

Covid CXR Hackathon Report

Team Name: MedXAI Group - BiND-Unipa

Authors: Prinzi, Francesco ¹; Militello, Carmelo ²; Currieri, Tiziana ¹; Gallo, Alberto ¹; Vitabile, Salvatore ¹

¹ Department of Biomedicine, Neuroscience and Advanced Diagnostics (BiND), University of Palermo, Italy

² Institute of Molecular Bioimaging and Physiology, Italian National Research Council (IBFM-CNR), Palermo, Italy

1. The Processing Pipeline

1.1 Lung ROIs Delineation and Radiomic Features Preprocessing

The delineation of lung regions of interest (ROIs) was performed by means of a manual procedure. CXR image and related segmentation mask were saved in .NIFTI format.

PyRadiomics was used for the extraction of radiomic features. A total of 1023 radiomic features, 107 original features (i.e. belonging to GLCM, GLRLM, GLSZM, GLDM, NGTDM categories) and 916 filtered features, obtained through LoG and wavelet filter applications (with Haar kernel) have been extracted. Moreover, in order to determine the optimal quantization level, different sets of features were extracted, as the 'binWidth' Pyradiomics parameter changes in {8, 16, 32, 64, 128, 256}.

1.1.1 Radiomic Features Selection

The features were skimmed considering the following steps: *i) binning analysis*: the quantization level $binWidth=32$ - maximizing the number of robust features (with $ICC>0.85$) - was chosen; *ii) near-zero variance analysis*: to eliminate feature without sufficient informative content (variance <0.01); *iii) correlation analysis*: to eliminate correlated feature (Spearman with cutoff=0.85); *iv) iterative feature selection*: to select only features useful for the problem under consideration. The remaining radiomic features were 16.

1.2 Clinical Features Selection and Missing Data Management

Starting with the initial clinical features (40), 'Hospital', 'Position', 'Death' and 'Prognosis' (4) are excluded a-priori; 'Fibrinogen', 'PCT', 'D_dimer', 'SaO2', 'Obesity' (5) were considered because with a missing data percentage over 50%; 'Ox_percentage', 'CardiovascularDisease', 'IschemicHeartDisease', 'AtrialFibrillation', 'HeartFailure', 'Ictus' (6) were excluded because not present on the test set. The remaining clinical features were 23. Missing values have been replaced with the mean value of each feature.

1.2.1 Selected Features for Modeling

A further iterative feature selection was performed considering the union of the 16 selected radiomic features and the 23 selected clinical features, to find eventual new relationships between clinical and radiomic features and to discard the redundant ones. At the end of this step 14 clinical and 6 radiomic features were selected, which are those used for the training of the final model.

1.3 Model Setup

Random Forest was used as a classifier. The model was setup as follows: *i) training phase*: 10-fold cross-validation (CV) was used, repeated 10 times; *ii) tuning phase*: the tuning of the hyper-parameters was carried out starting from the best model of the CV re-extracted on the entire training set; *iii) test phase*: the test was carried out on the final tuned model.

2. Experimental Results

In the training phase, the classifier was trained and evaluated by considering i) clinical features only; ii) radiomic features only; iii) both clinical and radiomic features.

Table 1. Accuracy obtained by the Random Forest classifier in the training set (10-fold CV).

	clinical features	radiomic features	clinical+radiomic features
accuracy \pm Std Dev	0.737 \pm 0.045	0.750 \pm 0.033	0.756 \pm 0.032

3. Discussion

Our work aims at the implementation of a fully Explainable Framework able to provide a multi-level explanation. We preferred the use of intrinsically explainable clinical and radiomics features excluding Deep Features in order to allow for a global and local explanation of the model. The Tree Explainer proposed by SHAP to interpret the features learned from the Random Forest algorithm was used. The proposed solution helps the developer in the debugging process, the physician to discover any clinical evidence and finally provide an explanation to the end user who receives the response of the system.

3.1 Global Explainability

Figure 1 shows the selected features having the highest impact on the trained model. The clinical features have high evidence and clinical validity: patients with high Lactate dehydrogenase concentration in blood (LDH) values are generally predisposed to severe diseases, while low values are predisposed to mild diseases [1]; Low Partial pressure of oxygen in arterial blood (PaO2) values are indicative of severe disease while high values are indicative of mild disease [2]; evidence confirmed that older subjects are more exposed to severe disease, as well as confirming the results found for glucose [3], sex [4], and RBC [5].

For the interpretation of the most discriminating radiomic features, filtered images were exactly calculated as described in PyRadiomics documentation, showing interesting visual results.

The most important features belong to the category of Gray Level Size Zone (GLSZM) that quantifies gray level zones in an image, where a gray level zone is defined as the number of connected voxels sharing the same gray level intensity. Higher value of the HighGrayLevelZoneEmphasis indicates a greater proportion of higher gray-level values and size zones in the image. In our case, high values means that the lung is much more uniform (large uniform regions) and no lesions are present. For ZoneEntropy, SEVERE

patients show a more heterogeneous texture. Hence, the behavior of ZoneEntropy is analogous to HighGrayLevelZoneEmphasis in the classification, as can be seen from the SHAP analysis in Figure 1.

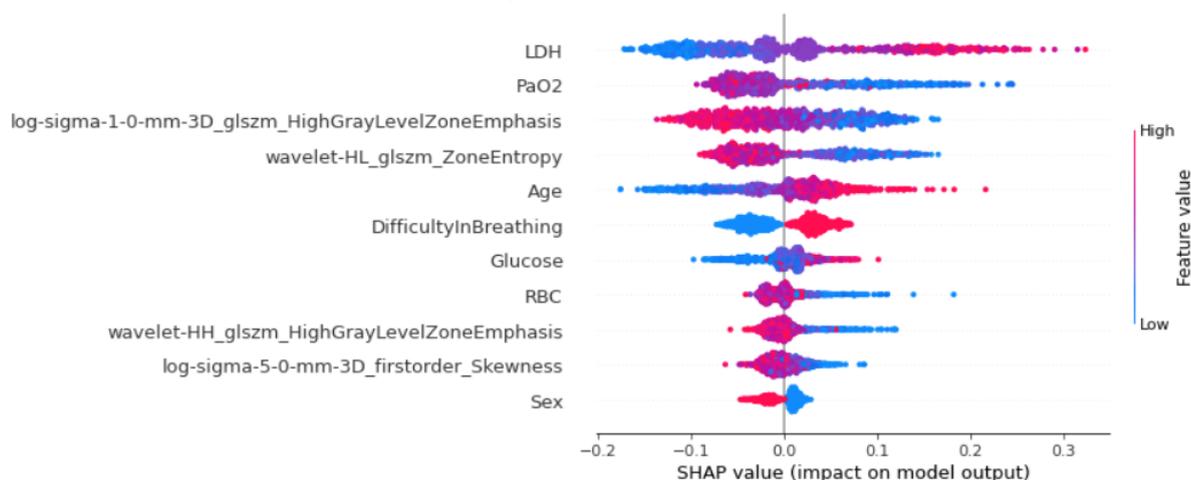


Figure 1. SHAP analysis - for simplicity have been plotted only the first eleven more important features for the trained model.

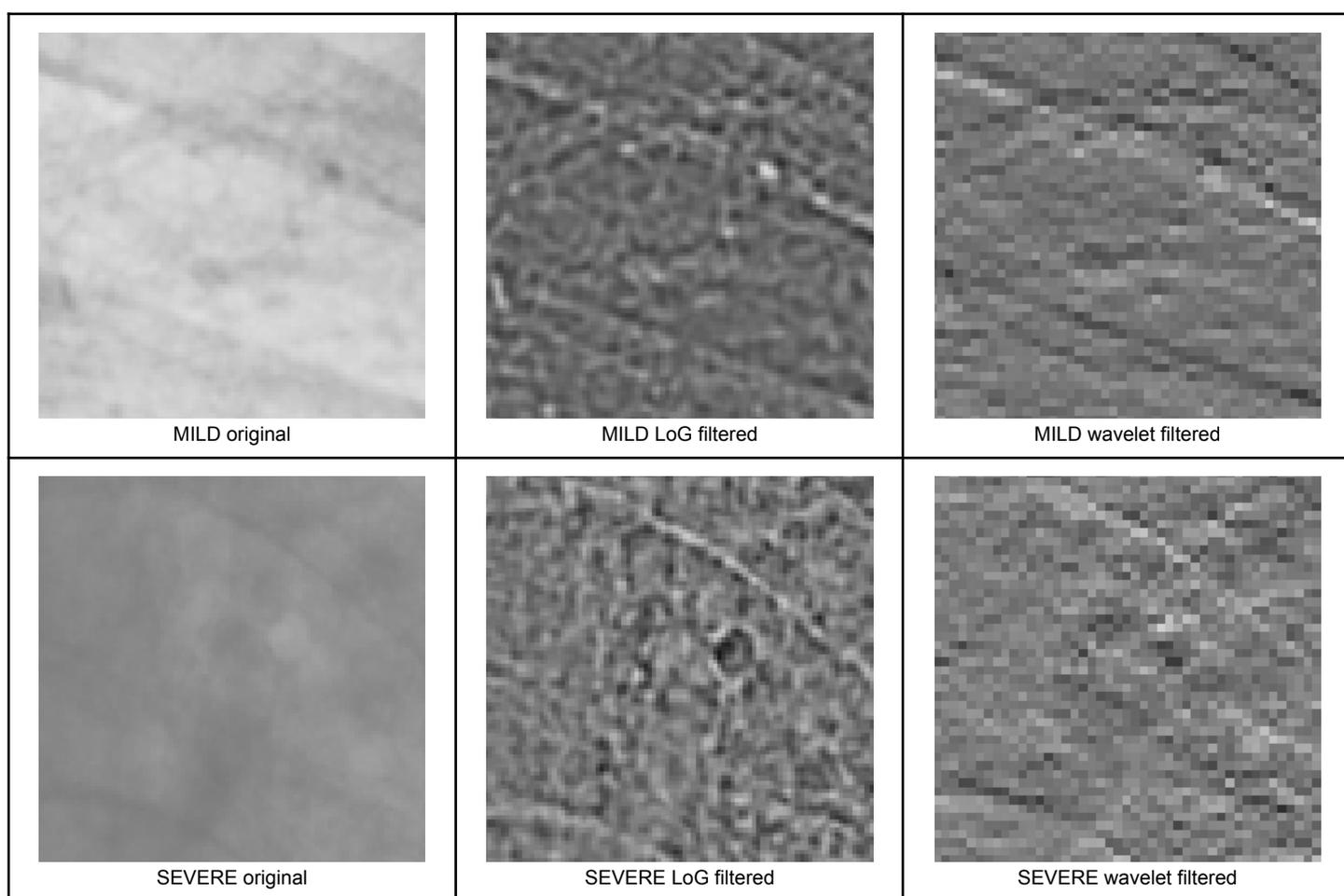


Figure 2. Graphics highlights the two most discriminant (i.e. HighGrayLevelZoneEnphasis and ZoneEntropy) radiomic features: in the upper row a MILD case; in the lower row a SEVERE case. The columns depict the original images (first column), the LoG filtered images (second column), and the wavelet filtered images (third and last column).

3.2 Local Explainability

Again using the Shapley values, it is also possible to obtain a local explanation, which is an interpretation of the result for a specific instance. Figures 3 and 4 show an explainability example for two patients predicted as MILD and SEVERE, respectively. For the first case, a low LDH value (299) and an absence of respiratory distress and imaging features, push the prediction towards a MILD prognosis. For the second case, a high LDH value (482) and an old age of the patient (82), lead to the prediction of SEVERE prognosis.

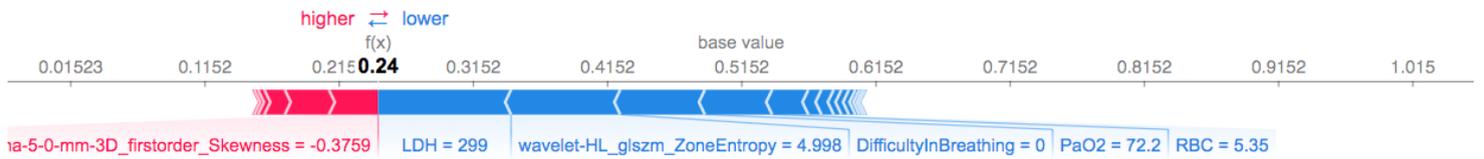


Figure 3. Local Explanation for a patient classified as MILD

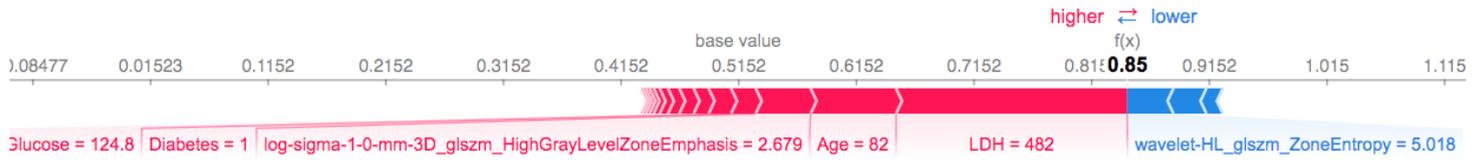


Figure 4. Local Explanation for a patient classified as SEVERE

References

1. Henry B.M., Aggarwal G., Wong J., et al. Lactate dehydrogenase levels predict coronavirus disease 2019 (COVID-19) severity and mortality: A pooled analysis. *Am J Emerg Med.* 2020; 38(9):1722-1726. doi:10.1016/j.ajem.2020.05.073
2. Tobin M.J., Jubran A., Laghi F. *European Respiratory Journal* 2021 57: 2100274; doi: 10.1183/13993003.00274-2021
3. Chen, Juan, et al. "The impact of COVID-19 on blood glucose: a systematic review and meta-analysis." *Frontiers in endocrinology* (2020): 732.
4. Nguyen, Ninh T., et al. "Male gender is a predictor of higher mortality in hospitalized adults with COVID-19." *PLoS One* 16.7 (2021): e0254066.
5. Kilercik, Meltem, et al. "A new haematocytometric index: Predicting severity and mortality risk value in COVID-19 patients." *PloS one* 16.8 (2021): e0254073.