Hautarzt 2022 · 73:323–325

<https://doi.org/10.1007/s00105-021-04940-z>

Accepted: 22 December 2021

Published online: 14 January 2022

© The Author(s), under exclusive licence to Springer Medizin Verlag GmbH, ein Teil von Springer Nature 2022

Rubrikherausgeber*in

Natalia Kirsten, Hamburg

Alexander Zink, München

Swarm learning for decentralized healthcare

Matthias Becker

Modulares Hochleistungsrechnen und künstliche Intelligenz, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany

Introduction

Machine learning is revolutionizing medicine by enabling novel applications as well as supporting physicians in routine tasks. These techniques rely on large datasets and collections for development, which are hard to acquire in the decentralized world of healthcare. Privacy and data safety are challenges which slow development in the medical field in comparison to others. Even if a sufficiently large dataset can be created, it is only a static snapshot and cannot reflect upcoming diseases. Machine learning is playing a growing role in dermatology [1, 2] and will become a major technology to assist (and not replace) physicians to improve the quality of care and reduce workload. Medicine itself is inherently based on learning from each other. The long-established mentor principle during training is a good example. Similarly, tools that support physicians in using machine learning techniques need to be trained as well. Unlike humans, these tools do not have a general understanding of the world (like a strong artificial intelligence) and need to compensate that by learning from a large number of training datasets. The performance of a machine learning model correlates with the amount of training data. Therefore, acquiring ever-growing training datasets is crucial. However, data privacy laws and patient consent limit the data collection and impact model performance.

What is new?

Training separate models at different sites, e.g., per hospital, was the first approach during the establishment of the use of artificial intelligence in medicine. To build more robust models or models for diseases with only few cases, central resources like a cloud [3] have been used. Here, the training data were accumulated at a central point and a single model was trained. While this offers improved model performance, increased interest in data privacy have made such centralized data accumulation more difficult. Laws like the US Health Insurance Portability and Accountability Act (HIPAA) or the EU General Data Protection Regulation (GDPR) [4] further limit this approach. Federated machine learning techniques [5] have been established to overcome this central player. Here, the training is performed at the participating sites, so called nodes, using only local data and the resulting local models are merged by a central instance. In this setup, all nodes need to trust the central player for the model parameter merging and distribution.

Swarm learning (SL) [6] aims to overcome the dependency on central components or participants. Unlike the previous approaches, there is no central player collecting either data or trained models. In SL, the nodes (participants in the swarm) train an agreed-upon machine learning model with their local data. After a specified number of training steps, the trained model parameters are shared with the swarm, merged, and then redistributed

Scan QR code & read article online

to all nodes. Merging parameters can be achieved by simple averaging, more complex functions, or code that might weight the different nodes by their size. The merging will be done by a single node; however, for each iteration, a new node is elected for this task. This removes the dependency on a central player and the risk of such a node manipulating the model parameters. A distributed ledger (based on the Ethereum blockchain) enforces a smart contract between the nodes. This contract can specify the conditions for joining and finishing a swarm experiment. Furthermore, it tracks the contributions of each node for proper attribution and retraceability of the learning process. The continuous updates between the nodes generate a shared model that encompasses the properties of all datasets of the nodes, e.g., device properties or study biases. Unlike a model that was trained in isolation, no transfer learning is needed to apply the swarm model to new datasets. To join the swarm, a node agrees first to the conditions and then begins training on the local data. Once the swarm has finished an iteration, all nodes share their model parameters and receive the merged parameters afterwards. The new node now has the same model parameters as all other nodes in the swarm and continues training on these. A swarm experiment does not require all nodes to have the same size or properties, e.g., in a multi-disease classifier, not all nodes need to have training data for all diseases.

Initially, swarm learning was tested on the classification of COVID-19 in transcriptomic profiles [6]. In these experiments, the swarm always showed better than or similar performance to individual nodes. These results have been demonstrated on swarms with up to six nodes and different prevalences at the nodes. Furthermore, two different techniques for acquiring the transcriptomic profiles were used at different nodes. The swarm was able to robustly merge the input from them. Next, it was shown that swarm learning works on a completely different data space as well by using a popular, openly available on Kaggle, X-ray chest image dataset [7] to successfully classify different lung diseases. Such an application uses the same underlying techniques as dermatological

applications as demonstrated in [1]. Swarm learning has also been successfully applied by other groups [8] in analyzing pathology images.

Health data is inherently distributed as each patient can be considered a single data source. Local health care centers, doctor's offices, or hospitals accumulate individual silos of data in many different modalities [2], including imaging, molecular profiles, and omics together with patient records, phenotypes, and diagnoses. Using traditional approaches, they cannot be easily pooled, but to advance the fields of personalized and precision medicine, these silos need to be tapped into. Swarm learning allows joint learning on this vast data pool without compromising patient data and has been shown to integrate different modalities and machine learning approaches. Tapping into all silos provides a dataset that helps to overcome study biases and differences from acquisition methods. Furthermore, rare diseases with insufficient local training data can be more robustly detected. A large swarm could continuously monitor health data from different countries worldwide and help in the early detection of pandemics [9]. Overall, medical research and clinical applications will benefit from swarm learning as collaboration can be done faster with reduced legal overheads, as data are not being shared.

So, how can swarm learning be used? There are two cases to be considered: first, you have a use case and want to initiate a swarm experiment, and second, you have data and want to join an existing swarm experiment. To create a swarm experiment, it is possible to follow the same steps as for a local machine learning experiment. For a specific use case, a fitting machine learning approach needs to be selected and sufficient initial training data are needed. The training data must be checked for their general properties and potential biases. If the data can be expected to be different at other nodes (e.g., different data acquisition protocols, devices), such data should also be included in training to estimate the generalization of the model. When performing the local training, the overall goal is not to get a perfect model, as only local data are used. It is rather intended to test the feasibility of the selected model

configuration. Next, the model needs to be swarm enabled. This is achieved by adding the swarm library and the necessary callbacks, which can be found in the documentation as well as in the examples [10]. Afterwards, the swarm needs to be deployed and as soon as the minimal number of peers have joined, the training will begin.

To join a swarm experiment, local data need to be prepared and then the swarm can be joined. The swarm experiment initiator defines the data requirements, which can be minimum quality requirements or resizing image data to a specific resolution. Additionally, the swarm-enabled model code will be distributed and only needs to be adjusted to load the correct data. Then the swarm can be joined by training the model.

All participants in the swarm will receive the merged model parameters after each iteration; this way everyone is using the same initial parameters for the next iteration. The smart contracts in the distributed ledger control start and stop conditions and log the contributions by each node, which makes the training transparent and accountable.

Conclusion for practice

Swarm learning is a novel collaborative machine learning approach to leverage health data acquired in decentralized environments. By using the power of the swarm, smaller nodes with fewer datasets can also contribute and obtain a robust model. The privacy-preserving properties of SL protect patient data. By exploiting an existing local machine learning model, new swarm experiments can be set up fast and with minimal code modifications, and joining an experiment only requires preparing the data.

Corresponding address

Dr. Matthias Becker

Modulares Hochleistungsrechnen und künstliche Intelligenz, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE)
Venusberg-Campus 1/99, 53127 Bonn,
Germany
matthias.becker@dzne.de

Declarations

Conflict of interest. M. Becker declares that he has no competing interests.

For this article no studies with human participants or animals were performed by any of the authors. All studies performed were in accordance with the ethical standards indicated in each case.

References

1. Liu Y, Jain A, Eng C et al (2020) A deep learning system for differential diagnosis of skin diseases. *Nat Med* 26:900–908. <https://doi.org/10.1038/s41591-020-0842-3>
2. Hogarty DT, Su JC, Phan K et al (2020) Artificial intelligence in dermatology—where we are and the way to the future: a review. *Am J Clin Dermatol* 21:41–47. <https://doi.org/10.1007/s40257-019-00462-6>
3. Dove ES, Joly Y, Tassé AM, Knoppers BM (2015) Genomic cloud computing: legal and ethical points to consider. *Eur J Hum Genet* 23:1271–1278
4. McCall B (2018) What does the GDPR mean for the medical community? *Lancet* 91:1249–1250
5. Kaissis GA, Makowski MR, Rückert D, Braren RF (2020) Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell* 2:305–311
6. Warnat-Herresthal S, Schultze H, Shastri KL et al (2021) Swarm learning for decentralized and confidential clinical machine learning. *Nature* 594:265–270. <https://doi.org/10.1038/s41586-021-03583-3>
7. Wang X et al (2017) ChestX-Ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR). IEEE, pp 3462–3471
8. Lester Saldanha O, Quirke P, West NP et al (2021) Swarm learning for decentralized artificial intelligence in cancer histopathology. *bioRxiv*. <https://doi.org/10.1101/2021.11.19.469139>
9. Luengo-Oroz M et al (2020) Artificial intelligence cooperation to support the global response to COVID-19. *Nat Mach Intell* 2:295–297
10. Hewlett Packard Enterprise (2021) A simplified library for decentralized privacy preserving machine learning. <https://github.com/HewlettPackard/swarm-learning>. Accessed 22 Dec 2021

Möchten Sie einen Beitrag für die Zeitschrift *Der Hautarzt* einreichen?

Wir freuen uns, dass Sie unsere Zeitschrift *Der Hautarzt* mitgestalten möchten. Um Ihnen bei der Manuskripterstellung behilflich zu sein, haben wir für unsere Autor*innen Hinweise zusammen gestellt, die Sie im Internet finden unter www.DerHautarzt.de (For Authors).

Bitte senden Sie Ihren fertigen Beitrag an:

Originalien/Übersichten/ Kasuistiken/Wie lautet Ihre Diagnose?

Prof. Dr. Daniela Bruch-Gerharz
Hautklinik des Universitätsklinikums
Düsseldorf, Moorenstr. 5,
40225 Düsseldorf
sdgehrke@googlemail.com

Dermatoskopie – Fall des Monats:

Prof. Dr. Andreas Blum
Hautarztpraxis Konstanz und Lehrpraxis der
Medizinischen Fakultät Universität Tübingen
Augustinerplatz 7
78462 Konstanz
a.blum@derma.de

One Minute Wonder:

PD Dr. Dr. Alexander Zink, Dr. Sarah Preis
Klinik und Poliklinik für Dermatologie und
Allergologie am Biederstein,
TU München

Bitte schreiben Sie bei Fragen an
die Verlagsredaktion:
julie.kind@springer.com

CME Zertifizierte Fortbildung:

Anfragen an:

Prof. Dr. Evelyn Gaffal
Universitätsklinik
Universitätsklinikum Magdeburg
Labor für Experimentelle Dermatologie
Leipziger Straße 44
39120 Magdeburg
evelyn.gaffal@med.ovgu.de

Prof. Dr. Sonja Ständer
Klinik und Poliklinik für Hautkrankheiten,
Universitätsklinikum Münster
Von-Esmarch-Str. 58
48149 Münster
sonja.staender@uni-muenster.de

Prof. Dr. Rolf-Markus Szeimies
Klinik für Dermatologie und Allergologie,
Knappschafts Krankenhaus Recklinghausen,
Klinikum Vest GmbH,
Dorstener Str. 151,
45657 Recklinghausen
dermatologie@kk-recklinghausen.de

PD Dr. Dr. Alexander Zink
Klinik und Poliklinik für Dermatologie und
Allergologie am Biederstein,
Technische Universität München
Biedersteiner Str. 29
80802 München
alexander.zink@tum.de