Probabilistic Inference of Comorbidities from Symptoms in Patients with Atrial Fibrillation: An Ontology-Driven Hybrid Clinical Decision Support System

Alexander Lacki[1], Diego Boscá Tomás[2], Antonio Martínez-Millana[1]

[1]ITACA Institute, Universitat Politècnica de València, Valencia, Spain
[2]VeraTech for Health, Valencia, Spain

Introduction
Atrial fibrillation (AF) is a cardiological disease, its risk factors and mechanisms are, however, often rooted in non-cardiological comorbidities, introducing complexity in patient treatment. We present a clinical decision support system (CDSS), which aims to mitigate potential negative impact of the cross-disciplinarity of AF, by predicting comorbidities related to AF mechanisms based on provided symptoms.

Methods
MIMIC-III, a critical care database containing electronic health records (EHRs) from 38,597 patients is used. Diagnostic codes are captured, and symptoms are identified from free-text clinical notes using named entity recognition, and entity linking. Diagnoses and symptoms are mapped to a domain ontology, acting as a knowledge base. The obtained dataset of symptoms and corresponding diagnoses for each patient is employed to train an ensemble of Naive Bayes classifiers, with each classifier predicting the probability of a given diagnosis using the present symptoms. Performance is assessed using top-k accuracy and 5-fold cross validation. A CDSS integrating the knowledge base, classifier ensemble, and a communication interface is implemented.

Results
10,277 patients with AF were identified in the database. Patients had 7.58+-4.19 symptoms, and 2.92+-1.96 comorbidities. The ensemble of Naive Bayes classifiers predicts the probability of 18 comorbidities from up to 43 symptoms with a top-k accuracy of 0.51, 0.66, and 0.77 for k = 1, 3, and 5 respectively.

Conclusions
The present work demonstrates the ability of probabilistic classifiers to identify AF related comorbidities from patient reported symptoms. The use of such systems could support clinicians in the identification of comorbidities and mechanisms driving the arrhythmia, thus providing support in the selection of the optimal treatment.