

Kunstig intelligens og norske helsedata

Kunstig intelligens er en av de mest lovende teknologiene for å forbedre helsetjenesten. Overordnet handler det om å gjøre maskiner i stand til å løse oppgaver som tidligere var forbeholdt mennesker.

Tidligere var programmerte, regelstyrte systemer rådende, men de siste årene har teknikker som gjør at maskiner kan lære selv blitt dominerende. Kjernen i slik maskinlæring er at algoritmer og data lærer opp en modell til å gjøre komplekse oppgaver og finne nye mønstre.

Muligheter for helsetjenesten

Presise diagnoser

Maskinanalyser av medisinske bilder kan bidra til raskere og mer presise diagnoser, som å tolke vevsprøver av [prostatakreft](#), finne [polypper og svulster i tarmen](#) ved skopiundersøkelser, og automatisk tolke røntgenundersøkelser av [lungene](#). Det første systemet som kan stille selvstendig diagnose av [øyesykdom](#) er nå godkjent i USA.

Persontilpasset behandling

Ved hjelp av maskinlæring på pasientjournaler kan man finne tidligere pasienter som likner en ny pasient, og foreslå den behandlingen som antakelig vil virke best. Maskiner kan også finne nye varianter av en sykdom som kan trenge forskjellig behandling. Svenske forskere har slik funnet [fem undergrupper av diabetes type 2](#).

Tidlig forebygging

Forutseende analyser kan oppdage risiko, som gjør det mulig å sette i gang tiltak tidlig og forebygge målrettet. Britiske forskere har utviklet en algoritme som kan oppdage risiko for [akutt nyresvikt](#) 48 timer tidligere enn med dagens blodprøvetester.

Selvhjelp

Persontilpassede apper kan hjelpe pasienter å håndtere

SAMMENDRAG

- » Kunstig intelligens kan gi både bedre forebygging, diagnose, behandling og selvhjelp.
- » En forutsetning er løpende tilgang på helsedata, som medisinske bilder, pasientjournaler og innbyggernes egne målinger. Det vil gjøre helseregistre relativt sett mindre viktige fremover.
- » Rapporten foreslår tiltak for å håndtere personvern, datasikkerhet, forskningsetikk og fordeling av verdiene som skapes når helsedata brukes i større omfang og på nye måter.

sykdom selv. Astmapasienter kan [håndtere sykdommen bedre](#) når de får daglige, persontilpassede råd om medisinbruk basert på maskinlæring av data fra bruk av astmainhalator, symptomer, vær og luftkvalitet.

Fra helseregistre til en strøm av data

Algoritmer blir vanligvis bedre jo mer data de får tilgang til. Norge har sterke helseforskningsmiljøer, blant annet på grunn gode helseregistre. Regjeringens hovedsatsning for helsedata er [Helseanalyseplattformen](#), som skal [forenkle](#) tilgangen til helsedata. Så langt er [helseregistre](#) [prioritert](#).

Sentraliserte registre er viktige, men vil likevel få mindre betydning fremover fordi maskinlæring ikke er avhengig av slike strukturerte data. Pasientjournalen gir et rikt bilde av hver pasient, og inneholder en stor variasjon av informasjon som medisinske bilder, prøvesvar, målinger, diagnoser og behandlinger over tid.

Innbyggerne kan dessuten frivillig supplere med egne målinger som aktivitet, hjerterytme og vekt.

De store mengdene varierte data som daglig produseres i klinikkene, og av innbyggerne selv, vil dermed bli den viktigste kilden for nye algoritmer. Dette vil utfordre dagens praksis for å håndtere personvern, datasikkerhet, forskningsetikk og fordeling av verdier.

Helsehjelp og personvern

Pasientjournalen blir opprettet for å gi en pasient direkte helsehjelp. Analyser av store samlinger med pasientjournaler kan imidlertid også gi bedre helsehjelp til både pasienten selv og andre pasienter. Denne måten å bruke pasientdata på er ikke uproblematisk, fordi helseopplysninger er sensitive og må vernes godt.

I takt med utviklingen av mer intelligent maskinlæring blir det mulig å utlede mer informasjon, spesielt ved å koble sammen datasett, som enkeltvis har god personvernbeskyttelse:

- *Anonymiserte datasett blir identifiserbare:* Anonymiserte datasett kan i dag brukes fritt. Forskere mener at [15 demografiske opplysninger](#) er tilstrekkelig for å identifisere 99,98 prosent av amerikanere fra åpne, anonymiserte datasett.
- *Ikke-sensitive data blir sensitive:* Nye sammenhenger i data som i utgangspunktet ikke er sensitive, kan likevel avsløre sensitiv informasjon. Facebook kan bl.a. [utlede](#) seksuell orientering, angst og risiko for selvmord ut fra brukernes nettaktivitet.

Motstridende hensyn

Hovedregelen i [personvernforordningen](#) (GDPR) er at den registrerte må gi et informert samtykke for at egne data skal kunne brukes til andre formål enn det opprinnelige.

Imidlertid er det åpninger i lovverket for at pasientjournalen uten samtykke fra pasienten kan utleveres til andre formål enn direkte helsehjelp, blant annet [helseforskning](#) og [kvalitetssikring av helsehjelpen](#). Forutsetningen er at bruken er av vesentlig interesse for samfunnet. Samtidig må hensynet til pasientens integritet og velferd være ivarettatt

Regelverket er imidlertid ikke tydelig på hvordan pasientinformasjon om andre pasienter enn den som er kommet til behandling kan brukes til å utvikle og bruke maskinlæring.

Risiko for høy bot for å komme i konflikt med personvernforordningen kan gi vegring mot å utlevere og behandle helsedata uten samtykke. Det kan hindre en viktig utvikling av helsehjelpen og føre til ulik praksis. Uklarhetene har blant annet hindret [Sørlandet sykehus](#) i

å ta i bruk et system som kan avverge alvorlige allergiske reaksjoner under operasjon.

Balanse mellom personvern og helsehjelp

Det vil bli behov for å balansere og forene hensynene til helsehjelp og personvern bedre. Mulige tiltak kan være å klargjøre regelverket, forbedre pasientenes muligheter til å involvere seg, samt å gjøre data tilgjengelig bare for maskiner.

Klargjøring av regelverket

Det danske parlamentet har [endret helseloven](#). Hensikten er å gi et klarere hjemmelsgrunnlag for behandling av helseopplysninger fra elektroniske systemer til bruk til beslutningsstøtte. Nærmere regler skal senere fastsettes i en forskrift. Norske myndigheter bør også gjennomgå helse- og personvernlovgivningen slik at maskinlæring kan tas i bruk til å forbedre helsetjenesten.

Reservasjons- og samtykketjeneste

Det vil bli meget krevende å håndtere forespørsler om et stort antall samtykker til maskinlæring i helsetjenesten fremover. I tillegg er det risiko for at mange enten ikke er i stand til, eller ikke bryr seg om å gi samtykke. Det kan gi mangler og skjevheter i datagrunnlaget som vil være uheldig for treffsikkerheten til systemene som læres opp fra dataene. Dette vil være en begrunnelse for å åpne for å bruke helsedata uten et eksplisitt samtykke.

England har etablert en sentral reservasjonstjeneste, [National data opt-out](#). Her kan innbyggerne reservere seg mot at deres helsedata brukes av organisasjoner som har lov hjemmel til å bruke helsedata til forskning, planlegging og forbedring. Økt åpenhet om hvordan helsedata brukes og bedre mulighet for å reservere seg, vil styrke innbyggernes selvbestemmelse og integritet.

En tilsvarende tjeneste kan etableres, også i Norge. Tjenesten kan bygges ut til å nå ut til innbyggerne med forespørsler om både samtykke og informasjon om prosjekter de kan reservere seg mot. Etter hvert kan innbyggerne få finmasket oversikt over hvilke helsedata som brukes av hvem og til hva.

Dette vil gjøre håndtering av samtykker og reservasjoner enklere for de som ønsker å bruke helsedata til helseforskning og utvikling av helsetjenesten.

Gjøre data tilgjengelig for mennesker

Det er mulig å tilgjengeliggjøre data slik at det bare er maskiner som kan behandle dem. Data kan for eksempel kapsles inn i såkalte [sorte bokser](#). Sykehus kan invitere forskere og utviklere til å utvikle algoritmer, uten at de faktisk får tilgang til dataene.

Statistiske metoder kan [oppdage](#) og [justere](#) skjevheter i datasettene. Data fra flere datasett kan dessuten [krypteres](#) og [kobles](#) slik at algoritmer kan analysere data fra flere



aktører på en sikker måte.

Maskinlæring og datasikkerhet

I [Helsenæringsmeldingen](#) ønsker Regjeringen å «gjøre det mulig å benytte helsedata mer aktivt i utviklingsforløpet til legemidler og medisinsk teknologi».

[Regjeringen foreslår](#) å tilgjengeliggjøre helsedata ved å samle dem i sentraliserte og sikre omgivelser. Når volumet øker, kan dette bli en sårbar strategi. Sentraliserte dataressurser er attraktive å angripe, og konsekvensene er store dersom noen med uærlige hensikter får tilgang til slike verdifulle og sensitive data.

Håndtering av kopier øker dessuten kompleksiteten og arbeidsbyrden med å holde dataene oppdaterte og korrekte. Det kan derfor være hensiktsmessig å vurdere en alternativ tilnærming til bruk av helsedata i tiden fremover.

Desentralisering som strategi

En metode som prøves ut på sykehus i bl.a. [Storbritannia](#) og [USA](#), er å dele algoritmer og modeller i stedet for dataene. Teknikken kalles [distribuert læring](#) (federated learning): en algoritme sendes for eksempel til CT-bildedatabasene på alle involverte sykehus. Her brukes de lokale CT-bildene til å lære opp hver sin versjon av en modell. Endringene på ett sted deles med de andre, som justerer seg slik at modellene lærer av hverandre. Dataene beholdes med andre ord lokalt, mens det er endringene i modellene som flytter på seg.

Distribuert læring kan gjøres uten å [utlede sensitiv informasjon](#). Metodene kan [kombineres med personvernfriende teknikker](#) og slik styrke personvernet.

Norske sykehus har allerede elektroniske pasientjournal-systemer, og kan med denne teknikken komme raskt i gang med å lære opp og validere algoritmer med helsedata fra egne pasienter. Samtidig kan algoritmene dra nytte av data fra populasjoner fra andre sykehus, noe som gjør dem mer robuste.

450 fastleger skal i løpet av 2020 kobles opp på [PraksisNett](#), som er en infrastruktur der forskere kan gjøre kliniske studier i norsk allmennpraksis. Her analyseres helsedata kun lokalt, og tilnærmingen passer dermed godt for å gjøre distribuert læring av algoritme på tvers av legekontorer.

Det krever at man tilrettelegger for maskinlæring når sykehusene, fastlegekontorene og kommunale helse- og omsorgstjenester skal oppgradere og kjøpe journalsystemer. Dagens planer tar ikke tilstrekkelig høyde for dette.

Pasienter som blir forskningsdeltakere

Maskinlæring kan dekke et stort spenn av kunnskapsutvikling i helsetjenesten, fra kvalitetssikring av helsehjelpen, til forskning og generering av ny kunnskap. Det blir viktig å beskytte pasientene mot skade og urimelige

belastninger dersom de også er forskningsdeltakere. Eksempelvis kan det få konsekvenser for pasientene dersom algoritmene viser at de tilhører en ny risikogruppe. Bør de få vite om det? Kan gruppen bli gjenstand for stigmatisering, manipulering eller nedprioritering i helsekøen?

Det er lovfestet at en regional forskningsetisk komité (REK) skal [forhåndsgodkjenne](#) at forskningen blir utført forsvarlig i alle [helseforskningsprosjekter](#), det vil si «virk-somhet som utføres med vitenskapelig metodikk for å skaffe til veie ny kunnskap om helse og sykdom».

Det er imidlertid ikke alltid mulig å si på forhånd om et maskinlæringsprosjekt vil komme frem til ny kunnskap, eller ikke. Metodikken for å utvikle maskinlæringsmodeller skiller seg dessuten fra tradisjonell vitenskapelig metodikk. I dagens lovverk er det et klart definert skille mellom helsehjelp og helseforskning. Maskinlæring gjør derfor dette skillet vanskelig.

På den ene siden vil det være uhenksiktsmessig å anse alle prosjekter som helseforskning, som dermed må forhåndsgodkjennes. Samtidig er det viktig å beskytte pasientene i de tilfellene det reelt sett dreier seg om forskning. Et bedre alternativ er å vurdere både formål og hvilken type maskinlæring som brukes.

Forenkling av arbeidsprosesser

I dag brukes maskinlæring for det meste for å forenkle eksisterende arbeidsprosesser, som å hjelpe leger å oppdage og bestemme funn i medisinske bilder. Her er det sjelden nødvendig med forskningsetisk vurdering.

Nye hypoteser

En maskinlæringsprosess som ser etter nye mønstre (også kalt ikke-veiledet læring) kan skaffe tilveie ny kunnskap, som å oppdage [fem undergrupper av diabetes type 2](#). Slike prosesser er hypotesegenerende, til forskjell fra tradisjonell, hypotesedrevet helseforskning. Det tilsier at de ikke passer inn i dagens system for forskningsetisk vurdering.

Det kan likevel reises forskningsetiske spørsmål knyttet til pasienter som befinner seg i nye risikogrupper. Håndteringen av dem bør gjøres til gjenstand for en forskningsetisk vurdering.

Uplanlagte funn

En modell som er tatt i bruk i vanlig drift, kan lære fortløpende av erfaringer fra hver ny pasient, slik at både pasienten selv og de neste pasientene får oppdatert og bedre tilpasset behandling. Slike dynamiske modeller kan derfor komme frem til ny kunnskap, selv om det ikke var formålet.

Helseinstitusjonen bør derfor gjennomgå modeller som lærer dynamisk i hele livsløpet og vurdere om de har

kommet frem til ny kunnskap. Hvis det er tilfelle bør det etableres et nytt helseforskningsprosjekt, for å validere funnene og sikre at forskningsetikken blir ivaretatt.

Dataforvaltning for fellesskapet

Regjeringens hovedmål i [Helsenæringsmeldingen](#) er å bidra til økt konkurransekraft i den norske helsenæringen og samtidig bidra til en mer bærekraftig helse- og omsorgstjeneste. Dersom Norge skal tiltrekke seg de beste forsknings- og utviklingsmiljøene, vil en utvidet tilgang til helsedata fra norske innbyggere og pasienter bli nødvendig.

Mange av de store globale selskapene, som Google, Apple og Amazon satser nå på helsemarkedet. Hvis data om norske pasienter og helseforhold skal deles med tredjeparter, bør fellesskapet få en merverdi i form av bedre offentlige helsetjenester, ny næringsutvikling, arbeidsplasser eller skatteinntekter.

Ved maskinlæring er det imidlertid ikke alltid forutsigbart hvordan dataene skaper verdi og for hvem. Maskinene kan lære på egenhånd og komme frem til nye sammenhenger som ikke var kjent tidligere, og læringen kan overføres fra ett system til et annet.

Dette gjør det komplisert å regulere partenes ansvar og rettigheter. Hvor mye skal for eksempel det offentlige betale for tjenester hvor det er tilgangen på offentlige helsedata som har muliggjort utviklingen?

I tillegg er innbyggerne part i saken, siden dataene er om og fra dem. Dette innebærer ekstra krav til forsvarlig bruk og innsyn dersom tilliten til datadelingen skal opprettholdes.

Datafond

Det er derfor nødvendig at myndighetene etablerer juridiske og institusjonelle rammer som gjør det mulig for både innbyggere og helseinstitusjoner å utveksle data trygt og som sikrer at rettigheter, verdier og ansvar fordeles rettferdig og likeverdig, også i fremtiden.

Storbritannia utvikler [datafond](#) (*data trust*) som en måte å øke tilgang til data og samtidig beholde tillit. Målet er å skape mest mulig samfunnsmessig verdi uten å gå på bekostning av etablerte vedtekter og etiske retningslinjer. Et slikt datafond kan inngå avtaler med forskere, utviklere og helseinstitusjoner på vegne av innbyggerne.

ANBEFALINGER

Tilgang til medisinske bilder, pasientjournaler og pasientenes egenmålte data vil være en forutsetning for å utvikle og bruke maskinlæring i den norske helsetjenesten fremover.

Helsemyndighetene bør derfor:

- » endre regelverket slik at sykehus og andre helseinstitusjoner er i stand til å bruke helsedata til utvikling og bruk av maskinlæring, samtidig som de ivaretar hensyn til personvern og forskningsetikk.
- » etablere en sentral reservasjons- og samtykke-tjeneste, slik at alle innbyggere blir informert og involvert i hvordan data om dem brukes.
- » legge til rette for at maskinlæring i hovedsak gjøres på helsedata som er lagret lokalt. Dette vil være smidigere og mer brukernært enn å benytte sentrale registre, og dessuten styrke datasikkerhet og personvern.
- » utrede datafond etter modell av britiske *data trust*, for å sikre at samfunnet får verdier tilbake fra norske helsedata, og dessuten sikre integriteten til innbyggerne.

Forfattere: Tore Tennøe / Hilde Lovett

Publisert: desember 2019

Kontakt: post@teknologiradet.no / www.teknologiradet.no

Ekspertgruppe: Erik Fosse (Oslo Universitetssykehus), Siri Hatlen (Teknologirådet), Steinar Madsen (Statens legemiddelverk), Damoun Nassehi (fastlege og Teknologirådet), Hans Olav Melberg (Universitetet i Oslo) og Michael Riegler (Simula og OsloMet).

Teknologirådet gir råd til Stortinget og regjeringen om ny teknologi