*Caring is Sharing – Exploiting the Value in Data for Health and Innovation*
*M. Hägglund et al. (Eds.)*

*doi:10.3233/SHTI230196*

# Secur-e-Health Project: Towards Federated Learning for Smart Pediatric Care

Rita RB-SILVA [a,1], Xavier RIBEIRO [b,c], Francisca ALMEIDA [b,c], Carolina AMEIJEIRAS-RODRIGUEZ [a], Julio SOUZA [a,d], Luis CONCEIÇÃO [b], Tiago TAVEIRA-GOMES [a,c], Goreti MARREIROS [b] and Alberto FREITAS [a,d]

[a] *Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS), Faculty of Medicine, University of Porto, Porto, Portugal*
[b] *GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, LASI – Intelligent Systems Associate Laboratory, Institute of Engineering – Polytechnic of Porto, Porto, Portugal*
[c] *MTG Research and Development Lab, Porto, Portugal*
[d] *CINTESIS@RISE, Faculty of Medicine, University of Porto, Porto, Portugal*

ORCiD ID: Rita Rb-Silva https://orcid.org/0000-0002-1422-0974; Xavier Ribeiro https://orcid.org/0000-0002-5316-0732; Francisca Almeida https://orcid.org/0000-0002-3872-0682; Carolina Ameijeiras Rodríguez https://orcid.org/0000-0001-8817-7772; Julio Souza https://orcid.org/0000-0002-8576-1903; Luis Conceição https://orcid.org/0000-0003-3454-4615; Tiago Taveira-Gomes https://orcid.org/0000-0002-0998-6000; Goreti Marreiros https://orcid.org/0000-0003-4417-8401; Alberto Freitas https://orcid.org/0000-0003-2113-9653.

**Abstract.** The application of machine learning (ML) algorithms to electronic health records (EHR) data allows the achievement of data-driven insights on various clinical problems and the development of clinical decision support (CDS) systems to improve patient care. However, data governance and privacy barriers hinder the use of data from multiple sources, especially in the medical field due to the sensitivity of data. Federated learning (FL) is an attractive data privacy-preserving solution in this context by enabling the training of ML models with data from multiple sources without any data sharing, using distributed remotely hosted datasets. The Secur-e-Health project aims at developing a solution in terms of CDS tools encompassing FL predictive models and recommendation systems. This tool may be especially useful in Pediatrics due to the increasing demands on Pediatric services, and the current scarcity of ML applications in this field compared to adult care. Herein we provide a description of the technical solution proposed in this project for three specific pediatric clinical problems: childhood obesity management, pilonidal cyst post-surgical care and retinography imaging analysis.

**Keywords.** federated learning; machine learning; privacy-preserving protocols; clinical decision support; pediatrics.

[1] Corresponding Author: Rita Rb-Silva, E-mail: rrsilva@med.up.pt

## 1. Introduction

In recent years, Federated Learning (FL) has gained increased attention from the academic, medical and data science communities due to its capability of training artificial Intelligence (AI) models using data from multiple sources while maintaining data privacy [1]. This Machine Learning (ML) paradigm enables collaborative algorithm training without exchanging or retrieving data from its sources, thereby addressing many of the existing barriers for achieving the full potential of AI applications in healthcare, namely those related to data governance and privacy [2]. Moreover, recent literature indicates that FL models can achieve performance levels that are comparable to those trained on centrally hosted datasets, and even superior to those estimated from isolated single-institutional datasets [3].

Amongst the plethora of clinical applications of AI models obtained through FL, there is still lack of research focusing on pediatric care, despite the current workforce challenges and rising demands on pediatric services due to increased patient complexity and comorbidity [4]. In this context, the Secur-e-Health project aims at developing a solution in terms of Clinical Decision Support (CDS) tools encompassing FL predictive models and recommendation systems for smart pediatric care. In this paper, we describe the technical solution proposed within the Secur-e-Health project for three specific pediatric clinical problems (childhood obesity treatment, pilonidal cyst post-surgical management and retinography imaging analysis).

## 2. Methods

A full CDS solution has been designed according to functional and technical requirements defined with the help of clinicians in order to satisfy current healthcare needs and gaps considering three specific clinical scenarios. Clinicians' feedback were obtained from virtual periodic meetings with two physicians and a nurse working in a Portuguese teaching hospital.

In the first scenario, related to overweight and obesity management in children, it is intended to build a predictive medical decision support system to aid clinicians in developing nutritional and exercise recommendations to maximize patient adherence in the long term and to anticipate circumstances that may reduce treatment effectiveness, considering the aspects of the child and its direct family socioeconomic, education and health conditions. In the second scenario, related to pilonidal cyst post-surgical care, a predictive medical decision support system is intended to predict optimal frequency visits following surgery and to estimate the optimum resource allocation for post-surgical treatments. In the third and last scenario, regarding smart retinography images analysis, a set of predictive models is proposed to detect relevant clinical phenotypes or comorbidities by means of retinal fundus imaging analysis, that would otherwise be identified later in the natural history of child development or non-communicable diseases.

## 3. Results

### *3.1.* Proposed architecture

Secur-e-Health system, being a federated learning model, is composed by two main elements: an aggregator and the remote distributed training sites (Figure 1). These two elements communicate through a study package, *i.e.* an algorithm that runs through the distributed data in a private and secure fashion, to determine a weight gradient (W) that will be used to feed the equation:

$$W_{t+1} = W_t + f(W^k)$$

and give the global results.

The main users of Secur-e-Health will be healthcare professionals. Specifically, the system will train predictive ML models with relevant data from EHR and provide the healthcare professionals with feedback and recommendations to maximize the adherence of overweight/obesity patients to the therapeutic plan, as well as to optimize the post-surgical care of patients with pilonidal cysts. Additionally, the healthcare professionals will be also provided with the output from predictive ML models that analyze retinography images and integrate the extracted quantitative features with clinical information.

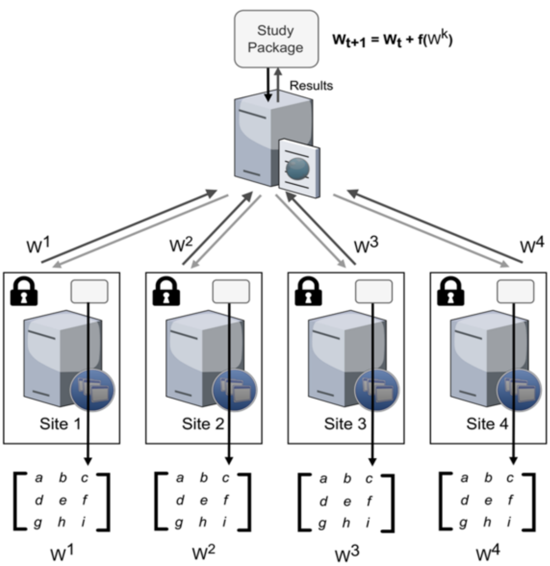


**Figure 1.** General architecture of the Secur-e-Health system.

### *3.2.* Requirements

The business and technical requirements shared by all the use cases are displayed on Table 1. Table 2 displays the specific business and technical requirements for each use case.

**Table 1.** Transversal business and technical requirements for the three pediatric use cases.

| Transversal business requirements | Transversal technical requirements |
|---|---|
| 1. To analyze structured and unstructured data from EHRs;<br>2. To assess and improve the quality of data through a process of data cleansing;<br>3. To display personalized feedback for healthcare professionals;<br>4. To ensure the privacy and security of patient data. | 1. Data gathering: the algorithm should be able to gather information from EHRs;<br>2. Data quality: the algorithm should be able to ensure data quality, by fixing inaccurate, incomplete, inconsistent and repetitive data, by removing duplicates and by adding the data necessary to complete the records;<br>3. Decision support: the algorithm should be able to incorporate all kind of characteristics, process them and output results in a natural medical language that help professionals to make decisions regarding a specific condition;<br>4. Privacy and security compliance: data used to train the models should be used complying with privacy and security norms concerning patient and family data;<br>5. Data access control: server software application programming interface (API) should provide user authentication and authorization. |

**Table 2.** Business and technical requirements for each use case addressing a specific pediatric clinical problem.

| Use case | Business requirements | Technical requirements |
|---|---|---|
| Childhood overweight and obesity management | 1. To identify and gather clinical information regarding pediatric overweight/obesity from EHRs;<br>2. To identify and gather information on lifestyle choices, family socioeconomic status, education level, eating habits and general health conditions;<br>4. To integrate clinical information with individual lifestyle choices and family context;<br>5. To use ML models to profile patient and family features;<br>6. To develop a CDS application to aid clinicians in developing recommendations;<br>7. To promote patient and/or caregivers' adherence to the lifestyle therapy plan;<br>8. To implement quality control over time and to use ML to improve the effectiveness of the CDS application. | 1. Data integration: the algorithm should find associations between clinical features and individual lifestyle choices and family context;<br>2. Individual feedback: the algorithm should output individual results per patient, in order to provide feedback to the named client and maximize the adherence of patients and/or caregivers to the therapy plan;<br>3. Quality control: the algorithm should constantly control its own quality and improve the effectiveness of the CDS application. |
| Pilonidal cyst surgery (PCS) | 1. To identify and gather relevant clinical information from EHRs regarding PCS;<br>2. To identify and gather information on education level, eating habits and general health conditions;<br>3. To identify and gather relevant clinical information on the surgical technique;<br>4.To identify and gather clinical information regarding follow-up visits;<br>5. To integrate clinical information with individual lifestyle choices, hereditary factors, issues related to the chosen surgical technique and anesthesia, and events at follow-up visits;<br>6. To develop predictive models (ML algorithms) to support medical decisions related to PCS;<br>7. To develop a stand-alone CDS system; | 1. Data integration: the algorithm must be able to integrate different types of variable values, regarding patients, at any point in time;<br>2. Data selection: the algorithm will filter the data that are relevant relative to patients submitted to SC surgery;<br>3. Model improvement: the algorithm must continually access the prediction results to constantly improve some of the parameters to be used;<br>4. Clinical assistance: the algorithm must output individual |

| | | |
|---|---|---|
| | 8. To promote patient and/or caregivers' adherence to post-surgical care of PCS. | results in order to maximize patient adherence and optimize the scheduling of further appointments. |
| Retinal imaging analysis | 1. To process retinography images;<br>2. To automatically assess the main factors interfering with the analysis of retinography images and to compare that assessment with the human assessment;<br>4. To identify the main anatomical features of the retina and to assess the blood vessels segmentation using retinography images;<br>5. To determine, gather and provide clinical quantitative features from retinography images;<br>6. To extract and gather morphometric features, comorbidities and other relevant clinical information from the EHRs;<br>7. To integrate quantitative features of retinographies with clinical, laboratory and/or medical imaging data;<br>8. To create predictive models for the presence of relevant clinical phenotypes or comorbidities;<br>9. To compute metrics for the treatment plan and goals assessment;<br>10. To use ML models to profile patient and family features and to help with the medical decision. | 1. Image analysis: the algorithm must analyze the retinography images and extract quantitative features, as well as assess factors interfering with the analysis;<br>2. Assessment similarity: the algorithm must be able to compare its automatic assessment with the human visual assessment;<br>3. Model improvement: supervised ML should be used to extract transferrable features from retinography data to increase model performance;<br>4. Predictive models: the algorithm must use techniques like linear regression, multiple regression, logistic regression, decision trees, random forests, data mining, and neural networks. |

## 4. Conclusions

The main goal of the proposed solution in Secur-e-Health is to optimize adherence to medical recommendations in pediatric care, and thus achieve greater treatment success rates, and healthcare cost savings in the long term. The proposed solution of Secur-e-Health will provide: (i) secure and privacy preserving cross-organizational analysis by providing MultiParty Computation, Federated Learning techniques and AI-driven algorithms; (ii) use of predictive models integrating a medical decision support system to aid pediatricians, namely in the scenarios of childhood obesity management, sacrococcygeal cyst surgical management and integration of complex retinal fundus imaging data to enrich the aforementioned models; (iii) further validation of the algorithms into AI-enabled digital care pathways.

## References

[1] Dayan I, Roth HR, Zhong A, *et al*. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med. 2021 Oct; 27(10):1735-1743. doi: 10.1038/s41591-021-01506-3

[2] Xu J, Glicksberg BJ, Su C, *et al*. Federated Learning for Healthcare Informatics. J Healthc Inform Res. 2021 Mar; 5(1): 1-19. doi: 10.1007/s41666-020-00082-4

[3] Rieke N, Hancox J, Li W, *et al*. The future of digital health with federated learning. NPJ Digit Med. 2020 Sep 14;3:119. doi: 10.1038/s41746-020-00323-1

[4] Clarke SL, Parmesar K, Saleem MA, Ramanan AV. Future of machine learning in paediatrics. Arch Dis Child. 2022 Mar;107(3):223-228. doi: 10.1136/archdischild-2020-321023