S. Volpe[1], L.J. Isaksson[1], M. Pepa[1], M. Zaffaroni[1], S. Raimondi[2], G. Lo Presti[2], C. Garibaldi[3], C. Rampinelli[4], G. Marvaso[1], S. Gandini[2], M. Cremonesi[3], B.A. Jereczek-Fossa[1]

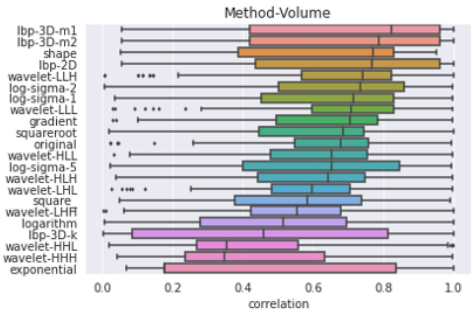
[1]Istituto Europeo di Oncologia IRCCS, Radiation Oncology, Milan, Italy; [2]Istituto Europeo di Oncologia IRCCS, Experimental Oncology, Milan, Italy; [3]Istituto Europeo di Oncologia IRCCS, Radiation Research Unit, Milan, Italy; [4]Istituto Europeo di Oncologia IRCCS, Radiology, Milan, Italy

Purpose or Objective
Radiomics is increasingly used to implement clinically-based prognostic models for non-small cell lung cancer (NSLCL). However, no evidence supports the choice of specific imaging pre-processing methodologies. Admittedly, dedicated investigations could contribute to refining both the reproducibility and the performance of radiomic studies. As the volume of the primary tumor is a well-recognized prognosticator, our purpose is to assess how pre-processing may impact the feature-volume dependency in computed tomography (CT) images of NSCLC patients treated with radiotherapy.

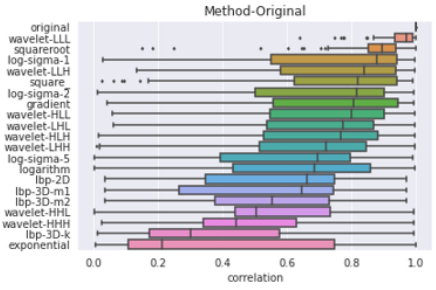
Materials and Methods
Images were retrieved from the publicly available repository NSCLC-Radiomics of The Cancer Imaging Archive (TCIA). Four hundred eighteen images were included in the analysis following manual inspection and editing of the segmentations; nodal disease- if any- was not included. Pyradiomics was used to extract 93 features; which were grouped as follows: first-order, shape-based (3D), shape-based (2D), gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighboring gray tone difference matrix (NGTDM) and gray level dependence matrix (GLDM). Twenty built-in pre-processing methods (filters) were applied, including wavelet and its possible permutations, Laplacian of gaussian, and local binary pattern (LPB); each feature except those belonging to the shape category was computed once per filter, and on the original CT image. The Spearman correlation coefficient (ρ) was used; with thresholds of ≥0.7 and ≤0.5 defining strong and weak correlations, respectively.

Results
Overall, features of the GLCM category were the least correlated with volume (ρ= 0.5). The feature/volume correlation was found to be filter-dependent: the highest correlation was found when lpb-3D-m1 was applied (ρ= 0.82), while the lowest correlations with volume were identified for the HHL and HHH-wavelet filters, and for the exponential method (ρ= 0.35, 0.30 and 0.18, respectively). These results were confirmed when features computed per each pre-processing modality were compared to the original image. An overview of the results is displayed in **Figure 1**.

Method-Volume correlation



Method-original correlation



Conclusion
Our results support the hypothesis that pre-processing does impact features values; and provide a proof of concept that further standardization is warranted for radiomic studies. Further- currently ongoing analyses- will focus on how these findings impact the performance of radiomic-based survival models.

PO-1766 Deep learning classifier for MGMT promoter methylation status in glioblastoma cancer

J. Barranco Garcia[1], D. Abler[2], M. Reyes[3], D. Voung[1], M. Guckenberger[4], S. Tanadini-Lang[4], A. Depeursinge[5]

[1]USZ, Radio Oncology Clinic, Zurich, Switzerland; [2]HES-SO Valais-Wallis, Institute of Information Systems, Sierre, Switzerland; [3]University of Bern, ARTORG Center for Biomedical Research, Bern, Switzerland; [4]USZ, Radio Oncology Clinic , Zurich, Switzerland; [5]HES-SO Valais-Wallis, Institute of Information Systems, Sierre, Switzerland

Purpose or Objective
O^6-methylguanine-DNA-methyltransferase (MGMT) promoter methylation status in glioblastoma cancer is accepted as a promising prognostic and predictive biomarker. We explore the possibility of deep learning algorithms to predict the presence of MGMT status in MRI imaging as a non-invasive method.

Materials and Methods
The RSNA and the MICCAI collaboration provided a dataset composed of 582 patients with four MRI modalities included (T1, T1ce, T2, FLAIR). MGMT status was encoded with 0/1. Out of the 582 patients, 306 were methylated and 276 not. The dataset was divided into training/validation (90%) and test (10%). using a random split. Then training and validation are splitted 80/20. The raw images are pre-processed: a) bias correction, b) normalization z-score and c) cropping and resampling to fit the entire brain (across all patients) into 144x144x144 voxels. To build the classifier we used two publicly available pre-trained image classifiers models to initialize the weights (ResNet50 and DenseNet121). Prior to training the networks, the 2D slice with the largest tumor area is selected in the horizontal view. For the largest tumor size, the surrounding bounding box is calculated and each image is cropped from the center of mass of the mask. This ensures that the tumor surrounding tissue is taken into consideration by the model. To combine the information of the different modalities the RGB channels were replaced with 3 modalities. Different techniques of data augmentation were used to prevent overfitting and improve performance. Affine transformations including horizontal and vertical translations and z-rotations were applied to the input images. The model was evaluated first for each modality independently using 5-fold cross validation.

Results
FLAIR obtained the best performance with DenseNet121 architecture with validation and test accuracies (0.7429,0.5953). We evaluate in groups of 3 modalities obtaining the best performance for the combination of (T1, T1ce and FLAIR) with validation and test accuracies of (0.7124, 0.6245) with the others combinations showing lower but close accuracies. Finally, data augmentation was performed during each epoch leading to similar results with the best combination again (T1, T1ce and FLAIR) and accuracies of (0.7035, 0.6355).

Conclusion
Deep learning classifiers shows promising results to predict the MGMT status in glioblastoma cancer. Combination of different modalities and data augmentation techniques improved the accuracy of the model.

PO-1767 Development of a MRI radiomic-based ML model to predict aggressiveness of prostate cancer

O. Rios Ibacache[1], P. Caprile[2], J. Domínguez[3], C. Besa[4]

[1]Pontificia Universidad Católica de Chile, Facultad de Física, Santiago, Chile; [2]Pontificia Universidad Católica de Chile, Facultad de Física, Santiago, Chile; [3]Pontificia Universidad Católica de Chile, Departamento de Radiología, Facultad de Medicina, Santiago, Chile; [4]ANID, Millennium Science Initiative Program – Millennium Nucleus for Cardiovascular Magnetic Resonance, Santiago, Chile

Purpose or Objective
The gold standard for the evaluation of prostate cancer (PCa) aggressiveness is the Gleason score (GS), which requires a histopathological analysis to discriminate between clinically significant (CS, GS≥ 7) and non-significant (non-CS, GS=6) cases. The aim of this study was to develop a non-invasive tool able to predict the GS classification of PCa, based on the information extracted from multiparametric magnetic resonance imaging (mpMRI), by using machine learning (ML) tools. Additionally, the impact on the model performance of the feature selection method, as well as the inclusion of clinical data and qualitative image information was assessed.

Materials and Methods
This retrospective cohort included 86 adult male patients with positive biopsy for PCa, made by fusion technique (mpMRI-ultrasound) at Hospital Clínico de la Pontificia Universidad Católica de Chile between 2017 and 2021, with lesions greater than 5 mm. 2D segmentations of the target prostate lesions were made by experienced radiologists in T2 weighted (T2w)/Apparent Diffusion Coefficient (ADC) map images at a 3T scanner. A radiomic analysis was performed considering first order, textural and shape features, besides clinical information, including qualitative image information such as PIRADS-v2. Splitting the dataset on train/test (80%) and validation sets (20%), univariate and multivariate models were built using manual and automatic feature selection algorithms. In order to evaluate the performance of the models, twofold cross-validation (CV) was employed with an 80%/20% split for the train/test groups respectively. In particular, we used the Repeated Stratified KFold CV technique with 1000 repetitions, with the Area Under the Curve (AUC) as the evaluation metric. The manual selection method was based on individual feature performance and correlation, using parametric and non-parametric statistical hypothesis tests, Pearson correlation, and predictive power with bootstrap AUC analysis. A comparison between models was performed using Frequentist and Bayesian correlated t-tests.

Results
The best model found was multivariate, obtained using the automatic feature selection algorithm Recursive Feature Elimination (RFE), with Logistic Regression as estimator with nine features, including image (T2w and ADC) and clinical information. The train/test mean AUC was 0.91 (0.06) [0.75−0.99] (p-value<0.05), with a validation AUC of 0.91 for a