

Automated prediction of COVID-19 severity upon admission by chest X-ray images and clinical metadata aiming at accuracy and explainability

ALEX OLAR, Eötvös Loránd University, Department of Complex Systems in Physics & Doctoral School of Informatics, Hungary

ANDRÁS M. BIRICZ, Eötvös Loránd University, Department of Complex Systems in Physics, Hungary

ZSOLT BEDÖHÁZI, Eötvös Loránd University, Department of Complex Systems in Physics & Doctoral School of Informatics, Hungary

BENDEGÚZ SULYOK, Eötvös Loránd University, Department of Complex Systems in Physics, Hungary

In the past three years COVID-19 posed a huge threat to healthcare systems around the world. One of the first waves of the pandemic hit Northern Italy severely resulting in high casualties and in the near breakdown of primary care. Due to these facts, the *Covid CXR Hackathon - Artificial Intelligence for Covid-19 prognosis: aiming at accuracy and explainability* has been launched at the beginning of February. In this short article we summarize our attempts at correctly diagnosing chest X-ray images created upon admission for severity of COVID-19 outcome. In addition to X-ray imagery clinical metadata was provided and the challenge also aimed at creating an explainable model. We created a best-performing as well as an explainable model that makes an effort to connect clinical metadata to image features whilst predicting the prognosis.

Additional Key Words and Phrases: covid-19, CXR, explainability, neural networks

1 DATA PROCESSING

The X-ray images and the clinical metadata from several hospitals in Northern Italy were provided. They were converted from different types of DICOM images ranging from 12-bit to 16-bit in precision and also digitized in various ways. The resulting images varied significantly in terms of quality: some images were rotated randomly by 90°, many were inverted and some contained a fully blacked-out or a gray margin on some edges of the scan. In order to deal with this, we selected the top two corners and the middle section of the scan in order to automatically decide whether the scan is inverted or not based on the mean brightness ratio between these regions; the inverted images have been reverted. We also manually annotated all images for inversion and concluded that our method was almost always (around 99%) accurate and corrected the rest of the images manually both in the training and the test sets. After finishing the inversion we followed the pre-processing steps from [Signoroni et al. 2021] by applying clipping between the 2nd and 98th percentile, quantizing the image to 8-bits, used contrast-limited adaptive histogram-equalization (CLAHE) and median filtering to reduce noise, finally we scaled the images to 512 x 512 pixels in order to be eligible for the next steps.

1.1 Additional pre-processing

From the 512 x 512 scans we created 2D Fourier images scaled to the range of 0-1 after taking the absolute value of the 2D Fourier-transform. From [Signoroni et al. 2021] we used their BSNet to align and segment all scans in order to create a more uniform dataset and later on provide additional information, in the form of segmentation masks, to our model. We also ran the BSNet model to provide the Brixia-scores for each image in order to append to the metadata. We also tried several pre-processing steps for the metadata in order to differently impute missing values, these attempts contained imputation based on age groups, hospitals, population means (medians for categorical values) and samples. We applied these pre-processing steps to the test and train sets with the accompanying metadata and started the modeling part afterwards.

2 PRECISION NETWORK

Having acquired additional features in the form of 2D Fourier transforms, Brixia-scores, segmentation masks and aligned images as well as the imputed metadata we started creating our precision network. We aimed at the integration of all the features into one single unit with several single-feature processing layers. A convolutional backbone processed the pre-processed image and the segmentation mask and after an average pooling created a feature vector, another smaller convolutional network processed the Fourier features while a fully connected network processed the imputed clinical metadata concatenated with the Brixia-scores. All the output feature vectors were then concatenated together and passed through a fully connected head that did the prognosis prediction in addition to the death prediction in order to do multitask-learning. In the inference phase we only used the prognosis output to predict the severity of COVID-19 for the test set. The major parameters, such as the type of the convolutional backbone, size of the last dense layer, image size, batch size, number of epochs, learning rate, etc. were tunable hyperparameters for which we did hyperparameter-optimization. Finally, we train several networks with different backbones and ensemble predict the mode of the outcomes from these to reduce variance.

3 EXPLAINABLE NETWORK

The explainable network has two main features, one convolutional sub-network processes the image data into an unpooled feature map (16 x 16 or 8 x 8) with many feature channels. The second sub-network is a small Transformer [Vaswani et al. 2017] for the clinical metadata without positional encoding, as the ordering does not matter in our case. Afterwards, a Bahdanau attention map is calculated between each processed, embedded meta feature and are stacked

Authors' addresses: Alex Olar, Eötvös Loránd University, Department of Complex Systems in Physics & Doctoral School of Informatics, Budapest, Hungary, 1117, olaralex@student.elte.hu; András M. Biricz, Eötvös Loránd University, Department of Complex Systems in Physics, Budapest, Hungary, ; Zsolt Bedöházi, Eötvös Loránd University, Department of Complex Systems in Physics & Doctoral School of Informatics, Budapest, Hungary, ; Bendegúz Sulyok, Eötvös Loránd University, Department of Complex Systems in Physics, Budapest, Hungary.

together into one large context vector, averaged along the feature maps spatial dimensions after applying the attentional weights. The context vector, still being a high-dimensional feature vector for each meta column, we process further with two layers of bidirectional LSTM [Hochreiter and Schmidhuber 1997] networks having the second only output the prognosis and death predictions in order to do multitask learning. This model is optimized via stochastic gradient descent with tunable hyperparameters for the convolutional backbone, number of attention heads, dimensionality of the Bahdanau attention mechanism [Bahdanau et al. 2014], Transformer encoding dimensions, etc.

4 RESULTS

We would like to sum-up our findings in an article after the end of this challenge, so we ran several models (with image data only, with meta data only and the image and meta data combined) with various architectures in each scenario and we can conclude that our final deep learning based models are on-par with the classical methods and it seems that the X-ray scans at admission do not improve the prognosis prediction significantly. We created averaged scans in both mild and severe categories of the aligned images in order to spot the differences between prognosis outcomes as seen in [Fig. 1].

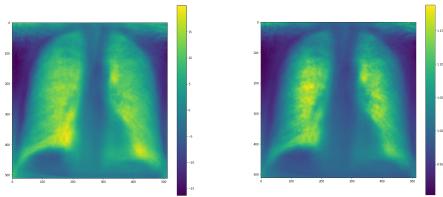


Fig. 1. (left) Difference image of SEVERE and MILD mean images, (right) ratio image of SEVERE and MILD mean images

For our explainable model we averaged the output attention maps of all test images and meta features (as well as training samples to be able to spot any difference) and created mean attention maps for each relevant meta feature with respect to the predicted outcome. These attention maps correlate well with some of the meta features, such as coughing, heart failure, obesity, cardiovascular disease, etc.

We superimposed the [Fig. 1 (left)] image behind these structures [Fig. 3]. This way we can see that some features that are related to the heart: heart failure, cardiovascular disease put high attention to that spatial region. Also the features that cannot be spatially localized tend to have spread-out attention, such as: obesity, dementia, glucose, days of fever. We can also see, that features that have very high correlation with the lungs themselves, have high attention on specific regions of the lungs (bottom or sides): arterial blood oxygen pressure, oxygen percentage, difficulty of breathing, coughing, etc.

In [Fig. ??] we show the same attention maps on the averaged difference image to highlight the corresponding regions. In our experiments we have also seen that we could optimize our large explainable models to almost the same level as our performance model, but we lost the explainable feature of the attention maps,

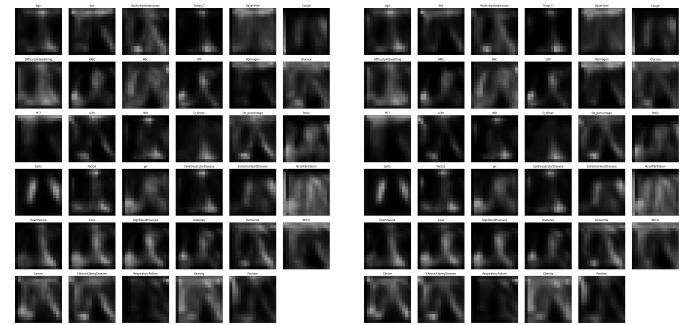


Fig. 2. (left) Mean attention maps on a random subset of the training data (right) mean attention maps on the testing data

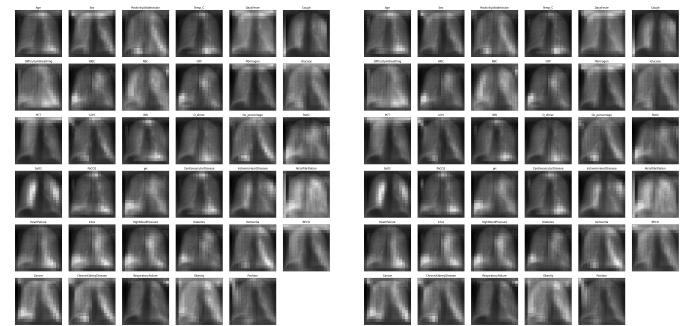


Fig. 3. (left) Mean attention maps on a random subset of the training data (right) mean attention maps on the testing data, both of the images have the difference of SEVERE and MILD average scans behind them.

therefore we restricted training time and chose hyperparameters carefully, in order to achieve better explainability maps.

5 ACKNOWLEDGMENTS

We would like to thank the National Research, Development and Innovation Office of Hungary grants 2020-1.1.2-PIACI-KFI-2021-00298 and the MILAB Artificial Intelligence National Laboratory Program.

REFERENCES

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- Alberto Signoroni, Mattia Savardi, Sergio Benini, Nicola Adami, Riccardo Leonardi, Paolo Gibellini, Filippo Vaccher, Marco Ravanelli, Andrea Borghesi, Roberto Maroldi, and Davide Farina. 2021. BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset. *Medical Image Analysis* 71 (2021), 102046. <https://doi.org/10.1016/j.media.2021.102046>
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).