

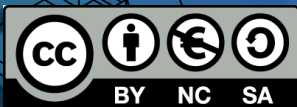
# KOSMOS

## MEDICINSK FYSIK

SVENSKA FYSIKERSAMFUNDETS ÅRSBOK 2022

## ARTIFICIELL INTELLIGENS INOM VÅRD OCH SJUKHUSFYSIK

© FREDRIK KALHOLM OCH WILLE HÄGER



Artikeln publiceras under Creative Commons-licensen CC BY-NC-SA 4.0  
För bildmaterial med källhänvisning  
gäller samma upphovsrättsliga  
regler som för källan.



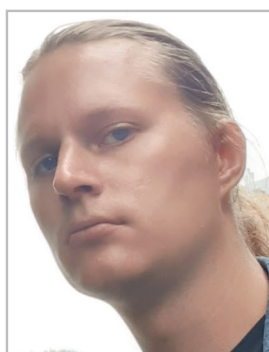


Artificiell intelligens gör sitt intåg i samhället, utan att vi alltid nödvändigtvis vet vad det innebär. I den här artikeln ger Fredrik Kalholm och Wille Häger exempel på hur tekniken kan komma till nytta också inom medicin, till exempel vid förbättring av medicinsk utbildning för såväl diagnostiska som terapeutiska tillämpningar.



**Fredrik Kalholm**

är doktorand inom medicinsk strålningsfysik vid Stockholms Universitet och Karolinska Institutet och forskar på hur strålskador från olika typer av partiklar och energier påverkar dödligheten för celler med simulering, modellering och AI.



**Wille Häger**

är doktorand inom medicinsk strålningsfysik vid Stockholms Universitet och Karolinska Institutet. Hans forskning inriktar sig på simulering av tumörers spridning i hjärnan med hjälp av fysikaliska modeller och AI.

*Bildtagning av 3D-objekt görs ett plan i taget, där data digitaliseras och transformeras matematiskt till verklighetstroga bilder. AI har potential att såväl förbättra bildkvaliteten som att effektivisera analysen av bilderna.*

# Artificiell intelligens inom vård och sjukhusfysik

Artificiell intelligens (AI) har revolutionerat många industrier och utför uppgifter som traditionellt krävt en människa för att lösa, såsom självkörande bilar, automatisk textöversättning, språkigenkänning och smarta assistenter såsom Apple Siri och Amazon Alexa. I takt med att även sjukvård och medicinsk forskning blir alltmer digitaliserad utgör AI ett viktigt verktyg också inom vården. Men vad är egentligen AI? Hur hamnade vi där vi är idag, där begreppet hörs nästan varje dag? Och mer specifikt, hur används AI inom medicinsk fysik?

## Historik

Artificiell intelligens inkluderar alla tekniker som möjliggör för datorer att efterlikna mänskligt beteende. Även om det är svårt att precisera exakt när fröna till AI såddes, kan det åtminstone härledas tillbaka till 1942 då den amerikanska författaren Isaac Asimov publicerade novellen *Runaround*. Handlingen innefattar en robot utvecklad av ingenjörerna Powell och Donovan och kretsar kring robotikens lagar: (i) en robot får aldrig skada en människa eller, genom att inte ingripa, tillåta att en människa kommer till skada; (ii) en robot måste lyda order från en människa, förutom om sådana order kommer i konflikt med första lagen; (iii) en robot måste skydda sin egen existens, såvida detta inte kommer i konflikt med första eller andra lagen. Detta, och andra av Asimovs verk, inspirerade många framtida forskare inom datavetenskap.

Maskininlärning (ML) innefattar AI-tekniker som har förmågan att lära sig från data utan explicit programmering och är därmed en undergrupp till AI. Forskaren Arthur Samuels kunde redan 1959 använda spelet dam (eng: *checkers*) för att visa att maskininlärning är möjligt: han utvecklade ett datorprogram som

kunde lära sig av olika positioner på brädet, istället för att helt förprogrammeras av en människa. I introduktionen till samma artikel citeras även Frank Rosenblatts då nypublicerade artikel om perceptronen: en algoritm designad att efterlikna neuroners funktion i hjärnan. Dessa algoritmer var för beräkningskrävande för att kunna användas i större omfattning med dåtidens teknik, men har blivit en grundläggande beståndsdel i modern ML.

De förväntade framtidsutsikterna för AI har varierat mycket genom åren, med flera perioder där såväl mängden anslag som antalet forskare minskat signifikant till följd av att höga förväntningar inte införlivats, s.k. AI-vintrar. Den första vintern kom i slutet av 1970-talet, då perceptronen dessförinnan varit upphaussad efter att den visat sig kunna användas för att modellera enklare system. Många forskare påpekade då begränsningarna med dåtidens algoritmer för att kunna använda perceptroner till mer komplexa uppgifter, samt framhävde hur små framsteg som faktiskt gjorts i förhållande till de stora förväntningarna. Den första vintern avslutades då Rumelhart, Hinton och Williams introducerade brantaste lutningsmetoden (eng: *gradient descent*) samt bakåtpropagering — metoder som bägge används för att optimera olika vikter använda i nätverken (mer om detta nedan).

Nästa AI-vinter började under tidigt 90-tal, främst orsakad av att dåtidens datorer inte var nog kraftfulla att hantera de komplexa nätverk som krävdes för att lösa de mer komplicerade uppgifterna som AI hade som mål. Under senare delen av 90-talet gjordes framsteg då en superdator tillverkad av IBM, *Deep Blue*, med hjälp av djupinlärning lyckades besegra Garry Kasparov, den då regerande världsmästaren i schack. Den andra vintern avslutades och hittills har en tredje vinter uteblivit. Senare framsteg i datorhårdvara, främst genom grafikprocessorer tidigare ämnade för tv-spel och videoredigering, samt en stor tillgång på träningsdata till maskininlärning, har nu gjort fältet mer aktivt än någonsin.

Varför har uppståndelsen kring AI, både nu och historiskt, varit så stor? Potentialen är inom många fält enorm, då AI i princip kan få mer erfarenhet och kunskap än även vad den mest intelligenta och erfarna människan har inom ett givet fält, genom att analysera och behandla mer data än samma människa kan se under flera livstider. Lägg därtill att en AI väl färdig och i drift aldrig (i dagsläget) kommer kräva ut någon lön från sin arbetsgivare, vilket inom många fält kan bli en stor kostnadsbesparing, speciellt då



många jobb som potentiellt kan ersättas är just höglönlade kunskapsbaserade yrken.

## Grundläggande principer

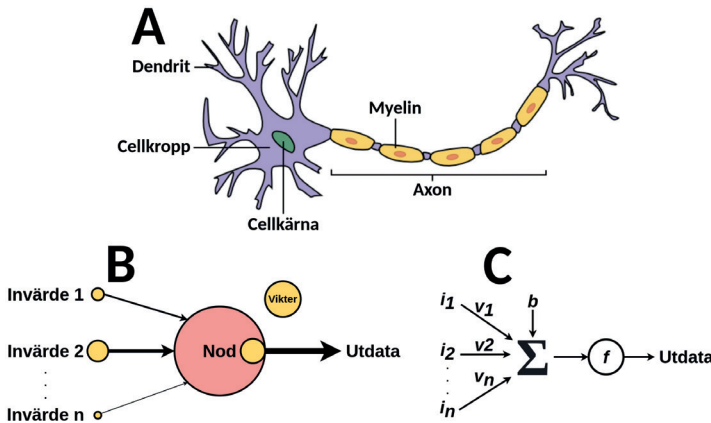
En AI utgörs av följande två komponenter:

- **Uppbyggnad av ett neuralt nätverk:** perceptronen (se Figur 1) tar in ett antal kontinuerliga invärdesparametrar ( $i_1, i_2, \dots, i_n$ ), med olika inkluderande vikter ( $v_1, v_2, \dots, v_n$ ), för att därefter summera dessa. Det slutliga värdet beräknas sedan genom att en icke-linjär aktiveringsfunktion ( $f$ , i Figur 1C) appliceras på denna summa. Genom att sedan sammankoppla flera perceptroner till varandra och därmed efterlikna en hjärnas struktur bildas ett s.k. neuralt nätverk (NN, se Figur 2). Detta kan ske i olika antal steg innan ett utvärde beräknas, dvs. en perceptron skickar in sitt utvärde som invärde till nästa perceptron osv. Varje sådant steg kallas ett lager (och behöver inte heller nödvändigtvis bestå av enbart perceptroner, andra algoritmer kan också förekomma). Lager som inte är direkt sammankopplade till invärden eller utvärden kallas gömda.

Ett NN är en metod för maskininläring, där en dator lär sig utföra en uppgift genom att analysera data. För s.k. övervakad inläring (eng: *supervised learning*) har dessa data ett rätt svar som algoritmen kan optimeras efter. Till exempel kan ett NN användas för att analysera bilder av olika djur (se Figur 2). I detta fall tränas det först av bilder där rätt svar också anges, för att därefter försöka avgöra vilket djur som finns på bilder där svaret är okänt. De NN som består av mer än ett gömt lager perceptroner kallas djupa, vilket också är vad djupinläring (eng: *deep learning*) syftar på.

- **Inläring:** för att uppdatera vikterna i nätverket, och därmed möjliggöra lärningsdelen i maskininläring används ofta algoritmer baserade på brantaste lutningsmetoden. Detta är en iterativ optimeringsalgoritm för att hitta ett lokalt minimum till en funktion. Algoritmen liknar Newtons metod för optimering, men skiljer sig från denna dels genom att funktionernas andraderivator inte behöver uppskattas (vilket reducerar beräkningstiden), dels genom att en hyperparameter  $\alpha$  som styr hur stora steg som tas i varje iteration

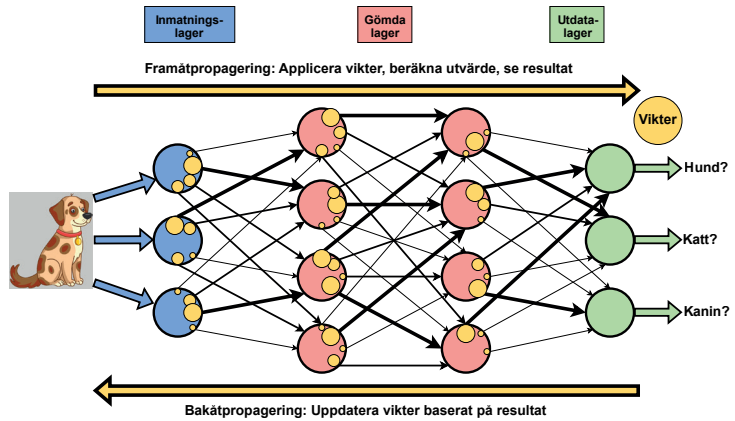
behöver anges. Detta kombineras även med bakåtpropagering: en metod för att också kunna uppdatera vikter i djupa neurala nätverk. Det görs genom att börja med vikterna närmast utvärdet, där den direkta återkopplingen ges genom den strafffunktion (eng: *cost function*) som jämför det förutspådda värdet av nätverket med det faktiska svaret från träningsdatan, för att sedan stegvis arbeta sig bakåt till de tidigare lagren. Målet är hela tiden att alla vikter i hela nätverket ska sättas så att värdet av den slutliga strafffunktionen minimeras, dvs. att de förutspådda värdena från nätverket i så stor utsträckning som möjligt efterliknar de faktiska svaren.



Figur 1: A: Schematisk bild av ett neuron (hjärncell), vilket perceptronen försöker efterlikna: dendriterna i bild A motsvarar invärdena i B och C, cellkroppen i A motsvarar noden i B och summering samt aktiveringsfunktionen i C, medan axonen i A motsvarar utvärdena i B och C. B: Perceptron, med illustrerade invärden med olika vikter som summeras, och producerar ett utvärde genom en aktiveringsfunktion. C: Perceptronen med bokstäver som beskriver begreppen:  $i_1, i_2, \dots, i_n$  representerar invärdena;  $v_1, v_2, \dots, v_n$  representerar vikterna. Dessa summeras tillsammans med en viktningsterm ( $b$ ), och den slutliga summan matas in i en aktiveringsfunktion som producerar ett utvärde.

## Faltningarnätverk

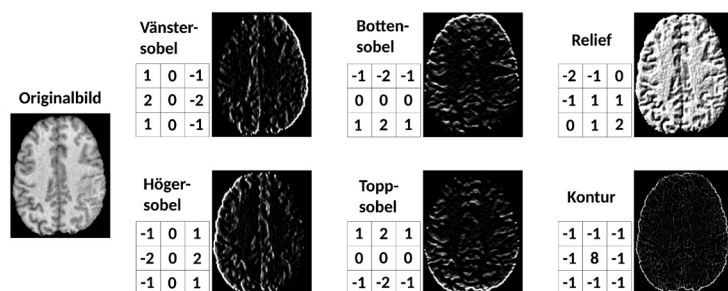
Majoriteten av AI-tillämpningar inom sjukhusfysik är relaterade till bildbehandling och bildanalys. Vanligaste typen av AI-metod



Figur 2: Schematisk illustration av ett fullt sammankopplat neuralt nätverk bestående av ett inmatningslager (blått), två gömda lager (rött), och ett utdatalager (grönt). Varje blå och röd nod motsvarar en perceptron.

för detta är användning av ett s.k. faltningsnätverk (eng: *convolutional neural network*). Exakt arkitektur av dessa nätverk skiljer sig åt beroende på tillämpningsområde, men delar ändå många gemensamma drag, såsom att tidiga lager tenderar att utföra operationer som komprimerar/reducerar mängden information i datan, för att efterföljande operationer ska bli mindre beräkningsintensiva. Dessa operationer innefattar:

- **Faltning:** olika filter används som framhäver eller reducerar vissa drag i bilden genom att denna faltas med originalbilden, se Figur 3. För varje pixel bestäms ett nytt värde genom att en viktad summa av pixeln tillsammans med omkringliggande pixlar beräknas, där vikterna i summan bestäms av en s.k. kernel. Dessa filter appliceras för att framhäva och förstärka olika egenskaper i bilden och på så vis reducera den till en form som är lättare för nästkommande lager att behandla. Ett filter består oftast av flera kernels, vilket resulterar i fler pixelkanaler (på samma sätt som olika färgkanaler).
- **Pooling:** detta har som mål att reducera storleken på bilderna för att underlätta beräkningsoperationer i nästkommande lager. Denna operation samplar datan på olika sätt, t.ex. genom att det nya pixelvärdet som ska representera en grupp av fyra gamla pixlar blir maxvärdet eller medelvärdet av dessa.



Figur 3: Illustration av faltningsoperationer med olika tillämpade kernels, där varje pixels nya värde bestäms genom ett viktat medelvärde av pixelns eget värde (mitten av kerneln) och närliggande pixlar (övriga delar av kerneln). Sobeloperatorn är vanligt förekommande för att detektera kanter, dvs. snabba förändringar av intensitet, i olika riktningar.

Dessa steg resulterar tillsammans i bilder med låg upplösning, fast med många kanaler per pixel. När slutmålet för nätverket är att känna igen ett visst objekt i bilden avslutas det ofta med fulla sammankopplade lager<sup>1</sup>. När målet istället är att skapa en ny bild kan dessa steg istället köras omvänt, dvs. faltningsoperationer som reducerar antalet kanaler i bilden samt uppskalning av upplösningen, vilket t.ex. görs i U-Net som utvecklades 2015 för biomedicinsk bildsegmentering, och som har blivit en mycket populär metod för liknande tillämpningar.

## Bildtagning

Då cancertumören sitter inuti kroppen, och inte kan ses med blotta ögat, krävs i princip alltid någon form av bildtagning för att ställa en korrekt diagnos och för att bestämma en optimal behandlingsplan. Beroende på var tumören sitter, och exakt vilken information man söker, finns olika metoder att tillgå:

- CT (datortomografi, eng. *computed tomography*): Bilder tas genom att ett röntgenrör och en detektor roterar runt patienten i olika vinklar, och olika pixlar i detektorn får olika stark signal beroende på hur mycket strålning som absorberas. Dessa 2D-bilder konverteras sedan till en gemensam 3D-bild. Här kan AI-metoder användas för att reducera bruset i rekonstruerade bilder.

<sup>1</sup>Se t.ex. [https://adamharley.com/nn\\_vis/cnn/2d.html](https://adamharley.com/nn_vis/cnn/2d.html), där ett faltningsnätverk används för att känna igen handskrivna siffror och varje steg på vägen är visualiserat.



- PET/SPECT (positronemissionstomografi, *single-photon emission computed tomography*): Dessa tekniker används när olika fysiologiska egenskaper hos tumören ska undersökas och utförs genom att radioaktiva atomer kemiskt fästs till molekyler som väntas upptas av tumören i högre grad än av frisk vävnad. Substansen intas av patienten, som efter en tid placeras i en PET/SPECT-kamera. Genom att detektera strålningen med den senare kan man bestämma var de radioaktiva atomerna inuti kroppen befinner sig. Signalen uppkommer alltså inifrån kroppen, snarare än genom extern strålning. Dessa typer av bilder är ofta långt överlägsna CT/MR för att detektera tidiga förändringar i celler och hitta mikroskopiska tumörer, men har betydligt sämre upplösning och går inte att använda för en generell överblick över hela anatomin (se även artikel av Nilsson & Ardenfors i denna upplaga av Kosmos).
- MRT (Magnetresonanstomografi): Teknik baserad på starka magneter och radiovågor. Maskinen består av en torusformad supraleddande elektromagnet som får de flesta av kroppens väteatomers magnetiska moment att linjeras upp med fältet. Därefter skickas pulser av radiovågor till olika delar av kroppen vilket gör att det magnetiska momentet en stund roterar runt fältets axel innan de efter en tid åter linjeras. Signalen i MR är dels styrkan av magnetiska oscillationer som uppkommer när momentet roterar runt magnetfältet efter radiopulsen (T2), och dels tiden det tar momentet att åter riktas upp (T1). Tekniken mäter alltså koncentrationen av väteatomer och är ofta överlägsen CT för kontrast av olika mjukvävnadstyper, som ofta absorberar ungefär samma mängd röntgen, men kan ha olika koncentration av väte (se även artikel av Ekman & Garpenbring i denna upplaga av Kosmos).

### Medicinska tillämpningar – Diagnostik

Radiomics är en av de enklaste metoderna att analysera bilddata. Metoden går ut på att utnyttja mängder av egenskaper (eng. *features*) hos bilder. Som exempel kan det inom onkologin vara egenskaper relaterade till tumörstorlek och form, men av vikt är

även egenskaper rörande bildens gråskaliga histogram, såsom medelvärde, kvadratisk medelvärde, och standardavvikelsen av alla pixelvärden. Endast fantasin sätter gränser på vad man skulle kunna använda, och i praktiken kan antalet egenskaper vara tusentals.

Av dessa egenskaper är 99% av låg vikt och kan ignoreras då modellen tillämpas, medan resterande 1% drar den tunga lasten. Egenskapernas vikter bestäms när algoritmen tränas på ett dataset bestående av ett träningsset och ett testset. Målet med radiomics är inte bara att förhoppningsvis få bättre diagnostisering, utan också att göra bildanalys mer vetenskaplig och distanserad från människans subjektivitet. Radiomics är en av de ytligare metoderna inom maskininlärning och dess användbarhet har varit debatterad. Storskaliga kliniska studier saknas och metoden har inte sett någon praktisk tillämpning på sjukhus. Metodiken bakom radiomics har beskrivits som en se-vad-som-fastnar-strategi, och intresset har klingat av en del till förmån för djupinlärningsmetoder såsom falttningsnätverk.

Integrering av AI har även potential att minska skillnaderna mellan utvecklings- och industriländer när det gäller tillgång till sjukvård. MoMic-projektet är just ett sådant projekt som drivs vid bl.a. Uppsala Universitet och Karolinska Institutet. Projektets mål är att utveckla mobila mikroskop som med hjälp av AI kan hjälpa att diagnostisera t.ex. infektionssjukdomar, även där det är brist på lokal medicinsk expertis. Ett ständigt hinder med tillämpning av AI i utvecklingsländer är bristen på träningsdata till algoritmerna. De begränsade data som finns att tillgå är inte alltid av tillräcklig kvalitet.

MoMic-projektet löser detta med internetbaserad träning: ett mikroskop uppkopplat till internet kan tränas på distans. Algoritmer, som ofta är beräkningsmässigt krävande, kan även utföras av datorer på andra platser i världen, vilket också gör mikroskopen billigare. Poängen med tekniken är att kringgå de mest resurskrävande momenten inom diagnostiskt arbetsflöde. Genom att minska behovet av högutbildade läkare öppnas möjligheter även för dem med sämre förutsättningar. Samtidigt krävs en medvetenhet om det vanskliga med att låta människor från utvecklingsländer agera försökspersoner, även om valet står mellan (experimentell) AI-baserad vård eller ingen vård alls.

AI har även använts inom andra delar av diagnostiskt arbetsflöde, t.ex. genom att analysera bilder och automatiskt upptäcka

anomalier som kan vara grunden till olika sjukdomstillstånd. En stor del av AI-relaterad forskning inom medicinsk fysik har också tillägnats behandlingsaspekten, där strålbehandling kanske varit det mest framträdande tillämpningsområdet. Detta breda fält innefattar allt från bildanalys för lokalisering av tumörer, till optimering av behandlingsplaner där hänsyn måste tas till patientens individuella anatomi och hur strålningen växelverkar med denna.

### Syntetisk CT

De två dominerande bildtagningstekniker som används på sjukhus är tveklöst röntgen- (t.ex. CT) och MR-baserade. Den stora fördelen med CT framför MR är att bildtagningen kan ske på sekunder istället för minuter. Utrustningen är också billigare och enklare att använda. Nackdelarna är att bildkvaliteten är sämre, speciellt för mjukvävnad, och att patienten oundvikligen exponeras för en liten men ej försumbar stråldos ( $\sim 10$  mSv). Vilken teknik som används varierar från fall till fall, men alla som kommer genomgå strålterapi måste först undersökas i en CT. Detta eftersom CT-bilderna innehåller mycket värdefull information bortom det rent visuella.

Eftersom CT-bilder tas genom att låta röntgenstrålar växelverka och attenueras av patientens kropp, mäts attenueringskoefficienterna i kroppens vävnader indirekt då den undersöks. Varje material har en attenueringskoefficient,  $\mu$  [ $\text{cm}^{-1}$ ] som anger hur starkt fotoner växelverkar med materialet. Värdet beror bl.a. på elektronkonfigurationen hos materialets atomer eller molekyler, samt på fotonens energi. Generellt minskar  $\mu$  exponentiellt då fotonenergin ökar (eftersom genomträngningsförmågan ökar), men sambandet gäller bara i intervallet  $\sim 1$  keV till 10 MeV, där de växelverkande kanalerna domineras av fotoelektrisk effekt och Comptonspridning. För riktigt låga fotonenergier gäller inte sambandet (t.ex. går ju en radiovåg enkelt genom kroppen, medan synligt ljus blockeras). Vid högre energier öppnas nya kanaler för partikelproduktion, som t.ex. parbildning och  $\mu$  tenderar att stagnera eller till och med öka med ökad fotonenergi.

För att dosplaneringen ska bli korrekt för en patient som ska genomgå strålbehandling är det av stor vikt att veta attenueringskoefficienterna. Ifall dosplaneringen baseras på en CT-bild blir koefficienterna automatiskt inbakade i ljusintensiteten hos pixlarna. Röntgenstrålarna i en CT har energier kring 50 keV. Pixlarnas ljus-

intensitet på den bild som fås motsvarar då attenueringskoefficienter för 50 keV. Vid strålterapi används snarare energier kring 10 MeV. Omvandlingen av koefficienter från 50 keV till 10 MeV är inte strikt linjär, men processen har ändå visat sig relativt enkel och effektiv.

Frågan har dock ställts om inte samma information kan utvinnas ur en MR-bild. Att CT-skanna patienter som redan har analyserats i en MR kräver extra resurser och utsätter patienterna för en stråldos. Konceptet syntetisk CT går ut på att man skapar en artificiell CT-bild utifrån MR-bilddata. Med andra ord, genom att direkt ta fram attenueringskoefficienter ur en MR-bild. Löst uttryckt mäter en CT elektrondensitet, medan MR mäter  $^1\text{H}$ -densitet. Fotonerna använda vid strålbehandling växelverkar i princip inte alls med väte, så i sig själva säger MR-bilder egentligen ingenting om hur fotoner kommer bete sig i vävnaden.

Problemet förenklas i praktiken. I grova drag är det endast fyra signifikanta materialtyper man behöver ta hänsyn till vid strålbehandling: mjukvävnad, ben, luft och vatten. Vissa patienter kan också ha tandfyllningar eller andra implantat av metall. Dessa undviks att placeras i strålfältet eftersom de orsakar påtaglig spridning av strålningen. Med kunskap om människans anatomi är det därför en tämligen enkel uppgift att bestämma innehållet på en MR-bild utifrån materialperspektiv. Trots detta (eller kanske på grund av detta) är uppgiften svår att automatisera. MR-bilder kan ha vitt skilda utseenden beroende på bildtagningsprotokoll samt MR-modell och magnetfältstyrka. Dessutom blir aldrig CT- och MR-serier identiska eftersom de måste tas vid separata tillfällen.

Att producera syntetiska CT-bilder har länge försökts, men resultaten har inte ansetts vara tillräckligt pålitliga för att implementeras kliniskt, speciellt med tanke på hur mycket forskning och pengar som läggs på att utveckla moderna och precisa strålkärl och sofistikerade dosplaneringsprogram. Approximerade CT-bilder blir således ett steg i fel riktning. Men här kan AI potentiellt hjälpa. På ett vanligt stort sjukhus tas hundratals CT- och MR-bilder av samma patienter varje dag. Datamängden är enorm, och uppgiften tydlig. Ämnet AI-genererad syntetisk CT exploderade åren 2019–2020, då nästan femtio artiklar kring ämnet publicerades. Konsensus löd: metoden är gångbar. Sedan dess har intresset svalnat, kanske p.g.a. den stagnation som följt i brist på fortsatt forskning inom klinisk tillämpning.



En syntetisk CT använder en bildtagningsteknik och simulerar vad motsvarande bild skulle bli med en annan modul. Forskning har även gjorts på att simulera bildtagning av liknande tekniker, där istället signalstyrkan i bilderna skiljer sig.

### **Bildförbättring av CBCT**

Innan ett behandlingstillfälle inom strålterapi påbörjas används alltid någon metod att positionera patienten rätt, för att säkerställa att stråldosen levereras enligt den ordinerade dosplanen. En vanlig metod är att ta en s.k. *Cone Beam Computed Tomography*-bild (CBCT) av en patient, när denne ligger i ungefärligt behandlingsläge. Detta liknar den CT-bild som tas inför dosplaneringen, men skiljer sig genom att:

1. Strålningen skickas i en konform genom patienten snarare än som en solfjäder, p.g.a. att behandlingsbordet är stillastående under bildtagningen, och strålfältet därmed behöver sträckas ut även i djupled för att täcka hela behandlingsområdet.
2. Den stråldos som ges under CBCT-bildtagningen är betydligt lägre än den CT som tas inför dosplanering, vilket medför försämrade bildkvalitet och brusigare bild.

Efter bildtagningen görs matchningar mellan CBCT- och dosplaneringsbilderna och behandlingsbordet flyttas så att patientpositionen efterliknar den under dosplaneringsbilden. Dessa matchningar görs nästan uteslutande baserat på benstrukturer, då dessa är de enda som säkert kan identifieras på CBCT-bilderna genom sin höga kontrast till mjukvävnad. Subtila skillnader mellan olika typer av mjuk vävnad går i regel ej att urskilja p.g.a. den undermåliga bildkvaliteten. Forskning har påbörjats för att undersöka huruvida AI kan användas för att förbättra bildkvaliteten på CBCT-bilderna. Potentiellt skulle detta kunna medföra att även mjukdelar och segmentering av riskorgan kan matchas mot den mer högupplösta dosplaneringsbilden.

Som data används bilder från patienter som redan genomgått strålbehandling, och har tagit cirka fem CBCT-bilder, och en diagnostisk CT-bild per patient. I flertalet publicerade studier används falttningsnätverk, ofta olika varianter av U-Net, där CBCT-bilderna är invärden och den diagnostiska CT-bilden är målet.

Merparten av de studier som gjorts visar på en stor potential att förbättra CBCT-bilderna med hjälp av AI. I vissa studier

har radiologer, utan att veta vilken bild de tittar på, fått bedöma kvaliteten och resultatet visar på en tydlig förbättring för de AI-processade bilderna. Radiologer har också blivit ombedda att markera tumören och riskorgan i bilderna, och även där ses en liknande förbättring. Dessa tekniker har hittills inte tagits i klinisk drift, men med tanke på de resultat som de AI-behandlade bilderna presterat i de studier som finns är det troligtvis bara en tidsfråga innan detta sker.

Ett kanske mer komplicerat problem än att förbättra enskilda statiska bilder är att uppskatta en cancertumörs placering när den befinner sig på olika platser i kroppen vid olika tidpunkter. Även här har det forskats på AI-tillämpningar.

### 4DCT lunga

En vanlig CT-bild innehåller alltid artefakter när andningsrörelser inkluderas i bildtagningsområdet. Då organ och/eller tumörer befinner sig på olika platser vid olika tider under bildtagningen blir resultatet suddigare bilder där det är svårare att urskilja organens konturer. Genom att mäta patientens andning under bildtagningen och sedan sortera ut bilderna efter andningsfas, kan dessa fenomen reduceras. Detta kan möjliggöra en lägre bestrålning till frisk vävnad då tumörens placering blir mer lättidentifierad och mindre säkerhetsmarginaler behöver tillämpas. Denna typ av CT kallas fyrdimensionell CT (4DCT), där den fjärde dimensionen avser tid.

Deformerbara bildregistreringsmetoder används konventionellt för att modellera dessa rörelser genom s.k. deformerbara vektorfält. Dessa beskriver rörelserna i varje bildserie i förhållande till en statisk referensbild, där prestandan brukar mätas i *Target Registration Error* (TRE) som mäter det genomsnittliga felet i en transformation. Metoderna kan dock vara långsamma, speciellt för större bildserier och/eller stora rörelser, och kan ibland leda till t.ex. orealistiskt deformerade revben och ryggrader

Flera olika forskarlag har undersökt möjligheten att tillämpa AI-baserade deformeringsalgoritmer för att uppnå bättre och snabbare resultat. Nederländska forskare föreslog 2019 användning av ett övervakat faltningsnätverk som liknar U-Net, med ett genomsnittligt TRE på 2,2 mm. Amerikanska forskare har utvecklat ett oövervakat nätverk (som alltså inte behöver ges ett rätt svar för att tränas), bestående av flera komplexa steg med olika falt-

ningsnätverk. Detta resulterade i ett genomsnittligt TRE på 1,6 mm: nästan lika bra som de bästa konventionella metoderna idag med ett TRE på cirka 1,3 mm för de bildserier som användes som jämförelse i studien. Med mer och mer tillgänglig data är det troligt att AI-metoderna om några år kan överträffa de mest sofistikerade konventionella metoderna.

Istället för att förutse en tumörs rörelse genom att kombinera flera statiska CT-bilder, eller försöka omvandla andra bildtagningssystem till CT-bilder, såsom syntetisk CT och CBCT-bildförbättring, kan även olika moduler sammanfogas till ett gemensamt koordinatsystem för att kunna se information från flera bildtyper samtidigt. Detta görs ofta kliniskt för att kombinera styrkan hos olika bildtagningstekniker och därmed skaffa sig en bättre helhetsbild av patientens sjukdomsläge och möjliggöra en bättre behandlingsplan.

### PET/CT-registrering

Olika kliniskt implementerade icke-AI-baserade algoritmer förekommer idag för att transformera tagna PET-bilder till CT-bilders koordinatsystem. Dessa metoder är dock ofta tidskrävande, och ger inte optimalt resultat för alla patienter. En grupp forskare föreslår en ny AI-baserad metod för att utföra denna transformation, som till stor del baseras på strukturen för ett tidigare föreslaget nätverk, DIRNet: ett oövervakat, djupt, faltningsnätverk för bildregistrering, med bl.a. skillnaden att transformationen utförs i 3D istället för 2D. Det föreslagna nätverket består av två delar. Först ett faltningsnätverk som genererar ett lågupplöst fält av 3D-vektorer som visar förflyttningen som ska utföras på PET-bilden. Sedan konverteras detta till en högre upplösning så att varje voxel<sup>2</sup> i bilderna tilldelas en egen förflyttningsvektor.

En strafffunktion som uppskattar likheten mellan de transformerade PET-bilderna med anatomin som ses i CT-bilderna används. Detta är alltså oövervakat, dvs. rätt svar för transformationen är ej känt på förhand. PET/CT bilder från 170 patienter användes. Nittio procent av dessa för att träna nätverket, och tio procent för att testa prestandan. Resultatet av studien är lovande, men det är förmodligen flera år kvar innan en klinisk implementering genomförs.

<sup>2</sup>voxel = tre-dimensionell pixel

## MR/CBCT-registrering

Vid strålbehandling av prostatacancer ges traditionellt en homogen dos till hela prostatan, då cancertypen anses vara multifokal, dvs. kan uppkomma på flera ställen i prostatan samtidigt, vilket medför en stor risk att mikroskopiska cancerceller missas om endast den synliga tumören behandlas. Vissa studier har dock visat att en ökad dos till s.k. DIL:s (eng: *dominant intraprostatic lesions*), dvs. delar av prostatan innehållande en högre koncentration av cancerceller, kan ge ett bättre resultat. Dessa områden kan dock vara svåra att urskilja på CT/CBCT-bilder, då både prostatan och DIL är mjukdelar med ungefär samma absorptionsegenskaper för röntgenstrålning.

Vid bildtagning med MR kan däremot dessa DIL ofta upptäckas, vilket ger möjlighet att segmentera områdena och inkludera dem i behandlingsplanen. Men då ingen MR-utrustning finns tillgänglig vid strålkanonen som utför behandlingen (såvida ingen modern MR-linac används) måste en bildregistrering mellan MR och CT utföras, dvs. en transformation av MR-bilden till CT-bildens koordinatsystem där DIL-områdena markeras. Dessa registreringar är, förutom den dåliga mjukvävnadskontrasten för CT, ofta svåra då olika områden ger olika stark signal för MR/CT samt därför att stora förflyttningar av prostatan kan ske mellan bildtagningstillfällen p.g.a. olika fyllnad i ändtarmen och urinblåsan.

En föreslagen AI-baserad metod att förbättra denna registrering använder sig av två olika faltningsnätverk för att segmentera prostatan i både CBCT- och MR-bilderna. Dessa segmenteringar konverteras sedan till 3D-punktmoln för prostatans hela volym. Efter detta steg används ett ytterligare nätverk, inspirerat av FlowNet3D med liknande struktur som U-Net, som deformerar ytan i MR-punktmolnet för att matcha den i CT-punktmolnet. När ytorna sedan matchats uppskattas motsvarande interna deformation i MR-bilderna, där förflyttningen av DIL-områdena också inkluderas. Därmed kan en CBCT-bild, där endast ytan av prostatan knappt kan urskiljas, användas för att uppskatta den dagliga positionen av DIL-områden, och därmed ge möjlighet att optimera behandlingsplanen för patienten. Som många andra AI-metoder visar denna metod på lovande resultat, men är ännu ej tagen i kliniskt bruk.

Vi vill också nämna forskningsläget för AI i ett av strålbehandlingens fundamentala steg, nämligen klassificering och kon-



turering av tumören och riskorgan. Det spelar ingen roll hur träffsäkert en önskad stråldos kan levereras av en avancerad strålkäna om det som tros vara en del av tumören faktiskt är något annat och vice versa. Uppgiften att urskilja och bestämma detta har traditionellt tillfallit onkologer, men mycket forskning har gjorts som undersöker om även AI-tekniker kan tillämpas.

## Bildanalys

Ett barn kan enkelt urskilja ifall en bild visar t.ex. en hund eller en katt. Denna för oss triviala uppgift kompliceras nämnvärt om vi istället låter en dator avgöra bildens innehåll. Man kan hävda att uppgiften är lättare för människor eftersom vi har erfarenhet av att observera hundar och katter. Grundläggande är att vi även kan extrapolera vår kunskap och dra slutsatser om nya fenomen. Ett barn som aldrig tidigare har sett en svart hund, men har sett många svarta katter, förstår ändå att djuret med den svarta pälsen är en hund med svart päls, och inte en katt. För oss är detta enkelt trots att bägge djur har fyra ben, en svans, två öron, utdragen nos, osv. Det finns stor variation bland t.ex. olika hundraser, men trots detta förstår vi instinktivt vad som karakteriserar en hund. Får vi däremot uppgiften att kvantifiera skillnaden i ord blir det betydligt svårare, och utan en tydligt formulerad definition kan vi inte översätta distinktionen till datorkod. Maskininlärning är ett försök att kringgå den nämnda problematiken och lösa problemet utan att förstå processen. Algoritmernas vikter blir snabbt omöjliga att tolka då man går djupare i nätverkets lager. Vår oförmåga att grundligt definiera visuella drag hos en hund har här sin motsvarighet hos maskinen.

Ett sjukhus är en plats där analys och klassificering av bilddata sker i stor utsträckning. Ända sedan första röntgenbilderna (se Figur 4) togs av Wilhelm Röntgen 1895 har bildanalys varit en del av en läkares vardag. Förmågan att se in i människokroppen utan kirurgiska ingrepp är något vi idag tar för givet, men konceptet var på Röntgens tid revolutionerande, och tekniken, som mer eller mindre har varit oförändrad i ett sekel, används än idag på sjukhus världen över. Under 1970-talet utvecklades dessutom MR. Kanske mer förvånande är att detta även är sant för analysen av bilderna. Än idag är det läkarnas uppgift att betrakta bilderna och genom sin erfarenhet tolka deras innehåll. Denna metod liknar därför mer en konstform än en hård vetenskap. Läkarnas kunskap och förmågor

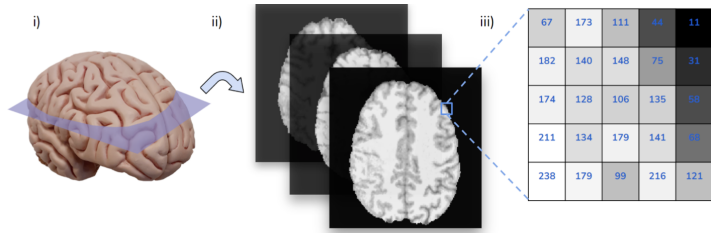


Figur 4: Världens första röntgenbild togs 1895 av Wilhelm Röntgen. Bilden föreställer Röntgens fru Anna Bertha Ludwigs hand med ringar. För sin upptäckt belönades Röntgen med det första Nobelpriset i fysik. Strålar-  
na bär dessutom hans namn i många språk såsom tyska och de nordiska språken. I sann vetenskaplig anda kallade Röntgen själv strålarna för X då han inte visste strålarnas natur, vilket återspeglas i t.ex. spanskan (rayos X), franskan (rayons X), och engelskan (X-rays). Röntgen ogillade starkt hans kollegors förslag att uppkalla strålarna efter honom. Tekniken Röntgen använde för hundra år sedan är mer eller mindre oförändrad idag. Detsamma gäller bildanalysen, som fortfarande är beroende av en läkares expertis.

är givetvis ovärderliga och resultaten talar för sig själva, men metodologin har två stora nackdelar:

1. tid spenderas på att diagnostisera snarare än att behandla patienter,
2. precisionen är och förblir begränsad av den mänskliga faktorn.

En onkolog lägger i snitt cirka 20 minuter av sin tid per patient på att konturera tumörer och andra organ. För att kunna strålbehandla en tumör måste relevanta objekt som tumörer och kritiska organ definieras som tredimensionella former. I sin digitala



Figur 5: i) Bildtagning av 3D-objekt görs ett plan i taget. Planens tjocklek bestäms i förväg. Till exempel har diagnostiska bilder en plantjocklek på cirka 2 mm. Detta sparar tid, men skador eller anomalier mindre än denna tjocklek missas. ii) Data digitaliseras och transformeras matematiskt till verklighetstrogna bilder. Planbilderna sparas i en bildserie där summan av serierna bildar hela den ursprungliga volymen. iii) Data består av gråskaliga bilder. Varje pixel (eller voxel) antar ett numeriskt värde, där högre värde motsvarar starkare ljusintensitet, och vice versa.

ursprungsform består ett organ bara av pixlar (eller voxel) med gråskalig intensitet (se Figur 5). För att dosplaneringsprogrammet skall kunna planera och räkna ut en optimal dosplan så måste en 3D-miljö av patienten och dennes organ skapas. Detta medför att onkologerna manuellt måste konturera upp till 50 objekt. Objektet består av t.ex. olika organ, tumörmassa, ödem, samt virtuella objekt som används vid dosplanering. Vid första ögonkastet verkar denna uppgift som gjord för ett datorprogram, men uppgiften försvåras avsevärt eftersom insidan av människokroppen är en komplex miljö som kan ändra form och utseende även hos en enskild individ. Två bildserier av samma person tagna vid två olika tillfällen kan se avsevärt olika ut, t.ex. om patienten befann sig i olika faser av andningscykeln då bilderna togs. En läkares erfarenhet av vad som finns i en kropp är här helt avgörande för att effektivt kunna konturera och definiera kroppens organ.

Precisionen av mänsklig konturering är, kanske överraskande nog, ett relativt utforskat ämne. De fåtal studier som gjorts kring ämnet pekar dock på samma sak: stor variation. Denna uppkommer inte bara p.g.a. den mänskliga faktorn, utan det finns även delade meningar kring kontureringsprotokoll kliniker emellan. Kontureringar går dessutom inte att verifiera med oberoende mätningar. Ett organs dimensioner erhålls av manuell mätning på patientbilder. Trots detta tolkas den mänskliga kontureringen som grund-

läggande facit när behandling planeras. Det faktum att det ännu inte finns några praktiska alternativ till manuell konturering har tvingat läkarna att bemästra konstformen på egen hand. I en nyligen publicerad artikel visades stor variation när neurokirurger från olika platser skulle konturera samma hjärntumör. Den hjärntumör som var objekt var visserligen av det mer komplicerade slaget, men marginalerna, som visade variation större än de tillåtna behandlingsmarginalerna, kom ändå som en kalldusch.

Här har AI ännu en utmaning. Då kontureringsalgoritmer utvärderas används i princip alltid konturer som skapats av ett lag läkare som grundsanning. Alltså, vi tränar algoritmerna att härma oss människor och bedömer dem utifrån vårt perspektiv. *The BraTS Challenge* är en årligt återkommande tävling där lag kämpar om att utveckla den bästa algoritmen för att automatiskt konturera hjärntumörer. De dataset som algoritmerna använder för att tränas, testas och bedömas, består av ett hundratal bildserier av patienter med hjärntumörer. Här har tumörerna manuellt konturerats och bedömts av en grupp läkare, och bedöms således hålla högsta kvalitet. Detta kan tolkas som orättvist mot maskinerna, men återigen uppstår problemet med brist på oberoende data. Att avgöra en AI:s prestanda är därför inte helt enkelt. Även om tid uppenbarligen kan sparas och variationen minska, återstår det att se huruvida algoritmernas resultat överensstämmer med verkligheten. Att släppa lös en sådan algoritm på riktiga patienter görs inte lättvindigt.

Huruvida AI kan ersätta läkare inom diagnostik har debatterats länge. Historiskt sett har teknologin helt enkelt inte varit tillräckligt bra för att kunna mäta sig med läkarnas kunskaper och erfarenheter, men idag har den genomsnittliga datorkraften blivit så pass kraftfull och algoritmerna så pass avancerade att resultaten sannolikt kan betraktas likvärdiga, om inte bättre, för de sistnämnda. Det naturliga steget framåt är att låta AI vara ett verktyg för läkare. Ett verktyg, precis som så många andra, som låter läkarna spara tid från mekaniska uppgifter och istället fokusera på patientens välmående. I blindtest har automatiska kontureringar genomförda av en AI ansetts ha mycket hög kvalitet av läkare, som själva inte känt behov att korrigera resultaten. Man kan även fråga sig huruvida det är moraliskt att låta en algoritm, vars detaljer ingen förstår, få styra en sådan kritisk uppgift, eller ifall det är mer omoraliskt att avstå från en metod som man tror är bättre.





## Vidare läsning

Asimov, I. (1942). Runaround, *Astounding Stories Magazine*.

Barragán-Montero, A. (2021). Artificial intelligence and machine learning for medical imaging: A technology review, *Physica Medica*, 83, 242.

Pyradiomics documentation (u.å.). *Radiomic Features*.  
<https://pyradiomics.readthedocs.io/en/latest/features.html>.

Siddique, S.; Chow, J. C. L. (2020). Artificial intelligence in radiotherapy, *Reports of Practical Oncology & Radiotherapy*, 25, 656.