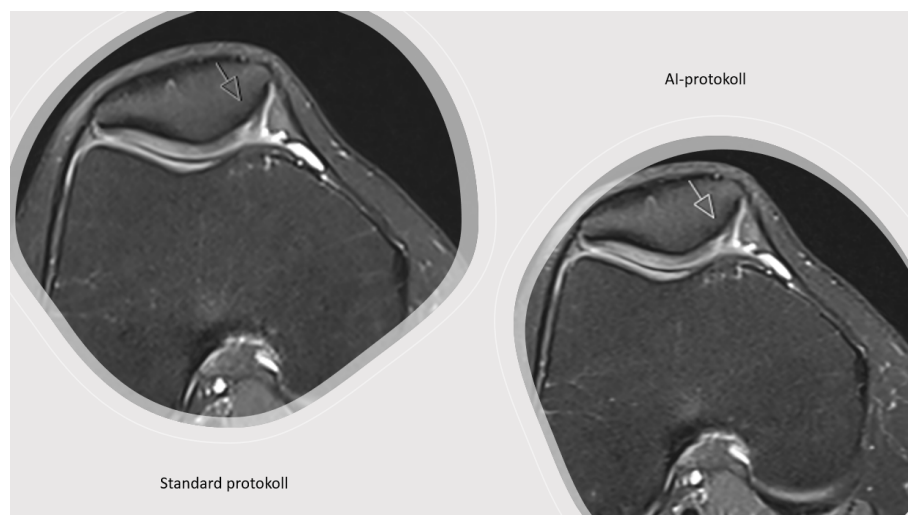


Anne Helene Reiten

# Bildeoptimalisering av MR kne, 2D TSE sekvenser ved bruk av AI basert bilderekonstruksjonsteknologi

Bildekvalitet, en subjektiv visuell vurdering av 4 radiologer

Masteroppgave i Master i medisinsk bildeteknologi (MMEDBT)  
Veileder: Mohammed Rasem Sadeq Sunoqrot  
Medveileder: Roar Remseth Pedersen og Rolf Svendsmark  
November 2022





Anne Helene Reiten

# **Bildeoptimalisering av MR kne, 2D TSE sekvenser ved bruk av AI basert bilderekonstruksjonsteknologi**

Bildekvalitet, en subjektiv visuell vurdering av 4  
radiologer

Masteroppgave i Master i medisinsk bildeteknologi (MMEDBT)  
Veileder: Mohammed Rasem Sadeq Sunoqrot  
Medveileder: Roar Remseth Pedersen og Rolf Svendsmark  
November 2022

Norges teknisk-naturvitenskapelige universitet  
Fakultet for medisin og helsevitenskap  
Institutt for sirkulasjon og bildediagnostikk



Kunnskap for en bedre verden



## Sammendrag

Kunstig Intelligens (AI) har siste årene blitt utviklet til bruk som akselerasjonsteknikk innenfor magnetisk resonanstomografi (MR). En programvare som har blitt utviklet til med hensikt for dette er Deep Resolve (Siemens Healthineers, Erlangen, Tyskland).

I studiet ble det undersøkt om det var mulig å bytte ut en AI-basert MR kneprotokoll, med å sammenligne protokollen med vår standard kneprotokoll. Hovedformålet ville være å redusere opptakstiden uten betydelig reduksjon i bildekvalitet. Opptakstiden ble redusert med 35 %, der radiologens vurdering av både standardprotokollen og den AI-baserte protokollen, var avgjørende for om AI-protokollen kan byttes ut. Resultatene tyder på at radiologene oppfattet standardprotokollen som bedre når det gjelder bildeklarhet, og kanskje var AI-protokollen bedre for mengde støy i bildene. Resultatene var imidlertid vanskelig å tolke, med en overraskende stor grad av variasjon av leserne. Selv med variasjon i vurderingen, ble forskjellen oppfattet for å være små, og den diagnostiske effekten av disse meningsforskjellene var ikke klare i resultatene. En annen studieutførelse og opplæring av radiologene, kunne ha gitt andre resultater og økt grad av enighet mellom radiologene. Av dette kan det reflekteres over, at det er stor variasjon i definisjon av bildekvalitet mellom radiologene.

## Summary

Artificial intelligence (AI) has the last year been developed for use as an acceleration technique within magnetic resonance imaging (MRI). A software that has been developed with this purpose in mind is Deep Resolve (Siemens Healthineers, Erlangen, Germany).

The study explored the possibility of applying an AI-based knee MRI protocol, comparing it to our standard knee protocol. The main purpose would be to reduce scan time without significant reduction in image quality. We reduced scan time by 35%, but the radiological evaluations of both standard and AI-based protocols were not conclusive. The results suggests that they perceived the standard protocol as better regarding image sharpness, and maybe the AI-protocol better regarding image-to-noise. However, the results were difficult to interpret, with a surprisingly large degree of variation. Even with variable responses, the differences were perceived to be small, and the diagnostic impact of these differences in opinion is not clear. Another study design and training of the radiologists could have generated other results and better degree of agreement between the radiologists. This probably reflects a large variation in the definition of image quality between radiologists.

## Forord

Det å få ta mastergrad ved siden av min stilling i Unilabs har vært en intensiv prosess, men en erfaring jeg ikke ville vært for uten. En stor takk til Unilabs som har gitt meg muligheten til dette, og støtte for å gjennomføre prosjektet for masterstudie. Uten ressurser fra firmaet, permisjon til studiedager, reiser til Trondheim, og tilrettelegging rundet det praktiske hadde ikke studiet vært mulig å gjennomføre. Jeg føler meg veldig heldig som fikk muligheten til å utføre et eget eksperiment i klinikken, på instituttet jeg jobber til daglig i min masteroppgave.

Denne oppgaven ble til i forbindelse med installasjon av en ny MR hos Unilabs Bryn sommeren 2021, MAGNETOM Altea 1,5 T. På denne maskinen ble det etter hvert tilgjengelig en AI-basert bilderekonstruksjonsteknologi, såkalte Deep Resolve. Jeg var så heldig å få teste denne programvaren fra Siemens, og den ble installert i februar/mars 2022.

Muligheten til å få utforske denne programvaren og å få skrive en masteroppgave om en AI-basert programvare var et privilegium. AI er i dagens søkelys innenfor MR utviklingen, og ser ut til å være et stort for å kunne innhente informasjon til MR bildene på redusert tid. Nysgjerrigheten i fagmiljøet rundt disse første erfaringer med teknologien, har vært til stor motivasjon i arbeidet med å utforske teknologien og oppgaven.

Fra Siemens så må jeg gi en takk til Terje Solheim, produktmanager i Siemens Healthineers, som jobbet for at vi fikk testet Deep Resolve i Unilabs. Fra Siemens må jeg også gi en stor takk til Rolf Svendsmark, produktspesialist innen MR i Siemens Healthineers, kontakt person fra Siemens til oppgaven, og veileder for masteroppgaven. Engasjert, alltid positiv og kunnskapsrik, å ha en slik veileder skal man ikke ta forgitt.

For hjelp til studieutforming, standardisering av protokoller som skulle sammenlignes i studiet, REK søknad, brev til pasienten og språkvask av masteroppgaven, må jeg gi en stor takk til Roar Remseth Pedersen. Fagansvarlig for muskelskjelettradiologi i Unilabs, og veileder i masteroppgaven.

En stor takk til radiologene som har vært med å vurdere bildekvaliteten til alle undersøkelsene, for protokollene som sammenlignes i studiet. Å få radiologer til å delta i dette studiet, var helt essensielt for at det lot seg gjennomføre med denne studieutformingen.

Jeg vil også benytte anledningen for å takke kollegaer og venner som har ligget i magnettomografen for å kunne utforske teknologien og optimalisere protokoller. Kollegaer som har oppmuntret meg i perioder med frustrasjon rundt studie og oppgaven. Det å gjennomføre oppgaven på så kort tid har vært krevende, fra august har vært noen slitsomme måneder, det skal jeg ikke legge skjul på. Uten deres forståelse og tålmodighet, i tider der tålmodigheten min ikke har vært så stor er prissatt. Alle innspill og støtte fra kollegaer som har stått meg nærmest i gjennomføringen av denne oppgaven, har vært til stor nytte og er verdsatt.

Fra NTNU må jeg få takke Mohammed R.S Sunqrot, for å ha vært hovedveileder til oppgaven, god hjelp til oppsettet til oppgaven og korrigerings underveis. Turid Follestad, rådgiver i statistikk/Førsteamanuensis ved NTNU, må jeg takke for gode innspill på valg av metoder til å analysere resultatene, og gjennomføringen av testene, og hjelp når det har vært uklarheter. Beathe Sitter, som er førsteamanuensis for institutt for sirkulasjon og bildediagnostikk, ved NTNU. Deg må jeg takke for tilrettelegging for å ta dette masterstudiet på komprimert tid, og veiledning for hva som kreves for å gjennomføre masterstudie.

En stor takk til alle sammen!

Oslo. 25 november 2022.

Anne Helene Reiten.



## Forkortelser og symbol.

MSK	Muskelskjelettplager
MR	Magnetisk resonanstomografi
PDFS	Fettsupprimerte proton vektet sekvens
TSE	Turbo spinn-ekko
SNR	Signal-to-noise ratio
PI	Parallel Imaging
IPAT	Integrated parallel acquisition technique
CS	Compressed Sensing
AI	Artificial intelligence
DL	Deep Learning
ANN	Artificial Neural Networks
CNN	Convolutional Neural Network
DR	Deep Resolve
FOV	Field of View
RF-puls	Radiofrequency Puls (eksitasjonspuls)
TR	Repetition time
TE	Time to-ekko
CNR	Contrast-to-noise ratio
RF-system	Radiofrequency- system
ETL	Echo train length
$\Delta$	Forandring
$\propto$	Tilnærmet lik
x	Multiplisert
ROI	Region of interest
NSA	Number of excitation
IES	Inter-ekko-spacing
IQM	Image quality metrics
ESSR	European Society of Skeletal Radiology Sports Sub-committee.
DICOM	Digital Imaging and Communications in Medicine

PACS	Picture archiving and communication system
IQR	Interquartile Range
CI	Confidence Interval
REK	Regionale komiteer for medisinsk og helsefaglig forskningsetikk
ACL	Anterior Cruciate Ligament (fremre korsbånd).

# Innhold

Sammendrag	i
Summary	ii
Forord	iii
Forkortelser og symbol.	v
Innhold	vii
1 Introduksjon	1
1.1 Bakgrunn	4
2 Teorigrunnlag, begrepsavklaring	5
2.1 MR	5
2.1.1 Bildedannelse, innhenting av informasjon til MR-bildet.	5
2.1.2 Bildekvalitet og optimalisering	7
2.1.3 Ulike aspekter ved bildekvalitet innenfor MR	8
2.2 Kunstig intelligens (AI, fra engelsk Artificial Intelligence)	12
2.2.1 Tilpasse dyplæringsmodell	14
2.2.2 Deep Resolve	15
2.3 Standardisering	17
2.4 Subjektiv vurdering av bildekvaliteten	17
3 Hensikt	18
4 Metode	19
4.1 Valg av forskningsdesign	19
4.2 Optimalisering og standardisering	19
4.3 Utvalg og utvalgskriterier	21
4.4 Testobjekter og variabler	22
4.5 Datainnsamling, gjennomføring av prosjektet	23
4.6 Eksperiment	24
4.7 Analyse	25
4.7.1 Valg av metoder for analysering av data	25
4.7.2 Kategorisering av datamaterialet	27
4.8 Etikk	28
5 Resultater	29
5.1 Resultater fra hver enkelt radiolog	30

5.2 Resultater fra alle radiologene innen hver enkelt kategori	37
5.3 Oppsummerende resultater fra alle radiologene innen alle kategoriene	38
5.4 Enighet mellom leserne, Inter-leser enighet	39
6 Diskusjon	41
6.1 Diskusjon omkring kvalitetspoeng	41
6.1.1 Støy	44
6.1.2 Skarphet	47
6.1.3 Anatomiske strukturer	49
6.2 I hvor stor grad er det grunnlag for å si om det er kvalitetsforskjeller?	51
6.3 Kappa, grad av inter-leser enighet	52
6.4 Standardisering og optimalisering av protokollene	55
6.5 En oppsummert evaluering av Deep Resolve	56
6.5.1 Fordeler og ulemper med bruk av AI protokollen sett i klinisk bruk	57
6.5.2 Begrensninger	57
7 Konklusjon	59
Referanseliste	60
Vedlegg.	63

# 1 Introduksjon

Ved utredning av muskelskjelettplager (MSK) er magnetisk resonanstomografi (MR) en veletablert og god undersøkelsesmodalitet. Knesmerter er en av de muskelskjelettplager man ser hyppigst der det henvises til MR for utredning (1). MR er i dag den foretrukne bildediagnostiske modaliteten for utredning av skjelett og bløtdelsskader. Modaliteten gir størst nøyaktighet innenfor bildediagnostikken, og har blitt en primær undersøkelse for utredningen av kneleddsmarter. Vevskontrasten som er mulig å oppnå med MR er unik for utredning av klinikk knyttet til kneleddet, samt diagnostisere skader i anatomiske strukturer i kneleddet som menisk, korsbånd, ligament, benmarg, leddbrusk og dessuten andre patologiske prosesser som kan oppstå i kneleddet. Det er ingen annen bildediagnostisk metode som kan sammenlignes MR når det gjelder å gjengi små strukturer med god vevskontrast (2).

Studier viser at utredning med MR kan ha stor betydning for videre valg av behandling for pasienten (2-3). MR-protokollene settes sammen av ulike sekvenser, basert på hva man ønsker å fremstille. Referansestandard for en MR kneprotokoll inneholder fettsupprimerte proton vektete (PDFS) - turbo spinnekk (TSE) i 3-plan, og T1-vektet turbo spinnekk (TSE) i et plan (4).

Bildeopptak med MR bygger på avanserte fysiske prinsipper, bildekvaliteten er følsom for parameterendringer, og bildekvaliteten er sårbar for indre og ytre påvirkninger (5).

Prinsippene bak innhenting av informasjon til MR-bilder gjør at teknikken er tidkrevende. For å oppnå god bildekvalitet er det nødvendig med tilstrekkelig informasjon i bildet, og godt nok signal-støy forhold (SNR, fra engelsk signal-to-noise ratio). Jo flere antall avlesninger i frekvens og faseretning for å fylle K-space (fra engelsk K-rommet), jo flere datapunkter er nødvendig for romlig koding av MR-signalet. Opptakstiden øker proporsjonalt med antall avlesninger i faseretning. Ulike parameterendringer for å redusere opptakstiden vil gi økt støy i bildet, eller gi lavere oppløsning som igjen gir redusert skarphet (6).

Parallell avbildning (PI, fra engelsk Parallel Imaging) er en av de første akselerasjonsteknikkene innen MR, og var et gjennombrudd for å innhente informasjon til MR-bildet med redusert opptakstid (7). Teknikken reduserer opptakstiden ved å redusere antall fasekodinger i K-rommet (fra engelsk K-space), såkalt undersampling. PI bruker spolens følsomhet, sensitivitetsprofilen til to eller flere spoleelementer langs faseretningen, til å trekke ut informasjon om undersamplete data. Teknikken krever to eller flere spoleelementer langs

faseretningen, der spoleelementene må ha individuelle utganger som behandler signaler parallelt. Informasjon om spolens sensitivitetsprofil blir brukt til å erstatte undersamplete fasekodingslinjer i K-space. Undersamplingen skjer ved at antall linjer i K-Space hoppes over i fyllingen av K-space, der antall linjer som hoppes over, øker med økt iPAT-faktor (fra engelsk Integrated parallel acquisition technique-factor). Med en iPAT-faktor 2 halveres opptakstiden (8).

Det finnes to typer PI, K-space basert og bildebasert. Ved K-space basert skjer digitaliseringen og rekonstruksjonen av manglende signaldata i «frekvensdomenet», før informasjon i MR-signalet blir transformert til et bilde. I den bildebaserte typen skjer rekonstruksjon av undersamplingen i «biledomenet», etter transformasjon av bildedata fra frekvensdomenet til tidsdomenet. Artefakter og støy i bildene setter begrensninger for hvor mye man kan øke akselerasjon ved bruk av PI (8).

Den neste akselerasjonsteknikken som kom på markedet var komprimert sensing (CS, fra engelsk Compressed Sensing). Med CS reduseres opptakstiden med en tilfeldig undersampling av k-space data. Manglende informasjon i bildet blir gjenopprettet i frekvensdomenet, ved bruk av matematiske modeller (9). De siste årene har bilderekonstruksjonsteknologi basert på kunstig intelligens (AI, fra engelsk Artificial intelligence) og dyplæring (DL, av engelsk Deep Learning) fått stor oppmerksomhet innen medisinsk bildebehandling, også som akselerasjonsteknikk for MR-opptak. Noen få studier og erfaringer så langt viser at dette kan være et stort steg videre for utviklingen innenfor MR (10).

Et av målene med å bruke DL og kunstige nevralt nettverk (ANN, fra engelsk Artificial Neural Networks) som bilderekonstruksjonsteknologi innenfor MR, er å konvertere undersamplete data til bilder med fullverdig informasjon, dvs. en slags gjenoppretting av undersamplingen. Undersamplingen gir manglende informasjon i K-space. For å fylle hele k-space blir det brukt DL og nevralt nettverk til å rekonstruere manglende informasjon i K-space. På den måten sparer man tid i bildeopptakene. Det finnes også dyplæringsalgoritmer som blir brukt til å fjerne uønsket støy i MR bildet, der DL har vist seg å være bedre til å gjenopprette lokale støyvariasjoner i MR-bildet, enn et konvensjonelt bildefilter (11).

AI har hatt en rask utvikling siste årene, der teknologien har blitt mer kompleks fra maskinlæring til dyplæring og bruk av nevralt nettverk. Dyplæringsalgoritmene er utviklet gjennom treningsdata, sammen med kraftig datakraft og signaloverføringer i nevralt nettverk. Dette gir teknologien potensiale til å gjenkjenne målrettede spesifikke oppgaver, kunstig

intelligens (12). Konvulsjon nevralt nettverk (CNN, fra engelsk Convolutional Neural Network) er en utvikling av ANN, som håndterer mer komplekse funksjoner. Denne typen nettverksarkitektur har gitt potensiale for å utvikle DL som bilderekonstruksjonsteknologi innen bildebehandling med gode resultater (13).

AI-baserte programvare som bilderekonstruksjonsteknologi, har nå blitt introdusert fra flere MR-leverandører. Resultatene som blir vist hos leverandørene er oppsiktsvekkende gode (14) (15). Andre studier som bruker AI som bilderekonstruksjonsteknologi i vurdering av kneleddet er referansene (11;16).

Frem til i dag har bilderekonstruksjonsteknologi basert på AI vært lite sett i klinisk praktisk bruk i Norge. Kunnskap om bilderekonstruksjonsteknologier ved bruk av dyplæring er til nå i stor grad basert på erfaring av leverandøren selv. Det er få vitenskapelige studier fra klinisk bruk. I dette studiet sees det på programvaren til Siemens, Deep Resolve (DR). Etter informasjon fra leverandøren og såvidt undertegnede vet, er dette det første vitenskapelige studien som er utført i klinikk med pasienter i Norge, neppe mange flere på verdensbasis. I litteraturen finner man imidlertid studier der det sees på DL-algoritmen. Allerede i 2018 publiserte Chaudhari med flere (17) lovende resultater, og etter den tid har teknologien utviklet seg med nye DL modeller og nevralt nettverk.

Et annet studie med bruk av bilderekonstruksjonsteknologien DL er studiet til Herrmann med flere (18), det oppgis det at studiet er utført på magnettomografer levert av Siemens. Dette studiet sammenlignet bilderekonstruksjonsteknologi ved bruk av DL på 1,5 Tesla og 3 Tesla magnettomografer. Teknologiens ytelse blir vurdert til å ha potensiale, med de beste resultatene på 1,5.Tesla. Det oppgis ikke hvilken teknologi som ble brukt, men det antas at det er Deep Resolve. I begge studiene var studiet rettet mot kneleddet, og utført på friske frivillige, og ikke til klinisk bruk.

I følge leverandøren skal programvaren Deep Resolve gi potensiale for bildeopptak med lav oppløselig bildeinformasjon, men med resulterende høy oppløselig bildeinformasjon ut, med undersampling der bildekvaliteten er beholdt (19). For å kunne vurdere ytelsen til DR, er det behov for mer erfaring og studier fra klinisk bruk.

## 1.1 Bakgrunn

En av ulempene med magnetisk resonanstomografi er at modaliteten er tidkrevende. De fysiske prinsipper MR bygger på setter begrensninger i det å redusere opptakstiden til sekvensene, hvis bildekvaliteten og den diagnostiske nøyaktigheten skal bevares (6;11).

På grunn av disse lange undersøkelsestider er derfor MR også en ressurskrevende undersøkelsesmodalitet, noe som bidrar til lange ventetider for å få time til undersøkelse. Lange ventetider kan føre til at alvorlig sykdom ikke oppdages i tide, eller at ventetiden forsinker behandlingen (20). Plager relatert til kneleddet er en hyppig plage, også hos barn og unge. Lange ventetider forsinker utredningen, som igjen påvirker pasientforløpet og behandlingen. I verste fall kan pasienten få kroniske plager som kunne vært unngått med en tidligere diagnostikk og behandling (21-22).

I Unilabs, der dette studiet ble gjennomført, kan raskere bildeopptak og reduserte undersøkelsestider potensielt gi økt effektivitet, potensiale for økt produksjon, reduserte ventetider og økte inntekter. Erfaring viser at kortere undersøkelsestid øker pasientkomfort, spesielt for klaustrofobiske eller smertepåvirkede pasienter. I tillegg vil det kunne redusere bevegelsesartefakter. Det foreligger studier som viser at redusert opptakstid gir mindre bevegelsesartefakter (23).

Dersom resultatene viser at bildeopptakstiden ved MR kan reduseres ved bruk av AI-basert bilderekonstruksjonsteknologi, kan flere tidsmessige utfordringer med MR som undersøkelsesmodalitet i bildediagnostikken reduseres.



## 2 Teorigrunnlag, begrepsavklaring

### 2.1 MR

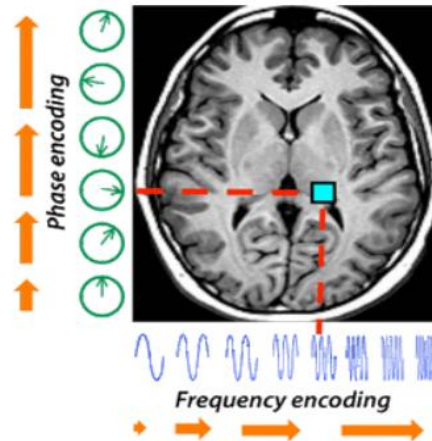
Magnetisk resonanstomografi, MR, er en undersøkelsesteknikk som ble først tatt i klinisk bruk i Norge i 1986. Kroppens biologiske vev og hydrogenprotonets magnetiske egenskaper utnyttes for å oppnå unik bildekontrast for ulike typer vev og patologiske prosesser i kroppen (24).

#### 2.1.1 Bildedannelse, innhenting av informasjon til MR-bildet.

For å få god bilde kvalitet er det nødvendig med tilstrekkelig informasjon i MR-bildet. Jo høyere romlig oppløsning man ønsker i bildet, jo mer informasjon er nødvendig for en gitt feltstørrelse, (FOV, fra engelsk Field of View). Bildeinformasjon innhentes ved å lage frekvens- og fasekodinger, ved bruk av feltgradienter (25).

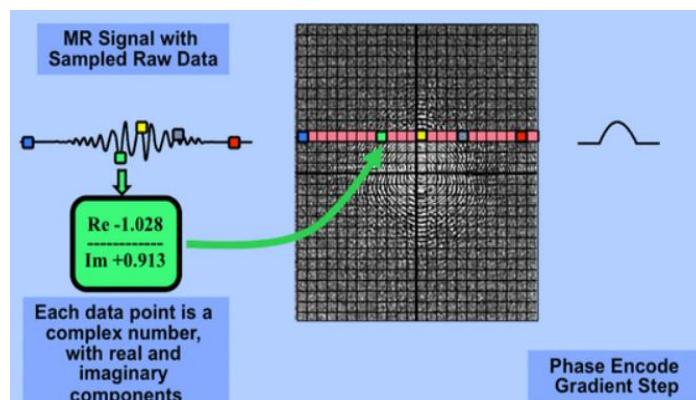
Bildesignalet inneholder informasjon som er nødvendig for plassering av signalet i MR-bildet, såkalt romlig koding. Hvert MR-signal har sin unike adresse i tre dimensjoner. Dette oppnås gjennom tre trinn i bildeopptaket; snittseleksjon, frekvenskoding og fasekoding, der alle tre trinnene i bildedannelsen gjøres ved bruk av magnetfeltgradientene (6).

Snittseleksjon bestemmer hvor i objektet snittet skal være og tykkelsen på snittet. Snittseleksjonsgradienten står på under eksitasjon, i det RF- pulsen (eksitasjonspuls) skrues på. For romlig koding må hydrogenprotonet ha en koding i to retninger innenfor det enkelte snitt, se figur 2.1. I en TSE-sekvens oppnås dette ved å sette på en fasekodingsgradient etter snittseleksjonsgradienten. Dette vil gi hydrogenprotonene en unik faseinformasjon. Fasekodingen koster tid, og opptakstiden øker med antall fasekodinger, som bestemmes av størrelsen på matrisen i faseretning. For 256 x 256 matrise, er det nødvendig med 256 fasekodinger. For koding i det andre planet, settes det på en lineært økende frekvenskodingsgradient, hydrogenprotonene vil oppleve en forskjell i magnetfeltstyrke langs frekvenskodingsgradienten. Hydrogenprotonene vil da få en frekvens som funksjon av posisjon langs feltgradienten. Frekvenskodingsgradienten står på under utlesningen av ekkoet, dette går så fort at det vil ikke påvirke opptakstiden i MR opptaket. Denne prosessen må gjentas for hvert snitt for å gi hvert proton sin unike frekvens og faseinformasjon, for å lage romlig koding av MR-signalet (26-27).



**Figur 2.1:** Bildet viser romlig koding i to retninger, for en gitt vokal i et MR bilde. Frekvens og fasekoding er oppnådd ved bruk av feltgradienter. Feltgradientene endrer magnetfeltstyrken langs feltgradientens retning innenfor 100 til 1000 Hz. Unik frekvens og fase til hydrogenprotonene gir informasjon om plassering av hydrogenprotonet i MR bildet. Innenfor hver vokal vil det være et gitt spekter av hydrogenproton med sin frekvens og fase, såkalt båndbredde per piksel (28). Figur hentet fra (28).

Frekvens - og faseinformasjon blir brukt til å bestemme plassering av MR-signalet i K-rommet, (fra engelsk K-space). Under Fourier-transformasjon går MR - signalet fra tidsdomene til frekvensdomene. I K-space lagres data systematisk etter frekvens og faseinformasjon, i hver sin dimensjon langs x og y aksene, se figur 2.2. Posisjon i MR-bildet og K-space er ikke visuelt sammenlignbare. Romlig oppløsning i bildet med en bildematrix på 256 x 256 vil kreve 256 fasekodingstrinn og 256 linjer i frekvensretning i K-space. Ut fra grunnleggende prinsipper for MR avbildning, vil økt antall avlesninger i frekvens og faseretning gi økt repetisjonstid mellom eksitasjon-pulsene, (TR, fra engelsk Repetition Time). En økning i TR vil øke opptakstiden. Endring i frekvens og faseavlesninger vil også endre tiden til ekkoene (TE, fra engelsk time to-ekko) (6).



**Figur 2.2:** Bildet illustrerer K-space som et rutenett med informasjon om plassering av MR signalet, for høyere romlig oppløsning må flere punkter i K-space fylles. Det er utviklet mange teknikker for å fylle K-space. Her fremstilles en kartetisk metode, linje for linje. Figuren illustrerer en linje langs fasekodingsgradienten, y-aksen. x-aksen inneholder ulike frekvensinformasjon, hvert punkt i K-space vil da bli et kompleks punkt med sin unike frekvens og faseinformasjon. I utregningen for hvert datapunkt, vil vi få kompleks tall med reale og imaginære komponenter. Amplituden på signalet, bestemmer hvor langs fasekodingsgradienten signalet skal plasseres, høye amplituder i senter, jo lavere amplitude jo lengre i ytterkant vil signalet plasseres (29).

Tilstrekkelig romlig oppløsning for å kunne gjengi små strukturer med et godt forhold mellom signal og støy kan være tidkrevende. Undersampling, en reduksjon av antall frekvens og fasekodinger, vil redusere opptakstiden, men da på bekostning av den romlige oppløsningen (6).

### 2.1.2 Bildekvalitet og optimalisering

God bildekvalitet oppnås gjennom optimalisering av bildeopptakene (sekvensene) og sammensetningen av disse, en protokoll. God bildekvalitet er forenlig med et bilde som har et godt signal-støyforhold, SNR og kontrast til støy (CNR, fra engelsk contrast-to-noise ratio), med en høy romlig oppløsning. For å oppnå dette, må man i bildeoptimalisering gjøre mange kompromisser for at ikke opptakstiden til sekvensen skal bli for lang (6).

God bildekvalitet er viktig for at all patologi skal bli fanget opp i bildet, og vil påvirke den diagnostiske verdien i bildet. Gode bilder forenkler jobben til radiologen, og vil være avgjørende for en god og sikker bildediagnostikk. Radiografens jobb for å oppnå god bildekvalitet er å gjøre parametervalgene i sekvensene til ønsket vektning, kontrast, skarphet, SNR og opptakstid. Litteraturen dokumenterer godt, og er relativt samstemt i parametervalg for ønsket vektning på MR-bildet (6;25).

Ulike prosedyrer, klinisk bruk og MR-system, gjør at protokollene må optimaliseres og tilpasses klinikken og det enkelte MR-system. For å oppnå god bildekvalitet er ikke alle parameterne i

en sekvens nødvendigvis overførbare og generaliserbare for god bildekvalitet til alle andre MR-system. Bildekvaliteten vil påvirkes av spesifikasjonene til MR-systemet. Fysiske prestanda som gradientytelse og linearitet, feltinhomogenitet, RF-system, spoler og ytelse på MR-systemet vil påvirke bildekvaliteten (30). Dette vil imidlertid ikke bli diskutert i denne oppgaven. Alle undersøkelsene er utført under samme spesifikasjoner og på samme MR-system, av samme radiograf og det sees kun på enkelte parametervalg i sekvensene som påvirker bildekvaliteten.

Bildekvaliteten påvirkes i stor eller liten grad av ulike parametre i sekvensen, og dette parametersamspillet er komplekst. Ulike MR leverandører, spoler, modeller, maskinvarefunksjoner, programvare og terminologi gjør at alle parametervalg er kun overførbart til et helt identisk MR-system for å oppnå god bildekvalitet (30).

Bildeoptimalisering ved bruk av AI som bilderekonstruksjonsverktøy, har i tidligere studier vist at det er viktig med en godt optimalisert protokoll som utgangspunkt. God bildekvalitet inn i dyplæringsmodellen, er viktig for å få gode bilder ut. Dyplæringsrekonstruksjon er utviklet på grunnlag av store mengder treningsdata. For at nevralt nettverk skal kunne gjenkjenne spesifikke oppgaver i bilderekonstruksjon, er det viktig at rådata er av god bildekvalitet. Dette bidrar til å unngå artefakter i bildet etter bilderekonstruksjon (31).

### 2.1.3 Ulike aspekter ved bildekvalitet innenfor MR

Det finnes en rekke parameter som kan endres i sekvensene. Grad av skarphet og støy i bildet er et samspill mellom parameter. Parametere som TE, TR, flipp-vinkel og ETL (ETL, fra engelsk echo train length) påvirker kontrasten i bildet og bildeopptakstiden. Mottakerbåndbredden, inter-ekko-spacing (IES) og oppløsningen påvirker skarpheten og mengde støy i bildet (6). Denne oppgaven har et for begrenset omfang til å kunne gå i detalj på alle parametre som påvirker bildekvaliteten i sekvensene. Det forutsettes at leseren har grunnleggende kunnskap om MR.

Kontrasten, CNR i bildene blir ikke vurdert opp imot bildekvalitet, da parametere som påvirker CNR holdes konstant mellom protokollene som sammenlignes. De relevante parametere; TE, TR og flippvinkel, er identisk i begge protokollene.

### Signal- støy forhold (Støy)

Støy er en faktor som har stor betydning for bildekvaliteten. En viss mengde bakgrunnsstøy fra elektroniske komponenter vil alltid være tilstede. Støy påvirker bildekvaliteten i negativ retning og er uønsket. For best mulig fremstilling av anatomiske strukturer, god kontrast og detaljoppløsning, bør bidraget fra støy være lavest mulig i MR-bildet (6).

Signal/støy, SNR, angir forholdet mellom gjennomsnittlig signalintensitet og støynivå. Det er ulike grunnleggende prinsipper som påvirker SNR i MR-bildet gjennom bildeoptimalisering for et gitt objekt. Signalet er tilnærmet proporsjonal med vokselstørrelsen som vist i ligning 2.1. En økning i frekvens og fasekodinger vil ikke bare endre den romlige oppløsningen, men også SNR i bildet. Dobler man antall fasekodinger vil man halvere antall protoner i en voksel, som vil gi en halvering av signal, dersom FOV holdes konstant. Med endringen vil støy reduseres med kvadratrotten av to, og nettoeffekten av SNR bli redusert med tilnærmet 30 % (6).

$$\text{Signal} \propto \Delta x \times \Delta y \times \Delta z \times F\text{sekvens}$$

**Ligning 2.1:** Ligningen viser en matematisk utregning for et tilnærmet forhold mellom signalet og størrelsesendring i vokselvolum, som er gitt av en endring i frekvens, fase eller snitt tykkelse. Der  $\Delta x$  og  $\Delta y$  er en endring i planet, og  $\Delta z$  er en endring gjennom planet, snitt tykkelse. Fsekvens er en sekvensavhengig signalutregningsfaktor, korreksjonsfaktor for den enkelte sekvens. (6). Ligning hentet fra (6 s.71).

Signal-støy forholdet er proporsjonalt med vokselvolum, kvadratrotten av antall målinger (NSA, fra engelsk number of signal averages) av bildeopptaket, og kvadratrotten av antall fasekodinger. Å øke antall målinger vil øke opptakstiden betydelig, gitt av ligning 2.2 (6). Se figur 2.3.

$$\text{Opptakstid} = \text{NSA} \times \text{TR} \times \text{NPE}$$

**Ligning 2.2:** Viser matematisk utregning av opptakstiden, der NSA er antall målinger av opptaket, TR er repetisjonstiden og NPE er antall fasekodinger. Formelen viser at opptakstiden øker proporsjonalt når en av disse parameter økes (6). Ligning hentet fra (6 s.71).

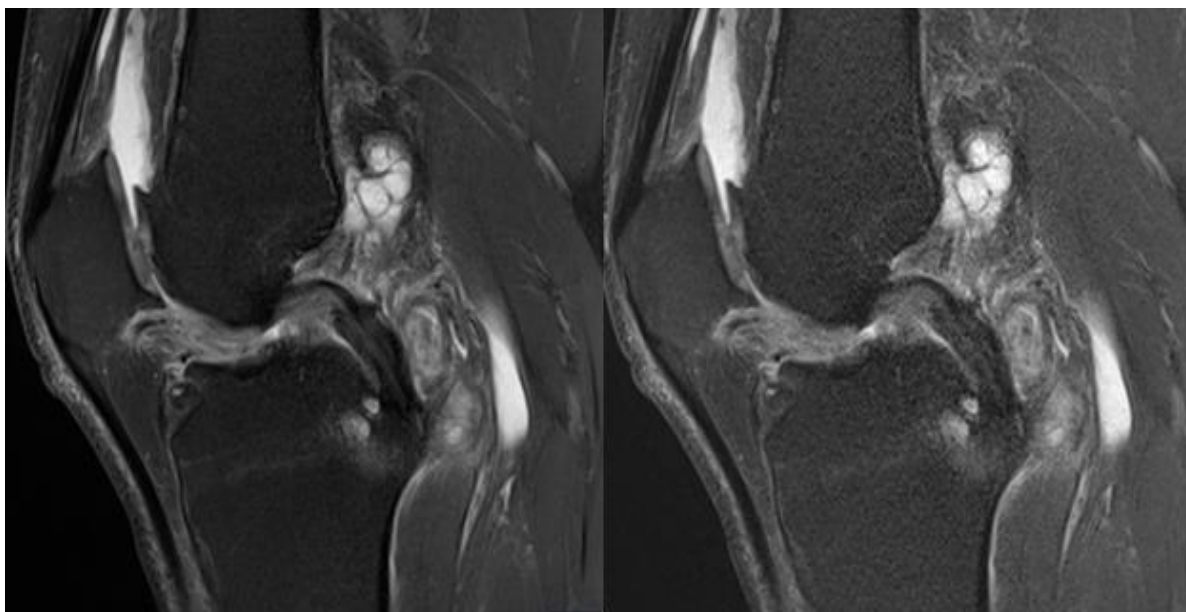
Støy blir ansett som en forstyrrelse i bildet, som kommer av tilfeldige spredte frekvenser i tid og rom. Generelt signal-støy forhold er definert av ligningen 2.3. Denne ligningen tar ikke hensyn til skjevfordeling av støy mellom objekt og utenfor objektet. Skjevfordelingen oppstår når MR-signalet går fra tidsdomene til frekvensdomene under bildedannelsen. Det vil da oppstå fouriertransformerte verdier som ikke kan bli negative, og det vil bli et kunstig lavt signal utenfor objektet, Rayleigh-fordeling. I objektet vil signalet være Gaussisk fordelt. For en mer korrekt signalintensitet i bildet, må dette korrigeres for å få et mer riktig SNR i bildet (6;32).

$$SNR = \frac{Signal}{Støy}$$

**Likning 2.3:** Likningen viser den matematiske formelen for generell signal-støy forhold i MR bildet, der SNR er gitt av forholdet mellom signal og støy i bildet (6). Ligning hentet fra (6 s.66).

Formelen i ligning 3.3 viser en generell utregning for SNR, men fordelingen av støy i bildet påvirkes av flere faktorer som spolegeometri, akselerasjonsteknikk og pasient, noe det ikke går nærmere inn på i oppgaven (32).

SNR avtar med avstand fra spolen i objektet. For å måle SNR i MR bildet, kan støy regnes ut ganske nøyaktig for et gitt område i MR bildet, ved bruk av ROI (fra engelsk, region of interest) målinger i bildet. Måling av SNR er komplisert, men kan forenkles for "enkel bruk", ved å sette to ROI- målinger i bildet, en satt i sentralt i vevet, «signal» og en i luft, «støy». Et gjennomsnittlig signal-støy-forhold i bildet beregnes ved at forholdet mellom disse målene (S/N) ganges med 0,66. Sistevente tall er Rayleigh distribution correction factor, og gir et mer nøyaktig mål på forholdet mellom signal og støy. Dette vil ikke bli gjennomgått i oppgaven, da det ikke ble utført i dette studiet (6).



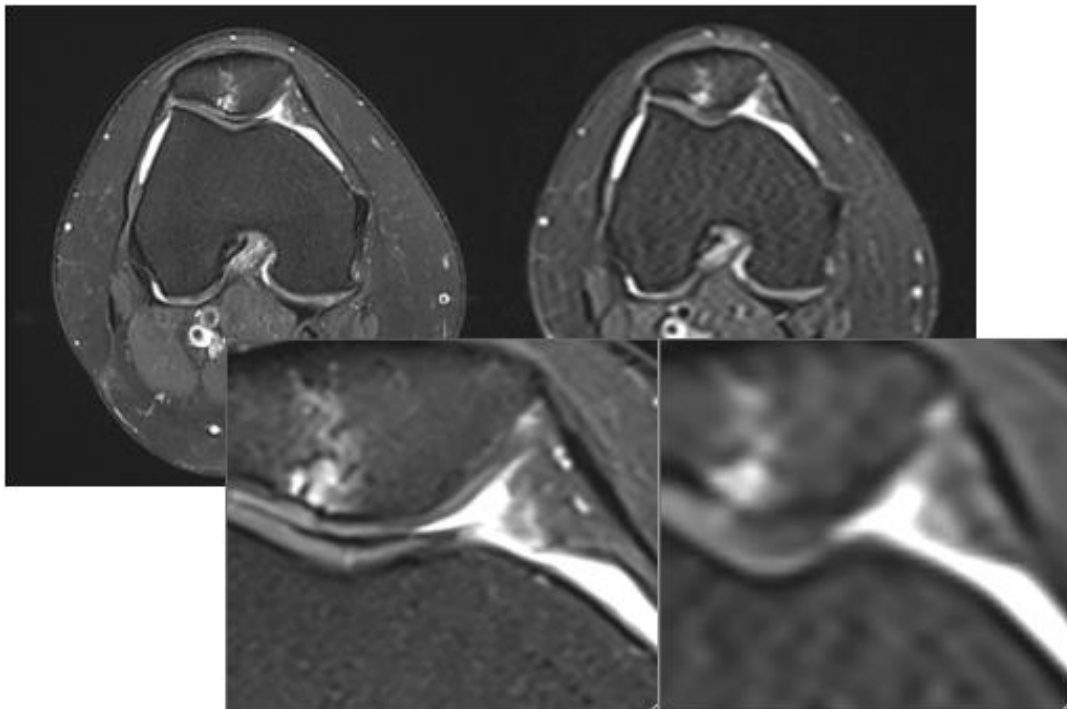
**Figur 2.3:** Viser bilde av to MR-opptak, det er gjort en endring mellom opptakene, antall NSA. Bilde til venstre NSA= 4. Bilde til høyre NSA = 1. Signal-støy forholdet er redusert med 50% som gjør at små strukturer blir vanskeligere å diagnostisere, og patologi kan bli oversett. (33).

Andre parameterendringer som vil redusere SNR er å øke mottakerbåndbredden, redusere faseoversampling, redusere antall målinger av bildeopptaket, og økt akselerasjonsfaktor. En endring

av fase - oversampling vil redusere eller øke opptakstiden på lik linje med antall fasekoder i bildeopptaket (32).

### Skarphet

Skarpheten i bildet påvirkes i stor grad av den romlige oppløsningen i bildet, synonymt med detaljoppløsningen. Den romlige oppløsningen angir hvor små detaljer MR-skanneren kan gjengi i et bilde, og er ifølge (6 s.69) definert som den minste synlige avstand mellom to punkter i et objekt. Dårlig romlig oppløsning gir et «blurry»-bilde og strukturene flyter sammen, mindre skarphet, som i figur 2.4. Jo lavere romlig oppløsning, jo dårligere evne vil bildet ha til å gjengi små detaljer, og anatomiske strukturer kan bli vanskeligere å vurdere diagnostisk (25). Bildeegenskapene for visualisering av detaljer og strukturer avhenger av tre faktorer. For det første må være nok kontrast mellom strukturene og omliggende vev. Den andre faktoren er oppløsning; er oppløsningen for lav, vil ikke informasjon i objektet bli gjengitt i bildet nøyaktig, og det kan oppstå feiltolkning under bildedannelsen. Den tredje er SNR og CNR, er disse to for lave, kan detaljer og strukturer bli skjult av bildestøy (6).



**Figur 2.4:** Bildene til venstre viser et bilde med høy oppløsning og fremstiller små strukturer, bildet under er et forstørret utklipp. Bildematrise 288 x 288. I bildet til høyre er bildematrisen redusert til 128 x 128 med samme FOV. Hvert bildepunkt er dermed større, og bildet har lav oppløsning. Det er dermed vanskelig å diagnostisere små strukturer med patologi i patella og leddbrusk.

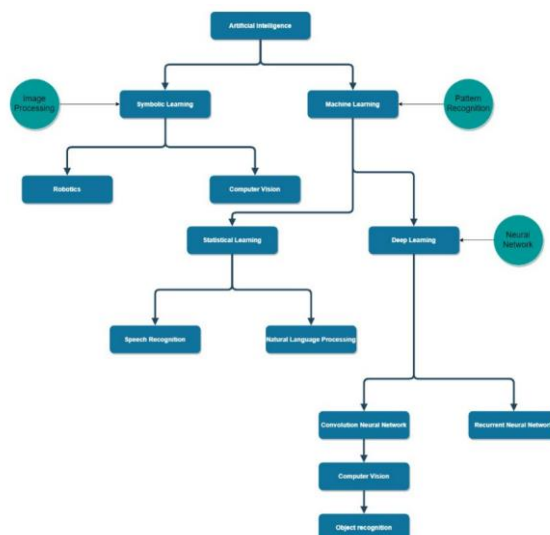
(33).

Romlig oppløsning er gitt av pikselstørrelse og vokselstørrelse. Pikselstørrelsen, hvert bildepunkt i planet, defineres av feltstørrelsen, FOV og bildematriksen i to retninger. Voksel størrelsen, gjennom planet, defineres av piksel størrelsen og tykkelsen på snittet, tre retninger. Størrelsen på bildematriksen vil bestemmes av antall frekvens og fasekoder i bildeopptaket, ved innhenting av informasjon til MR- bildet (6).

Skarpheten i bildet påvirkes også av IES, jo kortere IES, jo skarpere bilde. Kortere IES oppnås med å øke mottakerbåndbredden. Dette parametervalget vil imidlertid øke støy i bildet. I sekvensoptimalisering er det flere slike kompromisser som må gjøres, dette vil ikke diskuteres nærmere i oppgaven. Skarpheten vil også påvirkes av gradientlineariteten, men dette er også utenfor oppgavens omfang (6).

## 2.2 Kunstig intelligens (AI, fra engelsk Artificial Intelligence)

Det finnes mange ulike definisjoner på kunstig intelligens. Begrepet beskriver en datamaskin som modellerer intelligent atferd med minimal menneskelig innblanding. Utviklingen innen bruk av kunstig intelligens har gått raskt, det utvikles stadig nye «modeller» (34). Dyplæring er en undergruppe som er utviklet fra maskinlæring, og er en av mange undergrupper til AI, se figur 2.5 (35). De siste årene har DL og bruk av Convolutional Neural Network, CNN blitt brukt som bilderekonstruksjonsteknologi. Blant annet til å utvikle programvaren Deep Resolve (36).

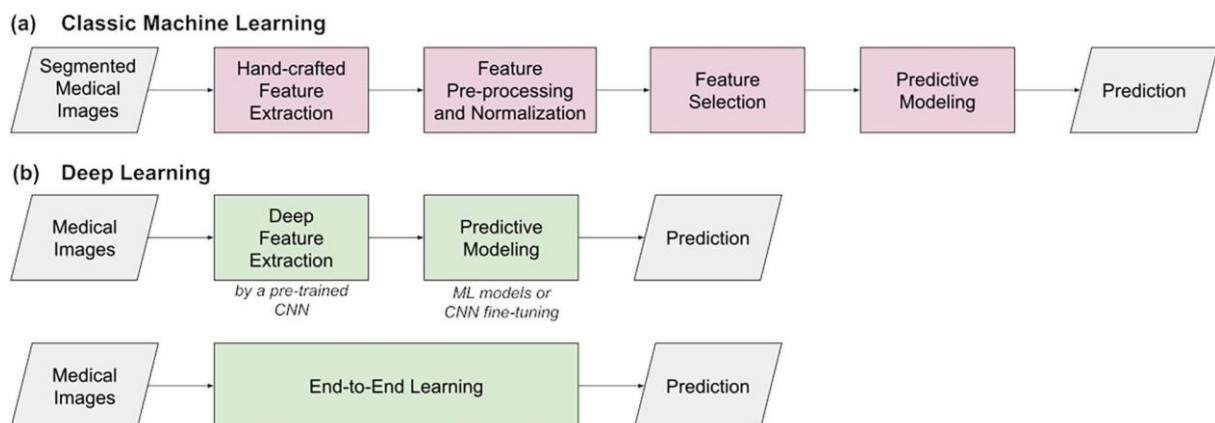


**Figur 2.5:** Illustrerer et hierarki innenfor AI gjennom et flytskjema. Der AI viser å ha utviklet seg til ulike grener med ulik funksjonalitet, basert på AI. Skjema viser at maskinlæring er en egen undergruppe til AI. DL og CNN som er brukt til å utvikle Deep Resolve, er en av mange undergrupper til maskinlæring (35). Figur hentet fra (35).



Artificial Intelligence er utviklet med hjernen til et menneske som utgangspunkt, teknologien baserer seg på menneskets intelligens. Den menneskelige hjernen er et nettverk av nevroner der signaler bli sendt i form av synapser gjennom aksoner, dette er måten mennesket lærer nye ting på. Det er på den samme måten nevralt nettverk er utviklet for å oppnå intelligenskapasitet. Når disse nettverkene blir satt sammen i flere skjulte lag, er det mulig å trene nettverkene til blant annet objekt gjenkjenning i flere dimensjoner, på bakgrunn av tidligere erfaring, det vil si store mengder treningsdata. Nettverket har da opparbeidet seg kunnskap gjennom en læringsprosess (35).

Generelt er AI basert på analytiske modeller som er utviklet for å utføre målrettede oppgaver, frembringe spådommer, regler, svar, anbefalinger eller lignende utfall. Disse er definert som systemer som utfører handlinger digitalt, basert på tolkning av treningsdata. Treningsdata behandles som strukturert eller ustrukturert data, i den hensikt å oppnå et gitt mål (34;37). Bruk av AI i bildebehandling er utviklet fra maskinlæring, der det har blitt tatt i bruk dyplæring for å spesifisere oppgaver som objekt gjenkjenning i flere dimensjoner og mønstergjenkjenning, se figur 2.6. Store mengder treningsdata blir brukt for å utvikle kunstig intelligens for dyplæringsalgoritmer med ulik arkitektur (38). Bruk av dyplæringsalgoritmer sammen med nevralt nettverk som er stimulert av datakraft, har vært kodeknekkere innenfor AI for bruk til bilderekonstruksjonsteknologi (39).



**Figur 2.6:** Illustrerer forskjellen mellom maskinlæring og dyplæring i kompleksitet. Under (a) og (b) i bildet vises to ulike arkitekturer ved bruk av AI innen medisinsk bildebehandling. (a) er en maskinlæringsmodell, med mange prosesseringstrinn, der det brukes håndlagde funksjoner, «randomik». (b) viser arkitekturen, til en DL modell, som gir muligheten til å gjøre automatiske trekk, og behandle rådata direkte, ved bruk av kunstige nevralt nettverk. Mange skjulte lag i de nevralt nettverket, gjør det mulig å utvikle prediktive DL modeller fra ende-til-ende, der postprosesseringsstrinn man ser i maskinlæremodellene er satt inn i det nevralt nettverket. (40) Figur hentet fra (40).

(40)

### 2.2.1 Tilpasse dyplæringsmodell

For bruk av dyplæring til en målrettet oppgave, er valg av dyplæringsmodell, arkitektur til modellen og modelltilpassing en del av utviklingen for nyskapende teknologi ved bruk av AI. Prosessen med å utvikle en dyplæringsmodell til en spesifisert oppgave, starter med å tilpasse treningsdata. Inn i modellen blir det matet inn strukturert og ustrukturert treningsdata. Ut ifra treningsdata utvikler modellen kunnskap. Et stort antall sett med treningsdata og variasjon i treningsdata vil være grunnlaget for å utvikle en robust modell (41). For store mengder treningsdata, og feil valg av treningsdata, kan gi lang beregningstid til modellen. Gode inputdata, betegnet som rensset data som representerer et mangfold, reduserer treningstiden og optimaliserer modellen til å gjenkjenne læringsmålet (42).

Modellens evne til å utvikle kunnskap om å utføre den spesifiserte læringsoppgaven blir omtalt som ytelsen til modellen. For å optimalisere modellen til å oppnå best mulig ytelse finnes det dedikerte dataverktøy (43).

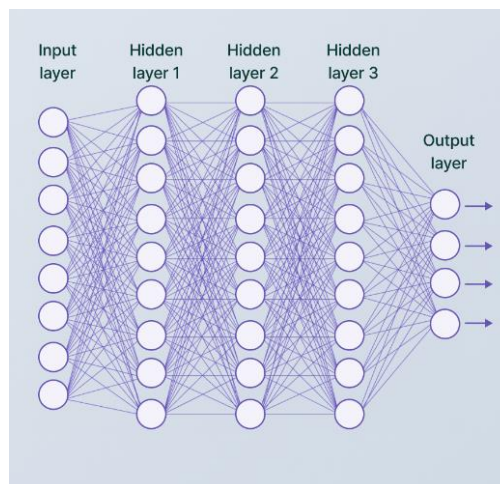
For å teste dyplæringsmodeller finnes det spesifiserte dataverktøy, som gir informasjon om treningsmodellens evne til å vise nøyaktighet, kvantifisering, sensitivitet og spesifisitet. Ut fra testresultatene til modellen blir treningsdata og funksjonsvalg justert, for å oppnå gunstig ytelse (43).

I testingen kan man også bruke en ROC- kurve for å måle modellens grad av sensitivitet og spesifisitet. Å måle outputbildet fra modellen opp mot en subjektiv radiologisk vurdering, er også en metode å vurdere ytelsen til modellen (43).

#### CNN

«Convolutional neural network» har flere lag enn ett nevralt nettverk (ANN), og benyttes for bildebehandling med dimensjoner, se figur 2.7. CNN spesifiserer dyplæringsmodeller med mer komplekse oppgaver. Det finnes flere ulike typer CNN, men felles for alle er at nettverket består av et stort antall skjulte lag. Dette gir mulighet til å skalere bildet i flere dimensjoner. Lagene i nettverket er bygd opp av små aktiveringsenheter som sender signaler mellom seg, som benytter prinsippene for lineær algebra, og er derfor godt egnet til å håndtere billedata for å trekke ut mønster. Det er i hovedsak tre lag som benyttes til å bygge CNN modeller: **Convolutional Layer**, **Pooling Layer** og **Full Connected Layer**. Kompleksiteten i lagene øker hvor dypere man kommer, første lagene benyttes til å identifisere enkle mønster, etter hvert øker kompleksiteten, der filter som beveger seg over rådata, med en binær tallmatrise, konvolusjon.

Tallverdiene i matrisen bestemmes av inngangsverdien, og en ny beregning blir gjort som utgangsverdi. Konvolusjon skjer mellom hvert lag. Signaloverføringer med input- og outputverdi mellom lagene brukes til å identifisere fremtredende strukturer. Totalen av beregninger med bruk av konvolusjon mellom lagene vil resultatet være utgangsvektor for output bilde (13).



**Figur 2.7:** Viser arkitekturen til et CNN-nettverk med flere lag, gjerne opp til 150 lag. I nettverket er konvolusjonslaget (input) og fullt komplett lag (output) et fullstendig sammenkoblet lag. Skjulte lag kan være en serie av konvolusjonslag og sammenslåingslag. Kompleksiteten i filtrering av data øker i dypet med lagene for å gjenkjenne et objekt eller forutsi en klasse. Mellom lagene sendes synapser, og vektoren i lagene er bestemt av inngangsverdien, og utgangsverdien er bestemt av beregninger i det enkelte lag. Flere slike lag, gir muligheten til å spesifisere bildet i flere dimensjoner og objektgjenkjenning (37;41). Figur hentet fra (37).

### 2.2.2 Deep Resolve

Deep Resolve kan benyttes på alle TSE-sekvenser i 2D. Programvaren er tilgjengelig for alle kroppsregioner, og for alle spoler som finnes på MR systemet. DR består av to teknologier, Deeo Resolve Gain og Deep Resolve Sharp. DR Sharp er den delen av teknologien hvor det er brukt dyplæring (19).

Deep Resolve må aktiveres før innhenting av rådata. Sharp-teknologien bygges på selve kjernen i frekvensdomene, der det er utviklet en målrettet DL modell for programvaren, etterfulgt med en type CNN (19).

DR er ikke mulig å kombinere med andre bildefilter i programvaren. Disse vil bli fjernet så snart man aktiverer DR (19).

### Deep Resolve Gain

DR Gain benyttes til støyreduksjon og gjenoppretter lokal støyvariasjon i MR-bildet. Støyen i MR-bildet er som regel ikke jevnt fordelt, på grunn av spolegeometri, og bruk av akselerasjonsteknikk som parallell imaging (PI). Lokal støyvariasjon kan ikke identifiseres og gjenoprettes av konvensjonelle bildefilter. DR Gain benytter individuelle støykart, som er basert på rådata. Denne informasjonen om lokal støyvariasjon, blir tatt direkte inn i rekonstruksjonsprosessen (36).

Områder i bildet med økt støy, får en større støyreduksjon enn i områder med lite støy. PI gir støyvariasjon i bildet, DR Gain skal minimere denne variasjon, og gi potensiale for å avbilde med økt PAT-faktor eller andre parameterendringer for å redusere opptakstiden, uten at det skal gå utover bildekvaliteten (36).

DR Gain kan graderes med faktor fra 1-8, der støyreduksjon øker med økt faktor. Enhancement er et konvensjonelt bildefilter for kantforsterkning og kan graderes fra 1-5. Både DR Gain og bildefilteret kan endres i etterkant av opptaket, forutsatt at DR er aktivert før opptaket (36).

### Deep Resolve Sharp

DR Sharp er utviklet med bruk av dyplæring og nevrale nettverk, data behandles i kjernen av frekvensdomene og må aktiveres før bildeopptaket. Ingen gradering, enten av eller på, interpoleringen ligger inne i selve algoritmen.

I utviklingen av teknologien, har det blitt utviklet dyplæringsalgoritmer på bakgrunn av treningsdata, et stort antall par av lavoppløselige og høyoppløselige datasett. Dyplæringsmodellen har blitt spesialisert for mønstergjenkjenning og objekt gjenkjenning i ulike dimensjoner. Deep Learning (DL) sammen med CNN skal kunne skape et høyoppløselig bilde, med lavoppløselig input. Gjenopprettingen for undersampling i bildeopptaket til informasjon i K-space, skal kunne gjenoprettes både i frekvens og faseretning. Dette skal gi potensiale til å akselerere bildeopptaket uten at det skal forringe bildekvaliteten (36).

Deep Resolve kan benyttes til tre ulike formål; oppnå bedre bildekvalitet på samme opptakstid, redusere opptakstid med samme bildekvalitet, eller øke bildekvaliteten noe, og redusere opptakstiden noe (44). I dette studiet er formålet å redusere opptakstiden og beholde samme bildekvalitet.

### 2.3 Standardisering

Å lage bilder med sammenlignbar bildekvalitet kalles en standardisering. Protokollen må optimaliseres og tilpasses med spesifikke parametere for det enkelte MR-system, og med kliniske hensyn til protokoll. Standardisering av protokoller er nødvendig for reproduserbarheten, og for å kunne gi en generalisert vurdering av bildekvaliteten (30).

Ved standardisering av protokoll er det viktig med kjennskap til vevsområdet og anatomien i den regionen som skal avbildes. En stor fordel med MR er at kontrasten i bildet kan manipuleres til å fremstille ulike vevstyper og eventuelle patologiske prosesser (45) Hvilke sekvenser, vektning og plan MR-protokollen skal inneholde begrunnes i hva man skal undersøke, og bestemmes av MR-radiologer. Radiografens oppgave er å avgjøre parametervalg for vektningene det bes om i protokollen, og lage best mulig bildekvalitet innenfor tidsrammen som er satt av til undersøkelsen (46).

### 2.4 Subjektiv vurdering av bildekvaliteten

Det kan være vanskelig å sette et standard mål for bildekvalitet, da dette er en subjektiv visuell vurdering fra radiologer. For å vurdere nye teknikker for innhenting av informasjon til MR-bildet, blant annet akselerasjonsteknikker, benyttes det gjerne både subjektiv og objektiv bildekvalitetsmålingsverktøy. For objektive mål blir det gjort beregninger ved hjelp av datamaskiner, såkalte bildekvalitetsmålingsverktøy (IQMs, fra engelsk Image quality metrics (47).

Hvor godt mål subjektiv visuell vurdering av bildekvaliteten er sammenlignet med bildekvalitetsmålingsverktøy, har det vært gjort studier på tidligere. Studiet til (47) viste at enigheten mellom radiologene var moderat. Gjennom arbeidserfaring blir radiologen lært opp til å ha et klinisk perspektiv i bildediagnostikken. Etter hvert oppstår ulike meninger om hva god bildekvalitet er blant radiologene, på bakgrunn av erfaring, miljø og at de er ulike individ. Erfaringene til radiologene, og utvalget av radiologer, vil påvirke resultatene ved bruk av subjektiv vurdering (47). Samsvar i resultatene til de som vurderer bildene, og enighet i vurderingen, viser variasjon mellom forskjellige studier. I studiet til (18) vises utmerket enighet mellom radiologene, der bildekvaliteten blir kun vurdert subjektivt.

### 3 Hensikt

Målet med studiet er å se om AI-protokollen kan erstatte standardprotokollen. En forutsetning for utskifting av protokollen er at bildekvaliteten er bevart, og at det ikke medfører noen betydning for diagnostikken. Gjennom subjektiv vurdering av bildekvaliteten til protokollene vil studiet evaluere ytelsen til Deep Resolve for å gjenopprette undersampling, og økt støyvariasjon i MR-bildet, i et klinisk miljø. I denne studien fokuserte vi på å evaluere bildenes generelle kvalitet etter gitte kriterier for studiet, og ut fra lesernes beslutning.

I dette studiet sammenligner vi bildekvaliteten til sekvensene i en standardprotokoll, med sekvensene til en AI-basert protokoll, der programvaren Deep Resolve er implementert. Bildekvaliteten ble vurdert kvalitativt av 4 radiologer med tanke på:

- 1) Å evaluere ytelsen til Deep Resolve for å opprettholde bildekvaliteten i AI-protokollen, sammenliknet med gjeldende standardprotokoll i klinikken.
- 2) Å observere variasjon i inter-leser-enighet for begge protokollene.

## 4 Metode

### 4.1 Valg av forskningsdesign

Kvantitativ utskiftningsstudie utført i klinikken, et eksperiment der to kneprotokoller, parvise variabler sammenlignes for bildekvalitet. Vurderingen av bildekvaliteten er en subjektiv kvalitativ tilnærming til resultatet, der leseren gir poeng på utvalgs-kriterier for bildekvaliteten, ved bruk av en 5-punkts likert-skala (1 = ikke-diagnostisk; 2 = lav bildekvalitet; 3 = moderat bildekvalitet; 4 = god bildekvalitet; 5 = utmerket bildekvalitet) (48). Det er også benyttet grad av blinding i studiet, der deltakerne ble blindet for hvilken protokoll som var Standard og AI, samt hverandres vurdering (49).

### 4.2 Optimalisering og standardisering

MR-protokollene ble optimalisert og standardisert før studiet startet av kandidaten for masteroppgaven, med samme innhold av sekvenser for vektning og plan for begge protokoller, se tabell 1. Instituttets standardprotokoll ble brukt som utgangspunkt for standardisering av protokoller til studiet. Denne protokollen ble optimalisert av kandidaten, og har vært i klinisk bruk i klinikken siden installasjon av MAGNETOM Altea, ca. 1 år. Før studiestart ble AI-protokollen optimalisert med utgangspunkt i standardprotokollen. I denne protokollen ble det implementert bruk av Deep Resolve og opptakstiden ble redusert med 35 %, se tabell 4.1. Som utgangspunkt for standardisering for sekvensvalg i protokollene ble det brukt European Society of Skeletal Radiology Sports Sub-committee (ESSR) sin veileder for MR kne protokoll (4). Protokollen ble endelig bestemt av fagansvarlig muskelskjelett-radiolog.

Parameterendringene som ble gjort mellom protokollene er vist i tabell 4.2. Før studiestart ble det gjort en liten pilotstudie med 10 pasienter, for å se om dette var to protokoller som kunne ha sammenlignbar bildekvalitet. MSK-ansvarlig radiolog godkjente disse protokollene før studiestart.

Parametere som påvirker kontrasten, TR, TE, flipp-vinkel og ETL ble holdt konstant mellom protokollene. Field of View (FOV), ble også holdt konstant, samt avstand mellom snittene, snittykkelse, og antall snitt, se tabell 4.3.

For å redusere opptakstiden ble antall fasekodinger og fase-oversampling redusert i PDFS sekvensene i AI-protokollen. Tiden ble ikke redusert med å øke PAT-faktor, da det ble sett at dette ikke ga sammenlignbar bildekvalitet mellom protokollene. Det ble redusert antall

avlesninger i frekvensretning, for å se på ytelsen til Deep Resolve for å kunne gjenopprette undersamplet data i K- space, både i frekvens og faseretning. For å finne grenseverdien for reduksjon i faseretning med oppløsningen i denne protokollen, ble det gjort en rekke tester fra 90 % ned til 70%. 80 % ble valgt for et passende mål, visuelt vurdert av kandidaten.

En reduksjon i både frekvens og faseretning ga en økt pikselstørrelse og økning i SNR i bildene. Etter en økning i SNR med økt pikselstørrelse, og en reduksjon i faseoversampling for å oppnå 35% redusert opptakstid, ble det relative SNR i AI-protokollen ca. 0.8, der standard har et relativt SNR på 1. I T1-sekvensen ble det gjort samme endringer for antall avlesninger i frekvens og faseretning. For å redusere tiden ble det der økt PAT- faktor, og SNR for sammenlignbar bildekvalitet ble justert med faseoversamplingen. Parametervalg i protokollene, se tabell 2. Opptakstid i ulike sekvenser i protokollene, se tabell 4.1.

Fysiske prinsipper innenfor MR gjorde at en endring av avlesninger i frekvensretning, endret TE i protokollen med DR. For å kunne holde TE konstant mellom protokollene i studiet, ble mottakerbåndbredden øket noe. Denne endringen medførte er noe lavere IES. Disse endringene var svært små, og det ble antas uten betydning for bildekvaliteten.

**Tabell 4.1:** Viser opptakstiden for sekvensene i begge protokollene, og den totale opptakstiden for hver enkelt protokoll.

Sekvens	Standard protokoll	AI-protokoll
Sag_PDFS	2.07 min	1.22 min
Cor_PDFS	2.03 min	1.20.min
COR_T1	46 sek.	29 sek
Axial_PDFS	2.03 min	1.20.min
Total opptakstid	7.00 min	4.30 min

**Tabell 4.2:** Viser parameterendringen som ble gjort mellom standardprotokoll og AI-protokoll.

Sekvens	Standard protokoll FOV 165x165				AI-protokoll FOV 165 x 165				
	Matrix.	PAT	NSA	Fase-oversampling	PAT	NSA	Matrix.	Fase-oversampling	DR
Sag_PDFS	256x256	2	1	120 %	2	1	240x192	75 %	ON
Cor_PDFS	256x256	2	1	110 %	2	1	240x192	75 %	ON
COR_T1	256x256	2	1	0 %	3	1	240x192	10 %	ON
Axial_PDFS	256x256	2	1	80 %	2	1	240x192	50 %	ON



**Tabell 4.3:** Viser parametervalg som holdes konstant mellom protokollene. Protokollene har samme innhold av sekvenser, FOV, gap, snittykkelse, TR, TE, ETL og antall snitt.

Sekvens	FOV	Gap	Snitt tykkelse.	TR	TE	ETL	Antall snitt
Sag_PDFS	165x165	10 %	3,5 mm	3370	40	8	28
Cor_PDFS	165x165	10 %	3,5 mm	2870	30	7	25
COR_T1	165x165	10 %	3,5 mm	454	8,8	3	25
Axial_PDFS	165x165	10 %	3,5 mm	3740	40	8	30

### Gradering av Deep Resolve Gain og bildefilter

For gradering av DR Gain ble det valgt faktor 6, som viste å gi best resultat for fettsupprimerte protonvektede sekvenser, og faktor 4 for T1. En for høy faktor av Gain gir økt støydemping, men viser seg også å gi mer uskarpe bilder. Bildefilteret for kantforsterkning ble satt til 4 på alle sekvenser.

### 4.3 Utvalg og utvalgsriterier

Det ble valgt ut 64 tilfeldige pasienter som var henvist til MR kne, og pasientene gjennomgikk to kneprotokoller under undersøkelsen. Det ble ikke sett på en spesiell klinisk problemstilling knyttet til kneleddet. Det ble heller ikke sett på beskrivelsene til undersøkelsene i dette studiet.

Utvalgsriterier for å bli inkludert, var at pasienten hadde blitt henvist til MR kne til klinikken der studiet ble utført. Alder mellom 18-90 år, uten kontraindikasjoner for MR. Utvalget omtrent likt fordelt mellom kjønnene, 34 undersøkelser kvinner og 30 menn.

Pasienter som ikke fikk plass i knespolen på grunn av størrelsen på kneet, og måtte undersøkes med fleks-spole ble ekskludert, fordi en fleks-spole gir vesentlig dårligere SNR enn dedikert knespole, og dermed ikke sammenliknbar. Undersøkelser som inneholdt bevegelsesartefakter, ble også ekskludert i datamaterialet, fordi artefakter var en bildekvalitetsparameter det ikke ble valgt å se på i dette studiet. For å unngå at det skulle påvirke vurderingen av bildekvalitet, ble undersøkelser med artefakter ekskludert.

Utvalget av radiologer for å delta i vurderingen av bildekvaliteten, ble tilfeldig valgt av MSK-ansvarlig radiolog i firmaet. Inklusjonskriteriet for å delta, var at radiologen dikterer til daglig MR kne i firmaet, og var kjent med standardprotokollen fra sin arbeidshverdag. Alle

undersøkelsene ble beskrevet som ordinært av en annen radiolog, før de utvalgte radiologene startet en systematisk gjennomgang av bildekvaliteten.

Radiologene har ulik erfaring og ulik bredde i kompetansen innenfor radiologien (se. tabell 4.4).

**Tabell 4.4:** Viser radiolog 1,2,3 og 4 i studiet sin erfaring som radiolog, antall års erfaring med å diktere MR kne, og gjennomsnittsvolum av antall kne undersøkelser radiologen har diktert per år. Tallene for gjennomsnittsvolum er gjort på bakgrunn av to siste årene. Lengre tilbake i tid, vil volumet være noe annet for enkelte radiologer som er inkludert.

Radiolog	Gjennomsnitt volum av antall diktete kne undersøkelser per år.	Antall år med praksis i å beskrive MR kne	Antall år med praksis radiolog
Radiolog 1	800 undersøkelser. Stor variasjon, bred kompetanse.	25 år	27 År
Radiolog 2	700 undersøkelser	13 år	13 år
Radiolog 3	1200 undersøkelser	18 år	18 år
Radiolog 4	1500 undersøkelser	2 år	6 år

#### 4.4 Testobjekter og variabler

Bildekvaliteten vil naturlig variere mellom pasientene, selv om det er brukt to standardiserte protokoller. Størrelsen på pasienten, benstruktur og patologi vil påvirke bildekvaliteten i stor eller liten grad mellom pasientene. Dette er variabler man alltid vil ha ved MR, derfor ble det inkludert minst 60 undersøkelser, for å kunne gi troverdige resultater for denne variabelen.

Man kan anta at hver enkelt radiolog som vurderte bildene har ulik oppfatning av bildekvalitet, og at det i så henseende er en variasjon mellom radiologene. Subjektiv vurdering vil gi resultater som påvirkes av individuelle ulikheter og preferanser mellom radiologene.

Objektposisjonering mellom venstre og høyre kne varierte mellom undersøkelsen. 15 cm høyre/venstre side, ut ifra isosenter.

#### 4.5 Datainnsamling, gjennomføring av prosjektet

Undersøkelsene ble utført hos Unilabs, avdeling Bryn, juli 2022. MR-skanner, Siemens MAGNETOM Altea 1,5 Tesla. Software XA31. 18 kanalers knespole (Tx/Rx Knee 18) ble brukt til alle undersøkelsene. Vurderingen av undersøkelsene etter kvalitetspoeng, ble utført i løpet av juli-august måned. Alle undersøkelsene som er inkludert i studiet var utført, før vurderingen av bildekvaliteten startet.

Begge protokollene ble utført under samme undersøkelse, pasient og bordposisjonering. For å sikre at sekvenser i samme plan, i og mellom protokollene ble utført identisk, ble det brukt MR-systemets kopieringsfunksjon for FOV og vinkling.

Alle undersøkelsene ble også utført av samme radiograf, kandidaten for masteroppgaven.

Radiografen markerte hvilken protokoll som ble utført først og sist på pasienten. Dette ble registret i et dokument der pasientopplysningene blir erstattet med en kode. Hvilken protokoll som ble utført først og sist ble ført inn i et skjema, uten innsyn for radiologene. Rekkefølge ble tilfeldig valgt for å unngå at radiologene kunne spore et system i billedtakingen. Hvilken kode pasienten har i studiet ble ført inn i et skjema, som har blitt oppbevart adskilt fra annen data. Skjemaet ble lagret i en kryptert mappe med passord på en USB-penn. Det er kun masterstudenten, veiledere og radiologer som er inkludert i studiet som har fått tilgangen til pasientens kode i studiet.

Alle bildene ble overført til klinikkens bildelagringsystem, Sectra PACS, fra modalitet umiddelbart etter at undersøkelsen var utført. Bildene har blitt oppbevart på vanlig måte i PACS. Bildene blir ikke tatt ut av datasystemet, og ligger i pasientens fil, som vanlig. Bildene ble hengt opp for visning i bildeprogrammet. En protokoll på hver skjerm, for direkte sammenligning. Protokollen som ble utført først på venstre skjerm, den andre på høyre skjerm, 4 delt skjerm. Radiografen noterte hvordan bildene ble hengt. Presentasjon for hengingen ble systemlagret. Det ble informert om at hengningene ikke kunne endres. Det ble også gitt informasjon, om at det ikke var lov for radiologen å gå inn på DICOM elementer på bildene.

Etter at undersøkelsen var utført, ble undersøkelsen diktert av en annen radiolog, før vurderingen av bildekvaliteten av utvalgte radiologer startet. Radiologene fikk tilsendt et skjema elektronisk i forkant, der det ble gitt instruks om kvalitetskriterier og poeng for bildekvalitet. I dette skjemaet ble pasientene identifisert med en kode, ikke personopplysninger. Radiologene utførte vurderingen på sin egen arbeidsstasjon, PACS.

Det ble gjort en systematisk vurdering av bildekvaliteten av alle radiologene, etter utvalgskriterier for bildekvalitet. Støy, skarphet, vurdering av menisk, korsbånd, ligament, benmarg og leddbrusk. Undersøkelsene ble nummerert kronologisk 1-64, der alle radiologene skulle vurdere i samme kronologisk rekkefølge 1-64. Etter endt vurdering fra radiologene på alle undersøkelsene, ble skjemaet med poeng, elektronisk sendt i retur til kandidaten for masteroppgaven.

Resultatene for hver enkelt radiolog ble deretter systematisert og satt inn i en tabell, etter protokoll med kategorier og poeng fra hver enkelt radiolog. Resultatene ble deretter analysert etter protokoll, kategori og radiolog.

#### 4.6 Eksperiment

Ut ifra formålet med studiet utledes følgende forskningsspørsmål:

*Vurdere om det er forskjell i bildekvaliteten mellom to MR kne protokoller. Opptakstiden i den ene protokollen reduseres med 35%, i denne protokollen benyttes programvaren Deep Resolve. Kan denne protokollen implementeres i klinikken til klinisk bruk, uten at det sees kvalitetsforskjeller i bildekvaliteten for diagnostikken?*

Av resultatet for kvalitetspoeng fra radiologene utføres en hypotesetest for å si om det er en signifikant kvalitets forskjeller i bildekvaliteten mellom protokollene.

*H0 = Det er ingen kvalitet forskjeller mellom standard protokoll, og AI.*

*H1= Det er kvalitetsforskjeller mellom standard protokoll og AI protokoll.*

## 4.7 Analyse

### 4.7.1 Valg av metoder for analysering av data

Poeng fra radiologene ga ordinale data som var datagrunnlaget for analysene. Statistiske analyser og grafiske illustrasjoner ble utført ved bruk av SPSS versjon 27 (IBM Corp). Deltakernes demografi og kliniske karakteristika ble oppsummert ved å bruke beskrivende statistikk og grafisk fremstilling. Resultatene fra radiologene viste seg å være skjevfordelte, og viste også svært liten grad av inter-leser enighet. Med ordinale data, som var skjevfordelte, ble det valgt å se på median og typetall for sentraltendens, for oppsummerende mål til resultatene. Med spredning i resultatene for kvalitetspoeng, og liten enighet blant leserne, ble det også valgt å se på Inter-kvartil-område (IQR, fra engelsk Interquartile Range), og 25 og 75 prosentilene for IQR, der boxplot viste at 25- og 75-prosentilene var mål av størst interesse, på bakgrunn av resultatene i dette studiet. Eksakte verdier for median og 25-og 75- prosentilene ble tatt ut fra SPSS i form av deskriptiv statistikk.

#### Hypotesetesting

For valg av hypotesetest ble det sett på studieutforming, variabler og datamaterialene i studiet. Med kun 4 radiologer med i vurderingen, kunne man ikke anta en normalfordeling. Radiologene var tilfeldig valgt, og ble ansett som uavhengige. Protokollene ble ansett som to parvise forsøk som ble utført gjentatte ganger, i dette tilfellet 64 undersøkelser. På grunnlag av data og variabler ble det valgt en ikke-parametrisk parvis t-test, der det ikke er krav om normalfordeling. En eksakt paret-sample Wilcoxon signed-rank test ble ansett som passende for studiet for å sammenligne protokollene for bildekvalitetspoeng fra radiologene (50) (51). Det ble valgt et signifikansnivå  $p < 0,05$  for en tosidig hypotesetest, med et konfidensintervall (CI, fra engelsk Confidence Interval) på 95% for å se om det var grunnlag i resultatene for å si om det var en signifikant kvalitetsforskjell mellom protokollene

#### Typetall og kvalitetsnivå for protokollene

Det ble gjort en grafisk fremstilling med stolpediagram, der det ble sett på typetall for å se hvor på skalaen radiologene hadde høyest frekvensen av gitt poeng. Typetall ble brukt for å se på kvalitetsnivået radiologene hadde vurdert protokollene til.

### Enighet mellom leserne - radiologene

Radiologene blir i dette studiet observatører sett fra et vitenskapelig perspektiv, og resultatene er basert på deres subjektive vurdering. I forskning blir grad av enighet mellom observatørene omtalt som grad av samsvar inter-leser-enighet (fra engelsk interrater agreement) (52). Det ble målt inter-leser enighet ved bruk av Fleiss Kappa, med et 95% konfidensintervall. For å tolke kappaverdier ble det brukt (53) sin anbefaling ble brukt til å tolke kappa-verdier, vist i tabell 4.1.

Fleiss Kappa, for å måle kappa-verdi, ble valgt ut ifra studieutformingen. Det ble gitt kvalitetspoeng ut ifra en likert skala 1-5, som ga ordinale data. Denne vurderingen ble gjort innen 7 kategorier av 4 lesere. Radiologene var tilfeldig valgt, og ingen av disse var unike innenfor gruppen de ble valgt. Radiologene ble sett på for å være uavhengige, som skulle vurdere samme undersøkelser. Det ble gitt informasjon om at det ikke var lov å diskutere vurderingen seg imellom i perioden vurderingen foregikk. Dette for å unngå at radiologene skulle påvirke hverandres resultater. Ut ifra dette ble Fleiss Kappa ansett som en passende test for å måle inter-leser-enighet.

Fleiss Kappa ble utført for å måle nivået, grad av enighet mellom disse radiologene, grad av overenstemmelse med et kjent utfall, 1-5 innen flere kategorier. Denne vurderingen ble gjentatt flere ganger, 64 undersøkelser. Kappa-verdier, og graden av enighet blir brukt til å vurdere påliteligheten til resultatene av samsvar.

At målemetoden er reliabel, gir godt samsvar og enighet mellom leserne er av betydning for generalisering og overførbarheten til resultatene (54). Større grad av enighet mellom radiologene for den ene protokollen, vil styrke resultatene for denne protokollen for å kunne være generaliserbare.

**Tabell 4.1:** Viser tolkning av verdier for Kappa i dette studiet, for grad av enighet mellom radiologene, med et konfidensintervall på 95%.

Kappa verdi	Tolkning
Nullhypotese Kappa = 0	Avtalen skyldes tilfeldigheter
0,01-0,020	Liten enighet
0,21-0,40	Rettferdig avtale
0,41-0,60	Moderat avtale
0,61-0,80	Betydelig avtale
0,81-1,00	Nesten perfekt avtale
Negativ (Kappa<0)	Enighet mindre enn forventet ved en tilfeldighet

#### 4.7.2 Kategorisering av datamaterialet

Resultatene for kvalitetspoeng ble først kategorisert for hver enkelt radiolog, etter protokoll og kategori. All kategoriseringen av data ble utført i Excel. Pasientskjemaet radiografen fylte ut under gjennomføringen av undersøkelsene med oversikt over hvilken protokoll som ble hengt på høyre og venstre skjerm, ble brukt til å kategorisere resultatene for riktig protokoll. Dette ble gjort for hver enkelt radiolog. (Se vedlegg A).

Etter denne kategoriseringen ble det laget en samlet kategori, anatomiske strukturer, med en sum av kategoriene menisk, korsbånd, ligament, benmarg og leddbrusk. Dette ble gjort for hver radiolog. Støy og skarphet ble beholdt som to separate kategorier. Videre ble data behandlet etter tre kategorier: 1) støy, 2) skarphet og 3) anatomiske strukturer.

Det kom raskt frem av resultatene at variasjon blant radiologene var stor, derfor ble det brukt noe mer tid på å analysere poeng fra hver enkelt radiolog separat enn hva det var tenkt ved oppstart av studien.

Det ble sett på oppsummerende mål for resultatene separat for hver enkelt radiolog, og i en samlet vurdering fra alle radiologene innenfor 3 kategorier. Til slutt ble det sett på resultatene samlet fra alle radiologene, der alle kategoriene ble slått sammen for hver enkelt protokoll.

I hypotesetestingen var den enkelte radiolog sin vurdering av hver enkelt undersøkelse datagrunnlaget for analysen i testen.

#### Kategorisering for oppsummerende mål som median, typetall, IQR og 25 og 75 prosentilene

For å se på resultatene fra alle radiologene samlet, innenfor 3 kategorier, ble resultatene fra alle radiologene satt under hverandre innenfor kategoriene støy og skarphet.  $4 \times 64$  undersøkelser/rader = 256 rader i Excel. For å få sammenlignbare resultater og for å kunne bruke samme metode i analysen for anatomiske strukturer, som for skarphet og støy, ble opprinnelige kategorier menisk, korsbånd, ligament, benmarg og leddbrusk, satt under hverandre i samme kolonne for anatomiske strukturer i 5 kategorier x 64 undersøkelser = 1280 rader i Excel.

For å sammenligne standard og AI-protokoll i en totalvurdering ble det gjort en ny kategorisering, der resultatene fra alle radiologene innenfor alle kategorier i hver protokoll, ble satt under hverandre. En kolonne for standard, og en for AI-protokollen, 2 kolonner 1792 rader i Excel innenfor hver protokoll,  $(64 \text{ undersøkelser} \times 4 \text{ radiologer}) \times 7 \text{ kategorier} = 1792$  rader.

### Kategorisering for å analysere kappa. Inter-leser enighet

Kappa, enighet blant leserne og grad av samsvar ble målt mellom leserne innenfor samme kategorisering som en analyse av resultatene for protokollene, støy, skarphet og anatomiske strukturer. For å få målbare resultater av vurderingen av 64 undersøkelser, ble poeng i opprinnelige kategorier som ligger under anatomiske strukturer slått sammen i en summasjon av kategoriene menisk, korsbånd, ligament, benmarg og leddbrusk. I denne analysen ble maks poeng for anatomiske strukturer 25 poeng.

### 4.8 Etikk

Det ble utført to kneprotokoller i hver undersøkelse på pasienten, som ble brukt til forskning. Prosjektet ble derfor meldt til Regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK). Av REK ble prosjektet vurdert til å gå utenfor helseforskningslovens virkeområde ref. (55).

Alle data som ble brukt i prosjektet ble anonymisert og pasienten koblet til en kode i prosjektet, som fulgte pasienten gjennom hele studiet. Det vil ikke være sporbart i publikasjon av masteroppgaven hvilke pasienter som har vært inkludert i studiet. Pasientene vil ikke ha mulighet til å gjenkjenne seg selv i publikasjon.

### Informasjon til pasientene

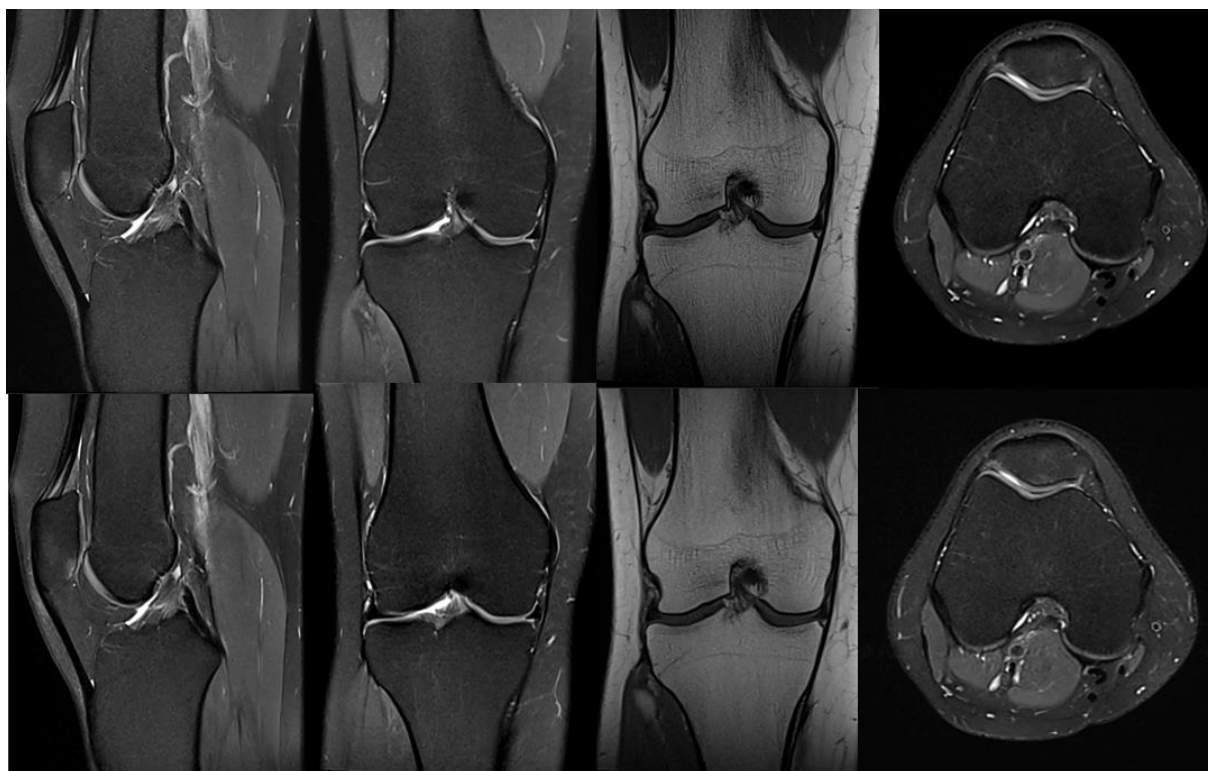
Det ble gitt informasjon til pasientene om deltakelse i studiet før undersøkelsen, i innkallingsbrevet, og samtykkeerklæring ble signert ved oppmøte til timen (Se vedlegg B). Ønsket ikke pasienten å delta i studiet, fikk pasienten tilbud om en ny time, til standard undersøkelse i Unilabs. Brevet til pasientene fulgte som vedlegg til REK søknaden. Alle studieprosedyrer ble utført i samsvar med de etiske standardene som er fastsatt i Helsinkierklæringen fra 1964 og senere endringer.



## 5 Resultater

Resultatene blir presentert i 3 bildekvalitetskategorier 1) støy, 2) skarphet og 3) anatomiske strukturer. Figur 5.1 viser bilder fra en pasient, med begge protokollene utført under samme undersøkelse. Øverst rad viser standardprotokoll, nederste rad viser bilder fra AI-protokollen.

Maks poeng for hver enkelt undersøkelse for kategoriene støy og skarphet var 5 poeng, i intervallet 1-5, hvor 5 er best. I kategorien anatomiske strukturer ble resultatene fra kategoriene menisk, korsbånd, ligament, benmarg og leddbrusk summert til en total poengsum for anatomiske strukturer. Maksimalt oppnåelig sum var 25 poeng for denne kategorien.

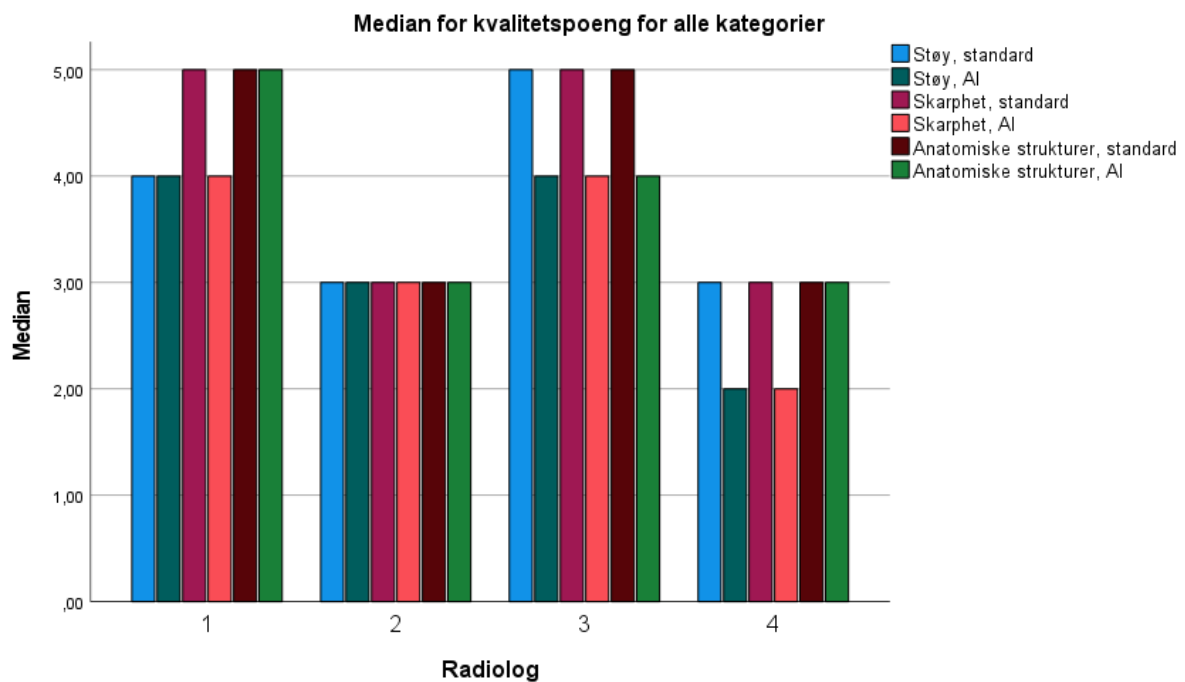


(33).

**Figur 5.1:** Viser bildene fra en undersøkelse i studiet, der man ser standardprotokollen i øverste rekke, og AI-protokollen nederste rekke. Denne undersøkelsen ble blitt gitt følgende kvalitetspoeng (standard – AI). Radiolog 1: Støy (3-4) Skarphet (5-4) Anatomiske strukturer (25-23). Radiolog 2: Støy (3-3) Skarphet (3-3) Anatomiske strukturer (11-23). Radiolog 3: Støy (5-4) Skarphet (5-4) Anatomiske strukturer (24-21). Radiolog 4: Støy (3-2) Skarphet (3-3) Anatomiske strukturer (15-14).

## 5.1 Resultater fra hver enkelt radiolog

Median for sentraltendens ble brukt som oppsummerende mål for resultatene i studiet. Resultatene for median for protokollene, fra hver enkelt radiolog innen kategoriene støy, skarphet og anatomiske strukturer er vist under i figur 5.2.



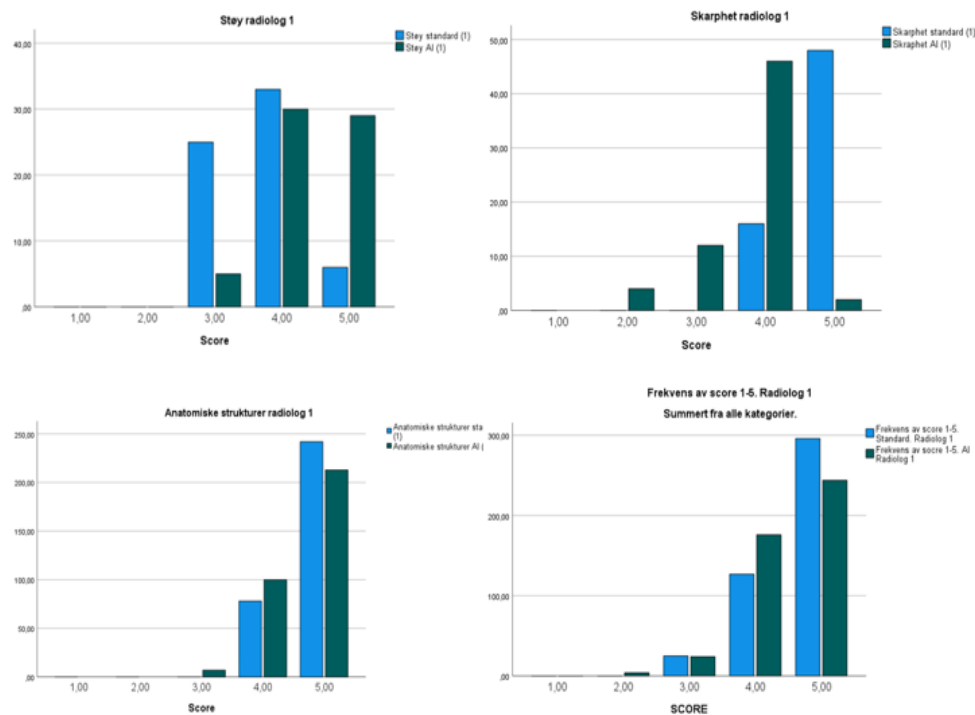
**Figur 5.2:** Søylediagrammet viser median for kvalitetspoeng til protokollene for alle radiolog, innen kategoriene støy, skarphet og anatomiske strukturer. 1, viser median for radiolog 1. 2, viser median for radiolog 2. 3, viser median for radiolog 3. 4, viser median for radiolog 4. Av søylediagrammet kan man se at resultatene for sentraltendens er noe lik for radiolog 1 og 3, det samme for 2 og 4.

### Radiolog 1

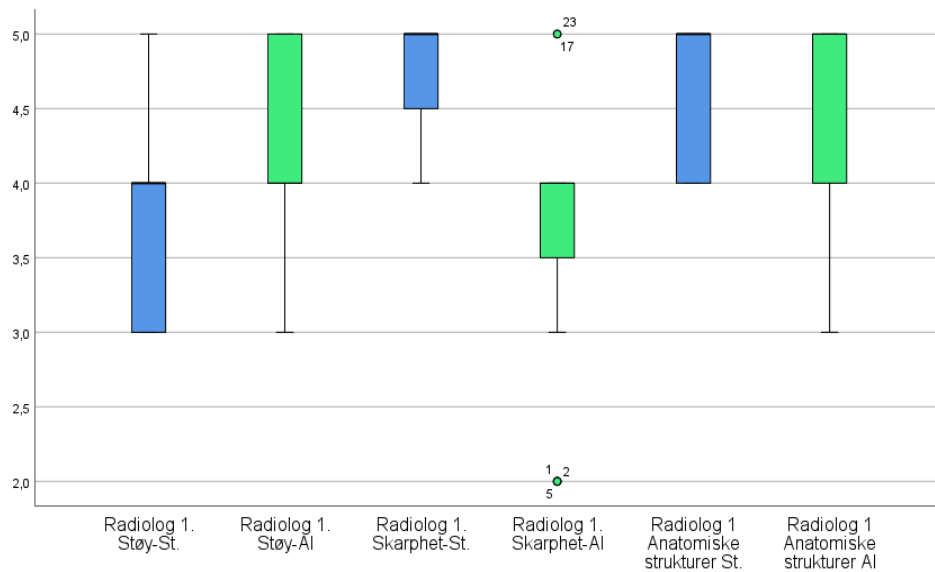
**Støy:** AI-protokollen ble vurdert av radiolog 1 til å ha best signal-støy forhold. Median standard - AI (4-4), se figur 5.2. Figur 5.4 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.3. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,000$ . AI-protokollen viste å være signifikant bedre enn standardprotokollen for mengde støy i vurderingen fra radiolog 1.

**Skarphet:** Standard protokollen ble vurdert av radiolog 1 til å ha best skarphet. Median standard - AI (5-4), se figur 5.2. Figur 5.4 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.3. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,000$ . Standardprotokollen viste å være signifikant bedre enn AI-protokollen for grad av skarphet i vurderingen fra radiolog 1.

Anatomiske strukturer: Standard protokollen ble vurdert av radiolog 1 til å ha best kvalitet for vurdering av anatomiske strukturer. Median standard – AI (5-5), se figur 5.2. Figur 5.4 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.3. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,002$ . Det er en signifikant forskjell mellom protokollene i vurderingen av anatomiske strukturer. P-verdien viste at kvalitetsforskjellen ikke stor i vurderingen fra radiolog 1, men sees det på typetall og boxplot ser man at den signifikante forskjellen der standard er best. Se figur 5.3 og 5.4.



**Figur 5.3:** Viser fire søylediagram for frekvensen av undersøkelser for gitt poengsum 1-5 fra radiolog 1, innen kategoriene støy, skarphet og anatomiske strukturer. Søylediagrammet nederst til høyre viser fordelingen av poeng på poengskalaen fra alle kategorier. **Typetall:** Standard – AI: Støy (4-4). Skarphet (5-4) Anatomiske strukturer (5-5) Total vurdering (5-5).



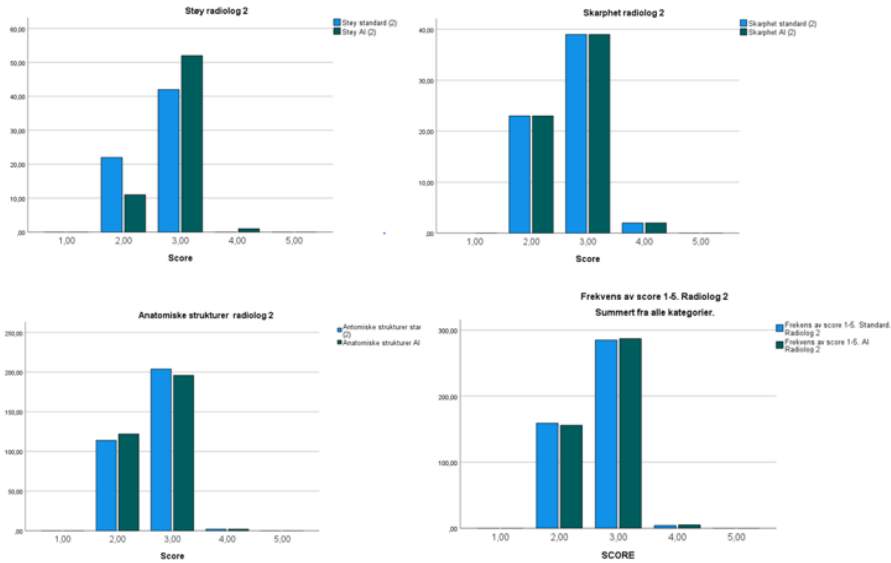
**Figur 5.4:** Viser boxplot for standard og AI protokoll innen 3 kategorier, støy, skarphet og anatomiske strukturer for Radiolog 1. IQR til protokollene standard -AI: Støy: (IQR 5-3/5-3). 25/75 prosentilene (3-4-/4-5). Skarphet: (IQR 5-4/5-2). 25/75 prosentilene (4,2-3,2-/5-4). Anatomiske strukturer: (IQR 5-4/5-3). 25/75 prosentilene (5-4-/5-5).

## Radiolog 2

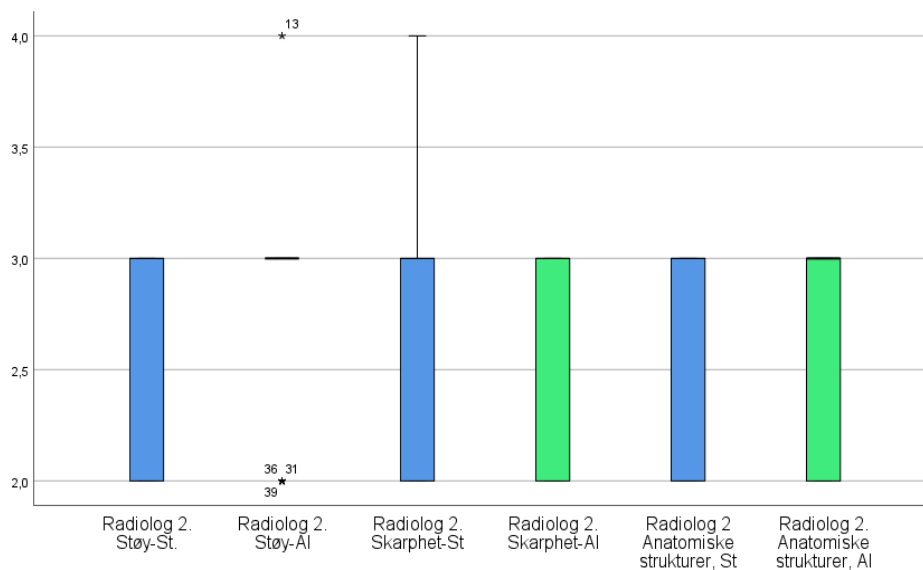
Støy: Ingen store kvalitetsforskjeller mellom protokollene for mengde støy, i vurderingen fra radiolog 2. Median standard - AI (3-3), se figur 5.2. Figur 5.6 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.5. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,029$ . Det er en signifikant forskjell mellom protokollene, forskjellen er ikke stor, men sees det på typetall og boxplotl, ser man at AI-protokollen er den protokollen som signifikant viser seg å være best. Se figur 5.5 og 5.6.

Skarphet: Ingen store kvalitetsforskjeller mellom protokollene for grad av skarphet, i vurderingen fra radiolog 2. Median standard - AI (3-3), se figur 5.2. Figur 5.6 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.5. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,585$ . Ingen signifikant forskjell mellom protokollene i grad av støy i vurderingen fra radiolog 2.

Anatomiske strukturer: Ingen store kvalitetsforskjeller mellom protokollene i diagnostikken for anatomiske strukturer, i vurderingen fra radiolog 2. Median standard – AI (3-3), se figur 5.2. Figur 5.6 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.5. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,585$ . Ingen signifikant forskjell mellom protokollene i kvaliteten for å diagnostisere anatomiske strukturer i vurderingen fra radiolog 2.



**Figur 5.5:** Viser fire søylediagram for frekvensen av undersøkelser for gitt poengsum 1-5 fra radiolog 2, innen kategoriene støy, skarphet og anatomiske strukturer. Søylediagrammet nederst til høyre viser fordelingen av poeng på poengskalaen fra alle kategorier. **Typetall:** Standard – AI: Støy (3-3). Skarphet (3-3) Anatomiske strukturer (3-3) Total vurdering (3-3).



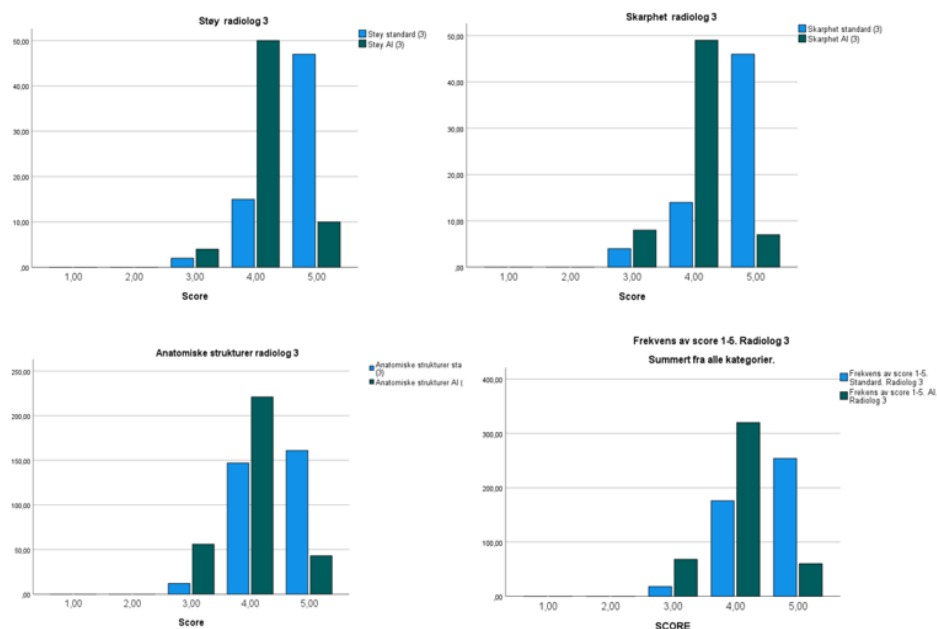
**Figur 5.6:** Viser boxplot for standard og AI protokoll innen 3 kategorier, støy, skarphet og anatomiske strukturer. IQR til protokollene Standard /AI Støy: (IQR 3-2/4-2) 25/75 prosentilene (2-3-/3-3). Skarphet: (IQR 4-2/3-2). 25/75 prosentilene (2-2-/3-3). Anatomiske strukturer: (IQR 3-2/3-2). 25/75 prosentilene (2-2-/3-3).

### Radiolog 3

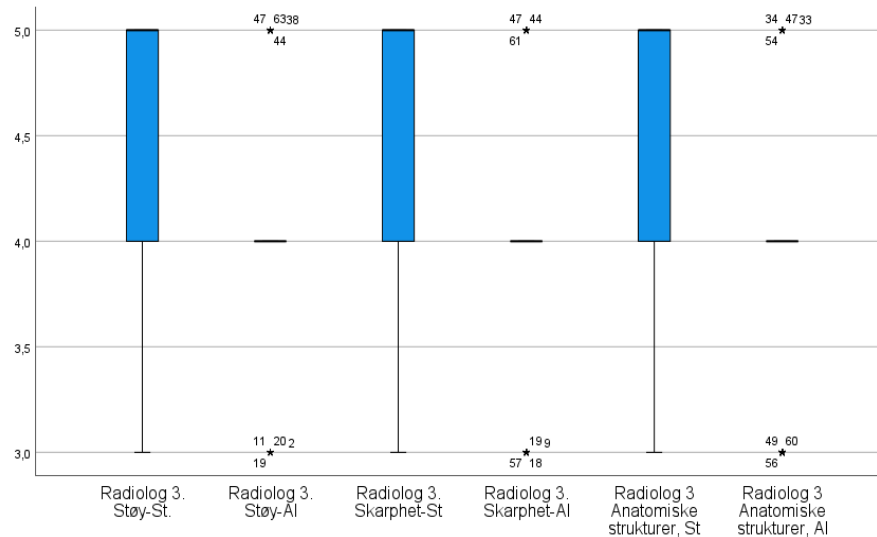
Støy: Standard protokollen ble vurdert av radiolog 3 til å ha best signal-støy forhold. Median standard - AI (5-4), se figur 5.2. Figur 5.8 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.7. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,000$ . Standard protokollen viste å være signifikant bedre enn AI-protokollen for mengde støy i vurderingen fra radiolog 3.

Skarphet: Standard protokollen ble vurdert av radiolog 3 til å ha best skarphet. Median standard - AI (5-4), se figur 5.2. Figur 5.8 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.7. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,000$ . Standardprotokollen viste å være signifikant bedre enn AI-protokollen for grad av skarphet i vurderingen fra radiolog 3.

Anatomiske strukturer: Standard protokollen ble vurdert av radiolog 3 til å ha best kvalitet for vurdering av anatomiske strukturer. Median standard - AI (5-4), se figur 5.2. Figur 5.8 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.7. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,000$ . Standardprotokollen viste seg å være signifikant bedre i vurderingen av anatomiske strukturer i vurderingen fra radiolog 3.



**Figur 5.7:** Viser fire søylediagram for frekvensen av undersøkelser for gitt poengsum 1-5 fra radiolog 3, innen kategoriene støy, skarphet og anatomiske strukturer. Søylediagrammet nederst til høyre viser fordelingen av poeng på poengskalaen fra alle kategorier. Typetall: Standard – AI: Støy (5-4). Skarphet (5-4) Anatomiske strukturer (5-4) Total vurdering (5-4).



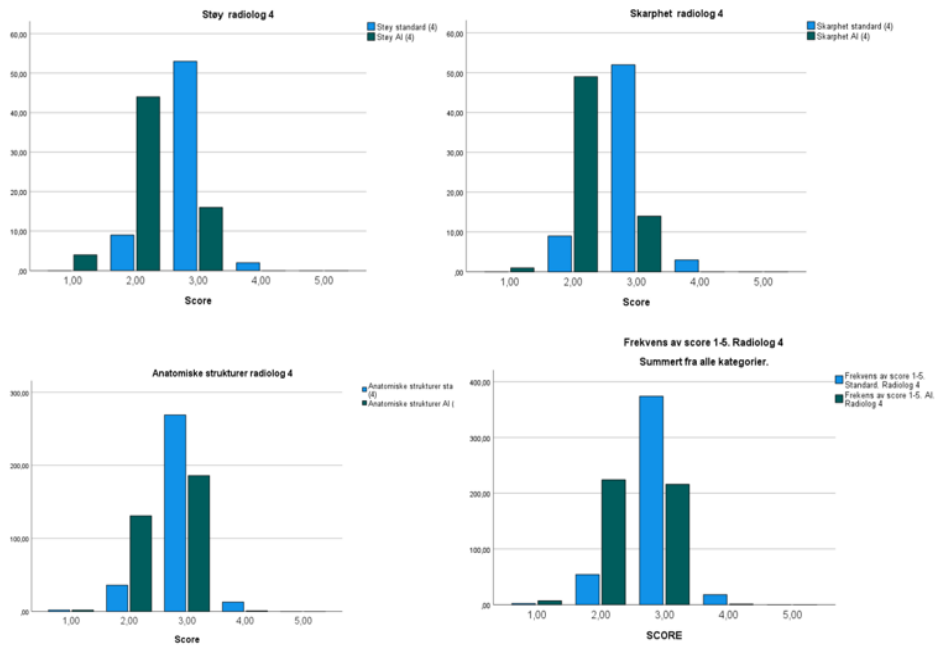
**Figur 5.8:** Viser boxplot for standard og AI protokoll innen 3 kategorier, støy, skarphet og anatomiske strukturer. IQR for protokollene standard /AI. Støy: (IQR 5-3/5-3). 25/75 prosentilene (4-4-/5-4). Skarphet: (IQR 5-3/5-3). 25/75 prosentilene (4-4-/5-4). Anatomiske strukturer: (IQR 5-3/5-3). 25/75 prosentilene (4-4-/5-4).

#### Radiolog 4

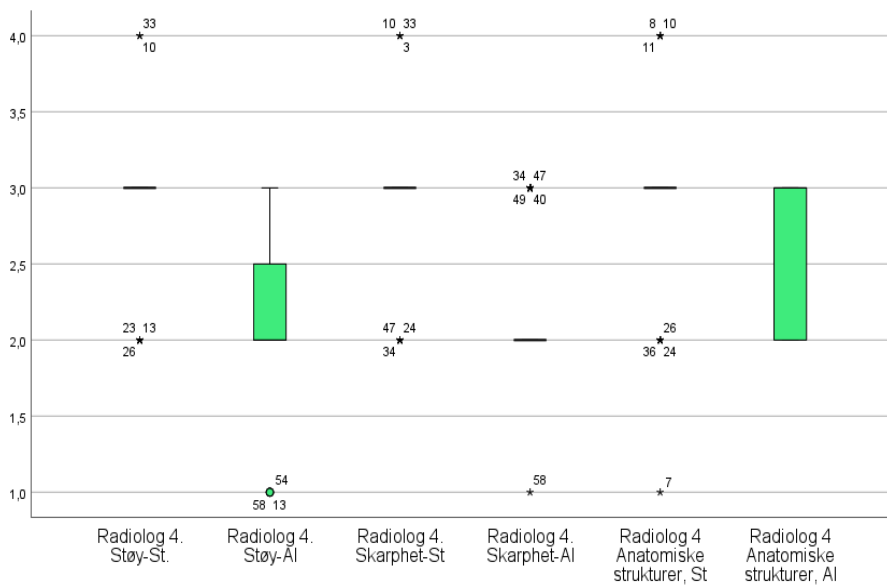
Støy: Standard protokollen ble vurdert av radiolog 4 til å ha best signal-støy forhold. Median standard - AI (3-2), se figur 5.2. Figur 5.10 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.9. To-side eksakt Wilcoxon-signed rank test viste en p-verdi < 0,000. Standard protokollen viste å være signifikant bedre enn AI-protokollen for mengde støy i vurderingen fra radiolog 2.

Skarphet: Standard protokollen ble vurdert av radiolog 4 til å ha best skarphet. Median standard - AI (3-2), se figur 5.2. Figur 5.10 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.9. To-side eksakt Wilcoxon-signed rank test viste en p-verdi < 0,000. Standardprotokollen viste seg å være signifikant bedre enn AI-protokollen for grad av skarphet i vurderingen fra radiolog 4.

Anatomiske strukturer: Standard protokollen ble vurdert av radiolog 4 til å ha best kvalitet for vurdering av anatomiske strukturer. Median standard – AI (3-3), se figur 5.2. Figur 5.10 viser resultater grafisk fremstilt i boxplot. Kvalitetsnivå, typetall og fordelingen av poeng 1-5 vist i figur 5.9. To-side eksakt Wilcoxon-signed rank test viste en p-verdi < 0,000. Standardprotokollen viste å være signifikant bedre i vurderingen av anatomiske strukturer i vurderingen fra radiolog 4.



**Figur 5.9:** Viser fire søylediagram for frekvensen av undersøkelser for gitt poengsum 1-5 fra radiolog 4, innen kategoriene støy, skarphet og anatomiske strukturer. Søylediagrammet nederst til høyre viser fordelingen av poeng på poengskalaen fra alle kategorier. Typetall: Standard – AI: Støy (3-2). Skarphet (3-2) Anatomiske strukturer (3-3) Total vurdering (3-2).



**Figur 5.10:** Viser boxplot for standard og AI protokoll innen 3 kategorier, støy, skarphet og anatomiske strukturer. IQR for protokollene standard /AI. Støy: (IQR 4-2/3-1). 25/75 prosentilene (3-2-/3-2,7). Skarphet: (IQR 4-2/3-1). 25/75 prosentilene (3-2-/3-2). Anatomiske strukturer: (IQR 4-1/4-1). 25/75 prosentilene (3-2-/3-3).

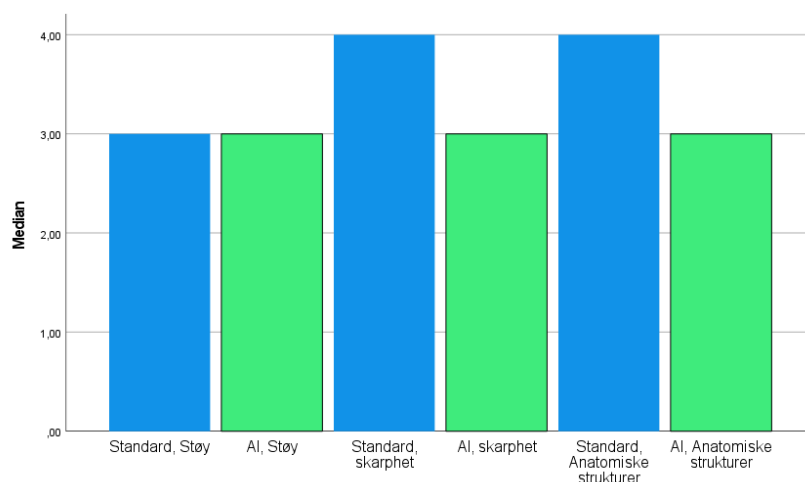


## 5.2 Resultater fra alle radiologene innen hver enkelt kategori

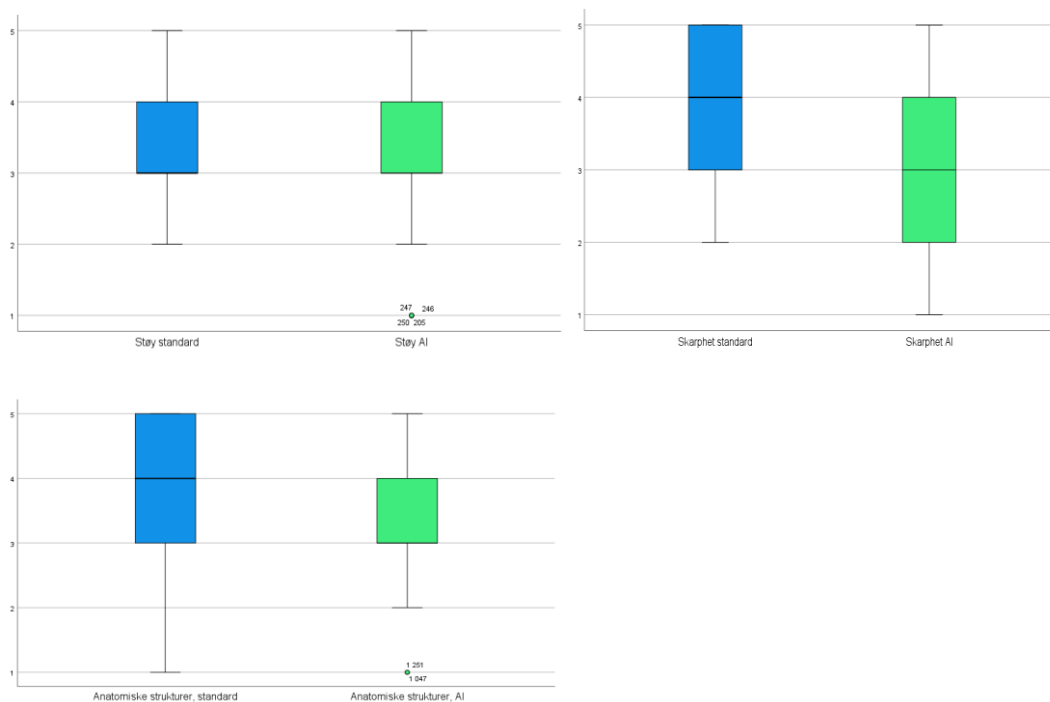
Støy: Ingen store kvalitetsforskjeller mellom protokollene i mengde støy i bildene, i resultatene basert på vurderingen fra alle radiologene. Median standard - AI (3-3), se figur 5.11. Figur 5.12 viser resultater grafisk fremstilt i boxplot. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,083$ . Ingen signifikant forskjell mellom protokollene, men hypotestesten viste en forskjell i favør AI-protokollen, der resultatene til alle radiologene er grunnlag for datamateriale.

Skarphet: Standard protokollen ble vurdert til å være best for grad av skarphet, i resultatene basert på vurderingen fra alle radiologene. Median standard - AI (4-3), se figur 5.11. Figur 5.12 viser resultater grafisk fremstilt i boxplot. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,000$ . Standardprotokollen viste seg å være signifikant bedre enn AI-protokollen for grad av skarphet, der resultatene fra alle radiologene er grunnlag for datamateriale.

Anatomiske strukturer: Standard protokollen ble vurdert til å ha best kvalitet for diagnostikken for anatomiske strukturer, i resultatene basert på vurderingen fra alle radiologene. Median standard – AI (4-3), se figur 5.11. Figur 5.12 viser resultater grafisk fremstilt i boxplot. To-side eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,000$ . Standardprotokollen viste seg å være signifikant bedre i vurderingen av anatomiske strukturer, i en vurdering der resultatene fra alle radiologene er grunnlag for datamateriale.



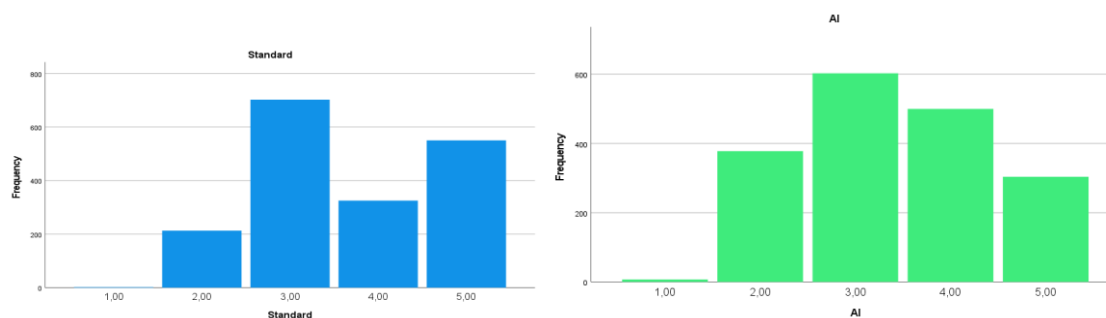
**Figur 5.11:** Viser median for standard og AI-protokollen, der resultatene fra alle radiologene er datagrunnlaget for median, i kategoriene 1) Støy. 2) Skarphet. 3) Anatomiske strukturer. Ut ifra figuren kan man se at standardprotokollen har median 4 innenfor kategoriene skarphet og anatomiske strukturer, mens AI-protokollen, har en median 3 innen alle kategorier.



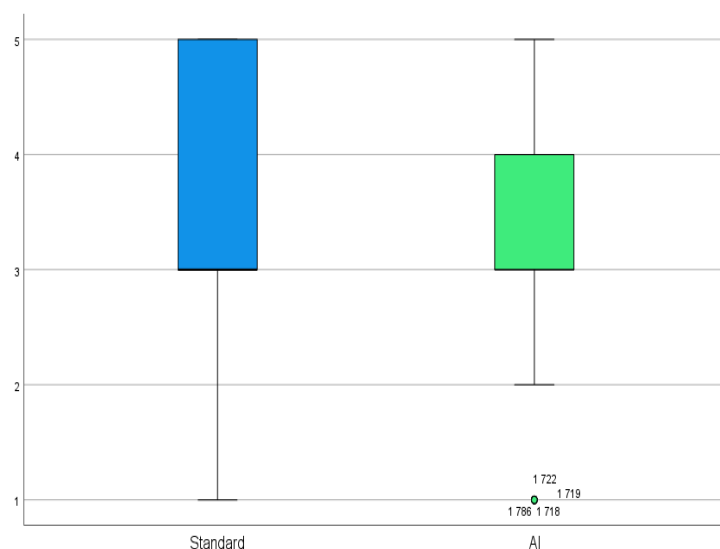
**Figur 5.12:** Viser boxplot for kvalitetspoeng fra alle radiologene innen kategorien støy, skarphet og anatomiske strukturer. Boxplot viser Støy: (IQR 5-2/5-2) 25/75 prosentilene (3-3-/4-4). Skarphet: (IQR 5-2/5-1). 25/75 prosentilene (3-2-/5-4). Anatomiske strukturer: (IQR 5-1/5-1). 25/75 prosentilene (3-3-/5-4). **Median** Støy: 3-3. Skarphet: 4-3. Anatomiske strukturer: 4-3. (AI protokollen viser 2 «outliers» innen kategoriene støy og anatomiske strukturer).

### 5.3 Oppsummerende resultater fra alle radiologene innen alle kategoriene

Resultatene oppsummert for begge protokollene, viste en to-sidet eksakt Wilcoxon-signed rank test viste en p-verdi  $< 0,000$ . P-verdien viser en signifikant forskjell mellom protokollene der standard var best. I en total vurdering fra alle radiologene innenfor alle kategorier mellom protokollene, viste median standard – AI (3-3), se figur 5.13. Kvalitetsnivå for protokollene og fordeling av poeng 1-5 vist i figur 21. Figur 5.14, viser grafisk fremstilling av resultatene gjennom boxplot.



**Figur 5.13:** Viser fire søylediagram for frekvensen av gitt poengsum 1-5 fra alle radiologer, for protokollene standard og AI. Typetall: Standard – AI: (3-3).

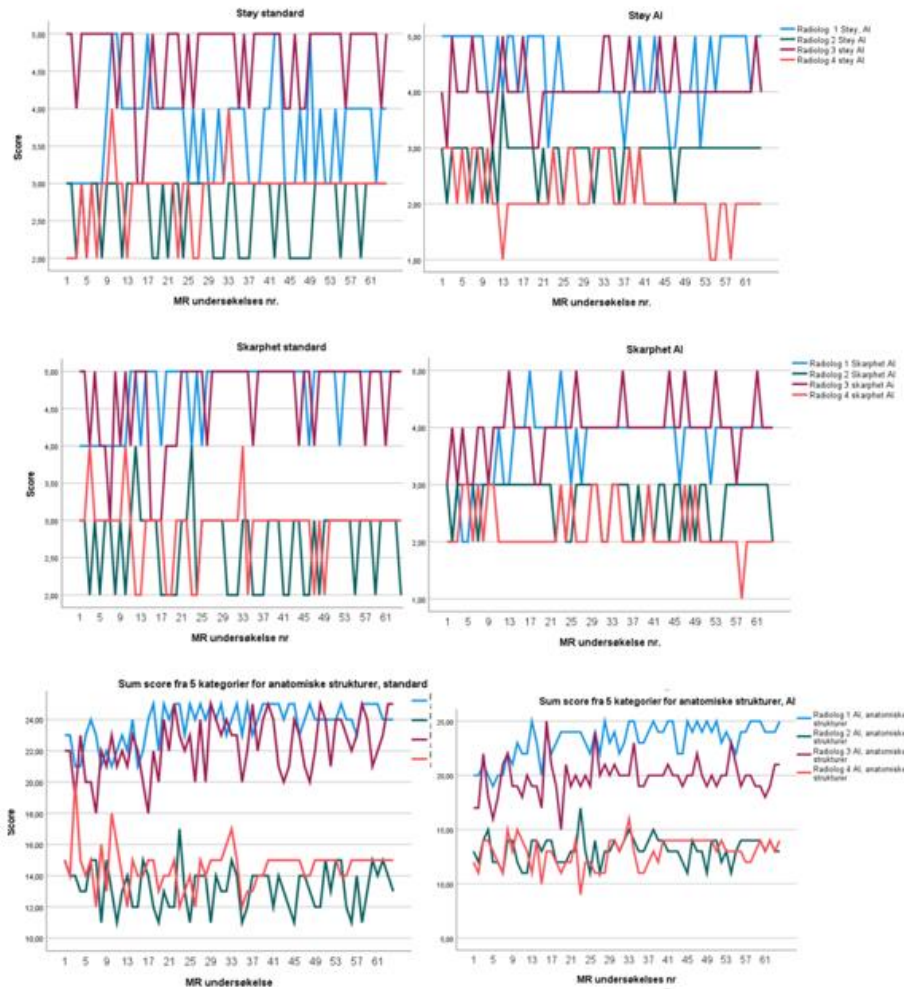


**Figur 5.14:** Viser boxplot for standard og AI protokollen, for poeng fra alle radiologene, samlet for alle kategorier for vurderingen av bildekvalitet. Boxplot viser IQR for standard-AI protokoll (IQR 1-5/1-5). Median (3-3). 25/75 prosentilene (3-3/5-4).

#### 5.4 Enighet mellom leserne, Inter-leser enighet

Kvalitetspoeng fra radiologene viste dårlig samsvar med negative kappa-verdier, med et CI på 95%, som tolkes som enighet mindre enn forventet ved en tilfeldighet, vist i tabell 4.1. Dette gir liten pålitelighet av samsvar i resultatene mellom leserne. Resultatene viser liten overenstemmelse i vurderingen blant radiologene.

Lite samsvar og enighet ble sett i vurdering av både protokoller, hver enkelt undersøkelse, og innen ulike kvalitetskriterier, kategorier. Kvalitetsnivået radiologene hadde vurdert protokollene for, viste også liten grad av samsvar. Under viser figur 5.15 linje-plot variasjon av kvalitetsnivå, og variasjon i vurderingen av hver enkelt undersøkelse for bildekvaliteten til. Protokollene, innenfor kategoriene støy, skarphet og anatomiske strukturer.



**Figur 5.15:** Viser linje-plot for kvalitetspoeng til 64 undersøkelser fra radiologene for protokollene standard og AI, innenfor kategoriene 1) Støy. 2).Skarphet 3) Anatomiske strukturer. Av linje-plot kan man se at det er lite enighet og samsvar mellom radiologene i vurderingen av bildekvaliteten, der bildekvaliteten er målt etter kvalitetspoeng 1-5, der 5 er best.

### Fleiss kappa

Graden av inter-leser enighet blant radiologene, med asymptotisk CI på 95% ga en kappa som vist i tabell 5.1 for grad av samsvar mellom radiologene.

**Tabell 5.1:** Viser grad av enighet mellom leserne. Alle kappa er negative. Lavest inter-leser enighet viser å være innen kategorien støy for AI protokollen, med en kappa verdi på -0,169. Enigheten viser å være størst innenfor kategorien skarphet for standardprotokollen, med en kappa verdi på -0,112.

Kategori og protokoll	Kappa
Støy, standard	-0,146
Støy, AI	-0,169
Skarphet, standard	-0,112
Skarphet, AI	-0,12
Anatomiske strukturer st.	-0,12
Anatomiske strukturer AI	-0,116

## 6 Diskusjon

### 6.1 Diskusjon omkring kvalitetspoeng

I dette studiet ble det sett på om vi kunne bytte ut en standard MR kneprotokoll, med opptakstid på 7.00 min, med en kneprotokoll der opptakstiden ble redusert til 4.30 min. I kneprotokollen med redusert opptakstid, AI-protokollen, ble programvaren Deep Resolve implementert.

Variasjon i vurderingen av bildekvaliteten var stor blant alle radiologer. Resultatene viste kappa-verdi lavere enn 0. I tolkningen av kappa, er det en enighet blant leserne som tolkes som «enighet mindre enn forventet ved en tilfeldighet». Det dårlige samsvaret i resultatene mellom radiologene ble sett i vurderingen av begge protokoller, innenfor alle kategorier for bildekvalitet i dette studiet, se figur 5.15. Kvalitetsnivået for den enkelte protokoll ble også vurdert ulikt mellom radiologene. I resultatene kom det frem to grupperinger av typetall og kvalitetsnivå blant radiologene for begge protokoller. Både standard og AI-protokoll ble av radiolog 1 og 3 vurdert til et høyere kvalitetsnivå enn fra radiolog 2 og 4. Kvalitetspoeng for den enkelte undersøkelse viste variasjon for begge protokoller, innen alle kategorier, se figur 5.15. Ut ifra resultatene sees det at radiologene har ulik preferanse og oppfatning av bildekvaliteten, og man ser at god bildekvalitet kan være et relativt begrep, med en subjektiv oppfatning av hva som defineres som god bildekvalitet.

Kvalitetsnivået til protokollene ble for både for standard og AI-protokoll vurdert til å være av god til utmerket kvalitet, med typetall 4-5 fra radiolog 1 og 3 innenfor innen alle kategoriene. Av radiolog 2 og 3 ble kvalitetsnivået vurdert til et annet nivå, moderat til lavt for begge protokoller, med typetall 2-3 i alle kategorier. Hva som er årsaken til denne store variasjon er vanskelig å begrunne, men viser tydelig at radiologene har ulik subjektiv oppfatning av hva som er god bildekvalitet. Dette blir nærmere diskutert under avsnittet «Enighet mellom leserne, Inter-leser enighet».

#### En oppsummering til resultatene for diskusjon

I resultatene fra en hypotesetest, to-side eksakt Wilcoxon signed rank test med et signifikansnivå på 5 %, CI 95%, der datagrunnlaget for testen var alle poengene gitt til protokollene innenfor alle kategorier fra alle radiologer, viste testen en p-verdi  $< 0,000$ . P-verdien med signifikansnivået satt for testen, viste at det var en signifikant forskjell mellom protokollene der standard var best. Andre oppsummerende mål med samme datagrunnlag viste

median, standard -AI protokoll (3-3). Typetall (3-3). (*IQR* 1-5/1-5), med 25/75 prosentilene (3-3/5-4).

Hypotesetest der det ble sett på resultatene fra alle radiologene innenfor kategorien støy, viste en p-verdi  $< 0,083$ . Det var ingen signifikant forskjell mellom protokollene, den lille forskjellen var i favør AI protokollen. Median for standard-AI innenfor denne kategorien var (3-3). (*IQR* 5-2/5-2) 25/75 prosentilene (3-3-/4-4).

Hypotesetest for resultatene innenfor kategorien skarphet der resultatene til alle radiologene ble tatt med i hypotesetesten, viste testen en p-verdi  $< 0,00$ . Det var en signifikant forskjell der standardprotokollen viste seg å ha bedre skarphet i bildene. Median for standard – AI-protokoll med samme datagrunnlag viste (4-3), (*IQR* 5-2/5-1) og 25/75 prosentilene viste (3-2-/5-4).

Hypotesetest der resultatene fra alle radiologene ble inkludert innenfor kategorien anatomiske strukturer, viste en p-verdi 0,00. Det var en signifikant forskjell, der standard var best. Median med samme datagrunnlag viste standard – AI protokoll (4-3), (*IQR* 5-1/5-1) og 25/75 prosentilene (3-3-/5-4).

I resultatene fra hver enkelt radiolog, og innen ulike kategorier viste hypotesetesten ulike utfall, om det var en signifikant forskjell eller ikke, hypotesetestene viste også variasjon i hvilken protokoll som signifikant viste seg å være best. Observatørene hadde ulik oppfatning av bildekvaliteten. Hadde det vært et annet utvalg av radiologer, der radiologen hadde preferanser om at god bildekvalitet, var forenlig med bilder som hadde godt signal-støy forhold, men skarpheten mindre essensiell for god bildekvalitet, kan det ikke utelukkes at resultatene ville vært annerledes. Kvalitetsforskjellene viser seg for å være små, variasjon i resultatene blant radiologene var stor. Med en lik median, ga resultatene lite grunnlag for å trekke konklusjoner ut ifra en hypotesetest alene, om at standardprotokollen har bedre bildekvalitet enn AI-protokollen, i en total vurdering av protokollene innenfor alle kategorier.

Forskjeller i skarphet og støy påvirker ikke nødvendigvis diagnostikk for anatomiske strukturer. Med så små kvalitetsforskjeller resultatene antyder, kan det ikke utelukkes at en overgang til AI-protokollen kan ha gjort diagnostikken vanskeligere for små strukturer. Men med variasjon som sees i vurderingen blant radiologene, kan det heller ikke utelukkes at i et gitt tilfelle i diagnostikken, eller for en enkelt undersøkelse ville en eller flere radiologer foretrukket AI-protokollen fremfor standardprotokollen. Er det en kvalitetsforskjell, vil små skader og forandringer som for eksempel bruskskader, bli vanskelig å avklare eller oppdage hvis bildene er uskarpe eller hvis det er mye støy. Om en endring til AI-protokollen vil medføre noen

oversette diagnoser er høyst uklart og usikkert. Det vil ikke nødvendigvis være mange diagnoser som går tapt eller som endres. Datagrunnlaget i dette studiet gir imidlertid ikke grunnlag for å si om det er diagnoser som vil gå tapt med å bytte ut standardprotokollen med AI protokollen.

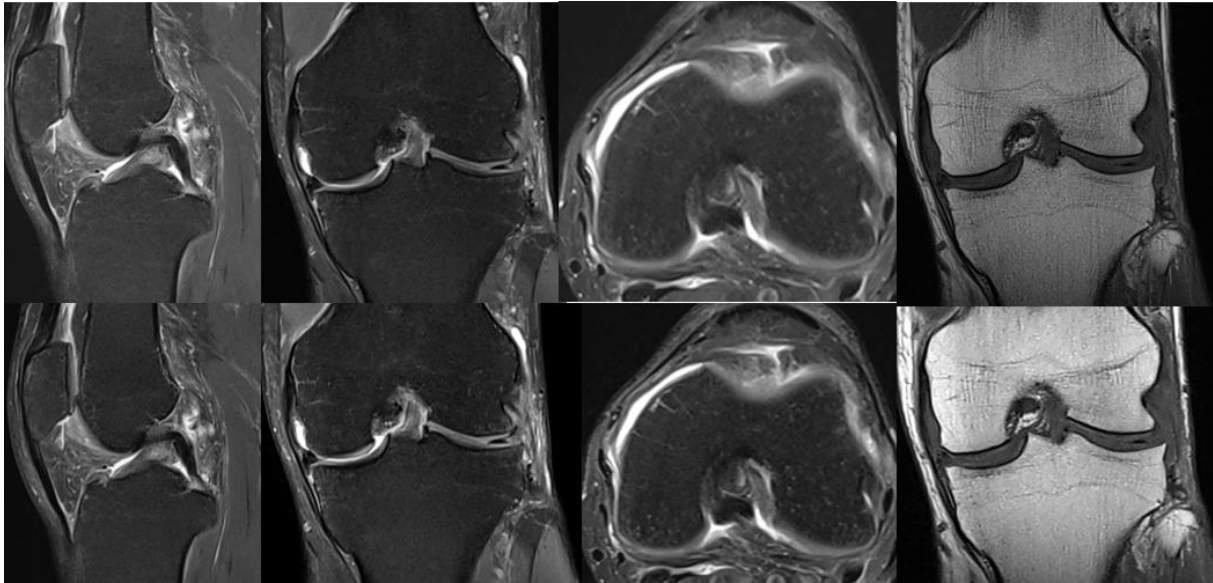
#### Resultatene sammenlignet med andre studier

I studiet til (18) der det også var gjort en subjektiv vurdering av bildekvaliteten, viste resultatene at AI-protokollen hadde en sammenlignbar eller bedre bildekvalitet for å diagnostisere anatomiske strukturer i kneleddet. Det kom ikke helt klart frem i dette studiet om det var en kvalitetsforskjell som vil ha påvirket diagnostikken, hypotesetesten viste en signifikant forskjell der standard var best, og man kan anta at dette utvalget av radiologer foretrakk standardprotokollen i diagnostikken. For mengde støy i bildene, fikk dette studiet samme resultater som i studiet til (18), AI protokollen viste seg å ha et bedre signal-støy forhold.

Opptakstiden til standardprotokollen i studiet til (18) var 11.56 min, der AI-protokollen hadde blitt redusert til 6.11 min, ca. 48%, en større reduksjon i opptakstid enn i dette studiet med utgangspunkt i en annen referanseprotokoll for sammenligning.

I resultatene til (18) viste bildene artefakter på 1,5 T. Artefakter i bildene ble ikke evaluert i dette studiet. Å evaluere grad av artefakter ble ikke sett på som relevant, da pilotprosjektet i forkant av studiet ikke viste noen artefakter i bildene. Generelt i klinisk bruk har ikke DR vist artefakter, på MR-systemet som undersøkelsene til dette studiet blir utført på. Det ble heller ikke rapportert om artefakter i bildene av radiologene som vurderte bildekvaliteten i studiet.

Variasjonen i bildekvaliteten vi så ved bruk av DR før studiestart, var ulik grad av skarphet og støy i bildene. Derfor ble det valgt å evaluere på radiologisk oppfatning av bildekvalitet innenfor kategoriene oppfattet støy, skarphet, og hvordan disse kvalitetsparameterne påvirket diagnostikken for anatomiske strukturer.



**Figur 6.1:** Viser bilder fra en undersøkelse, der standardprotokollen er vist øverst, med AI protokollen under til sammenligning på rad nr. 2. Bildekvalitetspoeng til denne undersøkelsen var en av mange der resultatene viste lite samsvar og enighet blant radiologene. Til denne undersøkelsen ble det gitt følgende kvalitetspoeng innenfor ulike kategorier til protokollene standard -AI.

Radiolog 1: Støy: (4-5) Skarphet: (5-4) Anatomiske strukturer: (24-23).

Radiolog 2: Støy: (3-3) Skarphet: (3-3) Anatomiske strukturer: (11-14).

Radiolog 3: Støy: (5-4) Skarphet: (5-4) Anatomiske strukturer: (23-20).

Radiolog 4: Støy: (3-1) Skarphet: (3-1) Anatomiske strukturer: (15-12).

### 6.1.1 Støy

I vurderingen for mengde støy i bildene hadde radiologene svært ulik oppfatning av protokollene, se figur 5.15. I linje-plot ser man at radiolog 1 og 2 har vurdert AI-protokollen til et høyere kvalitetsnivå for støy enn standardprotokollen. Radiolog 3 og 4 har vurdert standard til høyest kvalitetsnivå innenfor samme kategori.

I resultatene fra radiolog 1 var AI-protokollen signifikant bedre, det samme for radiolog 2, der forskjellen var litt mindre enn hos radiolog 1. Radiolog 3 og 4 viste resultatene en signifikant forskjell, der standard var best. Samlet var det ingen signifikant forskjell med en lik median, IQR og 25 og 75 prosentiler. Den lille forskjellen som kan sees om man ser på typetall, er at AI protokollen viser å ha best signal-støy forhold.

Gjennom resultatene kommer det frem at det er ulik oppfatning blant radiologene hva som er akseptabel støy, og bildekvalitet er svært subjektivt. Personlighetstyper kan ha påvirket resultatene, trolig har noen vurdert støy uavhengig om det er av betydning for diagnostikken eller ikke, mens andre har hatt diagnostikk som kriterium for hvilke karakterer de forskjellige parameterne skal få. Altså: "Har denne støyen betydning for diagnostikk, eller er den bare



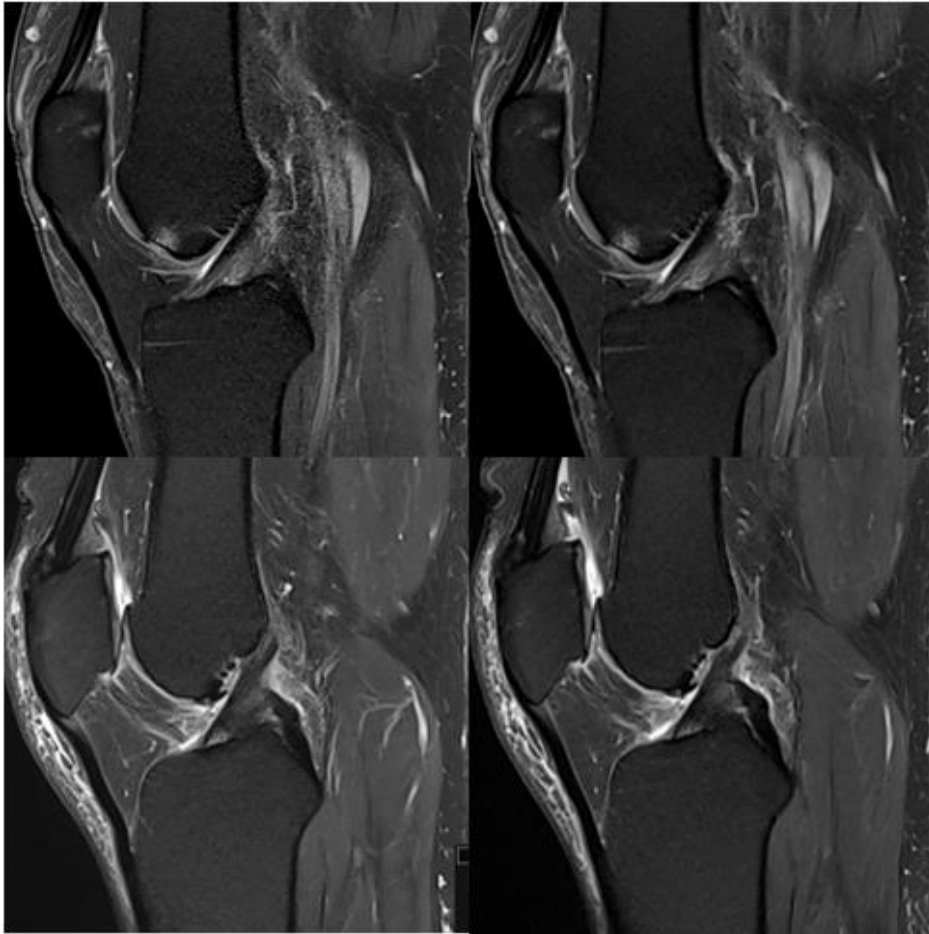
irriterende?" Hvilket utgangspunkt radiologen har hatt for vurderingen vil her gi forskjellige karakterer.

For å finne metode for å vurdere bildekvaliteten, ble det sett på hva som har blitt gjort i tidligere studier, der det forskes på Deep Resolve eller andre AI-baserte bilderekonstruksjonsteknologier. Det ble ikke funnet noen studier der det hadde blitt brukt objektiv vurdering i form av ROI-målinger. I tidligere studiet (17) og (18), der det forskes på AI-basert bilderekonstruksjonsteknologi, er det også kun brukt subjektiv radiologvurdering for bildekvaliteten. For å finne ut om det har blitt utført studier med bruk av ROI-målinger sammen med AI-basert bilderekonstruksjonsteknologi, ble det gjort søk i litteraturen i databasene Google Scholar, Scopus og Oria. I disse søkemotorene ble det ikke funnet bruk av objektivt bildekvalitetsmålingsverktøy for å måle bildekvalitet sammen med AI-baserte bilderekonstruksjonsteknologi innenfor MR. Bruk av ROI-målinger ville gitt et objektivt mål på SNR, og potensielt nyttig i evalueringen av SNR (støy) mellom protokollene. Dette ble imidlertid ikke utført i denne studien, av flere årsaker. Det primære målepunktet i studien var oppfattet bildekvalitet sett fra radiologens øyne. Videre ville ulik plassering av ROI medføre ulike resultater. Dette kunne vært standardisert ved at kandidaten gjorde disse målingene på alle undersøkelser. Målingene ville imidlertid ikke kunne sammenliknes fra pasient til pasient, kun mellom standard og AI-protokoll på samme sekvens. Likevel vil denne målemetoden ha svakheter. Metoden er designet for fantomer, ikke anatomiske strukturer. Videre ble parallell imaging og multielementspler benyttet i begge protokoller. Selv med korreksjonsfaktorer for dette vil reliabiliteten til målingene svekkes. For å kunne bruke ROI-målinger i dette studiet, var det også for lite kunnskap tilgjengelig, for hvor nøyaktig resultatene ble sammen med AI-basert bilderekonstruksjonsteknologi. For en god evaluering av teknologiene kan det for videre utvikling være nyttig med dedikerte objektive bildekvalitetsmålingsverktøy på AI-baserte MR-bilder.

#### *Parameterendringer som ble gjort, som påvirket SNR (Støy)*

Parameterendringer som ble gjort mellom protokollene i sekvensene, var blant annet en reduksjon i frekvens og fasekodinger i AI-protokollen. Disse parameterendringene alene vil gi et bedre SNR og mindre støy i bildene. For å redusere opptakstiden til 4.30 min i protokollen, ble fase-oversamplingen redusert. Etter optimaliseringen av AI protokollen var det relative SNR redusert til ca.20%. Ut ifra resultatene fra observatørene ser det ut til at Deep Resolve Gain har greid å gjenopprette denne reduksjon i SNR.

Økt sentral støy i MR-bildet sees i økt grad hos store pasienter der avstand til spolen og senter i bildet er større. Fordelingen av støy i bildet, vil også påvirkes av størrelsen på objektet. Med et stort objekt, vil bidraget av Gaussisk-støy i bildet øke innenfor et gitt FOV. Figur 6.2 viser samme bildeutsnitt fra protokollene til to pasienter, der den ene pasient er 115 kg den andre 72 kg. Bildene kan antyde at AI-protokollen gir et mer stabilt SNR for store pasienter, og kan bidra til en lik og sammenlignbar bildekvalitet til flere pasienter uavhengig av størrelsen på pasienten. Dersom dette er tilfelle, kan et annet utvalg av pasienter ha gitt andre resultater.



**Figur 6.2:** Utklipp av sag PDFS fra to ulike undersøkelser. Øverste bildene er fra en pasient på 115 kg, der man ser standardprotokollen til venstre, AI protokollen til høyre. Bildene under er fra en pasient på 72 kg, der standardprotokollen er vist venstre, AI protokollen til høyre.

### 6.1.2 Skarphet

I vurderingen for grad av skarphet til bildene viste hypotesetesten en signifikant forskjell mellom protokollene, der standard var best. Andre oppsummerende mål viste det samme. Median for standard - AI-protokoll med samme datagrunnlag viste (4-3). 25/75 prosentilene viste (3-2-/5-4).

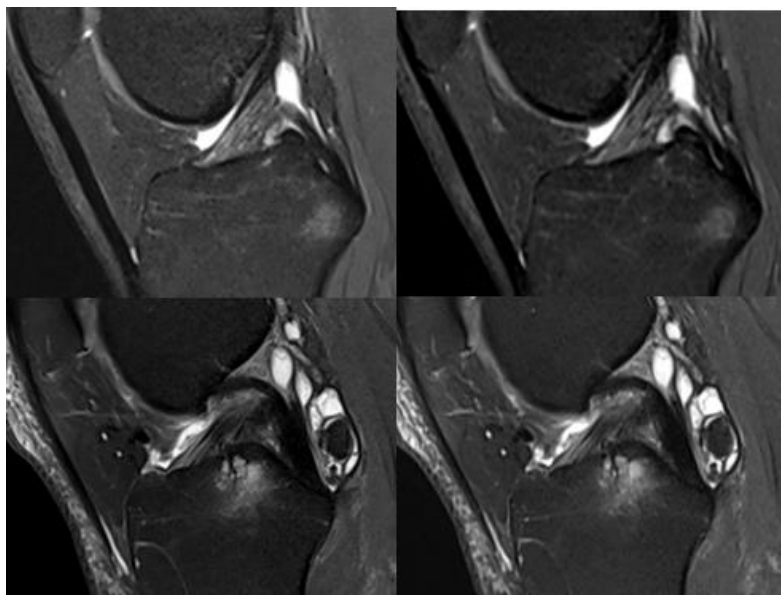
Ut fra resultatene kan man anta at det er en forskjell, og at standardprotokollen har en bedre skarphet og kanskje en bedre kvalitet for å gjengi små strukturer. Med størrelsen på utvalget av radiologer og resultater som viser liten grad av overenstemmelse i vurderingen, gir det ikke grunnlag for å trekke konklusjoner ut ifra resultatene. I vurderingen for skarphet, vil det trolig være samme faktorer innenfor individuelle preferanser, som har påvirket resultatene.

#### Parameterendringer som kan ha påvirket skarpheten

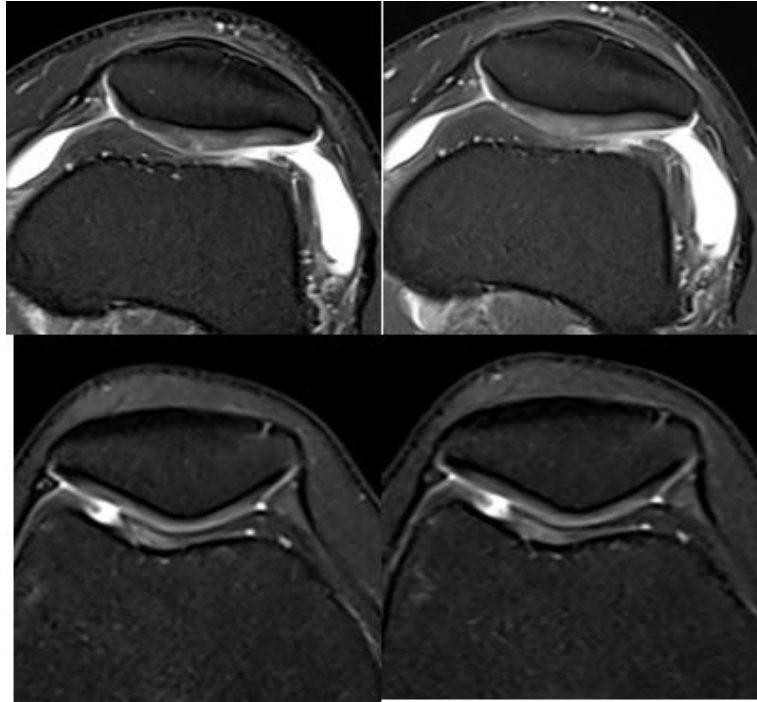
Skarpheten i protokollen påvirkes i størst grad av den romlige oppløsningen og bildenes evne til å gjengi små strukturer, noe som kan ha betydning for diagnostikken. I AI-protokollen ble det redusert antall frekvens og fasekodinger som ga protokollen en lavere romlig oppløsning. Studiets resultater i kategorien skarphet er derfor ikke overraskende. En del av hensikten med studiet var nettopp å se om DR Sharp har potensiale for å kunne kompensere for bildeopptak med lav-oppløselig input, og få høy-oppløselig output. På bakgrunn av resultatene i denne kategorien kan man antyde at DR Sharp ikke hadde god nok ytelse, til å kunne gjenopprette undersamplingen i protokollene som sammenlignes.

I standardprotokoll var matrisen 256 x 256 og FOV 165 mm dvs. pikselareal på 0,42 mm<sup>2</sup>. I AI-protokollen ble matrisen redusert til 240 x 192, dvs. pikselarealet på 0,59 mm<sup>2</sup>. Dette er en 30 % økning i piksel størrelse. Trolig ville en protokoll med høyere romlig oppløsning som referanseprotokoll gitt andre resultater for en 30 % økning i pikselvolum.. En høyere oppløsning ville gitt bedre input i dyplæringsalgoritmen. Høy oppløsning vil ha bedre detaljoppløsning og gi et bedre utgangspunkt for dyplæringsmodellen til å kunne gjenkjenne strukturer, og gi en bedre output etter bilderekonstruksjon. God input-data for å oppnå god output, er et av hovedprinsippene AI bygger på for å oppnå gode resultater (31). Dyplæringsmodellene er utviklet på bakgrunn av treningsdata, jo større nøyaktighet og romlig oppløsning inn i DL-modellen, jo bedre potensial har DL modellen gjenkjenne strukturer, mønster og objekt. Potensiale, med mål om å utføre spesifikke oppgaver, gjennom signaloverføringer i skjulte lag i det nevralt nettverket, øker med god input-data (11;31).

Under i figur 6.3 og 6.4 vises utklipp fra 2 ulike undersøkelser i studiet, der figurene er bildeutsnitt av protokollene. Grad av skarphet er essensielt for å vurdere anatomiske strukturer som vist i figurene. Av resultatene som forklart under figurene, kan man se en ulik vurdering fra radiologene med tanke på bildekvalitet knyttet til skarphet. Om det er en skarphetsforskjell som har noe å bety for diagnostikken, er det vanskelig å si noe om. En ting er om det har diagnostisk betydning for en enkelt radiolog. Bildene blir ofte sendt videre der pasienten skal behandles for kneplagene, i disse tilfellene er det også en ortoped som skal vurdere behandling ut ifra bildene. For en trent radiolog vil en redusert bildekvalitet kanskje ikke ha noen betydning, men for en ortoped kan kvalitetsforskjeller ha større betydning for vurdering av bildene, for eksempel i vurderingen av små bruskefekt. Et større utvalg av lesere med ulike kompetanse og preferanser, ville gitt en bedre vurdering på om AI-protokollen kan byttes ut med standardprotokollen for diagnostikken.



**Figur 6.3:** Øverst viser utklipp av sag PDFS-vektede sekvenser, fremre korsbånd til en pasient. Under vises utklipp av bakre korsbånd fra en annen pasient, i det sagitale planet. Standardprotokoll til venstre, AI til høyre. For grad av skarphet fikk disse undersøkelsene følgende vurdering av radiologene for bildekvaliteten til standard-AI protokoll: Radiolog 1 (5-5/5-4) Radiolog 2 (3-3/2-3) Radiolog 3(5-4/5-4) Radiolog 4 (2-2/3-2).



**Figur 6.4:** Øverst viser utklipp av patella PDFS-vektet sekvenser sett i det aksiale plan fra en undersøkelse. Under vises utklipp av patella i det aksiale plan fra en annen undersøkelse. Standardprotokoll vist til venstre til venstre, AI til høyre for begge undersøkelsene. For grad av skarphet fikk disse undersøkelsene følgende vurdering av radiologene for bildekvaliteten til standard-AI protokoll: Radiolog 1 (5-5/5-4) Radiolog 2 (2-3/3-3) Radiolog 3(5-4/5-4) Radiolog 4 (3-2/3-2).

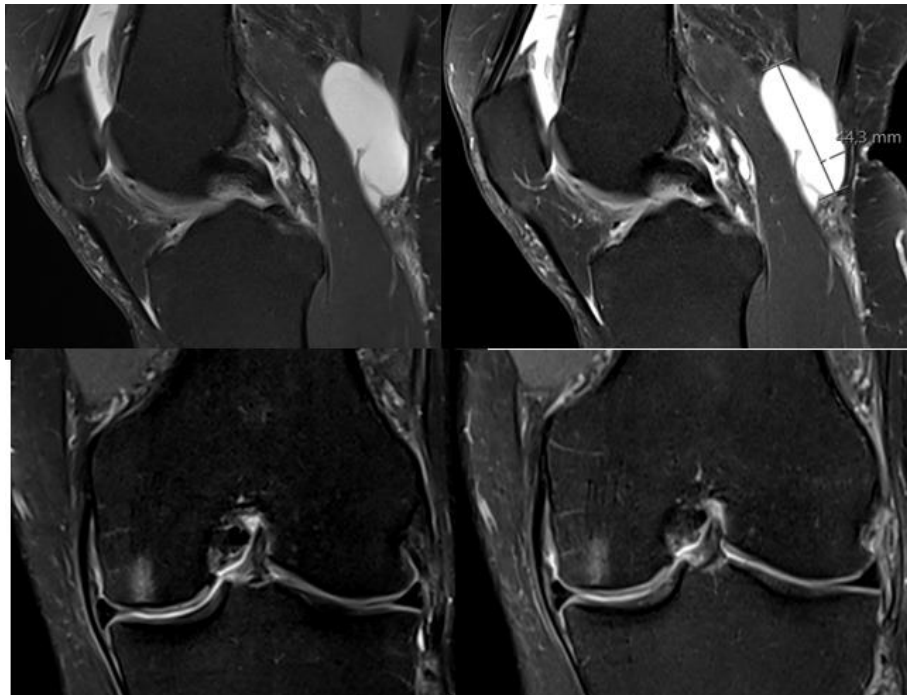
### 6.1.3 Anatomiske strukturer

Radiolog 1 hadde innen denne kategorien typetall 5 både for standard og AI, som indikerer en utmerket kvalitet til å vurdere anatomiske strukturer med bruk av begge protokollene. Radiolog 3 viste resultatene poeng innen samme enden av skalaen, men for radiolog 3 viste oppsummerende mål i resultatene at standard ble vurdert til å være best, med typetall 5 og en utmerket kvalitet, mens AI ble vurdert til typetall 4 og god kvalitet.

Radiolog 4 hadde kvalitetspoeng, der enkelte strukturer ble ansett til å ikke være diagnostisk med standardprotokollen, kvalitetspoeng 1. Ut ifra oppsummerende mål i resultatene kan man anta at radiolog 4 foretrekker standardprotokollen fremfor AI i diagnostikken, selv om det er flere undersøkelser som har fått kvalitetspoeng 1.

Innenfor kategorien er det undersøkelser radiolog 4 har vurdert for å ikke være diagnostiske, radiolog 1 og 3 har vurdert til utmerket bildekvalitet. Radiolog 2 har i denne kategorien høy frekvens av undersøkelser som er vurdert til moderat bildekvalitet innen begge protokoller, typetall 3. I figur 6.5 og 6.6 vises bildeutklipp fra 3 ulike undersøkelser. Poeng som er gitt for anatomiske strukturer viser liten grad av enighet. Det er tydelig at resultater basert på subjektiv

vurdering, påvirkes av den enkelte radiolog sin individuelle oppfatning om bildekvalitet. Den individuelle oppfatningen kan være påvirket av det radiologen er vant til å se gjennom arbeidserfaring. En årsak til at radiolog 4 har vurdert protokollene til et lavt kvalitetsnivå, kan være at denne radiologen til daglig dikterer i størst grad bilder fra Philips. Bildene fra Philips er ikke nødvendigvis av en bedre bildekvalitet enn Siemens. Uten å sammenligne bildekvalitet har bildene fra en Philips tomograf et annet bildeuttrykk, enn fra en Siemens tomograf. Dette kan ha påvirket oppfatningen av bildekvaliteten til protokollene.

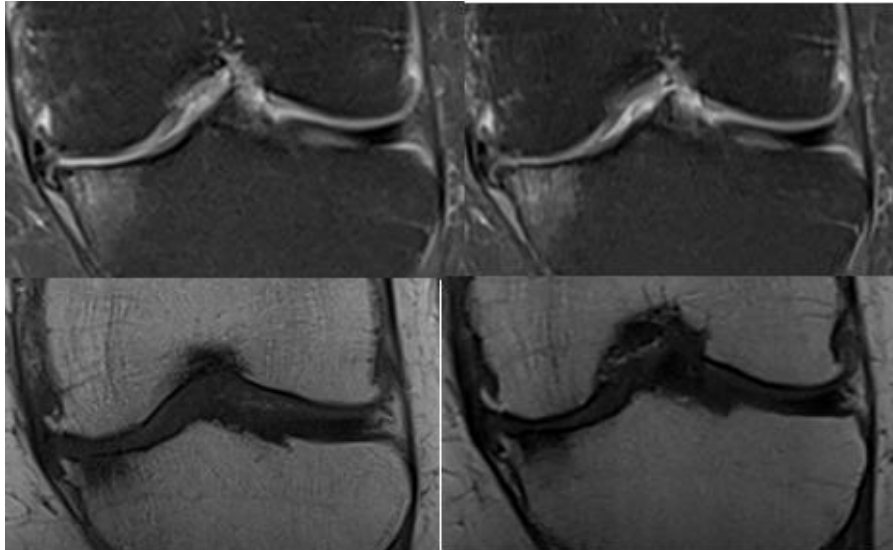


**Figur 6.5:** Viser PDFS fra to ulike undersøkelser. Øverst er det utsnitt av sag PDFS, der pasienten har en stor baker cyste bak kneleddet. Under er det utsnitt fra PDFS i det coronale planet, fra en annen undersøkelse, der pasienten har et benmargsødem i mediale femur condyle. Standardprotokoll vist til venstre, AI til høyre.

Radiologenes vurdering for anatomiske strukturer for disse undersøkelsene:

Radiolog 1 (24-23/25-25) Radiolog 2 (12-12/13-14) Radiolog 3 (23-22/25-23)

Radiolog 4 (12-11/13-12).



**Figur 6.6:** Viser PDFS og T1 i det coronale plan fra en og samme undersøkelse, standard til venstre, AI til høyre. Degenerative benmargsødem lags leddflaten i mediale tibia condyl. I denne undersøkelsen kan standardprotokollen kanskje vise å være skarpere i T1 vektet serie.

Radiologenes vurdering for anatomiske strukturer for denne undersøkelsen:

Radiolog 1 (25-24) Radiolog 2 (14-15) Radiolog 3 (24-21) Radiolog 4 (14-12).

## 6.2 I hvor stor grad er det grunnlag for å si om det er kvalitetsforskjeller?

En annen studieutforming kunne kanskje gitt et bedre svar på om det er bildekvalitetsforskjeller mellom protokollene som påvirker diagnostikken. I en utforming der radiologene hadde fått tildelt protokollene i to perioder med ca. 1 måned mellomrom, blindet for hvilken protokoll som ble tildelt i den enkelte perioden. Med denne studieutformingen hadde ikke radiologene hatt mulighet til å direkte sammenligne protokollene. Kvalitetspoeng gitt til protokollene uten direkte sammenligning, ville gitt en annet utgangspunkt for radiologene i å sette et mål på bildekvaliteten gjennom poeng. Med denne studieutformingen, ville det ikke ha vært like stor grad muligheten til å gjenkjenne et mønster i protokollene for radiologene.

For å se på en eventuell forskjell i diagnostikken, ville man fått en bedre tilnærming med at radiologene hadde diktert undersøkelsene i hver periode. Gjennom å analysere diktatene fra hver protokoll, ville det kanskje gitt et bedre svar på om det var en forskjell i diagnostikken mellom protokollene.

En studieutforming i likhet med dette ville krevd mer ressurser, begrensninger til dette studiet gjorde det ikke mulig å gjennomføre. Det kunne også gitt andre utfordringer i å tolke resultatene. Trolig ville man med en slik studieutforming se at radiologene hadde en større variasjon i kvalitetsnivå til de enkelte protokollene og den enkelte undersøkelse mellom periodene, da det ikke kan utelukkes at vurderingen av radiologen kan påvirkes av dagsform.

I hvor stor grad det er forskjell mellom protokollene for diagnostikken ut ifra resultatene i dette studiet er vanskelig å anslå, men ut ifra figur 6.7, kan det antas at det ikke er store forskjeller, selv der små strukturer skal diagnostiseres.



**Figur 6.7:** Viser bildeutsnitt av coronale PDFS vektet sekvens. Det er en skade i bakre horn av mediale menisk. Ut ifra dette bildet, kan det ikke sees store kvalitetsforskjeller mellom protokollene, selv der det skal sees på små strukturer. Standardprotokollen vist til venstre, AI til høyre.

### 6.3 Kappa, grad av inter-leser enighet

Fleiss Kappa ga kappa-verdier som var lavere enn 0, negative. I en tolkning av kappa for inter-leser enighet mellom leserne i dette studiet, tolkes dette som «enighet mindre enn forventet ved en tilfeldighet». (56) tolker kappa-verdier som vi fikk av Fleiss på lik linje med å kaste krone og mynt. I følge (56) er det sjeldent dette er tilfelle, hva som er årsaken til dårlig samsvar og enighet blant leserne i dette studiet er vanskelig å si. Gjennomsnitts kappa-verdier for protokollene viste å være lav for begge, standard - 0,135, og AI -0,126. Dette er tall som ikke viser noen differanse i gjennomsnittskappa mellom protokollene, som gir grunnlag for å si at påliteligheten til resultatene er større for den ene protokollen enn den andre.

En årsak til dårlig samsvar og lave kappa -verdier, kan i et tenkt tilfelle skyldes at radiologene har ulik arbeidserfaring og kompetanse. Gjennom arbeidserfaring har radiologene opparbeidet seg et ulikt klinisk perspektiv på bildediagnostikken, og oppfatningen av god bildekvalitet kan være ulik. På bakgrunn av erfaring og miljø, kan det ha utviklet seg egne meninger hos radiologene som har påvirket resultatene. Hva som er årsaken til variasjon for dette utvalget av radiologer er vanskelig å si, men det kan antas at personlighetstyper, og rett og slett forskjellige oppfatninger, vil være årsaken til variasjon i karaktersetningen for bildekvaliteten.



### Enigheten mellom leserne i dette studiet, sammenlignet med lignende studier

Av studier som det ble sett på for sammenligning, var det ingen med resultater som er forenelig med resultatene i dette studie. I studiet til (47) var det ikke bruk av AI-basert bilderekonstruksjonsteknologi, men bruk av subjektiv vurdering for bildekvalitet. Der ble det sammenlignet bruk av ulike objektive bildekvalitetsmålingsverktøy, sammenlignet med mål i form av subjektiv vurdering av 5 radiologer. Der viste den subjektive vurderingen av bildekvaliteten kappa-verdier 1- 0,66. I studiet til (18), der det også sees på AI basert bilderekonstruksjonsteknologi ble bildekvaliteten vurdert av 3 radiologer. Resultatene viser kappa-verdier mellom 0,61 – 1,00, med et CI på 95%.

Kappa-verdier fra  $< 0,00$  i dette studiet er en svakhet til resultatene. Påliteligheten svekkes, i en tolkning av kappa kan resultatene sees på som på som tilfeldige. Hva som er årsaken til et så dårlig samsvar, er ikke mulig å fastslå av informasjon som er tilgjengelig.

Leserne ble blindet for hvilken protokoll som var vist på høyre og venstre skjerm mellom hver undersøkelse. Radiologene gikk gjennom 64 undersøkelser. Gjennom vurderingen av undersøkelsene var det ikke mulig å unngå at radiologene fikk mulighet til å se karakteristiske forskjeller i protokollene. Personlige preferanser, og hva radiologene var vant til å se på gjennom arbeidserfaring, har trolig påvirket resultatene.

På grunn av dårlig enighet blant radiologene og lite samsvar i resultatene, kan disse resultatene kun sees på for dette utvalget av radiologer.

### Hva kunne gitt økt grad av inter-leser-enighet?

En bedre innføring til radiologene i å vurdere bildekvalitet etter likert skala. Til daglig ser radiologene etter eventuell patologi i bildene av kneleddet, for å stille en eventuell diagnose ut ifra kliniske opplysninger, det er ikke sikkert alle var forenelig med å sette karakter på bildekvalitet i form av poeng innen ulike kategorier ut ifra bilene. For sammenlignbar bildekvalitet og reproducerbar bildekvalitet, var det behov for standardisering av to protokoller. Informasjon om at det var standardprotokollen som ble brukt som referanseprotokoll, ville trolig vært av betydning for resultatene. I resultatene kan man se at det er ulik oppfatning om hva som er god bildekvalitet. Kvalitetsnivået til denne protokollen ble i dette studiet ansett som god bildekvalitet, og godkjent av MSK-ansvarlig radiolog i forkant av studiet. Det var klinikkens protokoll som var utgangspunktet for kvalitetsnivå. Pasienter som kommer til Unilabs for å ta en MR kne undersøkelse, skal ofte til førstegangsutredning, eller til kontroll for kjente plager. Etter klinikken sitt behov for protokoll, ble denne protokollen ansett for å ha god

bildekvalitet. En bedre informasjon om hva som ble betraktet som god bildekvalitet i dette studiet, vil trolig ha gitt bedre samsvar i resultatene for kvalitetsnivå for protokollene.

Støy kan ha blitt feiltolket, da støy er uønsket i bildet, og hvordan dette påvirker bildekvaliteten og diagnostikken kan ha blitt oppfattet ulikt blant radiologene. Det ble tatt utgangspunkt i at radiologene innehar grunnleggende kunnskap til å vurdere mengde støy i bilde, etter likert skala. Her kunne det ha vært en bedre presisering av å vurdere bildekvaliteten for mengde støy, der man setter et tall på dette.

### God bildekvalitet, relativt begrep i en subjektiv vurdering av bildekvaliteten

God bildekvalitet er et relativt begrep, og det finnes ulike subjektive meninger på hva god bildekvalitet er. Med utgangspunkt i en protokoll med høyere romlig oppløsning, et godt signal støyforhold og lengre opptakstid, ville bildekvaliteten vært bedre ut ifra fysiske prinsipper MR bygger på. Enhver radiolog ønsker seg nok i utgangspunktet bilder med en bildekvalitet, uavhengig av opptakstiden. Det finnes ulike oppfatning om hva som er godt nok bildekvalitet for diagnostikken, som trolig er årsaken til at radiologene har gitt protokollene ulikt kvalitetsnivå.

Sannsynligvis er en viktig årsak til resultatet, at bildekvalitet er et så flytende begrep hos radiologer at det er svært vanskelig å sette tall på det. Poenget med om det har diagnostisk betydning eller ikke er svært avgjørende for radiologer.

Utvalget av radiologer var tilfeldig, innen utvalget var det ulik grad av erfaring blant radiologene. Gjennom arbeidserfaring vil trolig radiologene ha opparbeidet seg et ulikt klinisk perspektiv i bildediagnostikken, og sin egen oppfatning av hva god bildekvalitet er. Behovet for radiologens preferanse om god bildekvalitet i diagnostikken, vil også være ulikt blant radiologene (47). I studiet til (57) ser man at en trent radiolog ikke har samme krav til bildekvalitet for å diagnostisere ACL rupturer, sammenlignet med en uerfaren.

God bildekvalitet gjør diagnostikken enklere og sikrere, uavhengig av erfaring og kompetanse til radiologen. Men det kan tenkes at en trent radiolog ikke er like avhengig av gode bilder. I resultatene til dette studiet, kan man antyde en sammenheng mellom vurderingen av kvalitetsnivå til protokollene og arbeidserfaring til radiologene. Radiolog 1 har lengst arbeidserfaring, og har vurdert begge protokollene til det høyeste kvalitetsnivået blant radiologene. Radiolog 3 med nest lengst arbeidserfaring, har vurdert begge protokollene til det nest høyeste kvalitetsnivået. Radiolog 2 har mindre arbeidserfaring enn både radiolog 1 og 3,

og vurdert kvalitetsnivået i begge protokollene til et lavere nivå, men høyere nivå enn radiologen 4, med minst arbeidserfaring, vist i tabell 4.4. Alle disse radiologene dikterer MR kne i sin arbeidshverdag, trolig er ikke arbeidserfaring det som har påvirket variasjon i resultatene i størst grad i dette studiet, men i et tenkt tilfelle kunne det vært årsaken. Trolig har personlige preferanser rundet bildekvalitet, og hva radiologene «liker» å se påvirket resultatene i større grad.

En annen årsak til resultatene knyttet til kompetanse, er at radiologene med lengst arbeidserfaring har vært med på utviklingen innen MR, der bildekvaliteten har utviklet seg i takt med utviklingen med nye MR-systemer. Bildekvaliteten innenfor MR har hatt en rivende utvikling, der fremskritt i teknologien har gjort MR til et unikt verktøy innen bildediagnostikken, for blant annet utredning av knesmerter (5) (10) (36). Bildekvaliteten da radiolog 1 begynte å beskrive MR kne, var på et langt lavere kvalitetsnivå enn det vi ser i klinisk bruk i dag. Arbeidserfaring med ulik bildekvalitet gjennom flere år, kan ha påvirket den subjektive oppfatningen av hva som defineres som god bildekvalitet, og kvalitetsnivået for den enkelte protokoll.

#### 6.4 Standardisering og optimalisering av protokollene

Standardprotokollen var utgangspunktet i optimaliseringen av AI-protokollen. Med utgangspunkt i en protokoll med lengre opptakstid og høyere kvalitetskriterier for romlig oppløsning, vil trolig ha gitt andre resultater med en 35 % reduksjon i opptakstid. Ser man på bruk av DL som bilderekonstruksjonsteknologi og prinsippene bak en dyplæringsmodell, etterfulgt med bruk av nevralt nettverk, vil data man putter inn i dyplæringsmodellen, påvirke det man får ut, output. Høyere oppløsning i bildene ville gitt andre inngangsverdier i DL-modellen og en bedre forutsetning for å trekke ut flere detaljer mellom lagene i det nevralt nettverket.

I dette studiet ble det brukt en protokoll som allerede var presset ned i tid, 7.min opptakstid, en reduksjon med 50 % her ville gitt en opptakstid på 3,5 min. Sammenlignet med studiet til (18) der opptakstiden ble redusert fra 11.56 min til 6.11. min, vil det gi et annet utgangspunkt for å få en 50% reduksjon av opptakstiden. AI-protokollen i studiet til (18) er nærmere lik opptakstid til standardprotokollen for dette studiet. Etter opptakstiden, kan det antas at AI-protokollen i studiet til (18) er en protokoll med høyere oppløsning enn protokollen her. Med høyere oppløsning, vil kanskje ikke en reduksjon i frekvens og faseretning og piksel volum, ha like stort utslag på bildekvaliteten. Det er vanskelig å evaluere og direkte sammenligne resultatene

i dette studiet opp mot resultatene i studiet til (18), da det er ulike protokoller som sammenlignes, utført på ulike MR-system.

Leverandørene reklamerer med opptil 50% reduksjon av opptakstid ved bruk av AI. Sees det på opptakstiden redusert i %, ser vi her at dette må tas i betraktning av hva det tas utgangspunkt i. Referanseprotokoll vil trolig sette noen begrensinger for hvor mye man faktisk kan redusere opptakstiden med.

### Parametervalg

I dette studiet ble det valgt å holde parametere som kan påvirke CNR i bildet konstant mellom protokollene, for å kun fokusere på kvalitetskriteriene støy, skarphet og anatomiske strukturer. Andre variabler som kan påvirket bildekvaliteten ble valgt å holdes konstant mellom protokollene.

En reduksjon i antall frekvens og fasekodinger som ga muligheten til å redusere TR. Dette ble ikke gjort, da det var en parameter som ble satt for å holdes konstant mellom protokollene. En endring av TR ville ikke ha kommet i konflikt med TR for ønsket vektning i bildene. Trolig vil ikke en endring i TR ha påvirket kontrasten i bildene av betydning. En reduksjon i TR kan potensielt ha gitt noe lavere SNR, og påvirket evaluering av støy. Hadde denne parameterendringen blitt gjort, ville opptakstiden blitt redusert med 40 % og et større potensialene av Deep Resolve ville blitt brukt, uten at det trolig har påvirket resultatene. Et minus med denne endringen, er at radiologene hadde hatt mulighet til å se en endring i TR mellom protokollene, og det ville være enklere å gjenkjenne protokollene.

### 6.5 En oppsummert evaluering av Deep Resolve

I dette studiet der opptakstiden reduseres med 35 %, viser resultatene at Deep Resolve har potensiale som akselerasjonsteknikk innenfor MR. AI-basert bilerekonstruksjonsteknologi ser ut til å kunne være med på å gjøre MR til en mindre tidkrevende undersøkelsesmodalitet, der bildekvaliteten sees å være lik med redusert opptakstid. I hvor stor grad opptakstiden kan reduseres uten at det sees endringer i bildekvaliteten, gir ikke dette noe studiet svar på. Ut ifra erfaring, teoretiske prinsipper MR og AI bygger på, samt tidligere studier kan det antas referanseprotokoll har stor betydning for i hvor stor grad opptakstiden kan reduseres, der bildekvaliteten er bevart.

Sannsynligvis vil en overgang til en AI-protokoll ikke bety noe for en erfaren radiolog, og den vurderingen vil ikke bety noe for en erfaren ortoped heller. Små kvalitetsforskjeller som

resultatene viser, kan det ikke utelukkes at AI-protokollen gir større utfordringer i diagnostikken for en uerfaren radiolog for små strukturer. Om kvalitetsforskjellene er så store at patologi kan gå tapt, gir ikke studiet svar på, trolig ikke.

Deep Resolve kan brukes i hensikt med å redusere opptakstiden. Det stilles stadig høyere krav til bildekvalitet innenfor MR i diagnostikken. Kanskje vil gevinsten med Deep Resolve være like stor for å øke bildekvaliteten, der protokollen allerede er presset ned i tid. Trolig vil teknologien kunne benyttes med ulike hensikter. For å se på anvendelsen av DR til å øke bildekvaliteten til protokoller med lik opptakstid, ble det ikke funnet noen studier i databasene Oria, Scopus og Google Scholar. Det kan tyde på at det er behov for mer forskning for å kunne gi svar på denne hypotesen.

#### 6.5.1 Fordeler og ulemper med bruk av AI protokollen sett i klinisk bruk

Trolig vil AI-protokollen ha flere positive egenskaper enn det som blir sett på i dette studiet. Redusert opptakstid minimerer sannsynligheten for bevegelsesartefakter i bildene. Kortere opptakstid vil også øke pasientkomforten med å slippe å ligge så lenge i magnettomografen, spesielt for pasienter med klaustrofobi vil dette være positivt.

Rekonstruksjonstiden til Deep Resolve etter at bildeopptaket er ferdig påvirker arbeidsflyten. Det tar litt tid før bildene kommer opp synlig på skjermen etter bildeopptaket er ferdig, dette kan ta opptil 15-20 sek. Tiden øker med økt oppløsning og bildemateriale som skal rekonstrueres, sees også kanskje for å ta lengre tid til fettsupprimerte sekvenser. Med en kortere rekonstruksjonstid hadde radiografen fått bildene raskere tilgjengelig.

#### 6.5.2 Begrensninger

For å kunne trekke konklusjoner og gi generaliserbare resultater er det behov for vurdering av et flere antall radiologer. En bedre introduksjon til radiologene i forkant av studiet vil nok være nyttig for generaliserbare resultater.

For å kunne trekke noen generelle konklusjoner om ytelsen til Deep Resolve, er det behov for flere studier der det sees på flere undersøkelsestyper, MR-system og spoler.

Andre parametervalg i optimaliseringen av AI-protokollen ville det trolig ha gitt andre resultater. I en optimalisering av AI- protokollen der antall avlesninger i frekvensretning hadde blitt beholdt fra referanseprotokollen, hadde det gitt en mindre reduksjon i romlig oppløsning og gjenoppretting for DR Sharp, dette ville potensielt gitt økt skarphet i bildene. Det ville heller ikke ha påvirket opptakstiden, men gitt et mindre pikselvolum og økt SNR i bildene som skulle

blitt gjenopprettet av DR Gain. Resultatene viser et godt signal-støy-forhold (støy) i AI-protokollen, kanskje ville DR Gain ha greid å gjenopprette denne økte støyvariasjonen. Årsaken til at det ble redusert antall avlesninger både i frekvens og faseretning i optimaliseringen, var for å se på om ytelsen til DR Sharp var til stede for å gjenopprette en undersampling både i frekvens og faseretning, ikke bare i faseretning.

I dette studiet ble det ikke sett på om det var den enkelte sekvens, som kunne ha påvirket vurderingen av protokollene. PDFS ble utført i tre ulike plan, T1-vektet sekvens var det gjort andre parameterendringer for reduksjon av opptakstid enn PDFS-vektede sekvenser. Det kan heller ikke utelukkes at Deep Resolve har en ulik ytelse på T1-vektede sekvenser sammenlignet med PDFS-vektede sekvenser. En metodetriangulering med en kvalitativ vurdering av sekvensene, ville kanskje gitt svar på om det var en enkelt sekvens som skilte seg ut i vurderingen av bildekvaliteten til protokollene.

Det kom heller ikke helt klart frem i dette studiet om standardprotokollen kan byttes ut med AI-protokollen i klinikken. Flere metoder for å vurdere bildekvaliteten enn subjektiv vurdering, ville økt troverdigheten til resultatene, trolig ville en annen studieutforming gitt svar på spørsmål som ikke kan besvares ut ifra dette studiet.

Det utvikles stadig nye DL-modeller og utviklingen innenfor AI går raskt. Siemens har allerede kommet med en ny versjon av Deep Resolve, Deep Resolve Boost. For å evaluere utviklingen ville en gjentakelse av dettes studiet med bruk av Deep Resolve Boost, kanskje gitt svar på om teknologien har kommet et steg videre i utviklingen.

## 7 Konklusjon

Resultatene antyder at standardprotokollen har en bedre bildekvalitet enn AI-protokollen. Oppsummerende mål i resultatene for sentraltendens, median og typetall viste seg å være lik for begge protokollene, med samme datagrunnlag for analysene. Det er ingen store kvalitetsforskjeller mellom protokollene. Deep Resolve viser å ha et stort potensial for å kunne brukes som akselerasjonsteknikk innenfor MR avbildning.

Kvalitetsforskjellen viste seg å være innenfor skarphet, der viste resultatene fra 3 radiologer at standardprotokollen hadde en signifikant bedre skarphet. Resultatene fra radiologene viste ikke enighet innenfor kategorien, en radiolog vurderte protokollene til å ikke ha noen signifikant forskjell i skarphet.

Innenfor kategorien støy, antyder AI-protokollen seg å være best, i en samlet vurdering fra alle radiologene.

Referanseprotokollen hadde en opptakstid på 7.00 min, der tiden ble redusert 35 %. Med en annen referanseprotokoll og andre parametervalg i optimaliseringen av AI protokollen, kunne man trolig fått andre resultater.

Med liten grad av inter-leser-enighet, med negative kappa-verdier for begge protokollene innen alle kategoriene, har ikke resultatene god nok pålitelighet til å være generaliserbare eller konklusive. Resultatene må sees på for kun dette utvalget av radiologer, et annet utvalg av radiologer ville trolig gitt andre resultater. Et av funnene i studiet var at subjektiv vurdering som mål for bildekvalitet kan gi stor variasjon i resultatene, og resultatene blir i stor grad påvirket av leserne.

I hvor stor grad kvalitetsforskjellene innenfor ulike kategorier påvirker diagnostikken gir ikke dette studiet svar på. Trolig vil ikke overgangen til AI-protokollen ha noen stor betydning for diagnostikken, men det er fortsatt behov for flere studier i klinikken.

Ser man på prinsipper MR og AI bygger på, i sammenheng med andre studier sammenlignet med dette, kan det antas at referanseprotokoll har stor betydning for i hvor stor grad Deep Resolve kan benyttes som akselerasjonsteknikk. Det kan tenkes at Deep Resolve har et like stort potensial for å kunne benyttes til å øke bildekvaliteten med samme opptakstid, til protokoller der tiden allerede er betydelig redusert.

## Referanseliste

1. Høltedahl R. Atraumatiske kneplager. Tidsskrift for Den norske legeforening. 2018.
2. Shapiro L, Staroswiecki E, Gold G, editors. MRI of the knee: optimizing 3T imaging. Seminars in roentgenology; 2010: NIH Public Access.
3. Stiris MG. Magnetisk resonanstomografi ved skjelett-og bløtdelstraumer. Tidsskrift for Den norske legeforening. 2000.
4. Sub-committee ESoSRS. Guidelines for MR imaging of Sports Injuries MRI protocols for ESSR: European Society of Skeletal Radiology Sports Sub-committee; 2016 [updated 191022]. Available from: <https://essr.org/content-essr/uploads/2016/10/ESSR-MRI-Protocols-Knee.pdf>.
5. Gjesdal K-I, Smith H-J. Magnettomografi—nyere bildeteknikker og fremtidsperspektiver. Tidsskrift for Den norske legeforening. 2000.
6. McRobbie DW, Moore EA, Graves MJ, Prince MR. MRI from Picture to Proton. 2 ed. Cambridge: Cambridge: Cambridge University Press; 2006.
7. Hamilton J, Franson D, Seiberlich N. Recent advances in parallel imaging for MRI. Progress in nuclear magnetic resonance spectroscopy. 2017;101:71-95.
8. AD E. Parallell bildebehandling Question and answers in MRI 2021 [updated 12.11.22]. Available from: <https://mriquestions.com/parallel-imaging-pi.html>.
9. Geethanath S, Reddy R, Konar AS, Imam S, Sundaresan R, DR RB, et al. Compressed sensing MRI: a review. Critical Reviews™ in Biomedical Engineering. 2013;41(3).
10. Chen Y, Schönlieb C-B, Liò P, Leiner T, Dragotti PL, Wang G, et al. AI-based reconstruction for fast MRI—a systematic review and meta-analysis. Proceedings of the IEEE. 2022;110(2):224-45.
11. Recht MP, Zbontar J, Sodickson DK, Knoll F, Yakubova N, Sriram A, et al. Using Deep Learning to Accelerate Knee MRI at 3 T: Results of an Interchangeability Study. AJR Am J Roentgenol. 2020;215(6):1421-9.
12. Sharma D, Kumar N. A review on machine learning algorithms, tasks and applications. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). 2017;6(10):2278-1323.
13. Chauhan R, Ghanshala KK, Joshi R, editors. Convolutional neural network (CNN) for image detection and recognition. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC); 2018: IEEE.
14. Tanenbaum LN. Artificial intelligence and medical imaging: image acquisition and reconstruction. Applied Radiology. 2020;49(3):34-5.
15. Healthcare-in-europe.com. Siste nytt diagnostisk bildediagnostikk 2022 [updated 10.10.2022]. Available from: <https://healthcare-in-europe.com/en/news/philips-ecr-ai-mri-imaging-speed-staff-burden.html>.
16. Pezzotti N, Yousefi S, Elmahdy MS, Van Gemert JHF, Schuelke C, Doneva M, et al. An adaptive intelligence algorithm for undersampled knee MRI reconstruction. IEEE Access. 2020;8:204825-38.
17. Chaudhari AS, Fang Z, Kogan F, Wood J, Stevens KJ, Gibbons EK, et al. Super-resolution musculoskeletal MRI using deep learning. Magnetic resonance in medicine. 2018;80(5):2139-54.
18. Herrmann J, Keller G, Gassenmaier S, Nickel D, Koerzdoerfer G, Mostapha M, et al. Feasibility of an accelerated 2D-multi-contrast knee MRI protocol using deep-learning image reconstruction: a prospective intraindividual comparison with a standard MRI protocol. European Radiology. 2022;1-15
19. Healthineers S. Deep Resolve. Erlangen: Clinical Application Job Aid. ; 2021.
20. Mork R. Forsinkelser ved utredning og behandling av kreft i ventrikkel, pancreas, colon og rectum: Universitetet i Tromsø; 2013.
21. Frihagen F, Kvakestad R, Melhuus K, Engebretsen L. Bruk av MR-undersøkelse ved kneskader. Tidsskrift for Den norske legeforening. 2003.
22. Lyng S. Retur til idrett og aktivitet minimum ett år etter fremre korsbåndskade hos barn 2021.



23. Tran HT. Radiologisk utredning av primære hjernesvulster, nytte av nyere MR-teknikker 2012.
24. Smith H-J, Gjesdal K-I. Magnetisk resonans–historikk og teoretisk grunnlag. Tidsskrift for den norske legeforening. 2000.
25. Westbrook C, Talbot J. MRI in Practice. 5th ed. Newark: Newark: Wiley; 2018.
26. AD E. Frequency Encoding Question and answers in MRI 2021 [updated 28.10.22. Available from: <https://mriquestions.com/frequency-encoding.html>.
27. AD E. Phase-encoding Gradients: Questions and answers in MRI 2021 [updated 28.10.22. Available from: <https://mriquestions.com/phase-encoding-gradient.html>.
28. AD E. Spatial Encoding Question and answer in MRI 2021 [updated 28.10.22. Available from: <https://mriquestions.com/how-to-locate-signals.html>.
29. AD E. Filling k-space Question and answers in MRI 2021 [updated 29.10.22. Available from: <https://mriquestions.com/data-for-k-space.html>.
30. Ashton E. Quantitative MR in multi-center clinical trials. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2010;31(2):279-88.
31. Jeelani H, Martin J, Vasquez F, Salerno M, Weller DS, editors. Image quality affects deep learning reconstruction of MRI. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 2018: IEEE.
32. MR-Technology. MR-tip, singal til støyforhold: MR-Tip; 2018 [updated 19.10.22. Available from: <https://www.mr-tip.com/serv1.php?type=db1&db=Signal%20to%20Noise%20Ratio>.
33. Unilabs. Siemens Altea. 2022.
34. Dick S. Artificial intelligence. 2019.
35. Devin. Forstå grunnleggende hierarkiet for kunstig intelligens. [Forum]. Medium.com 2020 [updated 29.10.22. Available from: [https://medium.com/@devin\\_chandula/understanding-the-ai-basic-hierarchy-b9d7438bff91](https://medium.com/@devin_chandula/understanding-the-ai-basic-hierarchy-b9d7438bff91).
36. Healtineers S. Technologies and innovations: Siemens Healtineers; 2021 [updated 06.08.2022. Available from: <https://www.siemens-healthineers.com/no/magnetic-resonance-imaging/technologies-and-innovations/deep-resolve>.
37. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electronic Markets. 2021;31(3):685-95.
38. Bayr U. Maskinl ring til bildeanalyse. 2020.
39. Liang D, Cheng J, Ke Z, Ying L. Deep MRI reconstruction: unrolled optimization algorithms meet neural networks. arXiv preprint arXiv:1907.11711. 2019.
40. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, et al. AI applications to medical images: From machine learning to deep learning. Physica Medica. 2021;83:9-24.
41. ProjectPRO. Topp 10 dypl ringsalgoritmer i maskinl ring [2022] 2022 [updated 29.10.22. Available from: <https://www.projectpro.io/article/deep-learning-algorithms>.
42. Van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging“how-to” guide and critical reflection. Insights into imaging. 2020;11(1):1-16.
43. Reinke A, Eisenmann M, Tizabi MD, Sudre CH, R dsch T, Antonelli M, et al. Common limitations of image processing metrics: A picture story. arXiv preprint arXiv:2104.05642. 2021.
44. Behl N. Mobilizing the power of networks. Siemens Healtineers; 2021.
45. Gold GE, Hargreaves BA, Beaulieu CF. Protocols in sports magnetic resonance imaging. Topics in Magnetic Resonance Imaging. 2003;14(1):3-23.
46. Reimer P, Parizel P, Stichnoth F. Clinical MR Imaging. 2nd Editio. Springer; 2006.
47. Mason A, Rioux J, Clarke SE, Costa A, Schmidt M, Keough V, et al. Comparison of objective image quality metrics to expert radiologists’ scoring of diagnostic quality of MR images. IEEE transactions on medical imaging. 2019;39(4):1064-72.
48. Joshi A, Kale S, Chandel S, Pal DK. Likert scale: Explored and explained. British journal of applied science & technology. 2015;7(4):396.

49. Polit DF, Beck CT. Generalization in quantitative and qualitative research: Myths and strategies. *International journal of nursing studies*. 2010;47(11):1451-8.
50. Oyeka ICA, Ebuh GU. Modified Wilcoxon signed-rank test. *Open Journal of Statistics*. 2012;2(2):172-6.
51. Xia Y. Wilcoxon signerte rangeringstest ScienceDirect: Board Review in Preventive Medicine and Public Health; 2017 [Available from: <https://www.sciencedirect.com/topics/medicine-and-dentistry/wilcoxon-signed-ranks-test>].
52. Lydersen S. Samsvar mellom observatører–eller interobservatørsamsvar? *Tidsskrift for Den norske legeförening*. 2020.
53. Jalal A. Fleiss Kappa rbiostatistics.com: Arif Jalal; 2022 [updated 15.11.2022. Available from: <http://rbiostatistics.com/node/67>].
54. Lydersen S. Intraklassekorrelasjon. *Tidsskrift for Den norske legeförening*. 2022.
55. Lovdata. Lov om medisinsk og helsefaglig forskning (helseforskningsloven): Lovdata.no; 2020 [updated 23.10.22. Available from: [https://lovdata.no/dokument/NL/lov/2008-06-20-44/KAPITTEL\\_1#%C2%A73](https://lovdata.no/dokument/NL/lov/2008-06-20-44/KAPITTEL_1#%C2%A73)].
56. Lydersen S. Cohens kappa–et mål på samsvar mellom observatører. *Tidsskrift for Den norske legeförening*. 2018.
57. Johre DL. Bakre meniskrotskade i knær med ACL-ruptur–Prevalens og MR-tolkning 2017.

# Vedlegg.

## Vedlegg A.

### Radiolog 1

Støy		Skarphet		Menisk		Korsbånd		Ligamenter		Benmarg		Leddbrusk		Ana.S	
Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI
3	5	4	2	4	3	5	4	4	5	5	4	5	4	23	20
3	5	4	2	4	3	5	4	4	5	5	4	5	4	22	21
3	5	4	3	4	3	4	5	4	4	4	5	5	4	23	20
3	5	4	2	4	3	5	4	4	3	4	5	4	5	21	21
3	5	4	2	4	3	5	4	4	4	5	4	5	4	21	20
3	5	4	3	5	4	5	4	5	4	4	4	5	4	23	19
3	5	4	3	5	4	4	4	5	4	5	4	4	4	24	20
3	5	4	3	5	4	4	4	4	5	4	4	4	5	23	20
4	5	4	3	4	4	5	4	5	5	4	4	4	4	21	22
5	4	4	3	4	5	4	4	4	5	5	4	4	5	22	21
5	4	5	4	4	4	5	5	4	5	5	4	4	4	21	23
4	5	5	3	4	4	5	5	5	5	5	4	4	4	22	22
4	4	4	3	5	5	5	5	4	5	4	5	4	5	23	22
4	5	5	4	5	4	5	5	5	4	5	5	4	5	22	25
4	5	5	4	4	4	5	5	4	4	4	4	4	3	24	23
4	4	5	4	4	5	5	5	4	5	5	5	4	4	21	20
5	4	4	5	5	4	5	5	5	4	5	5	4	4	22	24
4	5	5	4	5	4	5	5	5	5	5	5	5	4	24	22
4	5	5	4	4	5	5	5	5	5	4	5	4	4	25	23
4	5	5	4	5	5	5	5	5	5	5	4	5	5	22	24
4	5	5	4	5	5	5	4	5	5	5	5	4	5	25	24
4	3	5	4	5	5	5	5	5	5	5	5	5	4	24	24
4	4	4	5	5	5	5	5	5	4	5	5	5	5	25	24
4	5	5	4	5	5	4	4	4	5	5	5	5	4	25	24
3	4	4	3	5	4	5	4	5	5	5	5	5	4	23	23
4	4	5	4	5	5	5	5	4	5	5	5	5	4	25	22
3	4	5	3	5	4	5	4	5	5	5	5	5	4	24	24
4	4	5	4	5	5	5	5	5	5	5	5	4	5	25	22
3	4	5	4	5	4	5	5	5	5	5	5	5	4	24	25
3	4	5	4	4	5	5	5	5	5	5	5	5	4	25	23
4	4	5	4	5	4	5	5	5	5	5	4	5	4	24	24
3	4	5	4	4	4	5	5	5	5	4	4	5	5	25	22
4	4	5	4	4	5	5	5	5	5	5	5	5	5	23	23
4	4	5	4	5	5	5	5	5	5	5	5	5	5	24	25
4	4	5	4	4	4	5	5	5	5	5	5	4	4	25	25
4	4	5	4	5	4	5	5	5	5	5	5	5	4	23	23
3	3	5	4	4	5	5	5	4	5	5	5	5	4	25	23
3	4	5	4	5	5	5	5	4	5	5	5	5	5	23	24

3	4	5	4	5	5	5	5	5	5	5	5	5	4	24	25
4	5	5	4	5	5	5	5	5	5	5	5	5	4	25	24
4	4	5	4	5	5	5	5	5	5	5	5	5	5	25	24
5	4	5	4	5	5	5	5	5	5	5	5	5	5	25	25
5	5	5	4	5	4	5	4	5	5	4	5	5	4	25	25
3	4	5	4	5	5	5	4	5	5	5	4	5	4	24	22
3	4	5	4	5	5	5	5	5	5	5	5	5	5	25	22
3	3	4	3	4	5	5	5	4	5	5	5	5	4	25	25
4	3	5	4	4	5	5	5	5	5	5	5	5	5	23	24
3	4	5	4	5	5	5	5	5	5	5	5	5	4	24	25
5	4	5	4	4	5	5	5	5	5	5	5	5	5	25	24
3	4	5	4	5	5	5	4	5	5	4	5	5	5	24	25
4	5	5	4	5	5	5	5	5	5	4	5	5	5	24	24
3	3	4	3	5	4	5	5	4	5	5	5	5	4	24	25
3	4	5	4	5	5	5	5	5	4	4	5	5	5	24	23
4	5	5	4	5	5	5	4	5	5	5	5	5	4	24	24
3	4	5	4	5	4	5	5	5	4	4	5	5	4	25	23
4	5	5	4	5	5	5	5	4	5	5	5	5	4	24	22
4	5	5	4	5	5	4	4	5	5	5	5	4	5	24	24
4	5	5	4	5	5	5	5	5	5	5	5	5	5	23	24
4	5	5	4	5	5	5	5	5	5	5	5	5	5	25	25
4	5	5	4	5	5	5	5	5	5	5	5	5	5	25	25
4	5	5	4	5	4	5	5	5	5	5	5	5	5	25	25
3	4	5	4	5	4	5	5	4	5	5	5	5	5	25	24
4	5	5	4	5	5	5	5	5	5	4	5	5	4	24	24
4	5	5	4	5	5	5	5	5	5	4	5	5	5	24	24

## Radiolog 2.

Støy		Skarphet		Menisk		Korsbånd		Ligamenter		Benmarg		Leddbrusk		Anatom.S	
Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI
3	3	3	3	3	2	3	3	3	2	3	3	3	3	15	13
3	2	3	2	3	2	3	2	3	2	3	3	2	3	14	12
2	3	2	3	3	3	3	2	3	3	3	3	2	3	14	14
3	3	3	3	3	3	3	3	3	3	2	3	2	3	13	15
2	3	2	3	2	3	3	2	3	2	2	3	3	2	13	12
3	3	3	3	3	3	3	2	3	2	3	3	3	2	15	12
3	2	3	2	3	3	3	2	3	2	3	2	3	2	15	11
2	3	2	3	2	3	2	3	3	3	2	3	2	2	11	14
3	3	3	3	3	3	3	2	3	3	3	3	3	3	15	14
3	2	2	3	3	2	2	3	3	3	3	2	2	2	13	12
3	3	3	3	2	2	2	2	2	2	2	2	3	3	11	11
2	2	4	3	2	2	2	2	3	2	3	3	3	2	13	11
3	4	3	3	3	3	3	3	3	3	3	3	2	2	14	14
3	3	3	3	3	3	2	3	2	2	3	3	2	3	12	14
3	3	3	3	3	3	3	3	2	2	2	3	2	2	12	13
3	3	3	3	3	3	3	3	3	3	3	3	3	2	15	14
3	3	2	3	3	3	3	3	3	2	2	3	3	3	14	14
2	3	2	3	3	3	2	2	3	3	2	2	2	2	12	12
2	3	2	3	2	2	2	3	3	2	2	3	2	2	11	12
3	2	2	3	3	2	2	3	3	3	3	2	2	2	13	12
2	3	3	3	2	3	2	3	3	3	3	2	2	2	12	13
3	2	3	2	2	3	2	2	3	3	3	3	2	2	12	13
3	3	4	3	3	3	3	3	4	4	4	4	3	3	17	17
2	3	2	2	2	2	3	3	3	3	3	3	2	2	13	13
3	2	3	2	2	2	3	2	2	2	2	3	2	2	11	11
3	3	3	3	3	3	3	3	3	3	3	3	2	2	14	14
3	3	3	3	3	2	3	3	3	2	3	2	2	2	14	11
3	3	3	3	3	2	3	3	3	3	3	3	2	2	14	13
3	3	3	3	2	3	2	3	2	3	3	2	2	2	11	13
2	3	2	3	3	3	3	3	3	3	2	3	3	2	14	14
2	2	2	2	3	3	3	3	3	3	2	2	2	2	13	13
2	3	2	2	3	3	2	3	3	3	3	3	2	2	13	14
3	3	3	3	3	3	3	3	3	3	3	3	3	3	15	15
3	3	3	3	3	3	3	2	3	3	3	3	2	3	14	14
2	3	2	3	2	2	3	3	2	3	2	3	2	2	11	13
2	2	2	3	3	3	2	3	2	2	2	2	3	3	12	13
2	3	2	2	2	2	3	3	3	3	3	3	3	3	14	14
3	3	3	3	3	3	3	3	3	3	3	3	2	3	14	15
3	2	3	2	3	3	2	3	3	3	3	3	3	2	14	14
3	3	3	3	3	3	3	3	3	3	3	3	2	2	14	14
3	3	2	2	2	3	3	3	2	2	3	3	2	2	12	13
2	3	2	3	3	2	3	3	3	3	2	3	3	2	14	13
3	3	3	3	3	3	3	2	2	3	3	3	2	2	13	13

3	3	3	2	3	2	2	3	2	3	3	2	2	12	12
2	3	2	3	2	2	3	2	2	3	2	2	2	11	11
2	3	2	3	3	3	3	3	3	3	3	3	2	14	14
2	2	3	2	3	2	3	3	3	3	3	3	2	14	13
2	3	2	3	2	3	3	3	3	3	3	2	2	13	13
2	3	3	2	2	2	3	3	3	2	2	2	2	12	11
3	3	3	3	2	3	3	3	2	3	3	3	2	12	14
3	3	3	3	3	3	3	3	3	3	3	2	3	15	14
3	3	3	2	3	3	3	2	2	3	3	2	2	13	12
3	3	3	2	3	2	3	3	3	3	3	2	3	15	13
3	3	3	2	3	2	3	3	3	2	3	2	3	15	11
2	3	2	3	2	2	3	3	3	3	2	3	2	12	13
3	3	3	3	2	3	3	3	2	3	2	3	2	13	14
3	3	3	3	3	3	3	3	3	3	3	3	2	13	14
3	3	3	3	2	3	2	3	2	3	3	3	2	13	14
2	3	2	3	2	3	3	3	3	3	3	3	2	13	14
3	3	3	3	3	3	3	3	3	3	3	3	3	13	14
3	3	3	3	3	2	3	3	3	3	3	3	2	13	13
3	3	3	3	3	3	3	3	3	3	3	3	3	13	14
3	3	3	3	2	3	3	3	3	3	3	2	3	13	13
3	3	2	2	3	3	3	2	2	3	3	3	2	13	13

# Radiolog 3

Støy		Skarphet		Menisk		Korsbånd		Ligamenter		Benmarg		Leddbrusk		Anat.S	
Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI
5	4	5	3	4	3	4	3	5	4	5	4	4	3	22	17
5	3	5	4	4	3	4	3	5	4	5	4	4	3	22	17
4	5	4	3	4	4	3	4	4	5	4	5	4	4	19	22
5	4	5	4	5	4	4	3	5	4	5	4	4	3	23	18
5	4	4	3	4	3	3	3	4	3	5	4	4	3	20	16
5	4	4	3	4	3	4	4	4	4	4	4	3	4	20	18
5	5	3	4	3	4	4	4	4	4	4	4	5	3	4	18
5	4	5	4	5	4	4	4	4	4	5	5	4	5	22	22
5	4	4	3	4	4	4	5	4	3	5	4	4	3	21	19
5	4	5	4	5	4	5	4	4	4	5	4	4	3	23	19
4	3	4	4	4	3	4	4	4	4	5	4	4	3	21	18
5	4	5	4	4	4	5	5	4	3	5	4	4	4	22	20
5	5	5	5	4	4	4	3	4	4	5	5	4	3	21	19
5	4	5	4	5	4	4	4	5	4	5	4	4	3	23	19
3	4	3	4	5	4	4	3	4	3	4	3	5	4	22	17
3	4	3	4	4	5	4	5	4	5	4	5	4	5	20	25
4	5	3	4	3	4	3	4	4	4	4	5	4	4	18	21
5	4	4	3	5	4	4	3	4	4	5	4	4	4	22	19
4	3	4	3	4	3	4	3	4	3	4	3	4	3	20	15
4	3	4	4	5	4	4	4	5	5	5	4	5	4	24	21
5	4	5	4	5	4	4	4	4	3	5	4	4	4	22	19
5	4	5	4	5	4	5	4	5	4	5	4	5	4	25	20
5	4	5	4	4	4	4	3	5	4	5	4	5	4	23	19
4	4	5	4	5	4	4	4	4	4	5	4	4	4	22	20
5	4	5	4	5	4	4	3	5	4	5	4	4	4	23	19
4	4	4	5	4	5	4	4	4	5	4	5	4	5	20	24
5	4	5	4	5	4	5	4	4	4	5	4	5	4	24	20
5	4	5	4	4	3	3	4	4	5	5	4	4	5	20	21
5	4	5	4	5	4	5	4	5	4	5	4	5	4	25	20
5	4	5	4	5	5	5	4	4	4	5	4	5	4	24	21
5	4	5	4	5	4	4	4	5	5	5	4	4	3	23	20
5	4	5	4	5	4	5	4	5	5	5	4	4	3	24	20
5	5	5	4	5	5	4	3	4	4	5	4	5	4	23	20
5	5	5	4	5	5	4	4	4	5	5	4	5	5	23	23
4	4	4	5	4	3	4	4	4	4	5	4	3	4	20	19
5	4	5	4	4	4	4	4	5	4	5	4	4	3	22	19
4	4	5	4	5	4	5	4	5	4	5	4	5	4	25	20
5	5	5	4	5	4	4	4	4	4	5	4	4	4	22	20
5	4	5	4	5	4	5	4	4	4	5	4	5	4	24	20
5	4	5	4	5	4	5	4	5	4	5	4	5	4	25	20
5	4	5	4	4	4	5	4	5	5	5	4	5	4	24	21
5	4	5	4	5	4	4	4	4	4	5	4	3	4	21	20
5	4	5	4	3	4	4	4	4	5	5	4	4	3	20	20

4	5	4	5	4	4	4	4	4	4	5	4	4	3	21	19
4	4	5	4	5	4	5	4	4	4	5	4	5	4	24	20
5	4	5	4	5	4	4	4	4	4	5	4	5	4	23	20
4	5	4	5	4	5	4	5	4	5	5	4	4	3	21	22
4	4	5	4	4	3	4	4	4	5	5	4	3	4	20	20
5	4	5	4	4	3	4	4	5	4	5	4	4	4	22	19
5	4	5	4	5	4	5	4	5	4	5	4	5	4	25	20
5	4	5	4	5	4	4	3	5	4	5	4	5	4	24	19
5	4	5	4	4	4	4	5	4	4	5	4	4	3	21	20
5	4	5	5	5	4	4	3	5	5	5	4	5	4	24	20
5	4	5	4	5	5	5	4	5	5	5	4	4	5	24	23
5	4	5	4	4	4	5	4	5	5	5	4	4	4	23	21
4	4	5	4	4	3	3	4	5	4	5	4	5	4	22	19
5	4	4	3	5	4	4	3	5	4	4	5	5	4	23	20
5	4	5	4	5	4	5	4	5	4	5	4	5	4	25	20
5	4	5	4	5	4	5	4	5	4	5	4	4	3	24	19
5	4	5	4	4	3	4	4	4	4	5	4	4	4	21	19
5	4	4	5	4	3	4	4	4	3	5	4	5	4	22	18
5	4	5	4	5	4	4	3	4	4	5	4	5	4	23	19
4	5	5	4	5	4	5	4	5	5	5	4	5	4	25	21
5	4	5	4	5	4	5	4	5	5	5	4	5	4	25	21



# Radiolog 4

Støy		Skarphet		Menisk		Korsbånd		Ligamenter		Benmarg		Leddbrusk		Anatomiske strukturer	
Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI	Standard	AI
2	3	3	2	2	2	3	3	3	3	3	2	3	2	15	12
2	3	3	2	3	2	2	2	3	3	3	2	3	2	14	11
2	3	4	2	4	3	4	3	4	3	4	3	4	2	20	14
3	2	3	3	3	3	3	3	3	3	3	2	3	3	15	14
2	3	3	3	3	2	3	3	3	3	3	3	2	2	14	13
3	2	3	2	3	2	3	3	3	3	3	2	3	2	15	12
2	3	3	3	1	2	2	2	3	3	3	2	3	2	12	11
3	3	3	2	4	3	3	3	3	3	3	3	3	3	16	15
3	2	3	3	3	3	2	2	3	3	3	2	2	3	13	13
4	3	4	3	4	3	3	3	4	3	4	3	3	3	18	15
3	2	3	2	4	3	3	3	3	3	3	2	3	3	16	14
3	2	2	2	2	2	3	3	3	3	3	2	3	3	14	13
2	1	2	2	2	2	2	2	3	3	3	2	2	2	12	11
3	2	3	2	3	2	3	3	3	3	3	3	3	3	15	14
3	2	3	2	3	2	3	2	3	2	3	2	2	2	14	10
3	2	3	2	3	3	2	2	3	3	3	2	3	3	14	13
3	2	3	2	3	2	3	3	3	3	3	2	3	3	15	13
3	2	2	2	3	2	3	2	3	3	3	2	3	3	15	12
3	2	2	2	3	2	2	2	3	3	3	2	2	2	13	11
3	2	3	2	3	3	2	2	3	3	3	2	3	2	14	12
3	2	3	2	3	3	2	2	3	3	3	2	3	2	14	12
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
2	3	2	3	2	2	2	1	2	2	3	2	3	2	12	9
3	2	2	2	2	2	2	2	3	3	3	2	3	3	13	12
3	2	3	3	3	3	3	2	3	3	3	2	2	2	14	12
2	3	3	2	2	2	2	2	3	3	3	2	2	2	12	11
2	3	3	2	3	2	3	2	3	3	3	2	3	2	15	11
3	2	3	2	3	2	3	2	3	3	3	2	2	2	14	11
3	2	3	3	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	3	3	3	3	3	3	3	3	2	3	3	15	14
3	3	3	2	3	2	3	3	3	3	3	2	3	3	15	13
3	3	3	2	3	2	3	3	3	3	4	3	3	3	16	14
4	3	4	3	4	3	3	3	4	4	3	3	3	3	17	16
3	3	2	3	3	3	3	3	3	3	3	2	3	2	15	13
3	2	3	2	3	2	2	3	3	3	3	2	1	1	12	11
3	2	3	2	2	2	3	2	3	3	3	2	2	2	13	11
3	2	3	2	3	2	3	2	2	3	3	2	2	3	13	12
3	3	3	2	3	2	2	2	3	3	3	3	3	3	14	13
3	2	3	2	2	3	3	2	3	3	3	2	3	2	14	12
3	3	3	3	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14

3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	2	3	3	3	2	3	3	3	3	2	3	3	14	14
3	2	3	2	3	3	2	3	3	3	3	2	3	3	14	14
3	2	2	3	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	2	3	3	3	2	3	3	15	13
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	2	3	3	3	2	3	3	15	13
3	1	3	2	3	3	3	3	3	3	3	2	2	2	14	13
3	1	3	2	3	3	3	3	3	3	3	2	2	2	14	13
3	2	3	2	3	3	3	3	3	3	3	2	3	2	15	13
3	2	3	2	3	3	3	2	3	3	3	2	3	2	15	12
3	1	3	1	3	3	3	2	3	3	3	2	3	2	15	12
3	2	3	2	3	3	3	2	3	3	3	2	3	3	15	13
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	3	3	3	3	2	3	2	15	13
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14
3	2	3	2	3	3	3	3	3	3	3	2	3	2	15	13
3	2	3	2	3	3	3	3	3	3	3	2	3	3	15	14

## Vedlegg B.

### Vil du delta i et forskningsprosjekt?

#### Bildeoptimalisering av MR kne, ved bruk av AI basert bilderekonstruksjons teknologi.

Dette er en forespørsel om å delta i et forskningsprosjekt hvor formålet er å sammenligne to MR-kneprotokoller, en uten bruk av kunstig intelligens (AI), og en med. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

#### **Formål**

Forskningsprosjektet skal vurdere om bildekvaliteten på to ulike kne protokollene er av samme kvalitet, eller om den ene er bedre enn den andre. Protokollen med AI teknologi har en redusert undersøkelsestid med ca 35%. Kortere undersøkelsestider kan bidra til kortere ventetider for å få en MR- undersøkelse.

Dette er et forskningsprosjekt som er inkludert i et masterstudie. Resultatene vil bli publisert i en vitenskapelig artikkel. Det vil ikke være sporbart i artikkelen hvilke pasienter som har vært inkludert i studiet.

#### **Hvem er ansvarlig for forskningsprosjektet?**

NTNU er ansvarlig for forskningsprosjektet. Undersøkelsene vil bli utført hos Unilabs. AI-basert programvare Deep Resolve, som benyttes for å redusere undersøkelsestiden, er levert av Siemens.

#### **Hvorfor får du spørsmål om å delta?**

Du er henvist til MR kne hos Unilabs.

Kriterier for å delta er alder 18-90 år, ingen kontraindikasjon for MR. Det vil bli inkludert 60-80 pasienter, cirka halvparten av hvert kjønn.

Hva innebærer det for deg å delta?

Hvis du sier ja til å delta vil det bli utført 2 standardiserte kneprotokoller når du er til undersøkelse hos oss, begge blir utført ved samme besøket. Den totale undersøkelsestiden i MR-maskinen med begge kneprotokollene vil være ca 10 min.

I forkant av prosjektet er det kvalitetssikret at protokollene er gode nok for å gi en god klinisk vurdering, og den ene protokollen er vår standard kneprotokoll.

Deltagelse i studien vil ikke gi lengre svartid på undersøkelsen eller føre til andre ulemper.

#### **Det er frivillig å delta**

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du likevel når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger som har med studien å gjøre vil da bli slettet. Den planlagte MR-undersøkelsen som du allerede er henvist

til vil ikke bli endret på noe vis. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Deltagelsen vil ikke påvirke diagnostikken og beskrivelsen din.

Dobbelt bildesett vil ikke påføre deg noen ulempe eller skade, men du vil ligge litt lengre i MR-maskinen enn vanlig (totalt cirka 10 minutter). MR påfører deg ingen røntgenstråling.

Formålet er kun å se på bildekvalitet. Din MR av kneet vil bli vurdert på vanlig måte av våre radiologer, men studiedelen av undersøkelsen påvirker ikke diagnose eller eventuell videre behandling.

### **Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger**

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

Bildene blir oppbevart i datasystemet til Unilabs, Sectra PACS/RIS.

Personopplysninger på undersøkelsesdagen erstattes med en kode for videre analyse av undersøkelsen i studiet. Denne koden lagres på egen kryptert navneliste adskilt fra øvrige data. Det vil ikke være mulig for deltagerne å gjenkjenne resultater fra egen undersøkelse i den artikkelen som skal publiseres.

### **Hva skjer med personopplysningene dine når forskningsprosjektet avsluttes?**

Beskrivelsen fra radiologen blir liggende på undersøkelsen, som vanlig. Det er kun ansatte med brukernavn og passord til Unilabs' datasystem, som har tilgang til bildene, som vanlig. Bildene blir kun liggende på i din journal i datasystemet, som vanlig.

Personopplysningene blir ikke brukt videre i prosjektet etter undersøkelsesdagen, det vil derfor ikke være nødvendig å gjøre ytterligere tiltak etter at prosjektet er avsluttet.

### **Hva gir oss rett til å behandle personopplysninger om deg?**

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra NTNU og Unilabs har Personverntjenester vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

### **Dine rettigheter**

Så lenge du kan identifiseres i datamaterialet, har du rett til:

innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene å få rettet opplysninger om deg som er feil eller misvisende

å få slettet personopplysninger om deg

å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

Unilabs ved Anne Helene Reiten. E-post [anne.helene.reiten@unilabs.com](mailto:anne.helene.reiten@unilabs.com) eller hennes veileder til prosjektet Rolf Svendsmark. E-post [rolf.svendsmark@siemens-healthineers.com](mailto:rolf.svendsmark@siemens-healthineers.com)

Vårt personvernombud: [sett inn navn og kontaktopplysninger til personvernombudet hos behandlingsansvarlig institusjon]

Hvis du har spørsmål knyttet til Personverntjenester sin vurdering av prosjektet, kan du ta kontakt med:

Personverntjenester på epost ([personverntjenester@sikt.no](mailto:personverntjenester@sikt.no)) eller på telefon: 53 21 15 00.

**Vi håper at du vil delta i dette studiet. Ønsker du ikke å delta, kontakt Unilabs kundeservice på**

**Tlf. 21 000 800. Da vil du få tildelt en ny time.**

Med vennlig hilsen

Prosjektansvarlig

Eventuelt student

(Forsker/veileder)

---

Jeg har mottatt og forstått informasjon om prosjektet [sett inn tittel], og har fått anledning til å stille spørsmål. Jeg samtykker til:

å delta i prosjektet bildeoptimalisering av MR kne, med bruk av AI-basert bilderekonstruksjonsteknologi.

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

---

(Signert av prosjektdeltaker, dato)



