Prediction performance of radiomic features when obtained using an object detection framework

Hamza Chegraoui, Amine Rebei, Cathy Philippe, Vincent Frouin

▶ **To cite this version:**

Hamza Chegraoui, Amine Rebei, Cathy Philippe, Vincent Frouin. Prediction performance of radiomic features when obtained using an object detection framework. IEEE ISBI 2021 - International Symposium on Biomedical Imaging, Apr 2021, Nice, France. cea-03162316

HAL Id: cea-03162316
https://cea.hal.science/cea-03162316

Submitted on 8 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PREDICTION PERFORMANCE OF RADIOMIC FEATURES WHEN OBTAINED USING AN OBJECT DETECTION FRAMEWORK

Hamza Chegraoui, Amine Rebei, Cathy Philippe, Vincent Frouin

Université Paris-Saclay, CEA, Neurospin, 91191, Gif-sur-Yvette, France

ABSTRACT

Radiomic features analysis is a non invasive method for disease profiling. In the case of brain tumour studies, the quality of these features depends on the quality of tumour segmentation. However, these segmentations are not available for most cohorts. One way to address this issue is using object detection frameworks to automatically extract the area where the tumour is located in. The purpose of this study is to compare the quality of bounding-boxes based radiomics with manual segmentation, with regards to their performance in patient stratification and survival prediction.

1. INTRODUCTION

Radiomics is the extraction and analysis of high-dimensional features from clinical images. These features have the potential to characterise diseases, and improve diagnosis and prognosis [1].

Radiomics have been successfully used for the stratification and survival prediction of patients with brain tumours (and specially glioblastomas) [2]. One key step in radiomic analysis is the delineation of the tumoral volume of interest from which radiomic features will be extracted. However, the delineation of the tumour is not available in clinical routine care as well as in most clinical trial cohorts since it requires tedious manual work and thus automatic segmentation is eagerly expected. Most state-of-the-art automatic segmentation methods model tumoral heterogeneity and require the 4 MRI modalities (T1, T1 contrast enhanced, T2 and FLAIR), a condition not systematically verified in most of clinical cohorts. Finally, a pathology-independent and automatic technique of radiomic feature extraction is highly desirable in order to streamline processes and reduce operator biases.

Instead of another refined segmentation, we propose to study an alternate strategy for the tumor delineation issue consisting in finding bounding-boxes around tumours in images, as it is yielded by object detection approaches. Finding these bounding-boxes around the tumours - or tumor detection - is a simpler problem than a detailed segmentation. We hypothesise that reliable information on the tumour still remains in radiomics when extracted from a relatively imprecise box-delineation. Detection methods based on deep-learning techniques have shown wide success on finding multiple objects, of different classes, on a range of natural images. They have been shown to be robust across datasets and tasks, meaning they can be trained on a dataset to detect one type of object and then used, with minimal retraining, on a different dataset to detect other kinds of objects.

In this paper, we aim to assess the viability of using radiomic features based on bounding-boxes (BB) as opposed to segmentation (SEG) of a tumour in the context of glioblastoma multiforme (GBM). For this purpose, we will present and adapt an object detection framework to detect bounding-boxes around the tumours. In a test dataset, BB and SEG radiomic features are compared in regards of their prediction performance in two tasks.

2. MATERIALS AND METHODS

2.1. TCGA-GBM tumor images

The TCGA-GBM dataset [3] was used to assess the performance of the object-detection framework and to conduct the survival analysis. This dataset comprises 102 patients, diagnosed for GBM, each with 4-modality (T1, T1ce, T2, FLAIR) brain MRI volumes. Segmentation has been performed automatically and then manually corrected by a neuro-radiologist expert (`https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q`). Segmentation annotations comprise the gadolinium-enhancing tumour (ET), the peritumoral edema (ED), and the necrotic and non-enhancing tumour core (NCR). For this study, we merged all three annotations into a single one and generated the binary masks.

2.2. Tumor Detection

Image pre-processings: For detection purposes, all MRI images (BRATS and TCGA-GBM) received the same preprocessing as described for the BRATS 2019 dataset [4], i.e. re-oriented to LPS (left-posterior-superior) coordinate system, co-registered rigidly into their respective T1ce volume, interpolated to the (1 mm^3) isotropic resolution and skull-stripped. In order to discard grey level outliers, the images were then min-max normalised using the 5 and 95% percentiles and out-of-range values were capped. These transformations were applied to training and testing images.

Object Detection: The YOLO (You Only Look Once) framework [5] is used to detect objects on natural images. This is done by dividing the image into a $S \times S$ grid. Each grid predicts B bounding-boxes containing the object. Hyper-parameters ($B = 2$, $S = 7$) were used to detect the tumour defined as the union of the three annotations ET, ED and NCR as described above.

In most oncology datasets, images are acquired as part of routine practice and therefore do not present all MRI modalities expected for each patient. Often, the accessible modalities are the T1, T1ce and either T2 or FLAIR, although it is not the case in the currently studied dataset. We wanted to propose a solution to detect regions of interest around tumors applicable to cohorts of rare pathologies that do not have FLAIR. For this reason, for the detection, we decided to work only with the T1, T1ce and T2 and discarded the FLAIR for this step.

The implementation and trained weights of the YOLO network, available from Ultralytics [6] were used. Transfer learning was applied using BRATS dataset to fine tune the model for GBM detection. Because BRATS 2019 contains individuals from the TCGA-GBM cohort, we made sure to exclude these subjects from the object detection training and validation sets. Since YOLO only takes natural images as input, the red channel was fed with the 2D axial T1 images, the green channel with the 2D axial T2 images, and the blue channel with the T1ce 2D axial images.

Image post-processing: Since we used the YOLO model in a 2D setting, the continuity of the detected mask is not guaranteed along the axial direction. To tackle this issue, morphological closing and opening was applied to the 2D detected masks along the axial axis. This post-processing fills the holes and deletes isolated detected zones and produced the BB ROIS.

2.3. Classification and survival prediction with radiomics

The bounding-boxes are detected in the TCGA-GBM images using the finely tuned YOLO network. Using either BB or SEG ROIs, radiomic features are extracted from the TCGA-GBM images. The prediction performances are assessed according to two tasks: predicting patients' overall survival and classifying patients between good/poor prognosis. For the second task, we divided the patients as follows. We computed the median survival and its 95 confidence interval (MSCI). Patients with an overall survival above the upper bound of the MSCI are labelled with positive prognosis (GOOD). In the same way, patients with an overall survival below the lower bound of the MSCI are labelled with negative prognosis (POOR). Patients in MSCI (15 patients) are discarded during feature selection. To assess the performance, the patients were divided in a train and test set left apart. A feature selection is beforehand carried out by a 3 fold cross validation using train = (training+validation). From the 102 initial patients, 13 were removed because of the poor (see thereafter) tumour detection. We compared the SEG and BB approaches on 89 patients for the survival prediction and on 77 for the prognosis prediction. Figure 1 summarises all the steps described in this section.

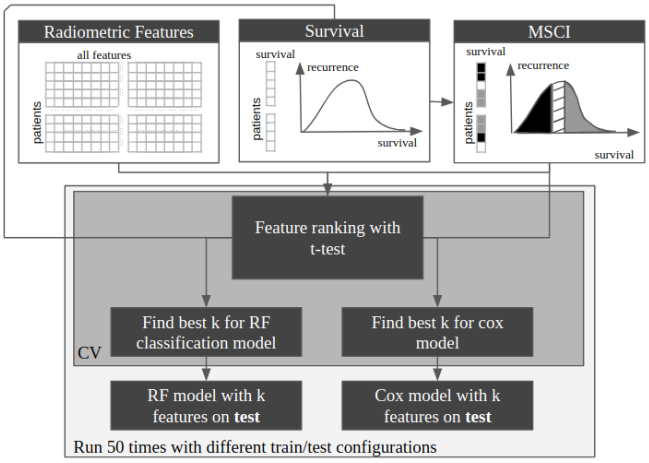


Fig. 1: Steps for features selection and survival prediction

Radiomic features extraction: All native images were subjected to radiomic feature extraction. First, N4 bias correction algorithm was applied. Then, an intensity normalisation was performed using the whitestripe algorithm to standardise intensity ranges for each imaging modality among all subjects [7]. Finally, we used PyRadiomics[8] to extract radiomic features (shape, first and second order) from original and wavelet-transformed images, resampled to a $2mm^3$ resolution, and with a 25 gray level bin size, yielding 3404 radiomic features for each patient.

Features selection: First, highly correlated features and constant features were removed ($\simeq$ 2000 features removed). We select variables sequentially. A variable is discarded if we have already selected a variable correlated to it ($r^2 > 0.95$). A feature is considered constant if $\sigma^2 < 0.0001$. Then, the features were scaled using the z-score. Finally, the k features that best correlate with the task main pattern were selected. A t-statistics was computed for each features to test the null hypothesis $H_0 : \mu_{POOR}(feature) = \mu_{GOOD}(feature)$. The corresponding p-values were used to rank the features. For both the classification and survival prediction task, we use the grid search technique to find the optimal number of ranked features k.

Assessment metrics of the classifier and survival predictor: We used the random forest implemented in Scikit-learn [9], with 1000 estimators and gini function as split criterion, as classifier for the GOOD/POOR prognosis task. Survival prediction was performed using a Cox proportional hazard model implemented in the Python lifelines module [10], with a 0.2 ridge penalty parameter. We scrutinised the

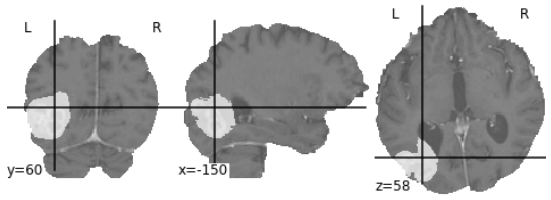
Table 1: Tumour detection results on the TCGA-GBM dataset. IOU : Intersection Over Union

	Recall	Precision	IOU
Mean	$0.88 \pm .14$	$0.62 \pm .1$	$0.69 \pm .12$
Median	0.93	0.64	0.73

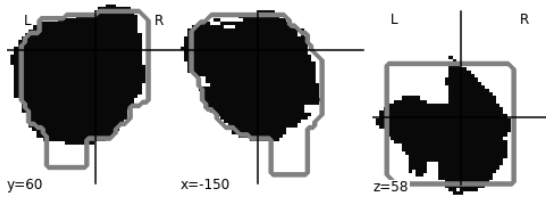
performance achievable by the classifiers when radiomics were generated either from BB or SEG ROIs. In the case of the GOOD/POOR classifier, we used area under the curves (AUC) and the balanced accuracy, defined as the average of recall obtained across the classes. Likewise, we used C-index for the survival predictor.

3. EXPERIMENTS AND RESULTS

To test our approach, we trained YOLO using BRaTS2019 dataset excluding TCGA-GBM images. Starting from YOLO networks and weights available in [6], we performed the transfer learning by training the network on 8211 images extracted from 126 High Grade Glioma volumes. We trained the network during 5 epochs.



(a) Tumour on brain



(b) bounding-box around the segmented tumour

Fig. 2: Examples of tumour detection (grey line) and ground truth (dark area).

Tumor detection results: Figure 2 give an example taken from the TCGA-GBM dataset of the bounding-boxes detected compared to the ground truth segmentation. All results shown correspond to the predicted masks after 3D morphological reconstruction. The results are presented in Table 1. The different metrics used are :

$$Recall = \frac{\#\ of\ tumoral\ voxels\ in\ predicted\ bbox}{\#\ of\ tumoral\ voxels}$$

$$Precision = \frac{\#\ of\ tumoral\ voxels\ in\ predicted\ bbox}{\#\ of\ voxels\ in\ predicted\ bbox}$$

$$IOU = \frac{predicted\ bbox \cap true\ bbox}{predicted\ bbox \cup true\ bbox}$$

We obtained scores aligned with other object detection benchmarks, which indicates our bounding-box predictions were successful. We also observed high recall score across all tumour compartments (0.97, 0.81 and 0.96 for ET, ED and NCR respectively). This means BB radiomic features will report on almost all the tumour. The recall on the ED is lower compared to the other parts of the tumours. This is due to the large size of the oedema compared the the whole tumour (50% of the tumour is oedema) and it is invisible on the T1 and T1ce images. The precision score is relatively low, but it must be put into perspective with the average tumour occupancy of 7% of the brain. Moreover, low precision scores were expected for our bounding-boxes because our regions are cubic while tumour shapes are random meshes. Therefore, radiomic features will also contain a significant amount of peritumoral information. The aim of this paper comes down to determining the impact of these added voxels on radiomic features predictive power.

Prediction results: The dataset was split into 70% training set and 30% testing set. The cross-validation was done using 3 folds within the train set. Moreover, to assess the stability of the results, we ran the whole pipeline 50 times with different train/test configurations.

The compared performance of the predictors with SEG and BB radiomic features was assessed for the classification and survival tasks. Table 2 gives a comparison between the two pathways when followed to classify GOOD and POOR prognosis individuals. Both types of features show stable results, with BB having slightly lower performance (not significant, $p = .05$).

Table 3 summarises results found for survival prediction. The two methods exhibit the same stability across validation and test sets. The results are not significantly different, but they are not very high to draw a final conclusion.

Finally, as our predictor rely on feature selection in a high-dimensional space, the stability of the extracted features set is not guaranteed. On average, we selected 12 features, while the the optimal number of features ranged from 5 to 20 across all the tests. Most selected features are based on co-occuring gray level matrix extracted from spatially filtered images. Both approaches selected variables extracted from the FLAIR. It must be noted that we did not select the same features from BB and SEG. Furthermore, correlation analysis between the two sources of the co-variate shows a strong correlation between most of the intensity based features. Most of selected features exhibits a strong correlation between the two sources. It should be noted that shape radiomic features are never selected either in the bounding-box strategy or in the segmentation strategy and do not seem to be a reliable survival predictor. 0.69

Table 2: GOOD/POOR prognosis classification results on 77 patients with 70% train, 30% test and 3-fold CV

	AUC		Balanced accuracy	
	CV	Test	CV	Test
BB	$0.73 \pm .06$	$0.61 \pm .06$	$0.67 \pm .07$	$0.61 \pm .07$
SEG	$0.79 \pm .05$	$0.62 \pm .09$	$0.76 \pm .04$	$0.62 \pm .08$

Table 3: Survival prediction results on 89 patients with 70% train, 30% test and 3 fold CV

	C-index	
	CV	Test
BB	$0.65 \pm .04$	$0.54 \pm .01$
SEG	$0.69 \pm .03$	$0.57 \pm .06$

4. CONCLUSION

We have shown that a general object detection framework for natural images tuned with transfer-learning can be used to reliably enclose tumours in clinical images. These bounding-boxes may be used as regions of interest for radiomics, but also to facilitate automatic segmentation in a second step. We studied the impact of choosing FLAIR instead of T2 for the fine tuning of the YOLO network. For the TCGA-GBM dataset, using only T1, T1ce and FLAIR images, we obtained very similar bounding-box performances (results not shown). The 3-channel input of the YOLO architecture does not seem to hamper the BB detection in a standard oncological MR dataset, even if this should be further investigated. Our detection approach exhibited insensitivity in regard of tumour size, shape and components, which makes it a reliable method. We think that object detection strategy, which is more robust than segmentation, is a true alternative to provide ROIs with sufficient quality to achieve radiomics analysis in cohorts of rare and poorly documented tumours without further tuning. In this paper, segmentation based radiomics did not show meaningful differences from bounding-boxes features in terms of prediction performance. Both approaches exhibited similar performance for stratification and survival prediction in patients with glioblastoma multiforme. Both approaches displayed commonalities in stability of selected features set and differences in regard of the specific chosen features, which deserve further investigations. With these results, bounding-box features' performance seems to be only slightly below segmentation's one; and we believe they are a viable alternative when segmentation is desired but labels are not available.

5. COMPLIANCE WITH ETHICAL STANDARDS

The present analyses were conducted with data obtained under regular access to the BRaTS & TCIA database which contains Information on the Compliance with ethical standards

6. ACKNOWLEDGMENTS

This work is partially funded by the Collaborative Research Agreement Ref: CJ2020-0306/37598 between Gustave-Roussy Villejuif and CEA Saclay.

7. REFERENCES

[1] H. W. L. Aerts et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach.," *Nature communications*, vol. 5, pp. 4006, 2014.

[2] P. Kickingereder et al., "Radiomic profiling of glioblastoma: Identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models.," *Radiology*, 2016.

[3] S. Bakas et al., "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, no. 1, pp. 170117, 2017.

[4] B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[5] J. Redmon et al., "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, 2015.

[6] G. Jocher et al., "ultralytics/yolov3: 43.1map@0.5:0.95 on coco2014," 2020.

[7] R. T. Shinohara et al., "Statistical normalization techniques for magnetic resonance imaging," *NeuroImage: Clinical*, vol. 6, pp. 9 – 19, 2014.

[8] J. J. M. van Griethuysen et al., "Computational radiomics system to decode the radiographic phenotype," *Cancer Research*, vol. 77, no. 21, pp. e104–e107, 2017.

[9] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[10] C. Davidson-Pilon et al., "Camdavidsonpilon/lifelines: v0.25.5," 2020.