



Universität Hamburg  
DER FORSCHUNG | DER LEHRE | DER BILDUNG



Source: <https://www.rmit.edu.au/study-with-us/biomedical-sciences>

Julian Matschinske, M.Sc. Computer Science, UHH

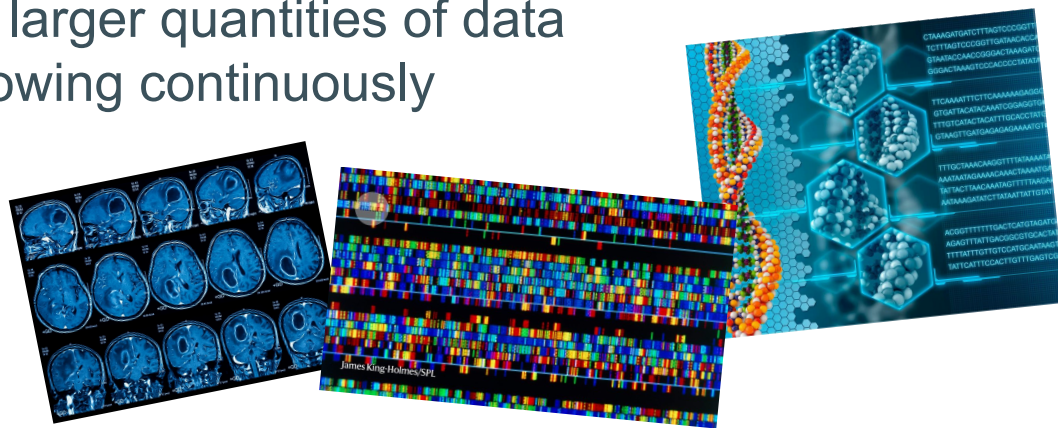
---

# Federated Learning - Introduction

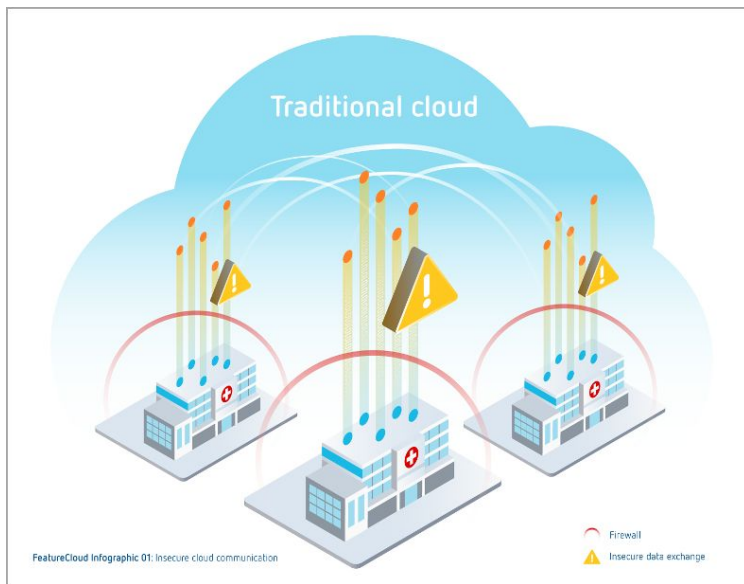
ISMB 2022, Federated Learning in Biomedicine (Tutorial)

# Large-scale AI in Biomedicine

- Machine learning (ML) allows for classification and predictions in different domains
- Benefits hugely from larger quantities of data
- Amount of data is growing continuously



# Large-scale AI in Biomedicine - Classical Approach



- Merge data into big datasets
- Train ML models on the whole set
- Share models with data contributors

# Problem: Privacy regulation



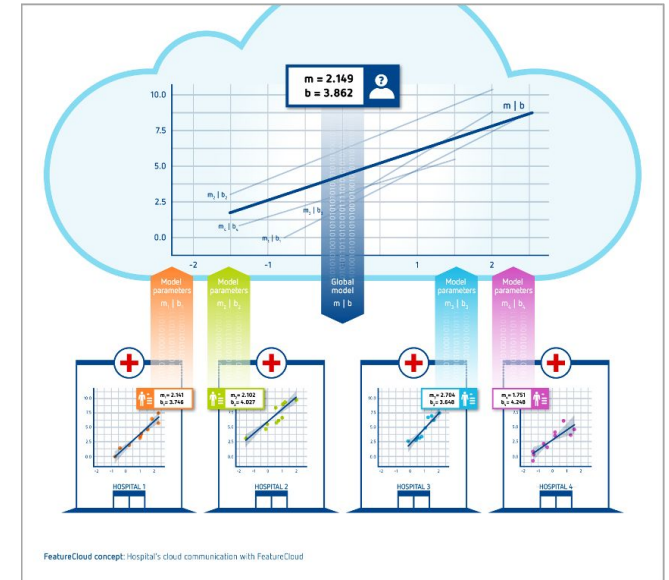
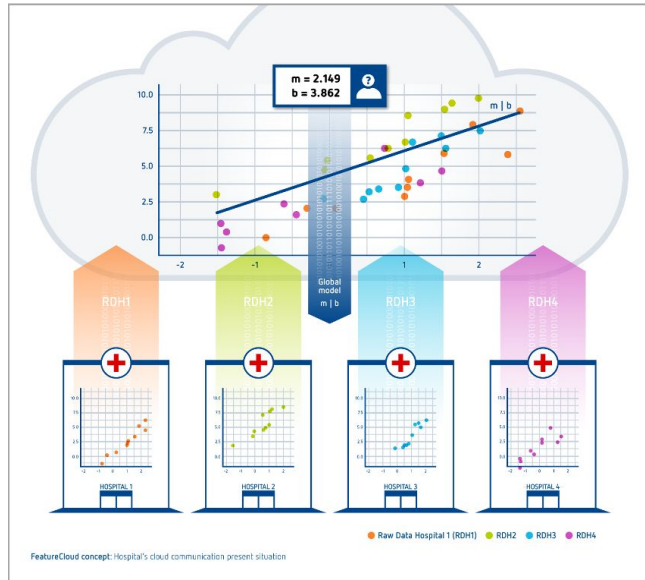
- General Data Protection Regulation (GDPR) and others
- Patients must have control over their data
- Data cannot leave the hospitals without big legal efforts

# Idea: Federated Learning

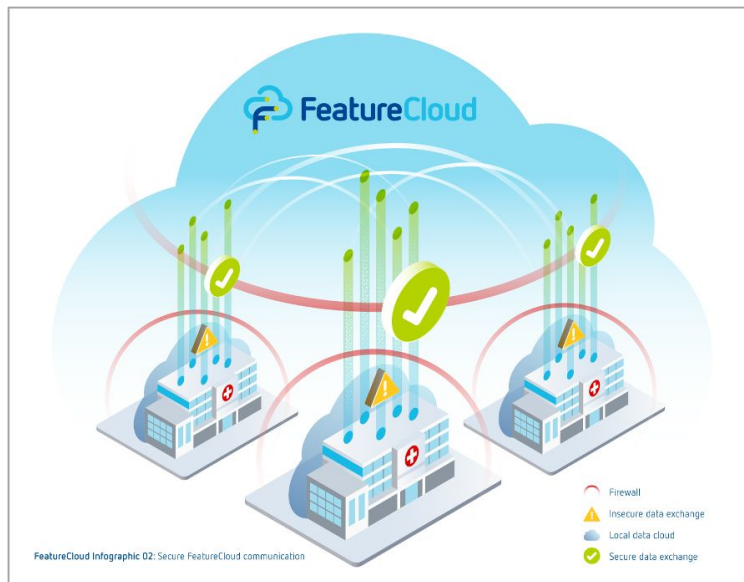
- Data remains where it has been collected
- Split training between local training and global aggregation

*“Bring the model training to the data, not the other way round”*

# Large-scale AI in Biomedicine - Classical Approach



# Federated Learning Platforms



- Isolate data
- Impose restrictions on communication
- Provide assistance for development of federated applications

# What's the price?

- Loss of performance due to communication overhead
- Loss of accuracy (in most cases) due to higher difficulty of generalization
- Less control over data quality

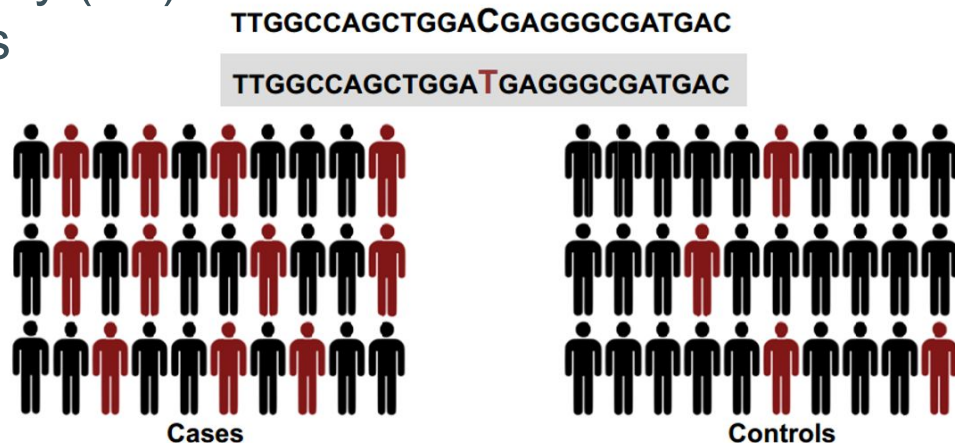


# Is Federated Learning privacy-preserving?

- Depends on the method and the data
- Privacy leaks can still occur (e.g., due to overfitting)
- But: huge improvement over sharing raw data
- Can be enhanced with additional techniques
  - Differential privacy
  - Secure multiparty computation
  - Homomorphic encryption

# Application example: GWAS

- Case/control studies on SNP data
- Each SNP examined individually (LR)
- Find SNPs related to diseases



# Application example: GWAS (Cont'd)

Data

TCGATCGATCGATCGA  
TCGATCGATCGATCGA

Age, gender, etc

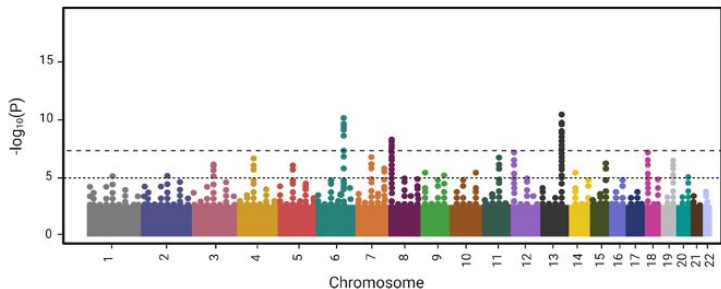
Case/control or  
quantitative

Association test

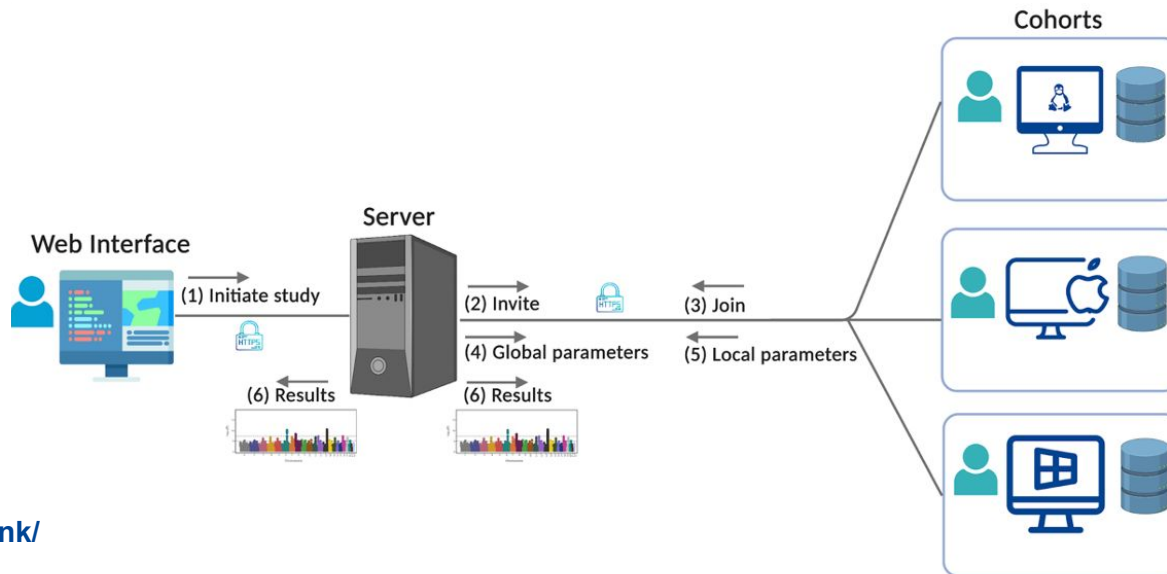


- Chi-square test
- Linear regression
- Logistic regression

p-values



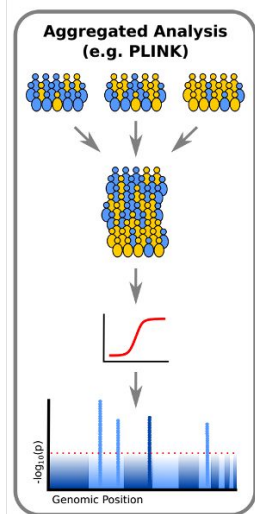
# Application example: GWAS (Cont'd)



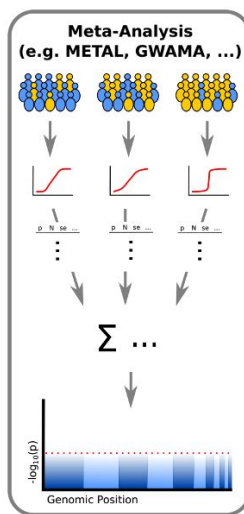
URL: [exbio.wzw.tum.de/splink/](https://exbio.wzw.tum.de/splink/)

*Reza Nasirigerdeh et al. (published in Genome Biology)*

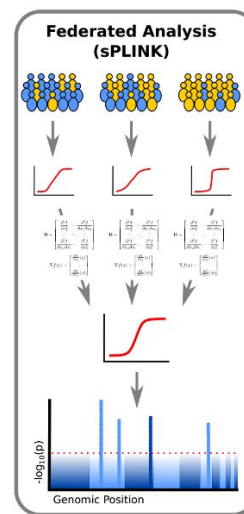
# Application example: GWAS (Cont'd)



- ✗ Privacy-preserving
- ✓ Robust to imbalance

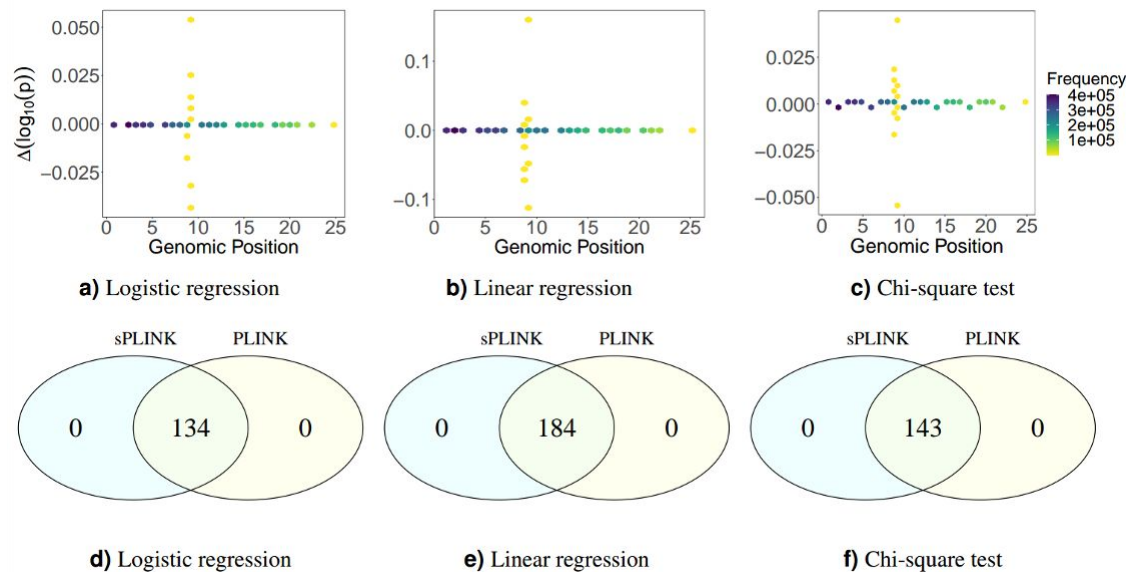


- ✓ Privacy-preserving
- ✗ Robust to imbalance



- ✓ Privacy-preserving
- ✓ Robust to imbalance

# Application example: GWAS (Cont'd)



# Other examples

## Flimma

Differential expression analysis based on the limma voom pipeline

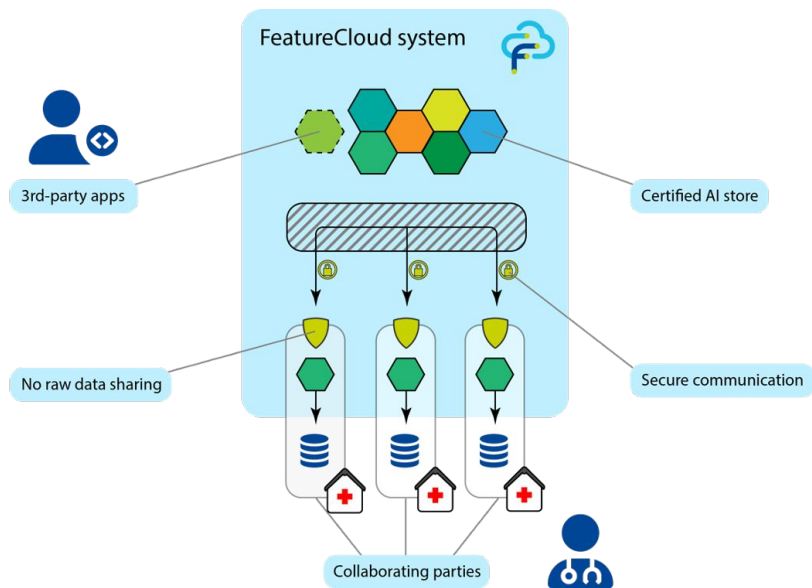
URL: [exbio.wzw.tum.de/flimma/](https://exbio.wzw.tum.de/flimma/) | *Olga Zolotareva et al.* (published in Genome Biology)

## Partea

Time-to-event studies

URL: [exbio.wzw.tum.de/partea/](https://exbio.wzw.tum.de/partea/) | *Julian Späth et al.*

# Platform example: FeatureCloud



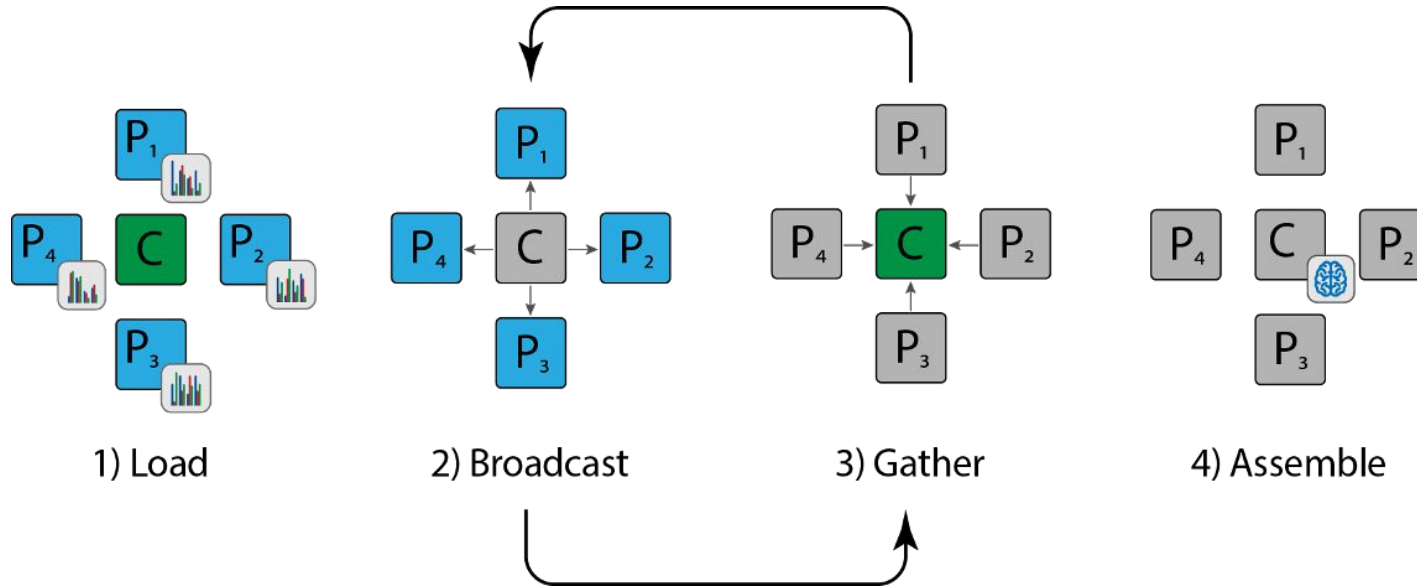
- Increase convenience for developers (including testing and deployment)
- Increase trust in apps (certification, consent management)
- Create a community (developers, researchers, patients)



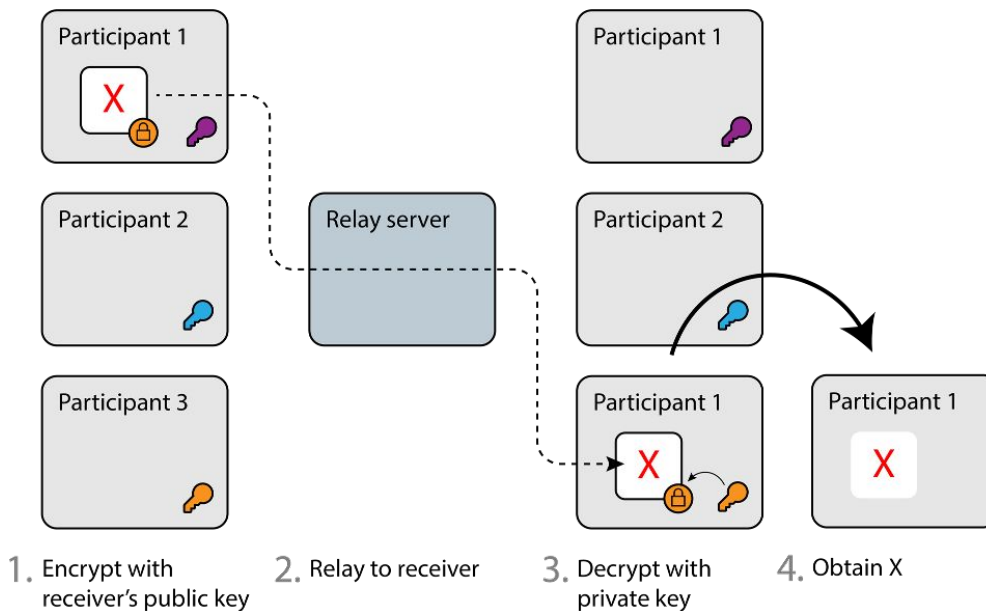
# App communication

- Star-based federated learning
  - Gather-broadcast
- Peer-to-peer
- Secure multi-party computation

# Star-based communication



# Peer-to-peer communication



# Continuous vs. discrete models

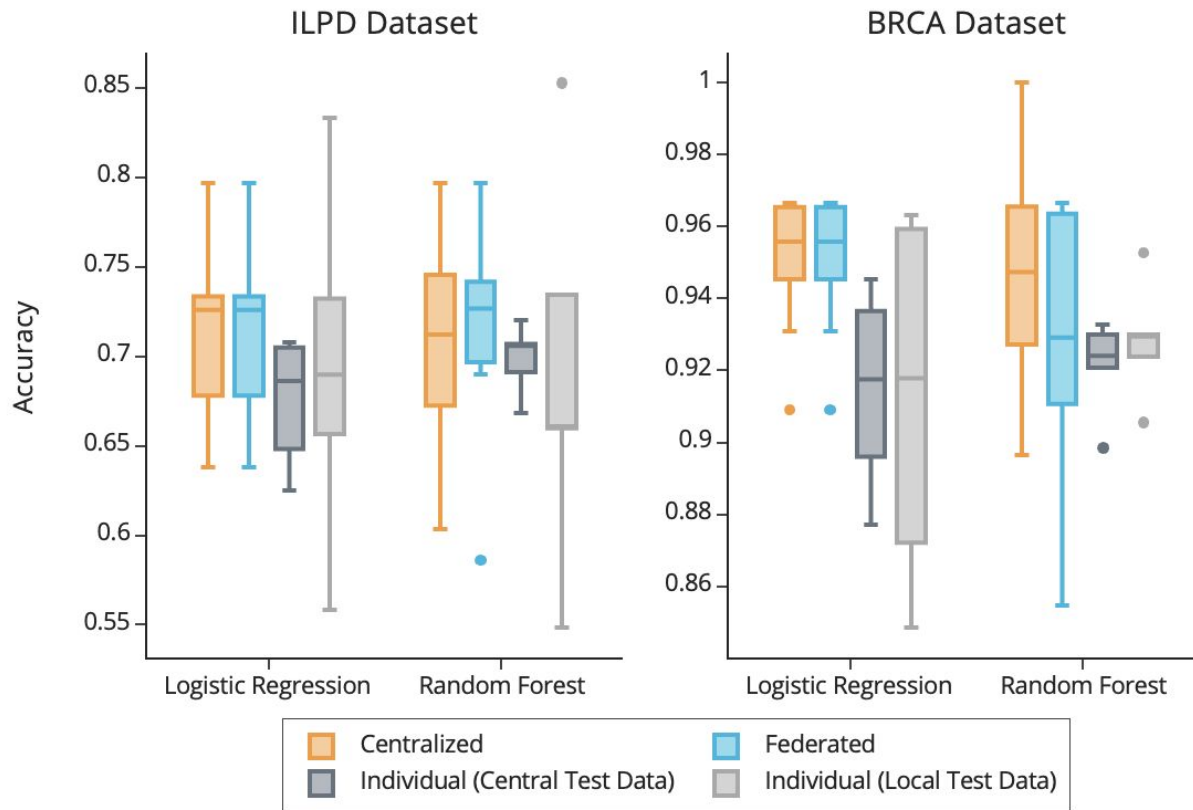
- Continuous models can usually be federated ‘straightforward’
  - Gradient descent
  - Federated average
  - One or multiple interactions
- Discrete models are more difficult
  - Example: Decision tree
  - Possible solutions: Ensemble techniques

## Example: Random forest

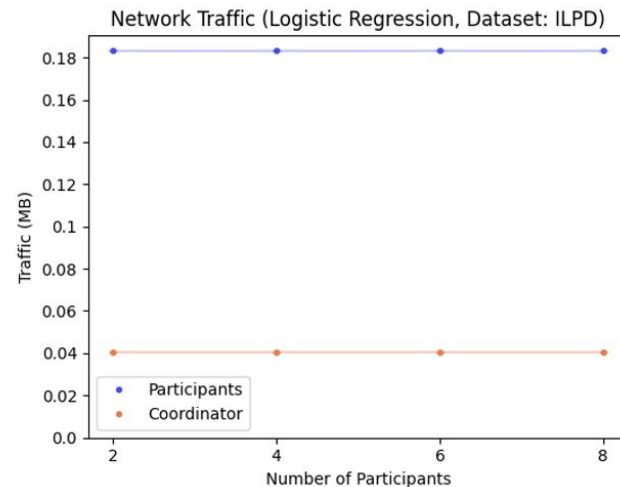
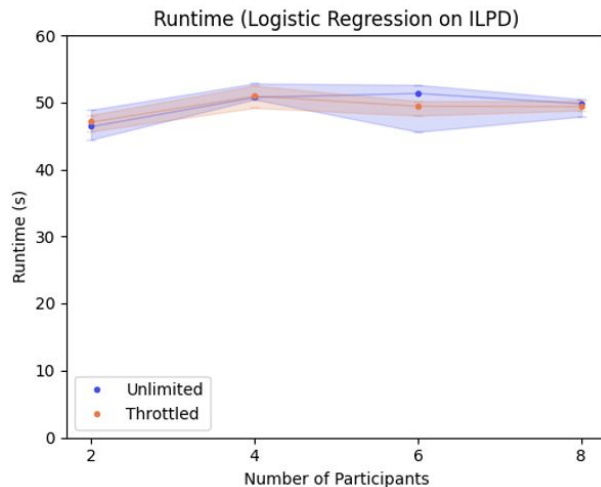
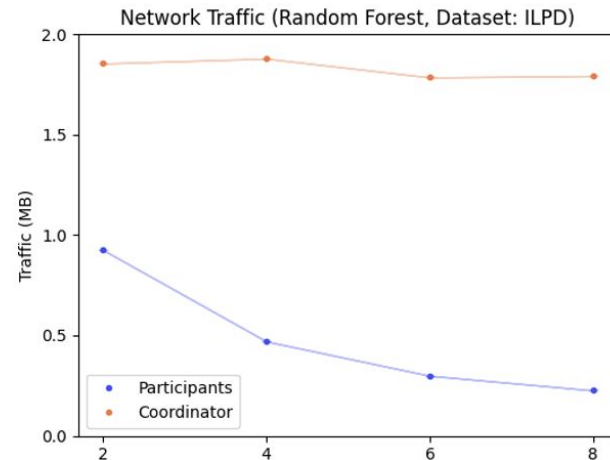
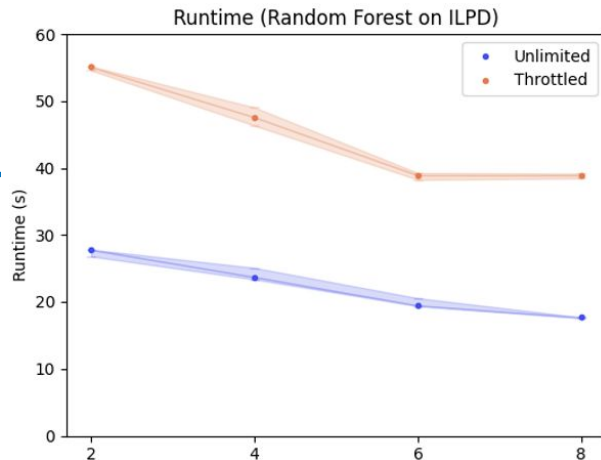
- Train decision trees using bootstrapping on each site
- Transmit models and merge into big ensemble → Random Forest

Federated Random Forests can improve local performance of predictive models for various healthcare applications, Anne-Christin Hauschild, Marta Lemanczyk, Julian Matschinske, et al., Bioinformatics, Volume 38, Issue 8, 15 April 2022, Pages 2278–2286

# Evaluation



# Evaluation





Universität Hamburg  
DER FORSCHUNG | DER LEHRE | DER BILDUNG

ISMB 2022 - Tutorial - Federated Learning in Biomedicine

---

**Thank you!**