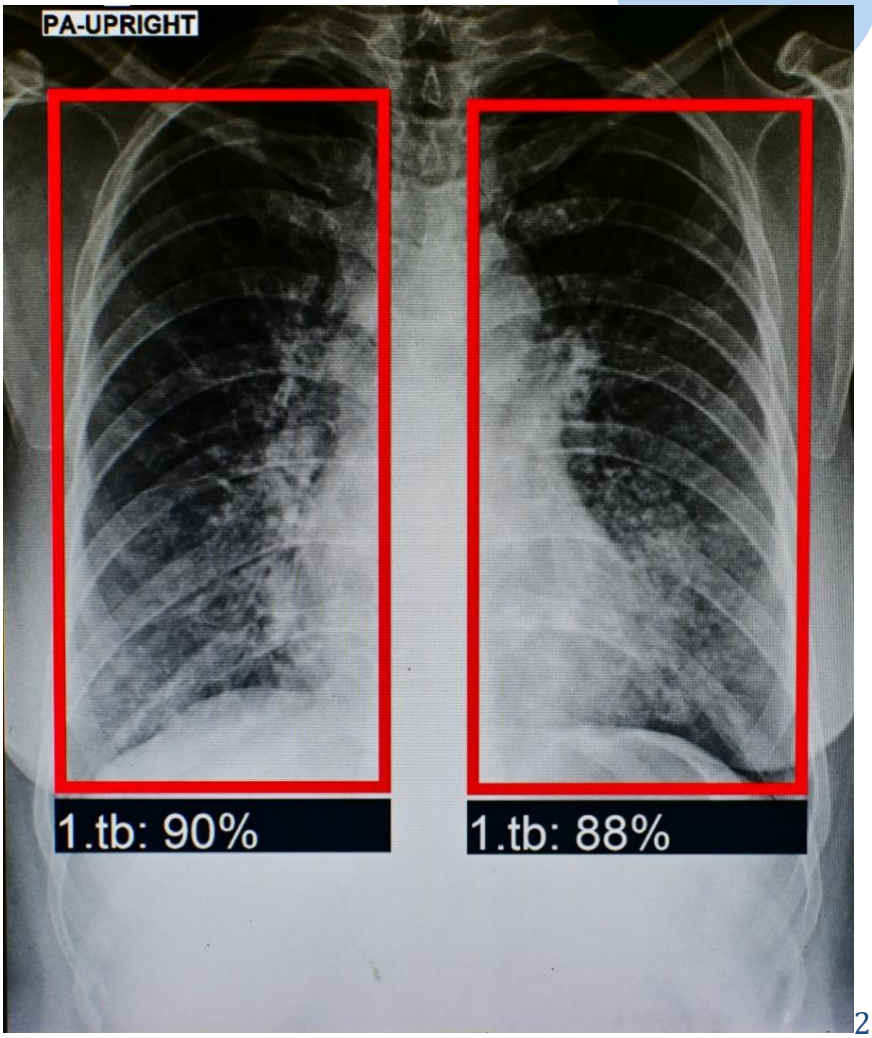
HELSE NORD | DAVVIDEARVVAŠVUOHTA VARRESVUOHTA NUORTTA HEALSOE NOERHTE

Kunstig intelligens innen radiologi i Helse Nord



2

Helse Nord RHF – november 2022

Forord

Da Helse Nord RHF i juni 2021 presenterte den overordnede strategien[1] for å ta i bruk kunstig intelligens i helsetjenesten, var vi først i løypa blant RHFene. I denne oppfølgende rapporten, der vi går mer detaljert og konkret inn på forutsetninger og muligheter for å implementere KI-løsninger innen radiologifeltet, har andre gått foran. Det gjelder særlig anskaffelsesprosjektet innen radiologi i Vestre Viken og Vestfold som har pågått over en periode på 17 måneder (mars 2021-august 2022) med signering av kontrakt ultimo august 2022. I vårt arbeid har vi hatt god og jevnlig kontakt med dette prosjektmiljøet, men av formelle og kontraktsmessige grunner fikk vi først tilgang til vurderingene og resultatene av dette arbeidet på sensommeren i år, etter at kontrakten var signert. I samråd med ledelsen i Helse Nord RHF, valgte vi derfor å utsette ferdigstillingen av vårt arbeid for å få tilgang til de viktige erfaringene fra Vestre Viken-prosjektet. Vi er svært takknemlige for den åpenhet og velvilje vi har møtt i kontakt med dette pionerprosjektet.

Vår tilnærmingsmåte er noe annerledes enn den som har vært valgt i Vestre Viken HF. Mens sistnevnte har gått direkte på en anskaffelse med utgangspunkt i en utarbeidet kravspesifikasjon, har vi i Helse Nord valgt å gå den noe lengre og mer omstendelige veien, med utforming av en generell overordnet strategi for KI og en oppfølgende konkretisert utredning om radiologifeltet. Vi har intet grunnlag for å påstå at vår tilnærmingsmåte er bedre, men ser fordeler i å vinne erfaringer med ulike arbeidsmåter innen et fagfelt som ennå må sies å være i en tidlig læringsfase. Samtidig skal Vestre Viken HF berømmes for at de går foran og viser vei.

For samtlige medlemmer av den tverrfaglige sammensatte arbeidsgruppen, har arbeidet med rapporten vært en søkende lærings- og modningsprosess. Det har i så henseende vært til nytte at denne prosessen har strukket seg over mange møter i løpet av de siste 11 måneder. Under arbeidet har det vært en viss arbeidsdeling mellom medlemmene primært ved at radiologer og radiografer har hatt hovedansvaret for behovsvurderinger og ellers håndtert spørsmål som krever domenekompetanse innen forannevnte fagfelt. Mer KI-tekniske og strategiske problemstillinger er primært håndtert av sekretariatet, men med deltagelse fra hele arbeidsgruppen; flere av medlemmene har også levert viktige skriftlige innspill til arbeidet. Hele arbeidsgruppen står slik sett bak de vurderinger og anbefalinger som gjøres i rapporten. Det meste av tekstene i dette dokumentet er ført i pennen av undertegnede (FHH og KØM) som dermed bærer et betydelig ansvar for dokumentets form og innhold. Under arbeidets gang har Sture Pettersen i Helse Nord RHF vært en utmerket tilrettelegger, møtereferent og samtalepart.

[1] Strategi for kunstig intelligens i Helse Nord 2022-2025. Helse Nord RHF juni 2022

De vurderinger som presenteres i rapporten bygger imidlertid på en rekke datakilder, i form av publikasjoner og brede kontakter med en rekke internasjonale og nasjonale miljøer, både blant leverandører og akademiske institusjoner. Det innebærer på ingen måte at det materialet vi bygger på, er uttømmende eller komplett. KI innen radiologi er et relativt nytt, men sterkt ekspanderende og dynamisk fagfelt der det hver måned kommer nye produkter og mange nye publikasjoner. Vi har etter beste evne, forsøkt å skissere status for dette feltet, slik det framstår for oss ved avlevering av rapporten, vel vitende om at flere av vurderingene vil ha en utløpsdato som ikke ligger langt fram i tid; det gjelder særlig i forhold til markedet for kommersielle løsninger. Rapporten bør derfor primært leses som en forenklet metodebok der vi, på grunnlag av andres erfaringer og litteraturstudier, skisserer en tilnærmingsmåte med fokus på noen viktige spørsmål og hensyn som bør vektlegges og avveies ved anskaffelser og implementeringer. Denne rapporten må ellers sees i sammenheng med den overordnede strategirapporten fra juni 2021. For flere av de mer generelle og felles problemstillinger som bl.a. gjelder referanse til juss, etikk, sikkerhet og anskaffelsesregimer, vises til strategirapporten. Noen av disse problemstillingene berøres likevel også i denne rapporten, men da med mer spesifikt fokus på KI innen radiologifeltet.

Blant de internasjonale miljøer vi har hatt kontakt med, fremheves særlig miljøet rundt Aiforradiology ved Radboud-universitetet i Nederland. Dette miljøet har etablert den viktige databasen som holder løpende oversikt over CE-merkede produkter innen radiologi, og utgir dessuten et månedlig nyhetsbrev med referanser til nye produkter og nye relevante publikasjoner om KI innen radiologi. Vi vil her særlig nevne lederen av Aiforradiology, Kicky van Leeuwen, som i møte med arbeidsgruppen også har gitt nyttige råd og innspill til vårt arbeid. Tim Govers ved Radboud-universitetet har gitt viktige innspill til avsnittet om helseøkonomiske vurderinger.

I det nasjonale miljøet vil vi nevne følgende personer som – primært direkte, men også indirekte - har deltatt med innspill vis a vis arbeidsgruppen:
Hauke Bartsch, MMIV, Atle Bjørnerud, Universitetet i Oslo, Pål Brekke, Oslo Universitetssykehus, Helga Brøgger, Helsetilsynet, Haakon Eisler, Helse Nord RHF, Siv Anette Fjellkårstad, Helsedirektoratet, Bjørn Anton Graff, Vestre Viken HF, Ove Granberg, HNIKT, Inge Groote, Sykehuset Vestfold HF, Kate-Mari Hagen, Sykehusinnkjøp, Einfrid Haldorsen, Helse Bergen HF, Mathias Hauglid, Norges arktiske universitet- UiT, Solveig Hofvind, Kreftregisteret, Hilde Margrethe Lovett, Direktoratet for ehelse, Njål Lura, Helse Bergen HF, Ida-Kristin Martinussen, Helse Nord RHF, Johan Rasch, Inger Merete Retzius, Sykehuspartner, Ann Elisabeth Rødvei, Helse Nord RHF, Helse Nord RHF, Erik Skjemstad, HNIKT, Olav Aune Thomassen, Finnmarkssykehuset HF, Binz Vallevik, DNV

På vegne av arbeidsgruppen,

Finn Henry Hansen	Karl Øyvind Mikalsen
(leder av arbeidsgruppen)	(faglig leder av sekretariatet)

Innhold

FORORD **2**

LESEVEILEDNING **7**

DEL 1: OPPSUMMERING **8**

1 OPPSUMMERING AV HOVEDELEMENTENE I RAPPORTEN **8**

DEL 2: BAKGRUNN **24**

2 INNLEDNING OG BAKGRUNN **24**

2.1 Bakgrunn og etablering av arbeidsgruppe **24**

2.1.1 Bakgrunn 24
2.1.2 Formål med arbeidet 24
2.1.3 Omfang, føringer og krav til gjennomføringen 24
2.1.4 Avgrensing og avhengigheter 25
2.1.5 Krav til resultat 26
2.1.6 Tidsramme og ressursbehov 26
2.1.7 Finansiering 27
2.1.8 Organisering og ansvar 27
2.1.9 Kommunikasjon og interessentoppfølging 27

3 HVORFOR RADIOLOGI VURDERES SOM ET PRIORITERT OMRÅDE FOR Å TA I BRUK KI-LØSNINGER **29**

3.1 Utviklingstrekk innen medisinsk radiologi **29**

***3.2* Kunstig intelligens som ett av flere virkemidler for å avhjelpe kapasitets- og kvalitetsutfordringene innen radiologien** **33**

3.3 Særskilte utfordringer for den radiologiske virksomheten i Helse Nord **34**

3.4 Status for bemanning av radiologer ved helseforetakene i Helse Nord per august 2022 ..... **37**

3.5 Radiologiske undersøkelser for ulike modaliteter i 2021 – fordelt på de fire helseforetakene og private aktører. **38**

4 DEN RADIOLOGISKE ARBEIDSFLYTEN **41**

DEL 3: BEHOVSSIDEN **45**

5 KARTLEGGING AV VOLUM/AKTIVITET I HELSE NORD PÅ UTVALGTE FAGOMRÅDER INNEN RADIOLOGI 45

6 BEHOVSKARTLEGGING 50

6.1 Behov innen radiologi 50
6.1.1 Konkretisering av behov med utgangspunkt i volum 51

6.2 Behov sett fra radiografers ståsted 56

6.3 Andre behov innen medisinsk bildeanalyse 58

DEL 4: TILBUDSSIDEN 60

7 STATUS FOR DEN KOMMERSIELLE TILBUDSSIDE INNEN FASE 0-3 60

8 DEN KOMMERSIELLE TILBUDSSIDEN INNEN BILDETOLKNING (FASE 4) 62
8.1.1 Abdomen 62
8.1.2 Mammografi 63
8.1.3 Hjerte 63
8.1.4 MSK (Muskel-skjelett) 64
8.1.5 Neuro 64
8.1.6 Thorax 65
8.1.7 Andre 66

8.2 Norske erfaringer med anskaffelse av verktøy for KI-støttet bildetolkning 66

DEL 5: ARBEIDSGRUPPENS ANBEFALINGER 70

9 ANBEFALINGER 70

9.1 Generelle vurderinger av markedet for KI-støttet bildetolkning 70

9.2 Prioriterte fagområder innen KI-støttet bildetolkning 72

9.3 Anbefalinger for UNN HF som har opsjon på avrop etter rammeavtalen som er inngått i Helse Sør-Øst 75

9.4 Anbefalt anskaffelsesprosedyre og hensyn som bør vektlegges i en kravspesifikasjon 77
9.4.1 Anskaffelsesprosedyre 77
9.4.2 Hensyn som bør vektlegges i en kravspesifikasjon 78

9.5 Fagområder der egenutvikling kan være aktuelt 79

DEL 6: ANDRE ANBEFALINGER OG FORHOLD 82

10 VALG AV PLATTFORM SOM STRATEGISK BESLUTNING 83

11 BETYDNINGEN AV TILRETTELAGT LOGISTIKK OG ARBEIDSFLYT 85

11.1 Overordnet om logistikk og arbeidsflyt 85

11.2 Ulike behov for ø-hjelp og elektive undersøkelser 86

12 INFRASTRUKTURELLE FORUTSETNINGER FOR IMPLEMENTERING AV KI-LØSNINGER INNEN RADIOLOGI 87

12.1 Regional IKT-plattform for å understøtte kunstig intelligens i Helse Nord (RIKTIG-prosjektet) 87

12.2 ROS-analyser mht. informasjonssikkerhet, personvern og pasientsikkerhet 88

13 REGULATORISKE ENDRINGER - FRA MDD TIL MDR 91

14 STUDIER, ERFARINGER OG KOST-NYTTE-ANALYSER 95

14.1 Helseøkonomiske analyser av KI-løsninger. 95

14.2 Implementering – studier og erfaringer 97

14.2.1 NSEs kunnskapsbaserte anbefalinger i lys av andre kilder og vurderinger som ligger til grunn for arbeidsgruppens rapport. 99

14.3 Nasjonale KI-Rapporter i regi av Helsedirektoratet og Direktoratet for e-helse 100

15 KOMPETANSEOPPBYGGING OG UTDANNING AV RADIOGRAFER OG RADIOLOGER 103

15.1 Kompetanseoppbygging 103

15.2 Utdanning 104

16 VALIDERING OG KVALITETSSIKRING AV KI-VERKTØY 106

16.1 Vurdering av behovet for å validere kommersielle løsninger mot egne pasientdata 106

16.2 Kvantitative metoder for å evaluere KI-algoritmer 107

16.3 Kriterier for vurdering av algoritmers kvalitet – radiologer som komparenter 109

16.4 Hvordan anvendes algoritmen? – arbeidsdeling mellom algoritme og fagperson 113

17 FORVALTNING 116

Leseveiledning

Denne rapporten er inndelt i seks deler med til sammen 17 kapitler.

Del 1 er en oppsummering av rapportens hovedelementer. Denne delen kan leses frittstående og gir en rimelig god oversikt over de de viktigste tema og problemstillinger som belyses, samt arbeidsgruppens anbefalinger. Oppsummeringen er derfor noe lengre enn det som kanskje er vanlig for en slik rapport. For mange av leserne vil det være tilstrekkelig å kun lese denne delen, mens fagpersoner innen feltet anbefales å lese hele rapporten.

Del 2-5 utgjør hoveddelen av rapporten. Denne bolken rundes av med kap. 9 som inneholder arbeidsgruppens viktigste anbefalinger. Del 6 er en samling av kapitler om en rekke temaer som er relevante ved innføring av KI innen radiologi. Mange av kapitlene i denne delen vil kunne leses frittstående for de av leserne som er spesielt interesserte.

Mer konkret, er Del 2-6 inndelt som følger:

Del 2 redegjør for bakgrunnen og mandat for dette arbeidet, oppsummerer noen generelle utviklingstrekk for det radiologiske fagområdet og drøfter mer spesifikt de utfordringer som preger dette fagområdet i Helse Nord. I denne delen redegjøres også for den radiologiske arbeidsflyten.

I del 3 presenteres en behovskartlegging utført av radiologene og radiografene i arbeidsgruppen, med vekt på områder der KI-løsninger kan tenkes å være potensielt nyttige.

Del 4 drøfter tilbudssiden for tilgjengelige kommersielle produkter i de ulike faser av den radiologiske arbeidsflyten, men med hovedvekt på produkter for tolkning av bilder; her redegjøres det også for norske erfaringer med anskaffelse av verktøy for KI-støttet bildetolkning, med hovedfokus på anskaffelsen i Vestre Viken HF.

I del 5 oppsummeres arbeidsgruppens generelle vurdering av markedet for KI-produkter innen dette feltet og arbeidsgruppens anbefalinger om prioriterte områder for anskaffelse i Helse Nord.

Del 6 er en samling av kapitler om nødvendige rammer og forutsetninger for trygg innføring av KI-løsninger, slike som: tilrettelagt logistikk/arbeidsflyt, infrastruktur, hensyn til regulatoriske forhold, kost/nytte-analyser, kompetanse, validering og kvalitetssikring. Det vises her også til noen viktige dokumenter fra det nasjonale koordineringsprosjektet for KI i helsetjenesten.

DEL 1: Oppsummering

1 Oppsummering av hovedelementene i rapporten

HOVEDDELEN AV RAPPORTEN

1.1 Innledning, bakgrunn og mandat

I kapittel 2 redegjøres for bakgrunn og mandat. I Helse Nords strategi for kunstig intelligens, vedtatt i styremøte 22.juni 2021, ble det anbefalt å videreføre arbeidet med KI gjennom en rekke oppfølgingstiltak; blant disse var å etablere en arbeidsgruppe med mandat til å utrede og konkretisere hvilke KI-løsninger innen radiologi som kan være aktuelle for implementering i Helse Nord. Formålet med arbeidet ble presisert som å fremme forslag om hvilke fagområder og modaliteter innen radiologien som de nærmeste årene ble vurdert å ligge best til rette for implementering av KI-løsninger som beslutningsstøtte i helseforetakene i Helse Nord.

I mandatet ble det gitt en rekke føringer for gjennomføringen av arbeidet. I arbeidet med å identifisere løsninger som ligger til rette for implementering, skulle det gjøres behovs-vurderinger (fagområder der KI-løsninger ville være avlastende for radiologene) som deretter skulle koples til fagområder der det var utviklet CE-merkede KI-algoritmer. Kriterier som behov, nytte og løsningenes modenhet skulle legges til grunn for anbefalingen om hvilke løsninger som bør prioriteres for implementering. Disse vurderingene skulle gjøres på ved hjelp av flere datakilder: gjennomgang av produkter og produkttyper, erfaringer fra anskaffelsen i Vestre Viken HF og fra begynnende implementeringer i andre land, samt forskningsbasert kunnskap gjennom studier publisert i tidsskrifter mv. Arbeidsgruppen skulle også avklare viktige forutsetninger og avhengigheter for implementering av løsninger: nødvendig infrastruktur, datafangst, regulatoriske ordninger, behovet for validering av kommersielle løsninger, ROS-analyser mv.

Til å gjennomføre arbeidet ble det etablert en arbeidsgruppe, ledet av Helse Nord, med til sammen ti medlemmer; samtlige helseforetak er representert med radiologer, mens to av HFene er representert med radiografer. Representanter fra brukere og arbeidstakere har også vært med i arbeidet. Det har vært avholdt i alt 11 møter, hvorav 10 har vært organisert som Teamsmøter.

1.2 Utviklingstrekk innen medisinsk radiologi og status i Helse Nord

Som en bakgrunn for vurderingen av behov for KI-løsninger, er det gjort en gjennomgang av noen generelle utviklingstrekk innen fagområdet medisinsk radiologi. Dette er globale trender som i hovedsak også manifesterer seg nasjonalt og regionalt. Radiologien som fagfelt har ekspandert kraftig i løpet av de senere tiår, både kvantitativt

og kvalitativt. Ny og forbedret teknologi har medført at det gjøres flere og mer komplekse undersøkelser for stadig flere sykdomsgrupper, og at radiologiske metoder i økende utstrekning tas i bruk for behandlingsformål (intervensjonsradiologi) og for å evaluere effekten av behandlinger. I tillegg kommer stadig mer tidkrevende deltagelse i MDT-møter (demonstrasjoner med tverrfaglig deltagelse, primært knyttet til pakkeforløp ved kreftsykdom).

Det foreligger foreløpig ingen oppdatert nasjonal oversikt over veksten i antallet radiologiske undersøkelser. Vi bruker derfor data fra Danmark – som ikke antas å avvike vesentlig fra Norge - for å belyse denne veksten. I perioden mellom 2011 og 2021 var den samlede veksten av undersøkelser på 25 %, mens den innen modalitetene CT og MR var på henholdsvis 82% og 62%. Undersøkelser innen disse modaliteter er dessuten mer komplekse og tidkrevende enn konvensjonell røntgen.

Ekspansjonen er ledsaget av en tiltagende arbeidsbelastning for radiologene, noe som både fører til økende forekomst av utbrenthet og økt risiko for feilvurderinger. Innen den offentlige helsetjenesten er det økende rekrutteringsutfordringer, og mange radiologer søker seg over til en ekspanderende privat sektor. Det er følgelig et stort behov for tiltak som kan avlaste radiologene, og det vurderes derfor i utgangspunktet som gunstig at dette er et av de fagområdene som ligger best til rette for implementering av KI-løsninger. Dette er Helse Nords arbeidsgruppes primære begrunnelse for å prioritere dette fagområdet i KI-strategien.

Flere av utfordringene for radiologifeltet i Helse Nord er belyst i en nylig avgitt ekstern rapport utarbeidet av et konsulenthus, på oppdrag fra Helse Nord RHF. Her påpekes både rekrutteringsproblemene og den høye gjennomtrekken i stillingene. I tillegg påpekes organisatoriske og ledelsesmessige utfordringer, samt for lite samordning og samarbeid både innen og mellom helseforetak.. Flere av arbeidsgruppens medlemmer stiller seg likevel kritisk til at rapporten blir noe overflatisk og ikke i tilstrekkelig grad har belyst de vesentlige forskjeller som foreligger både mht. tilstand og behov mellom de ulike HFene. Arbeidsgruppen stiller seg likevel positivt til etablering av et regionalt prosjekt i regi av Helse Nord RHF i samarbeid med HFene.

I kap. 3 redegjøres det avslutningsvis både for bemanningssituasjonen for radiologer i de enkelte HF i Helse Nord (og private aktører), og for volumet av radiologiske undersøkelser fordelt på de ulike modaliteter. Her framgår det at 16 prosent av alle undersøkelsene utføres av private aktører, mens denne andelen er hele 50 % for MR-undersøkelser. Hele halvparten av alle undersøkelsene utføres av de 9 lokalsykehusene.

Arbeidsgruppens vurderinger og anbefalinger:

- Situasjonen for radiologifaget i Helse Nord er kritisk mht. bemanning, kompetanse, organisering og regional samordning.

- Implementering av løsninger for kunstig intelligens vil på lengre sikt kunne bidra til å understøtte og avlaste radiologene
- Det anbefales at Helse Nord – gjennom organisering av det regionale radiologi-prosjektet (oppfølging av Karabin-rapporten)- innarbeider premisser fra denne rapporten bruk av KI innen radiologien

1.3 Den radiologiske arbeidsflyten

Som et hjelpemiddel for lesere som ikke er veldig godt kjent med radiologi har vi i kap. 4 kort beskrevet de ulike fasene i den radiologiske arbeidsflyten.

Det som skjer i forkant av billedtaking har vi valgt å referere til som fase 0:

0. Henvisning og timesetting

Det som skjer under selve bildetakingen kan deles i fire faser:
1. Planlegging av billedtaking; 2. Gjennomføring av billedtaking; 3. Rekonstruksjon og postprosessering av bilder; 4. Beskrivelse/tolkning av bilder.

I tillegg er det en siste fase som skjer i etterkant:

5. Kommunikasjon av resultat til kliniker / pasient.

1.4 Kartlegging av volum og aktivitet for det radiologiske fagområdet

I kap. 5 er det gjort en kartlegging av volum og aktivitet for ulike modaliteter og undersøkelsestyper innen det radiologiske fagområdet i Helse Nord. Blant annet fremstilles en oversikt over antall undersøkelser (volum) fordelt på ulike typer av sykehus, med vekt på å få fram forskjellene mellom sykehusgrupper på ulike funksjons- og spesialiseringsnivå. I denne rapporten vurderes det som særlig relevant å skille mellom aktiviteten ved UNN-Tromsø, Nordlandssykehuset-Bodø og de 9 lokalsykehusene. Oversikten viser imidlertid at det innen de fleste modaliteter foregår en betydelig aktivitet, også ved lokalsykehusene.

1.5 Behovskartlegging

I kapittel 6 gjøres en oppsummering av behovene, slik de er beskrevet av de radiologer og radiografer som er medlemmer av arbeidsgruppen. Radiologene kopler behovene til områder der arbeidsbelastningen er særlig stor: kreftutredninger og kontroller, tidkrevende beskrivelser basert på målinger og sammenligning med tidligere undersøkelser og bruddundersøkelser som har stort volum.

Med utgangspunkt i de innledende diskusjonene i arbeidsgruppa utformet radiologene en rangert liste som beskriver hvilke behov KI potensielt kan bidra til å avhjelpe:

Markering av områder med patologiske funn er gitt første prioritet:

a. Deteksjon, måling og kontroll av lesjoner, kreftsvulster og metastaser.
b. Tolking av skjelettrøntgen, særlig brudd-deteksjon
c. Tolking av røntgen-thorax, inkludert tuber og slanger
d. Oppfølging av MS-lesjoner (MR-Caput)
e. Slagdiagnostikk
f. Deteksjon av lungeemboli og målinger av kar
g. Tolking av MR-hjerte
h. Kvantitative beskrivelse av forløp for lungesykdommer
i. Differensialdiagnoser av kreftsvulster

Dernest er i følgende prioritert i rekkefølge; triagering av undersøkelser og henvisninger, støtte til både henviser og radiologisk avdeling for å sikre riktig og nødvendig informasjon, annotering/segmentering av anatomi, talegjenkjenning, automatisk skjelettalder, forbedret bildekvalitet og forbedret pasientlogistikk.

Denne rangerte listen ble sammenstilt med volum og aktivitet, hvilket har ledet til en konkretisering av behovene hvor det er skilt mellom elektiv og øyeblikkelig hjelp; dette fordi disse to hastegradskategorier er vesensforskjellige med tanke på funksjonalitet, nytte og gevinst.

Elektive undersøkelser

For elektive undersøkelser er det kreftdiagnostikk, utredninger og kontroller som utgjør hovedmengden av undersøkelser.

For elektive undersøkelser er behovet generelt:

1. Utredning (deteksjon patologi og differensiering, ekskludere patologi)
2. Kontroller (utvikling av patologi, deteksjon ny patologi)
3. Kvalitet (redusere feiltolkning og oversett patologi)
4. Objektivisering og kvantifisering (utvikling patologi over tid, leselige tabeller, RECIST)
5. Strukturert rapportering (automatisert overføring av mål og evt. endring over tid inn i beskrivelsen, f.eks. i tabeller; RECIST)
6. Effektivisering
7. Triagering

Aktuelle elektive undersøkelser der en ser behov: CT Thorax, CT/MR Abdomen, RG Thorax, RG Skjelettalder, MR Hjerte, MR Mamma, RG Mammografi, MR Prostata, MR Caput, MR Skjelett, Nukleærmedisin: f.eks. PET/MR prostata, UL thyreoidea og MR kolumna.

Øyeblikkelig hjelp

Ved øyeblikkelig hjelp undersøkelser er det flere områder med behov:

1. Triagering (hva haster mest å se på)
2. Avlastning på vakt (brudd-deteksjon)
3. Kvalitet (Redusere feiltolkning og oversett patologi)
4. Deteksjon (Påvise patologi)

Aktuelle øyeblikkelig hjelp undersøkelser der en ser behov: CTANG Angiografi av Thorax og CT Thorax, CT Caput, CT Abdomen, RG Skjelett og RG Thorax

Radiografers behov

Radiografene har utformet følgende liste over behov der KI potensielt kan bidra: Algoritmer som kan foreslå protokoll for undersøkelser, basert på informasjon i henvisning, bistand til posisjonering av pasient, bistand til å velge sekvenser og vinkling (MR), bruddalgoritme som kan avlaste radiolog på vakt, triagering, opplæring, automatisert link mellom prioriteringskode/protokoll og prosedyre, bistand til radiolog når undersøkelsesprotokoll skal velges.

1.6 Status for den kommersielle tilbudssiden i fase 0-3

Det er ikke funnet egnede kommersielle og sertifiserte KI-verktøy som kan bidra til oppgavene i fase 0 av den radiologiske arbeidsflyten.

Når det gjelder oppgavene i fase 1-3 er de tett tilknyttet maskinvaren som brukes for å gjennomføre skanningen. Det vil variere mellom maskinleverandørene i hvor stor grad de bruker KI-basert programvare for å eksempelvis øke produktiviteten ved å muliggjøre kortere skannetider eller forbedre bildekvaliteten gjennom å fjerne støy. KI-løsninger innenfor disse tre fasene, vil ofte være knyttet til den enkelte modalitetens leveranse og inngå som del av eller opsjon til programvaren som følger maskinen.

En investering i ny maskinvare for billedtaking vil derfor også være førende for hvilke muligheter som åpner seg mht. å få hjelp av kunstig intelligens til å planlegge og gjennomføre billedtaking. Denne avhengigheten gjør at fleksibiliteten reduseres; derfor vil anbefalinger rundt bruk av KI i disse tre fasene bli noe lavere prioritert i denne rapporten. Likevel vil arbeidsgruppen påpeke at det er viktig å ha fokus på også disse problemstillingene når helseforetakene i nord i tiden framover skal fornye og oppgradere maskinparken.

1.7 Status for den kommersielle tilbudssiden innen bildetolkning (fase 4)

Brorparten av det kommersielle tilbudet av sertifiserte KI-verktøy innen radiologi bidrar med støtte til bildetolkning / beskrivelse av bilder, som vi har referert til som fase

4 i den radiologiske arbeidsflyten. Arbeidsgruppen har brukt nettstedet www.AIforradiology.com som hovedkilde for å få oversikt over dette kommersielle markedet. I tillegg har de største leverandørene (Siemens, Sectra, GE Healthcare, Philips og AIdoc) blitt invitert inn i møter med arbeidsgruppen hvor satsningen på KI til de respektive selskapene har blitt presentert og diskutert.

Radiologene og radiografene i arbeidsgruppa har sittet sammen i grupper og gått gjennom tilgjengelige KI-produkter innenfor spesialitetene: abdomen, mammografi, hjerte, MSK, neuro og thorax. Totalt har arbeidsgruppa gått gjennom over 200 MDD og MDR-godkjente algoritmer og identifisert at ca. 25 av disse svarer til behovene vi har beskrevet i kap. 6. En fyldigere oppsummering av gjennomgangen er beskrevet i kap. 8.

I kap. 8.2 er det redegjort for anskaffelsen av KI-algoritmer innen radiologi i Vestre Viken HF. Anskaffelsen ble gjennomført i samarbeid med Sykehuset Vestfold HF, Sykehuspartner HF og Sykehusinnkjøp HF. Med konkurransepreget dialog som anskaffelsesform, ble det lyst ut en konkurranse våren 2021 og signert kontrakt i august 2022. På grunnlag av behovsvurderinger gjort av radiologene i helseforetaket, ble følgende fem områder innen bildeanalyse prioritert for anskaffelse: CT-thorax - lungenoduler/lungemetastaser, CT thorax - lungeemboli, MR Caput – MS-oppfølging, konvensjonell røntgen thorax, konvensjonell røntgen- skjelett/brudd. I løpet av dialogen ble det også avklart å inkludere plattform for KI-algoritmer i den endelige kravspesifikasjonen.

9 leverandører søkte om å få delta i denne konkurransen, men i løpet av dialogen ble dette antallet redusert slik at bare fire av leveandørne ble invitert til å levere tilbud. I tildelingskriteriene var kvalitet vektet med 70 % og pris med 30 %. Prismodellen som ble lagt til grunn var pris per undersøkelse, men med fallende enhetspris for høyere volum. For de aktuelle algoritmene samlet, var forskjellen mellom laveste og høyeste pris i tilbudene på ca. 25 prosent.

Erfaringene fra anskaffelsen i Vestre Viken HF kan oppsummeres slik:

- Om man hadde startet prosessen med en leverandørkonferanse, ville man både fått et bedre informasjonsgrunnlag for utforming av kravspesifikasjonen, og man ville også redusert noe av dialogtiden med leverandørene.
- Selv om konkurransepreget dialog var en transaksjonskrevende anskaffelsesform, bidro dialogene med leverandørene til mye læring og tilpasning underveis i prosessen.
- Ved både utforming av kravspesifikasjon og vurdering av tilbudene har samarbeidet med Sykehuspartner og Sykehusinnkjøp vært både nyttig og nødvendig.

Arbeidsgruppen anbefaler:

- At Helse Nord i sitt videre arbeid nyttiggjør seg av erfaringene fra anskaffelsen i Vestre Viken HF
- At Helse Nord høster ytterligere erfaringer fra den implementeringsfasen Vestre Viken HF nå går inn.

1.8 Arbeidsgruppens anbefalinger – KI-løsninger

Flere av aspektene i fase 0-3 i den radiologiske arbeidsflyten (se kap.4) treffer behovene som er beskrevet tidligere i denne rapporten. Vårt arbeid viser imidlertid at tilbudssiden er forholdsvis umoden når det gjelder mange av oppgavene i disse fasene. I tillegg er det slik at oppgavene i fase 1-3 er tett tilknyttet maskinvaren som skal brukes for å gjennomføre skanningen. Hovedfokuset for dette arbeidet har derfor vært rettet mot markedet for KI-algoritmer for tolkning av bilder (fase 4).

Arbeidsgruppen anbefaler:

- Tilgjengelig KI-programvare bør også være med i vurderingene når helseforetakene i nord i tiden framover skal fornye og oppgradere maskinparken.

Kapittel 9 fortsetter med en vurdering av markedet for KI-løsninger innen bildetolkning. Det konstateres at det i løpet av de senere år har skjedd en betydelig tilvekst av kommersielle CE-merkede KI-algoritmer, og at det det i tillegg er kommet en rekke løsninger for oppkopling av algoritmer, benevnt plattformer eller markedsplass. Innen de funksjonelle fagområder som er prioritert av arbeidsgruppen (kap. 6), er det identifisert ca. 25 aktuelle algoritmer (kap.8). Det foreligger derfor et markedsmessig grunnlag for å gjøre kommersielle anskaffelser, når dette skjer på en kontrollert og kvalitetssikret måte. Tiden er derfor moden for å gjøre de første anskaffelsene.

Arbeidsgruppen anbefaler:

- At det i forkant av kunngjøring om konkurranse, arrangeres en leverandørkonferanse for å bli oppdatert om løsninger og tjenester
- At det organiseres en felles anskaffelsesprosess i Helse Nord, med konkurransepreget dialog som anskaffelsesform
- At det legges opp til en felles plattformløsning for alle helseforetakene i Helse Nord
- At det legges opp til en rammeavtale som gir tilgang til aktuelle og prioriterte algoritmer gjennom avrop over flere år, tilpasset behov, økonomi og andre forutsetninger i de enkelte helseforetak

Arbeidsgruppen foreslår å prioritere følgende områder for anskaffelse:

1. Brudd: deteksjon og triagering (RG skjelett)
2. Kreftdiagnostikk
    a. Diagnostikk av lungekreft, inkludert tilfeldig påviste noduli (CT thorax)

   b. Diagnostikk av prostatakreft (MR prostata)
   c. Diagnostikk av brystkreft (MR mamma)
3. Tolkning, deteksjon av patologi og triagering for røntgen thorax
4. Slagdeteksjon (CT caput)
5. Deteksjon og triagering av lungeemboli (CT thorax)
6. Tolkning av lungefortetninger/ ILD (CT Thorax)
7. Tolkning av MR hjerte

1.8.2 Anbefalinger for UNN HF som har opsjon på avrop etter rammeavtalen som er inngått i Helse Sør-Øst

Som allerede nevnt, har UNN HF opsjon for å gjøre avrop på den rammeavtalen om anskaffelse av KI-algoritmer som Vestre Viken HF har inngått med valgt leverandør. Siden ingen av de øvrige helseforetak i Helse Nord har slik opsjon, er det følgelig ikke mulig for disse å gjøre avrop for innkjøp av de algoritmer (inklusive plattform) som inngår i denne avtalen. Det kan likevel være klokt å la UNN HF gjøre slikt avrop, selv om øvrige helseforetak i Helse Nord ikke kan inkluderes. Det presiseres at UNN HF som eget rettssubjekt står fritt til å gjøre avrop i henhold til nevnte avtale, uavhengig av hva Helse Nord RHF måtte mene om det. Samtidig vurderer arbeidsgruppen det som både ønskelig og hensiktsmessig at slike initiativ ved UNN skjer i samforstand med Helse Nord RHF og øvrige helseforetak i Helse Nord.

Arbeidsgruppen anbefaler at UNN HF, med sin nærhet til betydelige kompetansemiljøer innen kunstig intelligens og i kraft av sin funksjon som universitetsklinikk, starter med å implementere en eller flere løsninger før vi går videre med implementering i øvrige helseforetak. Dermed kan man vinne erfaring som gjør det enklere å implementere slike løsninger øvrige HF. En slik trinnvis strategi som starter med UNN HF, anbefales også ved oppstart på grunnlag av en felles anskaffelse som inkluderer alle fire helseforetak. Ved å gjøre avrop på ovennevnte rammeavtale, vil UNN raskt komme i gang med praktisk og klinisk rettet utprøving og implementering. Dermed vil Senter for pasientnær kunstig intelligens (SPKI) raskere utvikle sin kjernekompetanse knyttet til implementering, noe som er til nytte for hele foretaksgruppen.

Anbefalingen om å la UNN HF gjøre avrop, bygger imidlertid på at forutsetninger knyttet til forankring i ledelse og fagmiljø, algoritmenes kvalitet, planer om validering og følgeforskning, mv., er til stede. I så tilfelle, vil arbeidsgruppen anbefale at UNN HF prioriterer anskaffelse av den algoritmen for brudd-deteksjon som inngår i rammeavtalen med Vestre Viken HF.

1.8.3 Anbefalt anskaffelsesprosedyre og hensyn som bør vektlegges i en kravspesifikasjon

Arbeidsgruppen vil anbefale Helse Nord å vurdere *konkurransepreget dialog* som anskaffelsesform ved anskaffelse av kommersielle KI-algoritmer innen radiologifeltet.

På grunnlag av erfaringer fra Vestre Viken-prosjektet, vil vi også anbefale at det i forkant av utlyst konkurranse arrangeres en *leverandørkonferanse* for å skaffe seg mer oppdatert kunnskap om produktmarkedet og leverandørenes evne til å levere understøttende tjenester. En eventuell anskaffelse anbefales gjennomført samlet for hele foretaksgruppen i Helse Nord, med mulighet for hvert av HFene til å gjøre avrop på rammeavtale innen en 2-4 årsperiode. I den grad forholdene måtte ligge til rette for det (ved sammenfall i tid mv.), kan det vurderes å samarbeide om en anskaffelse med helseforetak utenfor Helse Nord.

For å ivareta behovet for innkjøpskompetanse på best mulig måte, anbefales etablert et best mulig samarbeid mellom ansvarlige for innkjøp i Helse Nord RHF og Sykehusinnkjøp. Den erfaring Sykehusinnkjøp har ervervet gjennom medvirkning i anskaffelsen i Vestre Viken HF, vurderes som så verdifull at den bør komme en eventuell anskaffelsesprosess i Helse Nord til gode.

Arbeidsgruppen har i kap. 9.4.2 redegjort for en rekke hensyn og krav (MDR-godkjenning, fagfellevurderte studier, mulighet for validering mv.) som anbefales vektlagt i en kunngjøring og kravspesifikasjon.

1.8.4 Fagområder der egenutvikling kan være aktuelt

Det er en rekke av behovene som er beskrevet i denne rapporten som ikke dekkes godt nok av det eksisterende tilbudet på markedet og der egenutvikling kan være aktuelt. Dette gjelder f.eks. talegjenkjenning, forbedring og triagering av henvisninger, forbedret pasientlogistikk og på generell basis flere løsninger som kan bidra til fase 0-3 i den radiologiske arbeidsflyten. Fremtidige KI-løsninger som eventuelt skal kunne bidra til å dekke disse behovene vil nødvendigvis måtte handtere andre inputdata enn bare bilder – og gjerne multimodale data (data fra flere kilder, f.eks. pasientjournal, selvrapporterte data og bilder). Dette bør derfor være et prioritert forskningsområde.

Abdomen er et eksempel på et stort og viktig område der vi har identifisert behov som er dårlig dekt. Foreløpig finnes det kun løsninger for MR prostata. Støtte til deteksjon av tumorer i organer som eksempelvis lever finnes det ikke gode nok kommersielle verktøy for. I tillegg ser vi at det finnes svært få kommersielle løsninger innen nukleærmedisin (PET) og få gode løsninger for ultralydbilder (f.eks. UL thyreoidea). Det anbefales derfor at disse områdene prioriteres i videre FoU-arbeid. Vi vil også påpeke at selv om en god del områder er dekt av kommersielle løsninger, betyr ikke nødvendigvis at løsningene er gode nok. For mange av løsningene er det fortsatt manglende evidens. Dette innebærer at også flere av disse områdene bør prioriteres i videre FoU-arbeid.

Medisinsk bildeanalyse er et stort og omfattende fagfelt. I denne rapporten har vi konsentrert oss om bildemodaliteter som beskrives av radiologer. Innen de andre områdene ser det ut til at markedet er mindre modent enn innen radiologi. Det skal

likevel nevnes at det innen noen av disse områdene finnes flere produkter på markedet med brukbar evidens. Spesielt gjelder dette oftalmologi (diabetisk retinopati), kardiovaskulær medisin og nevrologi (begge de to sistnevnte er delvis overlappende med radiologi og derfor i noe grad dekket av AIforradiology). De to førstnevnte faller utenfor mandatet til denne arbeidsgruppa. Vi vil likevel anbefale at behovet for KI innen disse områdene kartlegges nøyere av de respektive fagmiljøene.

Innen andre områder som dermatologi, odontologi, strålingsonkologi, gastroenterologi, gynekologi og digital patologi finnes det få gode kommersielle løsninger. Den senere tid har imidlertid vist at innovasjons- og FoU-arbeid kan lede til gode løsninger innen f.eks. strålingsonkologi og digital patologi. Flere av disse områdene bør også prioriteres i FoU-arbeid i Helse Nord.

Til sist nevnes at FoU-aktivitetene ikke utelukkende bør prioritere utvikling av verktøy. Kanskje like viktig er (følge)forskning med fokus på utprøving, validering og implementering. Slik kompetanse er viktig å bygge opp i fagmiljøer som SPKI.

ANDRE ANBEFALINGER OG BETRAKTNINGER

I den siste delen av rapporten tar vi for oss andre anbefalinger og forhold som har betydning for arbeidet med implementering og bruk av KI-løsninger innen radiologi i Helse Nord.

Dette dreier seg om

- Valg av plattform for KI-løsninger innen radiologi
- Betydningen av god logistikk og arbeidsflyt
- Krav til IKT-infrastruktur, ROS-analyser og informasjonssikkerhet
- Konsekvenser av den regulatoriske endringen med innføring av MDR.
- Analyse av kost-nytte og erfaringer med implementering nasjonalt og internasjonalt
- Behov for kompetanseoppbygging og utdanning
- Behov for rutiner og veiledning knyttet til validering og kvalitetssikring av KI-verktøy
- Forvaltning av KI-løsninger

1.9 Valg av plattform som strategisk beslutning

De store PACS- og maskinleverandørene holder nå på å posisjonere seg slik at de også kan tilby en KI-plattform (økosystem/appstore for KI-verktøy). I tillegg har enkelte spesialiserte selskaper innen medisinsk kunstig intelligens begynt å tilby slike plattformer, slik at det i dag er ca. 30 ulike plattformer i markedet. Disse KI-plattformene baserer seg på en kombinasjon av tredjepartsprodukter og egenutviklede

løsninger. Ved en anskaffelse av kommersielle KI-løsninger vil man derfor måtte ta stilling til om man også skal gå til anskaffelse av en KI-plattform eller om man skal starte i mindre skala og gå for enkeltstående KI-verktøy (algoritmer). Det er en rekke ulike faktorer som bør tas hensyn til ved en slik vurdering. Disse redegjøres for i kap. 10.

1.10 Betydningen av logistikk og arbeidsflyt

Betydningen av logistikk og arbeidsflyt dreier seg dels om teknisk tilretteleggelse av gode og brukervennlige grensesnitt mellom maskin, programvare og brukere, men også om organisering og tilpasning av rutiner for arbeidsflyt i det daglige arbeidet. Begge disse hensyn forutsetter godt samarbeid med og mellom ulike leverandører, med IKT-funksjonen og med helsepersonellet. Som sluttbrukere er helsepersonellets tilfredshet med innpasningen av KI-funksjonalitet i arbeidsflyten, av avgjørende betydning for å lykkes.

For å sikre at brukerne benytter seg av KI-løsninger i en travel hverdag, må slike løsninger – direkte eller via valgt plattform - være integrert i dagens arbeidsflyt i RIS/PACS med godt brukergrensesnitt. Videre må det utarbeides en arbeidsflyt som er tilpasset de ulike behovene innen de til dels svært forskjellige røntgenavdelingene ved våre fire helseforetak. Et annet aspekt knyttet til arbeidsflyt er behovet for å stratifisere funksjonalitet for ulike brukere

Logistikk og arbeidsflyt vil på overordnet nivå være ulikt for ø-hjelpspasienter og elektive pasienter. For ø-hjelpspasienter er hensikten til de aller fleste KI-applikasjoner å være en støttefunksjon for bedre triagering. Dette vil hjelpe vaktpersonell mht. hvilken undersøkelse de skal prioritere først å se på. Resultatet fra triageringen må vises på hensiktsmessig måte i radiologenes arbeidsliste. For elektive pasienter må det komme frem i arbeidslisten om undersøkelsen er gjennomgått av KI eller ikke, f.eks. med et eget elektivt KI-ikon. Positive funn her ville kunne presenteres på ulike måter, og i noen applikasjoner er det hensiktsmessig med strukturert rapport som lar seg integrere i beskrivelsen, f.eks. tabeller.

1.11 Infrastrukturelle forutsetninger og ROS-analyser

For å ivareta en god brukeropplevelse av IKT-tjenester innenfor bildediagnostikk må nødvendige samarbeidspartnere som Helse Nord IKT være rustet med både ressurser og kompetanse for å kunne levere nødvendige IKT-tjenester innen KI til ønsket tid. Ved implementering, drift, validering, m.m. av kommersielt anskaffede KI-verktøy kreves det at de infrastrukturelle forutsetningene som forvaltes av Helse Nord IKT er på plass. Dette gjelder bl.a. krav til lokal infrastruktur, skyløsninger, informasjonssikkerhet, osv.

Ved implementering av KI-verktøy må det tas stilling til om bilder skal overføres via skyløsning til leverandøren for prosessering, eller om det kan gjøres en lokal installasjon (On-prem) av KI-løsningen. Her varierer mulighetene for de ulike KI-verktøyene.

Når det gjelder ytterligere resonnementer og premisser rundt de infrastrukturelle forutsetningene, vil vi i denne rapporten nøye oss med å referere til kapitlene i den overordnete KI-strategien for Helse Nord (særlig kap. 6) der dette er mer utfyllende beskrevet. I tillegg har Helse Nord IKT ledet en konsept- og planfase om behovet for å opprette et nytt tjeneste- og fagområde i Helse Nord IKT for å understøtte både forskning og implementering av KI. Dette redegjøres kort for i kap. 12.

Tilfredsstillende informasjonssikkerhet er en forutsetning for å kunne ta i bruk KI-baserte beslutningsstøtteverktøy i Helse Nord. Innbyggere, pasienter og ansatte skal ha tillit til at Helse Nords sin behandling av helse- og personopplysninger gjennomføres med et forsvarlig sikkerhetsnivå. Vår anbefaling er at informasjonssikkerhetsaspektet ved KI inkluderes i det felles rammeverket for informasjonssikkerhet som nå er under gjennomføring i Helse Nord. I tillegg forventes det at den nasjonale Normen for informasjonssikkerhet utvides til å innbefatte KI-løsninger, og at denne legges til grunn i det videre arbeidet. Personvernforordningen krever at konfidensialitet, integritet, tilgjengelighet og robusthet skal være innebygget i løsningen. Innen KI for radiologi vil det være mest aktuelt at algoritmene kjøres i en ekstern skytjeneste. Dette stiller krav om anonymisering av data slik at det ikke er mulig å rekonstruere personsensitive opplysninger. Anonymisering av bilder tatt «inni kroppen» vil være vanskeligere å rekonstruere men må spesifikt adresseres i en ROS-analyse av det konkrete systemet. Ingen data skal lagres i den eksterne skyen. Det må også etableres retningslinjer om eierskap og rettigheter knyttet til bruk av bilder behandlet av algoritmen og om disse bildene kan brukes av selskapet for videre algoritmeutvikling.

Arbeidsgruppen anbefaler:

- Før KI-løsninger tas i bruk skal det gjennomføres en risiko og sårbarhetsanalyse, og iverksettes sikringstiltak som for andre systemer.

1.12 Regulatoriske endringer – fra MDD til MDR

I kapittel 13 er det redegjort for EU-kommisjonens regulatoriske grep ved å erstatte MDD (Medical Device Directive) med det nye MDR (Medical Device Regulative). I sistnevnte stilles det betydelig større krav til dokumentasjon og validering av løsningene, inkludert krav om at produsenten tar ansvar for oppdatering av produktets kvalitet i et livssyklusperspektiv. Det nye regulativet innebærer at nye løsninger for medisinsk utstyr og KI-algoritmer som skal inn på markedet etter mai 2022, må sertifiseres etter det nye regulativet, men innen mai 2024 må også tidligere MDD-sertifiserte produkter resertifiseres etter de nye krav i MDR. Det skjer også en endring i klassifiseringen av produkter etter risiko. Det får konsekvenser for et stort antall CE-merkede KI-produkter i radiologi som i henhold til MDD er sertifisert i klasse I ved selvdeklarasjon, men som i MDR må resertifiseres gjennom en samsvarsvurdering utført av eksternt Notified Body (klasse II). Det er i dette kapitlet redegjort nærmere for noen av de nye kravene som stilles i MDR.

Arbeidsgruppen anbefaler:

- At det ved anskaffelser av KI-algoritmer i Helse Nord stilles krav om at produktene er CE-merket i henhold til MDR før kontrakt signeres.

1.13 Helseøkonomiske analyser og implementeringsstudier

I kapittel 14.1 konstateres at det foreligger svært få HTA-analyser som sier noe om KI-løsningenes komparative effektivitet eller kost nytte i kliniske settinger. En fersk kunnskapsoppsummering av studier som tar for seg økonomisk evaluering av KI-algoritmer, identifiserte bare 20 studier som av tilstrekkelig kvalitet, vurdert etter kriteriene i det anerkjente helseøkonomiske rammeverket CHEERS; 9 av studiene var innen radiologi. I flertallet av disse studiene ble KI-løsningen vurdert som kostnadsbesparende, men det dreier seg om for få studier til å trekke generaliserende slutninger.

Det vises ellers til en kostnads-effektivitetsstudie av KI anvendt på deteksjon av karokklusjon ved akutte hjerneslag. Med empiri hentet fra det britiske hjerneslagregisteret, ble kvalitetsgevinstene mål ved QUALY (kvalitetsjusterte leveår). I sammenligningen med konvensjonelle arbeidsmåter, kom bruk av KI-algoritmen ut med en estimert besparelse på 11 millioner, dersom KI-løsningen ble tatt i bruk for denne pasientgruppen i hele Storbritannia.

Arbeidsgruppen anbefaler:

- At utprøving og implementering av KI-løsninger ledsages av egeninitiert helseøkonomisk følgeforskning
- At det etableres både nasjonal- og regional kompetanse på dette feltet, og at det organiseres et nasjonalt samarbeid om gjennomføring av følgeforskning ved innføring av KI i helsetjenesten.
- At det bør utvikles forenklede verktøy for kost-nytteanalyser – som en integrert del av Mini-HTA – og som kan tas i bruk av helseforetak og andre beslutningstakere i tilknytning til anskaffelser av KI-løsninger

I kapittel 14.2 tas utgangspunkt i den studien om implementering av KI-løsninger som Nasjonalt senter for ehelseforskning (NSE) nettopp har ferdigstilt. Denne studien er primært basert på intervjuer med en rekke aktører, både i Norge og utlandet, og på gjennomgang av aktuelle og relevante dokumenter og publikasjoner. Gjennom nevnte intervjuer, kom det fram at få virksomheter er kommet langt i implementeringen, og at en vesentlig del av de formidlede erfaringer er basert på pilotering og utprøving.

Basert på NSEs studie om implementering og andre kilder, anbefaler arbeidsgruppen at følgende hensyn legges til grunn ved implementering av KI-løsninger i Helse Nord:

- Forankring og støtte i både ledelse og fagmiljøer allerede fra starten av arbeidet
- Involvering av pasienter og pasientorganisasjoner i arbeidet
- Sørge for adekvat opplæring av det personell som skal delta i KI-implementering
- Tverrfaglig organisering av planleggings- og gjennomføringsprosessen
- Valg av algoritmer som er godt testet og validert, som har intuitive bruker-grensesnitt og som legger til rette for god arbeidsflyt
- Legge til rette for understøttende infrastruktur og god tilgang til relevante data

I kapittel 14.3 er det vist til det nasjonale koordineringsprosjektet for kunstig intelligens, ledet av Helsedirektoratet i samarbeid med andre direktorater, de regionale helseforetakene og KS. Vi har kort omtalt tre viktige dokumenter som er publisert i regi av dette prosjektet; det gjelder 1) *Tilrettelegging for bruk av kunstig intelligens i helsetjenesten - med utgangspunkt i det radiologiske fagområdet* (oktober 2021); *Kunstig intelligens i helsetjenesten – status og veien videre for det nasjonale koordineringsprosjektet* (oktober 2022); og *Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten (oktober 2022).*

Arbeidsgruppen anbefaler at:

- Personell og miljøer som skal planlegge, forberede eller ta i bruk KI i våre helseforetak setter seg inn i ovennevnte nasjonale rapporter.

1.14 Kompetansebygging og utdanning

Kapittel 15 dreier seg om utdannings- og kompetansebehov for å ta i bruk KI innen radiologifeltet. Det er både behov for å inkorporere kunnskap om KI i grunnutdanningen for radiologer og radiografer, og for å iverksette kompetanseutviklende tiltak overfor de mange som er ferdig med sin utdanning og er i aktiv tjeneste.

I kapitlet redegjøres for rapporten om *Styrking av utdannings- og opplæringstilbud innrettet mot kunstig intelligens i helsetjenesten,* i regi av UiT Norges arktiske universitet. Arbeidsgruppen vurderer forslagene i denne rapporten som noe passive, gitt de store kompetansebehovene som foreligger på dette feltet.

Arbeidsgruppen anbefaler:

- At kompetanse om KI bygges inn i grunnutdanningen både for radiologer og radiografer
- At personell som skal ta i bruk KI-løsninger ved røntgenavdelingene, gis adekvat opplæring i form av kurs (ordinære kurs og e-læringskurs)
- At det tas et regionalt ansvar for nødvendig kompetanseutvikling
- At det etableres et fagnettverk for KI - bestående av radiologer og radiografer – utgått fra det regionale forvaltningsforum for radiologi.

1.15 Validering av kvalitetssikring av K-verktøy

I kapittel 16 har vi samlet problemstillinger knyttet til valideringsbehov, målemetoder for validering og sammenligning av algoritmer med radiologer som komparenter.
Den viktigste begrunnelsen for å validere kommersielle radiologi-algoritmer før de tas i bruk, er at de er utviklet på grunnlag av data fra andre land, og – med få unntak – heller ikke validert mot norske pasientdata. Behovet for validering følger også av det faktum at en betydelig andel av de kommersielle algoritmene mangler vitenskapelig evidens i form av uavhengige fagfellevurderte studier. Valideringen bør skje på grunnlag av standardiserte og anerkjente metoder, og med både «ground truth» (evidens) og radiolog-komparenter som referanse.

Arbeidsgruppen anbefaler:

- I startfasen bør alle nye algoritmer som vurderes anskaffet, valideres mot retrospektive lokale pasientdata i dialogfasen av kvalifiseringsfase av anskaffelsen
- Validering bør gjennomføres på grunnlag av nasjonale, standardiserte retningslinjer
- Ved anskaffelser bør algoritmene testes mot både sensitivitet (andelen sanne positive) og spesifisitet (andelen sanne negative), og avveiningen mellom disse mål må tilpasses de respektive hensyn til både pasientsikkerhet og ressursbruk
- Algoritmenes kvalitet (treffsikkerhet) bør også vurderes og sammenlignes med gjennomsnittskvaliteten på radiologenes treffsikkerhet
- I den grad innkjøpsprosesser faller sammen i tid, bør det søkes samarbeid om validering med helseforetak utenfor Helse Nord.
- Fordi validering av stadig nye algoritmer vil kunne bli svært transaksjonskrevende, bør det etter hvert gjøres evaluering av behovet for validering av nye algoritmer mot egne pasientdata

1.16 Forvaltning av KI-verktøy

Innføring av kunstig intelligens krever god forvaltning av de ulike KI-løsningene. Løsningene som anskaffes vil være regionale, og forvaltningsansvaret bør ligge på et tverrfaglig regionalt forum bestående med representanter fra Helse Nord IKT, Felles funksjonell forvaltning av kliniske system og fagmiljøet. Det må foreligge klare funksjonsbeskrivelser samt beskrivelse av oppgaver mellom de ulike instansene i det tverrfaglige forumet.

De felles regionale kliniske IKT-løsningene forvaltes i dag i delt løsning mellom Felles funksjonell forvaltning kliniske system plassert i UNN HF og Helse Nord IKT (teknisk og merkantil forvaltning). Det er innenfor denne modellen etablert flere fora som Systemforvaltningsforum og arbeidsgrupper som kan inngå i en modell for forvaltning av KI-løsninger.

En viktig del av forvaltning er beslutningsmyndighet i forhold til for eksempel utvidelse av eksisterende løsninger eller oppgraderinger av disse. Dette innebærer ofte kostnader, ikke bare til leverandør, men også kostander tilknyttet oppgradering av nettverk med mer. Forvaltningsforumet må ha nødvendig innsikt og myndighet til å fatte beslutninger tilknyttet utvidelse av allerede eksisterende løsninger. Det er derfor særdeles viktig at deltakere i et slikt forum har tilstrekkelig med kompetanse og forståelse av hensikten med KI applikasjonen slik at man jobber mot et samstemt mål.

Del 2: Bakgrunn

Denne delen inneholder tre kapitler; først en innledning, deretter en beskrivelse av utviklingstrekk innen medisinsk radiologi og status for dette faget i Helse Nord, og til sist en beskrivelse av den radiologiske arbeidsflyten.

2 Innledning og bakgrunn

I dette innledende kapitlet beskrives bakgrunn, mandat, arbeidsform, mv.., for arbeidsgruppa som har vært med å utforme denne rapporten.

2.1 Bakgrunn og etablering av arbeidsgruppe

2.1.1 Bakgrunn

I Helse Nords strategi for kunstig intelligens, vedtatt i styremøte 22. juni 2021, ble det anbefalt å etablere en arbeidsgruppe med mandat til å utrede og konkretisere hvilke KI-løsninger som kan være aktuelle for implementering innen radiologifeltet.

2.1.2 Formål med arbeidet

Arbeidsgruppen skal utrede og fremme forslag om hvilke fagområder og modaliteter innen radiologien (inkludert KI-løsninger for doseplanlegging i tilknytning til stråleterapi og aktuelle KI-løsninger innen radiologi som anvendes innen persontilpasset medisin) som vurderes å ligge best til rette for implementering av KI-baserte løsninger som beslutningsstøtte i de enkelte HFene i Helse Nord RHF de nærmeste årene.

2.1.3 Omfang, føringer og krav til gjennomføringen

- I arbeidet med å identifisere løsningstyper som ligger til rette for implementering, må behovsvurderinger koples til områder der det er utviklet løsninger. En slik kopling forutsetter at arbeidsgruppens skaffer seg et mest mulig omforent og felles kunnskapsgrunnlag om hva KI kan bidra med, hvilke løsninger som foreligger og hvilke gevinster som kan høstes.
- Som et utgangspunkt, er det naturlig å starte med en forenklet kartlegging av aktuelt og fremtidig behov for radiologiske tjenester for befolkningen i Helse Nords nedslagsfelt. En slik kartlegging bør gjøres for ulike typer undersøkelser og ulike modaliteter (CT, MR, ultralyd mv.) og spesifiseres for hvert helseforetak. Denne kartleggingen vil i stor grad kunne gjenbruke og bygge på det arbeidet som skal gjennomføres i den regionale radiologiplanen for Helse Nord. Arbeidsgruppen bør identifisere områder som har tilstrekkelig volum til at KI-

løsninger kan bidra til å redusere arbeidsmengde og radiologitid per undersøkelse. Det bør også pekes på områder der det kan forventes at KI-løsninger kan bidra til å forbedre kvalitet og pasientsikkerhet.

- Kartleggingen av behov bør etterfølges av en kartlegging av tilbudet i markedet. Det bør primært søkes blant tilgjengelige CE-merkede/FDA-sertifiserte løsninger på markedet. I dette arbeidet bør det legges vekt på å identifisere områder der det foreligger et rimelig antall sertifiserte produkter. I denne søkeprosessen bør det også innhentes informasjon om volum og representativitet i det pasientunderlaget som er anvendt ved utvikling av løsningene.
- Når KI-løsninger på relevante fagområder er identifisert, bør det også innhentes erfaringer fra fagmiljøer som har tatt løsningene i bruk. Det er i denne sammenheng naturlig å innhente erfaringer fra de anskaffelses- og implementeringsprosesser som pågår i Vestre Viken HF, men slike erfaringer bør også innhentes fra fagmiljøer i Sverige og Danmark, eventuelt også andre land i Europa.
- Den overnevnte utredning og kartlegging skal munne ut i en anbefaling og konkretisering av hvilke typer KI-løsninger som kan være aktuelle for implementering innen radiologifeltet. I arbeidet bør informasjonen om tilbudet i markedet sammenholdes både med ovennevnte kartlegging av framtidige behov for radiologiske tjenester og kartlegging av erfaringer hos andre fagmiljøer. De delmengdene som er overlappende (løsningstyper med rimelig høyt volum på områder der det også er behov for økt kapasitet og kompetanse) gir et godt utgangspunkt for anbefaling av satsingsområder.
- Arbeidsgruppen bør også søke kontakt og samarbeid med Helse Sør-Øst som har fått i oppdrag å utrede et løp for anskaffelsesprosesser for medisinsk utstyr som inkluderer og kunstig intelligens. I omtalen av anskaffelsesprosesser forutsettes arbeidsgruppen å innarbeide de vurderinger som gjøres her.
- Det bør også utvikles en mal for ROS-analyser i samarbeid med ansvarlige for bygge på personvern og informasjonssikkerhet i de respektive helseforetak.

2.1.4 Avgrensing og avhengigheter

- I arbeidet bør kriterier som behov, nytte og løsningenes modenhet legges til grunn for å anbefale hvilke løsningstyper som bør prioriteres for implementering. De anbefalte løsningene bør bidra til å ivareta – og gjerne forbedre –kvalitet, pasientsikkerhet, og effektivitet.
- I samarbeid med fagmiljøene skal arbeidsgruppen også identifisere eventuelle hindringer for vellykket implementering, inkludert hensynet til god arbeidsflyt, nødvendig opplæring og kompetanseutvikling, samt harmonisering av undersøkelsesprotokoller og rekonstruksjoner.
- Arbeidsgruppen skal i sitt arbeid forholde seg til gjeldende regulatoriske begrensninger som eventuelt foreligger mht. å ta i bruk løsninger fra leverandører utenfor Europa (jf. utfordringer ved bruk av skytjenester). I sitt

arbeid bør derfor arbeidsgruppen ved behov søke kontakt og samarbeid med nasjonale regulatoriske og andre myndigheter (hvorav særlig Statens legemiddelverk, Helsedirektoratet, Direktoratet for e-helse). Det kan også være aktuelt å gjøre vurderinger av andre juridiske og etiske aspekter ved å ta i bruk slike løsninger.

- Arbeidsgruppen bør i sin gjennomgang av mulige produkter/produkttyper også ta i bruk aktuelle skriftlige kilder, i form av relevante tidsskrifter og nettsteder (eksemplifisert med aiforradiology.com) der ulike løsningstyper er forsøkt evaluert mht. kvalitet mv.
- Arbeidsgruppens anbefalinger kan danne grunnlag for en anskaffelsesprosess. Det bør derfor innhentes erfaringer fra andre norske og utvalgte europeiske (Sverige, Danmark, England og Nederland) miljøer som har gjennomført – eller er i ferd med å gjennomføre - tilsvarende prosesser på radiologifeltet.
- Arbeidsgruppen skal vurdere behovet for å validere anbefalte produkter mot norske data og hvordan disse valideringer tenkes gjennomført.
- Behov som identifiseres, men ikke møtes av eksisterende produkter, spilles inn til samarbeidende forskningsinstitusjoner med sikte på å vurdere potensialet for egenutvikling av løsninger.
- Parallelt med arbeidet vil det foregå en kartlegging av framtidige behov for radiologiske tjenester. Behov for KI-løsninger må avstemmes med dette arbeidet.
- Der det er mulig, bør det innhentes dokumentasjon om produktenes krav til infrastruktur og datafangst, og i hvilken grad slik programvare inngår i eller understøttes av produktet/leverandør av produktet. På dette punkt bør arbeidsgruppens vurderinger avstemmes mot HN IKTs utarbeidelse av konseptrapport om IKT-infrastruktur for KI.

2.1.5 Krav til resultat

Den ovennevnte utredning og kartlegging skal munne ut i en anbefaling og konkretisering av hvilke typer KI-løsninger som kan være aktuelle for eventuell anskaffelse og implementering innen radiologifeltet.

Utredningen innebærer ingen administrative og budsjettmessige konsekvenser som vil følge av en eventuell anskaffelsesprosess.
Utredningen leveres i form av en rapport til adm. direktør.

2.1.6 Tidsramme og ressursbehov

Arbeidet skal gjennomføres og rapporteres innen 1. juli 2022.
Arbeidet ledes fra Helsefagavdelingen ved fungerende forskningsdirektør.

2.1.7 Finansiering

Reise- og oppholdskostnader knyttet til arbeidet dekkes av den virksomhet medlemmene er ansatt i.
Senter for pasientnær kunstig intelligens, SPKI, ivaretar rollene som fag- og utredningsressurs ressurs på KI-feltet, med hovedansvar for utarbeidelse av rapport.

2.1.8 Organisering og ansvar

Rådgiver Finn Henry Hansen	Helse Nord RHF, Helsefagavd. (leder)
Bjørn-Petter Johannessen	LIS-lege radiologi, Finnmarkssykehuset, Hammerfest.
Ken-Inge Adolfsen	Radiograf UNN-Tromsø - deltok bare på første møte
Kurt Bøckman Gschib	Radiograf, UNN -Tromsø – øvrige møter
Tom-Vegard Markussen	Overlege radiologi, UNN-Tromsø
Heinrich Backmann	Overlege radiologi, Nordlandssykehuset, Bodø
Klaus Becker	Overlege radiologi, Helgelandssykehuset, Mosjøen
Line Horn Sivertsen	Radiograf Avd. leder, Nordlandssykehuset, Bodø
Randi Brendberg	Fagsjef, Helsefagavd., Helse Nord RHF
Arne Vassbotn	Regionalt Brukerutvalg Helse Nord
Martin Øien Jensen	Konserntillitsvalgt Akademikerne; Helse Nord

Fag- og utredningsansvarlig/sekretariat:
Rådgiver Finn Henry Hansen, Helse Nord RHF
Avdelingsleder Karl Øyvind Mikalsen, Senter for pasientnær kunstig intelligens (SPKI)
Sekretariats- og referatfunksjon ivaretas av seniorrådgiver Sture Pettersen, Helse Nord RHF

Arbeidet gjennomføres ved en kombinasjon av fysiske og digitale møter. Det legges til rette for digital deltagelse i samtlige møter.

2.1.9 Kommunikasjon og interessentoppfølging

Arbeidsgruppen vil utarbeide en interessentanalyse og kommunikasjonsplan som ivaretar øvrige interessenter.

2.1.10 Oppfølging av mandat og arbeidsform

Arbeidsgruppen har etter beste evne forsøkt å besvare de ulike punktene i mandatet, men erkjenner at det ikke har vært mulig å følge opp alle elementer i mandatets bestilling med samme grad av grundighet og dokumentasjon. Dette skyldes to forhold: 1) dels kapasitet i arbeidsgruppen, men 2) først og fremst har det vist seg krevende å fremskaffe dokumentasjon av erfaringer med gjennomførte implementeringsprosesser i andre land. Det henger i høy grad sammen med at ingen land er kommet langt i arbeidet med å ta i bruk KI-løsninger innen radiologi. Det foreligger derfor begrensede erfaringer å høste på dette stadium i feltets utvikling. KI-feltets dynamikk innen radiologien - og de mange utprøvende installasjoner som foregår - tilsier likevel, at dette forventes å endres

i løpet av de kommende år. Vi har derfor valgt å bygge implementeringskapitlet på den studien Nasjonalt senter for ehelseforskning (NSE) nylig har gjennomført. For øvrig bør de vurderinger og anbefalinger som gjøres i dette dokumentet, derfor suppleres gjennom etablering av en løpende følge-med-funksjon for oppdatering av nye erfaringer og ny kunnskap. I Helse Nord vil det være naturlig at SPKI ivaretar denne funksjonen.

I mandats punkt 2.1.3, kulepunkt 6, heter det at arbeidsgruppen i sin rapport skal innarbeide resultatene fra prosjekt i regi av Helse Sør-Øst RHF om anskaffelsesprosedyrer ved innkjøp av medisinsk utstyr som inneholder kunstig intelligens. Dette prosjektet ble av ulike grunner ikke gjennomført, med den konsekvens at dette punkt i mandatet ikke kan ivaretas.

Arbeidsgruppen har hatt 11 møter, hvorav hele 10 på Teams; det har altså bare vært mulig å samle arbeidsgruppens medlemmer til ett felles fysisk møte. Siden det for de fleste medlemmene dreier seg om nye og krevende problemstillinger, ville flere fysiske møter kanskje bidratt til en raskere progresjon i arbeidet. Mellom møtene i arbeidsgruppen har sekretariatet hatt utstrakt kontakt med ulike fagmiljøer og leverandører innen feltet, både nasjonale og internasjonale. Vi vil her særlig framheve kontakten med nettstedet AIforradiology i Nederland som både er vår viktigste datakilde til informasjon om tilgjengelige produkter i markedet, og som generelt har gitt oss viktig og relevant informasjon om KI innen radiologifeltet.

Det ble i mandatet satt en frist for ferdigstilling til 1. juli 2022[2]. Denne fristen er i forståelse med ledelsen i Helse Nord RHF utsatt til ultimo oktober 2022. Denne utsettelsen er begrunnet med nødvendigheten av å innarbeide de viktige erfaringene fra KI-anskaffelsen i radiologi i Vestre Viken. Opprinnelig skulle denne anskaffelsen være avsluttet ved årsskiftet 2021-2022, men kontraktarbeidet først ble avsluttet ultimo juni 2022 og kontrakten med leverandørene først ble undertegnet 19. august 2022.

[2] Ferdigstilling av rapporten ble ikke avsluttet før 15. november, som følge av nødvendig korrektur og kvalitetssikring

3 Hvorfor radiologi vurderes som et prioritert område for å ta i bruk KI-løsninger

3.1 Utviklingstrekk innen medisinsk radiologi[3]

Det er flere grunner til at radiologien som fagfelt har fått økt betydning innen helse-tjenesten; 1) Ny og forbedret teknologi innebærer at det gjøres flere og mer komplekse undersøkelser, men radiologiske metoder anvendes i økende utstrekning også til behandlingsformål (intervensjons-radiologi) og for å evaluere effekter av behandling.

Volumvekst: Siden 2008 har det ikke vært laget noen samlende oversikt over volum og typer av radiologiske undersøkelser i Norge. Nye systemer for dokumentasjon av aktivitet har lenge vært under planlegging. Det startet med innføring av et nytt kodeverk (NCRP)[4] i 2012, og bildediagnostikkmeldingen for innrapportering til NPR ble ferdigstilt i november 2016. I Helsedirektoratets strategi for rasjonell bruk av bildediagnostikk[5] (januar 2019) ble det annonsert at direktoratet skulle innhente aktivitets- og dosedata fra bildediagnostikk og stråleterapi i løpet av 2019 med sikte på rapportering fra og med 2020. Sommeren 2022 er slike løsninger fremdeles ikke på plass, og ifølge Helsedirektoratet vil rapportering av nasjonale aktivitetsdata tidigst kunne implementeres i løpet av 2023. Foreløpig har vi derfor ikke norske data som kan si noe om utviklingstrekkene de siste årene, verken totalt eller for de ulike modalitetene. Derimot har Danmark slike oversikter, og det er i utgangspunktet liten grunn til å anta at utviklingen i Norge avviker vesentlig fra den i Danmark. Nedenstående tabell viser utviklingen i Danmark for tiårs/perioden fra 2011 til 2021.

Tabell 3.1: Endring i antallet radiologiske undersøkelser i Danmark 2011-2021[6].

Proceduretype	2011	2021	Endring
UXA - Angiografier	45 645	39 060	-14 %
UXC - CT-skanninger	648 020	1 176 230	82 %
UXM - MR-skanninger	343 105	555 250	62 %
UXR - Røntgenundersøgelser	2 440 530	2 446 150	0 %
UXU - Ultralyds-undersøgelser	1 409 190	1 888 285	34 %
Totalsum	**4 886 490**	**6 104 975**	**25 %**

Vi ser av tabellen at det totale antallet undersøkelser har økt med 25 prosent i løpet av denne tiårsperioden, men at utviklingen er svært forskjellig for ulike modaliteter. Man

[3] Ingfrid Haldorsen og særlig Njål Lura, Helse Bergen, har gitt viktige innspill til hvordan radiologiske undersøkelser i løpet av det siste tiåret i økende utstrekning er tatt i bruk for nye diagnosegrupper. Leder av Norsk radiologisk forening, Helga Brøgger, har også gitt nyttige innspill til dette kapitel.
[4] NCRP= NOMESKO Classification of Radiology Procedures. Det er for øvrig en utbredt oppfatning I det radiologiske fagmiljøet at dette nye kodeverket er for grovt til å fange inn det faglige innholdet I radiologenes virksomhet. Det vurderes dessuten som lite egnet til å fange inn aktivitet og arbeidsmengde ved en gitt undersøkelse.
[5] Strategi for rasjonell bruk av bildediagnostikk. Forslag fra Helsedirektoratet 1. februar 2019.
[6] Kilde: https://www.esundhed.dk/Emner/Operationer-og-diagnoser/Radiologiske-undersoegelser

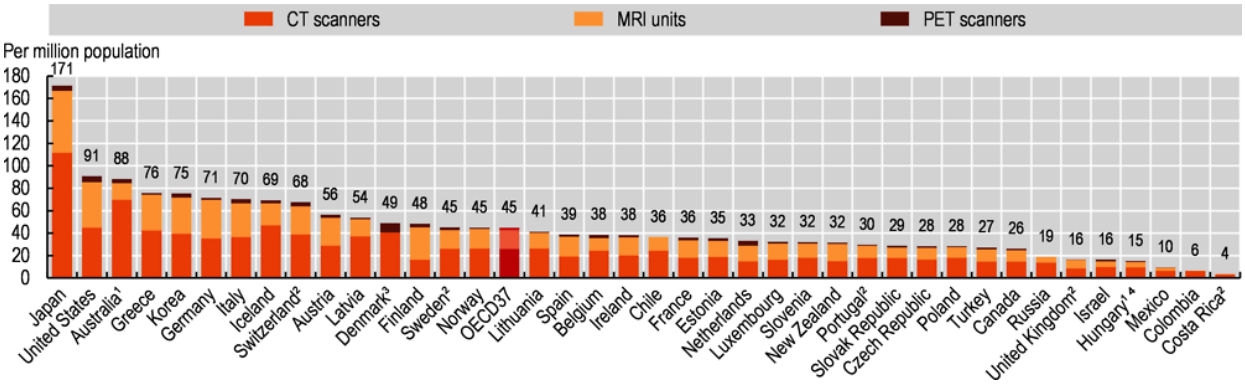
ser at antallet angiografier er redusert med 14 prosent, noe som kan ha sammenheng med redusert forekomst av hjerte-karsykdommer, men som kanskje i enda større grad er uttrykk for at CT-angiografi og MR-angiografi har erstattet konvensjonell diagnostisk angiografi. Det innebærer samtidig en større andel terapeutiske prosedyrer som er mer tidkrevende. Videre ser vi at antallet undersøkelser innen konvensjonell røntgen er uforandret i løpet av perioden, mens det er en betydelig vekst innen andre modaliteter; størst er veksten innen CT (82%), men den er også betydelig innen MR (62%) og ultralyd (34%). Tilsvarende utviklingstrekk for CT og ultralyd er dokumentert i en polsk studie som dekker samme periode[7].

At MR- og CT-undersøkelser i løpet av det siste 10-året tas mer regelmessig i bruk for flere diagnoser - både for diagnostikk og for å evaluere virkninger av behandlinger – er utvilsomt en av forklaringene på volumveksten. Radiologisk diagnostikk har fått en mer sentral rolle i primærutredning av de fleste onkologiske tilstander. Både CT og MR brukes for å kartlegge lokalutbredelse og å undersøke for eventuell metastatisk kreftsykdom. Positron emisjon tomografi (PET) har fått en spesielt viktig rolle for å utrede for metastatisk sykdom ved flere kreftsykdommer: Lungecancer, pankreascancer og cervix-cancer er eksempler på dette der tilstedeværelse av fjernmetastaser er helt avgjørende av behandlingsvalg. MR har fått en stadig mer sentral rolle i å redegjøre for lokalutbredelse av flere kreftsykdommer som rektum-cancer, cervix cancer/ endometriecancer, samt prostatacancer er eksempler på dette. MR er overlegent beste modalitet for å kartlegge både metastaser og primære neoplasmer i sentralnerve-systemet. MR har i økende grad også blitt tatt i bruk for å utrede kognitiv svikt og mistanke om ulike nevrodegenerative tilstander. I behandlingen av flere kreftsykdommer, har man også fått en økning av medikamentelle målrettede behandlingsmuligheter i form av enzymhemmere og immunterapi; dette er det flere eksempler på i behandling av lungekreft, ulike hematologiske kreftformer og melanom. Disse behandlingene er ofte kostbare og behandlingsformen krever hyppig oppfølging for å evaluere behandlingsrespons, samt den ønskede effekten, at flere pasienter lever vesentlig lengre. Disse forholdene fører til en vesentlig økning av antall undersøkelser pr pasient.

Prognosene for fremtiden tyder på at trenden med økende radiologisk diagnostikk vil fortsette. Dette vil antagelig drives av: 1) Teknologiske fremskritt og nye definerte bruksområder for modalitetene, for eksempel ulike MR-sekvenser; 2) En økende gjennomsnittsalder i befolkningen og dermed økende prevalens av kreftsyke og hjerte/kar-syke som trenger diagnostikk, behandling og oppfølging; 3) Screening-programmer – eksemplifisert med lunge-cancer screening - kan bli innført, noe som vil føre til en massiv økning i antall undersøkelser.

[7] Mateusz Winder et al.: Are We Overdoing It? Changes in Diagnostic Imaging Workload during the Years 2010-2020 including the Impact the Saars-CoV-2 Pandemic, *Health Care* 2021, 9, 1557, https://doi.org/10.3390/healthcare9111557

Internasjonale sammenligninger av forekomsten av radiologiske undersøkelser, viser at Norge ligger omtrent på gjennomsnittet i OECD-området, og med betydelig lavere forbruk av slike undersøkelser enn de land som ligger på topp: Japan, USA og Australia[8].



Figur 1: Forbruk av radiologiske undersøkelser innen utvalgte modaliteter i OECD-området i 2019.

Flere og mer komplekse bilder per undersøkelse

Veksten i antallet undersøkelser fanger imidlertid ikke opp tilsvarende vekst i arbeidsmengde, ettersom det særlig innen CT og MR tas langt flere og mer komplekse bilder per undersøkelse, noe som innebærer at radiologene bruker betydelig mer tid på å tolke og beskrive bildene fra disse undersøkelsene[9]:

> «*Present-day devices used in radiology provide higher spatial and contrast resolution of the visualized tissues. The greater detail of imaging tests allows for more accurate detection of diseases but requires greater accuracy from the radiologist and more time for analysis. The growing demand for imaging examinations, especially in oncology, is an important issue regarding the time that the radiologist must devote to describing the examination. Assessing disease response to treatment requires a comparison of the current and the previous studies, which in practice means re-evaluating several images at the same time*»,

Etablering av pakkeforløp for flere diagnosegrupper (særlig innen kreft), samt større fokus på ventetider og fristbrudd, har medført at radiologiske undersøkelser er blitt mer tidskritiske for logistikken i pasientforløpet. Dermed blir radiologene eksponert for mer krevende prioriteringer for å kunne overholde tidsfristene i pakkeforløpene. Radiologenes deltagelse i multidisiplinære teammøter (MDT-møter) knyttet til pakkeforløpene for kreft, medfører ikke bare betydelig tidsforbruk i selve møtene, men også tid til forberedelse og etterarbeid.

Samlet sett har den økende etterspørselen etter radiologiske tjenester ført til en global mangel på radiologer. Dermed har også de individuelle arbeidsbelastninger for radiologene tiltatt, med den konsekvens at det internasjonalt rapporteres økende

[8]StatLink https://stat.link/lgfjst
[9] Mateusz Wunder et al. Op.cit.:

forekomst av utbrenthet blant radiologer[10] [11]. I tillegg tar arbeidsoppgaver som ikke er direkte relatert til bildegransking en stadig større andel av arbeidstiden til radiologene. Her kan blant annet nevnes prioritering av stadig flere henvisninger, arbeidsprosesser knyttet til talegjenkjenning og tidkrevende korrekturlesning. Både arbeidsmengden i seg selv og den ledsagende risikoen for å gjøre feil som følge av overbelastning, bidrar til utbrenthet blant radiologer.

Den sterke volumveksten i radiologiske tjenester et imidlertid ikke bare drevet av teknologiske og faglige fremskritt. Den er også til dels drevet av svakt begrunnede henvisninger fra både sykehusleger og fastleger[12]; sistnevnte ofte under sterkt press fra nærmest «selvrekvirerende» pasienter. Fastlegene må dessuten forholde seg til et stadig mer teknologisk komplekst assortiment av diagnostiske muligheter og har dermed en svært krevende portvaktfunksjon. Trenden i retning av mer overdiagnostikk skal derfor på ingen måte belastes fastlegene alene.

Private forsikringsordninger[13] og private kommersielle aktører bidrar i høy grad også til den økende forekomst av «for sikkerhets-skyld-medisinen». Rekvirering og gjennomføring av radiologiske undersøkelser på sviktende indikasjon, er ikke bare et problem for helsetjenestens kapasitet og overbelastning av radiologene, men også for pasientene som eksponeres for unødvendig og risikofylt stråling. Som ledd i kampanjen Kloke valg, har derfor Norsk radiologisk forening i 2018 formulert en anbefaling i flere punkter for å redusere risikoen for overdiagnostikk[14].

I den forannevnte strategirapporten fra Helsedirektoratet, ble det skissert flere forslag om tiltak for mer rasjonell bruk av bildediagnostikk og for å redusere uønsket variasjon mellom ulike virksomheter og geografiske områder.[15] Her nevnes tre eksempler på slike tiltak:

- Etablering av felles bildelager, både for å unngå dobbeltundersøkelser og unødvendig strålebelastning, og for å få enklere tilgang til tidligere undersøkelser
- Interaktive henvisninger basert på omforente henvisningskriterier forankret i felles retningslinjer og beslutningsstøttesystemer
- Rutiner for analyser av aktivitetsdata for å kunne fange inn uønsket variasjon.

Den økende etterspørselen etter radiologiske tjenester og den ledsagende økte arbeidsbelastningen for radiologer, har medført betydelige rekrutteringsutfordringer for dette fagfeltet. Denne utfordringen er nå nasjonal, men med de vanlige geografiske

---

[10] Bruls, R.J.M.; Kwee, R.M. Workload for radiologists during on-call hours. Dramatic increase in the past 15 years. *Insight Imaging* 2020,11,121

[11] Sinsky, Harry E. et al: Physician Task Load and the Risk of Burnout Among US Phycisians in a National Survey Jt. Comm Jt Qual.Patient Saf 2021, 47,76-85

[12] Strategi for rasjonell bruk av bildediagnostikk

[13] Ca. 650 000 personer i Norge dekkes nå av private helseforsikringer

[14] www.legeforeningen.no/foreningsledd/fagmed/norsk-radiologisk-forening/artikler/foreningsstoff-fra-norforum/gjor-kloke-valg-bestill-plakater/

[15] Helsedirektoratet: Strategi for rasjonell bruk av bildediagnostikk 2019, side 36-37.

og funksjonelle gradientene, med større utfordringer i Nord-Norge og ved de små sykehusene. Ettersom det radiologiske fagfeltet blir stadig mer (sub)-spesialisert, blir det mer krevende å både utvikle- og opprettholde generalistkompetanse, med den konsekvens at det blir vanskeligere å rekruttere og holde på radiologer ved de små enhetene.

3.2 Kunstig intelligens som ett av flere virkemidler for å avhjelpe kapasitets- og kvalitetsutfordringene innen radiologien

De utviklingstrekk det er redegjort ovenfor, viser at radiologene har økende problemer med å håndtere den store økningen i arbeidsmengden. Det trengs flere avbøtende tiltak for å møte denne utfordringen, samtidig som det er viktig å ta i bruk nye og bedre diagnostiske metoder. Gitt denne situasjonen, er antagelig radiologifeltet et av de områdene innen medisinen som vil kunne ha størst nytte av nye metoder som kan avlaste, understøtte og kvalitetssikre radiologenes arbeid.

Mot denne bakgrunn, kan det derfor med en viss, om enn behersket optimisme konstateres, at radiologien er det medisinske fagfeltet som internasjonalt er kommet lengst i utviklingen av KI-algoritmer. Status per oktober 2022 er at hele 70 % av de i alt 521 FDA-klarerte KI-algoritmene fins innen det radiologiske fagområdet[16]; også blant CE-merkede KI-løsninger er dette fagområder dominerende med i alt 204 produkter. I alt fins det i dag ca. 200 slike kommersielle KI-løsninger innen radiologi[17]. Selv om det foreløpig er begrenset storskala implementeringer av slike løsninger, er det utvilsomt dette fagfeltet som i dag ligger best til rette for utprøving, validering og implementering av KI-løsninger. Dette er med andre ord det område der behov og potensielt tilbud vurderes som mest sammenfallende, noe som også er begrunnelsen for at Helse Nord i sin vedtatte KI-strategi har prioritert dette fagområdet for oppfølging og videre utredning. Iverksatt på en kontrollert og kvalitetssikret måte, er det arbeidsgruppens vurdering at KI-løsninger innen radiologi kan bidra med avlastende beslutningsstøtte, både for å håndtere volum, sikre og gjerne øke gjennomsnittskvaliteten på undersøkelsene. Selv om majoriteten av dagens KI-løsninger har diagnostikk og bildetolkning som formål, ser vi også potensialer for å ta i bruk KI for å forbedre logistikk og arbeidsflyt og for å prioritere henvisninger.

Arbeidsgruppen vil samtidig advare mot forestillinger om at KI alene kan løse kapasitets- og bemanningskrisen innen radiologifaget. Det vil fortsatt være nødvendig å satse på å styrke bemanningen, både gjennom flere LIS-stillinger (spesialistutdanning

---

[16] https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices?utm_source=The+Medical+Futurist+Newsletter&utm_campaign=528783dc60-EMAIL_CAMPAIGN_2021_9_28&utm_medium=email&utm_term=0_efd6a3cd08-528783dc60-420597065&mc_cid=528783dc60&mc_eid=765c84ee1d

[17] https://aiforradiology.com

for leger), bedre tilrettelagte utdanningsløp og flere overlegestillinger. Ved flere sykehus vil det også være nødvendig å styrke lab-kapasiteten, ettersom denne i betydelig grad også er en flaskehals i håndteringen av den økende arbeidsmengden; dette gjelder også enkelte sykehus i Nord-Norge. Andre utfordringer for den radiologiske virksomheten i Helse Nord, knytter seg til organisering, samarbeid og virksomhetsstyring innen dagens arbeidsmetodikk. Disse utfordringene, som omtales i neste avsnitt, må også adresseres for å kunne skape det organisatoriske «slack»[18] (frigjorte ressurser) som vil være nødvendig for å kunne ta i bruk maskinlæringsløsninger innen dette fagfelt.

3.3 Særskilte utfordringer for den radiologiske virksomheten i Helse Nord

Som del av et større oppdrag, gjorde Helse Nord RHF i 2021 en bestilling til konsulenthuset Karabin om å kartlegge status for radiologifeltet i foretaksgruppen, med særlig fokus på rekruttering og samarbeid på tvers. Kartleggingen skulle gi en bedre forståelse av både nåsituasjonen og framtidige behov, og peke på muligheter og tiltak som kunne bidra til forbedringer. Allerede før denne kartleggingen ble igangsatt, betegnet fagmiljøene situasjonen for radiologifeltet i nord som kritisk[19], noe som langt på vei bekreftes i den nylig avleverte rapporten om status og anbefalte tiltak[20]. De viktigste data som ligger til grunn for denne kartleggingen er innhentet gjennom en spørreundersøkelse og en rekke dybdeintervjuer med nøkkelpersoner (klinikksjefer, avdelingsledere, radiologer m.fl.).

I statusbeskrivelsen pekes det på en rekke utfordringer:

- Det er mangel på radiologer i alle helseforetak, høy turnover-rate (andel radiologer som slutter hvert år), og vanskelig å beholde nyutdannede og rekruttere nye radiologer.
- Mangelen på radiologer vurderes som langt større enn antallet ledige stillinger, og UNN HF og Nordlandssykehuset HF har vurdert underdekningen å tilsvare et minimumsbehov på 12 nye-, og et mer optimalt behov for 30 nye radiologer.

Rapporten peker også på sviktende organisering, planlegging og virksomhetsstyring som viktige deler av det samlede utfordringsbildet. Røntgenavdelingene i hvert av helseforetakene karakteriseres i rapporten som:

- Isolerte siloer, med liten samhandling på tvers av HFer, mangel på felles protokoller og med innbyrdes svært ulik håndtering av arbeidslister.

[18] Cyert, RM and March, JG: A Behavioral Theory of the Firm, John Wiley and Sons Ltd 1963
[19] Bekymringsmelding fra mai 2021 signert røntgenavdelingene i alle helseforetakene.
[20] Kartlegging av radiologi- og operasjonskapasitet i Helse Nord. Rapport med anbefalinger. Karabin 1. juli 2022.

I rapporten anbefales iverksatt sju tiltak som kan bidra til å øke utnyttelsen av kapasitet og kompetanse innen radiologifeltet. De to første av følgende tiltak foreslås iverksatt allerede høsten 2022:

- Etablere en koordinator med ansvar for etablering av prinsipper for planlegging av henvisninger og kapasitet innen radiologfeltet
- Sentralisere rekrutteringen innen radiologi for hele Helse Nord, noe som innebærer å etablere en sentral HR-funksjon for rekruttering av både LIS-leger og overleger
- Gjennomgang av avtalen med private aktører for å forbedre kvaliteten i arbeidet de leverer til helseforetakene, samt etablere systematikk for tilbakemelding.
- Utarbeide standardiserte pasientforløp for radiologiske undersøkelser
- Opprette et felles fagmiljø for radiologer i Helse Nord
- Innføre muligheten for å gjennomføre LIS-spesialiseringen innen Helse Nord
- Kartlegge muligheten for å opprette flere radiologihjemler til UNN-Tromsø og Nordlandssykehuset-Bodø.

Styret i Helse Nord RHF fikk forelagt rapporten i styremøte 28. september og fattet følgende vedtak:

1. Styret i Helse Nord RHF tar informasjonen om *Kartlegging av radiologi og operasjonsstuekapasitet* til orientering.
2. Styret ber adm. direktør komme tilbake med forslag om etablering av regionalt prosjekt innen desember 2022.

Siden de problemstillinger som adresseres i denne rapporten har relevante berøringspunkter til arbeidet med å forberede bruk av KI innen radiologien, har vi funnet det naturlig å inkludere - og kort kommentere - konklusjonene, også i vår rapport.

Det kan ikke stikkes under stol at flere av fagfolkene i vår arbeidsgruppe er kritiske til deler av rapporten. Blant annet savnes en mer differensiert vurdering av de betydelige forskjeller i volum og arbeidsoppgaver mellom de ulike helseforetakene, noe som også tilsier ulike organisatoriske løsninger for henholdsvis store spesialiserte sykehus og de mange og relativt små lokalsykehus. Det påpekte behovet for større grad av regional samordning av virksomhetene på tvers av helseforetakene får tilslutning i arbeids-gruppen, men det reises spørsmål om hvorvidt forslagene om å etablere en regional koordinatorfunksjon er det rette grepet for å møte utfordringene. I beskrivelsen av de forhold som har økt arbeidsbelastningen for radiologer, er ellers det omfattende og tidkrevende arbeidet med å delta i MDT-møter overhodet ikke er omtalt. Enkelte medlemmer i arbeidsgruppen peker også på en mulig sammenheng mellom bemanningsutfordringene og manglende (kapasitet) til planlegging og styring. Karabinrapporten er også vurdert av fagsjefene i helseforetakene som i rimelig høy grad deler situasjonsbeskrivelsen, men har til dels svært ulike vurderinger av de foreslåtte

virkemidlene; bl.a. er det tydelig uenighet om forslaget om å etablere en funksjon som koordinator på tvers av helseforetakene, og hvordan en eventuell slik funksjon i så fall skulle organiseres og hvor den skulle lokaliseres.

Arbeidsgruppen vurderer det likevel som både positivt og nødvendig at ledelsen i Helse Nord RHF har adressert den kritiske situasjonen innen radiologifaget og støtter forslaget om at det etableres et regionalt prosjekt for å utvikle nødvendige forbedringstiltak. Samtidig er det blant medlemmene uttrykt en forventning om at ledelsen i Helse Nord RHF i utformingen av mandat for det regionale prosjektet tar høyde for noen av de mangler og begrensninger som preger Karabin-rapporten.

Kunstig intelligens er så vidt nevnt i rapporten, men inngår ikke i de tiltak som foreslås. Karabin-rapportens sterke fokus på den betydelige kapasitetsutfordringene innen radiologen, aktualiserer likevel behovet for også å ta i bruk teknologiske virkemidler som avlastende beslutningsstøtte. Dernest redegjør rapporten for betydelige logistiske utfordringer knyttet til planlegging, arbeidsflyt og samarbeid; dette er problemstillinger der også KI-algoritmer etter hvert kan tenkes å være et av verktøyene i forbedringsarbeidet.

Arbeidsgruppen har forståelse for at det regionale radiologiprosjektet som planlegges etablert, vil ha hovedfokus på kapasitetsutfordringer, utdanning og rekruttering, planlegging, organisering og ledelse. På disse områdene trengs både kortsiktige og langsiktige tiltak. Forhåpentlig vil slike tiltak på lengre sikt kunne bidra til å skape nødvendig rom og kapasitet for at radiologene skal motiveres til å delta i utprøving og implementering av KI-algoritmer. Slike omfattende endringer lar seg imidlertid ikke gjennomføre over natten; de vil nødvendigvis ta tid, antagelig flere år.

Arbeidsgruppen vil anbefale at Helse Nord RHF i mandatet for det regionale radiologiprosjektet også inkluderer problemstillinger om å ta i bruk KI-algoritmer. Når det gjelder innføring av KI innen radiologien, er det på kort sikt neppe de store gevinster å høste. men utprøving i småskala er antagelig en forutsetning for å kunne ta de neste stegene. Det innebærer at vi allerede nå må investere i kompetanse på dette feltet, styrke infrastrukturen for sikker og kvalitetssikret datafangst og utvikle systemer for validering av løsninger. Arbeidsgruppen vil derfor anbefale at slike tiltak inngår i en samlet tiltaksplan for å styrke kapasitet, kvalitet og effektivitet innen den radiologiske virksomheten i Helse Nord. Gitt at forutsetningene legges til rette, vil gradvis innføring av KI-løsninger, også kunne bidra til å gjøre radiologifeltet mer interessant for unge leger, og dermed indirekte også fungere rekrutterende. I dagens ressurssituasjon for røntgenavdelingene innebærer det at startfasen for utprøving og implementering av KI-løsninger, må prosjektorganiseres og tilføres dedikerte ressurser.

3.4 Status for bemanning av radiologer ved helseforetakene i Helse Nord per august 2022

I samarbeid med helseforetakene og HR-seksjonen i Helse Nord RHF, er det laget en oppdatert oversikt over personellsituasjonen for radiologer ved helseforetakene i Helse Nord.

Tabell 3.4.1 Antallet budsjetterte og besatte årsverk innen radiologi per august 2022 fordelt på helseforetakene i Helse Nord

	FIN	%	UNN	%	NLSH	%	HSYK	%	Helse Nord totalt
Overleger									
Budsjetterte årsverk	10	13 %	43,2	55 %	16	20 %	10	13 %	79,2
Besatte årsverk, fastansatte	7,3		38,8		6,45		7,5		60,05
Besatte årsverk, vikar/byrå	1,1		4		5		2		12,1
Ubesatte årsverk	1,6		0,4		4,55		0,5		7,05
Behov			49,5		29				
Reelt ubesatt			6,7		17,55				
Leger i spesialisering									
Budsjetterte årsverk	4	15 %	11	41 %	9	33 %	3	11 %	27
Besatte årsverk	2		15		9		2		28
Ubesatte årsverk	2				0		3		5
Behov			18						
Reelt ubesatt			3						

Tabell 3.4.1 ovenfor viser at det budsjettert nesten 80 overlegestillinger ved helseforetakene i Helse Nord, men at bare ca. 75 prosent av stillingene er fast besatt. 12 av de «ledige» overlegestillingene er besatt av vikarer, inkludert bruk av vikarbyråer, mens 7 av stillingene er helt ubesatte. Samlet er det budsjettert 26 årsverk for LIS-stillinger innen radiologi; her er det oppgitt at 5 stillinger ubesatte, hvorav alle ved lokalsykehusene i Finnmark og på Helgeland. Samtidig har UNN 4 flere LiS-ansatte enn det er budsjettert med.

Tabell 3.4.2 Antallet budsjetterte og besatte årsverk innen radiologi per august 2022 fordelt på spesialiserte sykehus (UNN-Tromsø og NLSH-Bodø) og de 9 lokalsykehusene i Helse Nord

	UNN Tromsø	%	NLSH Bodø	%	9 lokalsykehus	%	Helse Nord
Overleger							
Budsjetterte årsverk	33,2	42 %	13,0	16 %	33,0	42 %	79,2
Besatte årsverk, fastansatte	28,8		6,5		24,8		60,05
Besatte årsverk, vikar/byrå	4,0		3,0		5,1		12,1
Ubesatte årsverk	0,4		3,6		3,1		7,05
Leger i spesialisering							
Budsjetterte årsverk	10,0	37 %	9,0	33 %	7,0	26 %	26
Besatte årsverk	14,0		9,0		-1,5		21,5
Ubesatte årsverk	-4,0		-		8,5		4,5

I tabell 3.4.2 er budsjetterte og ledige årsverk fordelt på de to spesialiserte sykehusene og de 9 lokalsykehusene i Helse Nord. Her ser vi at UNN-Tromsø har hele 42 % av de budsjetterte årsverkene for overleger, den samme andelen som de 9 lokalsykehusene Avviket mellom budsjetterte og fast besatte stillinger gir imidlertid et gunstigere bilde av bemanningssituasjonen enn det som er realiteten. Som følge av høy turnover og store problemer med å få søkere til utlyste stillinger, har sykehusene unnlatt å budsjettere med det antall års verk som gjenspeiler behovene. Det gjelder særlig de mer spesialiserte sykehusene UNN-Tromsø og NLSH-Bodø.

I realiteten er derfor utfordringen enda større, ettersom HFene – og særlig avdelingene i Tromsø og Bodø – har signalisert langt større behov for radiologer enn det som fremgår av budsjetterte stillinger.

3.5 Radiologiske undersøkelser for ulike modaliteter i 2021 – fordelt på de fire helseforetakene og private aktører.

Tabell 3.5.1 Antall radiologiske undersøkelser i Helse Nord 2021 fordelt på de fire helseforetakene og private aktører. Absolutte tall og prosentandel.

Antall undersøkelser fordelt på utførende foretak												
Antall	Helse nord totalt		Foretak									
Modalitet	Antall	Andel	FIN	Andel av total	UNN	Andel av total	NLSH	Andel av total	HSYK	Andel av total	Private	Andel av total
CT	54 983	100 %	8 280	15 %	17 498	32 %	15 287	28 %	10 196	19 %	3 722	7 %
CT ANG	8 220	100 %	1 099	13 %	3 360	41 %	1 996	24 %	1 545	19 %	220	3 %
MR	66 308	100 %	8 358	13 %	9 068	14 %	7 796	12 %	9 062	14 %	32 024	48 %
PET	1 147	100 %		0 %	1 147	100 %		0 %		0 %		0 %
RG	157 192	100 %	20 970	13 %	61 546	39 %	37 953	24 %	24 425	16 %	12 298	8 %
UL	34 006	100 %	4 320	13 %	13 030	38 %	8 897	26 %	4 985	15 %	2 774	8 %
Annet	39 182	100 %	5 081	13 %	12 267	31 %	10 253	26 %	5 707	15 %	5 874	15 %
Totalsum	**361 038**	**100 %**	**48 108**	**13 %**	**117 916**	**33 %**	**82 182**	**23 %**	**55 920**	**15 %**	**56 912**	**16 %**

Vi ser av tabell 3.5.1 at UNN HF utfører 33% av de vel 360 000 radiologiske undersøkelser som utføres i Helse Nord, mens Nordlandssykehuset HF utfører 24 % og de to lokalsykehusforetakene henholdsvis 13 % (Finnmarkssykehuset HF) og 15 % (Helgelandssykehuset HF). En betydelig del av undersøkelsene (16 %) utføres av private aktører som har avtale med Helse Nord, og ca. halvparten av alle MR-undersøkelsene utføres av disse private aktørene.

Tabell 3.5.2 Antall radiologiske undersøkelser i Helse Nord i 2021 for ulike modaliteter fordelt på spesialiserte og mindre spesialiserte sykehus. Absolutte tall og prosentandel

Antall undersøkelser fordelt på utførende enhet										
	Helse nord totalt		Spesialisert				Lokal sykehus		Private	
			UNN Tromsø		NLSH Bodø					
Modalitet	Antall	Andel	Antall	Andel av total	Antall	Andel av total	Antall	Andel av total	Antall	Andel av total
CT	54 983	100 %	10 376	19 %	9 128	17 %	31 757	58 %	3 722	7 %
CT ANG	8 220	100 %	2 306	28 %	1 036	13 %	4 658	57 %	220	3 %
MR	66 308	100 %	5 854	9 %	3 281	5 %	25 149	38 %	32 024	48 %
PET	1 147	100 %	1 147	100 %		0 %		0 %		0 %
RG	157 192	100 %	33 262	21 %	22 874	15 %	88 758	56 %	12 298	8 %
UL	34 006	100 %	8 832	26 %	6 088	18 %	16 312	48 %	2 774	8 %
Annet	39 182	100 %	7 258	19 %	7 033	18 %	19 017	49 %	5 874	15 %
Totalsum	**361 038**	**100 %**	**69 035**	**19 %**	**49 440**	**14 %**	**185 651**	**51 %**	**56 912**	**16 %**

I tabell 3.5.2 har vi gruppert sykehusene, ikke etter foretakstilhørighet, men etter spesialiseringsgrad. Dette gjøres for å fange inn fordelingen av undersøkelser mellom de mer spesialiserte sykehus UNN-Tromsø og Nordlandssykehuset-Bodø, og de 9 lokalsykehusene. Det viser seg da at lokalsykehusene samlet utfører ca. halvparten av alle de radiologiske undersøkelsene i Helse Nord, mens de spesialiserte sykehusene UNN-Tromsø og NLSH-Bodø til sammen utfører en tredel av undersøkelsene og privat sektor 16 prosent. Vi merker oss ellers den lave andel MR-undersøkelser hhv 8% og 5 %) ved de spesialiserte sykehusene, noe som må ses i sammenheng med at halvparten av disse undersøkelsene nærmest er «outsourcet» til privat sektor.

Tabell 3.5.3 Andelen elektive radiologiske undersøkelser i 2021 for ulike modaliteter fordelt på spesialiserte og mindre spesialiserte sykehus.

	Totalt Antall	Totalt Andel elektive	Spesialisert				Lokal		Privat
			UNN Tromsø		NLSH Bodø				Privat
Modalitet			Antall	Andel elektive	Antall	Andel elektive	Antall	Andel elektive	Antall
CT	68 853	72 %	14 361	66 %	12 621	72 %	34 641	68 %	7 230
CT ANG	4 324	42 %	1 267	25 %	442	59 %	2 588	46 %	27
MR	58 482	91 %	4 451	76 %	2 475	76 %	22 513	85 %	29 043
RG	149 879	68 %	29 874	54 %	20 325	69 %	87 625	69 %	12 055
UL	31 099	77 %	8 689	69 %	5 453	83 %	15 457	77 %	1 500
Annet	48 296	79 %	9 246	65 %	8 124	79 %	22 827	76 %	8 099
Totalsum	**360 933**	**75 %**	**67 888**	**61 %**	**49 440**	**73 %**	**185 651**	**72 %**	**57 954**

Av tabell 3.5.3 fremgår at ¾ av undersøkelsene er elektive, mens ¼ er ø-hjelpsundersøkelser. Andelen elektive undersøkelser er høyest (91%) for MR, noe som gjelder alle HFene; for private er denne andelen 100 %, ettersom alle undersøkelsene utført av private er elektive. CT angiografi har klart lavest andel elektive (42%).

Tabell 3.5.4 Antall radiologiske undersøkelser I 2021 henvist fra ulike opptaksområder til de mer spesialiserte sykehusene UNN-Tromsø og NLSH-Bodø

Antall undersøkelser fordelt på utførende enhet		Utførende enhet		
Opptaksområde for HF	Opptaksområde for sykehus*	UNN Tromsø	NLSH Bodø	Totalsum
NLSH	Bodø	1 469		1 469
	Lofoten	760	1 796	2 556
	Vesterålen	1 047	2 117	3 164
NLSH Totalt		**3 276**	**3 913**	**7 189**
FIN	Hammerfest	4 598	44	4 642
	Kirkenes	2 338	32	2 370
FIN Totalt		**6 936**	**76**	**7 012**
HSYK	Mosjøen	335	947	1 282
	Rana	787	2 256	3 043
	Sandnessjøen	596	1 361	1 957
HSYK Totalt		**1 718**	**4 564**	**6 282**
UNN	UNN Harstad	2 414	140	2 554
	UNN Narvik	2 313	96	2 409
	UNN Tromsø		129	129
UNN Totalt		**4 727**	**365**	**5 092**
Utenfor HNs opptaksområde	Utenfor HNs opptaksområde	2 428	1 237	3 665
Utenfor HNs opptaksområde Totalt		**2 428**	**1 237**	**3 665**
Totalsum		**19 085**	**10 155**	**29 240**

Tabell 3.5.4 viser antallet radiologiske undersøkelser som er henvist fra lokalsykehusene til utførelse ved de mer spesialiserte sykehusene i Tromsø og Bodø, samt antallet undersøkelser utført ved UNN-Tromsø for Nordlandssykehuset-Bodø. Sistnevnte antall er knappe 1500, mens UNN-Tromsø samlet har utført hele 14245 undersøkelser som var henvist fra de 9 lokalsykehusene, hvorav ca. halvparten fra de to sykehusene i Finnmark. Tilsvarende har Nordlandssykehuset-Bodø utført knappe 9000 undersøkelser som er henvist fra lokalsykehusene, primært fra Helgeland, Lofoten og Vesterålen.

Tabell 3.5.5 Antallet radiologiske undersøkelser i 2021 per 1000 innbyggere i ulike opptaksområder i Helse Nord

Antall undersøkelser per 1000 innbygger							
	Modalitet						
Opptaksområde for sykehus*	CT	CT ANG	MR	RG	UL	Annet	PET
Bodø	124	13	155	286	69	89	2
Hammerfest	113	15	108	333	73	82	3
Kirkenes	128	18	146	275	70	74	2
Lofoten	134	21	97	307	61	78	2
Mosjøen	135	21	131	351	83	107	2
Rana	140	24	121	360	79	95	2
Sandnessjøen	143	18	116	300	72	84	2
UNN Harstad	96	16	102	328	70	77	2
UNN Narvik	113	20	124	360	79	84	3
UNN Tromsø	88	14	153	313	59	73	3
Vesterålen	120	18	107	313	76	85	2

NB! Tallene er ikke kjønns- og aldersstandardiserte fordi datagrunnlaget ikke gjør det mulig.

Det har i det siste tiåret vært økt oppmerksomhet om forekomsten av ubegrunnet variasjon i forbruket av helsetjenester; både overforbruk, feilbruk og underforbruk er utfordringer, både mht. faglig kvalitet, ressursbruk og prioritering.
I tabell 3.5.5 er det derfor gjort en beregning av hvordan forbruket av undersøkelser innen ulike modaliteter varierer mellom de 11 opptaksområdene. Med de forbehold som følger av datagrunnlaget, merker vi oss det relativt lave forbruket av CT ved UNN-Tromsø og UNN-Harstad. Uten å kommentere de dokumenterte variasjoner mer inngående, inviteres fagmiljøene til å drøfte dokumenterte forskjeller i lys av kapasitetsutfordringer, indikasjonsstillinger og prioriteringer. Arbeidsgruppen ser det som ønskelig at problemstillinger om uønsket forbruksvariasjon også adresseres for de radiologiske undersøkelsene, og at framtidige analyser baseres på et datagrunnlag som muliggjør kjønns- og aldersstandardisering.

4 Den radiologiske arbeidsflyten

Den radiologiske arbeidsflyten kan deles inn i flere faser, og det er varierende i hvor stor grad kommersielle KI-verktøy kan bidra i de ulike fasene.

Den radiologiske arbeidsflyten i forkant av billedtakingen har vi valgt å referere til som fase 0. Denne fasen kan deles tre punkter:

a) Mottakelse av henvisningen
b) Prioritering av henvisningen (berettigelse, hastegrad, valg av undersøkelse og evt. protokoll). Denne oppgaven er ansvarsmessig plassert hos radiologer
c) Timesetting av undersøkelsen

KI vil potensielt kunne bidra til fase 0, men dette vil også kreve at verktøyet er i stand til å analysere andre data enn bare bilder (f.eks. fritekst).

Videre kan den radiologiske arbeidsflyten for selve billedtakingen grovt sett deles inn i fire faser (se figur under):

1. Planlegging av billedtaking
2. Gjennomføring av billedtaking
3. Postprosessering av bilder
4. Beskrivelse av bilder

Figur 2: De ulike fasene i den radiologiske arbeidsflyten.

I tillegg er det en siste fase som skjer i etterkant av billedtaking:

5. Kommunikasjon av resultat til kliniker / pasient.

Det anses foreløpig som uaktuelt bruk KI til dette.

Det er en stor forskjell i tilbudssiden mellom fase 1-3 og fase 4. Dette belyses i det følgende.

Fase 1 – 3

De to første fasene gjøres som regel av radiograf. Det er særlig innenfor modalitetene CT og MR man ser for seg at KI kan bidra med støtte. Løsninger som kan sikre at det velges

rett undersøkelsesprotokoll ut fra henvisningens problemstilling samt posisjonering av pasient i MR/CT-maskinen vil være verdifulle bidrag.

Ved gjennomføring av CT- og MR-undersøkelser vil løsninger som kan bidra til økt bildekvalitet være særlig aktuelt. Slike løsninger omfatter for eksempel mulighet til støtte for å legge inn vinkling i topogrammet (oversiktsbilde) i relasjon til organet/organene som skal undersøkes for å sikre et optimalt opptak. KI kan også bidra til å optimalisere sekvensene som kjøres slik at man kanskje klarer seg med kortere/færre sekvenser. Dette vil på sikt kunne bidra til bedre utnyttelse av maskinparken.

For ultralyd, hvor undersøkelsen i større grad tolkes og vurderes på undersøkelses-tidspunktet, kan KI bidra til å optimalisere undersøkelsen på lignende måter; f.eks. til lettere å gjennomføre størrelse- eller volummålinger for å oppnå mer representative mål. For konvensjonelle røntgenbilder, som for eksempel RG thorax eller skjelettundersøkelser er det antakeligvis større gevinster i å bruke KI til støtte ifbm. tolkning av bildene (fase 4).

Fase 3 involverer rekonstruksjon, postprosessering og visualisering av bildene. Med rekonstruksjon og postprosessering menes den prosessering og matematiske bearbeiding som skjer av råmaterialet fra en modalitet for at dette skal kunne framstilles som et tolkbart bilde. CT benytter detektorer som roterer rundt pasienten, og MR benytter seg av et kraftig magnetfelt hvor biologisk vev som plasseres i feltet avgir signaler ved hjelp av kortvarige radiobølgepulser. Felles for disse modalitetene er at dagens teknologi genererer store mengder data som skal bearbeides.

KI kan bidra til å forbedre postprosessering av data slik at man f.eks. kan få bedre bildekvalitet, redusere artefakter som følge av for eksempel støy, gi mulighet for reduksjon av intravenøs kontrastmengde og/eller stråledose, eller korte ned undersøkelsestiden sammenliknet med å kun benytte tradisjonelle algoritmer for disse stegene[21]. Samtidig må en også være klar over risikoen for at KI-baserte metoder for rekonstruksjon kan gi ustabile og uventede resultater[22]. Etter at bilder er rekonstruerte, kan de også bearbeides for å fremheve relevant informasjon. I sin enkleste form kan dette være f.eks. kantforsterkning eller justering av bildekontrast, eller metoder for å fjerne- eller fremheve kontrast, dersom intravenøs kontrast er gitt. Postprosessering kan ofte gjøres i leverandørspesifikk programvare til forskjell fra bildevisnings-programmet knyttet til det ordinære bildearkivet.

[21] Njølstad T og medforfattere, «Improved image quality in abdominal computed tomography reconstructed with a novel Deep Learning Image Reconstruction technique–initial clinical experience.,» *Acta Radiologica Open,* vol. 10, nr. 4, p. 20584601211008391, 2021.
[22] Antun V og medforfattere, «On instabilities of deep learning in image reconstruction and the potential costs of AI “,» *PNAS,* vol. 117, nr. 48, pp. 30088-30095, 2020.

Oppsummert er det altså slik at fase 1-3 involverer flere oppgaver hvor KI kan bidra. I fase 1 gjelder dette blant annet posisjonering av pasienten i CT og MR, og andre forberedelser til skanningen som å velge protokoll. I fase 2 gjelder dette blant annet innstillinger for scann, redusert stråling og bruk av kontrast, planlegging for rekonstruksjon av bilde, sjekke at kvaliteten på skanningen er god nok og raskere gjennomføring av sekvenser. I den tredje fasen kan KI bidra med f.eks. støyfjerning i bildene og mer effektive måter å fremstille bildene på.

Fase 4 – beskrivelse og tolkning av bilder

Radiologens hovedoppgave er å beskrive og tolke de medisinske bildene. I denne fasen kan KI bidra både til økt *effektivitet* og *kvalitet* på ulike måter:

Mer effektiv helsetjeneste:

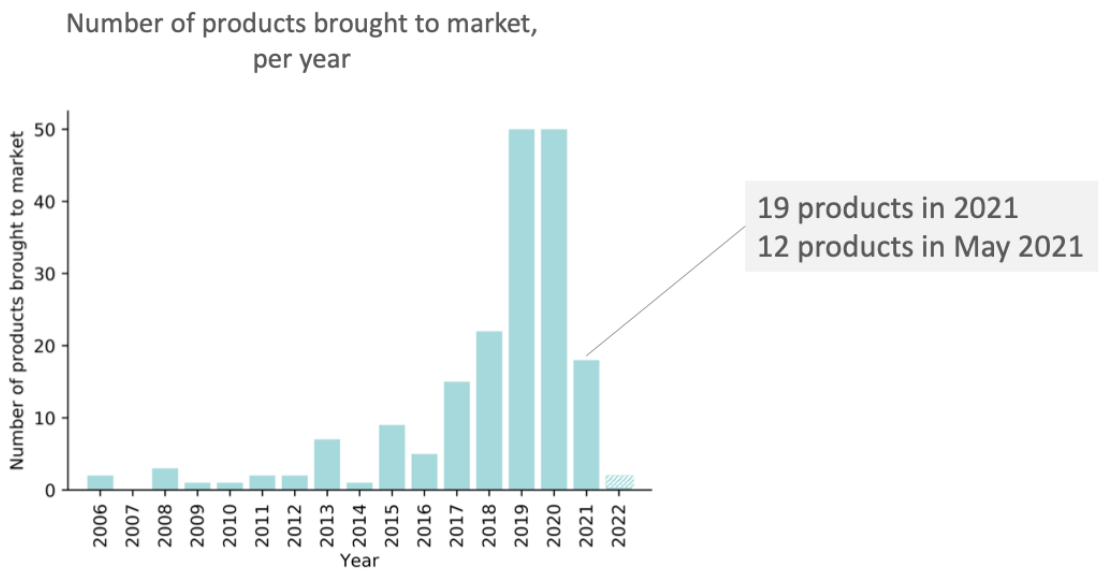
- Kortere lesetid (reading time)
- Triagering av undersøkelser: mer effektiv arbeidsflyt og allokering av ressurser/personell
- Konsistent/strukturert rapportering
- Redusert liggetid (LOS)

Forbedret kvalitet:

- Tidlig diagnostikk/deteksjon
- Forbedret diagnostisk treffsikkerhet (accuracy)
- Persontilpasset diagnostikk og behandling

Figur 3: Beskrivelse av hvordan KI kan forbedre både effektivitet og kvaliteten innen radiologi. Fra van Leeuwen, K.G.,. et al. How does artificial intelligence in radiology improve efficiency and health outcomes?. Pediatr Radiol (2021)

KI-verktøyene som kan brukes i denne fasen er ikke nødvendigvis tilknyttet verken maskinleverandøren eller PACS-leverandøren, men kan også være et verktøy fra en helt frittstående leverandør. I løpet av de siste årene har markedet for disse tolknings-verktøyene vokst ganske mye, før det riktignok ble en stagnasjon i veksten i 2021 da de nye reglene for medisinsk utstyr kom (MDR, se kap.13). Dette er illustrert under.



Figur 4: Oversikt over antall KI-produkter innen radiologi som har blitt introdusert på markedet de siste drøye 15 årene. Fra nettstedet [AI for radiology](#).

Det finnes nå over 200 CE-merkede KI-verktøy for radiologi på markedet (fase 4), og man ser en stor heterogenitet i utvalget. Arbeidsgruppen har gjennomgått hele dette markedet. En oppsummering av dette arbeidet følger i kap.8.

DEL 3: Behovssiden

5 Kartlegging av volum/aktivitet i Helse Nord på utvalgte fagområder innen radiologi

I kapittel 3.5 ble det redegjort for det samlede volumet av radiologiske undersøkelser innen ulike modaliteter fordelt på helseforetakene i Helse Nord. Formålet var her å fange inn de relative forskjeller i aktivitetsvolumet mellom de ulike helseforetak og sykehus. I dette kapitlet kombineres data om modalitet og undersøkelsestyper for å drøfte volum som kriterium i vurderingen av områder innen radiologien der KI-løsninger kan effektivisere driften ved å understøtte radiologene. Tersklene for kritisk volum vil variere mellom modaliteter og typer av undersøkelser, avhengig av kompleksitet og den tiden det tar å tolke bildene. Eksempelvis er tolking av CT-bilder vanligvis mer tidkrevende enn tolking av konvensjonelle røntgenbilder, og MR-bilder er gjennomgående mer krevende å tolke enn CT-bilder. Men også innen hver modalitet er det variasjon i tiden som medgår til tolkning for ulike typer undersøkelser.

Arbeidsgruppen har ikke funnet noe objektivt grunnlag for å fastsette slike minimumsterskler for volum, men har sett det som hensiktsmessig og nyttig å redegjøre for volum av ulike typer undersøkelser ved sykehusene i Helse Nord, slik det fremgår av tabell 5.1 på neste side.

I tabellen er antall undersøkelser (volum) fordelt på ulike typer av sykehus, med vekt på å få fram forskjellene mellom sykehusgrupper på ulike funksjons- og spesialiseringsnivå. I denne rapporten vurderes det som særlig relevant å skille mellom aktiviteten ved UNN-Tromsø, Nordlandssykehuset-Bodø og de 9 lokalsykehusene. Tabellen viser imidlertid at det innen de fleste modaliteter foregår en betydelig aktivitet, også ved lokalsykehusene.

Det finnes ca. 450 ulike undersøkelser i NCRP-kodeverket, fordelt på de ulike radiologiske modalitetene: konvensjonell røntgen (RG), Computer tomografi (CT), Magnet Resonans (MR), Ultralyd (UL) og Nukleærmedisin (NM). De fleste av undersøkelsene utføres ved radiologisk- eller nukleærmedisinske avdelinger, mens et mindretall utføres ved kliniske avdelinger, av klinikere eller annet personell.

Tabell 5.1: Utvalgte radiologiske undersøkelser innen ulike modaliteter i 2021 fordelt på helseforetak, spesialiserte sykehus, lokalsykehus og private aktører.

Antall undersøkelser fordelt på utførende enhet										
	Helse nord totalt		Antall Spesialisert		Sum	Lokal				
Utvalgte NCRP-er	Antall	Andel elektive	UNN Tromsø	NLSH Bodø	Lokal sykehus	FIN	UNN lokal	HSYK	NLSH lokal	Private
CT Abdomen m.fl.	10 562	62 %	1 674	1 293	6 973	1 848	1 429	2 322	1 374	622
CT caput	10 870	49 %	2 121	2 005	6 451	1 496	1 476	2 050	1 429	293
CT hjerte	1 914	92 %	632	243	167	167				872
CT kolumna m.fl.	2 977	62 %	676	355	1 710	381	403	546	380	236
CT pelvimetri	51	37 %	28	3	20	12	2	4	2	
CT samlekoder abdomen og toraks	12 828	79 %	2 422	2 556	7 463	2 142	1 510	2 323	1 488	387
CT skjelett	5 974	56 %	1 410	1 126	3 062	645	945	846	626	376
CT Toraks	9 807	81 %	1 413	1 547	5 911	1 589	1 357	2 105	860	936
CTANG angiografi av toraks	1 599	40 %	474	28	1 096	336	200	370	190	1
CTANG Hode/hals (Slagpasienter/blødninger)	2 725	43 %	793	414	1 492	266	436	430	360	26
CT Angiografi m/flere	3 896	73 %	1 039	594	2 070	497	418	745	410	193
MR abdomen m.fl.	3 963	82 %	754	499	1 816	456	361	620	379	894
MR Caput	13 235	86 %	2 051	1 344	5 946	1 648	1 172	1 968	1 158	3 894
MR Hjerte	342	74 %	218	124	-					
MR kolumna m.fl.	16 433	91 %	900	417	7 595	2 788	828	2 715	1 264	7 521
MR Mamma	423	92 %	267	154	-					2
MR Prostata	1 186	90 %	609	256	316	1	-	198	117	5
MR skjelett	30 726	94 %	1 055	487	9 476	3 465	853	3 561	1 597	19 708
PET	1 147	100 %	1 147		-					
RG skjelett	106 204	71 %	18 118	13 248	64 816	15 667	20 192	17 550	11 407	10 022
RG Toraks	43 675	62 %	11 756	7 077	22 809	5 303	6 983	6 851	3 672	2 033
RG mammografi	7 313	94 %	3 388	2 549	1 133		1 109	24		243
UL (resterende)	20 763	77 %	5 646	4 170	9 499	2 401	2 728	2 866	1 504	1 448
UL Abdomen	11 296	75 %	2 290	1 546	6 134	1 721	1 336	1 894	1 183	1 326
UL Tyreoidea	1 947	96 %	896	372	679	198	134	225	122	
Annet	39 182	78 %	7 258	7 033	19 017	5 081	5 009	5 707	3 220	5 874
Totalsum	**361 038**	**75 %**	**69 035**	**49 440**	**185 651**	**48 108**	**48 881**	**55 920**	**32 742**	**56 912**

Tabell 5.1 viser et uttrekk av relevante modaliteter og undersøkelser som er hensiktsmessig kategorisert i grupper/organområder for å belyse områder som kan være interessante for KI-applikasjoner. Det gjelder eksempelvis:

- «CT abdomen mfl.» og «MR abdomen mfl.» som bl.a. inkluderer lever abdomen, abdomen/Bekken, pankreas, nyrer, etc.
- Ved MR og CT benyttes det ofte samlekoder når en undersøkelse inkluderer flere organområder, for eksempel hals, bryst og mageregionen.
- RG Skjelett, CT Skjelett og MR Skjelett inkluderer RG/CT/MR av alle ulike knokler, f.eks. fot, hånd, Legg, lår, kne, hofte osv.
- CTANG inkluderer CT angiografier, en undersøkelse som kartlegger kroppens arterier. Eksempler på dette er undersøkelser med tanke på lungeemboli, hjerneslag og blødning.

Tabellen viser tall for kliniske mammografier men ikke for screeningsmammografier. Dersom KI skal bidra mtp mammografi så gjelder det hovedsakelig screening, som utgjør vel 25 000 undersøkelser per år i Helse Nord[23], tolket ved de brystdiagnostiske sentre (BDS) i hhv. Tromsø og Bodø.

[23] Det ble i 2021 utført 13005 (77,5% oppmøte) screeningundersøkelser i mammografi for Troms og Finnmark, mens tallet for Nordland i 2021 var 12633 (81,7% oppmøte). Kilde: Mammografiprogrammet ved Kreftregisteret.

«Øyeblikkelig hjelp» er en hastegrad for henvisninger som av ulike årsaker skal tas nokså umiddelbart, mens andre hastegrader som omtaler elektive undersøkelser kan planlegges til et tidspunkt. De fleste undersøkelsene med hastegraden «øyeblikkelig hjelp» har behov for umiddelbar gjennomføring og tolkning som følge av akutt terapeutisk konsekvens. Andre henvisninger har denne hastegraden fordi det haster å få tatt bilder, uten at det umiddelbart er behov for tolkning. Et eksempel på dette er når en kontroll skal tas før pasienten forlater sykehuset og det vil være meget ressurskrevende å måtte sende pasienten til kontroll når den først har blitt utskrevet. Et annet eksempel er kontroller i sammenheng med et poliklinisk besøk på sykehuset hvor bestilling av undersøkelsen i forkant har blitt avglemt. Øyeblikkelig hjelp vil derfor inkludere akutte undersøkelser så vel som undersøkelser utført av inneliggende og polikliniske pasienter med varierende klinisk hastegrad (halv akutt til elektiv innleggelse).

I store trekk utgjør rent elektive polikliniske undersøkelser normalt sett ca. 30%, mens ikke-elektive (inneliggende, øyeblikkelig hjelp, pakkeforløp, poliklinisk haster) utgjør 70% av undersøkelsene. Den observante leser vil derfor kanskje stusse over at den totale andelen elektive pasienter i tab. 5.1 er så høy som 75%. Dette skyldes at det er ikke mulig i datagrunnlaget å skille rent elektive polikliniske og ikke-elektive. Det finnes dessverre ingen mulighet i Helse Nords virksomhetsportal (SAS) å skille ulike grupper av øyeblikkelig hjelp-behov. Pasienter merket som «inneliggende», «pakkeforløp» og «poliklinisk haster» er derfor ikke kategorisert som «øyeblikkelig hjelp» i tab. 5.1. Tabellen viser imidlertid at enkelte kategorier av undersøkelser har en større representasjon av øyeblikkelig hjelp-undersøkelser enn andre, som f.eks. CTANG hode/hals. I kolonnene vises totalt antall undersøkelser både for hele Helse Nord og for de ulike helseforetakene og sykehusenhetene. Formålet med inndelingen er bla. å fange inn eventuelle forskjeller mellom de spesialiserte sykehusene UNN-Tromsø og NLSH Bodø, og de 9 lokalsykehusene (både samlet og hver for seg). Kategorien «Annet» inneholder undersøkelser som ikke vurderes som relevante for KI prosjektet, enten fordi volumet av undersøkelser er for lavt, eller fordi undersøkelser der ikke er relevante for KI.

Tabell 5.2 Totalt antall og andel (%) elektive undersøkelser i 2021 for ulike modaliteter fordelt på spesialiserte sykehus (UNN-Tromsø og NLSH-Bodø), lokalsykehus og private leverandører.

Antall undersøkelser fordelt på utførende enhet										
	Helse nord totalt		Spesialisert				Lokal		Privat	
			UNN Tromsø		NLSH Bodø					
Modalitet	Antall	Andel elektive	Antall	Andel elektive	Antall	Andel elektive	Antall	Andel elektive	Antall	Andel elektive
CT	54 983	67 %	10 376	56 %	9 128	68 %	31 757	67 %	3 722	100 %
CT ANG	8 220	57 %	2 306	41 %	1 036	68 %	4 658	60 %	220	100 %
MR	66 308	91 %	5 854	77 %	3 281	78 %	25 149	85 %	32 024	100 %
PET	1 147	100 %	1 147	100 %						
RG	157 192	69 %	33 262	58 %	22 874	71 %	88 758	69 %	12 298	100 %
UL	34 006	77 %	8 832	68 %	6 088	81 %	16 312	77 %	2 774	100 %
Annet	39 182	78 %	7 258	64 %	7 033	78 %	19 017	76 %	5 874	100 %
Totalsum	**361 038**	**75 %**	**69 035**	**61 %**	**49 440**	**73 %**	**185 651**	**72 %**	**56 912**	**100 %**

Tabell 5.2 viser antall radiologiske undersøkelser fordelt på modaliteter og variasjonen mellom spesialiserte sykehus som UNN Tromsø og NLSH Bodø kontra lokalsykehus og private aktører som f.eks. Evidia og Unilabs. Også her differensiert mellom øyeblikkelig hjelp og elektive undersøkelser.

Det er kjent at CT, PET og MR er mer ressurs- og tidkrevende å tolke enn konvensjonelle røntgen-undersøkelser (RG) og ultralyd, selv om sistnevnte også kan være tidkrevende siden både utførelsen og tolkningen foretas av radiolog. På den annen side er antall konvensjonelle røntgenundersøkelser (RG) atskillig større enn antall CT- og MR-undersøkelser. 10% økning av CT eller MR er ikke det samme som 10% økning av RG når det kommer til ressursbruk. En RG thorax tar mindre enn 5 minutter å tolke, mens en CT thorax tar ofte 15-45 minutter.

Tabell 5.1 og 5.2 illustrerer ikke kompleksiteten i de ulike undersøkelsen på tvers av HFene. Det er rimelig å anta at CT- og MR-undersøkelser som hovedregel vil være mer kompliserte ved spesialiserte sykehus enn lokalsykehus; mange av de kompliserte MR-undersøkelsene gjøres ikke ved lokalsykehus eller hos private leverandører, men henvises til de mer spesialiserte sykehusene. Det er dessuten slik at de spesialiserte sykehusene i hovedsak henviser enklere undersøkelser til private aktører. Videre er det rimelig å anta at elektive CT- og MR-undersøkelser i større grad enn akutte undersøkelser er av en rutinepreget karakter og således krever mindre legebasert stillingtaking til f.eks. protokoller og hvor beskrivelsen ikke krever særskilt prioritering.

I tabellene er ikke de intervensjonsrelaterte prosedyrene inkludert, dvs. der en bruker CT, RG eller UL som veiledning ved intervensjonsprosedyrer som biopsi, finnålsaspirasjon (FNAC), karintervensjon, ablasjon av lever og nyretumorer mv. Disse er ikke tatt med da de ikke har relevans til KI applikasjoner.

Tabell 5.3 Antall Second opinion og demonstrasjoner 2021 i de fire helseforetakene i Helse Nord

HF	UNN	NLSH	HSYK	FIN
Antall second opinion	1 476	278	107	104
Antall demonstrasjoner	5 970	3 230	334	149

Tabell 5.3 illustrerer forskjellen i antall second opinion og demonstrasjoner mellom de ulike HF. Hovedmengden av demonstrasjoner på UNN og NLSH vil være møter i multidisiplinære team (MDT). Dette er møter som utføres i henhold til nasjonale retningslinjer for kreft (Jf. pakkeforløp). Det er i MDT-møter behandlingsvalg avgjøres, og deltakere i MDT består av flere ulike spesialiteter med radiologen i en sentral rolle. Demonstrasjoner på lokalsykehus vil være mer den tradisjonelle «dagen derpå» demonstrasjoner. Forberedelse, gjennomføring og etterarbeid ved MDT-møter er mer krevende enn ved tradisjonelle demonstrasjoner. Ved abdomenseksjonen ved UNN-Tromsø går det med 2 stillinger bare til MDT. Second opinion er som regel en

revurdering av bildeundersøkelser på et høyere faglig nivå, men kan også være så enkelt som å vurdere bilder opp mot nye kliniske problemstillinger.

6 Behovskartlegging

I dette kapitlet følger en oppsummering av behovene som KI potensielt kan bidra til å løse, slik de er beskrevet av radiologer og radiografer ved de fire helseforetakene i Helse Nord som er representert i arbeidsgruppa.

6.1 Behov innen radiologi

Radiologene i arbeidsgruppa har beskrevet at de følgende utfordringene er felles for de fleste helseforetakene i regionen:

- For få radiologer og for stor arbeidsmengde
- Mest utfordrende undersøkelsestyper:
    - Kreftutredninger og -kontroller (MR, CT) – tidskrevende beskrivelser (målinger, sammenligning med eldre undersøkelser)
    - Bruddundersøkelser (røntgen) – stort volum.
- Prioritering av henvisninger, stort volum og mange er mangelfulle
- For liten maskinpark (UNN, NLSH)
- Utfordringer med talegjenkjenning

Dette er i tråd med beskrivelsen i de forrige kapitlene.

Basert på utfordringsbildet har radiologene utformet en rangert liste som beskriver hvilke behov KI potensielt kan bidra til å avhjelpe med, gitt at det finnes eller blir utviklet løsninger for disse behov:

1. Markere områder med patologiske funn:
    a. Deteksjon, måling og kontroll av lesjoner, tumorer, metastaser for forskjellige modaliteter og organgrupper (CT, MR, røntgen; thorax, abdomen, spesielt nevnes MR prostata, MR mamma og CT thorax (lungenoduli)). KI-verktøyene bør gi støtte til generell oppfølging og til å kunne bruke RECIST-kriteriene[24].
    b. Tolking av skjelettrøntgen, særlig bruddeteksjon.
    c. Tolking av røntgen thorax, inkludert fremheving av tuber og slanger
    d. Oppfølging av MS-lesjoner (MR Caput)
    e. Slagdiagnostikk
    f. Deteksjon av lungeemboli og målinger av kar.
    g. Tolking av MR hjerte
    h. Kvantitativ beskrivelse av forløp for lungesykdommer (f.eks. emfysem, fibrose).
    i. Differensialdiagnoser av tumorer (f eks ved integrasjon av radiomics)

[24] RECIST (Reponse Evaulation Criteria in Solid Tumors) er et system som kvantifiserer respons eller progress av kreftsykdom. Systemet baserer seg på at det skal være 20% størrelsesøkning av *mållesjonene*, dvs. kreftsvulstene, for at man skal kunne si at det en progress i sykdommen. Ved delvis respons av behandlingen skal man ha minst 30% reduksjon av kreftsvulstene.

2. Triagering av undersøkelser (f eks CT caput mtp slag eller hemoragi, evt. kandidater for trombektomi; skjelettskader; RG thorax, f eks pneumothorax, tumorer, pneumoni).
3. Triagering av henvisninger.
4. Støtte til både henviser og radiologisk avdeling for å sikre at riktig og nødvendig informasjon kommer på henvisning. Herunder er det aktuelt med støtte til å hente inn relevante kliniske opplysninger fra journal og til å vurdere om det er indikasjon for undersøkelse.
5. Annotering/segmentering av anatomi
6. Talegjenkjenning
7. Automatisk skjelettalder på konvensjonell radiologi.
8. Forbedret bildekvalitet (bedre bilderekonstruksjon).
9. Forbedret pasientlogistikk (lettere kontakt med pasient, mm.)
10. Strålereduksjon (RG, CT, …)

Utover behovene som er felles i hele regionen ble det i løpet av behovskartleggingen også identifisert noen spesifikke behov for enkelte helseforetak. Herunder nevnes:

- Måling av vinkler og lengder på RG og CT pelvimetri (12 undersøkelser).
- Radiomics/tekstur ved ultralydundersøkelser
- Myelomdeteksjon/forløp (CT, MR)
- Virtuelle endoskopier (bronkier, colon (brukes, men kan blir bedre))
- Umiddelbar tilgang til andre relevante casus eller relevant litteratur uten å forlate RIS/PACS
- Tolkningshjelp av ortopedi-MR

6.1.1 Konkretisering av behov med utgangspunkt i volum

Med utgangspunkt i volumet beskrevet i de to foregående kapitlene, har radiologene i arbeidsgruppa jobbet med å konkretisere behovene nevnt over. En beskrivelse av dette følger her.

Behov som skal dekkes av KI kan kategoriseres ut fra flere synsvinkler. For at KI kan integreres i arbeidsflyt og også bidra med ønsket funksjonalitet så bør de enkelte programmer kunne integreres mot den PACS og/eller RIS løsningen som benyttes av radiologene. Det er viktig for at arbeidsflyt av radiologene ikke forstyrres av å måtte forlate den vante arbeidsoverlaten og åpne tilleggsprogramvare. I tillegg ønskes funksjonalitet som mulighet for triagering og flagging av undersøkelser ut fra gitte kriterier (f eks spesielle typer av patologi) eller implementering av målinger eller andre funn fra KI via strukturert rapport inn i beskrivelsene. Også dette krever en integrasjon av KI inn i PACS eller RIS.

Angående faglige behov så omtales de enkelte organområder i det følgende. Overordnet kan det skilles mellom separate behov for undersøkelser i to hastegradskategorier; Elektiv og Øyeblikkelig hjelp; dette fordi disse to kategorier er vesensforskjellige med tanke på funksjonalitet, nytte og gevinst.

Elektive undersøkelser

For elektive undersøkelser er det kreftdiagnostikk, utredninger og kontroller som utgjør hovedmengden av undersøkelser.

For elektive undersøkelser er behovet generelt:

- Utredning (deteksjon patologi og differensiering, ekskludere patologi)
- Kontroller (utvikling av patologi, deteksjon ny patologi)
- Kvalitet (redusere feiltolkning og oversett patologi)
- Objektivisering og kvantifisering (utvikling patologi over tid, leselige tabeller, RECIST)
- Strukturert rapportering (automatisert overføring av mål og evt. endring over tid inn i beskrivelsen, f.eks. i tabeller; RECIST)
- Effektivisering
- Triagering

Aktuelle elektive undersøkelser der en ser behov:

- CT thorax (inkludert CT samlekoder Abdomen og thorax, CTANG thorax, CT Angiografi m. fl.)
- CT abdomen/MR abdomen m.fl. (inkl. CT Abdomen mfl., MR abdomen mfl., CT samlekoder abdomen og thorax, CT angiografi m.fl.)
- RG thorax
- RG skjelettalder
- MR hjerte
- MR mamma
- RG mammografi
- MR prostata
- MR caput
- MR skjelett
- Nukleærmedisin: f.eks. PET/MR prostata
- UL thyreoidea (UL hjerte er omtalt i kap. 6.3)
- MR kolumna

Innen thorax-feltet er deteksjon av noduli av størst interesse. Noduli er små runde prikker i lungeparenkymet som kan være solide eller subsolide, med eller uten solid andel. RG thorax kan detektere 50% av noduli med størrelse på 6-10 mm og noduli opptil 35 mm kan overses, men på CT thorax kan disse detekteres ned til 1-2 mm i størrelse. Betydningen av å detektere disse og følge opp med kontroller, følger av

dokumentert cancer-risiko (størrelse mv.) slik det fremgår av tabellene under. Men slike prikker kan også være metastatisk sykdom, og tidlig deteksjon av slike kan ha prognostisk betydning. Å detektere slike noduli er tidkrevende.

CT thorax og RG thorax er utfordrende undersøkelser, ikke bare for deteksjon av patologi, men også for differensialdiagnostiske vurderinger av lungeforandringer. KI for deteksjon og klassifisering av lungesykdommer som emfysem ville kunne hjelpe spesielt på avdelinger uten subspesialiserte thoraxradiologer. Både CT thorax og RG thorax er høyvolum undersøkelser. På en CT er lungene med flere typer CT-undersøkelser (jamfør punkt 1 over), summert til ca. 25.000 undersøkelser per år. RG thorax har et enda høyere antall– hele 44.000 undersøkelser. Her oppdages regelmessig uventet patologi, som også er malignitetssuspekt. Hos disse pasientene skal det umiddelbart startes pakkeforløp for lungekreft når radiologen oppdager patologi på RG thorax med konsultering av lungelege. Elektive RG thorax ofte kan ligge flere dager før de granskes. En fungerende KI som kan oppdage slik patologi og med funksjonalitet for triagering ville bidra til at flere slike tilfeller oppdages tidligere.

I tillegg er det slik at kreftpasienter følges ofte opp med regelmessige CT-thorax med tanke på å oppdage metastaser. For en slik undersøkelse gis intravenøs kontrast. Timingen på bildeopptaket er ikke optimalisert for fremstilling av blodpropper i lungearteriene, men disse kan likevel ofte være mulig å se. En slik undersøkelse kan ligge flere dager før den blir beskrevet av radiolog. Kreftpasienter har også økt risiko for slike blodpropper. Her ser man potensialet for en KI-løsning som kan finne blodpropper i lungene og flagge disse i arbeidslisten slik at pasienten kan få prioritert behandling. Videre, for pasienter som har etablerte metastaseforandringer og hvor man ønsker å følge antall og størrelse av disse for å vurdere behandlingsrespons, kan en KI-løsning som bidrar med automatisk telling og måling, redusere tidsbruken for radiologen.

Tabell 6.1 Risikotabeller for kreft eter størrelse på svulst og solid versus ikke solid noduli[25 26 27]

Solide noduli	Risiko cancer
< 5 mm	0,1-1%
5-10 mm	6-28%
>20 mm	64-82%

Subsolide noduli	Risiko cancer
Ikke solid andel	18%
Solid andel	63%

Volum dobblingstid(VDT)	Tolkning
< 20 dager	Infeksjon/inflammasjon
20-400 dager	Økt risiko for malign(10%)
400-600 dager	Intermediær risiko(4%)
>600 dager	Lav risiko – sannsynlig benign(0,8%)

I abdomen er det aktuell deteksjon og tolkning av patologi som er vesentlig, men også på en effektiv måte å analysere endring i størrelse av patologi over tid. Kreftleger har alltid

[25] Henschke CI, Yankelevitz DF, Mirtcheva R, et al. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. Am J Roentgenol. 2002;178:1053–1057.
[26] Wahidi MM, Govert JA, Goudar RK, et al. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer? ACCP evidence-based clinical practice guidelines (2nd edition). Chest 2007; 132: Suppl., 94S–107S.
[27] Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. Lancet Oncol 2014; 15: 1332–1341.

etterspurt evaluering av metastaser etter RECIST-kriterier med sine kriterier for regress eller progress, men dette har fra radiologisk side ofte ikke vært mulig på grunn av økt tidsbruk. En automatisert strukturert rapport med tanke på endring av størrelse, ville effektivisere arbeidet for både radiolog og onkolog, og resultatet ville være objektivt. CT Abdomen og MR Abdomen har samlet sett et høyt volum på ca. 28.000 undersøkelser per år.

MR hjerte har et lavt men stadig økende volum med 560 undersøkelser og utføres kun på UNN Tromsø og NLSH Bodø. Det er bare 2-3 radiologer som håndterer MR hjerte i Helse Nord. Undersøkelsen er tidkrevende både å gjennomføre og å tolke. Nøyaktig og korrekt segmentering og kvantifisering er avgjørende for korrekt diagnostikk. Det finnes kvantifiserings- og segmenteringsverktøy fra hardware-leverandørene både i Bodø og i Tromsø. Faglig vurdering av tilgjengelig verktøy i Bodø tilsier at softwaren fungerer såpass upålitelig at segmenteringen som regel må utføres manuelt med tilsvarende stor tidsbruk.

Nasjonalt er det stor mangel på mammaradiologer, og i Helse Nord er denne mangelen en enda større utfordring siden det ikke finnes fastansatte mammaradiologer ved UNN-Tromsø. Som ved all annen radiologi med stor «turn over» av undersøkelser, er det økt risiko for at patologi blir oversett. RG mammografi er en høyvolum-undersøkelse som følge av mammografiprogrammet (screening), og fremstiller et anatomisk lite område med standardiserte kriterier for bildeopptak. Således er mammografi et område som ligger bra til rette for utvikling av KI. Det foregår studier ved Kreftregisteret som ser på muligheter å implementere KI i mammografiscreening. Per i dag kan KI ikke bidra til å dekke behov for manglende mammaradiologer. Det bør avventes resultater fra kreftregisterets studier for det tas stilling til om bruk av KI i screening i sluttresultat vil føre til bedret pasientomsorg.

Som følge av Nasjonalt handlingsprogram for brystkreft har bruk av MR i diagnostiseringen av cancer mamma fått en økende rolle, selv om dagens antall undersøkelser er relativt lavt (423 totalt). Stadig flere brystkreftpasienter behandles neoadjuvant med behov for flere MR-kontroller gjennom forløpet. Bruk av KI vil kunne bidra til bedre kvalitet på tolkningen og lette arbeidsmengden for radiolog.

MR prostata har hatt en kraftig økning i antall undersøkelser de seneste årene pga. endring i handlingsprogrammet for prostatakreft. Også her er utfordringen at tolkning utføres av et fåtall radiologer og det er behov for tolkningshjelp, deteksjon og vurdering av eventuelle tumorer.

MR caput har et høyt volum (13.235 undersøkelser) med varierende problemstillinger, men det bør vurderes om KI-algoritmer for slagpasienter, demensutredning og multippel sklerose (MS) kan være aktuelle. For pasienter med MS er MR av hjernen viktig i oppfølging og terapivalg. Å identifisere nye MS-forandringer ("plakk") kan være

tidkrevende, spesielt for pasienter med mange eller små plakk. I dette arbeidet sammenlikner man også funnene med tidligere undersøkelser. En KI-løsning har i denne sammenhengen potensiale til å bidra til mer effektiv og presis diagnostikk ved å finne og segmentere plakk og gi informasjon om endringer sammenliknet med tidligere undersøkelser. Dessverre er det vanskelig å differensiere ut antall undersøkelser på bakgrunn av disse tre kliniske problemstillingene.

RG skjelettalder er ikke en høyvolumundersøkelse, men vurderingen av slike bilder er tidkrevende med sammenligning av bilder i bøker for å stadfeste skjelettalder. Det blir da også en subjektiv vurdering. I Helse Nord brukes for tiden en KI-applikasjon for dette som gjør det raskere og objektivt.

MR skjelett er en høyvolumundersøkelse, og blir av kapasitetsgrunner både på maskinparken og radiologsiden oftest overført til det private. Her kunne man sett verdi i algoritmer med tanke på anatomi og deteksjon av elementær patologi, men slik applikasjon finnes for tiden ikke.

Øyeblikkelig hjelp

Ved øyeblikkelig hjelp undersøkelser er det flere områder med behov:

- Triagering (hva haster mest å se på)
- Avlastning på vakt (brudd-deteksjon)
- Kvalitet (Redusere feiltolkning og oversett patologi)
- Deteksjon (Påvise patologi)

Aktuelle øyeblikkelig hjelp undersøkelser der en ser behov:

- CTANG angiografi av thorax og CT
- Thorax CT caput
- CT abdomen
- RG skjelett
- RG thorax

Triagering vil si at en KI-algoritme går igjennom undersøkelsen før radiolog har sett på den, og «flagger» ut patologi eller ikke patologi i forhold til hva KI-algoritmen ser på. På den måten kan et positivt algoritmefunn føre til at undersøkelsen kommer høyere opp på arbeidslisten og prioriteres raskere for tolkning av radiolog. Triageringshjelp er mest aktuelt for UNN Tromsø og NLSH Bodø på grunn av volum på vakt, men også for lokalsykehus på dagtid. Det kan dessuten tenkes at slik triagering også kan gjøres på elektive undersøkelser (se ovenfor), f.eks. en poliklinisk kontroll av en kreftpasient der KI-algoritmen har påvist lungeemboli eller lungetumor og at dette synliggjøres på arbeidslisten. Triagering vil gjelde alle undersøkelser nevnt over ved øyeblikkelig hjelp i tillegg til flere av de anførte elektive undersøkelser.

For CTANG av thorax og CT thorax kan KI bidra til deteksjon av lungeemboli, deteksjon og tolkningshjelp av fortetninger og deteksjon av noduli. Isolert sett for CT utføres cirka 1.000 CT per år med tanke på lungeemboli. Antallet er i seg selv ikke spesielt stort, men man må ta i betraktning at det ofte tas på vakt. I tillegg vil dette kunne være en algoritme som kan benyttes også for elektive CT thorax-undersøkelser, noe som innebærer at antallet undersøkelser blir betydelig høyere. For CT thorax er antallet cirka 6.200 undersøkelser.

Ved CT caput er det mest aktuelt med tanke på deteksjon av blødninger, infarkt og okklusjon av kar. Okklusjoner av kar kan være utfordrende å påvise, og er vesentlig å påvise uten forsinkelse, ettersom det kan være aktuelt med trombektomi som kan bedre prognosen. CT caput er høyvolumundersøkelse på vakt med 5.500 undersøkelser og CTANG hode/hals 1.600 undersøkelser.

Ved CT abdomen er de aktuelle problemstillinger fri luft, blødning, ileus, infeksjoner, nyrestein mv. Det fins i dag KI-algoritme som kan påvise fri luft med høy sensitivitet, og antall CT abdomen-undersøkelser er ca. 7.000.

Hensikten med KI-algoritme for RTG skjelett på vakt, er å påvise, flagge, respektive triagereundersøkelser hvor KI påviser eller mistenker fraktur. Dermed blir det mulig for vakthavende radiolog å prioritere sannsynlig patologiske undersøkelser. KI tenkes å avlaste vakthavende radiolog på to måter: 1) Ved bare å se på undersøkelser der KI har positive funn, og 2) Ved at radiolog ikke ser på undersøkelsen i det hele tatt og pasient blir henvist til ortoped ved positivt funn. Dette er høyvolumundersøkelser med ca. 31.000 undersøkelser.

RG thorax er en høyvolumundersøkelse på vakt med ca. 16.700 undersøkelser, og er - som tidligere nevnt - en av de rent billedmessig mest utfordrende å vurdere. Her kunne det være behov for både deteksjon av patologi så vel som tolkningshjelp.
Samlet sett - for øyeblikkelig hjelpundersøkelser - kunne KI bidratt til bedre vaktlogistikk, raskere tolkning av bilder med alvorlig patologi, mindre oversett avgjørende patologi og til generell hjelp. Triagering i arbeidslisten kan benyttes på alle overnevnte undersøkelser.

Det finnes KI-algoritmer som dekker flere av ovennevnte behov, men ikke alle, og for noen behov reises spørsmålet om volumet er for lite til at investering i en KI-algoritme vil gi kost-nytte; mer om dette kan leses i kapittel 8.

6.2 Behov sett fra radiografers ståsted

Radiografene i arbeidsgruppa har pekt på en rekke behov som potensielt kan dekkes av KI-verktøy, gitt at løsningene blir utviklet eller allerede eksisterer. Det presiseres at

noen av behovene nok ikke i dag kan dekkes av KI-verktøy, men potensielt av annen teknologi. Disse er likevel tatt med for helhetens skyld.

Liste over behov (urangert):

- *Foreslå hvilken protokoll som skal brukes*: Ved gjennomføring av en CT-undersøkelse skjer dette på bakgrunn av en protokoll som velges av radiolog basert på kliniske opplysninger i henvisning. Man kan se for seg en løsning hvor et KI-verktøy kommer med forslag til protokoller ut fra henvisningstekst når radiolog prioriterer undersøkelsen. Videre kan man se for seg at maskinen automatisk velger rett protokoll når radiograf henter opp pasientens henvisning i RIS.
- *Bistand til posisjonering:* F eks sentrering av pasient i CT – maskin.
- *Bistand til å velge sekvenser og vinkling på MR – undersøkelser.*
- *Brudd-deteksjon:* Algoritme for å avdekke brudd slik at radiograf kan sende pasient hjem eller videre i systemet. På den måten avlastes vakthavende radiolog og pasienten slipper ventetid. Radiolog beskriver bildene etter hvert.
- *Tolking av røntgen thorax:* Algoritme for å avdekke fortetninger på røntgen av lungene (for eksempel pneumoni) slik at pasient kan sendes hjem ved negativt svar på vakt for å følges opp dagen etter. Pasient slipper lang ventetid og radiolog kan fokusere på annet hastearbeid på vakt.
- *Triagering*: Mange ganger er det slik radiografene ser ting på røntgen som dekanskje tenker er alvorlig, men som ikke er akutt alvorlig. Eksempelvis kan det være mistanke om mulig kreftdiagnose. Dersom det er radiograf har slik mistanke, kan de benytte seg av KI som beslutningsstøtte for å vurdere om mistanken er reell eller ikke, og dermed kunne pushe beskrivelsen opp på radiologens liste. I dagens situasjon med stor mangel på radiologer og treghet på beskrivelse kan dette eventuelt bidra til at noen undersøkelser prioriteres i køen og beskrives hurtigere enn andre undersøkelser uten indikasjoner på mulig patologi?
- *Opplæring*: KI kan brukes til å lære radiografstudenter å se patologi på bildene
- *Automatisk link mellom prioriteringskode/protokoll og prosedyre* slik at prosedyre popper opp på skjermen på maskin hos radiograf
- *Bistand til radiolog når undersøkelsesprotokoll skal velges* ved at det kommer opp en forhåndsvisning av protokoll på skjermen når musepeker holdes over hver enkelt prosedyre.
- *Automatisk dataregistrering i RIS:* Data fra kontrastsprøyta som for eksempel mengde, NaCl og flow registreres automatisk i RIS (Sectra)
- *Flow og kontrastmengde popper automatisk opp på kontrastsprøyta* når man velger protokoll.

6.3 Andre behov innen medisinsk bildeanalyse

Medisinsk bildeanalyse er et stort og omfattende fagfelt. I denne rapporten vil vi først og fremst konsentrere oss om bildemodaliteter som beskrives av radiologer i Helse Nord, men for helhetens skyld tar vi med bildetyper som analyseres av andre fagspesialister (kardiologer, øyeleger, tannleger, osv.). Det er verdt å nevne at en god del KI-verktøy innenfor disse kategoriene vil kunne basere seg både på bilder og andre typer signaler.

- Dermatologi (hud)
  - Diagnostisere hudkreft og andre hudsykdommer
  - Foreløpig få produkter på markedet (to produkter godkjent av FDA per august 2022[28]).
- Oftalmologi (øye)
  - Detektere/klassifisere diabetisk retinopati og andre øyesykdommer som aldersrelatert makuladegenerasjon, glaukom, aldersrelatert eller medfødt katarakt, osv.
  - Finnes flere CE-merkede (7 FDA-godkjente) produkter på markedet for å detektere diabetisk retinopati.
- Odontologi (tann)
  - Automatisert tolkning av tannbilder (røntgenbilder, CBCT og MR-skanninger), behandlingsanbefalinger, predikere fremtidige tannsykdommer
  - Finnes produkter (noen få sertifiserte) som kan bidra med støtte til tolkning av tannbilder.
- Kardiovaskulær medisin (hjerte og blodårer)
  - Detektere hjertemuskelsykdommer, hjerteinfarkt, kransåresykdom, klaffesykdom, arytmier basert på ekkokardiografi, angiografi, EKG, osv.
  - Det finnes over 40 FDA-godkjente produkter innen denne kategorien. Langt fra alle benytter seg av bilder[29]. I oversikten AI for radiology finner man 14 CE-merkede algoritmer innen det kardiologiske fagfeltet.
- Strålingsonkologi
  - Stråleterapi, planlegging av behandling, predikere behandlingsrespons
  - Noen få sertifiserte produkter
- Gastroenterologi: endoskopi /koloskopi
  - Detektere adenoma og polypper, gastrointestinal blødning
  - Noen få sertifiserte produkter
- Hematologi og mikrobiologi
  - For eksempel analyse av blod, urin, sæd, andre kroppsvæsker.
  - Ca. 15 FDA-godkjente produkter.
- Nevrologi:

[28] https://www.medscape.com/viewarticle/978352
[29] https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

    - Forkalkning i halskar, deteksjon av autisme[30], deteksjon av epilepsi, mv..
    - I overkant av 10 FDA-godkjente produkter
- Gynekologi / fødsel / kvinnehelse
    - Ultralyd av eggstokker, osv. (utføres av gynekologer), tolking av prenatal ultralyd (jordmødre og leger)
    - Ett FDA-godkjent produkt[31]: undersøkelse av embryo ved kunstig befruktning.
- Kirurgi
    - Videoovervåkning av operasjonen
    - Augmented reality for skopi-kirurgi, inkludert anatomi-gjenkjenning, prosess-gjenkjenning, osv.
    - Noen få FDA-godkjente produkter.
- Digital patologi
    - F. eks. kreftdiagnostikk.
    - Er et prioritert område i KI-strategien og vil aktualisere seg når feltet blir mer modent.
    - Noen få CE-merkede produkter, deriblant ett fra Norge[32].

[30] https://www.fda.gov/news-events/press-announcements/fda-authorizes-marketing-diagnostic-aid-autism-spectrum-disorder
[31] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182798
[32] https://www.massdevice.com/domore-diagnostics-ce-mark-ai-colorectal-cancer/

DEL 4: Tilbudssiden

Som beskrevet i kap. 4 består den radiologiske arbeidsflyten av flere faser:

1. Henvisning og timesetting
2. Planlegging av billedtaking
3. Gjennomføring av billedtaking
4. Rekonstruksjon og postprosessering av bilder
5. Beskrivelse/tolkning av bilder

I de to påfølgende kapitler redegjør vi for status på den kommersielle tilbudssiden innen disse fasene.

7 Status for den kommersielle tilbudsside innen fase 0-3

Det er ikke funnet egnede kommersielle og sertifiserte KI-verktøy som kan bidra til oppgavene i fase 0. Vår behovsanalyse viser imidlertid KI-løsninger som kunne avlaste radiologene i arbeidet med å prioritere henvisninger, etterspørres.

Flere av aspektene i fase 1-3 treffer behovene som er beskrevet tidligere i denne rapporten. Radiografene i arbeidsgruppa har skissert behov blant annet når det kommer til posisjonering av pasienten og utvelgelse av protokoll. Videre er radiologene på generell basis opptatt av at bildene de skal beskrive er av så god kvalitet som mulig.

Imidlertid er det slik at oppgavene i fase 1-3 av den radiologiske arbeidsflyten er tett tilknyttet maskinvaren som skal brukes for å gjennomføre skanningen. Det vil variere mellom maskinleverandørene i hvor stor grad de bruker KI-basert programvare for å øke produktiviteten, eksempelvis ved å muliggjøre kortere skannetider eller forbedre bildekvaliteten gjennom å fjerne støy. Det vil også variere hvor automatisert den KI-baserte programvaren er. For eksempel vil en KI-algoritme for støyfjerning i bilder kunne gjøre det helautomatisk, mens et KI-basert verktøy for posisjonering av pasienten vil kunne brukes som et hjelpemiddel/støtteverktøy for radiografen.[33] KI-løsninger innenfor disse tre fasene, vil ofte være knyttet til den enkelte modalitetens leveranse og inngå som del av eller opsjon til programvaren som følger maskinen.

[33] Vi har ikke klart å finne en fullstendig oversikt over markedet for fase 0-3, men vil referere den interesserte leser til denne oversiktsartikkelen, som gir et inntrykk av hva status er per okt. 2022: Tadavarthi Y, Makeeva V, Wagstaff W, Zhan H, Podlasek A, Bhatia N, Heilbrun M, Krupinski E, Safdar N, Banerjee I, Gichoya J, Trivedi H. Overview of Noninterpretive Artificial Intelligence Models for Safety, Quality, Workflow, and Education Applications in Radiology Practice. Radiol Artif Intell. 2022 Feb 2;4(2):e210114. doi: 10.1148/ryai.210114. PMID: 35391770; PMCID: PMC8980942.

En investering i ny maskinvare for billedtaking vil derfor også være førende for hvilke muligheter som åpner seg mht. å få hjelp av kunstig intelligens til å planlegge og gjennomføre billedtaking. Denne avhengigheten gjør at fleksibiliteten reduseres; derfor vil anbefalinger rundt bruk av KI i disse tre fasene bli noe lavere prioritert i denne rapporten. Likevel vil arbeidsgruppen påpeke at det er viktig å ha fokus på disse problemstillingene når helseforetakene i nord i tiden framover skal fornye og oppgradere maskinparken.

8 Den kommersielle tilbudssiden innen bildetolkning (fase 4)

Arbeidsgruppen har brukt nettstedet AIforradiology som hovedkilde for å få oversikt over det kommersielle tilbudet av sertifiserte KI-verktøy innen radiologi. AIforradiology har følgende inklusjons- respektive eksklusjonskriterier:

Inklusjon

- Fokusert på radiologi eller nukleærmedisin
- Basert på maskinlæring
- Hovedsakelig beregnet for klinisk bruk
- CE-merket

Eksklusjon

- Plattformer og markedsplasser (se kap. 10)
- Forskningsprogramvare
- Bildeprogramvare brukt under billedtaking, intervensjon eller kirurgi
- Programvare for diktering

I praksis vil dette si AIforradiology beskriver tilbudssiden innen KI-støttet bildetolkning, som vi i denne rapporten har referert til som fase 4 i den radiologiske arbeidsflyten.

I tillegg til at vi har brukt AIforradiology, har de største leverandørene (Siemens, Sectra, GE Healthcare, Philips og AIdoc) blitt invitert inn i møter med arbeidsgruppen hvor satsningen på KI til de respektive selskapene har blitt presentert og diskutert.

Radiologene og radiografene i arbeidsgruppa har sittet sammen i grupper og gått gjennom tilgjengelige KI-produkter innenfor spesialitetene: abdomen, mammografi, hjerte, MSK, neuro og thorax.

Følgende generelle punkter er beskrevet i vurderingene under:

- Hvilket område dekker dette KI-produktet? Er det av interesse for vårt bruk?
- Mulighet for flagging (triagering) hvor det er aktuelt
- Mulighet for integrasjon i PACS/RIS arbeidsflyt og/eller via KI-plattform
- Mulighet for styring av arbeidsflyten i PACS/RIS, dvs. om det er styrbart når og hvem som vises resultat av KI-tolkning innen en radiologisk avdeling og innenfor en definert arbeidsflyt
- Mulighet for «strukturert rapportering»

8.1.1 Abdomen

Prostata

- 3 MDR-produkter og 2 MDD-produkt innen prostata

- Fagområdet dekker et stort behov
- Usikkerhet om mulighet for integrasjon i RIS/PACS arbeidsflyt
- Strukturert rapportering mulig

Abdomen (utenom prostata):

- 8 MDD-produkter. Områdene som dekkes er ikke av høy relevans for vårt prosjekt.

8.1.2 Mammografi

2 MDR-produkter

- Kan brukes innen screening og innen klinisk mammografi
- Kan integreres i PACS eller i RIS
- Mulighet for strukturert rapportering
- Ønskelig med evaluering av effekt i publiserte fagfelle-vurderte studier
- Forskjellig presentasjonsmåte av risikoscore, ønskelig med høyere grad av differensiering

8 MDD-produkter og CE-klasse II(a)

- 1 angår MR mamma: relevant for fagområdet og kan integreres mot PACS
- 1 angår kun tomosyntese: først relevant når screening foretas med tomosyntese
- 1 angår kun brysttetthet: løser for tiden ingen av de presserende problemer
- 5 for 2D-mammografi, hvorav 3 angir integrasjonsmulighet mot PACS eller RIS, kan være relevante

Usikkerhet om graden av nytte med dagens produkter, bør avvente videre studier fra kreftregisteret før løsninger på dette området eventuelt implementeres.

8.1.3 Hjerte

4 MDR-produkter

- 2 er aktuelle ut fra bruksområdene (MR hjerte)
- Mulighet for RIS/PACS integrasjon og for strukturert rapportering ser ut å være til stede

6 MDD produkter i klasse IIa

- 1 for ultralyd, ikke relevant her
- 1 for bedring av bildekvalitet for lav-dose CTer, ikke relevant her
- 1 for CT hjerte, kun calcium score, ikke relevant her
- 1 for CT hjerte: relevant, integreres via PACS, sannsynligvis ingen strukturert rapportering, mange publikasjoner

- 2 for MR hjerte: begge relevante, kan integreres enten via PACS eller RIS, strukturert rapportering støttes

MR hjerte KI-produkter kan muligens bidra med betydelig legetidsbesparelse per undersøkelse (kan ligge opp til 1 time), noe som i vesentlig grad avhenger av funksjonaliteten av automatisk segmentering av ventriklene, ventrikkelveggene og klaffe-nivåer, som må utprøves for hvert produkt

8.1.4 MSK (Muskel-skjelett)

4 MDR-produkter

- Har sine avgrensete snevre bruksområder
- Kan integreres i RIS/PACS
- Flagging/triagering og strukturert rapportering viktig, ikke alle som har den muligheten

8 MDD-produkter

- 4 produkter som dekker relevante områder, kan integreres via Ris eller PACS; de fleste kan flagge, ikke alle tilbyr strukturert rapportering (ikke avgjørende for dem som ikke kan), de fleste med fagfellevurderte publikasjoner, en av disse dekker benalder (NB! Helse Nords aktuelle løsning har ikke klasse II)
- 4 produkter dekker område av mindre relevans

8.1.5 Neuro

6 MDR-produkter

- Produktene har ulik bredde i anvendelsesområder og må vurderes opp mot behov av Helse Nord
- RIS/PACS integrasjon *ikke* mulig hos alle
- Strukturert rapportering mulig hos alle
- Flaggingsmulighet for enkelte
- Anvendelsesområder avgjørende, ikke spesifisert for alle produkter
- *29 MDD-produkter*
- Spredte anvendelsesområder, delvis meget spesifikke og mindre relevante
- 10 anses å mindre relevans eller det er ut fra informasjon ikke mulig å avgjøre
- 13 anses å dekke områder med klar relevans
- 11 av de relevante kan integreres mot PACS og/eller RIS
- Av disse 11 mangler 3 produkter mulighet for strukturert rapport, men for de aktuelle områder vurderes ikke dette som vesentlig
- Av disse 3 har 1 mulighet for flagging, som i varierende grad er en fordel

8.1.6 Thorax

Delt inn i:

- Røntgen thorax
- CT thorax

Røntgen thorax

10 MDR-godkjent (av disse 1 MDD og 5 FDA):

- Kun 6 av de 10 applikasjonene er aktuelle.
- Både front og side: 3 applikasjoner (synes å måtte være et absolutt krav)
- Bare front: 2 applikasjoner
- Ikke oppgitt (Front/side): 1
- Pneumothorax: 5 applikasjoner
- Fortetninger: 5 applikasjoner
- Kateter, dren og tuber: 3 applikasjoner
- MSK: 1 applikasjon
- Triage: 3 applikasjoner
- RIS/PACS integrasjon: 6 applikasjoner
- Strukturert rapport: 2 applikasjoner
- Av disse 6, er det 2 som skiller seg ut som mer aktuelle siden de er multifunksjonelle og kan detektere mange forskjellige patologier.

20 MDD-godkjent (6 FDA godkjent)

- Kun 8 av 20 aktuelle
- Både front og side: 3 applikasjoner (synes å måtte være et absolutt krav)
- Bare front: 4 applikasjoner
- Ikke oppgitt (Front/side): 1
- Pneumothorax: 8 applikasjoner
- Fortetninger: 8 applikasjoner
- Kateter, dren og tuber: 0 applikasjoner
- MSK: 1 applikasjon
- Triage: 8 applikasjoner
- RIS/PACS integrasjon: 8 applikasjoner
- Strukturert rapport: 5 applikasjoner
- Av disse 8, er det 2 som skiller seg ut som mer aktuelle siden de er multifunksjonelle og kan detektere mange forskjellige patologier.

CT thorax

10 MDR-godkjent (7 FDA godkjent)

- Kun 5 av 10 aktuelle
- Detektere og klassifisere noduli (inkludert subsolide noduli): 5 applikasjoner
    - Kan også detektere vekst av noduli: 4 applikasjoner

- Fortetninger/Interstitielle lungesykdommer (ILD): 1 applikasjon
- Lungeemboli: Ingen
- Case/faglitteratur støtte: 1 applikasjon
- Rapport: 4 applikasjoner.

20 MDD Godkjent (11 FDA godkjent)

- 17 av 20 aktuelle
- Detektere og klassifisere noduli (inkludert subsolide noduli): 8 applikasjoner
    - Kan også detektere vekst av noduli: 3 applikasjoner
- Fortetninger/ILD: 4 applikasjoner (gir støtte til analyse og kvantifisering, ikke diagnose)
- Lungeemboli: 3 applikasjoner
- Aortadisseksjon: 1 applikasjon.
- Emfysem/Kols/Astma: 7 applikasjoner

Mest aktuelt med 1 applikasjon for noduli, 1 for lungeemboli og 1 for fortetninger/ILD. De fleste har integrasjon mot RIS og/eller PACS. Varierende om de er «stand alone» eller tilgjengelig fra «marketplace». Kvantifisering av emfysem/ILD er ikke relevant fra at radiologisk ståsted, men klinisk, så her må man vurdere å konferere med lungeleger om dette er noe de har behov for.

8.1.7 Andre

1 interessant MDD-produkt

- MR lumbal columna (korsrygg), relevant fagområde, PACS-integrert og mulighet for strukturert rapport

8.2 Norske erfaringer med anskaffelse av verktøy for KI-støttet bildetolkning

I 2021 utlyste Vestre Viken HF, som det første helseforetak i Norge, en offentlig anbudskonkurranse om anskaffelse av verktøy for KI-støttet bildetolkning. På grunnlag av behovsvurderinger gjort av radiologene i helseforetaket, ble følgende fem områder innen bildeanalyse prioritert for anskaffelse:

- CT thorax - lunge noduler/lungemetastaser
- CT thorax – lunge emboli
- MR caput – MS-oppfølging
- Konvensjonell røntgen – thorax
- Konvensjonell røntgen - skjelett/brudd

Anskaffelsen ble gjort i samarbeid med Sykehuset Vestfold HF. I tillegg til de to helseforetakene bisto også Sykehuspartner HF og Sykehusinnkjøp HF i anskaffelsen. Det ble utarbeidet en kravspesifikasjon med tilhørende vekting av kvalitet (70%) og pris (30%). Konkurransepreget dialog ble valgt som anskaffelsesform. Konkurransen ble kunngjort (Doffin og TED) i mars 2021 med frist for leverandørene om å søke om deltakelse 23. april 2021. Det kom i alt søknader fra 9 leverandører. Av disse ble 3 avvist og 6 invitert til å delta i konkurransen. I den opprinnelige planen var det lagt opp til kontraktsignering innen årets utgang (desember 2021). Anskaffelsen tok av flere grunner lengere tid enn forutsatt:

- For det første var radiologene i helseforetaket så hardt presset med løpende driftsutfordringer, at de hadde vanskelig med å frigjøre tilstrekkelig med tid til å delta i dette arbeidet.
- Covid 19 medførte at samtlige møter med leverandører og andre, ble gjennomført digitalt. Blant aktørene på kjøpersiden kan det være noe ulike vurderinger av hvorvidt dette forsinket prosessen.
- I dialogen med leverandørene brukte man noe tid på å avklare fordeler og ulemper ved å kjøpe algoritmer via en markedsplass /plattform (se kap. 10) og å gjøre seg kjent med de ulike plattformene som ble tilbudt. Kravet til plattform ble deretter innarbeidet i endelig kravspesifikasjon, med krav om at plattformleverandørene måtte være i stand til å levere tredjeparts KI-applikasjoner.
- På et relativt sent tidspunkt i prosessen ble det fra Helse Sør-Øst RHFs side tatt initiativ til at anskaffelsen ikke bare skulle avgrenses til Vestre Viken HF og Sykehuset Vestfold, men inneholde opsjon for avrop fra alle helseforetak i Helse Sør-Øst. (Det kan her tilføyes at UNN HF også har opsjon for avrop på inngått kontrakt).
- Den valgte anskaffelsesformen, konkurransepreget dialog, var transaksjonskrevende, og det ble avholdt hele 18 møter med leverandørene. Vurderingen er likevel at man neppe hadde «kommet i havn» ved å bruke andre anskaffelsesprosedyrer, gitt den informasjonen man hadde tilgang til før kunngjøringen.
- En av de tapende leverandørene klagde på tildeling av kontrakt. Behandling av klagen gikk over sommerferien, noe som gjorde det umulig å signere i juli som opprinnelig planlagt.

Samlet bidro ovennevnte forhold til at rammeavtalen først ble signert ultimo august 2022. De helseforetakene som er inkludert i rammeavtalen – alle HF i Helse Sør-Øst samt UNN HF – kan gjøre avrop i en toårsperiode fra tidspunkt for signering av kontakt, altså innen ultimo august 2024. Etter dette tidspunkt er det mulig å forlenge rammeavtalen med 1+1 år.

Blant de opprinnelige 9 søknadene, måtte flere avvises i løpet av prosessen. Tre av søknadene ble avvist som følge av at kvalifikasjonskravene ikke var oppfylt. De

gjenværende seks leverandørene ble invitert til å levere første løsningsforslag innen 4. juni 2021. Etter mottak av første løsningsforslag, ble det klart at en av leverandørene måtte avvises fordi det obligatoriske kravet om at tilbudt(e) algoritme(r) skulle være «klar til å tas i bruk», ikke kunne oppfylles. De resterende fem leverandørene ble så invitert til et første dialogmøte og deretter til å levere nytt løsningsforslag. Det var i denne fasen kravet om å levere en markedsplass kom inn i kravspesifikasjonen. En av de gjenværende fem leverandørene kunne ikke tilfredsstille dette kravet, og dermed sto man igjen med fire leverandører som ble invitert til å levere ytterligere to løsningsforslag og delta i to dialogmøter og deretter til å inngi prøvetilbud før dialogen ble avsluttet og endelig tilbud levert inn. Fristen var 23. mai 2022. Etter at endelig tilbud var levert, viste det seg at en av leverandørene ikke kunne tilfredsstille et obligatorisk krav om lagringssted for krypteringsnøkler. De resterende tre tilbudene ble deretter vurdert på grunnlag av de tildelingskriterier som var oppgitt i endelig versjon av konkurransebestemmelsene. Kravene til kvalitet hadde flere elementer: generelle krav, funksjonelle krav, tekniske krav og krav til implementeringsplan (inkludert service-nivå (SLA)).

Kriteriet for vurdering av pris var kostnad per undersøkelse, spesifisert for volumterskler, slik at priser kunne differensieres avhengig av volum. Estimert totalkostnad per algoritme ble beregnet som et produkt av enhetspris og antatt volum (antall undersøkelser). Samlet sett for alle de fem algoritmene som ble evaluert, utfra estimert volum og enhetspris, var kostnadsforskjellen mellom høyeste og laveste tilbud ca. 25 prosent.

Prosjektledelsen har foreløpig oppsummert følgende læringspunkter knyttet til anskaffelsesprosessen:

- Det hadde vært ønskelig å starte hele prosessen med å arrangere en leverandørkonferanse der leverandørene ble invitert til både å presentere sine produkter, redegjøre for arbeidsmåter og hvordan de ville gjennomføre implementeringen. En slik fremgangsmåte ville på et tidligere stadium i prosessen gitt en bedre forståelse av markedet for produkter og tjenester, og kunne dermed redusert noe av den tiden som gikk med til avklarende dialoger med leverandørene. Dermed ville man også vært i stand til å spisse den opprinnelige forespørselen på måter som traff markedet bedre.
- Det hadde videre vært ønskelig at initiativet om å invitere samtlige helseforetak i Helse Sør-Øst til å omfattes av anskaffelsen, kom tidligere. Ved en slik utvidelse av mulige brukere/kunder, måtte prosjektledelsen sikre at løsningene som skulle anskaffes var kompatible med alle de ulike RIS- og PACS-løsninger som fins i HFene i Helse Sør-Øst.
- Selv om konkurransepreget dialog har vært en transaksjonskrevende anskaffelsesform, har den bidratt til læring og tilpasning underveis i prosessen, noe som vurderes som verdifullt når man er «først i løypa».

- I både utforming av kravspesifikasjon og vurdering av tilbudene, har samarbeidet mellom Vestre Viken, Sykehuset Vestfold, Sykehuspartner og Sykehusinnkjøp vært nyttig og nødvendig.

Prosjektledelsen er nå i ferd med å utforme en implementeringsplan som starter med en parallell gjennomføring av ROS-analyse, utforming av løsningsdesign og – i første omgang - utarbeidelse av mini-HTA for å ta i bruk noen av de fem algoritmene som er anskaffet. Deretter installeres plattformen (stipulert til en måneds arbeid), før en eller flere algoritmer installeres, først gjennom en fase med utprøving på retrospektive lokale data, før neste fase med prospektiv utprøving i parallell med ordinær drift. Ambisjonen er å være i gang med sistnevnte fase før årets utgang. Det er allerede etablert kontakt mellom Vestre Viken og andre HF i Helse Sør-Øst som har vist interesse for å gjøre avrop og samarbeide om implementeringen. I denne forbindelse vurderes det også å etablere et nettverk for erfaringsutveksling mellom alle de helseforetak som berøres av avtalen gjennom å planlegge- eller gjøre avrop for anskaffelse.

DEL 5: Arbeidsgruppens anbefalinger

9 Anbefalinger

Flere av aspektene i fase 0-3 i den radiologiske arbeidsflyten (se kap.4) treffer behovene som er beskrevet tidligere i denne rapporten. Vårt arbeid viser imidlertid at tilbudssiden er forholdsvis umoden når det gjelder mange av oppgavene i disse fasene. Arbeidsgruppen har heller ikke klart å oppdrive en samlet oversikt over markedet for fase 0-3 (for fase 4 finnes AIforRadiology.com). I tillegg er det slik at oppgavene i fase 1-3 er tett tilknyttet maskinvaren som skal brukes for å gjennomføre skanningen. Dette innebærer at en investering i ny maskinvare for billedtaking også vil være førende for hvilke muligheter som åpner seg mht. å få hjelp av KI til fase 1-3.

Samlet har disse faktorene ført til at anbefalinger rundt bruk av KI i fase 0-3 er blitt lavere prioritert i denne rapporten. Likevel vil arbeidsgruppen påpeke at disse faktorene bør vies oppmerksomhet når helseforetakene i nord i tiden framover skal fornye og oppgradere maskinparken. Dette feltet er også i rivende utvikling, og derfor må en også ha i mente at forutsetningene kan endre seg raskt. Hovedfokuset for dette arbeidet har derfor vært rettet mot markedet for KI-algoritmer for tolkning av bilder (fase 4).

Arbeidsgruppen vil påpeke at anvendelsen av KI-liknende analyseverktøy (CAD - Computer aided detection) som støtte ved tolkning av radiologiske bilder ikke er ny. Allerede på 1990- og 2000-tallet ble det tatt i bruk produkter som skulle bistå radiologer, f.eks. innenfor mammografi og thoraxradiologi. Imidlertid innfridde ikke disse løsningene de forventninger som ble skapt. Kort sagt, viste både erfaringer og studier at radiologene ikke fikk ønsket utbytte av å anvende CAD: Radiologens nøyaktighet kunne bli påvirket i negativ retning, føre til økt tidsbruk, og resultatet kunne føre til økte antall biopsier. En vesentlig andel av forklaringen lå i at CAD i de studiene med lovende resultat ble brukt i en klar definert arbeidsflyt, noe som ikke ble fulgt i dagligdags bruk av CAD.

Den utviklingen vi nå står foran, må ses i lys av denne erfaring. Samtidig som moderne maskinvare og bedre KI-algoritmer har gitt en renessanse til området, må vi ha realistiske forventinger til hva KI kan på kort sikt kan bidra med; det gjelder særlig med hensyn til helheten i den diagnostiske prosessen.

9.1 Generelle vurderinger av markedet for KI-støttet bildetolkning

Gitt erfaringene fra anskaffelsesprosjektet i Vestre Viken, ser denne arbeidsgruppen det også som hensiktsmessig å velge en felles plattform for håndtering av de algoritmene

som anskaffes og tas i bruk. Det fins i dag over 200 CE-merkede KI-algoritmer innen radiologi, mens det er registrert i overkant av 30 ulike plattformer. For ca. en tredel av algoritmene fins det oppfølgende valideringsstudier; det innebærer at de fleste algoritmer mangler kvalifisert ekstern og uavhengig validering. Foreløpig er det vanskelig å finne rapporter eller studier som evaluerer kvaliteten på plattformene. Det kan ha sammenheng med at de fleste plattformene enten er nylig utviklet eller under utvikling, og at det derfor foreligger begrenset erfaring med hvordan de fungerer i praksis. Selv om markedet for plattformer foreløpig kan synes å være noe umodent, vurderes dette som mindre kritisk mht. risiko, sammenlignet med valg av algoritmer. De foreløpige erfaringer fra anskaffelses-prosessen i Vestre Viken tilsier videre at det er relativt små forskjeller mellom de ulike plattformer mht. funksjonalitet.

De enkelte plattformleverandørene har til dels utviklet egne KI-algoritmer, men de har for en stor del også tredjepartsalgoritmer i sin portefølje. Mange av algoritmene er tilgjengelige hos flere plattformleverandører, men det er også forskjeller mellom leverandørene mht. hvilke og hvor mange KI-algoritmer de har i sin portefølje. De tre leverandørene som var med i sluttrunden i Vestre Viken, hadde mellom 20 og 32 algoritmer i sin portefølje.

De store og dominante leverandørene av maskinvare har alle utviklet egne plattformer, noe som også gjelder flere av PACS-leverandørene. Samlet innebærer dette at man ikke kan begrense seg til å vurdere enkeltprodukter i markedet, men også må ta hensyn til hvordan de ulike elementer i hele konfigurasjonen (maskinvare, PACS, RIS, plattform og algoritmer) spiller sammen. Siden PACS-et er radiologenes viktigste arbeidsflate, vil eksempelvis integrasjonen mellom PACS og KI-plattform vurderes som viktig, men uten at dette hensynet til integrasjon nødvendigvis tilsier valg av samme leverandør. Vestre Viken HF har riktignok valgt plattformtilbudet fra sin nåværende PACS-leverandør, men det vises også til eksempel i Nederland der samme plattform er integrert med annet PACS[34].

Oppsummerende er det arbeidsgruppens vurdering at det i løpet av de siste 2-3 årene har skjedd en betydelig utvikling av markedet for KI-algoritmer innen radiologi, både i form av antall leverandører og produkter, men antagelig også mht. kvalitet. Dette gjelder særlig om markedet avgrenses til CE-merkede algoritmer i klasse 2a og 2b (eksternt vurdert av «notified body»). I tillegg forventes de skjerpede regulatoriske krav som følger av MDR (se kap. 13), også å bidra til å heve kvaliteten på løsningene. En videreutvikling av markedet i retning av mer «modne» produkter, forutsetter antagelig også en mer utprøvende aktivitet i helsetjenesten, i form av begynnende validering i klinisk relevante settinger, basert på både retrospektive og prospektive data. Dette kan gi grunnlag for storskala implementering av KI-løsninger i samspill med fagmiljøene i klinikken, gjerne fulgt opp av uavhengige studier. Kunnskap om hvordan importerte

[34] Meddelelse fra prosjektledelsen i Vestre Viken/Vestfold-prosjektet.

kommersielle KI-løsninger innen radiologi virker når de anvendes på norske pasientpopulasjoner, kan bare erverves gjennom mer aktiv utprøving i helseforetakene. Dette hensynet er en selvstendig begrunnelse for å komme i gang med utprøving.

I vurderingen av markedet, er det ikke bare krav til produktenes tilgjengelighet, funksjonalitet og kvalitet som skal vektlegges. Pris og prismodeller er også viktige hensyn som må tillegges vekt. Foreløpig er volumet av anskaffelser såpass begrenset at prisfastsettelsen ikke nødvendigvis avstemmes ensidig av tilbuds- og etterspørsels-mekanismer; det innebærer at leverandørene vil kunne være åpne for å forhandle om pris. Erfaringene fra anskaffelsen i Vestre Viken viser at det særlig blant finalistene var signifikante prisforskjeller. Her ble samtlige leverandører invitert til å beregne pris per undersøkelse, med synkende pris for volumer over gitte terskler. For å vurdere samlet kostnad for å anskaffe og ta i bruk ulike algoritmer, må man følgelig estimere hvilke volum de enkelte algoritmer forventes å håndtere; her er det store forskjeller mellom ulike algoritmer innrettet mot ulike typer av undersøkelser. I vurderingen av terskler for volum ved anskaffelse av algoritmer, må det også tas hensyn til undersøkelsens kompleksitet og den radiologtid som medgår til hver enkelt undersøkelse.

9.2 Prioriterte fagområder innen KI-støttet bildetolkning

Arbeidsgruppen har sammenstilt fagområdene hvor det finnes behov (jfr. Del 3) med hvilke områder som er dekket på tilbudssiden (jfr. Del 4) og funnet at innen følgende fagområder er det overlapp:

- *CT thorax*:
  - ✓ Utredninger og kontroller av lungenoduli, lungeemboli, lungesykdommer (emfysem, fortetninger).
  - ✓ Elektiv og ø-hjelp.
  - ✓ Triagering både på vakt og til elektive undersøkelser.
  - ✓ Relevant for alle HF i Helse Nord.

- *MR prostata*:
  - ✓ Utredning og kontroller av prostatakreft.
  - ✓ Elektive undersøkelser
  - ✓ Hovedmengden på UNN HF gjøres av private aktører, men slike undersøkelser gjøres i de andre HFene og er svært relevant der.

- *RG mammografi*:
  - ✓ Utredninger og kontroller av brystkreft ifbm. screening.
  - ✓ Elektive undersøkelser.
  - ✓ Det pågår et eget løp for KI og screening i regi av Kreftregisteret. Det anbefales derfor å avvente resultatene av dette arbeidet som forventes å bli premissgivende for nasjonale løsninger på dette fagområdet.

- ✓ Relevant for NLSH og UNN.

- MR mamma:
  - ✓ Utredning og kontroller av brystkreft
  - ✓ Regelmessige undersøkelser av kvinner med genfeil
  - ✓ Spesielle tilfeller som f eks proteseruptur, «cancer of unknown primary»
  - ✓ Relevant for NLSH og UNN.

- *MR hjerte*:
  - ✓ Patologideteksjon, kvantifisering og segmentering.
  - ✓ Elektive undersøkelser
  - ✓ Det varierer hvor relevant dette er. UNN har kjøpt en maskin som har en innebygd software som allerede gjør at det spares mye tid. Ved NLSH er slike undersøkelser fortsatt meget tidskrevende, noe som gjør det svært relevant der. Ikke relevant for lokalsykehusene.

- *Konvensjonell røntgen* for *skjelett*:
  - ✓ Bruddeteksjon og triagering av undersøkelser
  - ✓ Øyeblikkelig hjelp
  - ✓ Relevant for alle HF.

- *Konvensjonell røntgen for thorax*:
  - ✓ Påvise patologi og tolkningshjelp.
  - ✓ Elektiv og ø-hjelp.
  - ✓ Triagering både på vakt og til elektive undersøkelser.
  - ✓ Relevant for alle HF.

- *MR/CT caput*:
  - ✓ Slag (okklusjon av kar, blødning og infarkt). Må vurderes om det bør avventes til resultat av pågående regional prospektiv studie foreligger før det eventuelt anskaffes en algoritme for dette[35].
  - ✓ Multippel sklerose (MS).
  - ✓ Demens.
  - ✓ Triagering for CT caput.
  - ✓ Elektiv og ø-hjelp
  - ✓ Relevant for alle HF. (MR/CT caput er høyvolumundersøkelser, men det er noe usikkert hvor stort volumet er for disse tre konkrete kliniske problemstillingene)

[35] https://www.spki.no/prosjekter/improving-stroke-care-in-north-norway-through-artificial-intelligence/

Arbeidsgruppen har gjort en prioritering av de overnevnte områdene basert på en rekke kriterier (som har vært drøftet i de ulike delkapitlene i denne rapporten). Herunder nevnes

- Grad av arbeidsbesparelse (opp mot volum, tid, brukervennlighet),
- Kvalitet (Mulighet for forbedret diagnostikk, forslag til lignende kasus)
- Grad av multifunksjonalitet
- Grad av evidens (studier, erfaringer fra andre),
- Pasientperspektivet (hvor akutt? hva hjelper pasienten mest?),
- Kost-nytte (vanskelig å vurdere på forhånd og få tilgjengelige studier),
- Regulatorisk betraktning (helst MDR-godkjenning),
- Kompleksitet (hvor lett å ta i bruk, hva kreves mht. opplæring, lokale forhold i de ulike fagmiljøene, er det gode sjanser for å få til en suksesshistorie?)
- Pågående prospektive valideringsstudier i regionen (mammografi og slag)

Basert på disse kriteriene har arbeidsgruppen kommet fram til følgende prioriteringsliste:

1. *Brudd: deteksjon og triagering (RG skjelett)*
   Positivt: Stort volum, god evidens, akutt, MDR-godkjente løsninger, lav kompleksitet. Også til hjelp for klinikere på vakt.
   Negativt: vanskelig å tallfeste kost-nytte.

2. *Kreftdiagnostikk*
   a. *Diagnostikk av lungekreft, inkludert tilfeldig påviste noduli (CT thorax)*
      Positivt: Tidsbesparelse, bedret diagnostikk, flere MDR-godkjente produkter, noen algoritmer har mulighet for multifunksjonalitet.
      Negativt: Ganske god evidens, men noe uklart med kost-nytte. Høyere kompleksitet enn bruddeteksjon.
   b. *Diagnostikk av prostatakreft (MR prostata)*
      Positivt: Tidsbesparelse, bedret diagnostikk, 2-3 aktuelle produkter er MDR-godkjente,
      Negativt: Middels grad av evidens og noe uklart med kost-nytte. Høyere kompleksitet enn bruddeteksjon.
   c. *Diagnostikk av brystkreft (MR mamma)*
      Positivt: Tidsbesparelse, bedret diagnostikk, en viss grad av multifunksjonalitet.
      Negativt: Kun ett aktuelt produkt per nov. 22. Ikke MDR-godkjent. Lav evidens og noe uklart med kost-nytte.
      NB: RG Mammografi kunne også vært høyaktuelt, men det anbefales å avvente til resultater av nasjonale studier i regi av Kreftregisteret foreligger

3. *Tolkning, deteksjon av patologi og triagering for røntgen thorax*
   Positivt: Høyt volum, både akutt og elektivt, MDR-godkjent, mulighet for

multifunksjonalitet. Også til hjelp for klinikere på vakt.
Negativt: Noe uklar grad av evidens og kost-nytte.

4. *Slagdeteksjon (CT caput)*
   Positivt: Høyt volum, raskere deteksjon, akutt, middels kompleksitet
   Negativt: MDD-godkjent, noe lav evidens, kost-nytte.
   Anbefales å avvente til resultater av studie[36] foreligger.

5. *Deteksjon og triagering av lungeemboli (CT thorax)*
   Positivt: Middels volum, men tas på vakt (akutt),
   Negativt: MDD-godkjent, noe lav evidens, kost-nytte.

6. *Tolkning av lungefortetninger/ ILD (CT thorax).*
   Positivt: Høyt volum, MDR-godkjent, multifunksjonalitet
   Negativt: Kun en aktuell løsning på markedet pr nå og kun en studie

7. *Tolkning av MR hjerte*
   Positivt: Potensiale for tidsbesparelse, flere funksjoner, MDR-godkjent, flere produkter tilgjengelig, ikke veldig høy kompleksitet
   Negativt: Middels grad av evidens. Middels volum og kun relevant for NLSH.

9.3 Anbefalinger for UNN HF som har opsjon på avrop etter rammeavtalen som er inngått i Helse Sør-Øst

Som allerede nevnt, har UNN HF opsjon for å gjøre avrop på den rammeavtalen om anskaffelse av KI-algoritmer som Vestre Viken HF har inngått med valgt leverandør. Siden ingen av de øvrige helseforetak i Helse Nord har slik opsjon, er det følgelig ikke mulig for disse å gjøre avrop for innkjøp av de algoritmer (inklusive plattform) som inngår i denne avtalen. Arbeidsgruppen vil her begrunne hvorfor det – under gitte forutsetninger - likevel kan være klokt å la UNN HF gjøre slikt avrop, selv om øvrige helseforetak i Helse Nord ikke kan inkluderes. Det presiseres at UNN HF som eget rettssubjekt står fritt til å gjøre avrop i henhold til nevnte avtale, uavhengig av hva Helse Nord RHF måtte mene om det. Samtidig vurderer arbeidsgruppen det som både ønskelig og hensiktsmessig at slike initiativ ved UNN skjer i samforstand med Helse Nord RHF og øvrige helseforetak i Helse Nord.

For det første vil arbeidsgruppen generelt anbefale at UNN HF – med sin nærhet til betydelige kompetansemiljøer innen kunstig intelligens og i kraft av sin funksjon som universitetsklinikk - starter med å implementere en eller flere løsninger før vi går videre

[36] https://www.spki.no/prosjekter/improving-stroke-care-in-north-norway-through-artificial-intelligence/

med implementering i øvrige helseforetak. Dermed kan man vinne erfaring som gjør det enklere å implementere slike løsninger øvrige HF. En slik trinnvis strategi som starter med UNN HF, anbefales også ved oppstart på grunnlag av en felles anskaffelse som inkluderer alle fire helseforetak.

Motargumentet for å la UNN gjøre avrop på Vestre-Viken-avtalen, er at man da kan risikere å få andre løsninger ved UNN HF enn de som senere måtte komme gjennom en felles anskaffelsesprosess. Men siden det dreier seg om et begrenset avrop – i form av 1-2 algoritmer – representerer dette en begrenset binding for framtiden. Det er neppe nødvendig å anskaffe en plattform for å ta i bruk 1-2 algoritmer, og dessuten følger det ingen stor kostnad ved anskaffelse av en plattform.

Den store fordelen ved å la UNN HF gjøre avrop på ovennevnte rammeavtale, er at man ikke blir liggende for lenge i planleggingsmodus, men kommer i gang med praktisk og klinisk rettet utprøving og implementering. Dermed vil Senter for pasientnær kunstig intelligens (SPKI) raskere utvikle sin kjernekompetanse knyttet til implementering, noe som er til nytte for hele foretaksgruppen. Ved at avrop tidsmessig skjer i nær tilknytning til tilsvarende implementeringer i andre deler av landet, får man også del i et felles læringsmiljø som vil berike utviklingen i Helse Nord. Det er allerede planlagt et felles nettverk for erfaringsutveksling mellom de helseforetak som gjør avrop på den inngåtte rammeavtalen. Vi ser det som svært nyttig om UNN HF og SPKI kunne bli en aktiv deltaker i dette nettverket.

Anbefalingen om å la UNN HF gjøre avrop, bygger imidlertid på noen forutsetninger som må oppfylles.

- For det første må både toppledelsen ved UNN HF og ledelse/fagmiljø ved Røntgenavdelingen stille seg bak en slik anbefaling. Det er generelt også viktig for å lykkes, at det miljøet som skal stå for implementeringen er motiverte for oppgaven og disponerer de ressurser som er nødvendige for gjennomføring
- Dernest må de algoritmene som inngår i rammeavtalen være relevante og prioriterte. I dette tilfellet er det heldigvis slik at samtlige algoritmer som inngår i denne rammeavtalen dekker områder som er prioritert av de radiologene som er medlemmer av arbeidsgruppen (Jf. kap. 6 foran i rapporten).
- Da gjenstår det i egenregi å kvalitetssikre hvilke av de algoritmene som inngår i rammeavtalen[37] som tilfredsstiller de kvalitetskriterier som er lagt til grunn i denne rapporten. Vi konstaterer at bare en av de fem valgte algoritmene er CE-merket etter de nye kravene i MDR (Jf. Kap. 13). To av algoritmene er i risikoklasse I, hvilket innebærer at de tidligere ikke har vært eksternt vurdert av «notified body». Det bør derfor sjekkes om det foreligger planer om slik godkjenning for de øvrige fire (MDD-godkjent algoritmene). (Det konstateres at

[37] Rammeavtalen dekker ikke bare de fem algoritmene som lå til grunn for forespørselen fra Vestre Viken, men hele porteføljen hos valgt leverandør bestående av 22 algoritmer.

leverandør av valgt algoritme for brudd-deteksjon allerede har sendt søknad om godkjenning etter MDR)

- Videre bør det sjekkes i hvilken grad algoritmene er eksternt validert gjennom eksterne uavhengige studier, og gjerne på ulike datasett.
- I den grad det gjøres avrop på algoritmer som berører UNN HFs regionfunksjoner – og dermed pasienter tilhørende andre HF i Helse Nord – bør det avklares i hvilken grad opsjonen i avtalen hjemler at øvrige HF som deltar i behandlingen av pasientene gis tilgang til de resultater som fremkommer ved bruk av algoritmen.
- En oppstart på UNN HF bør ledsages av følgeforskning, med vekt på dokumentasjon av både prosesser og resultater, der både logistikk, diagnostikk og effekter på klinisk bruk, er mulige endepunkt.
- Fagmiljøer i øvrige helseforetak bør inviteres til å følge med i en eventuell oppstart ved UNN HF, eksempelvis gjennom en referansegruppe der erfaringer formidles.

I den grad ovennevnte forutsetninger kan ivaretas, vil arbeidsgruppen anbefale at UNN HF prioriterer anskaffelse av den algoritmen for brudd-deteksjon[38] som inngår i rammeavtalen med Vestre Viken HF.

9.4 Anbefalt anskaffelsesprosedyre og hensyn som bør vektlegges i en kravspesifikasjon

Det presiseres at dette delkapitlet utelukkende må forstås som overordnete og generelle anbefalinger for en senere anskaffelsesprosess. Når det eventuelt gjøres vedtak i Helse Nord RHF om å starte en anskaffelsesprosess, skjer det følgelig med utgangspunkt i etablerte rutiner der ulike funksjonelle enheter og tilhørende kompetanser (helsefag, teknologi, juss, forretning og økonomi) samarbeider om hele prosessen. Det innebærer at de konkrete og detaljerte vurderinger som gjøres ved en eventuell anskaffelse, må tilpasses rådende omstendigheter og rammer på det aktuelle tidspunkt.

9.4.1 Anskaffelsesprosedyre

Arbeidsgruppen bygger her sine anbefalinger primært på den gjennomgangen av anskaffelsen i Vestre Viken HF som det tidligere er redegjort for i kap. 8.2. For å oppdatere denne kunnskapen, foreslås at Helse Nord innhenter erfaringer fra det videre forløp ved implementeringen i Vestre Viken, slik at også lærdommer i etterkant av

[38] Denne algoritmen har foreløpig bare godkjenning etter MDD, men det er sendt søknad til Notified Body for å få den godkjent etter de nye kravene i MDR. Det foreligger 6 fagfellevurderte studier av denne algoritmen. Her vises særlig til den ferskeste publikasjonen som viser at sensitiviteten ved bruk av algoritmen økte med 8-10 prosentpoeng sammenlignet radiologer, mest for radiologerunder utdanning. Nguyen T, Maarek R, Ducou H, et al. Assessment of an artificial intelligence aid for the detection of appendicular skeletal fractures in children and young adults by senior and junior radiologists, *Pediatric Radiology* **52**, 2215-2226 (2022)

anskaffelsen kan fanges opp. Det vises ellers også til kap. 9.4 i dokumentet *Strategi for kunstig intelligens i Helse Nord 2022-2025,* der regelverk for anskaffelse av kommersielle KI-løsninger er omtalt.

Vurderinger både fra prosjektgruppen i Vestre Viken HF og Sykehusinnkjøp HF – som bisto prosjektet med innkjøpskompetanse – synes å bekrefte at *konkurransepreget dialog* har vært en egnet anskaffelsesform for denne type anskaffelse. Siden markedet for slike produkter både er nytt og i rask endring, gir denne anskaffelsesformen gjennom en flertrinnsprosess mulighet for læring og tilpasning underveis. Arbeidsgruppen vil derfor anbefale Helse Nord å vurdere denne anskaffelsesformen ved anskaffelse av kommersielle KI-algoritmer innen radiologifeltet. På grunnlag av erfaringer fra Vestre Viken-prosjektet, vil vi også anbefale at det i forkant av utlyst konkurranse arrangeres en *leverandørkonferanse* for å skaffe seg mer oppdatert kunnskap om produktmarkedet og leverandørenes evne til å levere understøttende tjenester.

En eventuell anskaffelse anbefales gjennomført samlet for hele foretaksgruppen i Helse Nord, med mulighet for hvert av HFene til å gjøre avrop på rammeavtale innen en 2-4 årsperiode. I den grad forholdene måtte ligge til rette for det (ved sammenfall i tid mv.), kan det vurderes å samarbeide om en anskaffelse med helseforetak utenfor Helse Nord.

For å ivareta behovet for innkjøpskompetanse på best mulig måte, anbefales etablert et best mulig samarbeid mellom ansvarlige for innkjøp i Helse Nord RHF og Sykehusinnkjøp. Den erfaring Sykehusinnkjøp har ervervet gjennom medvirkning i anskaffelsen i Vestre Viken HF, vurderes som så verdifull at den bør komme en eventuell anskaffelsesprosess i Helse Nord til gode.

9.4.2 Hensyn som bør vektlegges i en kravspesifikasjon

Det presiseres også her at arbeidsgruppen ikke har noen intensjon om – enn si noe mandat til - å utarbeide noen tekst for kunngjøring eller kravspesifikasjon for en mulig anskaffelse. Her avgrenser vi oss til å presentere noen hensyn som anbefales vektlagt i en kunngjøring respektive kravspesifikasjon ved en framtidig anskaffelse. I tråd med de problemstillinger som er presentert og drøftet i denne rapporten, vil fokus her primært rettes mot noen utvalgte funksjonelle og tekniske krav til produkt og leverandør. Det innebærer at mer formelle juridiske og forretningsmessige sider ved en kravspesifikasjon ikke berøres her.

- De regulatoriske endringer som følger av MDR (Medical Device Regulative), tilsier at algoritmer som inngår i en anskaffelse må være sertifisert i henhold til dette regulativet før kontrakt signeres. Dette bør fremgå av kravspesifikasjonen
- Foretrukket prismodell (volum, enhetspris, abonnement, mv.)

- Med forbehold om vesentlige endringer i Helse Nords behovsvurderinger (Jf. kap. 6) eller endringer i tilbudet av produkter (jf. Kap. 8), anbefales rapportens prioriterte produktområder (Jf. Kap. 9) å legges til grunn for anskaffelsen.
- Det anbefales at anskaffelsen skjer gjennom en plattformleverandør med en relevant (Jf. prioriterte produktområder) og kvalitativt god portefølje av KI-løsninger, inkludert løsninger fra tredjepartsleverandører.
- Det må redegjøres konkret og spesifikt for algoritmens anvendelsesområde og hvordan den forutsettes brukt
- For å kvalifiseres til å delta i konkurransen skal leverandør dokumentere hvilke data som ligger til grunn for respektive utvikling og påfølgende testing av algoritmen.
- Det er ønskelig at leverandør kan vise til studier – gjerne i fagfellevurderte publikasjoner - der algoritmene er evaluert på grunnlag av statistiske mål som sensitivitet og spesifisitet. Grad av arbeidsbesparelse og forbedret kvalitet
- Algoritmer som er multifunksjonelle foretrekkes dersom de scorer likt på andre kriterier
- Det forutsettes god og enkel integrasjon mellom plattformen og etablert PACS-løsning.
- Det må redegjøres for krav til IKT-infrastruktur (skyløsninger, mv..) og hvordan personvern og informasjonssikkerhet ivaretas.
- Det må dokumenteres hvordan de tilbudte løsningene kan innpasses i radiologenes arbeidsflyt, med vektlegging av bl.a. strukturert rapportering.
- I løpet av anskaffelsesprosessen forventes at leverandører som kvalifiseres til å delta i konkurransen - i løpet av en nærmere angitt periode (eksempelvis tre måneder) - stiller aktuelle algoritmer til disposisjon for kjøper og legger til rette for validering mot lokale data.
- Mulighet for implementering av forskningsutviklede algoritmer på innkjøpt plattform bør telle positivt
- Det må redegjøres for egne prosesser for validering og innfasing av nye produkter

9.5 Fagområder der egenutvikling kan være aktuelt

Det er en rekke av behovene som er beskrevet i denne rapporten som ikke dekkes godt nok av det eksisterende tilbudet på markedet og der egenutvikling kan være aktuelt. Dette gjelder særlig disse behovene:

- Talegjenkjenning
- Triagering av henvisninger.
- Støtte til både henviser og radiologisk avdeling for å sikre at riktig og nødvendig informasjon kommer på henvisning. Herunder er det aktuelt med støtte til å hente inn relevante kliniske opplysninger fra journal og til å vurdere om det er indikasjon for undersøkelse.

- Forbedret pasientlogistikk (lettere kontakt med pasient, mm.)
- På generell basis: flere løsninger som kan bidra til fase 0-3 i den radiologiske arbeidsflyten.

Fremtidige KI-løsninger som eventuelt skal kunne bidra til å dekke disse behovene vil nødvendigvis måtte handtere andre inputdata enn bare bilder – og gjerne multimodale data (data fra flere kilder, f.eks. pasientjournal, selvrapporterte data og bilder). I dagens kommersielle marked finnes det svært få løsninger som handterer multimodale data. Dette er delvis en konsekvens av at de forskjellige datakildene typisk er organisert i siloer, noe som vanskeliggjør tilgangen. Dessuten er det vanskeligere å lage algoritmer som er gode nok til å utnytte heterogene data. Det er likevel en forutsetning for å lykkes med persontilpasset medisin, at man er i stand til å lage beslutningsstøtteverktøy som kan utnytte slike data. Dette bør derfor være et prioritert forskningsområde.

Det høyeste rangerte behovet er å bidra med støtte til å markere områder med patologiske funn i bilder; her nevnes særlig deteksjon, måling og kontroll av lesjoner, tumorer og metastaser for forskjellige modaliteter og organgrupper, samt tolking av skjelettrøntgen. En god del av dette behovet synes å være dekt av det kommersielle markedet, men mange av disse løsningene bør valideres grundigere.
De organgruppene og modalitetene som er nokså godt dekt er:

- Neuro (CT, MR)
- Thorax (røntgen, CT)
- MSK (røntgen)
- Hjerte (MR, CT)
- Mamma (MR og RG mammografi)

Her er det likevel nødvendig med en presisering: det faktum at disse områdene er dekt av kommersielle løsninger, betyr ikke nødvendigvis at løsningene er gode nok. For mange av løsningene er det fortsatt manglende evidens. Dette innebærer at også flere av disse områdene bør prioriteres i videre FoU-arbeid.

Abdomen er et eksempel på et stort og viktig område som er dårlig dekt. Foreløpig finnes det kun løsninger for MR prostata. Støtte til deteksjon av tumorer i organer som eksempelvis lever finnes det ikke gode nok kommersielle verktøy for.

I tillegg ser vi at det finnes svært få kommersielle løsninger innen nukleærmedisin (PET) og få gode løsninger for ultralydbilder (f.eks. UL thyreoidea). Det anbefales derfor at disse områdene prioriteres i videre FoU-arbeid.

Som nevnt i kap. 6.3, er medisinsk bildeanalyse et stort og omfattende fagfelt. I denne rapporten har vi konsentrert oss om bildemodaliteter som beskrives av radiologer. Innen de andre områdene ser man at markedet er mindre modent enn innen radiologi. Det skal likevel nevnes at det innen noen av disse områdene finnes flere produkter på markedet med brukbar evidens. Spesielt gjelder dette oftalmologi (diabetisk retinopati),

kardiovaskulær medisin og nevrologi (begge de to sistnevnte er delvis overlappende med radiologi og derfor i noe grad dekket av AIforradiology). De to førstnevnte faller utenfor mandatet til denne arbeidsgruppa. Vi vil likevel anbefale at behovet for KI innen disse områdene kartlegges nøyere av de respektive fagmiljøene.

Innen andre områder som dermatologi, odontologi, strålingsonkologi, gastroenterologi, gynekologi og digital patologi finnes det få gode kommersielle løsninger. Den senere tid har imidlertid vist at innovasjons- og FoU-arbeid kan lede til gode løsninger innen f.eks. strålingsonkologi[39] og digital patologi[40]. Flere av disse områdene bør også prioriteres i FoU-arbeid i Helse Nord.

Et annet prioritert forskningsområde bør være å kombinere medisinske bilder med helsedata fra andre modaliteter (pasientjournal, registre, mv..). En forutsetning for å lykkes med reell persontilpasset medisin, vil være at man klarer å utvikle slike multi-modale verktøy som utnytter all tilgjengelig informasjon om pasienten.

Til sist nevnes at FoU-aktivitetene ikke bare bør prioritere utvikling av verktøy. Kanskje like viktig er (følge)forskning på utprøving, validering og implementering. Slik kompetanse er viktig å bygge opp i fagmiljøer som SPKI.

[39] https://helse-mr.no/om-oss/nyheiter/nyheiter-2022/suksess-med-kunstig-intelligens-i-stralebehandlinga
[40] https://www.domore.no/

DEL 6: Andre anbefalinger og forhold

I den siste delen av rapporten tar vi for oss andre anbefalinger og forhold som har betydning for arbeidet med implementering og bruk av KI-løsninger innen radiologi i Helse Nord.

Dette dreier seg om

- Valg av plattform for KI-løsninger innen radiologi
- Betydningen av god logistikk og arbeidsflyt
- Krav til IKT-infrastruktur, ROS-analyser og informasjonssikkerhet
- Konsekvenser av den regulatoriske endringen med innføring av MDR.
- Analyse av kost-nytte og erfaringer med implementering nasjonalt og internasjonalt
- Behov for kompetanseoppbygging og utdanning
- Behov for rutiner og veiledning knyttet til validering og kvalitetssikring av KI-verktøy
- Forvaltning av KI-løsninger

10 Valg av plattform som strategisk beslutning

De store PACS- og maskinleverandørene holder nå på å posisjonere seg slik at de også kan tilby en KI-plattform (økosystem/appstore for KI-verktøy). I tillegg har enkelte spesialiserte selskaper innen medisinsk kunstig intelligens begynt å tilby slike plattformer, slik at det i dag er ca. 30 ulike plattformer i markedet. Disse KI-plattformene baserer seg på en kombinasjon av tredjepartsprodukter og egenutviklede løsninger. I prinsippet er økosystemene åpne for alle underleverandører, men i praksis vil nok markedsmessige og strategiske hensyn medføre at det vil variere hvilken type produkter som er tilgjengelige hos de ulike leverandørene av økosystemer. Det kan også variere hvilke transaksjonskostnader som følger av å kombinere ulike løsninger.

Ved en anskaffelse av kommersielle KI-løsninger vil man derfor måtte ta stilling til om man også skal gå til anskaffelse av en KI-plattform eller om man skal starte i mindre skala og gå for enkeltstående KI-verktøy (algoritmer). Det er en rekke ulike faktorer som bør tas hensyn til ved en slik vurdering. Vi redegjør for noen av de i det følgende:

- *Samspillseffekter og kompatibilitets-/kalibreringsutfordringer mellom komponenter (scanner, PACS, plattform, KI-algoritmer) fra ulike leverandører.* Er det problematisk å bruke ulike leverandører for maskinvare, PACS og plattform mht. KI-algoritmenes ytelse? Et faktum som taler for at dette kanskje ikke bør problematiseres for mye, er at etablerte standarder for dataflyt gir gode forutsetninger innenfor radiologi: Dataflyten mellom skanner, arkiv og programvare for bildevisning er basert på DICOM-standarden; denne sier noe om måten bildeformatet er, hvordan utstyr, arkiv og programvare kommuniserer innbyrdes, og hvordan informasjonselementer knyttet til bilder og undersøkelser, lagres. Slike informasjonselementer kan være f.eks. undersøkelsestidspunkt og skanner-innstillinger, men i tillegg kan også annoteringer i bilder eller strukturerte tekstrapporter lagres på denne måten. KI-algoritmer kan benytte denne informasjon i tillegg til å selv generere og lagre den, f.eks. ved å legge inn rapporttekst eller annoteringer i bilder. DICOM-standarden gjør integrasjon av KI-løsninger mindre avhengig av spesifikke leverandører, enn tilfellet er i andre deler av helsetjenesten. Til sammenligning er valg av standarder for lagring av digitale patologisnitt et særegent hinder for utvikling og implementering av ulike KI-produkter på dette feltet.
- Det er betydelig *variasjon mht. hvor stort antall produkter som er tilgjengelig på de ulike plattformene.* Et viktig hensyn å ta er hvor mange tredjepartsløsninger som tilbys.
- *Arbeidsflyt, standardisering, brukergrensesnitt:* Hvor enkelt er det å bruke plattformen og KI-applikasjonene? Er det muligheter for integrasjon opp mot eksisterende programmer og arbeidsflyt?
- Krav til *IKT-infrastruktur og informasjonssikkerhet.*

- *Pris og prismodeller* er også viktige hensyn som må vektlegges. Foreløpig er markedet såpass umodent at prisfastsettelsen ikke nødvendigvis avstemmes ensidig av tilbuds- og etterspørsels- mekanismer; det innebærer at leverandørene vil kunne være åpne for å forhandle om pris. Forretningsmodellene som tilbys varierer noe mellom de ulike leverandørene, men ofte fastsettes prisen settes som kombinasjon av faste kostnader og pay-per-use, med utgangspunkt i et estimert volum.
- For nyutviklede løsninger har leverandørene egne *prosesser for validering og innfasing* i økosystemene. Hvor grundig disse prosessene utføres, kan potensielt variere mellom de ulike leverandørene.
- *Skalerbarhet:* Hvor enkelt er det å starte smått og deretter skalere opp bruken?
- *System for testing, utprøving, validering av algoritmer.* Er det mulig å teste og validere en KI-applikasjon på egne data før man betaler for bruken av den?
- *En felles plattform i hele Helse Nord.* Vil det være en fordel å gå for en felles løsning i hele Helse Nord (tilsvarende løsningene for PACS og journal)?
- Er det en mulighet for å *integrere egne forskningsutviklede løsninger i plattformen?*
- *Muligheter for samspill med andre modaliteter utenfor radiologi?* Er det mulig å inkludere KI-løsninger som ikke kun baserer seg på bildene radiologene beskriver i plattformen (f.eks. journal, digital patologi, osv.)?
- Er det behov å anskaffe en plattform hvis man kun ønsker å bruke noen få *enkeltalgoritmer* i en startfase?

11 Betydningen av tilrettelagt logistikk og arbeidsflyt

11.1 Overordnet om logistikk og arbeidsflyt

Dette dreier seg dels om teknisk tilretteleggelse av gode og brukervennlige grensesnitt mellom maskin, programvare og brukere, men også om organisering og tilpasning av rutiner for arbeidsflyt i det daglige arbeidet. Begge disse hensyn forutsetter godt samarbeid med og mellom ulike leverandører, med IKT-funksjonen og med helsepersonellet. Som sluttbrukere er helsepersonellets tilfredshet med innpasningen av KI-funksjonalitet i arbeidsflyten, av avgjørende betydning for å lykkes.

For å sikre at brukerne benytter seg av KI-løsninger i en travel hverdag må slike løsninger – direkte eller via valgt plattform - være integrert i dagens arbeidsflyt i RIS/PACS med godt brukergrensesnitt. PACS må tilrettelegges på en slik måte at brukerne klart kan identifisere de undersøkelser der KI benyttes. Integrasjon i RIS/PACS er nødvendig for viktige ønskelige funksjoner som for eksempel triagering/flagging av undersøkelser i arbeidslistene. Det er derfor et overordnet behov for at valgt plattform lar seg integrere i Helse Nords RIS/PACS, og for at RIS/PACS åpner for at dette kan skje, slik at KI-applikasjoner lar seg implementere på en brukervennlig og hensiktsmessig måte.

Videre må det utarbeides en arbeidsflyt som er tilpasset de ulike behovene innen de til dels svært forskjellige røntgenavdelingene ved våre fire helseforetak. Det må særlig tas hensyn til forskjeller i størrelse og grad av subspesialisering ved de ulike avdelingene. Arbeidsflyten bør være så generisk og fleksibel at den enkelte avdeling kan ta i bruk den arbeidsflyten de finner mest hensiktsmessig og eventuelt gjøre nødvendige justeringer ved behov. Spesielt i starten må det forventes et stort behov for støtte vedrørende implementering og tilpasning. Sentral utarbeiding og oversikt over mulige varianter av arbeidsflyt vil kunne bidra til ønskelig harmonisering mellom foretakene samtidig som man sikrer optimal tilpasning til lokale behov.

Videre må det etableres rutiner i Helse Nord for hvordan svar som fremkommer etter bruk av KI skal formidles til klinikere eller pasienten. I første omgang er det mest aktuelt å bruke KI som beslutningsstøtte. Rutinene må derfor nødvendigvis også omtale hvordan man kombinerer informasjon fra KI og radiolog (Jf. senere kap. 16.4)

Et annet aspekt knyttet til arbeidsflyt er behovet for å stratifisere funksjonalitet for ulike brukere. For å sikre at bruken av KI ikke virker negativt inn på kompetanseutvikling hos radiolog, bør KI-løsninger kunne styres ut fra radiologens funksjon (LIS, spesialist, subspesialist). Videre bør man kunne velge når resultatet etter at KI-algoritme er brukt skal vises for radiologen. For eksempel kan det være ønskelig at KI-resultatet ikke vises før avsluttet diktat for LIS eller at resultatet ved bruk av KI under mammografiscreening ikke vises før undersøkelsen er gransket av en eller to radiologer.

11.2 Ulike behov for ø-hjelp og elektive undersøkelser

Logistikk og arbeidsflyt vil på overordnet nivå være ulik for ø-hjelpspasienter og elektive pasienter. Dette er fordi formålet, nytten og effekten vil være forskjellig mellom disse to pasientgruppene (som angitt i kapittel 6.1.1). Et vesentlig aspekt ved organisering av arbeidsflyt, er at det må forhåndsdefineres hvilke undersøkelser det skal utføres KI på i de to ovennevnte pasientgruppene; skal KI brukes for alle undersøkelser eller bare en selektert gruppe ved behov?

For ø-hjelpspasienter er hensikten til de aller fleste KI-applikasjoner å være en støttefunksjon for bedre triagering. Dette vil hjelpe vaktpersonell mht. hvilken undersøkelse de skal prioritere først å se på. Resultatet fra triageringen må vises på hensiktsmessig måte i radiologenes arbeidsliste. Det må fremgå at undersøkelsen er blitt behandlet av en KI-applikasjon og eventuelt hvor langt analyseringen er kommet, i tilfelle den tar litt tid (f.eks. en gjennom en «bar» i prosent). Når KI-applikasjonen er kjørt, bør det komme tydelig frem om det er positivt eller negativt svar. Ved positivt funn skal undersøkelsen flyttes opp på toppen av arbeidslisten, slik at den blir håndtert raskest mulig. I tenkt scenario med flere KI applikasjoner, vil det være hensiktsmessig at det i arbeidslisten fremkommer hva det positive funnet er, f.eks. «pneumothorax, brudd, lungeemboli». Hvis en slik triageringsfunksjon skal ha sin verdi må den være tilgjengelig for alle og lett å se i arbeidslisten.

Mange KI-applikasjoner viser positive funn med et slags «heatmap» der et bilde legges i bildearkivet til den aktuelle undersøkelsen. Her må det etableres egne hengingsrutiner (rutiner for hvordan bilder fysisk skal anordnes på skjermen) i PACS der «heatmap» bilde blir presentert tydelig for radiolog som skal jobbe med undersøkelsen. Hvis en KI-applikasjon har tilstrekkelig spesifisitet, kanskje mest med tanke på å utelukke brudd, kan en tenkt arbeidsflyt være at verken positive eller negative undersøkelser i vaktsituasjon ses på av radiolog, og på den måten er avlastende på vaktarbeidet og gir mindre forstyrrelser. Både positive og negative undersøkelser svares ut med en strukturert rapport eller standardsvar opprettet av radiograf, og signeres ikke ut før neste dag av overlege. Ved positivt svar henvises pasienten av radiografer videre til akuttmottaket. KI-applikasjonens positive svar vises i dette tilfellet også for ortoped som vil tilse pasienten.

For elektive pasienter er ikke triagering i arbeidslisten et viktig element, unntaket kan være hvis en ønsker å etablere en KI-applikasjon f.eks. for tilfeldig påvist lungeemboli. Da forutsettes det at disse undersøkelser kommer godt frem i arbeidslisten på lik linje som ø-hjelpspasienter og flyttes opp i prioriteringen. Generelt, må det for elektive pasienter komme frem i arbeidslisten om undersøkelsen er gjennomgått av KI eller ikke, f.eks. med et eget elektivt KI-ikon. Positive funn her ville kunne presenteres på ulike måter, og i noen applikasjoner er det hensiktsmessig med strukturert rapport som lar seg integrere i beskrivelsen, f.eks. tabeller. I andre tilfeller vil det være hensiktsmessig at

funn presenteres som bilde(r) i PACS. Det må etableres nye hengingsrutiner der dette blir brukt for å tydeliggjøre bildefunn til KI i PACS.

12 Infrastrukturelle forutsetninger for implementering av KI-løsninger innen radiologi

For å ivareta en god brukeropplevelse av IKT-tjenester innenfor bildediagnostikk må nødvendige samarbeidspartnere som Helse Nord IKT være rustet med både ressurser og kompetanse for å kunne levere nødvendige IKT-tjenester innen KI til ønsket tid. Ved implementering, drift, validering, m.m. av kommersielt anskaffede KI-verktøy kreves det at de infrastrukturelle forutsetningene som forvaltes av Helse Nord IKT er på plass. Dette gjelder bl.a. krav til lokal infrastruktur, skyløsninger, informasjonssikkerhet, osv.

Ved implementering av KI-verktøy må det tas stilling til om bilder skal overføres via skyløsning til leverandøren for prosessering, eller om det kan gjøres en lokal installasjon (On-prem) av KI-løsningen. Her varierer mulighetene for de ulike KI-verktøyene.

Når det gjelder ytterligere resonnementer og premisser rundt de infrastrukturelle forutsetningene, vil vi i denne rapporten nøye oss med å referere til kapitlene i den overordnete KI-strategien for Helse Nord (særlig kap. 6) der dette er mer utfyllende beskrevet. I tillegg har Helse Nord IKT ledet en konsept- og planfase om behovet for å opprette et nytt tjeneste- og fagområde i Helse Nord IKT for å understøtte både forskning og implementering av KI. Dette redegjøres kort for i neste avsnitt.

12.1 Regional IKT-plattform for å understøtte kunstig intelligens i Helse Nord (RIKTIG-prosjektet)

For å anerkjenne Helse Nords satsning på kunstig intelligens og persontilpasset medisin etablerte HN IKT på oppdrag fra HN RHF i 3. tertial 2021 prosjektet *Regional IKT-plattform for å understøtte kunstig intelligens (KI) i Helse Nord* (RIKTIG). Med bakgrunn i tilbakemeldinger fra fagmiljøene og de behovene de har til KI-støtte, fremla RIKTIG-prosjektet i februar 2022 en konseptrapport, med tilhørende prosjektbegrunnelse og prosjektforslag i mars. Dokumentene presenterte fire prinsipielt ulike konsepter hvor konsept 3 ble anbefalt. Det anbefalte konseptet gir mulighet for infrastruktur- og tjenestestøtte til KI-initiativ og prosjekter i prosjektperioden. I gjennomføringsfasen skal RIKTIG gi infrastruktur- og tjenestestøtte til 3 - 4 KI-initiativ /prosjekter etter oppdragsgivers/ prosjekteier (RHF) prioritering av tilvalg.

Prosjektproduktene fra RIKTIG kan hovedsakelig deles inn i to:

- Faglig fundament
- Teknisk grunnmur

Hovedproduktet *faglig fundament* innebærer etablering av en KI-tjeneste/KI-tjenesteområde og vil kreve kompetanse og roller som kan ivareta leveranser og rådgiving. I hovedproduktet *teknisk grunnmur* skal det utvikles og tilgjengeliggjøres teknologi for å supplere tjenesteområdet, herunder også bygge kompetanse rundt teknologiske KI-løsninger. Hovedproduktet *teknisk grunnmur* kan deles inn i *egenutvikling* og *implementering av kommersielle løsninger*, hvor begge kan understøtte kunstig intelligens innen forskning og/eller klinisk utvikling.

Helse Nord RHF vil i løpet av desember 2022 behandle forslagene i RIKTIG-prosjektet; denne behandlingen vil avklare hvilke rammer som eventuelt stilles til disposisjon, og dermed hvilken framdrift som er mulig i kommende budsjettår (2023).

12.2 ROS-analyser mht. informasjonssikkerhet, personvern og pasientsikkerhet

I det følgende redegjøres kort for aspekter ved informasjonssikkerhet og sikker informasjonsflyt som er relevante for bruk av KI-løsninger for radiologi i Helse Nord. For nærmere bakgrunn vises til kap. 9.2 i *Strategi for kunstig intelligens i Helse Nord 2022-2025.*

Tilfredsstillende informasjonssikkerhet er en forutsetning for å kunne ta i bruk KI-baserte beslutningsstøtteverktøy i Helse Nord. Innbyggere, pasienter og ansatte skal ha tillit til at Helse Nords sin behandling av helse- og personopplysninger gjennomføres med et forsvarlig sikkerhetsnivå. Vår anbefaling er at informasjonssikkerhetsaspektet ved KI inkluderes i det felles rammeverket for informasjonssikkerhet som nå er under gjennomføring i Helse Nord.
Grunnleggende krav for bruk av KI innen radiologi må følge personvernforordningen artikkel 5[41] der informasjonssikkerhet skal være innebygget i løsningen, og ivaretas slik at data og sikkerhetsrutiner:

- ikke blir kjent for uvedkommende, både internt og eksternt (konfidensialitet)
- ikke blir endret utilsiktet eller av uvedkommende (Integritet)
- er tilgjengelig for autoriserte ved behov (tilgjengelighet),
- og at organisasjonen og systemet evner å gjenopprette normaltilstand (robusthet)[42]

Helse Nord IKT vil være sentral leverandør av KI for radiologi til helseforetakene. Adm. direktør for Helse Nord IKT er databehandler for flere av disse løsningene og tjenestene, og skal oppfylle krav i lov og forskrift, regionalt styringssystem for

[41] Helseforskningsloven § 32 og personvernforordningen artikkel 5.
[42]Robusthet oppnås gjennom egnede tekniske og organisatoriske tiltak som muliggjør forebygging, deteksjon, skalerbarhet, håndtering og gjenoppretting.

informasjonssikkerhet og databehandleravtaler. Ansvar og myndighet for informasjonssikkerhet i Helse Nord er beskrevet i Regionalt styringssystem for informasjonssikkerhet. Ansvar, myndighet og oppgaver skal være ytterligere presisert og beskrevet i oppdaterte tjenesteavtaler og databehandleravtaler, som til enhver tid skal være oppdaterte.

Løsningen skal oppfylle alle krav i regionalt styringssystem for informasjonssikkerhet. Norm for informasjonssikkerhet og personvern i helse og omsorgssektoren, og Nasjonal sikkerhetsmyndighet (NSM) sine grunnprinsipper for IKT sikkerhet er førende for alt arbeid i Helse Nord. Det forventes at Normens innhold på sikt oppdateres også til å omfatte retningslinjer for å ivareta personvern og informasjonssikkerhet innen bruk av kunstig intelligens. Normen er kjent og utbredt i helse- og omsorgssektoren, og vil kunne ha stor betydning også innen KI for radiologi.

Risikoer for pasientens personvern kan variere avhengig av hvordan algoritmen er implementert og hvordan data lagres. Vanligvis kan KI-algoritmer kjøres på en av to måter, enten lokalt på servere installert i HN IKT's nettverk (lokal sky / on-prem) eller på data lastet opp til en ekstern skyinfrastruktur. Skyinfrastruktur er mer vanlig ettersom det reduserer implementeringskostnadene, gir bedre skalering og medfører enklere oppdatering av programvare. Leverandørene av KI-algoritmer har mulighet for å behandle data i algoritmer som kjøres både i lokal og ekstern skytjeneste. Uavhengig av gjennomføringsmetode, bør det være tydelig retningslinjer om eierskap og rettigheter knyttet til bruk av bilder behandlet av algoritmen og om disse bildene kan brukes av selskapet for videre algoritmeutvikling.

Opplysninger og vurderinger som kan knyttes til enkeltpersoner regnes som personopplysninger. En ekstern sky-basert tjeneste forutsetter at bildedata anonymiseres, krypteres når de overføres til en privat leverandør. Det er viktig at alle vilkår for anonymisering er oppfylt. Med anonymisering menes det at det ikke lenger er mulig å finne tilbake til personen som opplysningene angår, verken via en kodeliste, eller at opplysningene i seg selv avslører hvem personen er. Anonymisering av bilder tatt «inni kroppen» vil være vanskeligere å rekonstruere, men må adresseres i en ROS-analyse av det konkrete systemet.

Etter prosessering av algoritmen, krypteres det anonymiserte svaret før sending tilbake. I gjennomføring av en ROS-analyse er det særlig kommunikasjon mellom DICOM-server og gateway (hub) samt kommunikasjon mellom gateway til leverandørens skytjeneste som er kritiske. Data vil typisk ligge ca. ett minutt på DICOM-serveren og ingen data vil lagres i leverandørens sky. Forhold som omhandler konfidensialitet ansees derfor ikke til å være en uoverkommelig barriere for implementering.

I utgangspunktet er det ingen rettslige begrensinger for bruk av private leverandører for å levere en skytjeneste, forutsatt at det ikke behandles helse- og personopplysninger i

løsningen. Slik vi i dag har forstått de tilgjengelige løsningene, lagres ingen pasientdata hos leverandøren, men dette må kvalitetssikres i en ROS-analyse. Sykehuspartener HF arbeider nå med en ROS-analysert løsningsbeskrivelse for Vestre Viken HF, som vil være et godt utgangspunkt for i tilsvarende analyse i Helse Nord.

Det er et krav om å ha kontroll på informasjonsflyten mellom ulike deler av systemer, slik at enheter kun kan kommunisere sammen etter vedtatte regler. Virksomheten skal også ha kontroll på informasjonsflyten inn og ut av løsningen. KI løsningen for radiologi må oppfylle beste praksis for segmentering i nettverket slik at uvedkommende ikke får tilgang til data, og at den øvrige IKT-infrastrukturen ikke blir underlagt unødig risiko. Dataflyten via integrasjonene må underlegges samme krav til sikkerhet som for andre integrasjoner.
Bruk av KI i radiologi vil derfor ha noen ekstra trusler som må adresseres:

- Uautorisert tilgang til algoritmer, system eller data
- Skape feilaktige ut-data fra et KI-system basert på manipulerte treningsdata eller manipulert modell
- Hindre funksjonen til et KI-system

Alle tilganger og forsøk på tilganger skal logges, og det må iverksettes tilstrekkelig tiltak for at loggene følges opp. Helse Nord må legge til rette for å overvåke systemet, og skal samle inn relevante data for å oppdage sikkerhetshendelser, og legge et grunnlag for analyse av data.

Anbefaling:
Før KI-løsninger tas i bruk, skal det gjennomføres risiko og sårbarhetsanalyse, og iverksettes sikringstiltak som for andre IKT-systemer.

13 Regulatoriske endringer - fra MDD til MDR

Til tross for de siste års betydelige utvikling av KI-teknologier og det økte antallet CE-merkede algoritmer på markedet, har det parallelt vært reist kritiske spørsmål[43] [44] [45] om kvaliteten på den evidens som ligger til grunn for sertifisering av produktene. Det har blant annet vært rettet kritikk mot mangelen på ekstern validering av løsningene, ved at samme datasett brukes både til trening og validering; dermed risikerer man å overestimere den diagnostiske nøyaktigheten. Videre konstateres at det foreligger et begrenset antall randomiserte studier[46] og multisenter-studier, at det er en stor overvekt av retrospektive studier sammenlignet med prospektive studier i kliniske settinger, og at de mange ulike kriterier som anvendes for å evaluere løsningene gjør, sammenligninger vanskelige. I en relativt fersk gjennomgang av de 100 første produktene i databasen *AI for radiology*[47] konstateres at bare ca. en tredel av disse algoritmene kan vise til noen form for vitenskapelig dokumentasjon[48]. Disse begrensningene innebærer en risiko både for løsningenes kvalitet og pasientsikkerhet, noe som har stimulert akademiske miljøer til å utarbeide standarder for både utvikling og testing av algoritmer[49]; denne problemstilling har også mobilisert interesse hos regulatoriske myndigheter. I januar 2021 ble det i regi av FDA (Federal Drug Administration) i USA tatt initiativ til å utvikle metoder for å evaluere og forbedre maskinlæringsalgoritmer[50]; FDA har i det hele tatt vært mer framoverlent mht. å adressere problemstillinger knyttet til evaluering og validering av ML-algoritmer i helsetjenesten enn tilfelle har vært for tilsvarende EU-organer. Kommisjonens vedtak av det nye regulativet (MDR) – gjort gjeldende fra 26. mai 2021 – bidrar likevel til å skjerpe kravene til utvikling og sertifisering av medisinsk utstyr, inkludert programvare [51].

---

[43] Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic reviews of design, reporting standards and claims of deep learning studies. *Br.Med J* 2020; 368: m689.
[44] Vasey B, Utsprung S, Beddou B, et al. Association of clinician diagnostic performance with machine learning-based decision support systems: a systematic review. *JAMA Netw Open* 2021; 4:e211276
[45] Wu E, Wu K, Daneshjou, R, et al. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med* 2021;27: 576-584
[46] Plana D: Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review, JAMA Netw Open 2022, pubmed.ncbi.nhi.gov. I denne artikkelen identifiseres 41 RCTer etter en gjennomgang av over 19000 publikasjoner.
[47] www.aiforradiology.com
[48] Van Leeuwen KG, Schalekamp S, Rutten M, et al. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol* 2021;31: 3797-3804. DOI:10.1007/s00330-021-07892-z
[49] Sounderja, V, Ashrafian, H, Golub, RB et al.: Developing a reporting Guideline for artificial intelligence-centered diagnostic test accuracy studies: - STARD-AI protocol, BMJ Open 2021
Liu, X, Rivera, SC: Reporting Guidelines for clinical trial reports interventions involving artificial, intelligence: the CONSORT-AI extension), Nature Medicine 26, 1364-1374 (2020)
The DECIDE-AI Steering Group: DECIDE: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence, Nature Medicine 27, 186-187 (2021)
[50] US Food and Drug Administration: Artificial Intelligence Machine Learning , Volum 11, issuw 6 (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan 2021.
[51] https://legemiddelverket.no/medisinsk-utstyr/regelverk-for-medisinsk-utstyr/nytt-regelverk-om-medisinsk-utstyr

EU-Kommisjonen regulatoriske grep skjedde ved å erstatte MDD (Medical Device Directive) med MDR (Medical Device Regulative). Ingen av de tidligere kravene i MDD er blitt fjernet, men i MDR stilles betydelig større krav til dokumentasjon og validering av løsningene. Kommisjonen sier selv:

> «Compared to the current Directives, The MDR places more emphasis on a life-cycle approach to safety, backed by clinical data»[52].

Denne forordningen gjelder produksjon og omsetning av medisinsk utstyr, inkludert programvare som KI-algoritmer. Det nye regulativet innebærer at nye løsninger som skal inn på markedet etter 26. mai 2021, må sertifiseres etter de nye og skjerpede kravene i MDR, men også at produkter som tidligere er godkjent under MDD, må re-sertifiseres på grunnlag av de nye krav i MDR, senest innen 24.mai 2024, dersom de fortsatt skal ha markedsadgang.

I likhet med MDD klassifiseres produktene også MDR i fire ulike risikoklasser, fra klasse 1 (lavest risiko), via klasse 2a og 2 b til klasse 3 (høyest risiko). Produktene klassifiseres for risiko etter intendert bruk, og ikke etter anvendt teknologi. Kravene for klassifisering er noe endret, slik at mange av algoritmene i risikoklasse 1 flyttes til en høyere risikoklasse[53]. Dette er en vesentlig endring ettersom løsninger i klasse 1 kan selvdeklareres uten å gå veien om et faglig kontrollorgan (Notified body), mens produkter i høyere risikoklasser må gjennom en slik ekstern verifisering (samsvarsvurdering) av at kravene i regulativet er innfridd.

For KI-løsninger innen radiologi forventes dette å få betydelige konsekvenser, ettersom de i hovedsak vil bli klassifisert som klasse 2a eller 2b, og dermed må gjennom en slik ekstern verifisering (notified body)[54]. Blant CE-merkede algoritmer innen radiologi som er godkjent etter MDD, er det hele 68 av i alt 157 (43 %) produkter som tilhører klasse 1, og som altså ikke har vært gjennom noen ekstern verifisering. Det antas at disse produktene har en mer krevende prosess fram til reklassifisering enn produkter som allerede har vært gjennom en ekstern samsvarsvurdering[55].

Sammenlignet med MDD, skjerper MDR kravene til klinisk evaluering og legger mye større vekt på dokumentasjon av klinisk evidens etter at produktet er ute på markedet. Den kliniske evalueringen gjennomføres i flere steg[56]:

1) forberedelse av en klinisk evalueringsplan;

---

[52]European Commission. Factsheet for Manufacturers of Medical Devices, www.dnv.com

[53] Minssen T, Gerke S, Aboy M, et al. Regulatory responses to medical machine learning J.LawBiosc 220; 7:1-18.DOI:10.1093/jlb/Isaa02

[54] Niemiec, Emilia: Will the EU Medical Device Regulation help to improve the safety and Performance of medical ASI devices? *Digital Health* 2022 Volum .8;1-DOI:101177/20552076221089079journals.sagepub.com/home/dhj

[55] Epostutveksling med Kicky van Leeuwen i AIforradiology

[56] European Commission. Guidelines on medical devices. Clinical evaluation: a guide for manufacturers and notifies bodies under directives 93/42EEC and 90/385EECMEDDEV 2.7/l revision 4 2016

2) identifikasjon av relevante data, både fra leverandør og fra forskningslitteraturen;
3) vurdering av tilgjengelige data (inkludert metodologisk kvalitet, validitet og klinisk relevans);
4) for høyrisikoutstyr, skal det utpekes ekspertpanel for konsultasjon;
5) analyse av alle relevante kliniske data for å vurdere om de samsvarer med de generelle kravene til pasientsikkerhet og ytelse (performance).

Resultatene av den kliniske evalueringen skal presenteres i en klinisk evalueringsrapport som del av den tekniske dokumentasjonen. MDR forutsetter at leverandørene etablerer et system for overvåkning etter at utstyret er bragt i omsetning, der data om kvalitet, ytelse og pasientsikkerhet systematisk innsamles og analyseres over hele produktets levetid. Resultatene av disse analysene skal brukes til å evaluere bivirkninger (side effects) og eventuelle nye (emergent) risikoer ved bruk av produktene, slik at avveiningene av risiko og gevinster (benefits) avstemmes under hele produktets levetid.

/I MDR stilles det ikke bare skjerpede karv til leverandørenes dokumentasjon, men også til de eksterne kontrollorganene (Notified bodies). Det legges stor vekt på disse organenes integritet og uavhengighet, eksemplifisert ved at de ikke kan yte konsulenttjenester eller rådgivning til leverandører som har et produkt under vurdering for godkjenning.

Det er for tidlig å gi noen endelig og kvalifisert vurdering av hvilke virkninger innføringen av MDR vil ha, men det kan allerede konstateres at de skjerpede kravene i vesentlig grad har redusert tilveksten av nye CE-merkede produkter siden 26. mai 2021; per 1. november 2022 er det bare 39 av 204 produkter innen radiologi som er godkjent etter MDR, sammenlignet med det langt høyre antall nye produkter som i 2019 og 2020 ble godkjent etter reglene i MDD. Denne utvikling antas å være uttrykk for at terskelen for å dokumentere de nye kravene er blitt høyere og mer krevende for leverandørene; antagelig innebærer disse kravene også kostnader som kan være vanskelig å bære, særlig for små leverandører som foreløpig har dårlig inntjening.

MDR er et regelverk som primært adresserer krav til produsenter/leverandører og kontrollorganer (notified bodies). Det er likevel viktig at også helsetjenesten som brukere og innkjøpere er seg bevisst den innskjerping av kravene som følger av dette direktivet. Den norske loven om medisinsk utstyr, med tilhørende forskrifter, supplerer MDR for Norges del. Forskrift om håndtering av medisinsk utstyr[57] inneholder enkelte regler om bruk og anskaffelse av medisinsk utstyr.

En offentlig utlysning for kommersielle anskaffelse i Helse Nord antas tidligst å skje første halvår i 2023, med mulig beslutning i løpet av andre halvår. Avhengig av

[57] Forskrift 29. November 2013 nr.1373 om håndtering av medisinsk utstyr

vurdering av produktene i markedet, priser og økonomiske forutsetninger i foretaksgruppen, kan utlysning og anskaffelse eventuelt komme til å skje på et senere tidspunkt, eksempelvis i løpet av 2024. Siden gyldigheten for produkter godkjent etter MDD utløper 24. mai 2024, innebærer det i praksis at helseforetakene i Helse Nord må legge godkjenning etter kravene i MDR til grunn for kravspesifikasjoner som utformes i tilknytning til eventuelle anskaffelser. I den grad utlysning skjer før ovennevnte dato, kan det følgelig være flere av nåværende produkter i markedet som tidligere er CE-merket i henhold til MDD, som ikke er resertifisert i henhold til kravene i MDR. Noen av disse produktene kan være interessante og relevante for eventuell anskaffelse, forutsatt at de er i en prosess med resertifisering som kan være fullført innen beslutning om anskaffelse. Arbeidsgruppen ser det som viktig at disse hensyn ivaretas i den relativt korte overgangsperioden fram til 24. mai 2024, slik at vi ikke risikerer å måtte avskilte innkjøpte produkter etter kort tid. Etter nevnte dato i 2024, vil produkter som mangler godkjenning etter MDR ikke ha markedsadgang.

14 Studier, erfaringer og kost-nytte-analyser

14.1 Helseøkonomiske analyser av KI-løsninger.

Den betydelige veksten i kommersielt tilgjengelige og sertifiserte KI-løsninger innrettet mot helsetjenesten, gjør det desto viktigere å avklare hvilken merverdi slike løsninger kan levere til klinisk praksis. Status i dag er at majoriteten av løsninger i dag bare rapporterer algoritmenes nøyaktighet (accuracy) og presisjon[58]. Likevel er ikke perfekt nøyaktighet eller presisjon noen garanti for at løsningen tilfredsstiller kravet til klinisk effektivitet. Mot denne bakgrunn må det samtidig konstateres at det foreligger svært få HTA-studier som sier noe om KI-løsningenes komparative effektivitet eller kost-nytte i en klinisk setting. Beslutningstakere som skal vurdere implementering av KI-løsninger, vil ha et betydelig behov for slik kunnskap basert på helhetlige HTA-analyser der også kost-nytte perspektivet er ivaretatt.

Det er i inneværende år (2022) publisert en kunnskapsoppsummering om studier som tar for seg økonomisk evaluering av KI-løsninger, basert på studier utført mellom 1. januar 2016 og 1. april 2021.[59] Denne gjennomgangen henter data fra to databaser (PubMed og Scopus) og identifiserer hele 884 unike studier, men hvorav bare 20 ble inkludert i den endelige analysen. Nær halvparten av disse studiene – 9 i alt – gjaldt KI-løsninger innen radiologi. De inkluderte studienes metodologiske kvalitet, ble evaluert på grunnlag av det anerkjente rammeverket CHEERS (Consolidated Health Economic Evaluation Reporting Standards), utviklet i 2013[60] og oppdatert i 2022[61]. Dette rammeverket er utviklet og forvaltet gjennom ISPOR – The Professional Society for Health Economics and Outcome Research. CHEERS inneholder en sjekkliste bestående av 28 punkter som skal bidra til å standardisere og kvalitetssikre helseøkonomiske studier, noe som også vil gjøre studiene nyttige for beslutningstakere på ulike nivåer. Primært utfallsmål i studiene var kostnader per case, der de fleste studiene analyserte kostnader på sykehusnivå, mens noen studier også fanget inn samfunnsnivået. Grunnlaget for sammenligning av kostnader var KI-løsninger versus bruk av tradisjonelle metoder (radiologers vurderinger), og i de flere tilfeller ble KI-løsningene vurdert som kostnadsbesparende. Det dreier seg imidlertid om svært få studier, og bare fem av studiene skårer på mer enn 80 prosent av CHEERS-kriteriene.

En fersk kostnads-effektivitetsanalyse, som ikke var inkludert i forannevnte kunnskapsoppsummering, tar for seg bruk av kunstig intelligens ved deteksjon av

[58] Gore, JC. Artificial intelligence in medical imaging. Magn Reson Imaging 2020:68 A!-A4
[59] Voets, Adelon M. et al: Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare. The Tortoise and the Cheetah. Value Health 2022:25 (3):340-349
[60] Husreau, Don, Drummond, Michael, Pwetrou, Stavous et al: Consolidated Health Economic Evaluation Standards (CHEERS) – Explanation and Elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force, in Value Health. Mr-Apr 2013;16(2):231-50
[61] Huserau, Don, Drummond, Michael, Augustovsk, Fredricoi et al: Consolidated Health Economic Evaluation Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations, in BMJ, 2021:375: e0.67975

karokklusjon ved akutte hjerneslag[62]. Denne studien anvender tidlig HTA-metoden for å estimere potensielle gevinster ved å ta i bruk KI sammenlignet med tradisjonelle arbeidsmåter. Empirien er hentet fra det britiske slagregisteret (n=71840). Gevinstene måles ved QALY (kvalitetsjusterte leveår), og er vurdert både på kort og lang sikt og inkluderer følgelig samfunnsperspektivet – og ikke bare sykehusperspektivet – på gevinstrealisering. Ut fra de forutsetninger som er lagt til grunn i denne studien, er den årlige potensielle gevinsten ved å ta i bruk KI-løsninger for denne pasientgruppen i hele Storbritannia til ca.11 millioner pund. Det understrekes at dette er en analyse basert på retrospektive data, noe som tilsier behov for supplerende kontrollerte prospektive studier i en klinisk setting.

Vi må fortsette å følge med i og lære av internasjonale studier av kostnader og nytte ved å ta i bruk KI-løsninger. Arbeidsgruppen ser det likevel som minst like viktig at utprøving og implementering av KI-løsninger i vårt land ledsages av egeninitiert systematisk helseøkonomisk følgeforskning. Det vurderes som viktig at denne evalueringsforskningen gjennomføres etter anerkjente faglige metoder og på en standardisert måte, eksemplifisert med CHEERS. Dette tilsier at det utvikles en nasjonal tilnærming for slik forskning, etableres både nasjonal og regional (metode-) kompetanse på dette feltet, og at det organiseres et nasjonalt samarbeid om gjennomføring av følgeforskning ved innføring av KI-løsninger i helsetjenesten.

Gjennom slik følgeforskning basert på norske erfaringer med innføring av KI-løsninger, vil vi også få relevant empirisk kunnskap om pukkelkostnader i startfasen, og om skalaforutsetninger (volum mv.) for å kunne ta ut gevinster.
Det vil videre være en potensiell gevinst av slik følgeforskning, at den kan bidra til å utvikle forenklede analyseverktøy som kan tas i bruk av helseforetak og andre beslutningstakere ved vurderinger av kost-nytte, særlig ved storskala anskaffelser av KI-løsninger. Det er slik sett ikke bare forskerne som har behov for en felles metodikk, men også beslutningstakerne.

I tilknytning til anskaffelsen i Vestre Viken, ble det i vurderingsprosessen utviklet et regneark for å bergene pris per undersøkelse og totalkostnader (basert på forventet volum). Når dette helseforetaket i løpet av høsten gjennomfører en mini-HTA i tilknytning til implementeringen, planlegges forenklede kost-nytteanalyser som vi også kan lære av.

[62] Kicky G. van Leeuwen et al.: Cost-effectiveness of artificial intelligence aides vessel occlusion detection in acute stoke: an early health technology assessment, in Insight Imaging (2021) 12:133

14.2 Implementering – studier og erfaringer

Ettersom norsk helsetjeneste i løpet av de kommende år forventes å gjennomføre stadig flere anskaffelser med påfølgende implementeringer av løsninger for kunstig intelligens, er det viktig at dette forberedes på en kunnskapsbasert måte.

Nasjonalt senter for ehelseforskning (NSE) har nylig gjennomført en empirisk studie[63] av problemstillinger knyttet til implementering av kunstig intelligens i helsetjenesten. Materialet er hentet fra intervjuer med en rekke aktører[64] både i Norge og utlandet (9 nasjoner), og ved gjennomgang av aktuelle og relevante dokumenter. Gjennom kontakten med disse informantene, kom det fram at få virksomheter er kommet langt med implementering og at en vesentlig del av de formidlede erfaringer er basert på utprøving og pilotering. NSEs rapport er omfattende og favner bredt. I dette kapitlet begrenser vi oss derfor til å oppsummere et utvalg av de funn og konklusjoner som vurderes å ha særlig relevans for vår rapport.

Innledningsvis understrekes betydningen av å øke kunnskapen om KI gjennom ulike tiltak:

- Innarbeide kunstig intelligens i pensum og etablere dedikerte kurs for helsepersonell, etablere en nasjonal kompetansetjeneste som er tilgjengelig for ansatte på alle nivåer i helsetjenesten, og å fremme ulike tiltak for å øke kunnskapen om KI i befolkningen

Det understrekes videre at helsepersonell må konsulteres og involveres ved beslutninger om å ta i bruk KI, bl.a. ved å:

- Ta utgangspunkt i kliniske behov og å legge stor vekt på å tilpasse løsningene til den kliniske arbeidsflyten
- Involvere helsepersonell fra starten av implementeringsprosessen
- Inkludere pasientene i disse prosessen

For å få til vellykkede implementeringer, må det etableres gode samarbeidsrelasjoner til aktører både innenfor og utenfor virksomhetene; på *lokalt nivå* innebærer det bl.a. å

- Å etablere tverrfaglige team bestående av både teknologer og kliniske domeneeksperter, i samarbeid med administrative ledere, og med juridisk støtte
- Organisere implementeringen som en iterativ[65] prosess
- La kliniske nøkkelpersoner (clinical champions) drive KI-initiativene
- Oppnå felles forståelse med - og strategisk støtte fra – virksomhetens ledelse

På det *nasjonale plan* bør:

- Det etableres retningslinjer for implementering

---

[63] Makhlysheva A, Ruiz LM, Olsen, TS et al: Implementation of Artificial intelligence in Norwegian healthcare: The road to broad adoption: https://ehealthresearch.no/rapporter implementation-of-artificial-intelligence-in-norwegian-healthcare-the-road-to-broad-adoption

[64] 46 personer hvorav ca. halvparten fra utlandet, er intervjuet.

[65] Iterativ betyr repeterende prosesser i flertrinns justerende feedback-loops

- Samarbeidet mellom helseforetakene og mellom kommunene styrkes
- Det legges opp til å styrke den tverrfaglige kompetansen gjennom kompetansenettverk og digitale innovasjonshuber.

På det regulatoriske felt understrekes behovet for harmonisering av fortolkninger mellom ulike myndighetsorganer og helseorganisasjoner, og at det etableres samarbeidsgrupper på tvers av akademia, offentlige og private aktører for å understøtte behovet for regulatorisk kompetanse innen KI-feltet.

Når det gjelder regulering av anskaffelser, bør det legges til rette for bedre samarbeid mellom helseorganisasjoner som kjøpere og leverandørene. *Lokale anskaffelsesprosesser* bør understøttes av tverrfaglige team for å ivareta bl.a. følgende oppgaver og hensyn:

- Kartlegge klinisk arbeidsflyt for å avklare hvordan KI skal integreres og hvordan KI vil påvirke arbeidsflyten
- Integrere KI-systemene direkte i arbeidsflyten hvor det er mulig
- Etablere dokumentasjons- og kvalitetssikringssystemer som følger hele implementeringsprosessen fra start til avslutning
- Sørge for løpende tilbakemeldinger fra helsepersonell for å forbedre nødvendige tilpasninger i implementeringsprosessen
- Inkorporere medisinske retningslinjer (guidelines) og arbeidsrutiner i utvikling og bruk av KI, og sørge for en høy grad av datavisualisering gjennom intuitive brukergrensesnitt

Finansielle ressurser er viktige forutsetninger for implementering av AI, både på nasjonalt og lokalt plan. På *nasjonalt nivå* anbefales at det:

- Etableres finansielle incentiver for å få helseorganisasjoner til å utvikle og ta i bruk KI, og for å oppmuntre til samarbeid mellom akademia, industrien og helsetjenesten
- Etableres forsknings- og utviklingsfond for utprøvende KI-forsøk (trials)

På *lokalt nivå* forutsettes investeringer på andre felt for å ta i bruk KI:

- Plan for fornyelse av medisinsk utstyr
- Investeringer i IKT-infrastruktur, gjennom oppgraderinger og samarbeid med universitetene
- Etablere regionale datavarehus som deles med universitetene
- Finansiere regionale sentre for KI-kompetanse innen helsetjenesten

Testing og validering er viktige ledd ved implementering av KI-løsninger. Det anbefales derfor på *lokalt nivå* å:

- Utvikle kapasitet til å gjennomføre valideringsstudier
- Utføre kliniske studier for å evaluere effektene (nytte mv.) av AI-systemer for pasientene
- Evaluere de forandringer AI har medført for den kliniske arbeidsflyten

På *nasjonalt plan* anbefales:

- At det utformes nasjonale retningslinjer for validering av KI-systemer
- At det etableres en nasjonal valideringsplattform med et nasjonalt datasett for validering som kan muliggjøre kryssvalidering av algoritmene gjennom multisenter-studier
- Fremskaffe kompetanse og retningslinjer for avklaring av hvorvidt KI-systemer er trygge nok til å bli tatt i bruk

I rapporten fra NSE understrekes avslutningsvis at KI-systemer må understøttes av vedlikeholdssystemer som kontinuerlig monitorerer systemets ytelse i klinisk bruk. Både helseorganisasjoner og leverandører må være seg bevisst det ansvar hver av partene har for vedlikehold av KI-systemene, og at de respektive ansvarsroller kontraktfestes.

14.2.1 NSEs kunnskapsbaserte anbefalinger i lys av andre kilder og vurderinger som ligger til grunn for arbeidsgruppens rapport.

En rekke av de problemstillingene som angår implementeringsprosessen er allerede – direkte eller indirekte - berørt, både gjennom flere av denne rapportens kapitler, og i fjorårets rapport *Strategi for kunstig intelligens i Helse Nord. 2022-2025* (juni 2021). Blant kildene til disse rapportene inngår en ikke ubetydelig gjennomgang av internasjonalt publiserte KI-studier. Det er på dette grunnlag arbeidsgruppens vurdering at NSEs oppsummerende anbefalinger er gjenkjennbare fra den generelle litteraturen om utprøving, implementering og bruk av KI-algoritmer i helsetjenesten. NSEs rapport er utvilsomt mer systematisk og omfattende i drøftelsen av implementering, men det konstateres at flere av NSEs anbefalinger er sammenfallende med vurderinger som gjøres underveis i vår rapport.

Med støtte i bl.a. NSEs studie, vil arbeidsgruppens oppsummere følgende prioriterte anbefalinger for implementering av KI-løsninger:

- Allerede fra starten av, må anskaffelse og implementering av KI-løsninger forankres med støtte og tillit både i virksomhetsledelsen og i fagmiljøene
- Pasientene og pasientorganisasjonene må involveres i arbeidet med å ta i bruk KI-løsninger
- Ved innføring av KI-løsninger må det legges til rette for adekvat opplæring av det personell som skal delta i dette arbeidet
- Implementering av KI-løsninger bør organiseres som tverrfaglige prosjekter der teknologer, klinikere og andre yrkesgrupper samarbeider gjennom hele prosessen
- Ved innføring av KI-løsninger, bør det satses på algoritmer som er godt testet og validert på relevante pasientpopulasjoner, noe som øker mulighet for vellykket implementering

- Det bør satses på KI-løsninger med enkle og intuitive brukergrensesnitt som legger til rette for god arbeidsflyt, noe som også fasiliteter implementeringen
- Det må legges til rette for understøttende infrastruktur og god tilgang til aktuelle og relevante data

NSEs studie er gjort på grunnlag av informasjon et tidlig stadium i implementeringen av KI-algoritmer. Likevel har studien belyst viktige hensyn og forutsetninger som må ivaretas for å kunne lykkes i dette arbeidet. Ettersom det gjøres flere anskaffelser og utrullinger av KI-løsninger, vil det empiriske grunnlaget for evalueringer og kunnskapsoppsummeringer bli vesentlig bedre. Det er derfor grunn til å forvente at det i årene framover vil komme nye studier av implementering som vil gi oss nyttig og relevant kunnskap om kritiske suksessfaktorer for implementering.

14.3 Nasjonale KI-Rapporter i regi av Helsedirektoratet og Direktoratet for e-helse

Det nasjonale koordineringsprosjektet *Bedre bruk av kunstig intelligens,* er ett av over 20 prosjekter som ble opprettet for å følge med på implementeringen av *Nasjonal helse- og sykehusplan 2020-2023*. Helse og- omsorgsdepartementet ga Helsedirektoratet i oppdrag lede prosjektet, i samarbeid med andre direktorater (Direktoratet for e-helse, Legemiddelverket, Helsetilsynet og Folkehelseinstituttet), de regionale helseforetakene og KS. Samarbeidet mellom partene har primært vært organisert gjennom månedlige møter i en styringsgruppe, men også gjennom seminarer, arbeidsmøter og andre kontaktformer med relevante aktører i helsetjenesten. Helse Nord har, både gjennom formell representasjon i styringsgruppen og mer uformelle kontakter (SPKI m.fl.) gitt innspill til arbeidet i koordineringsprosjektet.

I regi av koordineringsprosjektet er det i løpet av de siste to-tre årene utformet flere rapporter. Her gis en kort omtale av tre rapporter som har relevans for de problemstillinger som adresseres i Helse Nords radiologirapport. Det gjelder

1) *Tilrettelegging for bruk av kunstig intelligens i helsetjenesten - med utgangspunkt i det radiologiske fagområdet* (oktober 2021)
2) *Kunstig intelligens i helsetjenesten – status og veien videre for det nasjonale koordineringsprosjektet* (oktober 2022)
3) *Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten (oktober 2022).*

Den første av ovennevnte rapporter, *Tilrettelegging for bruk av kunstig intelligens i helsetjenesten, har særlig fokus på det radiologiske fagområdet*, er å forstå som en underveisrapport fra arbeidet i koordineringsprosjektet. Denne rapporten er utformet som en slags katalog over aktuelle problemstillinger og spørsmål knyttet til bruk av kunstig intelligens i helsetjenesten, utformet på grunnlag av dialog med ledere, klinisk personell, anskaffere, brukerrepresentanter, pasientforeninger, næringsliv og forskere.

Her redegjøres for en rekke framtidige bruksområder for kunstig intelligens i helsetjenesten, eksemplifisert med problemstillinger knyttet til henvisning, beslutnings-støtte og hjelp til prioritering, kvalitetssikring, tidligere diagnostikk av alvorlig sykdom, og hjelp til å utføre tidkrevende og repeterende oppgaver. Mer konkret og spesifikt redegjøres for KI-applikasjoner innen radiologi som allerede er i bruk ved norske sykehus og for pågående forskningsprosjekter på dette feltet. I rapporten understrekes at det både ved kjøp av kommersielle CE-merkede produkter og ved egenutvikling av KI-løsninger, må sikres av løsningene understøtter reelle kliniske behov og at regelverk for utvikling, anskaffelse og bruk følges.

Kunstig intelligens i helsetjenesten – status og veien videre for det nasjonale koordineringsprosjektet, er en mer omfattende rapport som oppsummerer status for arbeidet med KI, både nasjonalt, i de fire regionale helseforetakene og i kommunene; i tillegg redegjøres det for de anbefalinger om oppfølgende tiltak som har fått tilslutning i styringsgruppen i koordineringsprosjektet. Denne rapporten er mer systematisk og bearbeidet enn ovennevnte, er grundig drøftet i styringsgruppen og dermed godt forankret blant partene som inngår i koordineringsprosjektet. I så måte er dette å betrakte som et referansedokument for arbeidet med KI i helsetjenesten i Norge.

Rapporten oppsummerer koordineringsprosjektets anbefalinger om videreføring av tiltak på åtte områder:

1. Nasjonal samhandling og koordinering
2. Veiledning og kompetanse på regelverket
3. Helsetjenestens validering av produkter før bruk
4. Metodevurdering
5. Finansiering og gevinster av KI-prosjekter
6. Kompetanse på kunstig intelligens
7. Tilgang til data
8. Lære av og samarbeide med andre land

Denne statusrapporten med koordineringsprosjektets anbefalinger er nå oversendt Helse- og omsorgsdepartementet for vurdering av de forslag som er fremmet.

Den tredje rapporten fra koordineringsprosjektet som omtales her, *Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten,* er utarbeidet av Direktoratet for e-helse. Hovedbudskapet i denne rapporten er at gode og godt tilrettelagte data, er avgjørende for at helse- og omsorgs-sektoren skal lykkes med å ta i bruk kunstig intelligens. Kvaliteten på anbefalingene fra et KI-system er nemlig avhengig av kvaliteten på de data som ligger til grunn for bruken. Tilgang på gode data skal samtidig ivareta de viktige hensynene til personvern og informasjonssikkerhet.

Det redegjøres for flere ulike datakilder: registerdata, medisinske bilder, data fra medisinsk utstyr, egenrapporterte data (PROM og PREM), gendata/omikkdata, lab-data,

legemidler, kurve (innlagte pasienter), pasientadministrative systemer og journalnotater (ustrukturerte). Dernest koples disse datakildene til ulike bruksområder mht. å ta i bruk KI-algoritmer; beslutningsstøtte, effektivisering, overvåking, prediksjon av risiko, bemanningsplanlegging mv.

Rapporten er inndelt i følgende temaområder:

1. Nasjonal koordinering og samarbeid
2. Kompetanseheving, veiledning og regelverk
3. Datahåndtering
4. Utvikling og forvaltning av standarder
5. Infrastruktur og analysekapasitet
6. Offentlig og privat samarbeid
7. Europeisk samarbeid

Denne rapporten er også oversendt Helse- og omsorgsdepartementet (20. oktober 2022).

Arbeidsgruppen for KI innen radiologi i Helse Nord vil anbefale at miljøer som skal planlegge, forberede eller ta i bruk KI i våre helseforetak, setter seg inn i ovennevnte nasjonale rapporter.

15 Kompetanseoppbygging og utdanning av radiografer og radiologer

Som kommende brukere av KI-verktøy trenger radiografer og radiologer opplæring om hvordan slike verktøy kan brukes på best mulig måte. Det må derfor sørges for at det innføres gode rutiner for kompetanseoverføring ved innføring av nye verktøy; det gjelder både kunnskap om hva KI er innen radiologien, og mer praktiske og konkrete ferdigheter som kan gjøre brukerne i stand til å ta i bruk slike løsninger.

Utover dette har radiologer og radiografer også behov for å at det gjennom grunnutdanningen formidles kunnskap som kan gi en kritisk forståelse av den nye teknologiens muligheter og begrensninger. Det anbefales derfor at utdanningen styrkes på dette området.

I dette kapitlet følger arbeidsgruppens vurderinger og anbefalinger når det gjelder utdanning og oppbygging av kompetanse for radiologer og radiografer i Helse Nord.

15.1 Kompetanseoppbygging

Teknologisk utvikling muliggjør mer komplekse bildeopptak som igjen stiller store krav til dagens radiologer og radiografer. Bruk av KI-løsninger innenfor radiologi vil derfor bli et viktig arbeidsverktøy for både radiografer og radiologer i tiden fremover. Samtidig foreligger det en risiko for at slike verktøy kan ha utilsiktede negative effekter, både for spesialistutdanningen (LiS) radiologer og for senere vedlikehold av denne kompetansen.

Ved innføring av KI-løsninger, er det derfor behov for en tydelig regional plan for implementering, der risiko for mulige negative bivirkninger ved bruk av KI vurderes og følges opp med tilpassede risikoreduserende tiltak.

Det er nødvendig, men ikke tilstrekkelig å bare styrke grunnutdanningene. For å lykkes med implementering av KI-løsninger innenfor radiologi, trengs dedikerte ressurser for å ivareta behovet for opplæring og kompetanseutvikling.. Slike tiltak bør søkes organisert på regionalt nivå, med støtte fra Senter for pasientnær kunstig intelligens (SPKI) og e-læringskurs. Allerede nå er det i avdelingene behov for hybridkompetanse. vil det være nødvendig at det settes av dedikerte ressurser for å styrke KI-kompetansen blant det personell som skal ta i bruk slike løsninger. Allerede nå er det i avdelingene behov for personell med hybridkompetanse. For å komme raskt i gang med oppbygging av kompetanse, bør det vurderes å etablere små, modulbaserte kurs innrettet mot helsepersonell ved røntgenavdelingene.

For å sikre ivaretakelse av medisinskfaglige aspekter samt ivareta læring på tvers av helseforetakene, bør det vurderes å opprette et fagnettverk for bruk av KI, bestående av radiologer og radiografer. Fagnettverket bør utgå fra det regionale forvaltningsforumet, slik at god samhandling mellom disse instansene sikres.

Lederne på alle nivåer i røntgenavdelingene må ta et ansvar for at verktøy og kunnskap om kunstig intelligens både anskaffes, implementeres og settes i system. Det bør derfor bygges kompetanse hos ledere, både om verktøyet og om endringsledelse, slik at arbeidsformene tilpasses nye verktøy.

15.2 Utdanning

I møte av 04.10.2021 mellom UiT og Helse Nord RHF/ UNN ble det besluttet at UiT tar ansvar for å nedsette en arbeidsgruppe for å «utrede mer detaljert hvilke konkrete tiltak som bør iverksettes for å styrke utdannings- og opplæringstilbud innrettet mot kunstig intelligens (KI) i helsetjenesten» i tråd med Helse Nord sin strategi kap. 1.4.3. Arbeidsgruppen som ble nedsatt leverte en rapport med navnet *Styrking av utdannings- og opplæringstilbud innrettet mot Kunstig intelligens (KI) i helsetjenesten* den 01.09.2022. Rapporten skal behandles i styret i Helse Nord RHF i løpet av høsten 2022.

I rapporten forslås følgende tiltak for å styrke grunnutdanningen av helsepersonell og LiS-utdanningen:

- *Arbeidsgruppen anbefaler å avvente oppstart av nye studietilbud/ revisjon av eksisterende studieplaner til de nye endringene er implementert og man har fått erfaringer med resultatene av disse.*
- *Arbeidsgruppen anbefaler at LiS utdanningen vurderes med tanke på om KI bør få en større plass her. De forskjellige LiS utdanningene vil trolig være den viktigste arenaen for å nå frem til riktig målgruppe innenfor de ulike anvendelsene av KI i medisin. Eksempelvis vil LiS utdanningen i radiologi og patologi kreve mer kompetanse på bildeanalyse (diagnostikk), mens klinikere som bruker KI i beslutningsanalyse og prediksjon vil trenge kompetanse for hvilke vurderinger som må gjøres, før KI kan bli et nyttig verktøy i klinikken.*
- *Arbeidsgruppen anbefaler at ELMED219 i større grad gjøres kjent for relevante målgrupper, og at UiT sine studenter gis mulighet til å delta på kurset i en overgangsperiode til et tilsvarende tilbud er blitt etablert ved UiT.*
- *På sikt er det nødvendig å etablere et eget elektivt kurs på masternivå ved UiT etter inspirasjon fra ELMED219. Et slikt kurs bør være åpent for en tverrfaglig studentmasse. Fagmiljøet som skal utvikle og gjennomføre kurset bør være tverrfaglig sammensatt for å sikre at både helsefag og teknologi blir godt representert. Arbeidsgruppen anslår at det vil være behov for to 50% stillinger i 6 måneder for å utvikle kurset. I tillegg må det påregnes ressurser til å drifte kurset. Da KI er et felt i rask endring vil det være behov for å trekke inn ekstern spisskompetanse i undervisningen. Det er viktig at UiT selv etablerer et permanent*

emne innen KI for å imøtekomme behovet for kompetanse lokalt innen dette fagfeltet. Å basere på andre universiteters kurstilbud innen KI vil bare kunne være en midlertidig løsning for å avhjelpe det behovet som allerede eksisterer ved UiT. Et nytt kurstilbud innen KI vil kreve både personellmessige og økonomiske ressurser, men etter arbeidsgruppens syn bør dette prioriteres.

Kommentar

Status er altså at det i løpet av det siste året har blitt innført noen endringer i grunnutdanningen for medisinere (ref. pkt. 1). Det er foreløpig uvisst i hvor stor grad disse endringene berører radiografutdanningen og om dette vil føre til at radiografene får tilstrekkelig med kunnskap om KI.

Radiologer vil gjennom de nye endringene bli introdusert til tematikken gjennom grunnutdanningen i medisin. Videre foreslås det i pkt. 2 at LiS-utdanningen vurderes revidert. Vår arbeidsgruppe støtter opp om at det gjøres en gjennomgang av LiS-utdanningen mht. å vurdere om KI bør få en større plass der.

I lys av det internasjonalt anerkjente fagmiljøet innen maskinlæring og den betydelige satsingen på forskning om kunstig intelligens ved UiT Norges arktiske universitet, finner arbeidsgruppen det noe overraskende at ovennevnte utredning ikke har vært mer offensiv i sine forslag til utvikling av nye utdanningstilbud innen dette feltet.

Arbeidsgruppen vil for øvrig anbefale at helsepersonell (radiologer, radiografer mv.) som berøres av beslutninger om å innføre KI-løsninger ved helseforetak i Helse Nord, bør tilbys opplæring der det gis relevant praktisk kunnskap om KI.

16 Validering og kvalitetssikring av KI-verktøy

16.1 Vurdering av behovet for å validere kommersielle løsninger mot egne pasientdata

Med unntak for den utprøving av kommersielle algoritmer innen mammografi som pågår i regi av Kreftregisteret og et fåtall enkeltalgoritmer som delvis er tatt i bruk ved norske sykehus, er så vidt vites - ingen av de øvrige nær 200 kommersielle CE-merkede algoritmene innen radiologi validert mot norske pasientdata. Disse algoritmene er heller ikke utviklet på grunnlag av norske data. Det er dessuten i kapittel 14 dokumentert at en vesentlig andel av disse kommersielt tilgjengelige algoritmene mangler vitenskapelig evidens i form av uavhengige studier. Mot denne bakgrunn, er det ønskelig at kommersielle algoritmer blir validert mot norske pasientdata før de tas i rutinemessig bruk. En slik validering vurderes å være særlig viktig i denne første fasen der man begynner å ta i bruk slike algoritmer. I denne fasen er det av avgjørende betydning for fagmiljøenes tillit til løsningene at rutinene for kvalitetssikring i forkant, er gode.

Ved anskaffelser vil arbeidsgruppen anbefale at samtlige algoritmer valideres mot egne retrospektive pasientdata gjennom en prekvalifiseringsfase, slik at bare leverandører som har latt sine algoritmer valideres mot lokale data, inviteres til å gi endelig tilbud. Det er derfor viktig å tilrettelegge egne pasientdata på de aktuelle fagområdene for slik validering i forkant av anskaffelsen. Ideelt sett, burde valideringen også gjøres på prospektive data, men arbeidsgruppens ser klare praktiske og tidsmessige utfordringer ved å gjøre dette i en anskaffelsesfase. Slik prospektiv validering anbefales likevel utført når løsningen er anskaffet og tas i rutinemessig bruk.

Validering av algoritmer vurderes å både være både kompetanse- og tidkrevende. I den grad det lar seg gjøre gjennom anskaffelsesprosesser som er rimelig sammenfallende i tid, vil det derfor være ønskelig å kunne gjennomføre slike prosesser i samarbeid med andre helseforetak. Slikt samarbeid kan tenkes organisert ved at samme algoritmer utprøves og valideres desentralt på ulike lokala datasett. På denne måten kan det gis et bredere empirisk (datamessig) grunnlag for valideringen enn ved å utelukkende basere seg på egne lokala data. Det vil også styrke valideringsprosessen om de metoder som benyttes er mest mulig harmoniserte; at valideringen skjer på samme måte, på grunnlag av felles protokoller. Det vises i denne sammenheng til brev[66] fra det nasjonale kompetansenettverket for kunstig intelligens (KIN) der det foreslås å utvikle felles nasjonale retningslinjer for validering av KI-algoritmer i heletjenesten.

[66] Brev av 30.september 2022 fra KIN, adressert til det nasjonale koordineringsprosjektet for kunstig intelligens i helsetjenesten v/Helsedirektoratet.

Anbefalingen om å validere alle algoritmer før de tas i bruk, gjelder særlig i startfasen, hvor mye kan stå på spill om man ikke lykkes. Ettersom antallet algoritmer som anskaffes og tas i bruk øker i antall, vil det kunne bli svært ressurskrevende å gjennomføre slike valideringer for alle nye algoritmer. Erfaringene med de valideringsrutiner som etableres må derfor evalueres, og kanskje også sammenlignes med prosesser der anskaffelse og implementering skjer uten slik forutgående validering, I den grad aktuelle algoritmer er utviklet og/eller validert på andre populasjoner som i liten grad avviker fra norske pasienter (inkludert pasienter i Helse Nord), vil behovet for ny validering tenkes å bli redusert. Ettersom flere kommersielle algoritmer tas i bruk i andre europeiske land, vil antallet publikasjoner om oppfølgende studier av algoritmenes kvalitet (med vekt på både sensitivitet og spesifisitet) forventes å øke; disse antas også å gi relevant informasjon om validering. Behovet for validering må i framtiden også vurderes med referanse til de skjerpede kravene om kvalitet som er nedfelt i MDR. I mai 2024 kreves det at alle CE-merkede produkter er godkjent etter kravene i MDR. Det er grunn til å anta at algoritmer godkjent etter disse nye kravene vil ha en høyere gjennomsnittskvalitet enn tilfellet var for algoritmer godkjent etter MDD.

16.2 Kvantitative metoder for å evaluere KI-algoritmer

Når en algoritme skal valideres som ledd i en kvalitetssikringsprosess, er det viktig at det gjøres med grunnlag i anerkjente faglige standarder og målemetoder. Det gjelder så vel ved validering av innkjøpte kommersielle algoritmer som egenutviklede. I utgangspunktet må KI-algoritmer valideres på samme måte som mer konvensjonelle medisinske metoder, slik at resultatene kan sammenlignes med disse, med referanse til «ground truth»[67] (empirisk evidens ved observasjon og måling). Slike målemetoder anvendes ikke bare for å sammenligne en KI-algoritme mot konvensjonelle medisinske metoder (f.eks. vurdering av radiolog), men også for å sammenligne kvaliteten på ulike produkter (KI-algoritmer) innrettet mot samme medisinske problemstilling. Dette siste hensynet blir stadig viktigere ettersom markedet forsyner oss med et økende antall konkurrerende algoritmer med samme formål.

Før man starter en slik validering, er det nødvendig å klargjøre et datasett som samsvarer med den pasientpopulasjonen algoritmen er utviklet for og som inneholder de relevante data og resultatmål.

I vurderingen av kvaliteten på en algoritme, er det blant brukere vanlig å fokusere begrepet nøyaktighet (accuracy) hvis algoritmen brukes for klassifisering (f.eks. til å diagnostisere/detektere kreft vs. ikke-kreft). Dette begrepet tolkes ofte som andelen av tilfeller der algoritmen gir korrekt svar. Begrepet er imidlertid for generelt - og dermed for upresist – til å fange inn de relevante dimensjoner ved en algoritmes kvalitet eller

[67] Wikipedia: «Ground truth is a term in statistics and machine learning for accuracy against the real world. The term is borrowed from meteorology where ground truth refers to information on site.

ytelse (performance). For å fange inn de mer komplekse egenskapene ved en algoritme, er det derfor hensiktsmessig å anvende de klassiske statistiske og epidemiologiske begrepene sensitivitet og spesifisitet.

Sensitivitet måler andelen av pasienter som er korrekt diagnostisert (teller) sammenlignet med totalt antall pasienter som de facto har diagnosen (ground truth). Matematisk kan det settes opp slik;

SP/SP+FN, der SP=sanne positive og FN=falske negative.

Ved KI-analyse av radiologiske bilder, er sensitivitet uttrykk for i hvilken grad algoritmen korrekt kan identifisere hvilke pasienter som har en gitt sykdom. Dette er utvilsomt en god egenskap og dermed et viktig suksesskriterium for å evaluere en algoritme, men denne egenskapen må avveies mot andre egenskaper.

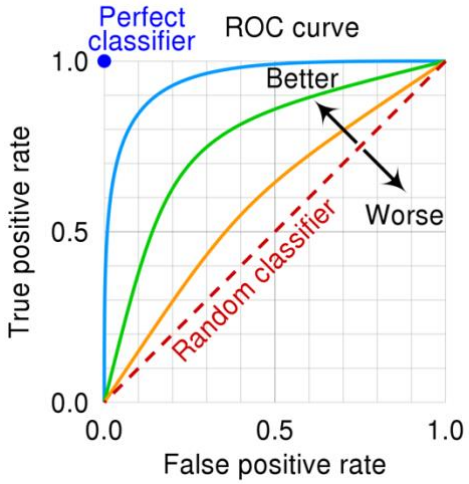
Derfor er det også nødvendig å teste algoritmens *spesifisitet,* operasjonalisert som andelen sanne negative tilfeller diagnostisert (ikke syke) sammenlignet med totalt antall pasienter som de facto er negative (ikke syke). Matematisk kan dette settes opp slik:

SN/SN+FP, der SN=sanne negative og FP=falske positive.

Hvis det er en stor andel falske positive, vil følgelig spesifisiteten reduseres vesentlig, noe som kan ha flere negative konsekvenser. Det er et krav til en god algoritme, at den skårer høyt både på sensitivitet og spesifisitet. Det fins likevel ingen eksempler på at algoritmer skårer fullt ut (100%) på begge disse parameterne, og ofte vil svært høy sensitivitet medføre lavere spesifisitet. En algoritme som identifiserer alle faktisk syke (Sensitivitet =100%), vil med stor sannsynlighet også fange inn mange falske positive; det følger av at terskelen for å identifisere sykdom er for vid. I den virkelige verden vil det følgelig være nødvendig å avveie hensynet til sensitivitet mot hensynet til spesifisitet.

Formålet med en algoritme vil i noen grad være premissgivende for den relative vektlegging av disse to dimensjonene: hvor formålet er å unngå overbehandling, vil man vektlegge spesifisitet, mens hensynet til høy sensitivitet vil bli prioritert dersom man vil redusere risikoen for å overse pasienter som de facto er syke. Dette er krevende avveininger, for eksempel ved diagnostikk av kreftsykdom, der man helst vil identifisere alle med faktisk kreftsykdom. I slike tilfeller kan det være naturlig å akseptere flere falske positive enn falske negative. Medisinskfaglige vurderinger vil ofte være tungtveiende i slike avveininger, men det er heller ikke mulig å se helt bort fra samfunnsmessige hensyn, eksempelvis til kostnader ved oppfølging av en stor andel falske positive.

Gjennom ROC-kurver[68] kombinerer man data om sensitivitet (y-aksen) og spesifisitet (x-aksen) i et diagram. Slike kurver ligger til grunn for en samlet kvalitetsparameter for å evaluere kvaliteten på en algoritme; jo større AUC (arealet under kurven) er, desto bedre samlet treffsikkerhet har algoritmen. AUC kan teoretisk variere mellom 0 og 1, der en verdi på 0,5 (der kurven faller langs diagonalen), er uttrykk for at algoritmen ikke treffer bedre enn ved tilfeldighet. Verdier på mellom 0,8 og 0,9 vurderes som svært gode, verdier i intervallet 0,6-0,8 som middels gode, mens verdier mellom 0,5 og 0.6 vurderes som dårlige. ROC-kurven gir også en god grafisk illustrasjon for sammenhengen mellom sensitivitet og spesifisitet.



Figur 5: Illustrasjon av en ROC-kurve. Hentet fra: cmglee, MartinThoma - Roc-draft-xkcd-style.svg, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=109730045

I tillegg til klassifisering er en av de vanligste oppgavene til en KI-algoritme innen radiologi å gjøre segmentering av patologi, altså å automatisk identifisere hvilket piksler i bildet som viser f.eks. en svulst. Dice-koeffisienten er et mål på graden av overlapp mellom en automatisk segmentering (av f.eks. svulster) og referansesegmentering utført på konvensjonell, manuell måte. Jo nærmere koeffisienten er 1 (full overlapping), jo bedre er den automatiske segmenteringen. For større strukturer gir imidlertid Dice-koeffisienten liten mening, siden den ikke spesifiserer hvordan grensene overlapper. I slike tilfeller anbefales heller målet gjennomsnittlig overflatedistanse (surface distance) og Hausdorff distanse for å evaluere en KI-algoritme.

16.3 Kriterier for vurdering av algoritmers kvalitet – radiologer som komparenter

Det er naturlig at KI-algoritmer blir sammenlignet med beslutninger fattet av eksperter, i dette tilfelle radiologer. Sammenligningen gjelder både spørsmålet om hvordan en

[68]Wikipedia: En ROC-kurve er et grafisk plott som illustrerer den diagnostiske evnen til et binært klassifiseringssystem når dets diskrimineringsterskel varierer.

algoritme kommer fram til et resultat og selve resultatet. Det er både rimelig og forståelig at det stiles kritiske spørsmål til den manglende transparens som hefter ved de beregninger som skjer i «black-boxen», og til den usikkerhet som følger av manglende innsikt om hvordan en algoritme vil «oppføre seg» i nye uventede kontekster. Det er også en selvfølge at resultatene av en ny metode sammenlignes med anerkjente etablerte metoder (=ekspertvurderinger). Hva som er mer overraskende, er at det samme kritiske blikk i begrenset grad har vært rettet mot prosesser og resultater knyttet til menneskelige (ekspert)-beslutninger, i dette tilfellet radiologenes beslutninger. I en publikasjon fra 2021 reises derfor det retoriske spørsmålet om det stilles strengere krav og forventninger til KI-algoritmer enn til radiologer.[69]

I lys av nyere forskning innen kognitiv psykologi, nevrovitenskap mv., er det gode grunner til å modifisere våre overdrevne forestillinger om menneskehjernens evne til å håndtere store om komplekse informasjonsmengder og til å foreta «korrekte» vurderinger og beslutninger. Nobelprisvinneren Kahneman og medarbeideres epokegjørende studier om skjevhet (systematiske avvik)[70] og støy (tilfeldig spredning)[71] i beslutninger, er utvilsomt blant de viktigste oppsummerende bidrag som belyser begrensninger og feilkilder ved menneskelig beslutningsatferd. Flere av de studiene det refereres til, omhandler vurderinger og beslutninger foretatt av eksperter, inkludert medisinere og radiologer.

At det fins betydelige innslag av uberettiget variasjon i medisinsk diagnostikk og behandling, har i flere tiår vært dokumentert gjennom helseatlas-prosjektet ved Dartmouth College[72] i USA. Radiologenes diagnostiske beslutninger er neppe preget av mer støy enn beslutninger innen andre deler av medisinen, men den lar seg enklere dokumentere ved å gå tilbake til lagrede skanninger, bilder og beskrivelser, og vurdere dem på nytt. Dermed kan man ikke bare sammenligne variasjon i diagnostisk treffsikkerhet mellom radiologer (interpersonlig støy), men også variasjon i den enkelte radiologs vurdering av det samme bilde ved ulike tidspunkter (intrapersonlig støy). Interpersonlig støy innen medisinen kan bl.a. måles ved hjelp av *korrelasjons-koeffisienten kappa*[73]: jo høyere kappa, desto mindre støy. En kappaverdi på 1 angir således perfekt samsvar, mens en kappa verdi på 0 gjenspeiler «nøyaktig så mye samsvar som kan forventes mellom aper som kaster piler mot en liste over mulige diagnoser»[74].

---

[69] Mazurowski, MA: Do We Expect More from Radiology AI than from radiologists? *Radiology Artificial Intelligence 2021:3 (4)e200221,* https://doi.org/10./1148/ryai.2021200221
[70] Kahneman, Daniel: Thinking, fast and slow. Farrar, Straus and Giroux, 2011.
[71] Kahneman, D, Sibony, O and Sunstein, C: Noise – a flaw in human judgement, 2021
[72] Wennberg, John: Tracking Medicine: A Researchers Quest to Understand Health Care. 2011
[73] Wald, R, Bell CM, Nisenbaum, R et al: Interobserver Reliability of Urine Sediment Interpretation, *Clinical Journal of the American Society of Nephrology* 4, no 3(March 2009)
[74] Hahneman, D, Sobony, O og Sunstein CR: et al: Støy – når dømmekraften svikter, Pax Forlag 2021(norsk versjon) side 279

Allerede i 1997 ble den betydelige variasjonen mellom ulike radiologers diagnosesetting karakterisert som deres «achilleshæl»[75]. I deres siste bok (2021) referer Hahneman og medarbeidere til en rekke studier som dokumenterer slik variasjon for flere ulike diagnosegrupper. Det er flere grunner til slik interpersonlig variasjon; forskjeller i radiologenes individuelle ferdigheter påpekes i flere studier. I en studie av lungebetennelsesdiagnoser der det ble konstatert høy grad av støy, kunne variasjon i ferdigheter forklare hele 44 prosent av variasjonen[76] i diagnosesettingen. Dette viser at seleksjon, opplæring og erfaring kan være viktige virkemidler for å redusere støy.

Stress og tretthet blant radiologene kan også påvirke støyen. Som tidligere omtalt i rapporten (kap. 3), bidrar den økende arbeidsbyrden blant radiologene til at de ofte kommer på etterskudd i arbeidet, noe som fører til å mer stress og tretthet, med ledsagende økt risiko for økt variasjon i diagnosesetting, også blant høyt kompetente fagfolk. Det er eksempelvis dokumentert variasjon i diagnosesetting mellom undersøkelser foretatt tidlig og sent på dagen[77]. Gjeldende kunnskapsstatus innen et diagnostisk område, er en tredje faktor som kan påvirke støyen; på områder med vage kriterier og kompliserte vurderinger, kan den interpersonlige variasjonen bli vesentlig større enn der kunnskapsgrunnlaget er godt.

Størrelsen på variasjonen varierer både mellom diagnosegrupper og mellom ulike studier og mellom nasjoner. For mammografiscreening er det dokumentert at radiologer i utviklede land detekterer over 80 prosent av brystkrefttilfellene, mens tilsvarende tall i utviklingsland bare ligger på 40 prosent[78]. Den gjennomsnittlig høyere diagnostiske presisjon i land med et velutviklet helsevesen, ledsages imidlertid av en til dels betydelig variasjon mellom radiologer. De mest ekstreme variasjoner vi har funnet dokumentert, er fra en studie av mammografiscreening i USA fra 1997; her varierte andelen falske negative funn fra 0 til 50 prosent, mens andelen falske positive funn varierte mellom 1 og 64 prosent[79]. De fleste nyere studier viser lavere, men likevel signifikante og til dels vesentlige variasjoner mellom radiologer. Om vi holder oss til diagnostikk av brystkreft, konkluderer en omfattende studie (n=216 783 mammogrammer) fra 2016 som følger:

> *«…there is a wide variation among radiologist in the percent of mammograms rated dense (ranging from 6,3 % 84,5 %), which persist after adjustment for patient factors. Additionally, 17,2 % of women (more than 1 in 6) with consecutive*

[75] Robinson, PJ: Radiology's Achilles Heel: Error and Variation in the Interpretation of the Rontgen Image. *British Journal of Radiology,* 70 (1997)
[76] Chan, DC et al: Selection with Variation in Diagnostic Skill: Evidence from Radiologists. National Bureau of Economic Research NBDR working document 26467, november 2019
[77] Backmann HA, Larsen M, Danielsen AS, Hofvind S: Time of day and mamograhic reader performance in a population-based breast cancer screening program. *Journal of Medical Screening* 2021:28 (3): 295-301, doi:10.1177/096914132953206
[78] Bray, F et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J.Clin. 68, 394-424*
[79] Craig, A Beam et al. Variability in the Interpretation of mammograms by US Radiologists, *Archives of Internal Medicine 156 (1996)*

mammograms interpreted by different radiologists during a short time period, are re-classified in dense versus non-dense categories»[80]

En annen studie (n=318 366 radiologirapporter) tar for seg variasjon i radiologenes oppfølgende (follow-up) bildetaking (tilleggsundersøkelser) på tvers av modaliteter. Resultatene viser en variasjon på 1:2,8 -1:6,7, avhengig av radiologens subspesialitet[81]. Det er bekymringsfullt dersom funn med indikasjoner for nye bilder ikke blir fulgt opp, men antagelig minst like problematisk at radiologer tar mange tilleggsbilder på sviktende indikasjon. Det innebærer en sløsing med knappe radiologressurser med konsekvenser både for helseøkonomien og ventetiden for nødvendige undersøkelser.

Slike variasjoner som følge av støy, er på ingen måte unik for radiologer eller andre spesialiteter innen medisin. Hahneman og medarbeidere viser i sin oppsummerende bok til en rekke undersøkelser fra andre yrkesgrupper (eksempelvis dommere, personalledere mv.) der variasjonen er minst like stor. I denne boken fremheves faktisk legene (inkludert radiologene) for det betydelige støyreduserende arbeid som er nedlagt gjennom utforming av retningslinjer[82] og protokoller. Hovedfunnet i denne forskningen om skjevhet og støy, er at regler i de fleste tilfeller gir bedre beslutninger enn skjønnsutøvelse, også for eksperter.

Vurderinger av algoritmers kvalitet, i betydningen treffsikkerhet (accuracy), skjer vanligvis ved å sammenligne resultatene fra algoritmene med en fasit (ground truth) som kan framkomme på flere måter (histopatologi mv.). Samme fasit legges til grunn for å vurdere treffsikkerheten for de resultater erfarne og kompetente radiologer kommer fram til. Dermed har man en referanse for å vurdere algoritmens kvalitet; man sammenligner med andre ord en ny metode (KI-algoritmer) med en konvensjonell eller tradisjonell metode (= faglige vurderinger av radiologer). Vanligvis vil man ikke anbefale eller ta i bruk en ny metode, med mindre den er likeverdig eller bedre enn etablerte metoder, eventuelt gir bedre kost-nytte. Men i lys av den forskningen om variasjon i diagnosesetting som vi nettopp har redegjort for, er det neppe rimelig å la fagelitens diagnostiske treffsikkerhet være målestokk for evaluering av en algoritme. En mer rettferdig, realistisk og objektiv sammenligning ville være å vurdere resultatene fra algoritmene mot de vurderinger som gjøres av et representativt utvalg av radiologer. I erkjennelsen av at også radiologer gjør feilvurderinger, bør følgelig gjennomsnittsverdier for radiologenes diagnosesetting legges til grunn ved evaluering av en algoritme.

---

[80] Sprague BL, Conant EF, Onega T et al. Variation in mammographic breast density assessments among radiologists in clinical practice: Findings from a multicenter observational study. Ann Intern Medicine 2016, Oct 4; 165 (7):457-464

[81] Cochon, LT, Kapoor, N, Carrodeguas, E et al. et al. Variation in Follow-up Imaging in Radiology Reports: Patient, Modality and Radiologits Predictors. Radiology, Vol 291, No 3, 2019 https://doi.org/10.1148/radiol.2019182826

[82] Hahneman et al, op.cit. (side 281 i norsk utgave)

I den forannevnte publikasjonen om hvorvidt vi forventer mer av algoritmene enn av radiologene, oppsummerer forfatteren avveiningen slik:

> *«On the one hand, we should, to a reasonable extent, mistrust AI to make correct decisions regardless of the setting. Therefore, it is crucial that we test the algorithms in real world scenarios and on datasets that are diverse in terms of patient populations, scanner parameters, imaging protocols, technologists who acquire the images, and any other parameters that may reasonably affect the decision.*
>
> *On the other hand, putting an undue burden on the algorithms because we trust ourselves as human beings may not be a correct policy. The goal should be to implement a radiologic practice that provide the most benefit to the patients. This needs to be based on a reasonable mistrust toward the algorithm, but it should not be based on an unjustified trust toward our own human ability to perceive and make decisions. »*

16.4 Hvordan anvendes algoritmen? – arbeidsdeling mellom algoritme og fagperson

Grensesnittene mellom menneske og maskin er flere; noen av disse gjelder tilpasningen av teknologi til fagfolkenes arbeidsprosesser og arbeidsflyt, slik det er redegjort for i kapittel 11. Her fokuserer vi imidlertid på hvordan en KI-algoritme anvendes i samspill med fagfolk (radiologer). Bruk av KI-algoritmer må tilfredsstille lovgivningens krav til faglig forsvarlighet.[83] I mange tilfeller vil forsvarlighet forutsette en viss menneskelig kontroll med bruken av algoritmene, særlig ved viktige avgjørelser, som diagnostikk. Derfor vil det ofte være mest nærliggende å bruke KI-algoritmer som beslutningsstøtte. Bruk av KI som beslutningsstøtte kan imidlertid innebære ulik grad av menneskelig involvering og kontroll. Hvilken type beslutningsstøtte som er forsvarlig, må vurderes konkret i de ulike situasjoner KI skal benyttes.

Beslutningsstøtte innebærer en arbeidsdeling der KI tas i bruk på måter som avlaster radiologene ved å automatisere deler av det samlede arbeidet med å tolke bildene. Dette kan imidlertid gjøres på ulike måter. Den enkleste - og vel vanligste - måten har vært å la algoritmen være «second reader» i parallell med en radiolog. For å unngå at radiologen påvirkes av algoritmen i sin vurdering, bør dette helst skje ved at radiologen gjør sin vurdering uten på forhånd å kjenne «svaret» fra algoritmen, og at eventuelle diskrepanser avklares av en tredje fagperson før endelig diagnose fastsettes. En slik bruk antas å ivareta forsvarlighetskravet, ettersom alle pasienter vurderes av en fagperson.

[83] Se helsepersonelloven §§ 4 og 16, spesialisthelsetjenesteloven § 2-2 og helse- og omsorgstjenesteloven § 4-1.

En annen anvendelse som i noen tilfeller antas å være forsvarlig, er å bruke en sertifisert og kvalitetssikret algoritme for triagering av pasienter under vakt. Algoritmen identifiserer da pasienter med sannsynlig sykdom, slik at disse kan prioriteres for tolkning av den radiologen som er på vakt. De pasientene som algoritmen identifiserer som sannsynlig negative, kan da håndteres manuelt på et senere tidspunkt, eksempelvis påfølgende dag. Dermed kan algoritmen både avlaste vaktlegen og bidra til at vedkommende prioriterer oppmerksomheten mot pasienter med større hastegrad og behov. For pasienter som ankommer en akuttenhet for å få avklart mulige bruddskader i ekstremiteter, vil en slik bruk av algoritmen innebære liten risiko og samtidig være til gavn for pasienten som slipper å sitte i mange timer i påvente av at vakthavende radiolog skal få tid til å granske et bilde. Om falske negative ble håndtert dagen etter, ville det sjelden ha noen negative konsekvenser for behandlingen.

Om derimot en algoritme skulle anvendes på tilsvarende måte for å triagere og identifisere pasienter med hjerneslag med sikte på å identifisere mulig blødning, ville risikoen være svært stor ved falske negative funn. Fordi pasienter med blødning ikke skal behandles med trombolyse, kan det i dette tilfelle reises spørsmål om forsvarlighetskravet blir ivaretatt ved å bruke algoritmen for triagering.

Imidlertid er det i de senere år presentert studier der arbeidsdelingen er mer kompleks og der de respektive fortrinn hos henholdsvis radiolog og algoritme kombineres på måter som – foreløpig gjennom retrospektive studier[84] – har vist å gi like bra eller bedre resultat enn ved bruk av radiolog alene eller algoritme alene. Slike studier er gjort i mammografiscreening hvor algoritmen gir mammogrammene en skåre som sorterer kvinner etter grad av risiko for brystkreft. Deretter «henviser» algoritmen de pasientene som algoritmen har klassifisert som sannsynlig syke og de som befinner seg i gråsonen, for undersøkelse av radiolog.

I en lignende retrospektiv studie[85] av brystkreftscreening, var fokus særlig rettet mot intervallkrefttilfeller. Også i denne studien ble risikostratifisering lagt til grunn for arbeidsdelingen mellom algoritmen og radiologene. Dermed kan radiologressursene som settes inn tilpasses de ulike risikogruppene. I denne studien ble det konkludert med at om lag 40% av intervallkrefttilfellene som er tydet negativt av radiologene, ble gitt en høy score av KI; noe som dermed åpnet mulighet for å kunne oppdage disse på i et tidligere stadium. Begge studiene postulerer betydelig potensial for å redusere

[84] Leibig C, Brehmer M, Bunk S, et. al. Combining the strength of radiologists and AI for breast cancer screening: a retrospective analysis, *Lancet Digit Health 2022: 4:e507-19*

[85] Dembrower K, Wåhlin E, Liu Y et. Al: Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologists workload: a retrospective simulation study. *The Lancet Digital Health,* September 2020 DOI:https://doi.org/10.1016/S2589-7500(20)30185-0

radiologens tydevolum. Slike design for arbeidsdeling mellom algoritmer og radiolog, er utviklet gjennom forskningsprosjekter, og foreløpig ikke utprøvd i prospektive kliniske studier. Det foreligger imidlertid aktuelle planer om gjennomføring av prospektive randomiserte studier innen mammografiscreening, både i Norge og andre land[86]. I Norge starter slike randomiserte studier ved de fire brystdiagnostiske sentre i Helse Vest i løpet av første halvår 2023, og deretter med planlagt oppfølging i Helse Midt-Norge, Helse Nord og Helse Sør-Øst. En tilsvarende studie pågår i Malmø i Sverige, mens Københavnregionen i Danmark har implementert KI som verktøy i tolking av screeningmammogrammer. Andre, mindre studier pågår i ulike land, mens større, randomiserte studer planlegges i England og Tyskland.

Det foregår også innovativ forskning[87] på partnerskapet mellom menneske og maskin, der radiologer gjennom virtuelle nettverk - designet som «human swarms»[88] - samarbeider innbyrdes og i samspill med dyplæringsalgoritmer; gjennom iterative synkrone prosesser justeres resultatene fra algoritmen gjennom feedback i flere runder. Ved analyse av konvensjonelle røntgenbilder av thorax, var resultatene av dette samspillet mellom radiologer - organisert som «human swarms» - og algoritmer, klart bedre enn ved bruk av algoritmene alene, eller kollektiver av radiologer alene.

> *«Human-in-the loop» (HITL) AI may offer advantages where both radiologists and machine-learning algorithms fall short. This paradigm allows leveraging all the advantages of AI models (i.e., rapid automated detection) but having a human at various checkpoints to fill gaps where algorithms are not confident in their probabilities or where they may fall short due to underlying biases.»*[89]

Det presiseres at dette dreier seg om forsknings- og utviklingsarbeid som foreløpig ikke er implementert i kliniske settinger. Disse studiene kan likevel være et frempek mot framtidige løsninger og arbeidsprosesser som kombinerer algoritmer og humane faglige vurderinger på nye måter.

Arbeidsgruppen vurderer det som viktig at det skjer en tilpasning mellom kvalitetssikret faglig utvikling og de lover og forskrifter som skal regulere bruken av algoritmene. Det innebærer at begrepet beslutningsstøtte må tolkes på en dynamisk måte som fanger inn nye og forbedrede metoder for anvendelse av kunstig intelligens i helsetjenesten. Kravene til å ta i bruk nye metoder, må være at de ivaretar hensynet til forsvarlighet og pasientsikkerhet minst like godt som eksisterende metoder.

---

[86] Meddelelse fra professor Solveig Hofvind, leder av Mammografiprogrammet ved Kreftregisteret
[87] Patel BN, Rosenberg L, Willcox, G et al. Human-machine partnership with artificial intelligence for chest radiograph diagnosis, *npj Digital Medicine* **2,** Article number: 111 (2019)
[88] «Human swarms» henter i generisk forstand ideer fra «biological swarms», men der sistnevnte er biologisk programmert for innbyrdes samarbeid som styrker kollektiv samordning og atferd, vil førstnevnte være designet må måter som styrker kollektiv intelligens.
[89] Patel BN og medarbeidere (2019), opcit.

17 Forvaltning

Innføring av kunstig intelligens krever god forvaltning av de ulike KI-løsningene. Løsningene som anskaffes vil være regionale, og forvaltningsansvaret bør ligge på et tverrfaglig regionalt forum bestående med representanter fra Helse Nord IKT, Felles funksjonell forvaltning av kliniske system og fagmiljøet. Det må foreligge klare funksjonsbeskrivelser samt beskrivelse av oppgaver mellom de ulike instansene i det tverrfaglige forumet.

De felles regionale kliniske IKT-løsningene forvaltes i dag i delt løsning mellom Felles funksjonell forvaltning kliniske system plassert i UNN HF og Helse Nord IKT (teknisk og merkantil forvaltning). Det er innenfor denne modellen etablert flere fora som Systemforvaltningsforum og arbeidsgrupper som kan inngå i en modell for forvaltning av KI-løsninger.

Forvaltning innebærer blant annet å sørge for at løsningene som skal benyttes til enhver tid er oppdaterte, og har siste versjon fra leverandør. Forvaltning innebærer også å følge opp feil og mangler som brukerne kan oppleve mot leverandør. For å lykkes med implementering av KI – løsninger er det viktig at brukerne opplever hurtig feil-søkning og løsning, og at de kommer til rett instans for dette ved første kontakt. Det bør etableres regionale «Help- desk» som kjenner de ulike KI -løsningene godt, og som også kjenner hverdagen i røntgenavdelingene.

En annen viktig del av forvaltning av regionale løsninger er utarbeidelse av arbeidsflyt. Det bør utarbeides regionale retningslinjer hvor hvordan KI – løsningene skal tas i bruk i helseforetakene. Et viktig element i dette er opprettelsen av definerte bruker – roller som kan tildeles den aktuelle radiograf/ radiolog ved hvert foretak. Siden foretakene har forskjellige behov bør de definerte bruker-rollene inneha en viss fleksibilitet slik at hvert enkelt foretak kan tilpasse disse til egne behov. I tillegg må det etableres superbrukere i avdelingene som har tilstrekkelig med avsatt tid, kunnskap og tilganger til å kunne bidra til at de ulike løsningene «skreddersys» til hvert enkelt foretak. Samtidig må forvaltningsforumet bidra med overordnede føringer slik at det foreligger et godt system rundt bruk av slike løsninger med tanke på for eksempel opplæring (se kap. 15) og vedlikehold av denne.

En viktig del av forvaltning er beslutningsmyndighet i forhold til for eksempel utvidelse av eksisterende løsninger eller oppgraderinger av disse. Dette innebærer ofte kostnader, ikke bare til leverandør, men også kostander tilknyttet oppgradering av nettverk med mer. Innen rammen av godkjent budsjett, anbefales at Forvaltningsforumet gis myndighet til å fatte beslutninger tilknyttet utvidelse av allerede eksisterende løsninger. Det er derfor særdeles viktig at deltakere i et slikt forum har tilstrekkelig med kompetanse og forståelse for hensikten med KI applikasjonen slik at man jobber mot et samstemt mål.

Litteraturliste:

AIforradiology.com

Antun V og medforfattere, «On instabilities of deep learning in image reconstruction and the potential costs of AI» *PNAS,* vol. 117, nr. 48, pp. 30088-30095, 2020.

Artificial Intelligence Models for Safety, Quality, Workflow, and Education Applications in Radiology Practice. Radiol Artif Intell. 2022 Feb 2;4(2):e210114. doi: 10.1148/ryai.210114. PMID: 35391770; PMCID: PMC8980942.

Backmann HA, Larsen M, Danielsen AS, Hofvind S: Time of day and mamograhic reader performance in a population-based breast cancer screening program. *Journal of Medical Screening* 2021:28 (3): 295-301, doi:10.1177/096914132953206

Bray, F et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J.Clin. 68, 394-424*

Bruls, R.J.M.; Kwee, R.M. Workload for radiologists during on-call hours. Dramatic increase in the past 15 years. *Insight Imaging* 2020,11,121

Chan, DC et al: Selection with Variation in Diagnostic Skill: Evidence from Radiologists. National Bureau of Economic Research NBDR working document 26467, november 2019

Cochon, LT, Kapoor, N, Carrodeguas, E et al. et al. Variation in Follow-up Imaging in Radiology Reports: Patient, Modality and Radiologits Predictors. Radiology, Vol 291, No 3, 2019 https://doi.org/10.1148/radiol.2019182826

Cyert, RM and March, JG: A Behavioral Theory of the Firm, John Wiley and Sons Ltd 1963

Craig, A Beam et al. Variability in the Interpretation of mammograms by US Radiologists, *Archives of Internal Medicine 156 (1996)*

Dembrower K, Wåhlin E, Liu Y et. Al: Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologists workload: a retrospective simulation study. *The Lancet Digital Health,* September 2020 DOI:https://doi.org/10.1016/S2589-7500(20)30185-0

European Commission. Factsheet for Manufacturers of Medical Devices, www.dnv.com

European Commission. Guidelines on medical devices. Clinical evaluation: a guide for manufacturers and notifies bodies under directives 93/42EEC and 90/385EECMEDDEV 2.7/l revision 4 2016

FDA: Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices, October 5, 2022, Update.

Gore, JC. Artificial intelligence in medical imaging. Magn Reson Imaging 2020:68 A!-A4

Helsedirektoratet: Strategi for rasjonell bruk av bildediagnostikk 2019

Helseforskningsloven § 32

Helse- og omsorgstjenesteloven § 4-1.

Helsepersonelloven §§ 4 og 16

Henschke CI, Yankelevitz DF, Mirtcheva R, et al. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. Am J Roentgenol. 2002;178:1053–1057.

Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. Lancet Oncol 2014; 15: 1332–1341.

Husreau, Don, Drummond, Michael, Pwetrou, Stavous et al: Consolidated Health Economic Evaluation Standards (CHEERS) – Explanation and Elaboration: a report of the
ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force, in Value Health. Mr-Apr 2013;16(2):231-50

Huserau, Don, Drummond, Michael, Augustovsk, Fredricoi et al: Consolidated Health Economic Evaluation Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations, in BMJ, 2021:375: e0.67975

Kahneman, Daniel: Thinking, fast and slow. Farrar, Straus and Giroux, 2011.

Kahneman, D, Sobony, O og Sunstein CR: et al: Støy – når dømmekraften svikter, Pax Forlag 2021(norsk versjon)

Karabin: Kartlegging av radiologi- og operasjonskapasitet i Helse Nord. Rapport med anbefalinger. 1. juli 2022.

KIN**:** Forslag om etablering av arbeidsgruppe for utvikling av retningslinjer for validering av AI-verktøy til bruk i helsetjenesten – innspill fra sekretariatet i KIN. Brev av 30.09.2022 adressert til Helsedirektoratet

Leibig C, Brehmer M, Bunk S, et. al. Combining the strength of radiologists and AI for breast cancer screening: a retrospective analysis, *Lancet Digit Health 2022: 4:e507-19*

Liu, X, Rivera, SC: Reporting Guidelines for clinical trial reports interventions involving artificial intelligence: the CONSORT-AI extension), Nature Medicine 26, 1364-1374 (2020)

Makhlysheva A, Ruiz LM, Olsen, TS et al: Implementation of Artificial intelligence in Norwegian healthcare: The road to broad adoption: https://ehealthresearch.no/rapporter/implementation-of-artificial-intelligence-in-norwegian-healthcare-the-road-to-broad-adoption

Mazurowski, MA: Do We Expect More from Radiology AI than from radiologists? *Radiology Artificial Intelligence 2021:3 (4)e200221,* https://doi.org/10./1148/ryai.2021200221

Mateusz Winder et al.: Are We Overdoing It? Changes in Diagnostic Imaging Workload during the Years 2010-2020 including the Impact the Saars-CoV-2 Pandemic, *Health Care* 2021, 9, 1557

Minssen T, Gerke S, Aboy M, et al. Regulatory responses to medical machine learning J.LawBiosc 220; 7:1-18.DOI:10.1093/jlb/Isaa02

Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic reviews of design, reporting standards and claims of deep learning studies. *Br.Med J* 2020; 368: m689.

Niemiec, Emilia: Will the EU Medical Device Regulation help to improve the safety and Performance of medical ASI devices? *Digital Health* 2022 Volum .8;1-DOI:101177/20552076221089079journals.sagepub.com/home/dhj

Nguyen T, Maarek R, Ducou H, et al. Assessment of an artificial intelligence aid for the detection of appendicular skeletal fractures in children and young adults by senior and junior radiologists, *Pediatric Radiology* **52**, 2215-2226 (2022)

Njølstad T og medforfattere, «Improved image quality in abdominal computed tomography reconstructed with a novel Deep Learning Image Reconstruction technique–initial clinical experience.,» *Acta Radiologica Open,* vol. 10, nr. 4, p. 20584601211008391, 2021.

Patel BN, Rosenberg L, Willcox, G et al. Human-machine partnership with artificial intelligence for chest radiograph diagnosis, *npj Digital Medicine* **2,** Article number: 111 (2019)
Personvernforordningen artikkel 5.
Plana D: Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review, JAMA Netw Open 2022, pubmed.ncbi.nhi.gov
Robinson, PJ: Radiology's Achilles Heel: Error and Variation in the Interpretation of the Rontgen Image. *British Journal of Radiology,* 70 (1997)
Sinsky, Harry E. et al: Physician Task Load and the Risk of Burnout Among US Phycisians in a National Survey Jt. Comm Jt Qual.Patient Saf 2021, 47,76-85
Sounderja, V, Ashrafian, H, Golub, RB et al.: Developing a reporting Guideline for artificial intelligence-centered diagnostic test accuracy studies: - STARD-AI protocol, BMJ Open 2021
Spesialisthelsetjenesteloven § 2-2
Sprague BL, Conant EF, Onega T et al. Variation in mammographic breast density assessments among radiologists in clinical practice: Findings from a multicenter observational study. Ann Intern Medicine 2016, Oct 4; 165 (7):457-464
Strategi for kunstig intelligens i Helse Nord 2022-2025. Helse Nord RHF juni 2021
Tadavarthi Y, Makeeva V, Wagstaff W et al. Overview of Noninterpretive Artificial intelligence Models for Safety, Quality, Workflow, and Education Applications in Radiology Practice. Radiol Artif Intell. 2022 Feb 2;4(2):e210114. doi: 10.1148/ryai.210114. PMID: 35391770; PMCID: PMC8980942.
The DECIDE-AI Steering Group: DECIDE: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence, Nature Medicine 27, 186-187 (2021)
US Food and Drug Administration: Artificial Intelligence Machine Learning, Volum 11, issue 6 (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan 2021.
Van Leeuwen KG, Schalekamp S, Rutten M, et al. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol* 2021;31: 3797-3804. DOI:10.1007/s00330-021-07892-z
Van Leeuwen, KG et al.: Cost-effectiveness of artificial intelligence aides vessel occlusion detection in acute stoke: an early health technology assessment, in Insight Imaging (2021) 12:133
Vasey B, Utsprung S, Beddou B, et al. Association of clinician diagnostic performance with machine learning-based decision support systems: a systematic review. *JAMA Netw Open* 2021; 4:e211276
Voets, Adelon M. et al: Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare. The Tortoise and the Cheetah. Value Health 2022:25 (3):340-349
Wahidi MM, Govert JA, Goudar RK, et al. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer? ACCP evidence-based clinical practice guidelines (2nd edition). Chest 2007; 132: Suppl., 94S–107S.
Wald, R, Bell CM, Nisenbaum, R et al: Interobserver Reliability of Urine Sediment Interpretation, *Clinical Journal of the American Society of Nephrology* 4, no 3(March 2009)
Wennberg, John: Tracking Medicine: A Researchers Quest to Understand Health Care. Oxford University Press 2010

Wu E, Wu K, Daneshjou, R, et al. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med* 2021;27: 576-584

Vedlegg 1: Ordliste - radiologirapporten

Abdomen: Buk
Algoritme: Algoritme er i matematikk og databehandling en fullstendig og nøyaktig beskrivelse av fremgangsmåten for løsning av en beregningsoppgave eller annen oppgave.
Angiografi: Konvensjonell angio hvor man går inn i lyske med kateter
Annotering: Merking av røntgenbilde
Augmented reality: utvidet virkelighet, der man legger et lag med digitalt tilleggsinnhold oppå et bilde, gjerne levert direkte fra et kamera
Cancer mammae: Brystkreft
Caput: hode
CE-merking: Er produktmerking som bekrefter at et produkt er i samsvar med alle relevante EU-direktiver og regelverk hva angår produktsikkerhet,
Cervix: livmor
CHEERS: (Consolidated Health Economic Evaluation Reporting Standards): Et internasjonalt rammeverk for helseøkonomiske analyser
CT-angiografi: Fremstilling av blodkar ved hjelp av CT- eller MR-teknologi og kontrast
PET/CT: Kombinasjon av positron emisjons tomografi og computerized tomography Bilder fra PET-skanning vil avsløre opphopning av celler i høy aktivitet, slik som kreftceller. CT-skanning viser hvor i kroppen denne opphopningen befinner seg, samt omfanget av dette.
DICOM: Etablerte standarder for lagring og overføring av medisinske bildedata og informasjon knyttet til disse.
Differensialdiagnose: Betegnelse på alternative sykdommer på et spesielt symptom eller funn
DoMore: Digitalt patologi prosjekt ved OUS som bruker KI. Er nå kommersialisert
FDA: (Food and Drug Administration): Medisinsk sertifiserings- og godkjenningsinstans (USA)
Flow: Hvor fort kontrasten sprøytes inn i kroppen
Fraktur: brudd
Gateway (hub): Sammenkopling av flere datanettverk
GDPR = Personvernforordningen (Forordning 2016/679, på engelsk General Data Protection Regulation, forkortet GDPR).
Ground truth: Informasjon som er bekreftet sann ved observasjon eller måling
HF: Helseforetak (f.eks. Finnmarkssykehuset HF, Helgelandssykehuset HF)
HTA: Helseteknologivurdering (Health Technology Assessment)
Ileus: tarmslyng
Iterativ: Betyr repeterende prosesser i flertrinns justerende feedback-loops
KI: Kunstig intelligens; (Engelsk AI – Artificial Intelligence). Bruk av algoritmer til å analysere data i sanntid og løse problemer basert på maskinlæring eller dyp læring.
Kryptering av data: Kryptering er en sikkerhetsmetode som konverterer data/tekst til en hemmelig kode for å unngå avlytting eller utilsiktet offentliggjøring.

Lesjon: Fellesbetegnelse for alle lokaliserte skader på kroppen, f.eks. ytre vold, svulst, infeksjon
Lungenoduli: fortetning på et bilde som kan være normalt eller patologi (f.eks. kreft)
Lungeemboli: Blodpropp i lungene
Maskinlæring: Dataprogram som lærer og løser oppgaver uten å være forhåndsprogrammert til å gjøre det. Maskinen lærer automatisk (dynamisk prosess) å forbedre prosessene for å løse problemene.
Metastaser: Spredning av kreft
MDD (Medical Device Directive): EU-forordning for godkjenning av medisinsk utstyr
MDR (Medical Device Regulative): avløser MDD for godkjenning av medisinsk utstyr
Modalitet: Brukes her om ulike typer undersøkelsesmetoder (CT, MR, ultralyd PET mv.),
MMIV: Mohn Medical Imaging and Visualization Centre, lokalisert i Bergen
MS: Multippel sklerose
MSK: Muskel/skjelett
NCRP: (Nomesko Classification of Radiology Procedures): Klassifikasjonssystem for radiologi
NSE: Nasjonalt senter for ehelseforskning ved Universitetssykehuset Nord-Norge (UNN)
Omikk; -omics: Suffiks som sier noe om at man studerer helheten av noe til forskjell fra Enkeltelementer. I denne sammenhengen menes somoftest biologisk informasjon på molekylært nivå f.eks.: genetisk informasjon (Genomikk; Genomics), eller informasjon fra proteiner (proteomikk),
PACS: Picture Archive Classification System: bildelagringssystem for røntgen
Pancreas: Bukspyttkjertel
Pelvimetri: Måling av fødselskanalens rommelighet ved måling avbestemte diametre
PET: (Positron Emission Tomography): Bildeundersøkelse som bruker radioaktive isotoper
Pneumothorax: Punktert lunge
Postprosessering: Databehandling av røntgenbilder for f.eks. å fjerne støy
Presisjonsmedisin: Forebygging, diagnostikk, behandling og oppfølging tilpasset biologiske forhold hos den enkelte og ikke gjennomsnittet til en større populasjon.
Prospektive studier: Studier der en ser på utvikling over tid for å fastslå utfallet av en intervensjon hos individene i studien (f.eks. virkning av et legemiddel).
Rekonstruksjon: CT- og MR-bilder kan rekonstrueres for å få bilder i forskjellig plan eller tykkelse
Retrospektive studier: Studier der man studerer historiske data og relaterer disse til et allerede etablert utfall, f.eks. en sykdom pasienten allerede har.
RIS: (Radiology information system) Journalsystem for røntgen
ROS-analyse: Risiko- og sårbarhetsanalyse basert på sannsynlighet og konsekvens
Screening: Masseundersøkelse av nærmere avgrenset befolkningsgruppe
Segmentering: Klippe ut aktuell anatomi i et røntgenbilde
Sky; Skyløsning: ved skyløsning innebære at man trenger ikke laste ned og installere programvare på egen datamaskin, men løsningen kan brukes slik den er på nettet
Slack: Her brukt i betydningen organisatorisk slack, som betyr ledige, tilgjengelige ressurser
SPKI: Senter for pasientnær kunstig intelligens, ved UNN-Tromsø
Thorax: Lunge

Thyroeidea: Skjoldbruskkjertel
Tomosyntese: Tredimensjonal avbildningsteknikk ved diagnostikk av brystkreft
Treningsdata: Data som maskinlæringsalgoritmer bruker for å lære og justere operasjoner
Triagering: Vurdering av hastegrad
UiO: Universitetet i Oslo
UiT: Universitetet i Tromsø – Norges arktiske universitet
UNN = Universitetssykehuset Nord-Norge
OUS: Oslo Universitetssykehus
Valideringsdata = Data som maskinlæringsalgoritmer bruker for å teste og validere sine predikasjoner

Vedlegg 2:

Til Helsedirektoratet/koordineringsprosjektet
v/Siv Anette Fjellkårstad **Oslo 30.09.2022**

Forslag om etablering av arbeidsgruppe for utvikling av retningslinjer for validering av AI-verktøy til bruk i helsetjenesten – innspill fra sekretariatet i KIN.
Det antas i dag å være i overkant av 400 registrerte FDA-sertifiserte/CE-merkede AI-algoritmer med helsetjenesten som målgruppe. Oversikten over FDA-godkjente algoritmer er bedre enn for CE-merkede, men de mange KI-initiativ innen EU-systemet forventes å gi bedre oversikt og dokumentasjon også av CE-merkede produkter. Oversikten over CE-merkede algoritmer er best innen radiologi, takket være nettstedet AIforradiology.com. FDA rapporterte i oktober 2021 at det da fantes ca. 350 FDA godkjente algoritmer innrettet mot helse, hvorav 70% innen radiologi[90]. Det amerikanske og europeiske markedet er i stor grad overlappende, men vi har grunn til å anta at det fins et - kanskje ikke helt ubetydelig – antall CE-merkede produkter som ikke er FDA godkjente.

Det foreligger i varierende grad eksterne fagfellevurderte studier som har validert de sertifiserte produktene. En studie i regi av AIforradiology fra 2021 viser for eksempel at dette bare gjaldt en tredel av de CE-merkede algoritmene innen radiologi[91]. At mange av de CE-merkede produktene mangler vitenskapelig dokumentasjon, er utvilsomt en utfordring for tilliten til produktene i diagnostiske og kliniske miljøer som har kunnskapsbasert medisin som ledetråd for sitt arbeid. Samtidig vites at disse kravene om transparens og dokumentasjon er blitt skjerpet i det nye MDR som fra og med mai 2024 vil gjøres gjeldende for alle CE-merkede produkter.

Selv om det foreløpig er et begrenset antall av de FDA/CE-sertifiserte algoritmene som i noen betydelig skala er tatt i bruk i helsetjenesten, er dette antallet klart økende. En fersk oversikt fra Nederland[92] viser at stadig flere sykehus der nå har gjort kommersielle anskaffelser av KI-algoritmer. I Norge har Vestre Viken HF/Sykehuset Vestfold HF nettopp gjort en anskaffelse for løsninger innen radiologi, der det er mulig for samtlige helseforetak i Helse Sør-Øst og UNN HF i Helse Nord å gjøre avrop. Det innebærer at det i løpet av de kommende år med stor sannsynlighet vil bli implementert en rekke KI-algoritmer på dette feltet, også i Norge.

Det er avgjørende for tilliten til innføring av slike løsninger i norsk helsetjeneste, at de er kvalitetssikret før de implementeres i storskala. Slik kvalitetssikring vil være særlig

[90] https://www.iqvia.com/locations/united-states/blogs/2021/10/fda-publishes-approved-list-of-ai-ml-enabled-medical-devices
[91] Van Leuween, KG og medforfattere: Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *European Radiology*, pp 1-8, 2021
[92] Informasjon fra Nederland er hentet i Newsletter fra AIforradiology, 31.august 2022

kritisk i den første fasen av implementering som vi nå går inn i. Det kan likevel kunne være noe ulike vurderinger i fagmiljøene om i hvilken grad det er behov for ytterligere kvalitetssikring og eventuell validering av løsningene på lokale data.

Verktøy som benyttes til beslutningsstøtte el. som gir konsekvens for pasientbehandling blir sertifisert i henhold til EU-forordningen om medisinsk utstyr (CE-merket). Fagmiljøet rapporterer i dag usikkerhet omkring behov for ytterligere kvalitetssikring når verktøy skal tilpasses til lokale forhold – populasjon, kohortseleksjoner, behandlingsalgoritmer, dataformater og arbeidsflyt eller annet bruksscenario. Lokal kvalitetssikring kan gjøres ved at de anskaffede verktøyene prøves ut (valideres) mot lokale/regionale norske pasientdata, både i retrospektive og prospektive design. Dette er imidlertid kostbare og tidkrevende prosesser, og vil kunne medføre betydelige utsettelser av implementeringen. Det er behov for samarbeid mellom store og små aktører, samt en avklaring av hvilke forhold som eventuelt utløser behov for ytterligere kvalitetssikring. Metodene må være kunnskapsbaserte og standardiserte slik at resultatene lar seg gjenbruke nasjonalt, men som også kan gjennomføres innen rimelige tidshorisonter.

Mot denne bakgrunn foreslår derfor KIN at det opprettes en arbeidsgruppe med mandat til å utarbeide felles veileder for kvalitetssikring som en støtte til klinikere i helsetjenesten i tilknytning til anskaffelser av KI-løsninger. For å sikre nasjonal forankring, bør et slikt prosjekt utgå fra våre sentrale helsemyndigheter, for eksempel i tilknytning til koordineringsprosjektet som ledes av Helsedirektoratet i samarbeid med andre direktorater i helsetjenesten. Arbeidsgruppen bør være tverrfaglig sammensatt, primært med fagfolk fra våre ledende fagmiljøer innen feltet og gjerne med geografisk spredning. Både teknologimiljøene og de helsefaglige miljøene bør være representert i arbeidet, ideelt sett også med fagpersoner med hybridkompetanse (som kombinerer kunnskap om teknologi og helse på dette feltet). Arbeidsgruppen forutsetter å forankre sine vurderinger i relevant internasjonal forskningslitteratur og relevante retningslinjer/protokoller/standarder[93] fra andre land. Det pekes bl.a. på miljøer som NHS-X i England og Radboud-universitetet i Nederland. KIN ser det som ønskelig at sekretariatfunksjoner for nevnte arbeidsgruppe ivaretas av personer med tilkytning til de direktorater som samarbeider i koordineringsprosjektet for KI i helsetjenesten.
Det foregår allerede prosjekter som gjennomfører slike utprøvinger av KI-algoritmer innen helse, både i mammografiprosjektet ved Kreftregisteret og ved senteret MMIV i Bergen, og vi finner andre viktige og relevante kompetansemiljøer ved Senter for pasientnær kunstig intelligens (SPKI) i Tromsø, NTNU, Oslo universitetssykehus mv. KIN

[93] Sounderja, V, Ashrafian, H, Golub, RB et al.: Developing a reporting Guideline for artificial intelligence-centered diagnostic test accuracy studies:- STARD-AI protocol, BMJ Open 2021
Liu, X, Rivera, SC,: Reporting Guidelines for clinical trial reports interventions involving artificial, intelligence: the CONSORT-AI extension), Nature Medicine 26, 1364-1374 (2020)
The DECIDE-AI Steering Group: DECIDE : new reporting guidelines to bridge the delvelopement-to-implementation gap in clinical artificial intelligence, Nature Medicine 27, 186-187 (2021)

ser det som viktig at ressurspersoner fra slike kompetansemiljøer inkluderes som medlemmer i den foreslåtte arbeidsgruppen.
Fra sekretariatet i KIN spiller vi inn følgende ressurspersoner – hvorav de fleste har sagt seg villige til å deltas - om aktuelle og mulige kandidater for å delta I dette arbeidet:

- Førsteamanuensis Alexander Lundervold, MMIV-Bergen,
- Overlege Njål Lura, Helse Bergen
- Lege/PhD-kandidat Ishita Barua, Deloitte
- Radiolog/PhD-kandidat Tomas Sakinis, OUS (ikke forespurt)
- Forskningskoordinator Inge Groote, Sykehuset Vestfold/OUS
- Professor Solveig Hofvind, Kreftregisteret /UiT
- Leder Registerinformatikk Jan Nygård, Kreftregisteret/UiT
- Forsker Vibeke Binz Vallevik, DNV/forsker Courtney Nadeau, DNV
- Avdelingsleder Karl Øyvind Mikalsen, Senter for pasientnær kunstig intelligens
- Forsker Mattijs Elschot, NTNU
- Professor Frank Lindseth, NTNU
- Forsker Maryam Tayefi, NSE/ Forsker Phuong Dinh Ngo, NSE

Utfra forestående anskaffelser og implementeringsprosjekter innen KI-feltet, vurderer KIN det som viktig at arbeidet med den foreslåtte arbeidsgruppen kommer i gang snarest mulig. I den grad dere har spørsmål av utdypende karakter som dere ønsker avklar, må dere gjerne ta kontakt.

På vegne av sekretariatet i KIN
Finn Henry Hansen (leder)

Kunstig intelligens
i norsk helsetjeneste