IL NUOVO CIMENTO **46 C** (2023) 142
DOI 10.1393/ncc/i2023-23142-6

Communications: SIF Congress 2022

Robustness and predictivity of MRI-based radiomic features in glioma grade discrimination(*)

L. Ubaldi(1)(2), S. Saponaro(3)(4), A. Giuliano(5), C. Talamonti(1)(2) and A. Retico(3)

(1) *National Institute for Nuclear Physics (INFN), Florence Division - Florence, Italy*
(2) *Department Biomedical Experimental and Clinical Science "Mario Serio", University of Florence - Florence, Italy*
(3) *INFN, Pisa Division - Pisa, Italy*
(4) *Department of Physics, University of Pisa - Pisa, Italy*
(5) *Unit of Medical Physics, Pisa University Hospital "Azienda Ospedaliero-Universitaria Pisana" - Pisa, Italy*

received 3 April 2023

Summary. — Analysis pipelines based on Radiomics are widely used exploration tools in medical imaging. This study aims to define a robust processing pipeline based on the computation of radiomic features on multiparametric Magnetic Resonance Imaging data to make a Machine Learning classification between two diagnostic categories. As a case study, we considered the discrimination between high-grade and low-grade gliomas. The impact of intensity normalization techniques and different settings in image discretization on classification performances was studied. A set of MRI-reliable features was defined by selecting the most appropriate normalization and discretization settings. The results in glioma grade classification showed that the use of MRI-reliable features improves discrimination performances.

1. – Introduction

Nowadays, analysis techniques based on Radiomics and machine learning (ML) offer a great potential in medical imaging research, since they have the capability to derive large amounts of quantitative features from images and to process them to produce clinically meaningful output. One of the main challenges for the clinical applicability of Radiomics is the robustness of the radiomic features [1,2]. There is no consensus nowadays regarding

(*) This article is a short overview of Ubaldi L., Saponaro S., Giuliano A., Talamonti C. and Retico A., *Phys Med.*, **107** (2023) 102538, `https://doi.org/10.1016/j.ejmp.2023.102538` © 2023 Associazione Italiana di Fisica Medica e Sanitaria, published by Elsevier.

the most repeatable, reproducible and robustness features [3]. There are several steps in a typical radiomic workflow where different choices of procedures and parameters can be made, thus affecting the robustness of the extracted features. An international collaboration, the Image Biomarker Standardization Initiative (IBSI) was established with the purpose of standardizing the procedures according to which radiomic features should be defined and extracted [4]. However, feature calculation settings and software versions, which are fundamental aspects of the radiomic workflow [5], are not included in the IBSI. Since normalization [6] and intensity discretization affect the robustness of radiomic features, in radiomic studies it is important to carefully report the steps and settings used to implement the radiomic workflow [7].

We studied these topics in our previous work [8]. In this overview, we describe how different choices in normalization and intensity discretization parameters influence the robustness of radiomic features and impact on their predictive power.

2. – Materials and methods

The data used in this study are a subset of two datasets of multiparametric MRI scans of patients with brain tumor that were made publicly available on The Cancer Imaging Archive (TCIA) [9-11]. We included 102 and 65 scans from the TCIA-GBM and TCIA-LGG collections, respectively. For each subject, we have analyzed the T1-weighted, T2-weighted, FLAIR, T1-Gd modality, considering the entire tumor volume.

A typical radiomic and ML analysis workflow, based on a radiomic feature extraction step followed by a machine learning classification, has been followed in this study, as depicted in fig. 1. In particular, we focused on the image normalization and features extraction steps.

To compare the gray value distributions of images acquired with the same MRI sequence across different subjects, an intensity normalization procedure can be adopted. In our study, we implemented the following three different types of normalization algorithms on the images:

- Norm_MinMax: We rescaled the voxel intensity values between the maximum and minimum gray value;
- Norm_RobustScaler: We rescaled the voxel intensity values by subtracting the median value and scaling data according to the quantile range;
- Norm_Brainstem: For this normalization, we selected a ROI in the brainstem and rescaled the intensity value by subtracting the median value and then dividing

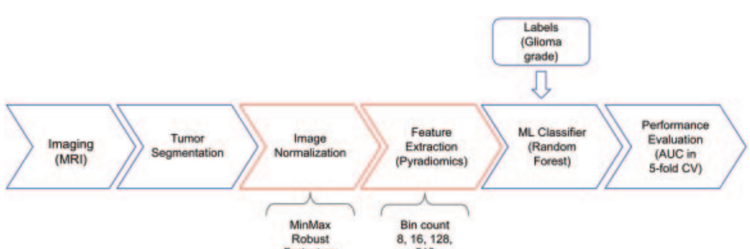


Fig. 1. – A typical radiomic and ML workflow is shown.

by the IQR of the intensity values of the brainstem. We expect the intensity of the brainstem region to be more homogeneous, which we believe makes this normalization more robust.

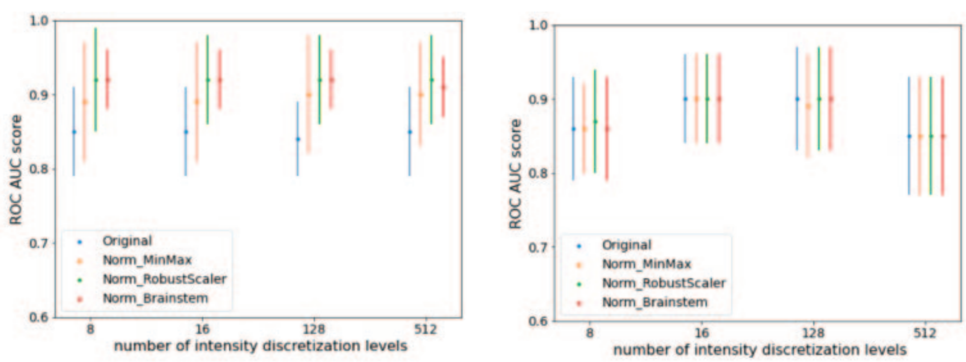
The computation of the radiomic features on the multiparametric MRI images was performed with the open source package *Pyradiomics* (v3.0.1) [12], IBSI compliant. We extracted 93 features consisting of: 18 histogram-based features and 75 texture-based features. The extraction of the texture-based and some of the intensity features requires binning the intensity histogram. We evaluated the influence on the robustness of radiomic features of choosing different total number of bins (8, 16, 128 and 512), which is the parameter that determines the dynamic range of the discretized gray values of the images, as recommended by the IBSI when dealing with non-quantitative data.

We evaluated the predictive power of radiomic features in the categorization between LGG and HGG, by using random forest (RF) classifiers. The binary classification performances have been evaluated across the various image normalization methods and the different settings in the image discretization procedure. The RF model has been trained according to a stratified 5-fold cross-validation (CV) scheme. Results across the 5 test folds were collected to calculate the average AUC and its standard deviation.

The open-source Python package *Pingouin* [13] was used for statistical analysis. The Intraclass Correlation Coefficient (ICC) was considered as a measure of the robustness of the radiomic features. In particular, we selected the two-way mixed effects model, with average raters and absolute agreement [14]. The ICC values range between 0 and 1, with values closer to 1 representing higher robustness [15]. We studied the effect of normalization and the effect of image intensity discretization on the robustness of features.

3. – Results and conclusion

All the normalization techniques implemented improved similarity among the image histograms. However, the *Norm_Brainstem* normalization procedures appear as the best method. The classification performances obtained considering only the intensity features (fig. 2(a)) depend on the normalization technique applied but are almost stable across



(a) Intensity features (b) Texture features

Fig. 2. – Performances achieved in LGG vs HGG discrimination using different numbers of gray levels for intensity discretization and different normalization methods on intensity (a) and on texture (b) features.

the different choices of discretization levels. On the contrary, as shown in fig. 2(b), the performances obtained considering only the texture features are not influenced by the normalization strategy but a variation in the number of intensity discretization levels, for the same normalization, leads to different performances. The trend of the AUC values for the different choices in the number of intensity discretization levels suggests that 16 and 128 levels are good choices for the extraction of informative texture-based radiomic features.

Further, we found out that, when varying the normalization method, the subset of the most robust features (ICC > 0.9) is composed of 16 intensity features and, when varying the image intensity discretization strategy, it is composed of 43 texture features.

The subset of robust features allowed obtaining stable classification performance in the LGG vs HGG discrimination (AUC $= 0.83 \pm 0.08$), regardless of the settings chosen for image normalization and discretization of image intensity levels. Moreover, the performance in glioma grade discrimination is enhanced when the set of features defined as MRI-reliable is used (AUC $= 0.93 \pm 0.05$). Despite the reliability of the performances obtained in the CV evaluation, a limitation of this study could consist in the absence of a validation of the model on external data, which could be introduced as a future improvement. Our results highlight that image normalization and intensity discretization are a fundamental step in MRI analysis via Radiomics and Machine Learning. Due to the strong impact it has on the performance of a ML classifier, special attention should be taken in the image preprocessing step before typical radiomic analysis are performed.

* * *

This work has been carried out within the Artificial Intelligence in Medicine (next_AIM, `https://www.pi.infn.it/aim`) project funded by INFN (CSN5) and within the FAIR-AIM project funded by Tuscany Government (POR FSE 2014-2020).

REFERENCES

[1] Lambin P. *et al.*, *Eur. J. Cancer*, **48** (2012) 441.
[2] Gillies R. *et al.*, *Radiology*, **278** (2015) 151169.
[3] Traverso A. *et al.*, *Int. J. Radiat. Oncol. Biol. Phys.*, **102** (2018) 1143.
[4] Mitchell-Hay R. N. *et al.*, *J. Magn. Reson. Imaging*, **56** (2022) 1559.
[5] Fornacon-Wood I. *et al.*, *Eur. Radiol.*, **30** (2020) 6241.
[6] Saltybaeva N. *et al.*, *Phys. Imaging Radiat. Oncol.*, **22** (2022) 131.
[7] Hoebel K. *et al.*, *Radiol.: Artif. Intell.*, **3** (2020) e190199.
[8] Ubaldi L. *et al.*, *Phys. Med.*, **107** (2023) 102538.
[9] Clark K. W. *et al.*, *J. Digit. Imaging*, **26** (2013) 1045.
[10] Bakas S. *et al.*, *Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection* (Data Set), *The Cancer Imaging Archive*, `https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF` (2017).
[11] Bakas S. *et al.*, *Sci. Data*, **4** (2017) 170117.
[12] Van Griethuysen J. J. *et al.*, *Cancer Res.*, **77** (2017) e104.
[13] Vallat R., *J. Open Source Softw.*, **3** (2018) 1026.
[14] Liljequist D. *et al.*, *PLOS ONE* , **14** (2019) e0219854.
[15] Koo T. K. and Li M. Y., *J. Chiropr. Med.*, **15** (2016) 155.