering. Thus sDB1...sDB6 constitute the different sub datasets for comparison after feature selection. SMOTE was applied on the training data set and a random forest classifier (4) was trained and tested 500 times.

II. Methodology

Fig. 1 depicts the devised workflow in our study. Tumor features were extracted from multimodal MR Images (yielding 273 features including geometrical, first order, gradient-based and second order)(1). Data were split into training (80%) and test (20%). The workflow was applied to the training set. Thus, after the computation of similarity across the features (2), different feature selection methods were applied, namely, unsupervised (3) and supervised (4) clustering versus randomly selected features. SMOTE was applied and random forest classifiers were trained with each one of the resulted feature set (sdb1 to sdb6). Finally, the model was used to predict on the test "never seen" dataset. This procedure was repeated 500 times.

A. DATA

Eighty-one consecutive patients, treated in two clinical centers, underwent SBRT for inoperable HCC and were retrospectively analyzed. For each patient 7 MRI modalities namely, T1 enhanced (Gadolinium), T1 late phase, T2-weighted, 3 Diffusion-weighted with different b-value and ADC maps.

B. Feature extraction

The radiomics features were automatically extracted from the tumor, manually delineated by a clinical expert on multi-parametric MRI. We developed the whole feature extraction algorithms in agreement with the radiomics guidelines described in [5]. Thus, 39 quantitative features per modality were obtained belonging to 4 categories as follows:

- Shape of the tumor: Volume, Surface, Surface to area ratio, Sphericity, Compactness, Major/Minor/Least axis length, Elongation, Flatness.
- First order features: Number of voxels, Min, Max, Mean, Std., Skewness, Kurtosis.
- Gradient-based features: Sobel, Canny, Laplacian of Gradient, Gabor.
- Second order features: Haralick's features [13] calculated from the Grey-Level Co-Occurrence Matrix (GLCM), yielding the frequency of co-occuring adjacent voxel pairs intensity.

C. Similarity metric

In order to compute the distance across the features on the population we use Spearman correlation computed as:

$$\rho = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} \quad (1)$$

Where n is the number of observations and D is the difference between the two ranks of each observation.

D. Clustering strategies

1) Unsupervised clustering: Three unsupervised clustering methods were implemented, namely k-means, hierarchical clustering and Affinity Propagation. k-means method is a widely used clustering technique that seeks to minimize

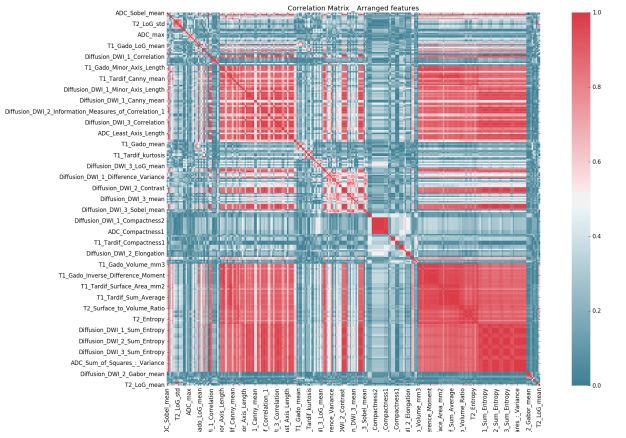

Fig. 2. Similarity Matrix computed from the whole 273 features using the spearman correlation.

the average squared distance between points in the same cluster. In hierarchical clustering, the goal is to group features according to a linkage criterion. In this study a complete-linkage criterion was used:

$$max\{d(a,b) : a \in A, b \in B\} \tag{2}$$

Where a and b are two sets of observations.

The unsupervised algorithm implemented was Affinity Propagation. This algorithm is based on passing messages between data points [10]. It proceeds by alterning two messages passing:

- The responsibility matrix $\boldsymbol{R}$ has values $r(i,k)$ that quantify how well-suited the feature k is to become an exemplar for the cluster i.

$$r(i,k) \leftarrow s(i,k) - max\{a(i,k') + s(i,k')\forall k' \neq k\} \tag{3}$$

- The availability matrix $\boldsymbol{A}$ has values $a(i,k)$ that quantify how well-suited the cluster i could chose the feature k.

$$a(i,k) \leftarrow min\{0, r(k,k) + \sum_{i's.t.i' \notin [i,k]} r(i',k)\} \tag{4}$$

The algorithm performs several iterations until convergence:

$$\begin{cases} r_{t+1}(i,k) = \lambda \cdot r_t(i,k) + (1-\lambda) \cdot r_{t+1}(i,k) \\ a_{t+1}(i,k) = \lambda \cdot a_t(i,k) + (1-\lambda) \cdot a_{t+1}(i,k) \end{cases} \tag{5}$$

The exemplars are extracted from the final matrices as those whose responsibility + availability for themselves is positive as $(r(i,i) + a(i,i)) > 0$.

The number of exemplars is driven by the input preferences value (features with larger values of preferences are more likely to be chosen as exemplars). By default the preferences are set to the median of the feature similarities.

2) Supervised clustering: We used the minimum-redundancy and maximum-relevancy algorithm (mRMR) [11] for selecting features according to the endpoint. This algorithm aims at maximizing the mutual information between a feature x and the cluster c but knowing the output :

$$max\{MI(S,c)\} \quad \text{where} \quad MI = \frac{1}{|S|} \sum_{x_i \in S} I(x_i, c) \tag{6}$$

adding a constrain to minimize the redundancy between intra-cluster features:

$$min\{R(S)\} \quad \text{where} \quad R = \frac{1}{|S^2|} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{7}$$

E. Validation

We fixed the same number of clusters n for all the methods by using the output given by the affinity propagation optimized results ($n = 7$ clusters in our study). Thus, different sub data-bases of patients with reduced number of features were obtained from the training data set. SDB1 is equivalent to original database, the reduced data-sets by clustering were labeled in Fig. 1 as sDB2 to sDB6:

- sDB1 : The whole data set with all the features.
- sDB2 : individuals with n randomly selected features.
- sDB3 : individuals with n randomly selected features from each one of the clusters obtained with k-means clustering.
- sDB4 : individuals with n randomly selected features from each one of the clusters obtained with Hierarchical clustering.
- sDB5 : individuals with n selected features with Affinity Propagation (exemplars).
- sDB6 : individuals with the n best ranked selected features from mRMR.

This procedure of learning and testing was replicated 500 times(Fig. 3). To over-sample the minority class we used SMOTE [12] on the training database. The new synthetic individuals are thus created in the feature space, allowing a more reliable and robust training. For the test step only the original "non seen" individuals were used. The Area under the curve (AUC) was computed to assess the performance of the classifier after using each sub-database.

III. Results & Discussion

Fig.2 depicts the similarity measure matrix obtained across the whole set of features. We optimized the affinity propagation algorithm to obtain the optimal number of exemplars, yielding 7 different clusters. We then adapted the feature selection algorithms to obtain an equivalent number of clusters. We applied the mrMR algorithm obtaining the 7 most relevant features. Figure 3 depicts the Area Under The ROC Curve (AUC) for prediction on the test data sets after using different sub-data (sDB1 to sDB6) for training. As it is shown, the AP algorithm outperforms the other unsupervised feature selection method (p-value < 0.001). This suggests the use of the exemplars as representative feature of each cluster. Moreover, the sensitivity of the AP algorithm is 0.52 and the specificity is 0.71. For the other methods the sensitivity is around 0.26 and the specificity 0.73.

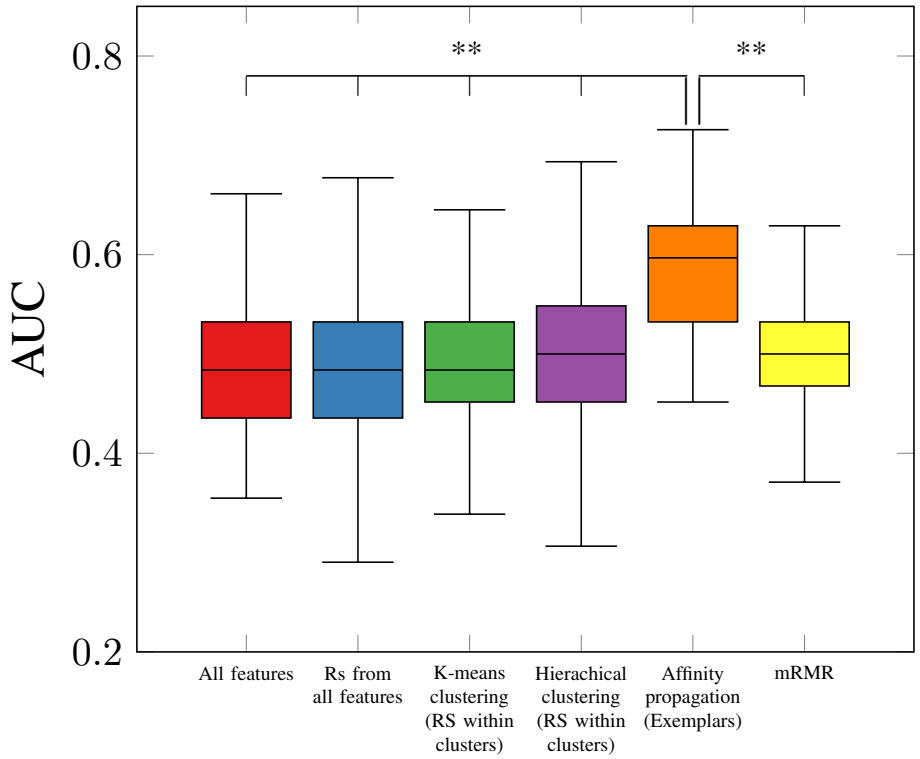

Fig. 3. Results (Area under the curve) of overall survival prediction with SMOTE/Random Forest (RF) workflow after different features selection strategies. All the features (sDB1). $n = 7$ Random Selected (RS) features from the whole data (sDB2), with k-means (sDB3), Hierarchical Clustering (sDB4), Exemplars from Affinity Propagation (sDB5) and mRMR selection (sDB6). The training and validation steps of the RF model was performed 500 times. **p < 0.01

IV. Conclusion

We compared in this paper different supervised and unsupervised feature pre-selection methods within a radiomics workflow, together with compensation for imbalanced data, which achieved improved prediction capabilities for hepatocellular carcinoma overall survival from multi-parametric MRI. Unsupervised clustering using affinity propagation emerged as the best method allowing to select exemplars representative of each class reducing redundancy while increasing the predictive capabilities. The method outperforms not only other clustering methods but also a supervised method based on mRMR. These results showed the interest of an unsupervised feature selection step for dimensionality reduction within the radiomics workflow. This work provides a way forward to larger studies with more patients in order to build robust methods for radiomics applied to the research of imaging biomarkers in different clinical problems. The relatively low AUC compared to the literature could be due to the step of normalization which is not maybe not optimal or even the pre-treatment step. These methods require however larger external validation before being applied in order to build robust models.

References

[1] European Association For The Study Of The Liver et al., "Easl–eortc clinical practice guidelines: management of hepatocellular carcinoma," *Journal of hepatology*, vol. 56, no. 4, pp. 908–943, 2012.

[2] Jonathan Klein and Laura A Dawson, "Hepatocellular carcinoma radiation therapy: review of evidence and future opportunities," *International Journal of Radiation Oncology* Biology* Physics*, vol. 87, no. 1, pp. 22–32, 2013.

[3] Wen-Yen Huang, et al., "Stereotactic body radiation therapy in recurrent hepatocellular carcinoma," *International Journal of Radiation Oncology* Biology* Physics*, vol. 84, no. 2, pp. 355–361, 2012.

[4] Robert J Gillies, et al., "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2015.

[5] Alex Zwanenburg, , et al., "Image biomarker standardisation initiative," *arXiv preprint arXiv:1612.07003*, 2016.

[6] Chintan Parmar, et al., "Machine learning methods for quantitative radiomic biomarkers," *Scientific reports*, vol. 5, pp. 13087, 2015.

[7] Xi Zhang, et al., "Radiomics assessment of bladder cancer grade using texture features from diffusion-weighted imaging," *Journal of Magnetic Resonance Imaging*, vol. 46, no. 5, pp. 1281–1288, 2017.

[8] Chintan Parmar, et al., "Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer," *Scientific reports*, vol. 5, pp. 11044, 2015.

[9] Ryan Clay, et al., "Computer-aided nodule assessment and risk yield (canary) may facilitate non-invasive prediction of egfr mutation status in lung adenocarcinomas," *Scientific reports*, vol. 7, no. 1, pp. 17620, 2017.

[10] Brendan J Frey and Delbert Dueck, "Clustering by passing messages between data points," *science*, vol. 315, no. 5814, pp. 972–976, 2007.

[11] Hanchuan Peng, Fuhui Long, and Chris Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[12] Nitesh V Chawla, et al., ¡ "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[13] Robert M Haralick, Karthikeyan Shanmugam, et al., "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, , no. 6, pp. 610–621, 1973.