

# Automated prediction of COVID-19 severity upon admission by chest X-ray and explainability based on imagery and meta data

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 BEDŐHÁZI, Eötvös Loránd University, Department of Complex Systems in Physics, Hungary

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

In the past three years COVID-19 posed a huge threat to the healthcare systems all around the world. One of the first waves of the pandemic hit Northern Italy severely resulting in high casualties and almost in the 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 sum up our attempts to correctly diagnose chest X-rays created upon admission for severity of COVID-19 outcome. In addition to X-rays clinical metadata was provided and the challenge also aimed for 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-rays and the clinical metadata were provided from several hospitals in Northern Italy. They were converted from different types of DICOM images ranging from 12-bit to 16-bit in precision and also digitized in many different ways. This resulted in images that varied significantly, contained some images that were rotated randomly by 90°, many of them 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 to automatically decide whether the scan is inverted or not based on the mean brightness ratio between these regions. We also manually annotated all the 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 to be able to apply 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 in order to align and segment all the 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 have more metadata. We also tried several pre-processing steps for the meta data 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 pre-processed this way 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 original (and already imputed) metadata we started to create our precision network. With this we aimed for integration of all these features into one single unit with several single-feature processing layers. A convolutional backbone processed the aligned image and the segmentation mask and after an average pooling created a feature vector, another convolutional, but smaller network processed the Fourier-features while a fully connected net processed the imputed clinical metadata concatenated with the Brixia-scores. All the output feature vectors then were concatenated together and passed through fully connected head that did the prognosis prediction in addition to the death prediction 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 backbone, size of the last dense layer, image size, batch size, number of epochs, learning rate, etc. were tunable hyper-parameters for which we did hyper parameter-optimization.

## 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 meta data 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 together into one large context vector, meaned along the feature maps spatial dimensions after applying the attentional weights. The context vector, still being a many features vector for each meta column, we process it further with two layers of bidirectional LSTM [Hochreiter and Schmidhuber 1997] networks having the second only output the prognosis and death predictions to also do multitask learning. This model is optimized via stochastic-gradient descent with tunable hyper-parameters for the CNN backbone, number of

---

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 Bedőházi, Eötvös Loránd University, Department of Complex Systems in Physics, Budapest, Hungary, ; Bendegúz Sulyok, Eötvös Loránd University, Department of Complex Systems in Physics, Budapest, Hungary.

attention heads, number of Bahdanau [Bahdanau et al. 2014] attentions dimensions, Transformer encoding dimensions, etc.

## 4 RESULTS

TBD

## 5 ACKNOWLEDGMENTS

We would like to thank...

## 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).