The 66th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

803.EMERGING TOOLS, TECHNIQUES, AND ARTIFICIAL INTELLIGENCE IN HEMATOLOGY

An Artificial Intelligence-Based Federated Learning Platform to Boost Precision Medicine in Rare Hematological Diseases: An Initiative By GenoMed4all and Synthema Consortia

Gianluca Asti, MSc [1], Saverio D'Amico, MSc [1,2], Luciana Carota, PhD [3], Davide Piscia, MSc [4], Francesco Casadei, MSc [5], Nono Saha Cyrille Merleau, PhD [6,7], Patricia Alonso de Apellániz, MSc [8], Anne Sophie Kubasch, MD [9], Silke Gloaguen, MSc [9], Nicole Modler, MSc [9], Catalina Gonzalez Martin, MSc [10], Cesare Rollo, PhD [11], Juan Parras, PhD [8], Mattia Delleani [1], Elena Zazzetti, MSc [1], Elisabetta Sauta, PhD [1], Nico Curti, PhD [12], Nicolas Riccardo Derus, PhD [12], Gianluca Carlini, PhD [5], Daniele Dall'Olio, PhD [13], Claudia Sala, PhD [3], Lorenzo Dall'Olio, PhD [12,5], Luca Lanino, MD [14], Giulia Maggioni, MD [1], Alessia Campagna, MD [1], Marilena Bicchieri, PhD [1], Arturo Bonometti, MD [1], Cesare Lancellotti, MD [15], Daoud Rahal, MD [1], Luca Di Tommaso, MD [1,16], Tiziana Sanavia, PhD [11], Piero Fariselli, PhD [11], Matteo Zampini, PhD [17], Matteo Brindisi [14], Victor Savevski, MEng [1], Uwe Platzbecker, MD [18], Maria Diez-Campelo, MD PhD [19], Lin Pierre Zhao, MD PhD [20], Pierre Fenaux, MD [20], Torsten Haferlach, MD [21], Vincent Planat, PhD [10], Raffaella Colombatti, MD PhD [22], Maria del Mar Mañú-Pereira, PhD [23], Anders Krogh, PhD [24], Silvia Uribe, PhD [25], Santiago Zazo, Prof.,PhD [8], Enrico Giampieri, PhD [3], Gastone Castellani, PhD [26], Matteo Giovanni Della Porta, MD [1,16], Federico Alvarez, Prof, PhD [8]

[1] IRCCS Humanitas Research Hospital, Rozzano, Italy
[2] Train s.r.l., Rozzano, Italy
[3] University of Bologna, Bologna, Italy
[4] Centro Nacional de Análisis Genómico - CNAG, Barcelona, Spain
[5] Data Science and Bioinformatics Laboratory, IRCCS Institute of Neurological Sciences of Bologna, Bologna, Italy
[6] Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
[7] Center for Scalable Data Analytics and Artificial Intelligence, University of Leipzig, Leipzig, Germany
[8] Information Processing and Telecommunications Center, Universidad Politécnica de Madrid, Madrid, Spain
[9] Medical Clinic and Policlinic 1, Hematology and Cellular Therapy, Leipzig, Germany
[10] Dedalus Spain, Madrid, Spain
[11] Computational Biomedicine Unit, Department of Medical Sciences, University of Torino, Turin, Italy
[12] Department of Physics and Astronomy, University of Bologna, Bologna, Italy
[13] IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy
[14] IRCCS Humanitas Research Hospital, Rozzano, Italy
[15] Struttura Complessa di Anatomia Patologica, Policlinico di Modena, Modena, Italy
[16] Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Italy
[17] IRCCS Humanitas Research Hospital & Humanitas University, Rozzano - Milan, Italy
[18] Department for Hematology, Cell Therapy, Hemostaseology and Infectious Diseases, University of Leipzig Medical Center, Leipzig, Germany
[19] Hospital Clínico Universitario de Salamanca, Salamanca, Spain
[20] Hôpital Saint-Louis, Université de Paris 7, Paris, France
[21] MLL Munich Leukemia Laboratory, Munich, Germany
[22] Department for Woman's and Child's Health, University of Padova, Padova, Italy
[23] Translational Research in Cancer and Blood disorders in Children, Valll d'Hebron Research Institution, Barcelona, Spain
[24] University of Copenhagen, Copenhagen, DNK
[25] Escuela Técnica Superior de Ingeniería de Sistemas Informáticos, Universidad Politécnica de Madrid, Madrid, Spain
[26] Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy

BACKGROUND. Most oncological and non-oncological hematological conditions fall under the category of rare diseases. Rare diseases present unique challenges due to the limited availability of data, which impacts diagnostic rates and the generation of clinical evidence. Overall, they constitute a public health concern, highlighting the urgent need to develop new

methods for improving data accessibility. In this context, Federated Learning (FL) is a Machine Learning approach that allows multiple centers to collaborate on complex research questions without the need to centralize or share data.

This project was conducted by the Genomed4all and Synthema consortia with the goal of developing an innovative FL platform for rare hematological diseases. This platform enables the development of novel Artificial Intelligence (AI) models for personalized medicine without data sharing, to be implemented in the referral centers of EuroBloodNET, the European Reference Network for rare hematological diseases. The aims of the project were: 1) to develop robust federated models for personalized prediction using multicentric, real-world datasets; 2) to protect patients' privacy; and 3) to enhance collaboration between institutions while avoiding the creation of centralized data repositories.

METHODS. The FL platform includes a manager node (MN) and multiple worker nodes (WN). Users upload their model to the MN, which distributes it to WNs for local training. Trained weights are returned to the MN for aggregation, repeating until training is complete. The MN is hosted at the Humanitas Research Hospital, and the 3 WNs are located at Humanitas, Universität Leipzig, and the University of Bologna.

The testing approach is based on MOSAIC, an AI-based framework for multimodal analysis in rare cancers (PMID 38875514). We focused on myelodysplastic syndromes as a use case, since they are a rare hematological disease with high clinical and genomic heterogeneity (which represents a challenging scenario). The dataset includes comprehensive clinical and genomic data, with information on treatments and clinical outcomes. Each of the 3 WNs has a different sample size of patient data, with 2656, 1328, and 443 samples. Our goal was to build an AI-based model for prediction of survival probability, combining molecular information with clinical data. We planned 3 scenarios for testing: a best-case scenario using all available features, a worst-case scenario excluding up to 50% of genomic information, and a random scenario where an incremental percentage of the genomic variables were randomly removed. For each setting, a DeepSurv model is trained using 80% of local data for training and 20% for testing.

RESULTS. In the best-case scenario, the model achieved a concordance index (c-index) of 0.75. The node with the smallest dataset initially registered the lowest c-index value (0.4); however, after the first training step, its performance began to follow the trend at levels comparable to the nodes with larger samples (0.54). In the random scenarios, the model still achieved a c-index of 0.75 but required more training rounds to converge. The same pattern was also observed in the worst-case scenario that was able to achieve a c-index of 0.74. The most relevant difference between the tested scenarios is the increased training time required to reach a good and stable performance when missing information is present. The developed technology is fully compliant with European GDPR regulations.

In the upcoming months, the FL platform will be implemented in the referral centers of the EuroBloodNET clinical network. In addition to tabular data, we plan to address the federation of features extracted from medical images, including both histological slides and radiological scans (CT, MRI and PET-CT).

CONCLUSION. The FL platform for rare hematological diseases allows multicentric training of AI models without sharing sensitive patient data. This approach ensures data privacy and security, addressing specific challenges associated with rare diseases. Moreover, this platform manages scenarios with missing data or variables, maintaining robustness and accuracy of predictive models. Overall, FL approach is expected to facilitate the development of more advanced and reliable healthcare solutions, paving the way toward personalized medicine in hematology.