The key to research: Using data without borders by applying a ML approach to generate useful in-silico data in virtue of privacy and GDPR compliance

ABSTRACT

In this research, we investigate and design a standardised Federated Learning (FL) architecture for medical data, addressing the significant gaps identified in the current literature. In our research, there is a particular focus on tabular data, a less explored area compared to image data in FL studies. The proposed architecture comprises several key components, including privacy mechanics, FL algorithms, and model specifications. Each component is meticulously evaluated to ensure optimal performance with a strong emphasis on privacy protection. To address the gap in the underexplored area of tabular data, we investigated traditional and novel methods of leveraging tabular data. Notably, we introduce "FedDeepInsight" an innovative method we devised to transform tabular data into images and then use neural net- works for training. This approach has demonstrated the potential to enhance model performance while strengthening data privacy. Furthermore, our extensive research on the key components establishes a blueprint to design a standardised federated learning architecture tailored for medical data.

KEYWORDS

federated learning, deepinsight, tabular data, image data, architecture, privacy protection, gdpr

Description and purpose of the project: You will be working with pre-fitted data to create an approach based on your proposed methods and elucidate a model that can fit into the existing privacy and GDPR pipeline. The objective is not only to develop something new you would like to research about but also be able to complement a tool that could be useful for researchers later.

Expectations: All our students usually end with at least one publication of their work in a peer-reviewed journal or conference. You are expected to be able to explain your results/methods in Layman speech (able to explain to a kid in normal everyday words).