



Innhold i Helsearkivregisteret

Tema : Helsedata og forskning med kunstig
intelligens



Foto : Vegard Breie/Arkivverket

Fra pasientarkiv til helsedata

Norsk helsearkiv (etablert i 2019) mottar, digitaliserer og bevarer alle pasientjournaler fra avdøde pasienter i spesialisthelsetjenesten og tilgjengeliggjør dem, primært for forskning. Norsk helsearkiv forvalter Helsearkivregisteret, som er et av Norges sentrale helseregistre. Vårt overordnede mål er å være et anerkjent helseregister for ny kunnskap og bedre folkehelse.

Tall per 01.06.23:

Antall digitaliserte pasientjournaler: 935 719

Antall digitaliserte sider : 97 190 000

Kunstig intelligens og ustrukturerte helsedata

Det er gjennomført noen prosjekter med kunstig intelligens og naturlig språkanalyse av ustrukturert journaltekst i Norge, men av personvern hensyn har det vært vanskelig å få bruke pasientjournaler til dette. For å gjøre det enklere å bruke pasientdata ble det i 2021 gjort en endring i Helsepersonelloven, som forhåpentligvis vil stimulere utviklingen av slike løsninger. Nylenna m.fl. påpeker at det er viktig at norske helsemyndigheter kartlegger og oppmuntrer til utvikling av naturlig språkanalyseprosjekter i Norge (1).

Helsearkivregisteret bruker nå kunstig intelligens for å gjøre det enklere å generere data til forskningsbruk.



Foto : Bing image creator

Kunstig intelligens åpner nye dører for forskning i Helsearkivregisteret

Helsearkivregisteret har strukturerte metadata som er lett tilgjengelige og lett søkbare. Dette inkluderer bl.a. navn, personnummer, dato for fødsel, død og innleggelse, og diagnoser i form av både tekst og ICD-koder. I tillegg utgjør hele journaler store mengder ustrukturerte data, som er utfordrende å finne frem i. Hvis en forsker ønsker et datasett med for eksempel spesifikke blodprøvesvar finner vi ikke frem til disse ved hjelp av metadata, og manuell letning vil i praksis være umulig med våre datamengder.

Alle digitaliserte pasientjournaler blir imidlertid OCR-lest, noe som gir mulighet til å finne frem i den ustrukturerte delen av pasientjournalen - friteksten i journalen. Derfor har Helsearkivregisteret inngått et innovasjonspartnerskap med Anzyz Technologies og videreutvikler nå et forskningsstøtteverktøy basert på en språk algoritme (Natural Language Processing), som åpner helt nye muligheter. I første omgang brukes forskningsstøtteverktøyet på digitaliserte pasientjournaler, men kan også benyttes på elektroniske pasientjournaler, som Helsearkiv-registeret etter planen skal starte mottak i løpet av 2023.



Foto : Eksempel på ustruktureerte helsedata/sider i pasientjournaler

Ved hjelp av KI (kunstig intelligens) vil Helsearkivregisteret lage og levere verdifull statistikk på ulikt innhold. Noe som også betyr mer presis identifikasjon av journaler eller journalsider som har det spesifikke innholdet forskere søker etter. Dette kan eksempelvis være; skjema for prøvesvar, observasjon, rekvisisjoner, epikriser, spesifikke symptomer, legemidler, behandlinger, osv. Det kan også være bredere konsepter, som for eksempel bruk av tvang, og fenomener som gjenspeiles i språk, som makt og holdninger. Innhold som kan danne grunnlag for både kvantitativ og kvalitativ forskning.

For eksempel kan bruk av et legemiddel enkelt identifiseres og kobles med andre metadata fra pasientjournalen. Dette åpner, blant annet, muligheter for å generere kontrollarmer til planlagte forskningsprosjekter. I tillegg kan enkeltdokumenter som inneholder spesifikt innhold hentes ut og enkelt anonymiseres for utlevering til forskning, i motsetning til anonymisering av hele pasientjournaler som i praksis er umulig. Produksjon av syntetiske data (kunstige data generert fra originale data) er også en hittil uutnyttet mulighet i Helsearkivregisteret.

Helsearkivregisteret har store mengder data som inneholder over 100 år med erfaring og kunnskap. Ikke bare de få utvalgte målepunkt som tradisjonelt sett rapporteres til helseregistre, men også all den finmaskede informasjonen som finnes i pasientjournalen. Kunstig intelligens gir innsikt og legger til rette for at det rike datamaterialet kan gjenbrukes på helt nye måter (2).



Prosjekter med data fra Helseregisteret

Obduksjonsrapporter

Obduksjoner har helt siden antikkens tid vært viktige for vår forståelse av menneskekroppens bygning og hvordan den forandres ved sykdom. Den er en viktig kilde for medisinsk forskning, obduksjonsresultatet utgjør en viktig del av datagrunnlaget i dødsårsaksstatistikken og ansees som nødvendig for at dødsårsaksstatistikken skal kunne betraktes som pålitelig (3,4,5,6). I et samarbeid med Anzyz technologies vil kunstig intelligens, en algoritme som kan utføre språkanalyse av store mengder ustrukturerte data, testes for sin evne til å identifisere obduksjonsrapporter i materialet ved Helsearkivregisteret.

Et pilotprosjekt ble utført i 2022 og språkalgoritmen klarte å identifisere obduksjonsrapporter i journalene mye raskere og med like god treffsikkerhet som mennesker (7). Resultatene kan overføres til andre dokumenttyper og konsepter i pasientjournalene ved Helsearkivregisteret og vil åpne nye dører for forskning. Obduksjonsrapportene og de tilhørende journalene skal benyttes videre i et ph.d.- prosjekt. Med iskemisk hjertesykdom som fokusområde skal diagnoser som pasientene fikk etter sin død, ved obduksjon, sammenlignes med kliniske diagnoser gitt mens de levde.

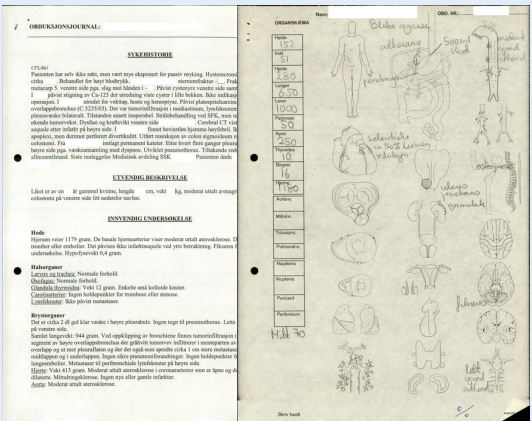


Foto : Eksempler på sider i en obduksjonsrapport



Klokketesten

Klokketesten er en kognitiv screeningtest som i flere tiår har vært benyttet som ledd i utredning og forløpskontroll av demens, hjerneslag o.l. (8). I forbindelse med et prosjekt ved Universitetet i Tromsø er det søkt om utlevering av anonyme klokketester fra Helsearkivregisteret. Klokketestene skal benyttes for å utvikle en algoritme der maskinen setter score etter å ha analysert bildet. Videre vil testens reliabilitet og validitet vurderes, samt hvilke aspekter ved tegningen som bidrar mest til scoren, slik at det kan skapes en ny tegnetest, fordi den gamle testen blir vanskeligere å bruke når det kommer nye generasjoner som ikke er like vant med gamle klokker.

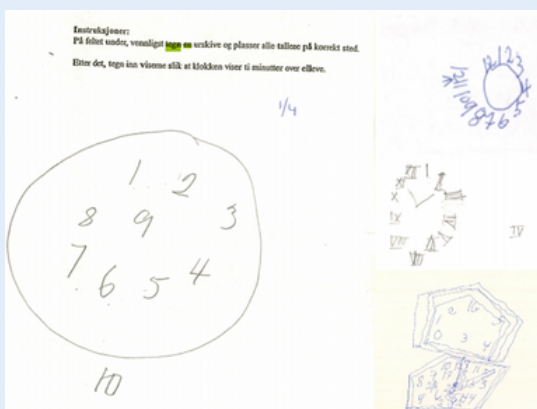


Foto : Eksempler på klokketester. Til venstre: klokketest scoret med 1 av 4 mulige poeng

Aidentifisering ved hjelp av maskinlæring?

Helsearkivregisteret har levert ut 3000 journaler til forskningsprosjektet CLEANUP, i regi av Norsk regnesentral (9). CLEANUP sitt hovedformål er å utvikle maskinlæringsmodeller som automatisk skal gjenkjenne personopplysninger i tekst, og redigere dokumentene for å maskere disse opplysningene. Det skal også evalueres hvor godt disse anonymiseringsmodellene fungerer, blant annet i hvilken grad de bidrar til å redusere avsløringsrisikoen knyttet til ulike tekstdokumenter. Dersom man lykkes med automatisk aidentifisering av journalsider vil det redusere det manuelle arbeidet ved utlevering av forskningsmateriale.

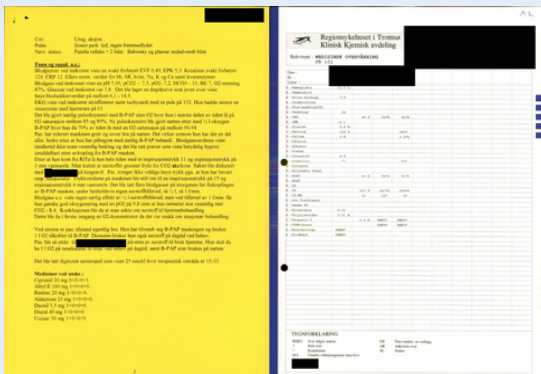


Foto : Eksempler på avidentifiserte dokumenter

Ønsker du å vite mer?

Ta gjerne kontakt med oss for å avtale et møte. Vi kommer gjerne til ditt arbeidssted for å holde en presentasjon om innholdet i Helsearkivregisteret.
<https://helsedata.no/no/forvaltere/arkivverket/helsearkivregisteret>

helsearkivregisteret@arkivverket.no

Referanser

1. Pasientjournalen – språk, dokumentasjon og helsekompetanse.
<https://www.michaeljournal.no/journal/1000/31>
2. Helsearkivregisteret åpner nye dører for forskningen.
<https://tidsskriftet.no/2022/11/tra-fagmiljoene/helsearkivregisteret-apner-nye-dorer-forskningen>
3. Organisering og nytteverdi av obduksjon.
<https://tidsskriftet.no/2012/01/originalartikkel/organisering-og-nytteverdi-av-obduksjon>
4. 75 års jubileum 1923–1998 [Jubileumsbok].
<https://patologi.com/jubileumsbok%201998.pdf>
5. Dødelighet og dødsårsaker i Norge gjennom 60 år 1951-2010.
<https://www.fhi.no/globalassets/dokumenterfiler/helseregistre/dar/dodelighet-og-dodsarsaker-pdf.pdf>
6. NOU 2011: 21. Når døden tjener livet – Et forslag til nye lover om transplantasjon, obduksjon og avgivelse av lik.
<https://www.regjeringen.no/no/dokumenter/nou-2011-21/id666923/?ch=1>
7. Identifying autopsy-reports in unstructured, digitized patient records: A pilot-study from the Norwegian Health Archives Registry.
<https://ebooks.uis.no/index.php/USPS/catalog/book/195/>
8. Norsk revidert klokketest (KT-NR2)
<https://web.archive.org/web/20150526023316/http://www.aldringoghelse.no/ViewFile.aspx?ItemID=6466>
9. The CLEANUP project, Norsk regnesentral.
<https://cleanup.nr.no/>