

Tungregning og Big data

Forandrer me disinen

Med tungdata og enorme beregningsmuligheter er biologi og medisin i ferd med å endres.

Tekst: Mats Schjølberg Ulshagen

Framskrittene innen medisin kommer stadig oftere og svært ofte, om ikke alltid, er de knyttet til ny teknologi. Der man tidligere brukte uker og måneder, og tverrfaglige samarbeidsprosjekter på tvers av landegrensener for å analysere en enkelt celle, kan man i dag lese av DNA-koden til en person på et laboratorium på under et døgn. Forskere i dag er i ferd med å kartlegge mange av hjernens funksjoner og hemmeligheter, og vi vet i dag hvordan et enormt antall celler arbeider sammen og kan sammenligne det med hundrevis av andre mennesker og deres celler. Disse funnene genererer enormt mye data, og behovet for datakapasitet og tungregning blir bare større og større.

Samarbeid

Marianne Fyhn er førsteamanuensis på Institutt for biovitenskap ved Universitetet i Oslo og forsker på hjernen. Hun var med på å finne stedsansen i hjernen sammen med nobelprisvinnerne May-Britt og Edvard Moser, og jobber nå med å forske på hvordan hjernen er i stand til å lagre nye minner samtidig som den beholder gamle minner.

– Hjernen må være statisk for at minner skal kunne lagres i lang tid; samtidig må den være så fleksibel at vi er i stand til å lære nye ting, forteller hun.

Hjernen består av hundre milliarder celler, der hver av dem har tusenvis av kontakter.

– Samme hvor mange elektroder vi setter inn i hjernen for å måle hjernesignalene, får vi bare et glimt av hva som skjer. Skal vi forstå samspillet mellom hjernecellene trenger vi matematiske modeller til å fylle inn alt vi ikke greier å måle, forklarer hun.

Slike eksperimenter er gjerne dyre og krevende, og Fyhn forteller at matematiske modeller kan være med på å sørge for at eksperimentene man foretar er mer målrettede.

– Nå har vi fått instrumenter som gjør at vi kan se på mange hjerneceller som virker i en prosess samtidig, og ved hjelp av simuleringer



laget på basis av matematiske modeller kan vi se på det i en stor tidsopløsning, forklarer hun.

– Dette skaper et behov for å samarbeide med andre grener av vitenskapen

– Vi samarbeider med andre miljøer, for eksempel på informatikk og fysikk, for å finne sammenhenger i dataene, sier hun.

Kraftigere modeller

Marianne Fyhn har også fått med seg Anders Malthe-Sørensen, professor på Fysisk institutt ved Universitetet i Oslo, på laget. Han blir regnet som ledende på beregninger i geofysikk i Norge.

– Det vi gjør er å lage matematiske modeller av hvordan ulike ting fungerer, ut fra den informasjonen vi har tilgjengelig. Vi kombinerer moderne eksperimenter med kraftigere beregningsmetoder, og så tilpasser vi modellen til resultatet underveis, forklarer han.

Mer nøyaktige data og bedre kunnskap om hvordan ting fungerer, gjør at disse modellene blir bedre.

– Ta trening som eksempel. Tidligere kunne jeg analysere en treningsøkt, eller kanskje noen

« Samme hvor mange elektroder vi setter inn i hjernen for å måle hjerne-signalene, får vi bare et glimt. Skal vi forstå samspillet mellom dem trenger vi matematiske modeller til å fylle inn punktene vi ikke greier å måle.

få, og ut fra det forsøke å finne ut hvordan kroppen min reagerer på ulike former for trening. I dag kan jeg sammenligne alle målinger fra alle treninger jeg har gjort de siste par årene, i tillegg til treningene andre har gjort, i lignende eller forskjellige situasjoner. Da har jeg etter hvert mye mer informasjon å basere vurderingene på. Dersom jeg i tillegg måler flere ting samtidig, som blodstrøm, hjerte og svettegrad, og så setter alt dette sammen, vil det gi en langt kraftigere matematisk modell for hvordan trening fungerer og hvilken effekt det har på min kropp, forklarer han.

Slik er det også med hjerneforskningen. – Hvilke grunnleggende prosesser som skjer i hjernen er et stort og utfordrende problem, kanskje er det uløst. Men etter hvert som vi får mer data får vi kraftigere modeller, som gjør at vi vil få et mer nøyaktig bilde av det, forteller han.

Tunge beregninger

Også innen forskning på kreft har datamaskiner, tungregning og big data en stadig viktigere funksjon.

– Det er to helt forskjellige verdener. Vi har fått muligheter vi ikke kunne forutse for bare få år siden. Og det er ikke bare big data, men også utvikling i lab-teknologi som har gjort at vi kan gjøre analyser på bare få timer som vi tidligere trengte år til. Dette setter oss i fare for å drukne i alle data, og det er her supercomputere kommer inn i bildet, forteller Rolf Skotheim, forsker ved Avdeling for molekylærmedisin, Radiumhospitalet og førsteamanuensis ved Institutt for informatikk ved Universitetet i Oslo. Forsk-

ningsgruppen hans leter etter den genetiske forklaringen på tarmkreft og prostatakreft, og de sammenligner genmaterialet i svulster med genmaterialet i friske celler.

– Helt konkret, så forsøker vi å finne varianter i arvestoffet RNA og DNA, som vi utelukkende finner i kreftceller. Altså finne feil i arvestoffet til kreftcellene, forteller han.

RNA er enkelt forklart en kopi av DNA, som brukes når kroppen produserer proteiner. I kreftceller skjer det noe galt med RNA-kopieringen, som gjør at det dannes feil mengde av visse typer proteiner eller at egenskapene til proteinene blir feil. Forskerne undersøker friskt og sykt vev hos 550 pasienter med prostatakreft, på jakt etter avvik i RNA. En av laboratorieanalysene de bruker gir imidlertid omtrent 100 millioner svar på sammensetninger av ulike RNA-molekyler i kreftprøvene.

– Slike enorme mengder rådata krever avanserte analyser og beregninger på kraftige datamaskiner, forklarer Skotheim.

Flere blir friskere

En rekke internasjonale miljøer arbeider med lignende forskning, og deling av data mellom miljøene er viktig for å gjøre framskritt innen forskningen.

– På verdensbasis er det flere tusen forskere som jobber med kreftgenomikk, noen av dem gigantiske internasjonale prosjekter, og det er utstrakt datadeling, forklarer han.

Skotheim forteller at de har hatt konkrete resultater av forskningen.

– Det har allerede gitt en del ny og viktig kunnskap. Vi har funnet flere gener som er



BILDETEKST:
Rolf Skotheim, forsker ved Avdeling for molekylær-onkologi, Radiumhospitalet og førsteamanuensis ved Institutt for informatikk ved Universitetet i Oslo.

viktige når friske celler blir kreftceller, noe som igjen vil kunne gi forbedringer i oppdagelse av kreft, og nye medisiner, forklarer han.

Han er ikke i tvil om at avansert matematikk og kraftige datamaskiner har eskalert kreftforskningen.

– Det har den. Det løses stadig vekk større og mindre kreftgater, og det har gått raskere etter at genomteknologien har blitt tilgjengelig. Men det er dessverre ikke slik at vi finner svaret

« Slike enorme mengder rådata krever avanserte analyser og beregninger på kraftige datamaskiner.

på alt, selv om stadig flere blir friske og det er større tempo på framskrittene, forteller han.

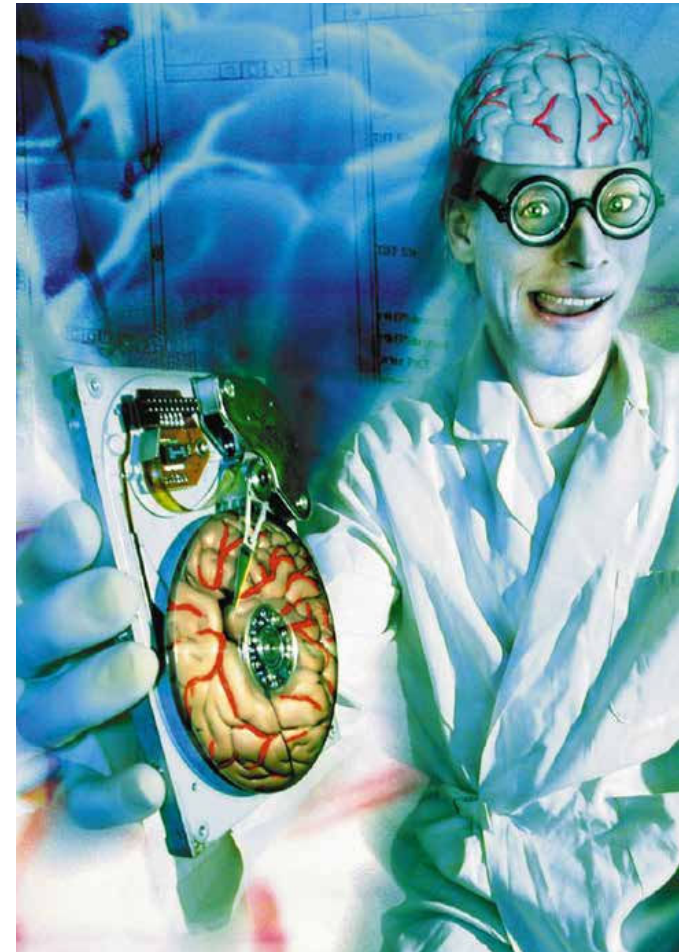
Forskjellig kompetanse

Anders Malthe-Sørensen trekker også fram datadeling som en viktig faktor innenfor hjerneforskningen.

– Vitenskapen har en tradisjon for samarbeid. Når det gjelder hjerneforskning er jeg imponert over hvor mange ulike initiativer som deler alt. The Allen Institute for Brain Science har for eksempel som prinsipp at de deler alt, og legger det åpent ut, forteller han.

Han har også tro på at et tverrfaglig samarbeid mellom ulike vitenskapsgrener er veien å gå.

– Det er en del utfordringer i forbindelse med å samarbeide slik, som hvordan man jobber med noen som er opplært til å tenke annerledes enn en selv. Men det er historisk svært mange oppdagelser som har kommet i grensenivået av forskjellige vitenskaper, og jeg tror ikke at man innen forskning kan regne med gjennombrudd fra en enkelt person. Skal man finne løsninger trenger man team av personer med forskjellig kompetanse. Det tror jeg er fremtidens forskning, avslutter han. □



Verdens beste datamaskin

Ikke bare gjemmer hjernen på en rekke hemmeligheter, den er også til stor inspirasjon for dem som jobber med å lage bedre datamaskiner.

Hjernen vår er utrolig effektiv til å prosessere og rekonstruere enorme mengder informasjon, og det forskes på hvordan vi kan lære av metodene den bruker til å skape bedre datamaskiner.

– Det er en retning innen forskning på datamaskiner som kalles deep learning, der man forsøker å lage mer avanserte systemer basert på hvordan hjernen virker. Etter hvert som vi får mer informasjon om hvordan hjernen virker, vil vi kunne designe maskiner som fungerer mer lik hjernestrukturer, forteller Anders Malthe-Sørensen.

– Hjernen er utrolig effektiv til å prosessere; kan man lære noe av det, kan man kanskje forbedre kapasiteten på harddisker. Det hjernen er i stand til, er å trene nervecellenettverket til å behandle og rekonstruere det som kommer inn. Ta synssystemet for eksempel, der er det enorme mengder med data som kommer inn. Bare du åpner øynene, mottar hjernen informasjon tilsvarende mer enn 10 bøker. Denne informasjonen nedskaleres trinnvist. Kan vi registrere de ulike trinnene, og putte dem inn i matematiske modeller, kan vi kanskje få datamaskiner til å lære seg å gjøre det samme, tror Marianne Fyhn.



HJERNEFORSKER:
Marianne Fyhn, førsteamanuensis på Institutt for biovitenskap ved Universitetet i Oslo



BEREGNINGER:
Anders Malthe-Sørensen, professor på Fysisk institutt ved Universitetet i Oslo

Endrer faget

– Biologistudenter trenger å lære mer matematikk, tror Marianne Fyhn ved Institutt for biovitenskap ved Universitetet i Oslo.

– Biologistudenter er ofte ikke kjent for å være de mest matematikkinteresserte, men det er et faktum at biologien og medisinen er i ferd med å bli mer og mer regnetung, og hele utdanningen trengs å endres, forteller hun.

Ettersom samfunnet blir stadig mer avhengig av at man kan bruke dataverktøy, som blir stadig mer avansert, må biologiutdannelsen endres.

– I dag gjenspeiler ikke utdannelsen utviklingen, den er lik som den var for 20–30 år siden. Det er ikke optimalt. Biologistudenter må lære seg å bruke matematiske modeller. Ellers gjør vi de som studerer en bjørnetjeneste, mener Marianne Fyhn.

Anders Malthe-Sørensen, professor på Fysisk institutt ved Universitetet i Oslo er enig.

– Dette gjelder ikke bare i medisinen. Framtiden kommer til å kreve at man kan tenke med tall, og på tall, og kunne bruke algoritmer til å beskrive prosesser. Man trenger ikke kalle det matematikk, men abstraksjonsnivå er blitt viktig i mange utdannelser. Derfor tror jeg vi må bygge en utdannelse som gir de studerende muligheten til å kunne hanskens med morgendagens utfordringer, der man må kunne bruke PC-en som et verktøy til å styrke tenkningen, ikke bare som en wordprocessor, sier han.