Federated Machine Learning to support Diagnostic Imaging for next generation AI-powered healthcare

1 CHALLENGE

The use of Artificial Intelligence - and Machine learning in particular - on health data holds great promises for the medicine of the future. If the analysis of health data can have impact on diagnosis definition or new treatments1, this data is also extremely important to the research. But as a special category of personal data, health data brings a lot of challenges especially as it is difficult to concentrate large amounts of quality data for the development of AI methodologies, mainly for security and privacy reasons. Musketeer platform offers to tackle those challenges in order to leverage AI capabilities using large health datasets in a secure way, namely:

- **Demonstrate the application of the Artificial Intelligence methodologies and technologies developed by the project, enabling access to vast amounts of distributed medical imaging data to train and improve the learning algorithms**

Health data is a special category of personal data that encompasses an extreme value for the data subject, considering its own health and wellbeing, and for the healthcare practitioners who should decide on the correct diagnosis and care pathways to achieve the best patient outcomes. Health data is also extremely important to the research, development and validation of new technologies, procedures and care pathways to improve the diagnosis, prognosis and treatment of diseases.

The welding quality assessment can be improved using machine learning algorithms which support the machinery status monitoring. But a single factory might offer too few data to create such algorithms. It requires accessing larger datasets from Comau's robots located in different places to boost the robustness and quality of the machine learning model.

However, Comau's customers can be competitors. Those companies do not intend to share data with competitors and simply waive their data sovereignty. With federated machine learning techniques, Comau can offer an appropriate level of security for customers and save them costs at the same time. Besides, the aforementioned data might include personal information regarding operators working in the manufacturing plant which can raise additional privacy concerns that have to be tackled by the solution.

[1] https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-cons-predictive-maintenance.pdf

Although 70% of world biobanks are located in Europe, until recently, imaging data coming from sources such as magnetic resonance imaging (MRI) or computed tomography (CT) were not included in such biobanks [2]. Projects have been launched to acquire large repositories of image data, but in 80% of cases the access to imaging biobanks is restricted to research and clinical reference.

Besides traditional in-house PACS systems, multi-tenant and multi-datacentre cloud solutions for medical imaging management, analysis and reporting, have been used in clinical practice for radiology and tele-radiology for a few years. These cloud solutions have been used by public hospitals to organise networked, collaborative reporting services, and by private practices to improve the productivity on large distributed groups and on small clinics. Vast amounts of medical imaging data are collected and reported using these cloud solutions, however, each organisation accesses only its own data.

The pressure for productivity is increasing due to the lack of Radiologists and the growing demand for medical imaging services. Key driving factors are the rise in prevalence of chronic diseases, technological advancements in diagnostic imaging modalities, increasing number of imaging procedures, rising awareness among the patients about early diagnosis of clinical disorders and rise in base of aging population. In addition, increasing demand from emerging countries, improved government funding towards chronic disorders, increasing investment in public private organizations, and increasing disposable income among the population will further expected to drive the market in the coming years [3].

The European Union, the US and many other countries have been focusing their public health policies and research efforts on personalised medicine and evidence based clinical pathways to improve patient outcomes and effectiveness of care. The solution lies in providing powerful tools to support the radiologists to take faster and more accurate decisions for diagnosis and prognosis. As some research projects have been indicating, AI and Deep Learning (DL) are the disruptive technologies that will enable develop these powerful tools leading to improvements to the clinical protocol pathways and conducting to better efficiencies and better patients' outcomes.

Privacy and security concerns and barriers to overcome:

- Privacy of personal data;
- Data Localization;
- Information Leakage;
- Data Standardization;
- Data Untrustworthiness.

In addition, we can take into account the possibility of adversarial attacks. A data poisoning attack can result in a useless predictive model.

This project can solve these barriers. MUSKETEER allow machine learning over datasets allocated in different locations (thus removing the data localization barrier) where the privacy preserving analytics remove any chance of information leakage and with mechanisms to provide standardisation among different partners. In addition, the adversarial attack detection and mitigation strategies will be capable to detect data poisoning attacks and alert the other hospitals.

2 ACTION

Collection of health data in hospitals and clinics using MRI scanners at Hygeia Hospital in particular to support the training of AI algorithms for the detection of prostate cancer, in the federated and privacy-preserving learning environment of the MUSKETEER platform.

Being such a vast area, with several imaging modalities applying to different human body parts to analyse distinct conditions, we shall restrict the demonstration to one specific type of study. The main objective is the training of AI algorithms for support the detection of prostate cancer.

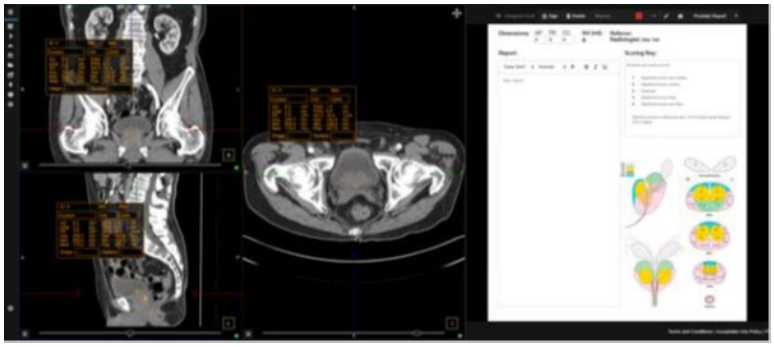

FIGURE 1 PROSTATE IMAGE VISUALISATION AND REPORT

The medical imaging data, consisting of multi-parametric Magnetic Resonance Imaging (mp-MRI) is collected in hospitals and clinics using MRI scanners. Imaging data is stored in PACS systems, as Biotronics3D' s 3Dnet, for visualization, analysis and reporting of findings. Patients suspected of prostate cancer have a biopsy that allows to confirm the diagnosis.

The partners are collecting data at Hygeia Hospital and reusing publicly available data from The Cancer Imaging Archives (TCIA) to train AI algorithms for supporting the detection of prostate cancer, in the federated and privacy-preserving learning environment of MUSKETEER.

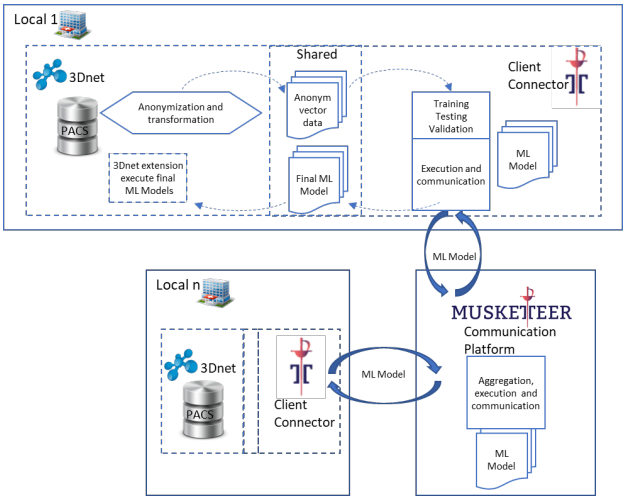

FIGURE 1 DATA FLOW IN MUSKETEER HEALTHCARE PILOT

3 RESULT

Implementation of a machine learning algorithm able to identify lesions and classify prostate cancer on medical images.

The combination of large amounts of medical imaging data and biopsy reports from patients suspected of prostate cancer enables the creation of AI models. Once the models are trained and have a satisfactory accuracy, thanks to the federated approach, it shall be possible to provide identification and classification of prostate cancer lesions directly from the imaging data.

This leads to numerous advantages over the existing procedures:

- Support Radiologists and Urologists establishing a diagnosis based on medical imaging in a faster and more accurate manner;
- Avoid invasive procedures, as biopsy, for patients with confirmed negative results;
- Federated learning allows training on multiple datasets maintaining the sovereignty of each participant, and preserving data privacy;
- Similar approaches can be adopted to support diagnosis of different diseases based on non-invasive medical imaging data.

REFERENCES

[1] D. Bos (Erasmus MC, Rotterdam, NL), Strategy/pipeline to develop an imaging biobank, Vienna, ECR2018

[2] ESR Position Paper on Imaging Biobanks, Insights Imaging, 2015

[3] Medical Imaging Market - Global Industry Analysis, Size, Share, Trends and Forecast, 2015 – 2023