



Commentary

Bring the model to the data: The Deep Learning Epilepsy Detection Challenge

Annika Reinke

German Cancer Research Center DKFZ, Division of Computer Assisted Medical Interventions, Heidelberg, Germany



ARTICLE INFO

Article History:

Received 19 March 2021

Accepted 19 March 2021

Clinical trials are the state-of-the-art methods to assess the effect of a new medication or treatment in a comparative manner. In the machine learning community, the equivalent are international competitions, so-called challenges. Challenges aim to compare the performance of various algorithms under identical conditions, especially using the exact same data set for validation. They are an important tool to address clinical questions, compare method performances and assess algorithms' capabilities of being translated into clinical practice. Similar to clinical trials, transparent reporting of the challenge's design and outcome is the key aspect for a challenge to be interpretable and of clinical value.

The development of machine learning algorithms requires large amounts of data [1]. While this is possible for common computer vision approaches, acquiring large data sets in the biomedical domain is a difficult task. Patient data is extremely sensitive and its usage therefore strictly regulated [2]. This puts an additional burden on researchers and tremendously complicates the task of algorithm development. Federated learning is a concept trying to address this problem "by training algorithms collaboratively without exchanging the data itself" [3]. This research area recently attracted the interest of the biomedicine community for several applications, for instance for brain tumor segmentation [4], predictions from electronic health records [5] or EEG signal classification [6].

The article by Roy et al. in this issue of *EBioMedicine* [7] tackles the problem of data privacy with a comparable approach for their Deep Learning Epilepsy Detection Challenge. The authors present a "model-to-data" platform, implemented within the IBM cloud. Challenge participants developed their methods within the platform without being able to see the data at any point. Algorithm training and validation was enabled by direct feedback in the shape of an intermediate scoring, allowing for further tuning and development. The final model submission was evaluated against a held-out test set,

leading to the final ranking. Besides honoring privacy concerns, this approach embraced further advantages: Overfitting to the training data could easily be avoided as well as potential cheating and tuning of results by participants, such as manually revising submitted test cases. Furthermore, the platform facilitated participation in the challenge, as only basic programming and domain knowledge was required. The platform granted easy algorithm development by providing starter kits for modeling and for pre- and post-processing steps.

Although data scarcity and algorithm tuning are well-known problems in the biomedical community, surprisingly few challenges apply such an approach as presented in the article. However, in past years, some challenges introduced the concept of docker container submission to keep the test set private, while the training set was still provided to the participants (e.g., [8]). The presented model-to-data platform has the advantage of keeping the complete data set private. Although participants' domain knowledge cannot directly be applied without seeing the data, recent research has indicated that model-to-data approaches can lead to comparable performances to methods trained with public data [4].

Besides their excellent platform, the authors present the results of their challenge, which aimed to automatically detect seizures from EEG data. They followed the recommendations of the Biomedical Image Analysis ChallengeS Cha(BIAS) guideline for biomedical challenges [9] and reported their challenge design in a transparent fashion by presenting the organization, data sets, metrics, results and other important aspects of the challenge. The first run of the challenge only included participants from IBM locations. Therefore, the impact of the challenge could potentially be further increased by reorganizing the challenge within a broader audience, for example on a conference like the International Symposium on Biomedical Imaging (ISBI).

A key goal of the competition was the focus on clinically relevant properties such as high sensitivity and low false alarm rates. Although many challenges target clinical questions, the actual goals are often not well reflected in the performance metrics. However, the authors of this article chose the evaluation metrics to reflect the clinical properties needed and restricted the eligibility for the final scoring by defining a necessary minimum sensitivity goal.

The best performing algorithms were able to reduce the amount of raw data to be reviewed by human experts by a substantial factor of 142. The authors therefore conclude that deep learning systems should be used in combination with the usual manual human

DOI of original article: <http://dx.doi.org/10.1016/j.ebiom.2021.103275>.E-mail address: a.reinke@dkfz.de<https://doi.org/10.1016/j.ebiom.2021.103323>2352-3964/© 2021 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

labeling processes, leading to a heavily reduced time burden for manual annotations. This is an exemplary way of practically implementing machine learning algorithms in clinical practice, making use of their computational strengths and overcoming their weaknesses in accuracy through human support.

Declaration of Competing Interest

The author has no conflict of interest to disclose.

Acknowledgments

The author is supported by the Helmholtz Association of German Research Centers in the scope of the Helmholtz Imaging Platform (HIP).

References

- [1] Syeda-Mahmood T. Role of big data and machine learning in diagnostic decision support in radiology. *J Am Coll Radiol* 2018;15(3):569–76.
- [2] Van Panhuis WG, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014;14(1):1–9.
- [3] Rieke N, et al. The future of digital health with federated learning. *NPJ Digit Med* 2020;3(1):1–7.
- [4] Li W, et al. Privacy-preserving federated brain tumour segmentation. In: *Proceedings of the international workshop on machine learning in medical imaging*. Cham: Springer; 2019.
- [5] Brisimi TS, et al. Federated learning of predictive models from federated electronic health records. *Int J Med Inform* 2018;112:59–67.
- [6] Ju Ce, et al. Federated transfer learning for EEG signal classification. In: *Proceedings of the 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*. IEEE; 2020.
- [7] Roy S, et al. Evaluation of an artificial intelligence system for assisting neurologists with fast and accurate annotation of scalp electroencephalography data. *EBioMedicine* 2021. doi: 10.1016/j.ebiom.2021.103275.
- [8] Roß T, et al. Comparative validation of multi-instance instrument segmentation in endoscopy: results of the ROBUST-MIS 2019 challenge. *Med Image Anal* 2020;101920.
- [9] Maier-Hein L, et al. BIAS: transparent reporting of biomedical image analysis challenges. *Med Image Anal* 2020;66:101796.