NEUR CRITICAL
CARE SOCIETY

LETTERS TO THE EDITOR

Comment on "Machine Learning for Early Detection of Hypoxic-ischemic Brain Injury After Cardiac Arrest"

Noah S. Molinski[1*] , Aymen Meddeb[2], Martin Kenda[3] and Michael Scheel[1]

With great interest, we have read the article by Mansour et al. [1], reporting on the use of deep transfer learning to identify early signs of hypoxic-ischemic brain injury (HIBI) on head computed tomography (HCT) scans. The authors report a very high accuracy (0.94) of their model with respect to the detection of HIBI signs on HCT scans performed within hours after the return of spontaneous circulation. The authors conclude that "Deep transfer learning reliably identifies HIBI in normal appearing findings on HCT performed within 3 h after ROSC in comatose survivors of a cardiac arrest" [1]. This interpretation is likely too optimistic.

Deep learning networks show poor classification results and tend to be overfitted when trained on a very small data set [2]. A medical imaging data set of 54 HCT scans is a very small training data set. Further, we think that the following methodological issues could also contribute to overfitting in this study: (1) choice of the network, (2) the training pipeline (data augmentation, early stopping), and (3) principal component analysis (PCA) and repeated data usage.

No justification was given for why a VGG19 network was chosen, although it has a significantly worse accuracy in the analysis of CT data than, for instance, ResNet-50 or DenseNet-201 networks [3]. At the same time, it

remains unclear why only ImageNet data and no medical imaging data were pretrained. The natural images from ImageNet differ in many aspects from clinical imaging data: image shape, colors, resolution, and dimension. Therefore, the network is trained on parameters that are irrelevant for its purpose, which may interfere with an accurate analysis.

Furthermore, it was not mentioned whether regularization methods such as transformations of the raw data (e.g., resizing, rotations, flipping, intensity shifting and/or scaling, Gaussian noise, zooming), weight constraints, or activity regularizations were used for reducing overfitting [4]. It remains unclear how many epochs the final model has been trained for. "Early stopping" (monitoring of the model performance on a validation set and then stopping training when the performance degrades) has become universally established to keep weights small during training and reduce the risk of overfitting [4].

Another aspect is the use of PCA. Because PCA is a linear algorithm for dimensionality reduction, the question arises on which basis a linear relationship between the detected features can be assumed. Given the complexity of the present data in terms of possible blurring or degradation due to fluctuating contrast, it is problematic to make such assumptions on the basis of the representation of shapes and images using smooth manifolds. Nonlinear methods (manifold learning), such as kernel PCA, t-distributed Stochastic Neighbor Embedding, or Multidimensional Scaling, could be applied instead. Moreover, the authors write "single-scan testing was repeated so that each of the 54 scans served as the test scan exactly one time" [1]. Although the leave-one-out cross validation described above improves model quality, the multiple repeated uses of the same data as training data can strongly facilitate overfitting.

*Correspondence: molinski@ardenne.de
[1] Department of Neuroradiology, Charité-Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, 10117 Berlin, Germany
Full list of author information is available at the end of the article

Springer

As the authors reported that early HIBI signs were due to "subtle changes that evade the detection threshold of the human eye" [1], it would have been desirable to visualize by using heat maps or GradCAM, in which the subtle changes in the brain could start [5]. Those are important tools to plausibly illustrate the "thinking process of AI" to the readers.

The authors used a very small data set ($n = 16$) for validation. On this data set, the positive predictive value was 0.5, indicating that in the validation set a prediction of severe HIBI from early HCT had a 50% chance of being correct.

In conclusion, we agree that machine learning is an attractive new tool that may help to better predict severe HIBI from early HCT scans in cardiac arrest survivors in the future. The study by Mansour et al. [1] is a first step, but further studies on larger cohorts are necessary before it can be safely concluded that "deep transfer learning reliably identifies HIBI" from early HCT scans.

Author details
[1] Department of Neuroradiology, Charité-Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, 10117 Berlin, Germany. [2] Department of Radiology, Charité-Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Hindenburgdamm 30, 12203 Berlin, Germany. [3] Department of Neurology, Charité-Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Augustenburger Platz 1, 13353 Berlin, Germany.

Author contributions
NSM, AM, and MS were involved in conceptualization. NSM and AM were involved in formal analysis and drafted the article. NSM, AM, MK, and MS were involved in critical revision of the article for important intellectual content. All authors approve of the final manuscript.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Mansour A, Fuhrman JD, Ammar FE, Loggini A, Davis J, Lazaridis C, et al. Machine learning for early detection of hypoxic-ischemic brain injury after cardiac arrest. Neurocrit Care. 2021. https://doi.org/10.1007/s12028-021-01405-y.
2. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci. 2004;44(1):1–12. https://doi.org/10.1021/ci0342472.
3. Pham TD. A comprehensive study on classification of COVID-19 on computed tomography with pretrained convolutional neural networks. Sci Rep. 2020;10:16942. https://doi.org/10.1038/s41598-020-74164-z.
4. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: MIT Press; 2004.
5. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE international conference on computer vision (ICCV). IEEE; 2017. p. 618–26. https://doi.org/10.1109/ICCV.2017.74.