FONDAZIONE
PER LA RICERCA FARMACOLOGICA
GIANNI BENZI
ONLUS

December 10th, 2021
Virtual meeting

XIV FORESIGHT TRAINING COURSE
The health emergency: regulatory crash and future perspectives

Federated Learning as a Tool for Gathering Knowledge from Multiple Data Sources

Dr. George Drosatos

Researcher, Institute for Language and Speech Processing,
Athena Research Center, Greece
https://www.drosatos.info .gr .eu

ATHENA
Research & Innovation
Information Technologies

Fondazione per la Ricerca Farmacologica Gianni Benzi onlus

Via Abate Eustasio, 30 – 70010 Valenzano (BA) Tel.: +39 080 2052499
www.benzifoundation.org

FONDAZIONE
PER LA RICERCA FARMACOLOGICA
GIANNI BENZI
ONLUS

Presentation overview

- What is federated learning?
- How does traditional centralized machine learning work?
- How does federated learning work in general?
- Federated learning variants
- How can the provided privacy be improved?
- Why is it an excellent tool in the field of health?
- Examples of use cases in digital health

FONDAZIONE
PER LA RICERCA FARMACOLOGICA
GIANNI BENZI
ONLUS

What is federated learning?

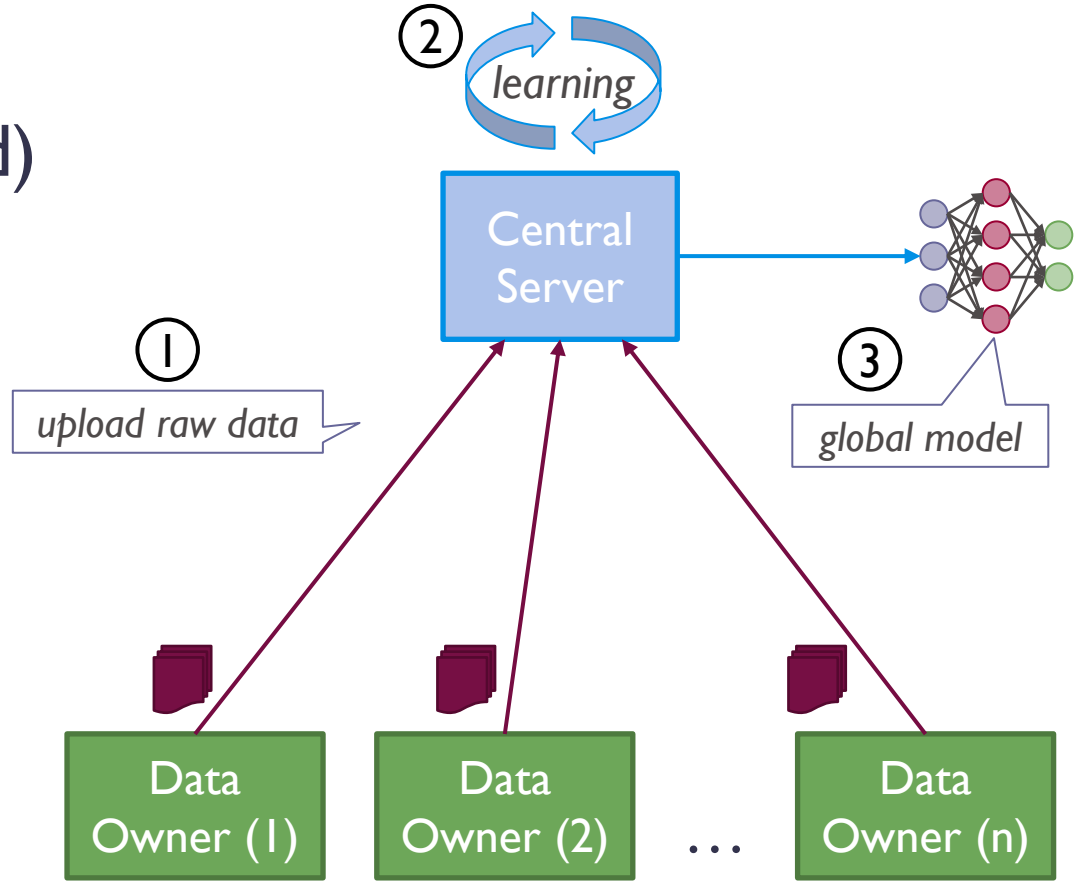
- It is a machine learning technique that trains an algorithm across multiple decentralized local data samples without exchanging them
- Local data samples can be located on edge devices (e.g., mobile phones) or servers representing specific entities (e.g., hospitals)
- It enables multiple actors to build a common, robust machine learning model without sharing data, allowing to address issues such as:
  - data privacy, data security, data access rights, and access to heterogeneous data
- Currently, its applications are spread over a number of industries including:
  - defense, telecommunications, IoT, and digital health

FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

How does traditional centralized machine learning work?

Basic steps in a centralized machine learning approach:

1. Local raw data is uploaded by data owners to a central server (e.g., cloud)
2. The central server performs machine learning for a specific prediction task*
3. The global model is ready for use by any third party



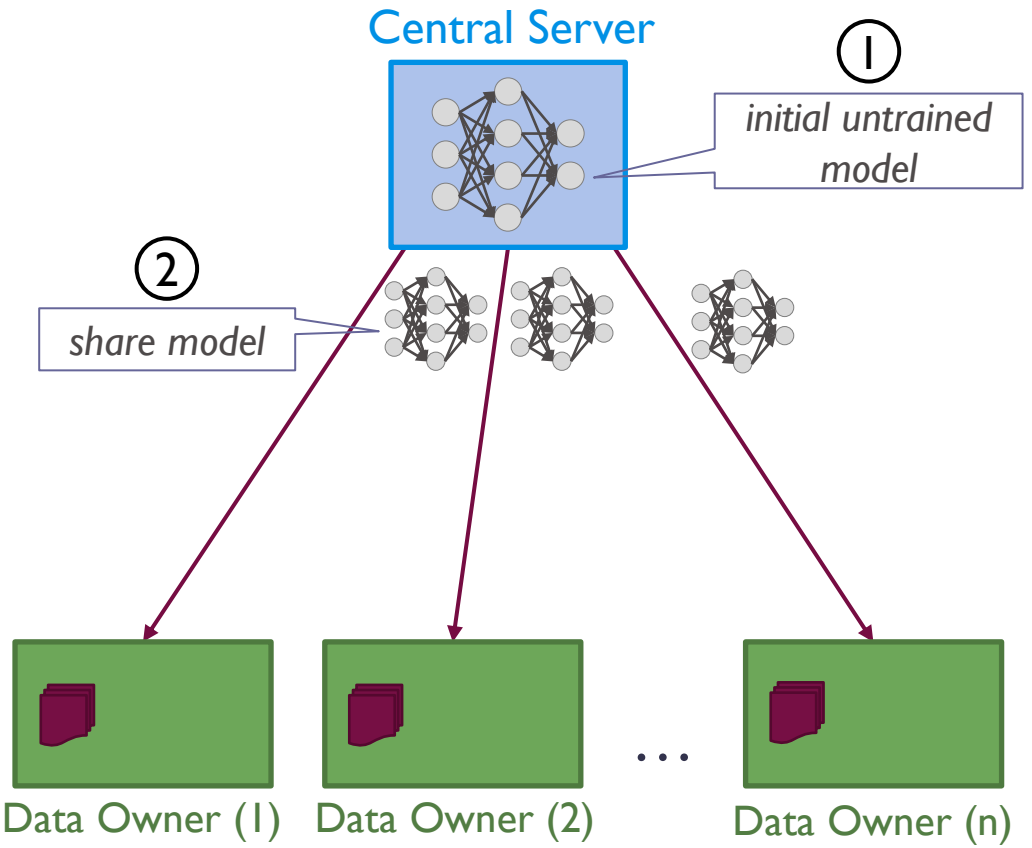
*e.g., *classification, regression, recommendation, decision support, etc.*

FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

How does federated learning work in general?

Basic steps in a federated learning approach:

1. The central server initiates an untrained model
2. This model is shared with data owners

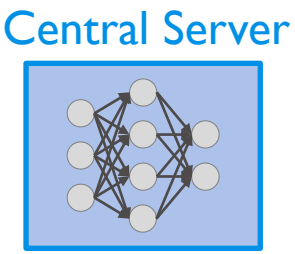


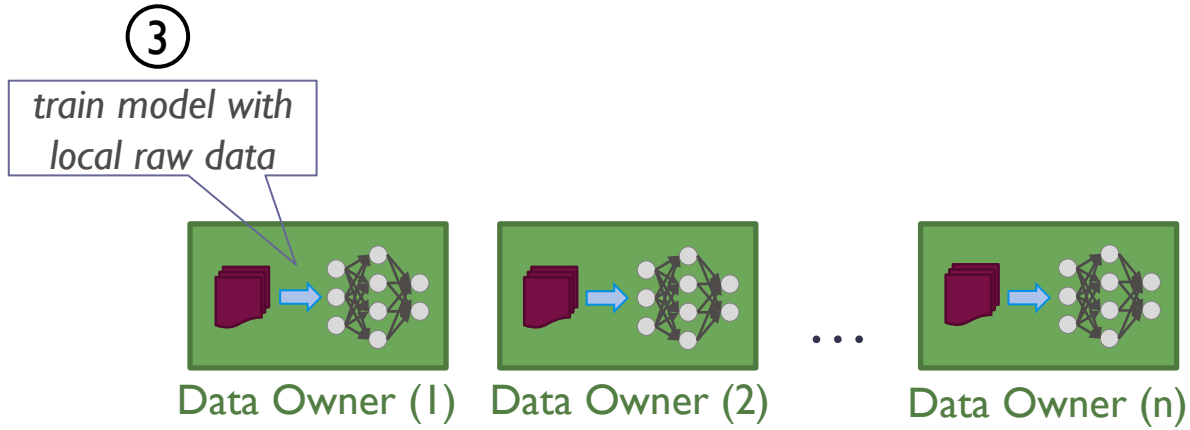
FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

How does federated learning work in general?

Basic steps in a federated learning approach:

1. The central server initiates an untrained model
2. This model is shared with data owners
3. The data owners train the model with local raw data



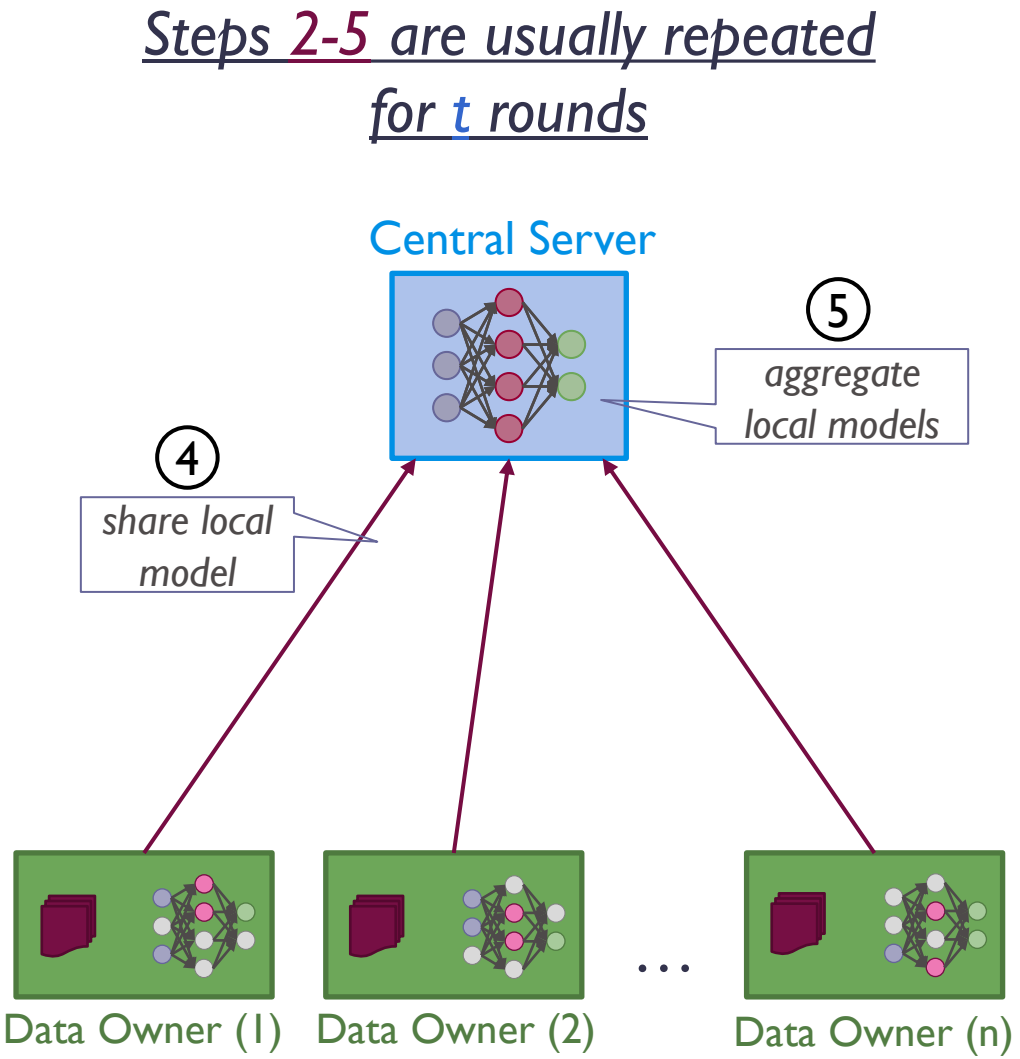


FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

How does federated learning work in general?

Basic steps in a federated learning approach:

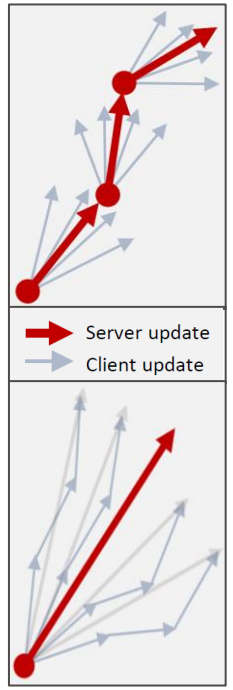
1. The central server initiates an untrained model
2. This model is shared with data owners
3. The data owners train the model with local raw data
4. The local models are shared with the central server
5. The central server aggregates local models into a global model



FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

Federated learning variants

- The exact aggregation method can be differentiated. There are two main approaches:
  - **Gradient-based aggregation (e.g., FedSGD):** The gradients are computed using a random fraction C of the nodes and using all the data on this node. Then, the gradients are averaged by the server proportionally to the number of training samples on each node.
  - **Weights-based aggregation (e.g., FedAvg):** It is a generalization of gradient-based aggregation, which allows local nodes to perform more than one batch update on local data and exchanges the updated weights rather than the gradients.
- There are approaches that do not require a central server



FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

How can the provided privacy be improved?

- The main advantage of using federated approaches to train machine learning models is to ensure data privacy or data secrecy
- However, the local models that are shared with the central server, or even the learning parameters, leak information about the underlying local data
- To avoid such leaks, federated learning is usually combined with the following privacy techniques:
  - **Differential privacy (DP):** This technique adds noise during the training process (e.g., on the local data, local training, or even after aggregation step). Thus, it is more difficult to distinguish the real raw data
  - **Homomorphic encryption (HE):** The learning parameters are encrypted before sharing between learning rounds and it is possible to make computations on the encrypted data without decrypting them beforehand
  - **Secure multi-party computation (SMC):** It is a cryptographic method, such as secure aggregation, that allows parties to jointly compute a function over their inputs while keeping those inputs private

FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

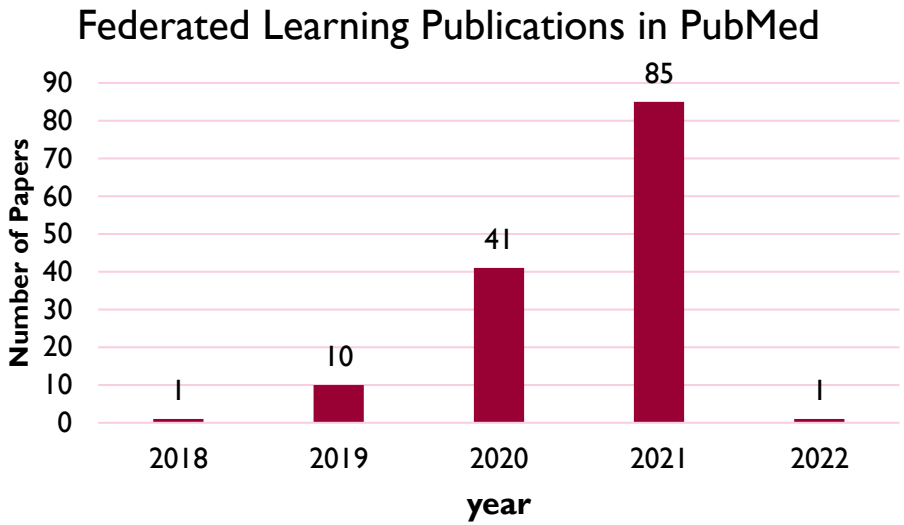
Why is it an excellent tool in the field of health?

- Federated learning seeks to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself
- Today's standard approach of centralizing data from multiple health centers comes at the cost of critical concerns regarding patient privacy and data protection
- To solve this problem, the ability to train machine learning models at scale across multiple medical institutions without moving the data is a critical technology that is possible using federated learning
- The processing of this sensitive data, and in compliance with the GDPR, is done by the health centers that manage the patient data and for reasons of public welfare

FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

Examples of use cases in digital health

- Identifying potential risk variants in ankylosing spondylitis utilizing genotyping data of patients from cross-institutional partnerships [1]
- COVID-19 diagnosis using region segmentation recognition on chest CT scans from multi-national data [2]
- Predict mortality in hospitalized patients with COVID-19 within 7 days using clinical data across multiple institutions [3]
- Detecting medication errors in the general internal medicine clinics [4]
- Predicting adverse drug reactions on distributed health data [5]
- Collaborative drug discovery among pharmaceutical institutions [6]



Search: "federated learning" [Title/Abstract]

FONDAZIONE PER LA RICERCA FARMACOLOGICA GIANNI BENZI ONLUS

References

1. Wu, X., Zheng, H., Dou, Z., Chen, F., Deng, J., Chen, X., ... & Xu, H. (2021). A novel privacy-preserving federated genome-wide association study framework and its application in identifying potential risk variants in ankylosing spondylitis. Briefings in Bioinformatics, 22(3), bbaa090. doi: 10.1093/bib/bbaa090
2. Yang, D., Xu, Z., Li, W., Myronenko, A., Roth, H. R., Harmon, S., ... & Xu, D. (2021). Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan. Medical image analysis, 70, 101992. doi: 10.1016/j.media.2021.101992
3. Vaid, A., Jaladanki, S. K., Xu, J., Teng, S., Kumar, A., Lee, S., ... & Glicksberg, B. S. (2021). Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: Machine learning approach. JMIR medical informatics, 9(1), e24207. doi: 10.2196/24207
4. Chin, Y. P. H., Song, W., Lien, C. E., Yoon, C. H., Wang, W. C., Liu, J., ... & Bates, D. W. (2021). Assessing the international transferability of a machine learning model for detecting medication error in the general internal medicine clinic: Multicenter preliminary validation study. *JMIR Medical Informatics*, *9*(1), e23454. doi: 10.2196/23454
5. Choudhury, O., Park, Y., Salonidis, T., Gkoulalas-Divanis, A., & Sylla, I. (2019). Predicting adverse drug reactions on distributed health data using federated learning. In *AMIA Annual symposium proceedings* (Vol. 2019, p. 313). American Medical Informatics Association. PMID: 32308824
6. Chen, S., Xue, D., Chuai, G., Yang, Q., & Liu, Q. (2020). FL-QSAR: A federated learning-based QSAR prototype for collaborative drug discovery. Bioinformatics, 36(22-23), 5492-5498. doi: 10.1093/bioinformatics/btaa1006

FONDAZIONE
PER LA RICERCA FARMACOLOGICA
GIANNI BENZI
ONLUS

EYXAPIΣTΩ/GRÀCIES/HVALA/DĚKUJI/TAK/DANK
JEWEL/AITÄH/KIITOS/MERCI/DANKE/KÖSZNÖNÖ/
GRAZIE/PALDIES/AČIŪ/GRAZZI/TAKK/DZIĘKUJĘ
OBRIGADO/MULTUMESC/СПАСИБО/ХВАЛА
HVALA/БЛАГОДАРЯ/THANKYOU/TAK/GRACIAS
/KIITOS/TACK/TEŞEKKÜREDERİM/СПАСИБІ/
JUFALEMINDERIT/EYXAPIΣTΩ/DANKJEWEL/TAK
TACK/GRAZZI/DANKJEWEL/MULTUMESC/AITÄH
KÖSZNÖNÖ/СПАСИБО/ХВАЛА/AČIŪ/THANKYOU

Thank you!

Fondazione per la Ricerca Farmacologica Gianni Benzi onlus

Via Abate Eustasio, 30 – 70010 Valenzano (BA) Tel.: +39 080 2052499
www.benzifoundation.org