Collaborative learning without sharing data

Accurate and fair medical machine learning requires large amounts and diverse data to train on. Privacy-preserving methods such as federated learning can help improve machine learning models by making use of datasets in different hospitals and institutes while the data stays where it is collected.

One of the most promising and active application areas for machine learning is in healthcare, where trained models can, for example, help clinicians make diagnoses based on medical images or other healthcare data, or decide on treatment plans. However, an increasingly raised concern[1] is that machine learning models often do not go beyond the research development phase and do not translate well outside the settings of the model development. Typically, machine learning models are trained and evaluated on a limited amount of data with known, as well as unknown, biases.

This problem has been observed most notably in the context of COVID-19 research, which has seen a large number of reports of quickly developed diagnostic and prognostic machine learning models in the past year. A recent Analysis[2] surveyed many of these reports and found that they often made use of the same publicly available datasets. This data monoculture can become problematic in cases of 'shortcut learning'[3], where datasets that are collected by a single, or only a few institutes contain confounding patterns, such as patient positioning in medical X-ray imaging. Machine learning models will exploit such 'shortcuts' and can appear to be very accurate — only to fail when tested on data from different hospitals.

Ideally, machine learning model developers can make use of many different patient datasets from around the world. However, the sharing of clinical databases requires significant care given the challenges in data protection and privacy issues. The large-scale aggregation and exchange of patient datasets comes with risks, as in recent years it has become clear that re-identification attacks[4] on anonymized health data can be very effective, in particular by cross-referencing between different datasets.

One line of approach to make safer use of datasets from different hospitals and institutes is to implement federated learning schemes, where data remains in the institute where it is collected, and instead a machine learning model is shared and collaboratively trained.

The term 'federated learning' was coined in 2016[5] and originally developed for improving AI systems in consumer applications, such as Siri and Google Assistant. In these approaches, AI algorithms learn locally on mobile devices — 'on the edge' — and send model updates, rather than the data, back to a central server.

The potential for leveraging such schemes in medical applications, so that models can learn from data at different institutes without sharing data, was quickly realized[6]. Several data protection challenges would be resolved if machine learning models were trained on multi-institutional patient and medical datasets while keeping data decentralized. However, federated learning is not immune to risk of attacks, and several encryption techniques are still necessary to apply federated learning to practical, real-world applications with large clinical datasets.

In an Article in this issue, Braren and colleagues report a framework called PriMIA (Privacy-preserving Medical Image Analysis) for secure federated learning with encrypted inference on medical imaging data. They demonstrate it in a case study, training an 11.1 million parameter ResNet18 convolutional neural network over public Internet on the paediatric pneumonia dataset by Kermany and colleagues[7]. The model learns the challenging task of classifying images into normal (no sign of infection), viral pneumonia and bacterial pneumonia.

Although the data in this proof-of-principle study is fully available for research purposes, it is used here in a mock setting in which three hospitals want to collaborate to train a diagnostic model without sharing sensitive patient data. Using differential privacy, each hospital can share updates to the model with a reduced risk that other parties can infer sensitive information from this data. The combined model is then trained separately in each hospital and the updates are shared again, repeating the cycle until the performance for each of the collaborating partners is satisfactory. While this procedure imposes some additional cost to the training process compared to a model that is trained at a single institution, the increased access to data can lead to much more robust models.

Further work needs to validate the approach on real-world tasks involving several clinics. The framework is open source so it can be extended to other use cases or adapted to the requirements of a specific group of hospitals.

Federated schemes can be extended to full decentralization, where the machine learning model is not shared either, promising full confidentiality. Such an approach is demonstrated by an international team of researchers in a recent paper in *Nature*[8]. They report on the feasibility of collaboratively developing disease classifiers for detecting several diseases including COVID-19, tuberculosis and leukaemia. Rather than aggregating updates to a model in a central server, an audit trail, or blockchain, contains all information and is exchanged between the different partners, the data holders.

Performing computations locally rather than in data centres could be good news for efficiency. On the other hand, such distributed learning approaches are inherently iterative processes, which could be costly in terms of learning time and energy consumption. As for all machine learning developments, in future work it will be important to evaluate and report on computational costs. ❐

Published online: 17 June 2021
https://doi.org/10.1038/s42256-021-00364-5

References

1. Hu, Y. et al. *Nat. Mach. Intell.* **2**, 298–300 (2020).
2. Roberts, M. et al. *Nat. Mach. Intell.* **3**, 199–217 (2021).
3. DeGrave, A. J., Janizek, J. D. & Lee, S.-I. *Nat. Mach. Intell.* https://doi.org/10.1038/s42256-021-00338-7 (2021).
4. El Emam, K., Jonker, E., Arbuckle, L. & Malin, B. *PLOS ONE* **6**, e28071 (2011).
5. Kairouz, P. & McMahan, H. B. (eds) *Advances and Open Problems in Federated Learning* (Now, 2021).
6. Rieke, N. et al. *npj Digit. Med.* **3**, 119 (2020).
7. Kermany, D. S. et al. *Cell* **172**, 1122–1131 (2018).
8. Warnat-Herresthal, S. et al. *Nature* https://doi.org/10.1038/s41586-021-03583-3 (2021).