Federated Cell type classification

Supervisor: Dr. Anne Hartebrodt

Federated learning has been promoted as a privacy-aware data analysis strategy for the processing of biomedical and proprietary data. It has generally been shown to achieve similar performance compared to centralised machine learning, depending on the Application.

Federated learning can contribute to better classification performances in areas where people are not free to share their data, such as in the medical domain, but are willing to share statistical information.

The goal of this project is to create an online (pseudo)-federated cell type classifier, based on previous approaches. In particular, the challenge will be harmonizing datasets with the same cell types but different annotations. Concretely, there will be an initial training phase, where the classifier is trained on the annotated data set with unified annotations (to be created). In the subsequent online phase, users can download the model and use the pretrained model as a classifier for their data, if they expect cell types to be annotated in the initial training data. However, if they expect to have new cell types, they can use the pretrained weights, but add a new class. Subsequently, they can upload their new model to the server, if the performance does not significantly drop on the validation data located at the server.

Requirements:

- Python
- DL Framework
- Molecular biology basic knowledge