



Thesis Project Form

Title (tentative): Development of NLP and Machine Learning Tools for the Automatic Extraction of EQUAL Candida Score Components from Italian Clinical Notes

Thesis advisor(s): Giacomini Mauro, Daniele Roberto Giacobbe

E-mail: Mauro.Giacomini@unige.it

Address: Via Opera Pia 13

Phone: (+39) 010 33 56546

Description

Motivation and application domain

The timely and accurate assessment of candidemia management quality is essential for critically ill patients. Automating the calculation of the EQUAL Candida Score using AI applied to clinical notes could significantly enhance clinical decision support systems in infectious disease management.

General objectives and main activities

This thesis aims to extend the previous work of our research group on the automated detection of central venous catheter (CVC) events (which are part of the information needed for automatically calculating the EQUAL Candida Score) by developing methods to extract additional components of the EQUAL Candida Score from unstructured clinical notes written in Italian. The student will:

- Conduct a literature review on the EQUAL Candida Score and its components (e.g., blood culture timing, antifungal treatment, echocardiography, ophthalmologic exams);
- Analyze the available dataset of ICU patient clinical notes and annotate relevant concepts;
- Develop a pipeline using Natural Language Processing (NLP) techniques to preprocess, tokenize, and embed the texts;
- Train and evaluate machine learning and transformer-based models to classify clinical events or extract key temporal information;
- Compare model performance across different EQUAL Score items and optimize the pipeline for integration into real-time monitoring systems.

Training Objectives (technical/analytical tools, experimental methodologies)

The student will gain practical experience in biomedical text mining, clinical NLP in the Italian language, machine learning for healthcare, and evaluation of AI tools in a clinical setting. They will also develop skills in interdisciplinary research between engineering and medicine.

Place(s) where the thesis work will be carried out: DIBRIS, IRCCS Policlinico San Martino

Additional information

Maximum number of students: 1