RESEARCH TOPIC DASME17
Synthetic data generation by artificial intelligence to accelerate research and precision medicine in onco-hematology

Curriculum DASME standard

Research Area
Onco

Laboratory name and address
AI Center, Humanitas University

Datascience Supervisor
Rosanna Asselta rosanna.asselta@hunimed.eu

Research Supervisor
Matteo Della Porta matteo.della_porta@hunimed.eu

Abstract
Synthetic Data (SD) is artificially-generated data that mimics real data by capturing the statistical distribution and dependencies of the original datasets. The adoption of SD in healthcare is driven by the scarcity of openly available data, concerns about patient privacy preservation, and the potential risk of re-identification associated with traditional methods, as well as the need to address limited availability of data for training AI models through SD augmentation or for accelerating new drug development (e.g., by leveraging synthetic control arms in clinical trials). This PhD project aims to develop and validate reliable SDG tools and methods, providing evidence of their suitability for specific applications in the onco-hematology field.
The project will enhance the methodological and technical aspects of SD, by developing new techniques and advancing already well-established techniques for various data modalities (clinical, -omics, imaging, textual, longitudinal, etc.). Moreover, the project will then focus on assessing the quality/fidelity, applicability/utility, legal status and privacy aspects of the generated SD.

Main technical approaches
Experience in developing machine learning and deep learning techniques and algorithms (such as k-NN, Naive Bayes, Support Vector Machines, Random Forests, etc) in healthcare, also applied to time-series/longitudinal data;

Good knowledge of generative AI and LLM;

Experience in developing generative models (e.g. statistical, GAN, VAE, etc.) applied to medical data for synthetic data generation and digital twins;

Humanitas University
Via Rita Levi Montalcini, 4
20072 Pieve Emanuele (MI) Italy
Tel +39 0282241 - Fax +39 0282242394
info@hunimed.eu
hunimed.eu
CF 97692990159

Good knowledge of Computer Vision and/or NLP is appreciated;
Understanding of data structures, data modeling and software architecture;

Experience in applied statistics skills, such as distributions, statistical testing, regression, etc;
Good scripting and programming skills;

Good proficiency in Python, R programming languages;

Experience with data science frameworks (e.g. tensorflow, pytorch, scikit learn, scipy, pandas, numpy) and visualization frameworks (e.g. plotly, seaborn, matplotlib);

Experience with cloud (GCP, AWS, Azure) and/or distributed computing is appreciated;

Knowledge of database systems and data lakes, good knowledge of SQL is appreciated;

Knowledge of MLOps practices, IT infrastructures, back-end frontend development is appreciated;

Master (PhD would be a plus) in a STEM discipline;

Fluent in written and spoken English and Italian;

Scientific references

D'Amico Saverio et al. "Synthetic data generation by artificial intelligence to accelerate research and precision medicine in hematology" JCO Clinical Cancer Informatics 7, e2300021

D'Amico Saverio et al. "Multi-Modal Analysis and Federated Learning Approach for Classification and Personalized Prognostic Assessment in Myeloid Neoplasms" Blood 140, 9828-9830

Jacobs Flavia, D'Amico Saverio et al. "Opportunities and Challenges of Synthetic Data Generation in Oncology" JCO Clinical Cancer Informatics 7, e2300045

Asti Gianluca, D'Amico Saverio et al. "Synthetic Histopathological Images Generation with Artificial Intelligence to Accelerate Research and Improve Clinical Outcomes in Hematology" Blood 142, 902

Asti Gianluca, D'Amico Saverio et al. "Clinical Text Reports to Stratify Patients Affected with Myeloid Neoplasms Using Natural Language Processing" Blood 142, 122

Humanitas University
Via Rita Levi Montalcini, 4
20072 Pieve Emanuele (MI) Italy
Tel +39 0282241 - Fax +39 0282242394
info@hunimed.eu
hunimed.eu
CF 97692990159

Type of contract
Scholarship of € 25.000 gross per year awarded by Istituto Clinico Humanitas. This sum is subject to IRPEF income tax and exempt from social security contributions.

Borsa di studio pari a € 25.000 annui lordi erogata da Istituto Clinico Humanitas. Importo soggetto a tassazione IRPEF ed esente da contribuzione previdenziale.

Humanitas University
Via Rita Levi Montalcini, 4
20072 Pieve Emanuele (MI) Italy
Tel +39 0282241 - Fax +39 0282242394
info@hunimed.eu
hunimed.eu
CF 97692990159