# SUMMARY PAPER-2
# The future of digital health with federated learning

Research on artificial intelligence (AI), and particularly the advances in machine learning (ML) and deep learning (DL) have led to disruptive innovations in radiology, pathology, genomics and other fields.Federated learning (FL) is a learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself. Originally developed for different domains, such as mobile and edge device use cases, it recently gained traction for healthcare applications. FL enables gaining insights collaboratively, e.g., in the form of a consensus model, without moving patient data beyond the firewalls of the institutions in which they reside. Instead, the ML process occurs locally at each participating institution and only model characteristics (e.g., parameters, gradients) are transferred as depicted.However, FL still requires rigorous technical consideration to ensure that the algorithm is proceeding optimally without compromising safety or patient privacy. Nevertheless, it has the potential to overcome the limitations of approaches that require a single pool of centralised data.

## DATA-DRIVEN MEDICINE REQUIRES FEDERATED EFFORTS

ML and especially DL is becoming the de facto knowledge discovery approach in many industries, but successfully implementing data-driven applications requires large and diverse data sets. However, medical data sets are difficult to obtain.The need for large databases for AI training has spawned many initiatives seeking to pool data from multiple institutions. This data is often amassed into so-called Data Lakes. These have been builtwith the aim of leveraging either the commercial value of data, e.g.IBM's Merge Healthcare acquisition.Centralising or releasing data, however, poses not only regulatory, ethical and legal challenges, related to privacy and data protection, but also technical ones. Anonymising, controlling access and safely transferring healthcare data is a non-trivial, and sometimes impossible task.

## Current FL efforts for digital health

Since FL is a general learning paradigm that removes the data pooling requirement for AI model development, the application range of FL spans the whole of AI for healthcare.It is worth noting that FL efforts require agreements to define the scope, aim and technologies used which, since it is still novel, can be difficult to pin down. In this context, today's large-scale initiatives really are the pioneers of tomorrow's standards for safe, fair and innovative collaboration in healthcare applications.By linking healthcare institutions, not restricted to research centres, FL can have direct clinical impact. The on-goingHealthChain project52, for example, aims to develop and deploy a FL framework across four hospitals in France. This solution generates common models that can predict treatment response for breast cancer and melanoma patients. It helps oncologists to determine the most effective treatment for each patient from their histology slides or dermoscopy images.The aim is to improve tumour boundary detection,including brain glioma, breast tumours, liver tumours and bone lesions from multiple myeloma patients. Another area of impact is within industrial research and translation. FL enables collaborative research for, even competing, companies. In this context, one of the largest initiatives is the Melody project.

## Federated learning definition

FL is a learning paradigm in which multiple parties train collaboratively without the need to exchange or centralise datasets.each participant typically obtains and refines a global consensus model by conducting a few rounds of optimisation locally and before sharing updates, either directly or via a parameter server. The more rounds of local training are performed, the less it is guaranteed that the overall procedure is minimised.

## Challenges and considerations

Despite the advantages of FL, it does not solve all issues that are inherent to learning on medical data. A successful model training still depends on factors like data quality, bias and standardisation.These issues have to be solved for both federated and unfederated learning efforts via appropriate measures, such as careful study design, common protocols for data acquisition,structured reporting and sophisticated methodologies for discovering bias and hidden stratification.upon the key aspects of FL that are of particular relevance when applied to digital health and need to be taken into account when establishing FL. For technical details and in-depth discussion, we refer the reader to recent surveys.

ML, and particularly DL, has led to a wide range of innovations in the area of digital healthcare. As all ML methods benefit greatly from the ability to access data that approximates the true global distribution, FL is a promising approach to obtain powerful,accurate, safe, robust and unbiased models. By enabling multiple parties to train collaboratively without the need to exchange or centralise data sets, FL neatly addresses issues related to egress of sensitive medical data. As a consequence, it may open novel research and business avenues and has the potential to improve patient care globally. However, already today, FL has an impact on nearly all stakeholders and the entire treatment cycle, ranging from improved medical image analysis providing clinicians with better diagnostic tools, over true precision medicine by helping to find similar patients, to collaborative and accelerated drug discovery decreasing cost and time-to-market for pharma companies. Not all technical questions have been answered yet and FL will certainly be an active research area throughout the next decade 12. Despite this, we truly believe that its potential impact on precision medicine and ultimately improving medical care is very promising.