UNIVERSITY OF TWENTE.

# Exploiting Natural Language Processing for Improving Health Processes

M. van Keulen[1], J. Geerdink[2], G. Linssen[2], R.H.J. Slart[1] and O. Vijlbrief[2]

[1] University of Twente, {m.vankeulen;r.h.j.a.slart}@utwente.nl

[2] Hospital Group Twente, {j.geerdink;g.linssen;o.vijlbrief}@zgt.nl

## A wealth of textual data in EPDs

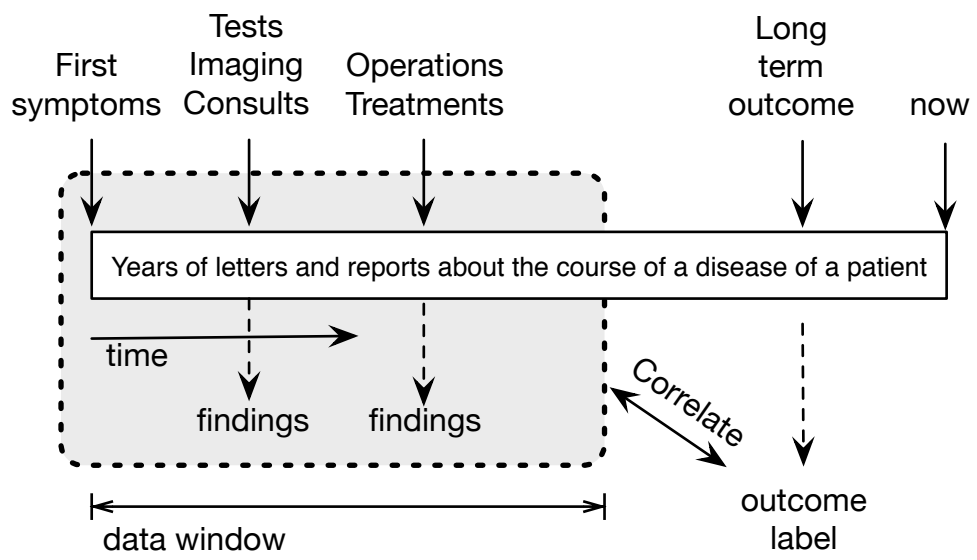
A care process often consists of many steps involving many actors. At every stage, clinicians document observations and interpretations of consults, tests, images, treatments, and operations in reports. Our main question is: Can NLP exploit this wealth? By extracting structured data and using it as features for machine learning, a wide variety of process improvements become possible. This poster sketches a few prominent process improvements that we plan to research.

## Improving Complex Diagnostic Processes

Many conditions are hard to diagnose. For example, Giant Cell Arteritis (GCA) is an immune mediated vasculitis characterized by inflammation of the large and medium sized arteries. It is uncommon (25–70 per 100.000) and presents itself only with atypical signs and symptoms such as low-grade fever, malaise, and weight loss. Its diagnosis is difficult and can be missed easily. Untreated it may lead to loss of eye sight and cerebral stroke. Several diagnostic modalities are essential for diagnosis, but machine learning may pick up subtle patterns of clues scattered over various reports containing conclusions from these modalities.

Chemo and radiation therapy in cancer patients damage heart and bloodvessels. It is important to identify those patients who are at high risk for long term complications. They need tailored medical therapy to prevent cardiovascular diseases. Machine learning may aid in identification, avoiding or suggesting such measures.

## Data windows




By varying the data window and assessing prediction performance, one may determine from which step in the care process a sufficient prediction performance may be expected. A learned model may be unobtrusively incorporated in existing EPD software warning the clinician of substantial risk for uncommon conditions like GCA.

## Dynamic Shortening of Care Processes

Actions like tests, imaging, treatments, etc. are preceded by consults. Patterns may be discovered between data in letters and reports from before a consult, that sufficiently predict a subsequent action. In such cases, one may dynamically adapt the process to already plan this action before the consult, at the limited expense of some cancellations due to mispredictions, or one may even omit the consult altogether.

## Quality Assurance

Hospitals often enforce strict protocols as a means for quality assurance. NLP techniques may aid in checking adherence to protocols and supporting clinicians in upholding the standards. For example, the radiology diagnostic reports for breast cancer in the Hospital Group Twente contain among other things a standard risk classification, called “Breast Imaging Reporting and Data System”. Whether or not these risk scores correlate well with actual outcomes is largely unknown. Being able to check this is important for quality assurance. Moreover, discovering under which circumstances deviating risk assessments are given, may aid in improving care processes with better-informed decisions.

| BIRADS score | Malignant |
|---|---|
| 1 & 2 | 0% |
| 3 | 0-2% |
| 4a | 2-10% |
| 4b | 10-50% |
| 4c | 50-59% |
| 5 | >95% |
| 6 | proven |