

# An Annotated Corpus of Textual Explanations for Clinical Decision Support

Roland Roller<sup>1,2</sup>, Aljoscha Burchardt<sup>1</sup>, Nils Feldhus<sup>1</sup>, Laura Seiffe<sup>1</sup>, Klemens Budde<sup>2</sup>, Simon Ronicke<sup>2</sup>, Bilgin Osmanodja<sup>2</sup>  
<sup>1</sup>German Research Center for Artificial Intelligence (DFKI), Berlin, Germany, <sup>2</sup>Charité – Universitätsmedizin Berlin, Germany

## Abstract

Machine Learning for Clinical Decision Support (CDS): a good performance might be essential, but the aspect of trust should not be underestimated. For the treating physician using such a system and being (legally) responsible for the decision made, it is particularly important to understand the system's recommendation.

**Contribution of this work:** a) Analysis of how explanations of physicians look like (also in comparison to an ML system), b) publishing a novel annotated dataset in German

## Background

-Testing the performance of an ML-based CDS system against/with physicians to forecast outcomes of transplanted kidney patients.

-Predicting the likelihood of a negative outcome (0-100)

-ML-System uses explanations in form of local/global features

-Physician (junior/senior) made the same estimation as ML-system & wrote a short justification

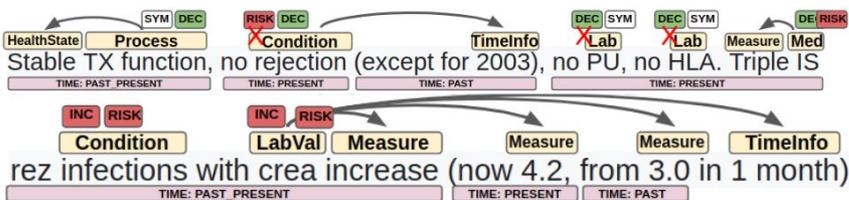
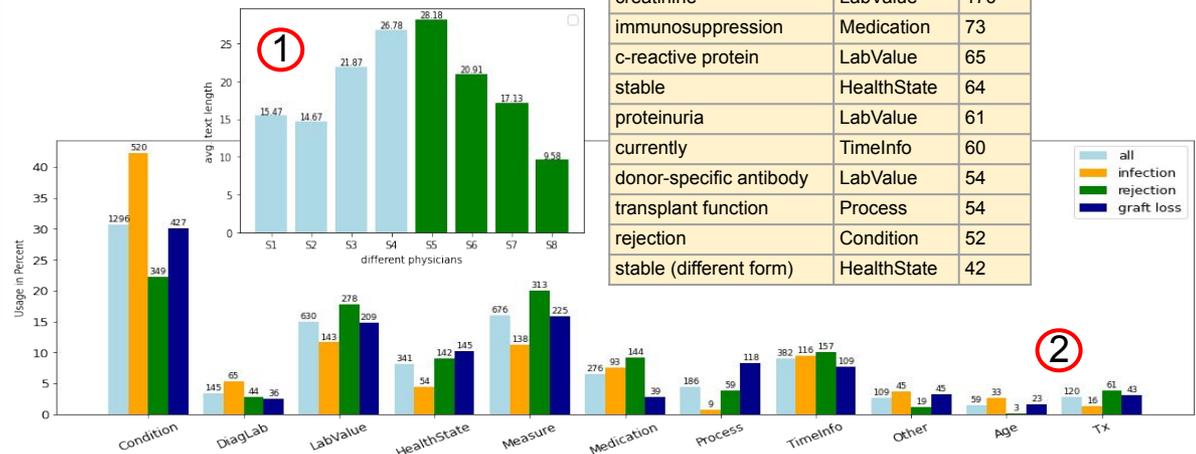


Figure: Upper part is partially extracted from a positive (task=rejection, score=0) and the lower one partially from a negative (task=graft loss, score=78) explanation.

## Annotations

To analyse the collected explanations, annotations were carried out on different levels: a) Entities (e.g. Med-Condition, Medications), b) Relations (e.g. hasMeasure), c) Temporal Aspects (e.g. past, past-present), d) factuality, e) progression (e.g. increase/decrease)

## Quantitative Analysis



Object 1: Chart shows the avg. text length of the 8 physicians (left=junior; right=senior); Object 2: Shows the usage of entities in percent, overall, and for the 3 tasks; Object 3: Shows the Top-10 annotations

## Comparing ML-based and human explanations

**Object 4 (top): Top-10 frequently used local features of the ML-based CDS;**

Feature	#
body size	134
blood pressure (diastolic)	118
last creatinine value	94
# hospitalizations in last year	67
mean CRP value	59
current weight	57
last hsthp value	55
age	54
mean creatinine value	53
body temperature	48

Object 4 (top): Top-10 frequently used local features of the ML-based CDS; Object 5 (bottom): Top-5 global features for the prediction of rejections.

**Object 5 (bottom): Top-5 global features for the prediction of rejections.**

Feature	Value	Import
last creatinine value	float	12.78
months since transplantation	int	7.66
had rejection in last 180 days	binary	6.69
days since transplantation	int	6.69
#lab values in the last 60 days	int	3.18

## Aspects of human explanations

We observed different explanation patterns in the human explanations, some might be easy to implement for a CDS, and some more complex

- 1) Explicit Description;
- 2) Tendencies and Fluctuations;
- 3) Factuality;
- 4) Values In-/Outside the Norm;
- 5) Interpretation;
- 6) High Level Interpretation

**Acknowledgement:** This research was supported by the German Federal Ministry of Education and Research (BMBF) through the project **VALID** (01GP1903A).