Distributed learning for optimal radiomics knowledge

Impact paragraph

Artificial intelligence demands a large amount of high-quality data to build reliable and generalizable models. Conventionally data analysis models are learned using centralized datasets. In a centralized setting, data is collected and shared to one database. Data analysis models such as machine/deep learning learn directly from the data. Data sharing and centralization is however strictly regulated by legal and ethical considerations making the process of publicly sharing the data time consuming and costly.

Distributed and federated learning infrastructures allow to apply machine/deep learning algorithms, in addition to other statistical data analysis in multicentric settings without patient data and information ever leaving the individual clinical institutes. These infrastructures are designed to preserve patient data privacy. These infrastructures however suffer from luck of traceability and trust. Despite that, to apply these infrastructures in personalized medicine, and for them to last and be part of the medical device development process, the following points need to be addressed: 1) intensify traceability of patient data, 2) boost the trust amongst distributed learning partners, 3) provide evide nce regarding the added value of distributed learning on providing optimal treatment options for individual patients.

Peer-to-peer distributed frameworks such as blockchain (Ethereum in our study), provide a secure decentralized ledger that allows individuals to securely transfer and exchange value. Ethereum is an open-source public blockchain, that enables smart contact development. The computer programs of the smart contracts enforce the execution of the terms of legal agreements (contracts).

In our study, we integrated blockchain to distributed learning and proposed a fully decentralized infrastructure. The integration of Blockchain to distributed learning infrastructures is of high importance as it 1) eliminates the need to trust a third parity that aggregates the learning process (fully decentralizes the learning process), 2) provides means to trace partner involvement in the learning process (insured by the immutability of the transaction history), and 3) sets a transparent environment for the commercialization and intellectual property agreements related to the medical devises learned in a distributed fashion.

Scientific impacts

1. All our studies are published in medical and technical journals (such as: Medicinal Research Reviews, JCO Clinical Cancer Informatics, IEEE Access, and Computers in Biology and Medicine).

2. All our studies are open access.

3. The main contribution of this thesis is the application of blockchain technology to distributed learning infrastructures (decentralized learning). The studies carried out throughout this thesis are a baseline and starting point to investigate multiple diagnosis and prognosis clinical and/or imaging analysis. Additionally, decentralized learning, by exposing the model to more data, can help in improving the predictive value of existing and new models such as radiomics signatures as well as automated segmentation models. This in turn will help to reduce the time of the decision-making process.

4. **Chapter 2** presents an overview of radiomics (as one of the research fields that utilized distributed learning), assessed its challenges, pitfalls, and future directions.

5. **Chapter 3** appraises the existing implementations of conventional distributed learning frameworks and discusses possible future pathways.

6. **Chapter 4** showed that utilization blockchain technology for distributed learning is feasible and is essential to decentralize the learning process, enforce trust, immutability, transparency, traceability, and security.

7. **Chapter 5** in dictated that distributed learning is feasible and promising and in case of the diseases with a low prevalence.

8. **Chapter 6** evaluated the infrastructure proposed in chapter 4 in a distributed learning network.

Social impacts

1. Distributed learning, especially decentralized learning, has the potential to expose predictive models to more data, hence, help in improving the predictive value of existing and new radiomics signatures as well as the automated segmentation models. This in turn will help to reduce the time of decision-making process and reduce medical practitioners' workload.

2. Continuous improvement of diagnosis and prognosis tools can improve patient care systems and reduce mortality rates.

3. Fast and accurate predictive models may save time and reduce costs for patients.