Federated Medical Data

How much can Deep Learning Models benefit?

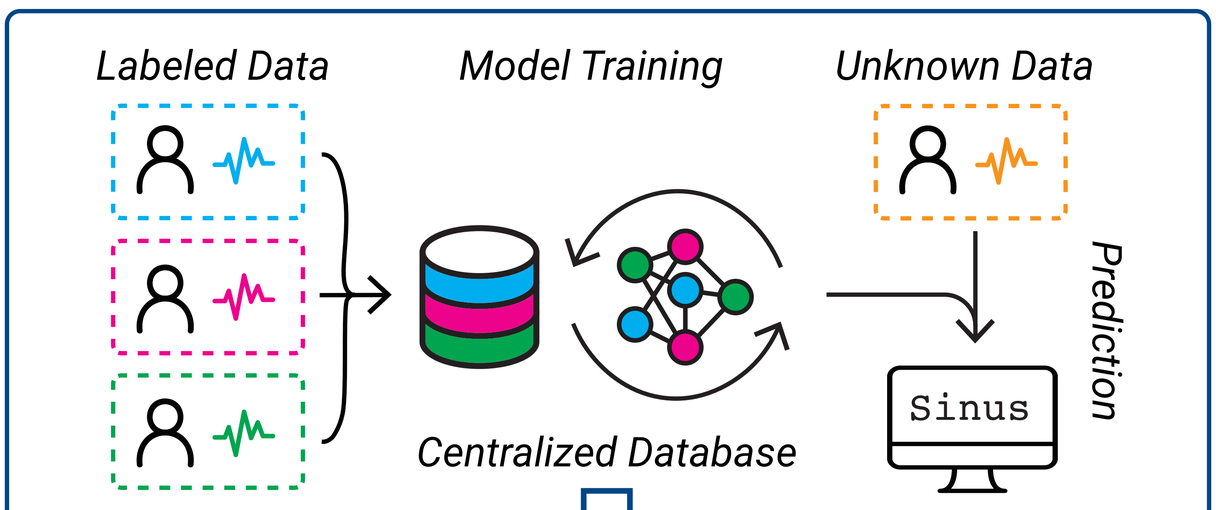
Universität Augsburg University

Software Methodologies for Distributed Systems

Fabian Rabe, Fabian Stieler, Bernhard Bauer

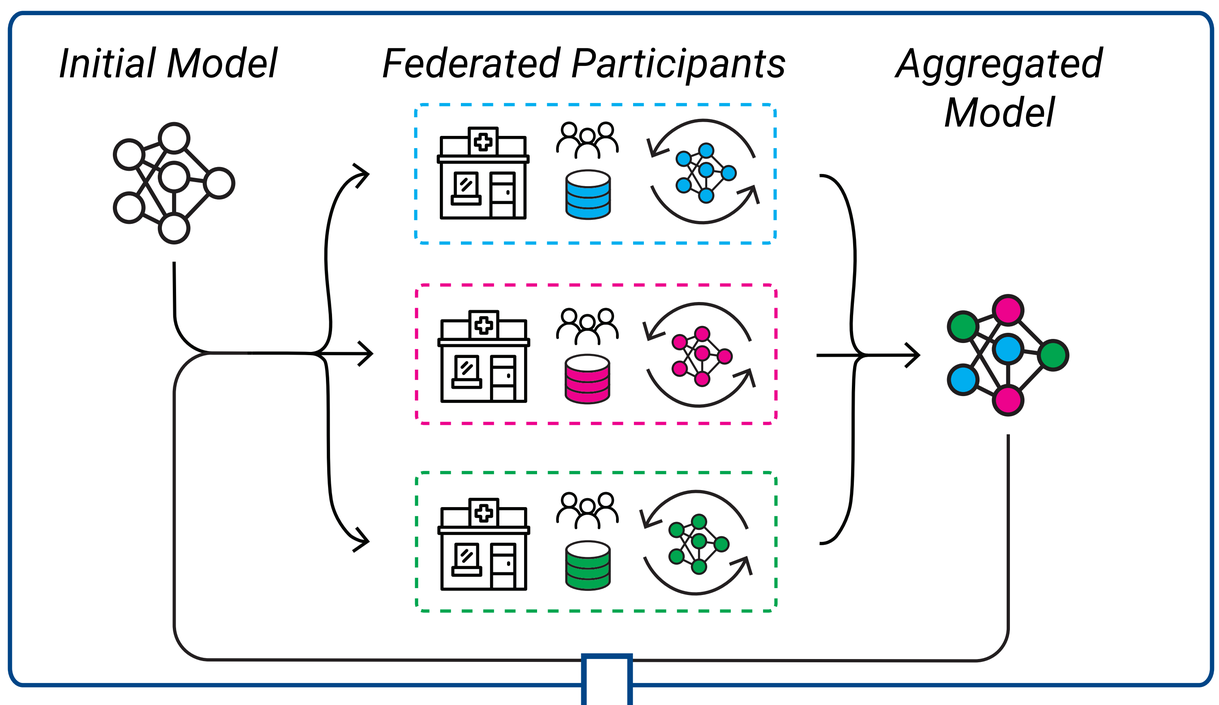
Background

Many machine learning algorithms, like **supervised Deep Learning**, assume that Training Data are available in a single database.



Privacy Risks
e.g., data from different institutions in a central database.[1]

Federated Learning[2] trains a model at each institution locally, aggregates and share only the model, not the patient data.



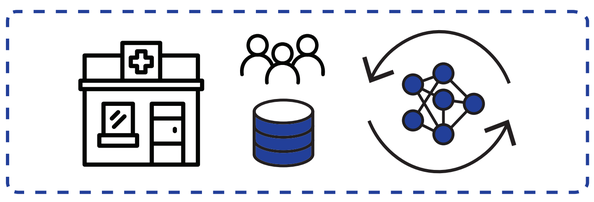
Federated Benefit
The iterative process allows each model to benefit from knowledge learned in other institutions, without sharing the data.

Method

Using the **TensorFlow Federated**[3] Framework and data from the **MIT-BIH Electrocardiogram**[4] database, we simulate two scenarios of an arrhythmia classifier, with:

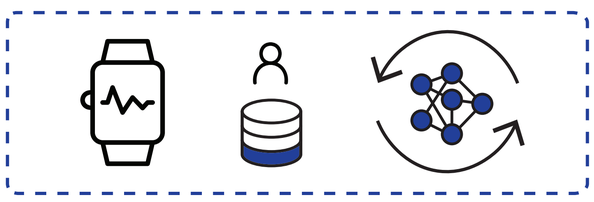
- different ammounts of *Federated Participants* (FP)s and
- different *patient data* per FP

Hospital Scenario



- 20 FPs (hospitals)
- 3,000 heart beats per FP
- **Independent and identically distributed data (IID):**
- different patients per FP
- different diagnosis per FP

Smartwatch Scenario



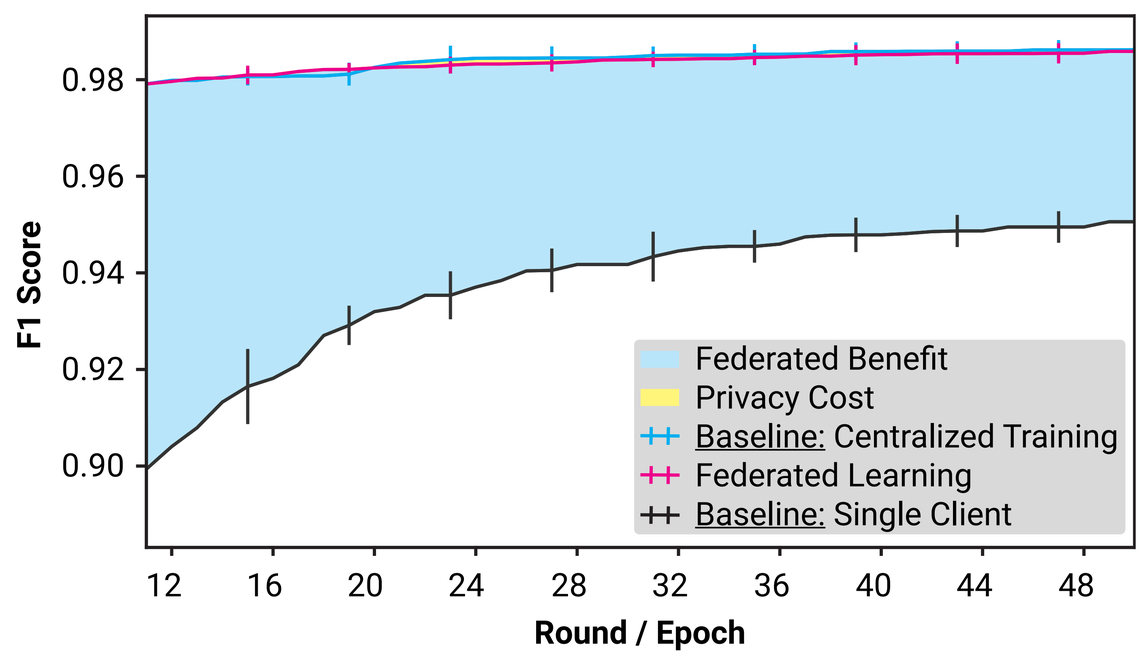
- 2,000 FPs (smartwatches)
- 30 heart beats per FP
- **Non-IID:**
- 1 patient per FP
- 95% FPs: sinus rhythms only

Results

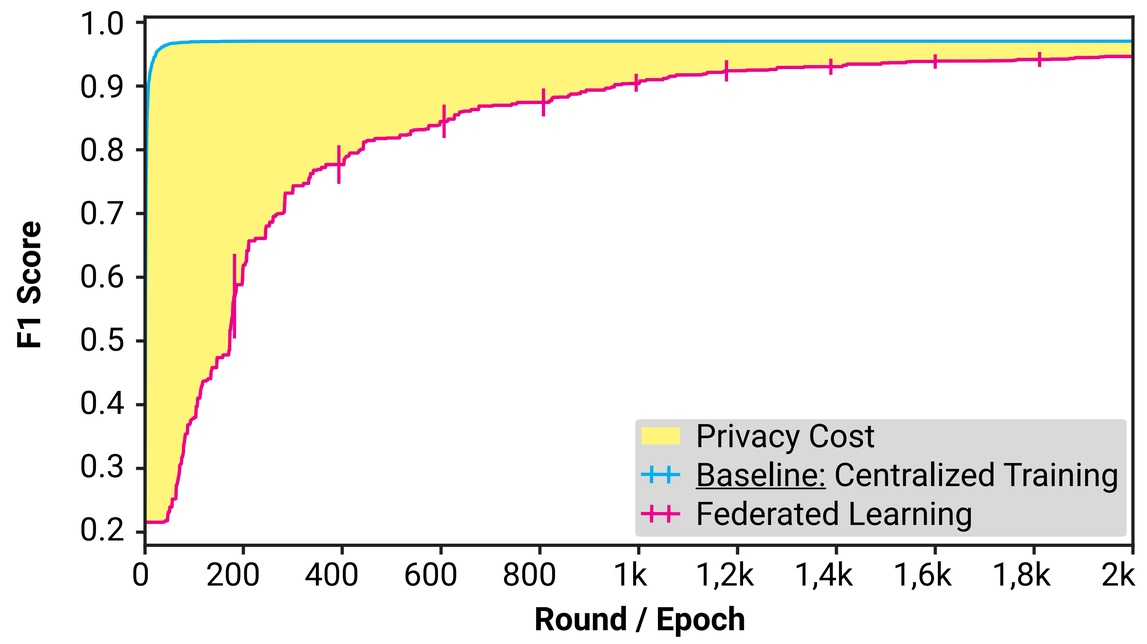
Model quality is measured via the F_1 score on a validation data set. We define

Privacy Cost (PC) := $F_{1(\text{centralized})}$ - $F_{1(\text{federated})}$
Federated Benefit (FB) := $F_{1(\text{federated})}$ - $F_{1(\text{single client})}$

Hospital Scenario: We achieve **PC of 0.03%** and **FB of 3.53%** after 49 federated rounds (training iterations).



Smartwatch Scenario: The **PC are 2.33%** after 1,999 training rounds. A model trained on a single smartwatch isn't feasible, since it contains not all considered arrhythmia-classes. Hence, the **FB is being able to learn at all**.



Conclusion

- From a machine learning perspective Federated Learning instead of centralized model training is **a promising option**.
- Models can be trained to **comparable performance** without the risk a centralised patient data base poses.
- There are scenarios in which federated medical data **enable a deep learning model**, which only emerge through the iterative model aggregation of various FPs.

References

[1] John (Xuefeng) Jiang and Ge Bai. "Evaluation of Causes of Protected Health Information Breaches". In: JAMA InternalMedicine 179 (Feb. 1, 2019), pp. 265–267.
[2] Brendan McMahan et al. "Communication-Efficient Learning of Deep Networks from Decentralized Data". In: vol. 54. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, Apr. 20, 2017, pp. 1273–1282.
[3] TensorFlow Federated. TensorFlow. url: https://www.tensorflow.org/federated (visited on 04/2020).
[4] G.B. Moody and R.G. Mark. "The impact of the MIT-BIH Arrhythmia Database". In: IEEE Engineering in Medicine and Biology Magazine 20 (June 2001), pp. 45–50.

AMIA 2020 Virtual Clinical Informatics Converence,
May 19-21