



ArgumeNtaTion-Driven explainable artificial intelligence fOr digiTal mEdicine (ANTIDOTE)

CALL CHIST-ERA 2019 - XAI

Coordinators:

Elena Cabrio, Serena Villata



Scientific background

Motivations:

- in neural architectures the correlation between internal states of the network (e.g., weights assumed by single nodes) and the justification of the network classification outcome is not well studied;
- high quality explanations are crucially based on argumentation mechanisms (e.g., provide supporting examples and rejected alternatives);
- in real settings, providing explanations is inherently an interactive process involving the system and the user.

ANTIDOTE fosters an **integrated vision of Explainable AI** (XAI), where low level characteristics of the deep learning process are combined with higher level schemas proper of the human argumentation capacity.

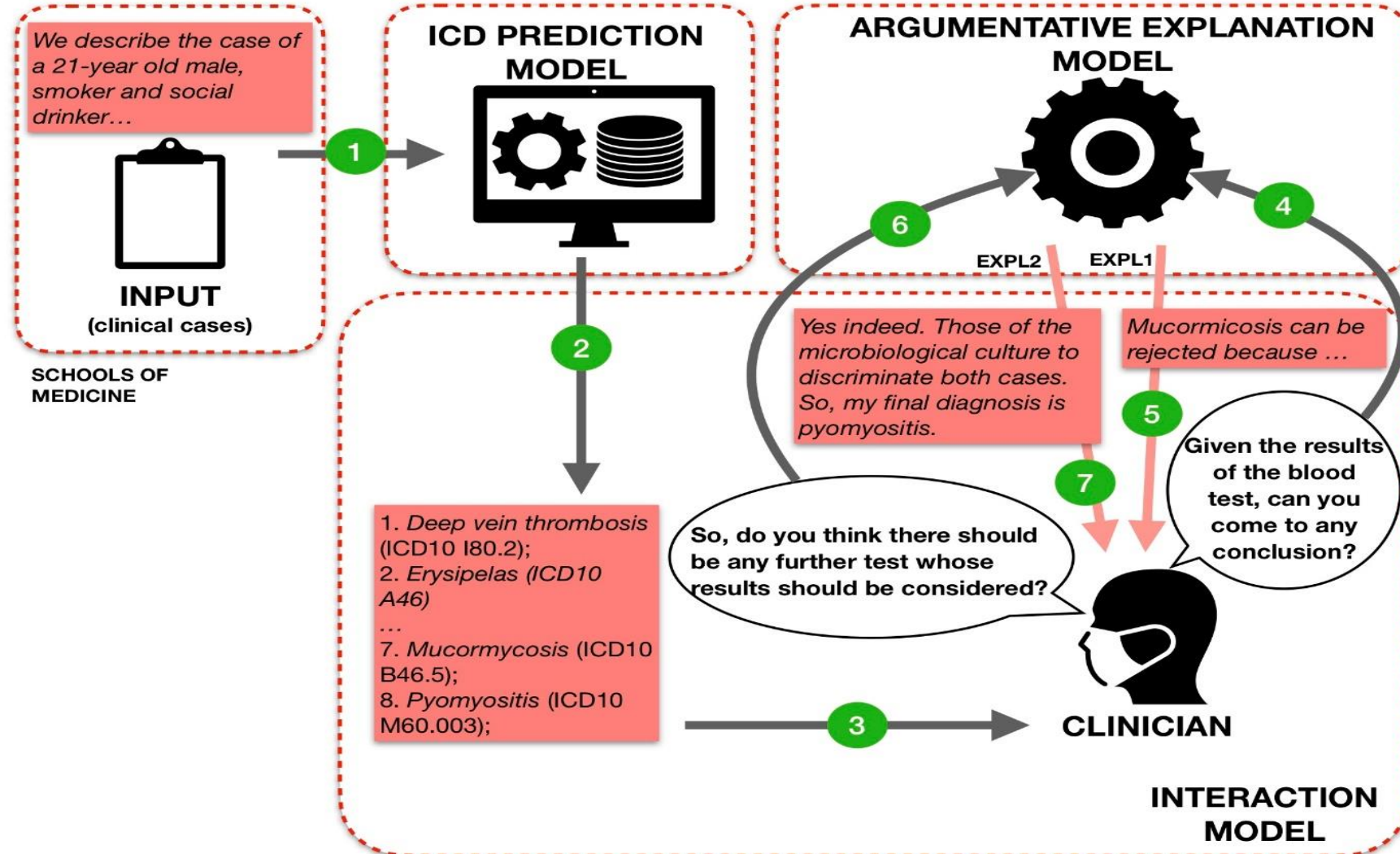
Objectives

- **Focus:** a number of deep learning tasks in the **medical domain**, where the need for **high quality explanations for clinical cases deliberation** is more critical than in other domains
- **General objective:** providing a **unified computational framework for jointly learning clinical predictions and the associated argumentative justifications**, fostering a natural interaction with clinicians through explanatory dialogues.

Key challenges

- Take advantage of past research on both theoretical and computational models of **human argumentation** to complement state-of-the-art approaches on explainability
- **Automatically learn explanatory argumentation schemas** in Natural Language (NL) and effectively combine evidence-based decision making with high level explanation schemas
- Combine a **task-specific prediction model** and a **general argumentation model** to produce explanatory dialogues

Proposed methodology



Consortium



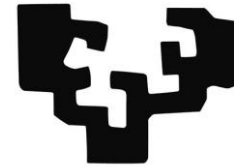
Elena Cabrio, Serena Villata,
Project Coordinators

Argument Mining and Generation (+ Medical school)



Bernardo Magnini

Dialogues, conversational agents



UPV EHU

Rodrigo Agerri

*Information Extraction,
Medical text mining
(+ Medical school)*



Marie Francine Moens

Machine Learning for NLP



Marcin Lewiński

Argumentation theory



Expected impact

- ***Scientific impact:*** results will be published in top-ranked conferences and journals
- ***Technological impact:*** adoption of AI technologies in healthcare (support from schools of medicine and health institutions)
- ***Societal impact:*** a decision support tool will be made at disposal of practitioners (SDG goal 3, related to health and well-being)
- ***Training impact:*** applied to training, it could be useful for students to test their knowledge