covidcxr-hackathon: Solution description

Authors: Coppola Edoardo, Ferrari Damiano

Phase 1: Image pre-processing and BSNet

We have pre-processed the images with the following techniques: resizing to 512x512, normalization in a 0-1 range, adaptive histogram equalization (CLAHE, clip: 0.01), median filtering (kernel size: 3) and clipping outside the 2nd and 98th percentile. However, there are other problems with the images that concern their inversion and rotation. In order to solve them, we have fed the segmentation layers of BSNet [1] with the images and then we have analysed the segmentation probability mask produced by the network. If the input image is not in the correct format (w.r.t. rotation and inversion), the output mask is basically noise. Therefore, we have defined two tests that the mask should pass to correctly classify the input image. The first test verifies the existence of two regions of a certain size (the lungs) within the mask; the second test checks the shape and direction of the lungs. If the image passes the two tests, then it is used to compute the Brixia-score[1], otherwise it is first inverted and then rotated until both tests are satisfied. If the image does not pass the tests in any of the eight possible combinations of inversion and rotation, then it is discarded and a message feedback regarding its quality is returned to the user[2]. Finally, the computed Brixia-score, and its associated confidence, are used as additional features for the machine learning model.

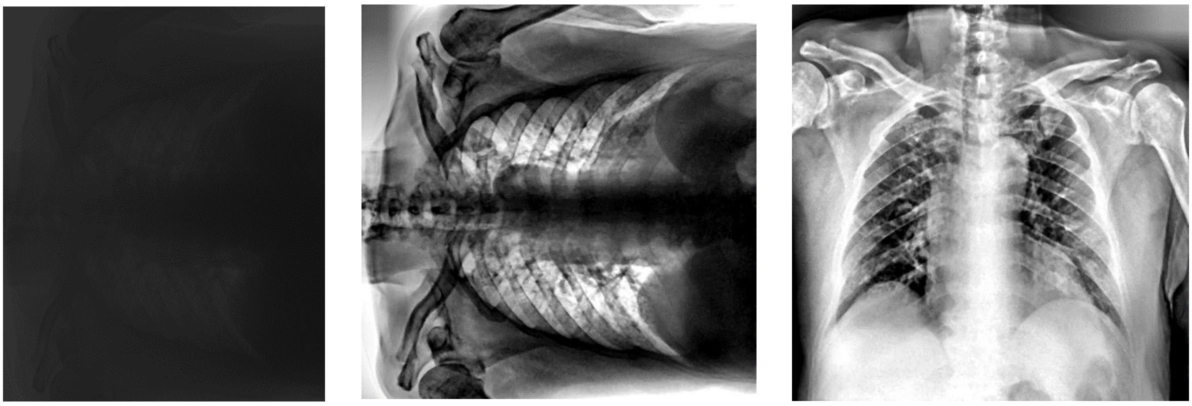

Figure 1 - These three images show, from left to right, the original CXR, its preprocessed version and the final corrected image with respect to rotation and inversion. They refer to P_1_60.png from the training set.

Phase 2: Machine Learning Model

From the first steps of the pre-processing phase we have tried to put ourselves in conditions as similar as possible to those of the test set records so, since these instances lack information about *Oxygen percentage*, *Fibrinogen* and *Position*, we have dropped such features. Missing values have been handled both using a KNN imputation and a constant filling strategy. Then, we have eliminated the records associated with the images discarded during the process described in the previous section (they represent just 1.72% of the training records). Subsequently, in order both to identify significant features for the explicability of the predictions and to reduce the dimensionality of the dataset, we have conducted several model-based feature extractions varying both the threshold of feature importance and the maximum number of attributes to be selected. The selection that has led to the best performances, in terms of balanced accuracy, includes *Hospital, Age, Sex, Temp_C, DaysFever, DifficultyInBreathing, WBC, RBC, CRP, Glucose, PCT, LDH, D_dimer, PaO2, SaO2, PaCO2, pH, Brixia_score_confidence* and *Brixia_score*. Finally, we have trained models like Decision Trees, Support Vector Machines, Extra Trees, XGBoost classifiers and Random Forests. The latter performances are as follows: *Sensitivity*: 78.95%, *Specificity*: 80.43%, *Balanced Accuracy*: 79.69%. Since Random Forests have shown themselves as promising classifiers for this task, we have decided to conduct a new hyper-parameter tuning based on records whose Hospital value is equal to F. Finally, the tuned model has been re-trained over the whole training set. This classifier is what we have used to predict test set patients' prognoses.

[1] A scoring system which helps both quantifying and locating lung abnormalities. See [2]

[2] Despite the poor quality of the CXR, its Brixia-score and the associated confidence are computed anyway.

Explainability of the model

Since the system will always try to adjust the input image with respect to its inversion and/or rotation, if the CXR has an excessively low resolution, it is completely black or the lungs are not quite visible, a negative feedback will notify the user about the poor quality of the radiograph. The Brixia-score produced by BSNet explains the lung abnormalities of the CXR. Each lung is subdivided into three regions (upper, middle and lower) and each one has an associated severity in a scale from 0 (no lung abnormalities) to 3 (interstitial, and dominant alveolar infiltrates). More details about the scoring system can be found in [2]. The lungs explainability maps, of which we can find an example in Figure 2, offer a deeper description of the Brixia-score (See [1], section 6.8). In the example we can see that the black lower region of the right lung has a strong influence (it is darker) on the score with class 3 (whose colour is black), while the upper green part of the same lung slightly impacts (it is lighter) the score associated with class 0 (whose colour is green).

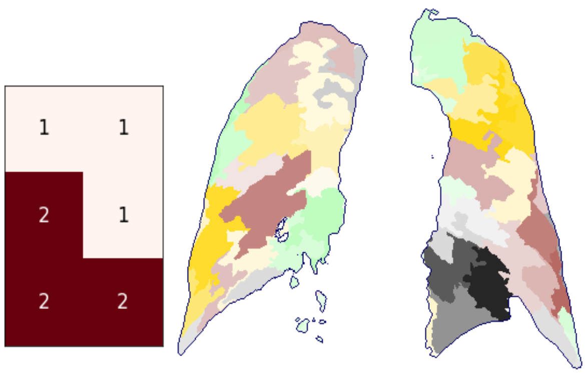

Figure 2 – Brixia-score and associated lungs explainability maps

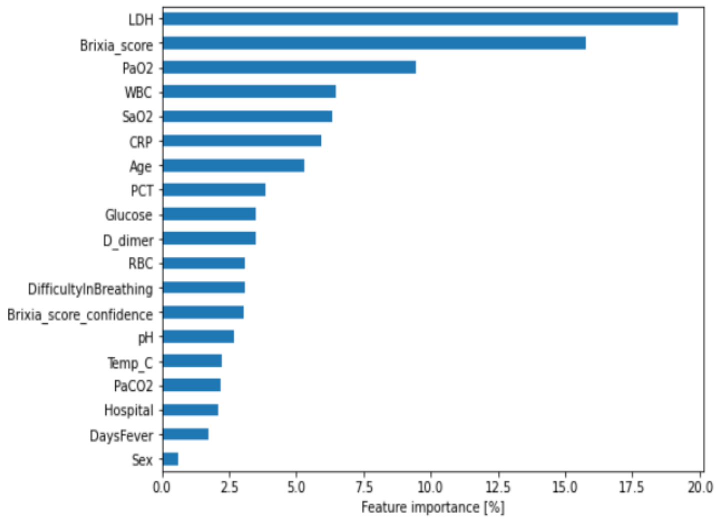

Figure 3 – Importance of patient's features to the model

Random Forest also provides some form of explainability that can be used as a hint for doctors in the prognostic process. This is represented by the importance given by the model to the features used for predicting a patient's prognosis, as depicted in Figure 3.
The last piece of information that may be useful for doctors is given by the chart in Figure 4. It shows a single patient's feature values whose true class is *severe*. Since the model output f(x) is equal to 0.64 and it is greater than the base value of 0.5, the predicted class will be *severe*. For this particular patient, the features in red are those that force the model prediction to be *severe*, vice versa the blue features drag the prediction towards *mild*. In the example, the patient's advanced age, his difficulty in breathing and his Brixia-score of 9 force the model to predict a *severe* prognosis.

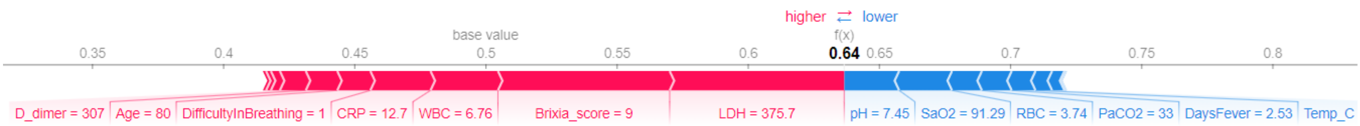

Figure 4 – Impact of a single patient's features on output

Acknowledgements

We would like to thank Prof. A. Signoroni and Dr. M. Savardi of the University of Brescia for supporting and advising us in this hackathon.

Other datasets

We have not explicitly used other datasets, but we have used the pre-trained model of BSNet which was trained on Cohen et al. dataset with Brixia-score annotation. Other segmentation datasets include Montgomery County, Shenzhen Hospital and JSRT database.

Bibliography

[1] A.Signoroni, M. Savardi et al. "BS-Net: learning COVID-19 pneumonia severity on a large Chest X-Ray dataset", Medical Image Analysis, 2021
[2] A. Borghesi, R. Maroldi "COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression", La radiologia medica, 2020