Executive Summary

Third reporting period (1st July 2021 – 30th June 2022)

The digital revolution, in particular big data and artificial intelligence (AI), offers new opportunities to transform healthcare. However, it also harbours risks to the safety of sensitive clinical data stored in critical healthcare information and communication infrastructure. Data exchange over the internet is perceived as insurmountable, posing a roadblock hampering big data-based medical innovations. FeatureCloud's transformative security-by-design concept minimises the cyber-crime potential and enables secure cross-border collaborative data mining endeavours. FeatureCloud is implemented as a community-extendible software platform to substantially reduce cyber risks to healthcare infrastructure by employing the world-wide first privacy-by-architecture approach with two key characteristics: (1) no sensitive data is communicated, and (2) data is not stored in one central point of attack. Federated machine learning (for privacy-preserving data mining) integrated with relevant privacy-enhancing technology, like secure multiparty computation or differential privacy, will safely apply next generation AI technology for medical purposes. Our ground-breaking cloud-AI infrastructure only exchanges learned model representations which are anonymous by default. Collectively, our highly interdisciplinary consortium, from IT to medicine, covers all aspects of the value chain: assessment of cyber risks, legal considerations and international policies, development of federated AI technology, extendible app store and user interface design, implementation as prognostic medical devices, evaluation and translation into clinical practice, commercial exploitation, as well as dissemination and patient trust maximisation. FeatureCloud's goals are bold, necessary, achievable, and pave the way for a socially agreeable big data era of the Medicine 4.0 age.

Work performed from the beginning of the project to the end of the period covered by the report and main results achieved so far

We implemented and disseminated a stable production version of the FeatureCloud AI Store, including software development packages for corresponding computer-computer interfaces running as web servers and fostering an advanced, user-extendible app store functionality. It was developed using documented bi-weekly platform developer online conferences with all source code stored in Git repositories. Software development was organised into sprints and extended by several app development hackathons, including one for external participants, to stress-test and improve the platform regarding app development. We have added functionalities to further secure the platform by providing secure multiparty computation and differential privacy capabilities/packages. More than 30 apps for federated learning are now available, from standard statistics to advanced apps on federated principal component analysis (PCA; published at ICDM), artificial neural networks, random forest classifiers and survival time predictors. In total, we prepared five live demos to illustrate FeatureCloud's capabilities. Using the first FeatureCloud apps, we demonstrated the power of federated machine learning coupled to relevant, related privacy-enhancing technologies. Specifically, we worked on typical medical application scenarios. We began with a federated genome-wide association study (GWAS) tool: "sPLINK", which mimics the non-federated standard GWAS tool "PLINK" (published in Genome Biology). We demonstrated that currently available distributed GWAS software (so-called meta-analysis tools) massively loses accuracy when the data suffers heterogeneously distributed outcomes or confounders. In contrast, "sPLINK" gives the exact same results as the gold-standard tool PLINK, and thus has the potential to become the new standard tool for genotyping as it does not require any exchange of raw data between participating institutions/hospitals while maintaining the same level of accuracy as state-of-the-art centralised tools. "sPLINK" implements federated Chi-squared tests, as well as federated multimodal linear and logistic regression models. Likewise, we developed the first software for federated survival analysis: PARTEA. It combines federated statistical modelling and differential privacy approaches based on Laplacian noise to generate privacy-preserving Kaplan-Meier plots. A corresponding paper is under review. In addition, we developed, evaluated and published "flimma", a federated gene expression data analysis tool (published in Genome Biology). The FeatureCloud AI store itself was also evaluated and tested in real-world scenarios, and a corresponding publication is under review (see preprint).

Progress beyond the state of the art and expected potential impact (including the socio-economic impact and the wider societal implications of the action so far)

FeatureCloud contributes significantly to all three expected impacts mentioned in the work programme:

- **Improved security of Health and Care services, data and infrastructures.**

  By addressing the evident roadblock in medical data mining – centralised data mining but distributed clinical data – we improve the cyber security of computational health care services, patient data and communication infrastructure. FeatureCloud's federated machine learning and SMPC engines erase the necessity to share sensitive data via a cloud.

- **Less risk of data privacy breaches caused by cyberattacks.**

  FeatureCloud significantly reduces the risk of data privacy breaches caused by cyberattacks on health cloud services or on the communication channels between hospital and cloud. Instead of bringing the data to the AI, we bring the AI to the data.

- **Increased patient trust and safety.**

  Based on trusted authority technology, like blockchains, we work on ensuring patients' full control over access rights to their own sensitive data combined with the guarantee that no sensitive data is exchanged in order to learn the federated AI model that could be traced back to individual patients. This strategy will increase patient trust and safety significantly. Our FeatureCloud platform is designed to be in accordance with E.U. GDPR and NISD policies, and it is developed with respect to the criteria for software-supported medical devices of the FDA and EMA, respectively. FeatureCloud furthermore contributes to the following most significant impacts not mentioned in the work programme:

- **The novel FeatureCloud technology will create new market opportunities.**

  The FeatureCloud's AI Store for client-side machine learning tools will have an enormous impact worldwide and foster pan-European business, e.g., with spin-offs and start-ups because of a huge emerging market in privacy-aware machine learning.

- **The European society will benefit from new levels of personalised medicine, new possibilities for research of complex diseases like e.g., cancer, and lower costs of medical research.**

  FeatureCloud enables open science without boundaries, cross-domain and pan-European, which will particularly allow new levels of cancer research because FeatureCloud apps address current privacy, ethical, security, and safety restrictions at the core, and will thus increase overall quality of medical care while slowing increases of health care costs in Europe.

FeatureCloud Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078. This periodic summary report reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

Address (URL) of the action's public website

https://featurecloud.eu/