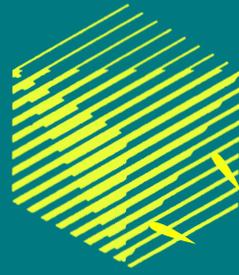# ERCIM NEWS

Special theme:

eXplainable AI

Also in this issue

Research and Innovation:

Fibre-Optic Sensing for Road-Traffic Monitoring in Remote Areas

ence between normal and abnormal heartbeats: the lower and evident V-shape in the rightmost part of the time series. The saliency map confirms this observation by highlighting the last observations of the time series. The rules further illustrate the differences between classes, with the factual rule indicating the presence of a specific subsequence in normal instances and the counterfactual rule indicating its absence in abnormal instances. While the saliency map and rules may not cover the exact same areas, they provide complementary insights into the discriminative features of the time series.

Overall, LASTS represents a significant advancement in the field of time series explanation, with promising potential for future research and practical applications. In our future research, we plan to explore several directions to enhance LASTS. Firstly, we intend to evaluate the framework on longer and more complex real-world time series datasets, aiming to validate its performance in challenging scenarios. Additionally, we aim to extend LASTS to other types of sequential data, such as trajectories, text, and shopping transactions, in order to broaden its scope of applicability. Secondly, we will delve deeper into the relationship between the latent and subsequence spaces, conducting further investigations to gain a comprehensive understanding of their interactions. Finally, we intend to conduct human decision-making tasks guided by LASTS explanations, offering practical evaluation and valuable insights into the effectiveness of the explanations in real-world decision scenarios.

This article is coauthored with Mirco Nanni, Fosca Giannotti, and Dino Pedreschi (CNR-ISTI, Scuola Normale Superiore, Università di Pisa).

Links:
[L1] https://artificialintelligenceact.eu/
[L2] https://kwz.me/hxK

References:
[1] A. Theissler, et al., "Explainable AI for time series classification: a review, taxonomy and research directions," IEEE Access, 2022.
[2] R. Guidotti, et al., "Explaining any time series classifier," 2020 IEEE 2nd Int. Conf. on Cognitive Machine Intelligence (CogMI), IEEE, 2020.

Please contact:
Francesco Spinnato
Scuola Normale Superiore and CNR-ISTI, Pisa, Italy
francesco.spinnato@sns.it
Riccardo Guidotti
University of Pisa, Pisa, Italy
riccardo.guidotti@unipi.it

Explaining Ensemble Models for Lung Ultrasound Classification

by Antonio Bruno, Giacomo Ignesti and Massimo Martinelli (CNR-ISTI)

Correct classification is the main aspect in evaluating the quality of an artificial intelligence system, but what happens when you reach top accuracy and no method explains how it works? In our study, we aim at addressing the black-box problem using an ad-hoc built classifier for lung ultrasound images.

In the last few years, the novelties of artificial intelligence (AI) and computer vision (CV) significantly increased, allowing new algorithms to obtain meaningful information from digital images. Medicine is a field in which the use of this technology is experiencing fast growth. In 2020, in the USA alone, the production of 600 million medical images was reported, and this number seems to increase steadily. Robust and trustworthy algorithms need to be developed in a multi-disciplinary collaboration.

During the SARS-CoV-2 pandemic, a fast and safe response became even more necessary. The use of point-of-care ultrasound (POCUS) to detect SARS-CoV-2 (viral) pneumonia and the bacterial infection emerged as one of the most peculiar emerging case studies, involving the use of on-site ultrasound examinations rather than a dedicated facility. As well as being faster, safer and less expensive, lung ultrasound (LUS) also appears to detect signs of lung diseases as well as or even better than other methods, such as X-ray and computed tomography (CT).

The employment of lung POCUS seemed an optimal solution for both quarantined and hospitalised subjects. CT and magnetic resonance imaging (MRI) are far more precise and reliable examinations, but both have downsides over mass screening. In our study, an efficient adaptive minimal ensembling model was developed to classify LUS using the largest publicly available dataset, the COVID-19 lung ultrasound dataset [L1], composed of 261 ultrasound videos and images from 216 different patients. The General Data Protection Regulation
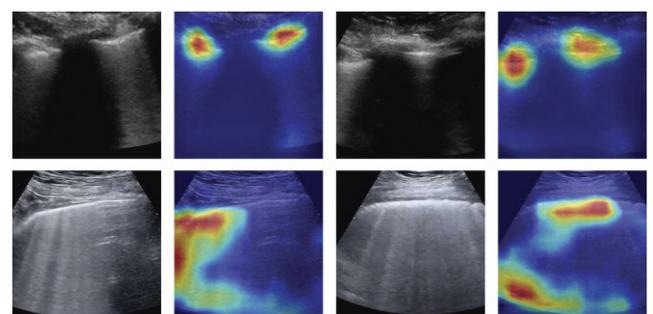

Figure 1: COVID-19 – Original and Grad-CAM-processed samples are shown for subjects with COVID-19; different images within different subjects show similar activation maps.

(GDPR) and the European Committee AI Act focus on the intent that an AI-deployed system should be trustworthy and fully explainable.

Several explainability approaches arose from the scientific community and new ones are under development. Focusing on image interpretation, a debate about which approach should be used is ongoing. As the core model, we selected EfficientNet-b0 [1] because of its good accuracy/complexity trade-off.

The efficiency of this architecture is given by two main factors:
- the reduced number of parameters given by compound scaling, by which input scaling (i.e. input size), width scaling (i.e. convolutional kernel size) and depth scaling (i.e. number of layers) are performed in conjunction because they are dependent
- the low number of FLOPs (floating point operations) of the inverted bottleneck MBConv (first introduced in MobileNetV2, an efficient model designed to run on smartphones) as a main constituent block.

The greatest contribution of our study is given by the introduction of an ensembling strategy. Due to its resource-consuming nature and the exponential growth of model complexity, ensembling is scarcely used in computer vision, but we demonstrate how to perform it in an adaptive and efficient way:
- using only two weak models (minimality, efficiency)
- performing the ensemble using a linear combination layer, trainable by gradient descent (adaptivity)
- performing it using the features instead of the output, excluding redundant operations (efficiency)
- fine-tuning the combination layer only (efficiency).

The linear combination layer used to perform the adaptive ensemble can be described by the following equation:

$$Feat_{comb} = W_{comb}Feat_{concat} + b_{comb} \ ,$$

	Class	Recall	Precision	F1-Score
InceptionV3 Acc.: 89.1% (89.3%) #Param.: 23.9M FLOPs: 6G	COVID-19 Pneumonia Healthy	0.864±0.036 0.908±0.025 0.907±0.026	0.901±0.031 0.842±0.037 0.918±0.021	0.880±0.030 0.871±0.025 0.911±0.021
DenseNet-201 Acc.: 90.4% #Param.: 20M FLOPs: 4.29G	COVID-19 Pneumonia Healthy	0.892 0.903 0.850	0.918 0.610 0.842	0.905 0.728 0.846
Light Transformer Acc.: 93.4% #Param.: 0.3M FLOPs: 1G	COVID-19 Pneumonia Healthy	0.958±0.025 0.948±0.013 0.877±0.034	0.958±0.012 0.951±0.038 0.912±0.037	0.951±0.017 0.949±0.020 0.894±0.036
Weak model (our) Acc.: 98.7% (98.3%) #Param.: 5M FLOPs: 0.39G	COVID-19 Pneumonia Healthy	0.984±0.004 0.997±0.005 0.999±0.003	0.993±0.004 0.991±0.006 0.993±0.003	0.990±0.004 0.991±0.007 0.995±0.004
Ensemble (our) Acc.: 100% (100%)	COVID-19	1.000±0.000	1.000±0.000	1.000±0.000

Table 1: Test set comparisons (on 5-fold cross-validation), with metrics for each class, of the proposed model with the SOTA. Our solution outperforms the SOTA on all metrics and has the lowest complexity.

where:
- $Feat_{concat} = feat_{weak1} \cdot feat_{weak2}$ is the concatenation of the weak features
- $feat_{weak1}$, $feat_{weak2}$ are the features obtained by the two weak models.

Table 1 shows that the ensemble further reduces the variance and improves the generalisation power (i.e. performance on validation and test dataset), outperforming the state-of-the-art (SOTA) with lower complexity (moreover, the complexity of this ensemble strategy is equal to the complexity of a single EfficientNet-b0, since the processing of the weak models is independent and parallelised).

Even if our model can give extremely accurate and fast responses, it is crucial that it also is secure and understandable. To this aim, we applied Grad-CAM, since this method uses gradients with respect to a particular convolutional feature map (in our case the last convolutional layer of the model) to identify the regions on the input that are more discriminative for the classification results (i.e. higher gradient value means higher contribution to the classification).
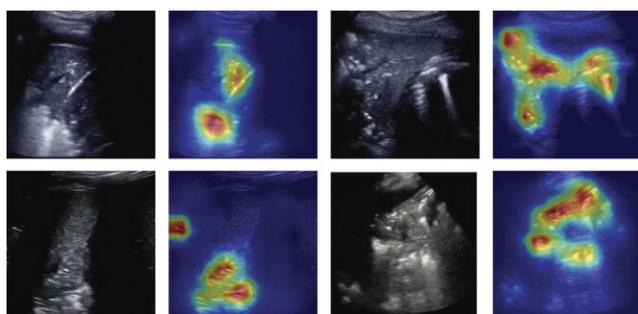

Figure 2: Pneumonia – Original and Grad-CAM-processed samples are shown for subjects with pneumonia; the attention of the classifier is on different regions of the images in contrast to the other two classes.
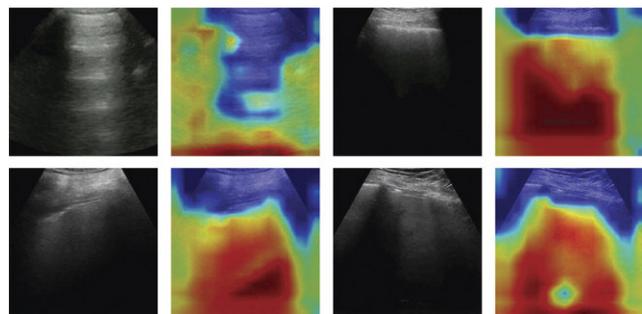

Figure 3: Healthy – Original and Grad-CAM-processed samples are shown for healthy subjects; the classifier focuses its attention all over the image or outside; it seems it does not find relevant information, unlike the cases with a pathology.

The application of this technique to our model gives reasonable results. A first non-trivial result is that for the images classified in the same way, a similar gradient is activated, which in turn originates from similar areas in the images. This result was further investigated by comparing the saliency map of the same correctly classified images between the EfficientNet-b0 and the ensemble models. While both identify similar parts, the ensemble model seems to maintain attention on a more concentrated area.

Figures 1, 2 and 3 show that there are typical signs of evidence for each class:

- COVID-19 (Figure 1) – usually, more concentrated and relatively large, mainly over the pleural line and on the "edges"
- Pneumonia (Figure 2) – evidence mainly below the pleural line, with widespread area having spots
- Healthy (Figure 3) – mainly the very expanded, healthy part of the lung (black).

Even if this study seems to provide robust and interpretable results, it lacks in-depth research on the effective stability of the explanation. To further test our method, in an ongoing telemedicine project [2], in close collaboration with specialist sonographer staff, we are going to use further explainability methods on other important signs that can be present in LUS (e.g. A-lines, B-lines, thickness), in order to improve the correlation with the ground truth.

Link:
 [L1]
https://github.com/jannisborn/covid19_ultrasound/blob/master/data/README.md

References:
[1] Mingxing Tan and Quoc V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114, 2019. Available at: http://proceedings.mlr.press/v97/tan19a.html
[2] G. Ignesti et al., "An intelligent platform of services based on multimedia understanding and telehealth for supporting the management of SARS-CoV-2 multi-pathological patients," in Proc. 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2022, pp. 553–560. Available at: https://doi.org/10.1109/SITIS57111.2022.00089

Please contact:
Massimo Martinelli, CNR-ISTI, Italy
massimo.martinelli@isti.cnr.it

A Governance and Assessment Model for Ethical Artificial Intelligence in Healthcare

by Luigi Briguglio, Francesca Morpurgo and Carmela Occhipinti (CyberEthics Lab.)

How can clinicians be deemed responsible for basing their decisions on diagnoses generated by artificial intelligence and derived in a way that cannot be fully understood? How can patients rely and accept decisions if they are based on "black boxes" of data and algorithms? In the context of the MES-CoBraD project, CyberEthics Lab. defines a model for governing and assessing "Ethical Artificial Intelligence" (ETHAI).

Innovative technologies and approaches have enabled the evolution of many sectors and the increased well-being of our society, reducing time to produce and make available solutions to the public, and improving quality of life. Healthcare services have been a core part of this process, achieving impressive milestones in less than a hundred years. The COVID-19 pandemic has shown how reliable scientific research can be, producing vaccines in a relatively short time and under difficult working conditions [1].

Successful clinical care requires that all people have access to good-quality health care and that they can comply with recommended treatments. However, the ability to respond to and support the health demands of citizens depends on both the health system of a country and the economic and health structure of society (e.g. an ever-growing rate of ageing population means more attention).

This reflects in the global trend of innovating health care by introducing digitised and distributed "neighbourhood healthcare centres" capable of offering wider access to care for citizens. Delivering better health care means a more efficient and effective intervention in diagnosis, and thus prevention and early detection of non-communicable diseases can be better tracked. Ensuring health and promoting well-being for all and for all ages is set out as the Goal 3 of the 2030 Agenda for Sustainable Development of the United Nations [2].

In the digital era, healthcare (eHealth) and prevention actions will be characterised by: (i) preventive health care based on diffused remote monitoring through connected devices, that is, the Internet-of-Things (IoT) and predictive disease detection through computing capabilities, that is, Artificial Intelligence (AI); (ii) prevention actions, continuous contact with healthcare providers and smarter medication; (iii) remote and in-hospital assistance and surgery with dedicated robotics; (iv) asset- and intervention-management in hospitals and healthcare centres [3].

In this context, researchers of the Multidisciplinary Expert System for the Assessment & Management of Complex Brain Disorders (MES-CoBraD) project [L1] are working together to exploit the potential of data and AI to develop a common innovative protocol for the accurate diagnosis and personalised