Data Science for Health at FBK: AI needs and challenges

AI, FHIR, EHDS & the AI Act: Impacts on Healthcare Software and Standards

Flavio Ragni, PhD, *Fondazione Bruno Kessler, Trento, Italy*

FONDAZIONE BRUNO KESSLER

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

PNC
Piano nazionale per gli investimenti complementari al PNRR
Ministero dell'Università e della Ricerca

D3 4 HEALTH
Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care

Profile

Flavio Ragni, PhD
Researcher

Profile
Neuroscientist (MSc, PhD), Data Scientist

Expertise
Development of AI models for a wide range of diseases (neurodegenerative diseases, ophthalmology, cancer, …) and clinical goals
(diagnosis, prognosis, patient stratification, survival).
Analysis of diverse data types (EHR, brain imaging, eye imaging, omics, …).

Stefano Bovo, PhD
Researcher

Profile
Biomedical Engineer, PhD in Neuroscience

Expertise
Analysis of different data types: EHR, imaging (MRI, PET, CT), signals (EEG, ECG, TMS-EMG/EEG)
Focus on neurodegenerative diseases and various types of cancer
Designing projects and building collaborations

DSH
DATA SCIENCE FOR HEALTH

Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
PNC
Piano nazionale per gli investimenti complementari al PNRR
Ministero dell'Università e della Ricerca
D3 4 HEALTH
Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care

Main projects and activities

D³4HEALTH
Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care

***Transforming** and **advancing** current **methodologies** regarding the **diagnosis**, **monitoring** and **therapy** of the reference diseases, facilitating the application of **precision medicine** approaches through the **development** of the **digital** and biological **twins**.*

D³4Health five reference diseases:

- Metastatic Colon Cancer
- Liver and Bile Duct Cancer
- Central Nervous System Cancer
- Type I Diabetes
- Multiple Sclerosis

Main projects and activities

FONDAZIONE BRUNO KESSLER

NET-2018-12366666
NeuroArt P3

*Enhance the understanding of disease progression and predict patient care and disease trajectories, by applying **Machine Learning** and **Deep Learning** techniques to patient's data.*
*Easily **accessible** clinical data collected during routine clinical practice to **improve algorithms usability** in a wider range of contexts.*

NeuroArt P3 reference diseases:

- Parkinson's disease
- Alzheimer's disease
- Multiple Sclerosis
- Central Nervous System Cancer

OSPEDALE POLICLINICO SAN MARTINO
Sistema Sanitario Regione Liguria

Azienda Provinciale per i Servizi Sanitari
Provincia Autonoma di Trento

FONDAZIONE DON CARLO GNOCCHI ONLUS

Università di Genova

GASLINI

I.R.C.C.S. Ospedale San Raffaele
Gruppo San Donato

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

PNC
Piano nazionale per gli investimenti complementari al PNRR
Ministero dell'Università e della Ricerca

D3 4 HEALTH
Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care

ML/DL workflow

1

Problem Definition

2

Data Gathering

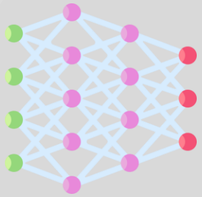
3

Data Preprocessing

4

Data Splitting

5



Model selection and training

6

Model evaluation

7

Model interpretation

8

Deploy

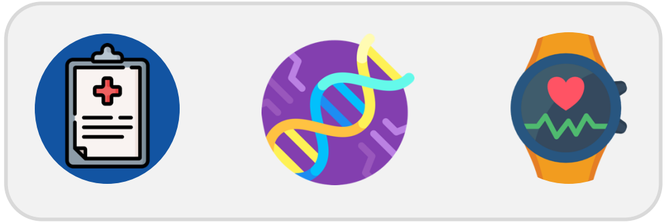
Data types and challenges

FONDAZIONE BRUNO KESSLER

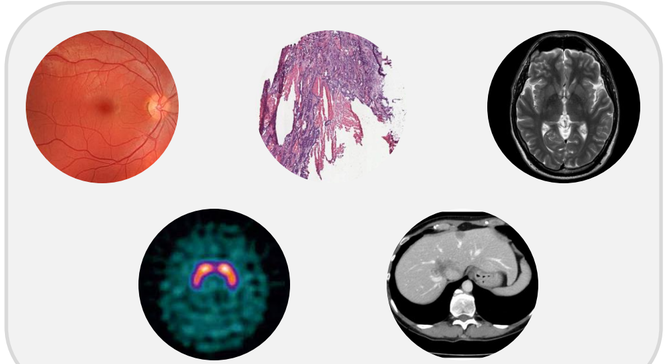
Data Gathering

Data Preprocessing

TABULAR DATA



IMAGING DATA



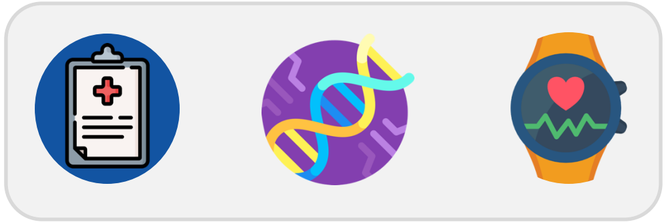
Data types and challenges

FONDAZIONE BRUNO KESSLER

Data Gathering

Data Preprocessing

TABULAR DATA



IMAGING DATA

Lack of standardization

We receive data in non-standardized formats: file types, variable names, and units of measurement often differ across data providers (e.g., hospitals, research institutes, registries).

Fragmented understanding of clinical concepts

We must rely on local codebooks (when available) and clinicians' expertise to interpret variables and clinical definitions.

Limited data quality assurance

Most quality checks happen ad-hoc during preprocessing, and it is often difficult to distinguish true outliers from data errors due to limited domain knowledge and missing validation rules.

Inconsistent pseudonymization practices

There is no shared approach to pseudonymization, leading to heterogeneous procedures and complicating safe data linkage across sources.

Project-specific preprocessing pipelines

Each project requires its own way to read, harmonize, and preprocess data; a single generic pipeline is rarely suitable, making the process time-consuming and hard to reuse.

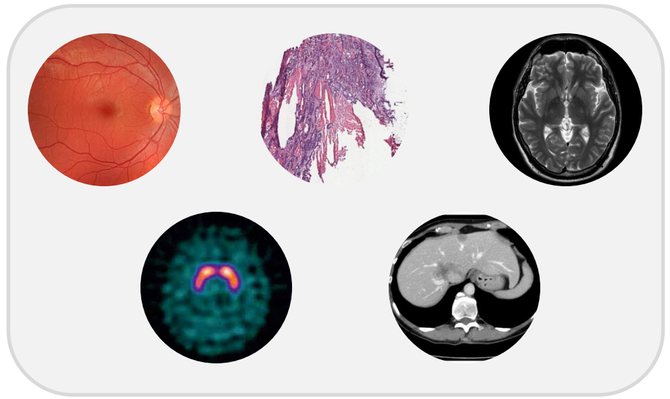
Data types and challenges

Data Gathering

Data Preprocessing

TABULAR DATA

IMAGING DATA



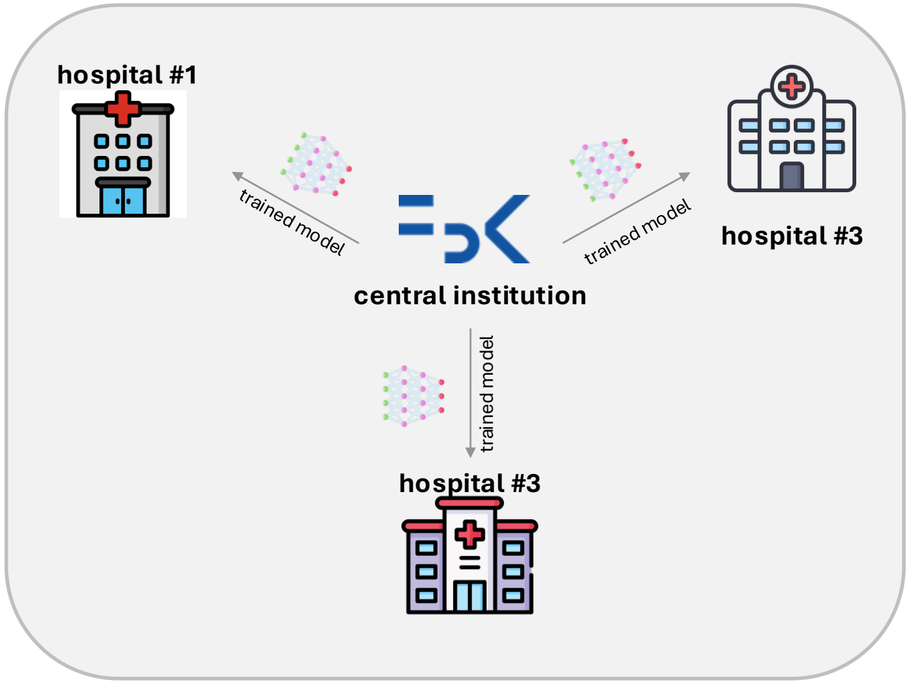
Limited visibility on prior image processing.
Most information needed to preprocess and analyze imaging data is stored in the image header, but previous processing steps (e.g., denoising, normalization, resampling, cropping) are often undocumented or incomplete.

Proprietary and heterogeneous imaging formats.
In some domains (e.g., ophthalmology, digital pathology), images are stored in vendor-specific formats. Metadata and acquisition parameters may be saved in different locations, using different codes and structures.

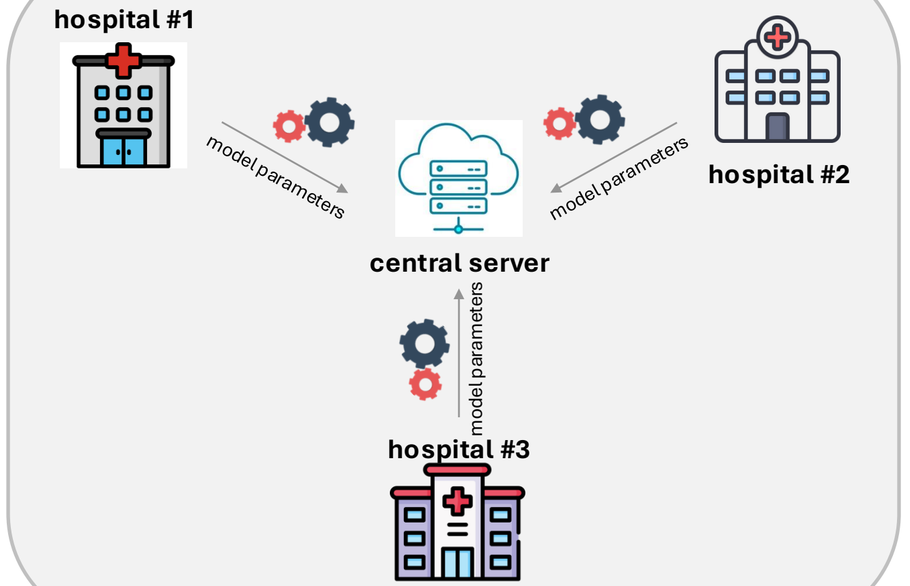
Inconsistent pseudonymization and de-identification.
Pseudonymization procedures for imaging and headers vary widely across sites, making safe data sharing and linkage difficult to standardize.

Deploy

DECENTRALIZED EXTERNAL VALIDATION



FEDERATED LEARNING




FONDAZIONE
BRUNO KESSLER