A New Era for Health Research by Using AI

Kjersti Aasgård
Gina Helstad
A New Era for Health Research by Using AI

Results from a pilot study demonstrating how AI can analyze historical patient records

Gina Helstad and Kjersti Aasgård
Film om NHA?
Introduction

- Part of the National Archives of Norway
- Established in 2019
- Two purposes:
  - Making health data available for research
  - Securing and preserving patient records from deceased patients
- 36 million digitized pages pr year
- All hospitals are required by law to transfer all patient records to The Norwegian Health Archives after the patient is dead
- First national health archive in the world

Photo: The National Archives of Norway and Vegard Breie
"To become a recognized health registry for new knowledge and better public health"

Strategic goal
INPUT:
• 15 million paper-based patient records
• 150,000 shelf meters in total
• Electronic patient records: backlog of 2 million

PROCESS:
• Digitization and disposal of the paper
• Metadata registration:
  • Personal information
  • Diagnoses
• Digital preservation system
• OAIS compliant

OUTPUT:
• Digital access
• Statistics and complete records (images and OCR-text)
• Access only by application
• Researchers and next of kin
The Norwegian Health Archive Registry
Content overview September 2022

- 700,000 patient records
- 53 hospitals
- 3,700,000 diagnoses
- 1871 – 2016: current time span
- 244 million files
- 100 TB data
- 85 million pages digitized

Photo: Vegard Breie
The Health Archives Registry

HARI

Metadata
Patient ID, diagnoses
structured

Digital Preservation System

Images – digital format
OCR text
unstructured
The Health Archives Registry contains both **structured** and **unstructured data**

- The **structured** data are systemized, «tidy», easy to use and well suited for statistical analysis.

- The **unstructured** data can be more challenging to handle and access, but that does not mean the data are less valuable.

Photo: Pixabay and Makhbubakhon Ismatova
Klokketest

Instruksjoner:
På fôret under, vennligst tegn en urkive og plasser alle tallene på korrekt sted.
Etter det, tegn inn viserne slik at klokken viser 11 minutter over elleve.
Innovation partnership with Anzyz Technologies

- Use artificial intelligence to find relevant information in big data
- Test and develop AI to find information that is irretrievable using metadata
- OCR is a prerequisite
The pilot dataset and workflow

- 15,000 patient records from a university hospital
- Representative for the registry
- OCR-text is the input to the CCL algorithm
- AI-base established on random sample of data set
- Define statistical categories in the Anzyz Dashboard
- Output is a statistical report (.csv)
Build statistical categories

Figures: Anzyz Dashboard/ The Norwegian Health Archive Registry
Verify results and gain insight in the data set

Figures: Anzyz Dashboard/ The Norwegian Health Archive Registry
Autopsy reports
Penicillin allergy
Antibiotic resistance
Urinary tract infection and antibiotic use
Autopsy reports
### Results of statistical concept autopsy reports

**Total records in data set: 15 092**

<table>
<thead>
<tr>
<th>Type</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>F1-measure (%)</th>
<th>F2-measure (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCL identified as autopsy reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True autopsy reports</td>
<td>669</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False positives</td>
<td>86%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False negatives</td>
<td>14%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negatives</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Type**

- **PHR which included Autopsy reports**
  - Recall (%): 100
  - Precision (%): 86
  - F1-measure (%): 92
  - F2-measure (%): 97

**Results of statistical concept autopsy reports**

- **Total records in data set: 15 092**
- CCL identified as autopsy reports: 669
- True autopsy reports: 86%
- False positives: 14%
- False negatives: 0%
- True negatives: 100%
First PhD project hosted by The Norwegian Health Archives Registry

› An interdisciplinary project in three parts that combines:
  › resources and data from NHA
  › modern technology
  › public health knowledge.

› In the first article, AI technology will be further tested in regard to identifying autopsy reports in HARI.

Figures: The Norwegian Health Archive Registry
I hope to demonstrate how we can use AI methods to build relevant statistics and gain valuable insight into our extreme amounts of unstructured health data.

This way we can offer researchers better and more precise data products.

Elin Bjørnstad-Tuveng, PhD student