



NO9700091

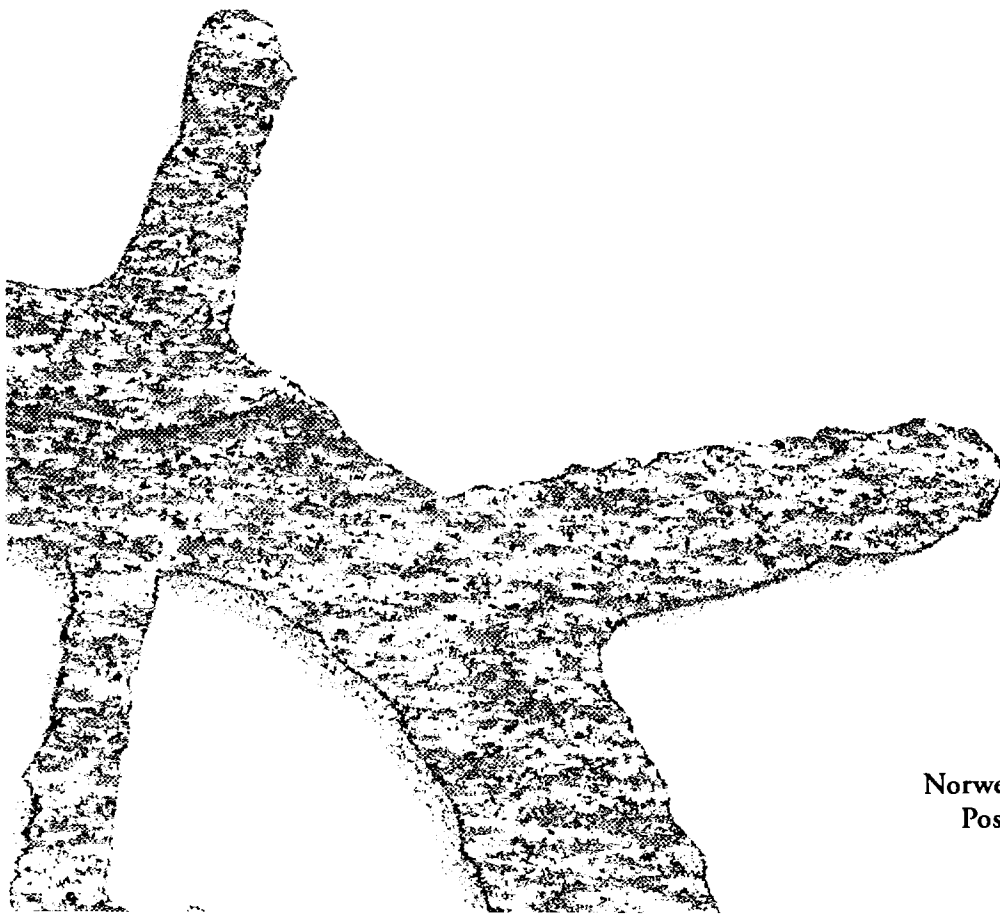
StrålevernRapport 1997:4

Mammografivirksomhet i Norge

Bildekvalitet og totalytelse

Utarbeidet ved Det Norske Radiumhospital

Jan B. Olsen - Arne Skretting - Anders Widmark



Statens
strålevern

Norwegian Radiation Protection Authority
Postboks 55 · N-1345 Østerås · Norway

Referanse:

Olsen JB, Skretting A, Widmark A. Mammografivirkosomhet i Norge: Bildekvalitet og totalytelse. StrålevernRapport 1997:4. Østerås: Statens strålevern, 1997.

Emneord:

Mammografi. Bildekvalitet. Fantomer. ROC.

Resymé:

Rapporten beskriver en metode for vurdering av totalytelse ved mammografi-virkosomhet basert på Receiver Operating Characteristic (ROC) analyse. Metoden har blitt brukt til å måle totalytelsen for samtlige 45 norske mammografi-laboratorier som var i drift i perioden desember 1993 til mars 1994. Samtidig er flere fantomer benyttet for å kartlegge bildekvaliteten til laboratoriene.

Reference:

Olsen JB, Skretting A, Widmark A. Mammography in Norway: Image quality and total performance. NRPA Report 1997:4. Østerås: Norwegian Radiation Protection Authority, 1997. Language: Norwegian.

Key words:

Mammography. Image quality. Phantoms. ROC.

Abstract:

This report describes a method for assessing the total performance in mammography based on Receiver Operating Characteristic (ROC) analysis. In the time period from December 1993 to March 1994, the method was applied to assess the total performance of all the 45 Norwegian mammography laboratories operative at that time. Image quality characteristics in each laboratory was established by use of well-known phantoms.

Prosjektleder: Jan B. Olsen.

Godkjent:



Gunnar Saxebøl, avdelingsdirektør, Avdeling Tilsyn.

Jan B. Olsen og Arne Skretting, Avdeling for medisinsk fysikk og teknikk, Det Norske Radiumhospital, Montebello, 0310 Oslo

Anders Widmark, Avdeling Tilsyn, Statens strålevern, Postboks 55, 1345 Østerås

73 sider. Utgitt 1997-04-04. Opplag 500.

Form, trykk: Falch Hurtigtrykk as, Oslo

Bestilles fra:

Statens strålevern, Postboks 55, 1345 Østerås.

Telefon 67 16 25 00, telefax 67 14 74 07

ISSN 0804-4910

Sammendrag

Statens strålevern foretok i samarbeid med en prosjektgruppe ved Det Norske Radiumhospital en gjennomgang av samtlige 45 norske mammografilaboratorier som var i drift i perioden desember 1993 til mars 1994. Del 1 av prosjektet var et ordinært tilsyn, med hovedvekt på apparaturens tekniske ytelse og dosemålinger. Resultatene fra denne delen av prosjektet finnes i StrålevernRapport 1995:5 [1]. Hovedmålet for del 2 var å teste ut en metode for vurdering av totalytelse ved mammografivirksomhet basert på registrering av Receiver Operating Characteristic (ROC). Samtidig ble det gjort en kartlegging av bildekvaliteten for det enkelte laboratorium. Denne rapporten omhandler kun prosjektets del 2.

Vanligvis har røntgenutstyr blitt vurdert på grunnlag av fysikalske målinger, som ikke enkelt kan relateres til klinisk praksis. I de senere år har derfor slike vurderinger i større grad også inkludert bildekvaliteten, vanligvis vurdert på grunnlag av bilder tatt av ulike fantomer. Enkelte egenskaper kan bestemmes kvalitativt og/eller kvantitativt, mens andre kun fremkommer som følge av mer eller mindre subjektive vurderinger. Det har vist seg at heller ikke disse egenskapene alltid kan relateres til funn i kliniske bilder. Ved Radiumhospitalet er det gjennom en årrekke arbeidet med å utvikle metodikker som kan gi et mer representativt bilde av laboratoriets kvalitet. Som et ledd i dette er det utviklet fantomer bestående bl.a. av brystvev. De resulterende bildene har mange likhetstrekk med kliniske bilder, noe som gjør det enklere å utnytte radiologenes kunnskaper og erfaringer i vurdering av bilder og bildekvalitet.

I 1992-1993 ble det utviklet et nytt fantom, som skulle kunne brukes til såkalte Receiver Operating Characteristics (ROC) studier. Det ble i denne forbindelse utviklet en teknikk for å produsere "patologiske" strukturer (kalkkorn og fortetninger) med et stort spenn i både dimensjoner og tetthet (kontrast). Disse strukturene ble så støpt inn i tynne epoxyplater, som plasseres oppå en blokk i fantomet som inneholdt brystvev. I denne undersøkelsen ble kun overlageringsplater inneholdende "kalkkorn" benyttet. "Kalkkornenes" størrelse og tetthet spente fra det som ingen laboratorier var i stand til å finne med optimal bildekvalitet, til det alle var i stand til å finne selv med suboptimal bildekvalitet.

Det var store variasjoner i bildekvaliteten, og ved så mange som 28 laboratorier (62 %) var kvaliteten av minst ett av tre bilder av et bestemt fantom (RMI-fantomet) under grensene for godkjenning i henhold til de amerikanske kravene. Bilder tatt av et fantom som hovedsakelig består av brystvev, viste at det er nødvendig å bruke rasterteknikk for å kunne oppnå tilfredsstillende bildekvalitet i mammografi.

Resultatene fra ROC-eksperimentene viser at metoden fungerte tilfredsstillende. De gjennomgående høye verdiene for arealet under ROC-kurven (gjennomsnitt ved kontaktteknikk var 0,879) indikerer imidlertid at størsteparten av "kalkgruppene" var lettere å oppdage enn ønskelig sett fra et metodesynspunkt. Dette reduserte metodens sensitivitet, dvs. dens evne til å kunne skille mellom gode og dårlige laboratorier. På tross av dette ble det funnet store variasjoner i arealet under ROC-kurven for de ulike laboratoriene.

Observatørene ved størstedelen av laboratoriene utnyttet hele spekteret av konfidensnivåer (1-4) slik metoden forutsetter, og de resulterende ROC-kurvene synes meningsfulle sett i forhold til de aktuelle svarskjemaene. Det ble utviklet et program for tilpasning av dataene og beregning av ROC-kurvene. Dette programmet fungerte tilfredsstillende, og metoden for tilpasning av dataene resulterte i ROC-kurver som samsvarte bedre med målepunktene enn den vanligste metoden referert i litteraturen. Med få konfidensnivåer (4) og relativt få "kalkgrupper" (40) er det statistiske grunnlaget ikke det aller beste, og modelltilpasningen ga i noen tilfeller lite meningsfulle kurver. I fremtidige eksperimenter vil det derfor være naturlig med minst 5 konfidensnivåer. Videre bør antallet "kalkgrupper" økes, og disse bør være vanskeligere å oppdage for observatørene. I forbindelse med modelltilpasningen gjøres det i dag ingen beregning av standardavvik eller konfidensintervall for arealet under ROC-kurven, og en metodikk for å bestemme disse størrelsene må utvikles.

Det ble funnet store forskjeller i totalytelse (areal under ROC-kurven) mellom laboratoriene. Størsteparten av de laboratoriene som viste dårlig totalytelse hadde samtidig tilfredsstillende bildekvalitet. I enkelte tilfeller var totalytelsen åpenbart avhengig av betrakningsforhold og/eller observatør: til tross for teknisk tilfredsstillende bildekvalitet klarte ikke laboratoriets radiolog å oppdage tilstrekkelig mange "kalkgrupper".

En sammenligning med kontaktteknikk viste at forstørrelsesteknikk ga både økt sensitivitet og økt spesifisitet for deteksjon av "kalkgruppene". Dette viser at forstørrelsesteknikk må inngå i den kliniske utredningen, noe som dessverre ikke alltid var tilfelle.

Summary

In co-operation with The Norwegian Radium Hospital (NRH), the Norwegian Radiation Protection Authority (NRPA) carried out site visits in all the 45 Norwegian mammography laboratories that received patients in the period December 1993 to March 1994. Parts of the results, with focus on the technical performance of the equipment and on dose measurements, were published in NRPA Report 1995:5 [1, in Norwegian]. In the present report, we describe a method for evaluation of the total performance in mammography by means of Receiver Operating Characteristic (ROC) methodology, and its application in a nation-wide survey. In each laboratory, the image quality was also studied by means of several image quality phantoms.

Traditionally, the performance of x-ray equipment has been assessed by physical measurements that cannot easily be related to the diagnostic imaging. Evaluation of the image quality by means of image quality phantoms has been an important part of such traditional surveys. Some of the image characteristics may be measured qualitatively and/or quantitatively, while others are based on more or less subjective evaluations. Several projects have therefore been carried out at the NRH, in order to develop methods to measure the quality of a laboratory in a way that can be related to clinical practice. A key development has been the construction of phantoms containing fixed mammary tissues. The resemblance between the images obtained with such phantoms and those encountered in clinical practice, enables the radiologist to use his/her knowledge and experience in evaluation of image quality.

In 1992-1993, a new phantom was developed that was designed to be used in Receiver Operating Characteristic (ROC) studies. "Pathological" structures simulating calcifications and masses were manufactured and embedded in overlay plates to be inserted on top of a tissue block. In this survey, only overlay plates containing "calcifications" were applied. The sizes and densities of these structures varied in such a way that some would be perceived even with poor image quality, while others were beyond the limit of detection even in the high quality images.

There were large variations in image quality, and in as many as 28 of the laboratories (62 %), the quality of at least one out of three images obtained with the RMI-phantom fell below the American criteria for accreditation of a mammography facility. Images of a phantom containing fixed mammary tissue, demonstrated the need for anti-scatter grids to obtain acceptable image quality in mammography.

The results of the ROC tests demonstrated that the method worked satisfactory. In general, the area under the ROC curve was high (mean value for contact technique: 0,879). This fact indicates that most of the "calcifications" were too easily detected. The ability of the method to separate good laboratories/observers from slightly worse ones may therefore be improved by a more careful selection of sizes and densities of the "calcifications". However, even with the actual design, large variations in total performance between the different laboratories were found.

The observers in most of the laboratories were able to make use of the entire range of confidence levels (scores 1-4) as presumed in ROC methodology. The resulting ROC curves appear to be meaningful in relation to the reported scores in most cases. A computer program was developed to fit the data to an ROC model and calculate the ROC curves. The program worked well, and the resulting ROC curves agreed better than reported methods with the true and false positive sets obtained at the different confidence levels. However, in some cases, because of the limited number of confidence levels (4) and the relative few groups of “calcifications” (40), the fit to the model gave less meaningful curves. Future experiments should therefore use at least 5 different confidence levels. Furthermore, the number of groups of “calcifications” should be increased, and these should be more difficult for the observers to perceive. The present computations do not include estimation of the standard deviation on the ROC-areas or confidence intervals for these. Further work is needed to develop such methods.

The total performance showed large variations among the laboratories/observers. The majority of laboratories that had sub-optimal total performance also had low quality images. In some cases the sub-optimal total performance was dependent on the viewing conditions and/or the radiologist: the image quality was acceptable, but too few “calcifications” were perceived.

A comparison between contact and magnification technique showed that the latter gave markedly higher sensitivity and specificity for detection of the “calcifications”. It is therefore recommended that magnification technique should be part of the mammographic work-up of suspicious findings.

Innholdsfortegnelse

| | |
|---|------------|
| SAMMENDRAG | iii |
| SUMMARY | v |
| INNHALDSFORTEGNELSE | vii |
| 1 INNLEDNING | 1 |
| 2 RECEIVER OPERATING CHARACTERISTICS (ROC) METODIKKEN | 3 |
| 3 FANTOMER | 7 |
| 3.1 EKSPONERINGSBETINGELSER | 7 |
| 3.2 NRT-FANTOMET | 8 |
| 3.3 RMI-FANTOMET | 9 |
| 3.4 NRH-FANTOMET | 10 |
| 3.5 NRH-ROC-FANTOMET | 11 |
| 3.5.1 <i>Tilpasning til to normalfordelinger</i> | 13 |
| 3.5.2 <i>Tilpasning til rett linje i normalavviks koordinater</i> | 15 |
| 3.5.3 <i>Ikke-parametrisk modell – Wilcoxon statistikk</i> | 16 |
| 4 RESULTATER | 18 |
| 4.1 RESULTATER FOR NRT-FANTOMET | 18 |
| 4.2 RESULTATER FOR RMI-FANTOMET | 19 |
| 4.3 RESULTATER FOR NRH-FANTOMET | 19 |
| 4.4 RESULTATER FOR NRH-ROC-FANTOMET | 20 |
| 5 DISKUSJON | 23 |
| 5.1 KARTLEGGING AV BILDEKVALITET | 23 |
| 5.2 METODE FOR VURDERING AV TOTALYTELSE | 24 |
| 5.3 SAMMENLIGNENDE UNDERSØKELSE AV TOTALYTELSE | 24 |
| 6 KONKLUSJONER | 27 |
| REFERANSER | 28 |
| VEDLEGG I. SYKEHUS/INSTITUTTER SOM DELTOK I UNDERSØKELSEN | 30 |
| VEDLEGG II. MATEMATISK TILPASNING AV ROC-KURVENE | 31 |
| II.1 TILPASNING TIL TO NORMALFORDELINGER | 31 |
| II.2 TILPASNING TIL RETT LINJE I NORMALAVVIKSKOORDINATER | 33 |
| II.3 IKKE-PARAMETRISK MODELL - WILCOXON STATISTIKK | 34 |
| VEDLEGG III. RESULTATENE FOR NRT-FANTOMET | 36 |
| VEDLEGG IV. RESULTATENE FOR RMI-FANTOMET | 40 |
| VEDLEGG V. RESULTATENE FOR NRH-FANTOMET | 43 |
| VEDLEGG VI. RESULTATENE FOR NRH-ROC-FANTOMET | 47 |
| VEDLEGG VII. OPPSUMMERING AV RESULTATENE | 65 |

1 Innledning

Mammografi er røntgenundersøkelse av bryst. Brystet består hovedsakelig av kjertelvev og fettvev. Radiologiske tegn på patologiske forandringer i brystet kan være fortetninger eller små kalkkorn. Fortetningene skiller seg lite fra det normale brystvevet i absorpsjonsegenskaper, noe som stiller store krav til kontrast i bildene. Forkalkningene kan være små, og stiller strenge krav til både skarphet og kontrast. Mammografi er blant de mest krevende disipliner innen radiologien, og det stilles strenge krav til bl.a. personalets kompetanse, utstyrets funksjon og ytelse og kvalitetskontrollrutinene for å oppnå tilfredsstillende bildekvalitet.

Prosjektet "*Mammografivirksomhet i Norge*" er et samarbeid mellom Statens strålevern og en prosjektgruppe ved Det Norske Radiumhospital. Samtlige norske mammografi-laboratorier ble besøkt i perioden desember 1993 til mars 1994. Til sammen 45 mammografiapparater fordelt på 42 sykehus og institutter inngår i undersøkelsen. Disse er ført opp i vedlegg I, sammen med den type mammografiapparat og film som ble brukt. I denne rapporten er laboratoriene anonymisert, og det enkelte laboratorium omtales kun ved en to-bokstavers kode.

Selve prosjektet var todelt. Del 1 var et ordinært tilsyn, med hovedvekt på apparaturens tekniske ytelse og dosemålinger. Hovedmålet for del 2 av prosjektet var å teste ut en metode for vurdering av totalytelse. I denne sammenheng ble det også gjort en kartlegging av bildekvaliteten for hvert enkelt laboratorium. I denne rapporten vil hovedsakelig det som har med vurdering av bildekvalitet og totalytelse bli diskutert. Samtlige laboratorier har tidligere mottatt en egen rapport [1] som oppsummerer resultatene fra de tekniske målingene.

Vurdering av røntgenapparat har tradisjonelt vært gjort av medisinske fysikere på grunnlag av fysikalske målinger. Slike målinger kan være vanskelige å relatere til klinisk praksis. I de senere år har derfor slike vurderinger i større grad også inkludert bildekvaliteten. Dette har som oftest vært gjort på grunnlag av bilder tatt av ulike fantomer. Etter avbildning av fantomet, vurderes fremstillingen av de enkelte strukturer. Enkelte parametre kan bestemmes kvalitativt og/eller kvantitativt, mens andre kun fremkommer som følge av en mer eller mindre subjektiv vurdering.

De kommersielt tilgjengelige fantomene inneholder vanligvis strukturer som linjetest-rastere, lavkontraststrukturer og mikropartikler. Bruken av slike fantomer gir imidlertid ikke alltid informasjon som lett kan relateres til funn i kliniske bilder. Dette kan i verste fall resultere i at man trekker feil konklusjoner. En av de mest fundamentale svakhetene ved slike fantomer er at de ikke simulerer de komplekse mønstrene man finner i naturlig brystvev. Et annet problem ved disse fantomene kan være at de gir for liten mulighet for å skille mellom bilder som bare oppviser små forskjeller i kvalitet. Dette gjelder spesielt ved sammenligning av ulike mammografiapparater eller film-folie-kombinasjoner, eller i forbindelse med optimalisering.

Det ble i 1986 utviklet et mammografifantom, NRH-fantomet [2], ved Radium-hospitalet bestående hovedsakelig av brystvev (ca. 8x10x4 cm). Dette ble fiksert i

formalin og lagt i en hermetisk lukket polyetylenboks. På toppen av vevet ble det lagt inn fire grupper med små partikler som simulerer forkalkninger. Fantomet tillater kun subjektiv vurdering av bildekvalitet, men studier viser ingen betydelige intra- eller interobservatørvariasjoner [3]. Dette er en viktig forutsetning for at metoden kan anses pålitelig og realistisk. Fordelene er at man på grunnlag av legenes vurdering av bilder tatt med dette fantomet bedre kan relatere resultatene til klinisk praksis og at det er følsomt for svært små forskjeller i bildekvalitet. Etter flere års bruk av fantomet kan man konkludere med at: i) resultatene lett kan overføres til klinisk praksis, ii) metoden har høy sensitivitet sammenlignet med bruk av andre kjente fantomer, iii) resultatene avhenger i relativt liten grad av radiologen som vurderer bildene, iv) fantomet har muliggjort et samarbeid på tvers av de yrkesgruppene som er involvert i mammografi-virkomheten.

En svakhet ved NRH-fantomet er at det ikke inneholder strukturer med kjent størrelse og tetthet (kontrast). Basert på erfaringene med fantomet ble det i 1992 utviklet et nytt fantom [4,5]. Dette skulle kunne brukes til såkalte ROC-studier (Receiver Operating Characteristics). For å muliggjøre dette, ble det utviklet en teknikk for å produsere "patologiske" strukturer (forkalkninger og fortetninger) med eksakte dimensjoner og valgfrie attenuasjonsegenskaper. Disse strukturene ble så støpt inn i tynne epoxyplater. Slike plater plasseres inne i fantomet, oppå ulike bokser inneholdende brystvev. For å kunne gjennomføre ROC-eksperimenter var det nødvendig å lage strukturer med et stort spenn i både størrelse og tetthet (kontrast). Disse måtte spenne fra det som ingen laboratorier er i stand til å finne med optimal bildekvalitet, til det alle er i stand til å finne selv med suboptimal bildekvalitet.

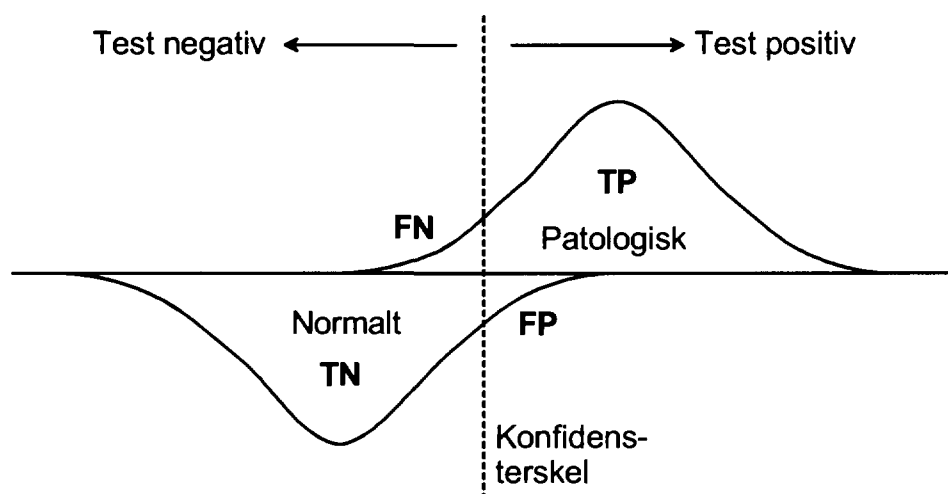
Et forprosjekt vinteren 1993 dannet grunnlaget for prosjektet "*Mammografivirkomhet i Norge*". Det ble på grunnlag av erfaringene fra forprosjektet utviklet en ny versjon av ROC-fantomet.

Prosjektet "*Mammografivirkomhet i Norge*" har til hensikt å bruke ROC-fantomet for evaluering av laboratoriekvalitet (totalytelse) ved mammografivirkomhet og sammenholde resultatene med vurdering av bildekvalitet og fysikalske målinger.

2 Receiver Operating Characteristics (ROC) metodikken

Å relatere bildekvalitet kun til fysikalske egenskaper ved det bildedannende system gir to ulike metodiske problemer; i) å lage et altomfattende mål for bildekvalitet er umulig, og ii) slike sammenhenger inkluderer ikke persepsjonsevnen til observatøren. Selve den diagnostiske prosessen kan deles inn i følgende trinn: i) deteksjon (noe i bildet), ii) gjenkjenning (størrelse og form) og iii) identifisering (patologi, sykdomsmønster). Det finnes ingen enkel metode for å evaluere den diagnostiske prosessen, og heller ikke for å vurdere bildekvalitetens betydning for det diagnostiske arbeidet. Hvis problemstillingen forenkles til kun å omfatte deteksjonsprosessen, finnes det imidlertid metodikker tilgjengelige. Med bruk av den såkalte *Receiver Operating Characteristics* (ROC) metoden [6,7,8], kan man korrelere en observatørs beslutning (“funn/deteksjon”) med den “sanne diagnosen” (etablert/bekreftet med en uavhengig metode).

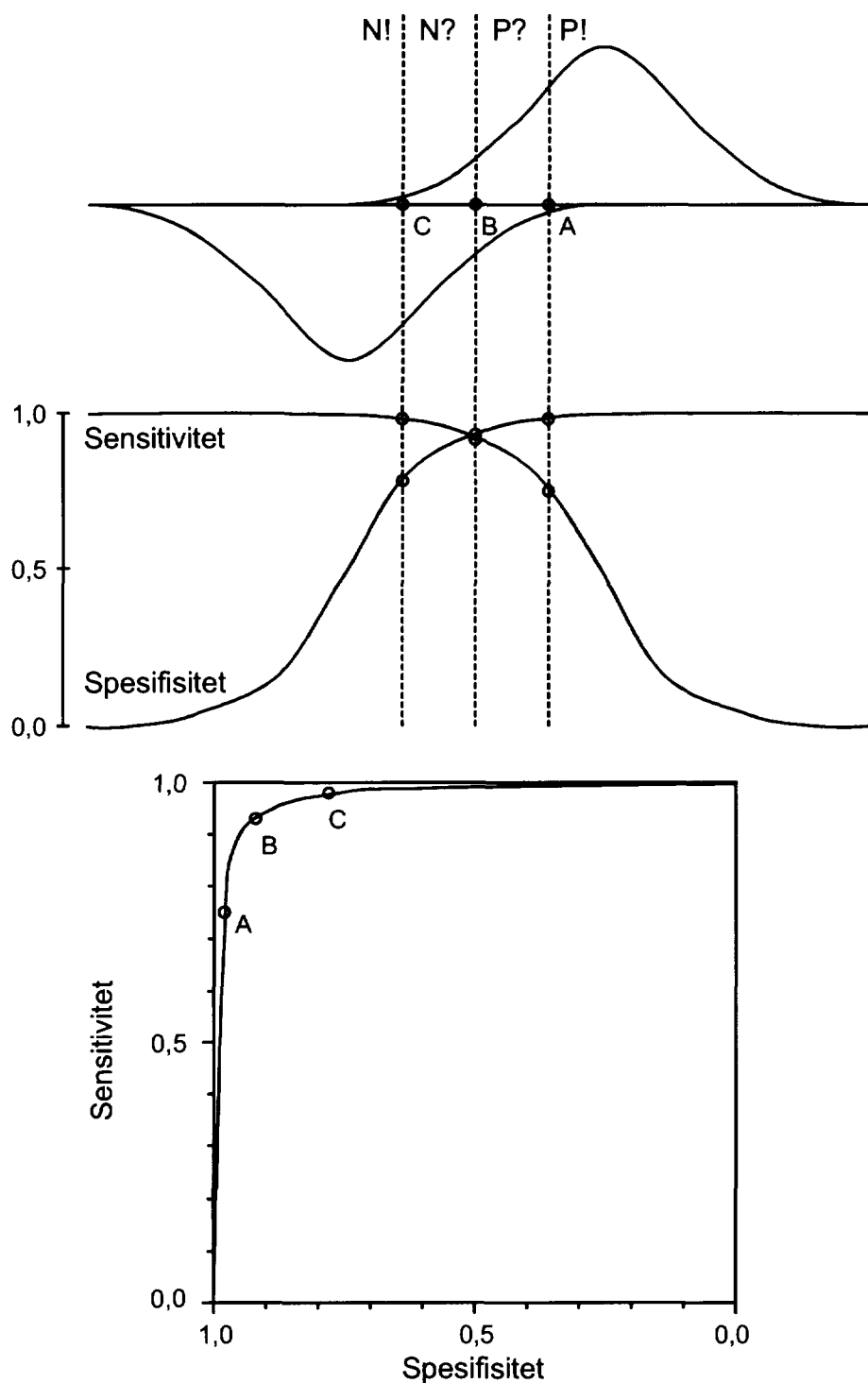
For å skissere fremgangsmåten ved et ROC-eksperiment, kan man langs en akse (figur 1) tegne to histogrammer (som f.eks. kan være bildeintensiteten til et område). Histogrammet over aksens kan representere fordelingen av verdier for patologiske områder, mens histogrammet under aksens kan representere fordelingen av verdier for normale områder. Den ideelle situasjon ville vært at disse to histogrammene ikke overlappet hverandre, noe de i praksis alltid vil gjøre i større eller mindre grad.



Figur 1. Prinsippet for et ROC-eksperiment. En patologisk fordeling deles i sanne positive (TP) og falske negative (FN), og en normal fordeling deles i falske positive (FP) og sanne negative (TN) av en konfidens-terstel.

Hvis det kun tillates å svare enten negativ (N) eller positiv (P), vil observatøren sette den såkalte *konfidens-tersekelen* et sted på aksens, slik at N og P separeres. Settes terskelen helt til høyre på aksens vil alle tester komme ut som N. Tilsvarende vil en terskel helt til venstre på aksens føre til at alle tester kommer ut som P. Normalt settes terskelen et sted midt mellom disse ytterpunktene, f.eks. som vist ved den stiplede linjen på figur 1. Konfidens-tersekelen deler den normale populasjonen i en sann negativ (TN) og en falsk positiv (FP) gruppe, og den syke populasjonen i en sann positiv (TP) og en falsk negativ (FN) gruppe.

I et ROC-eksperiment vil man arbeide med flere muligheter enn kun negativt og positivt. Det er vanlig å benytte seg av 4 eller 5 nivåer: *sikkert patologisk*, *sannsynligvis patologisk*, evt. *usikkert funn*, *sannsynligvis normalt* og *sikkert normalt*. Avhengig av de



Figur 2. Prinsippet for å bestemme en ROC-kurve. Man presser observatøren til å sette ulike konfidensskler (angitt ved A, B og C). P! betyr sikkert patologisk, P? sannsynligvis patologisk, N? sannsynligvis normalt og N! sikkert normalt. For hvert konfidensnivå får man en sensitivetsverdi og en spesifisetsverdi. Ved å plote sensitivitet som funksjon av spesifisitet får man det man kaller en ROC-kurve.

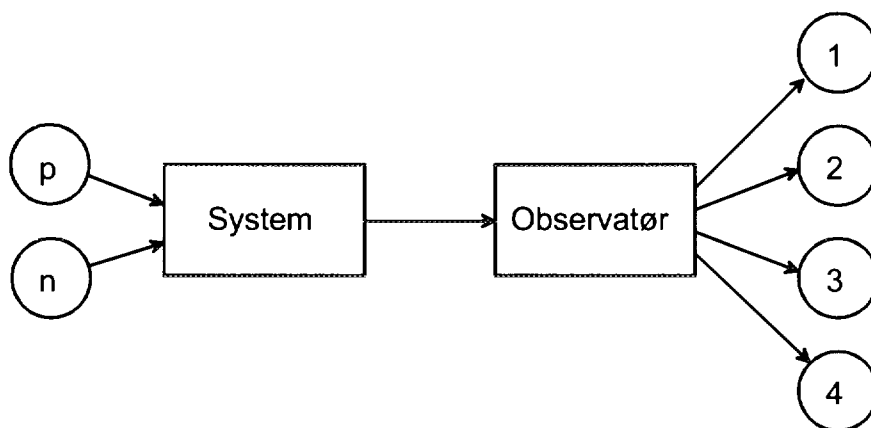
kliniske konsekvensene vil konfidensterskelen flyttes bevisst eller ubevisst. Størrelsene *sensitivitet* og *spesifisitet* er ofte benyttet for vurdering av diagnostiske metoder. Med sensitivitet menes testens evne til å plukke ut de syke blant de syke, og med spesifisitet menes testens evne til å utelukke sykdom blant de friske. Sensitiviteten er definert som $\frac{TP}{TP+FN}$, og spesifisiteten er definert som $\frac{TN}{TN+FP}$. Fra figur 2 ser man at når konfidensterskelen flyttes fra venstre mot høyre vil sensitiviteten gå fra 1 (ingen falske negative) til 0 (ingen sanne positive), mens spesifisiteten vil gå fra 0 (ingen sanne negative) til 1 (ingen falske positive). Man ser av figuren at sensitivitet og spesifisitet er to helt uavhengige størrelser. Det ville derfor vært meningsløst å oppgi en sensitivitetsverdi uten også å oppgi den tilhørende spesifisitetsverdien.

Sensitivitet og spesifisitet, slik de er definert her, er det som betegnes pretest sannsynligheter. Hvis man vet testens sensitivitet vet man m.a.o. sannsynligheten for at testen vil bli positiv under forutsetning av at det er sykdom tilstede. Ofte ønskes egentlig svar på det motsatte, nemlig sannsynligheten for at det er sykdom tilstede når den diagnostiske testen er positiv. Denne sannsynligheten kalles for *Positive Predictive Value* (PPV), og er definert som $\frac{TP}{TP+FP}$. Den tilsvarende sannsynligheten for at det ikke er sykdom tilstede når testen er negativ, kalles for *Negative Predictive Value* (NPV), og er definert som $\frac{TN}{TN+FN}$.

En entydig kombinasjon av sensitivitet og spesifisitet er definert for hver posisjon av konfidensterskelen. Ved å plote hver sensitivitetsverdi mot dens tilsvarende spesifisitetsverdi, får man en kurve som i figur 2 (nederst). I praksis kan man ikke lage eksperimenter hvor konfidensterskelen flyttes kontinuerlig. Dette kan kun gjøres i diskrete trinn, der hvert trinn svarer til de ulike svaralternativene nevnt tidligere. For hvert slikt alternativ får man et sett med sammenhørende sensitivitetsverdi og spesifisitetsverdi, som da svarer til et punkt på kurven. Denne sensitivitet-spesifisitetskurven er, bortsett fra navnene på aksene, identisk med en ROC-kurve. ROC-kurver presenteres vanligvis med en absisseakse som representerer den falske positive fraksjonen (FPF), definert som $1,0 - \text{spesifisiteten}$, mens ordinataksen representerer den sanne positive fraksjonen (TPF), som er identisk med sensitiviteten.

Hvis antall svaralternativ er N, får man N-1 punkter på ROC-kurven. I tillegg vet man at kurven starter i det nedre, venstre hjørnet og slutter i det øvre, høyre hjørnet. Figur 3 viser forløpet i et ROC-eksperiment. Input til systemet kan f.eks. være et antall pasienter (p+n), hvorav noen (gjerne halvparten) har en sykdom (p). Systemet kan være et røntgenlaboratorium inkludert operatør (i dette tilfellet en radiograf), en film-foliekombinasjon og en framkallermaskin. Ut fra systemet kommer det bilder for hver enkelt pasient. Observatøren (radiolog i dette tilfelle) skal da for hvert bilde sette et tall 1, 2, 3 eller 4 med følgende betydning:

- 1 - sikkert normalt
- 2 - sannsynligvis normalt
- 3 - sannsynligvis abnormalt
- 4 - sikkert abnormalt



Figur 3. Skjematisk fremstilling av et ROC-eksperiment.

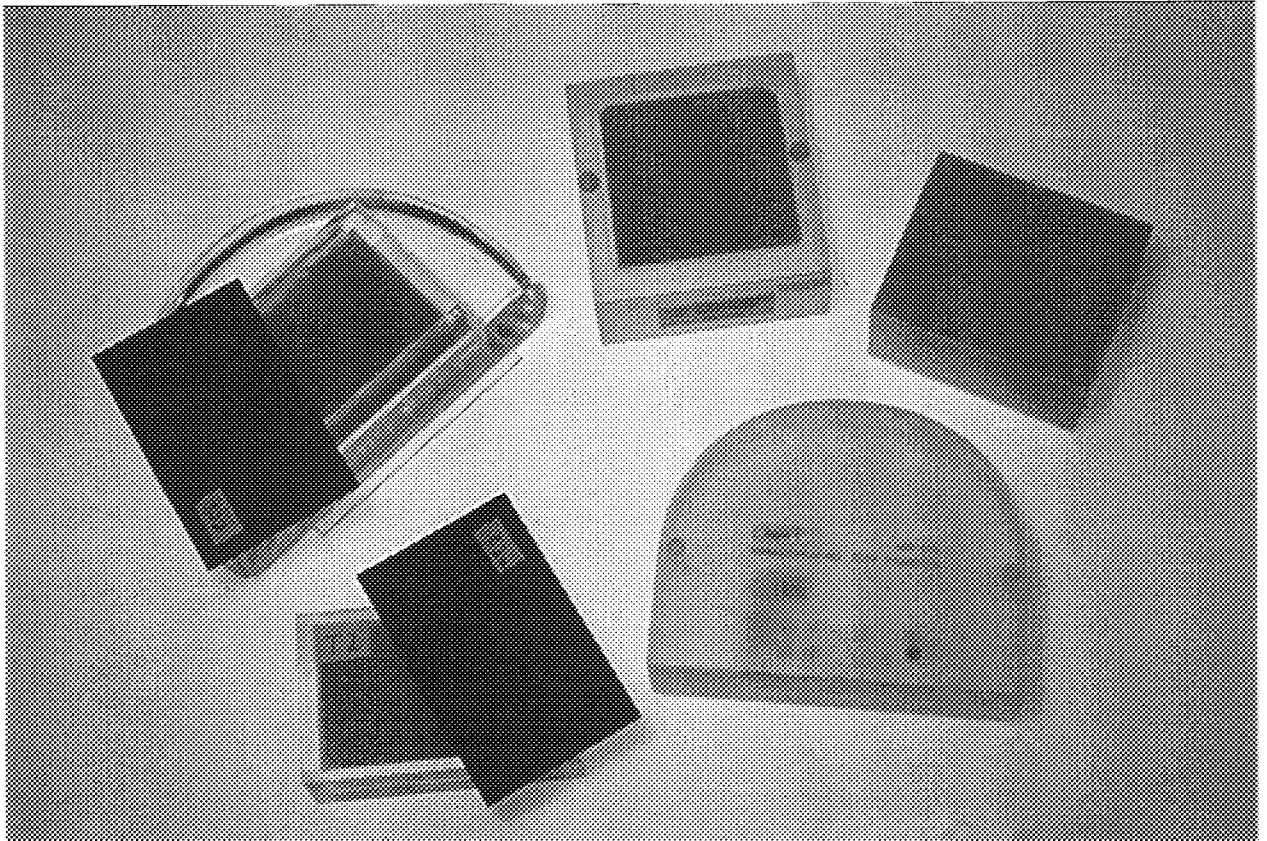
For å vise hvordan hvert enkelt punkt på en ROC-kurve fremkommer, kan man tenke seg at både (p) og (n) på figur 3 er 100. Punktet A på figur 2 fremkommer ved å telle opp antall 4'ere for gruppen (p) og (n). Ettersom antallet 4'ere for gruppen (p) er 75 (tabell 1) og gruppen bestod av 100 pasienter, får vi $TPF_A = 0,75$. Antallet 4'ere for gruppen (n) er 2, noe som gir $FPF_A = 0,02$. Det neste punktet, B, får vi ved å summere antallet 4'ere og 3'ere for hver av de to gruppene (p) og (n). Det er til sammen 93 4'ere og 3'ere for gruppen p, og TPF_B blir dermed lik 0,93. Tilsvarende blir FPF_B lik 0,07. Den samme prosedyren følges for punkt C. Det siste punktet, som vi får ved å telle opp antallet 4'ere, 3'ere, 2'ere og 1'ere for hver gruppe (p) og (n), vil alltid få FPF og TPF lik 1,0 (øvre, høyre hjørne), og gir derfor ikke noen informasjon om ROC-kurvens forløp. Vi får derfor N-1 punkter på en ROC-kurve, hvor N er antallet kategorier. I dette eksemplet var N lik 4 og vi fikk 3 punkter på den resulterende kurven.

| Pasientgruppe | 4'ere (P?) | 3'ere (P?) | 2'ere (N?) | 1'ere (N?) |
|---------------|------------|------------|------------|------------|
| p | 75 | 18 | 5 | 2 |
| n | 2 | 5 | 14 | 79 |

Tabell 1. Resultatene fra et tenkt ROC-eksperiment (kurven som er vist i figur 2).

3 Fantomer

Et fantom er et objekt som plasseres i strålegangen, og som skal etterligne den aktuelle kroppsdel (her: brystet) i form og størrelse og mht. absorpsjon og spredning av røntgenstrålingen. Samtidig bør fantomet inneholde strukturer som ligner de man finner i den aktuelle kroppsdel. De fantomene som inngår i denne undersøkelsen (figur 4) oppfyller bare delvis disse kravene.



Figur 4. De fire fantomene som har vært benyttet i dette prosjektet. Til venstre ses NRH-ROC-fantomet med klossene med brystvev og overlagringsplatene. Til høyre foran ses NRT-fantomet, i midten bak RMI-fantomet og bak til høyre NRH-fantomet.

3.1 Eksponeringsbetingelser

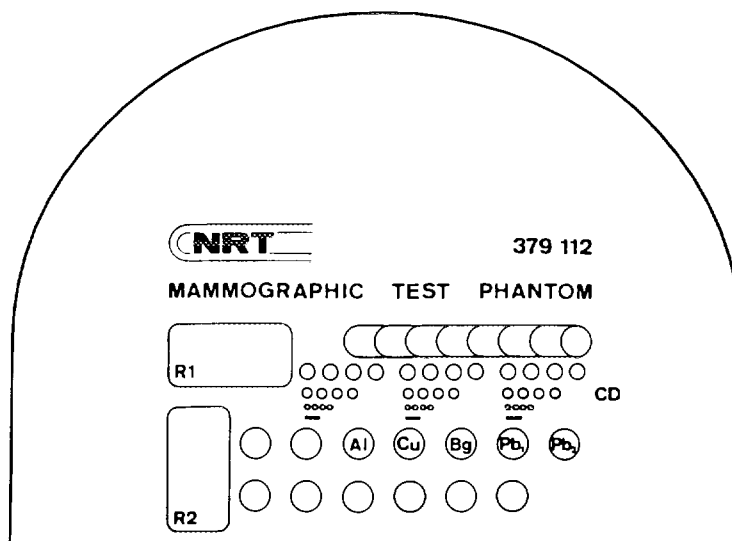
Tre ulike eksponeringsbetingelser har blitt benyttet: 1) anvendt teknikk, 2) referanseteknikk og 3) forstørrelsesteknikk. Anvendt teknikk og referanseteknikk er begge kontaktteknikk (fantomet plassert direkte oppå bucky eller kassett), men ulike rørspenninger er benyttet. Ved anvendt teknikk ble det benyttet den kV som normalt ble brukt ved laboratoriet, mens det ved referanseteknikk ble brukt den kV som svarte til en målt kV_p nær 28 [1]. For mange av laboratoriene ble referanseteknikk identisk med anvendt teknikk. Ved forstørrelsesteknikk ble de normale innstillingsverdiene ved laboratoriet benyttet. Ved å bruke de normale innstillingsverdiene får man informasjon som er mest mulig klinisk relevant, mens man ved referanseteknikk muliggjør sammenligning av utstyrets tekniske ytelse ved tilnærmet like betingelser.

I de følgende avsnittene gis det en kort beskrivelse av hvert enkelt fantom benyttet i dette prosjektet. Alle filmsvertninger er avlest med et og samme densitometer (Tobias Associates Model TBX; Ivyland, PA).

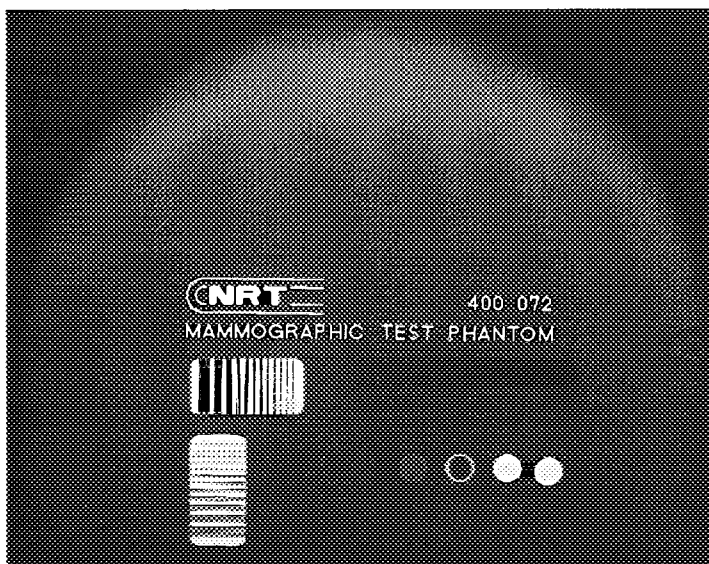
3.2 NRT-fantomet

NRT-fantomet (figur 4) (NRT Mammographic Test Phantom; Nordic Roentgen Technology, Hasselager, Danmark), er omtrent identisk med CGR-fantomet [9], og er det vanligste fantomet ved norske mammografilaboratorier.

Figur 5 viser en skisse av NRT-fantomet, med de strukturer som finnes innebygd, og figur 6 viser et røntgenbilde av fantomet. Kun de strukturer som har blitt benyttet i prosjektet, er identifisert på skissen av fantomet (figur 5). I beskrivelsen nedenfor er henvisning til disse strukturene angitt med kursivert skrift.



Figur 5. Skjematisk fremstilling av NRT-fantomet.



Figur 6. Røntgenbilde av NRT-fantomet.

Området *Bg* er for måling av filmsvertning, og fantomet ble eksponert slik at filmen fikk en svertning mellom 1,15 og 1,25 OD. Områdene *Al* og *Cu* (folier av aluminium respektive kobber i fantomet) er utformet for beregning av to kontrastindekser. Disse er gitt som $C_{Al} = \frac{D_{Bg} - D_{Al}}{D_{Bg}}$ og $C_{Cu} = \frac{D_{Bg} - D_{Cu}}{D_{Bg}}$, hvor D_{Bg} , D_{Al} og D_{Cu} er filmsvertningen målt i områdene *Bg*, *Al* og *Cu*.

Linjetestrasterne R_1 og R_2 er for bestemmelse av oppløsningsevnen. R_1 angir oppløsningen i fokusets bredderetning, (vinkelrett på anode-katode-retningen), og R_2 i lengderetningen (parallelt med anode-katode-retningen).

S/P (forholdet mellom spredt stråling og primærstråling) er gitt som $S/P = \frac{D_{Pb1} - D_{Pb2}}{D_{Bg} - D_{Pb1}}$, hvor D_{Bg} , D_{Pb1} og D_{Pb2} er filmsvertningen målt i områdene *Bg*, *Pb₁* og *Pb₂*.

Seksjonen *CD* er for bestemmelse av kontrast-detalj-oppløsningen. *CD*-seksjonen består av tre grupper med objekter med forskjellig kontrast (3 %, 2 % og 1 % målt ved 20 keV monoenergetisk). Hver gruppe består av fire rader, hver med fire objekter av samme diameter (3 mm, 2 mm, 1 mm og 0,5 mm). Det ble gitt et bedømmelsestall (score) for hvert enkelt objekt på grunnlag av kriteriene gitt i tabell 2, og kontrast-detalj-oppløsningen er gitt som summen av alle objektenes score.

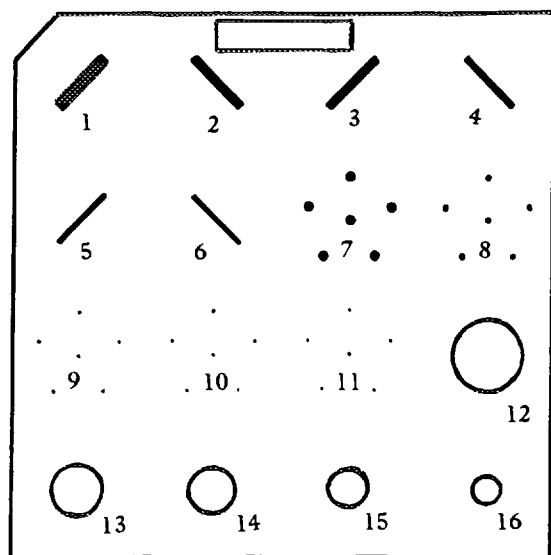
| <i>Fremstilling av objekt</i> | <i>Score</i> |
|-------------------------------|--------------|
| God kontrast, sirkulær | 1.00 |
| God kontrast, diffus kant | 0.50 |
| Diffus og uskarp, ikke rund | 0.25 |

Tabell 2. Bedømmelseskriterier for kontrast-detalj-oppløsningen.

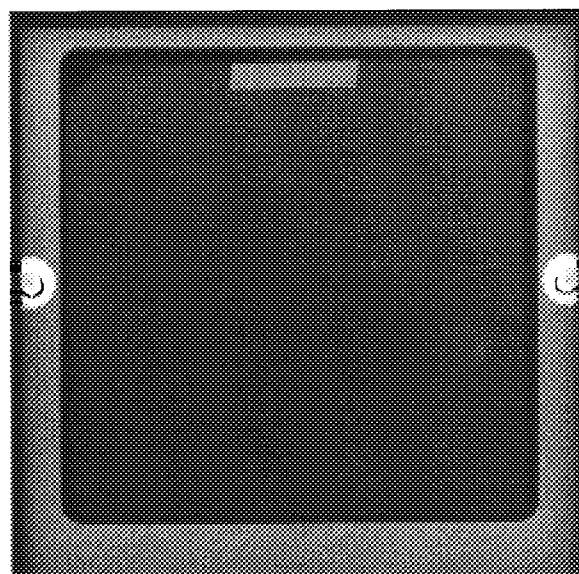
3.3 *RMI-fantomet*

RMI-fantomet (figur 4) [10,11] (Mammographic Accreditation Phantom, Model 156; Radiation Measurements Inc., Middleton, WI) har inngått i ACR's (American College of Radiology) frivillige akkrediteringsordning for mammografilaboratorier i USA, og inngår nå i FDA's (Food and Drug Administration) godkjenningsordning (som ble gjort gjeldende fra 1. oktober 1994). Det spesielle med dette fantomet er at det for hver type struktur er helt bestemte krav til hvor mange strukturer som skal kunne sees i bildene for å oppfylle kravene til god bildekvalitet. RMI-fantomet ble kun eksponert ved anvendt og referanseteknikk.

Figur 7 viser en skisse av RMI-fantomet, med de strukturer som finnes innebygd. Strukturene simulerer bindevevsdrag, kalk og fortetninger, noe som fremgår av røntgenbildet av fantomet (figur 8). Fantomet består av en tynn voksskive som inneholder de ulike strukturene (6 fibriller, 5 grupper med kalk, 5 fortetninger), en 43 mm tykk perspeks-kloss med rom for voksskiven og et lokk. Dette tilsvarer absorpsjonen for et 45 mm tykt komprimert bryst av normal tetthet. Fantomet ble eksponert slik at svertningen avlest i senter av fantomet ble mellom 1,10 og 1,40 OD. Tabell 3 angir de respektive strukturenes dimensjoner (diametere for fibrillene, kornstørrelsen for



Figur 7. Skjematisk fremstilling av RMI-fantomet.



Figur 8. Røntgenbilde av RMI-fantomet.

kalkgruppene og tykkelsen for fortetningene) og numrene henviser til skissen av fantomet (figur 7). For at de amerikanske kravene til god bildekvalitet skal være oppfylt, må 0,75 mm fibrillen, 0,32 mm kalkgruppen og 0,75 mm fortetningen kunne sees.

3.4 NRH-fantomet

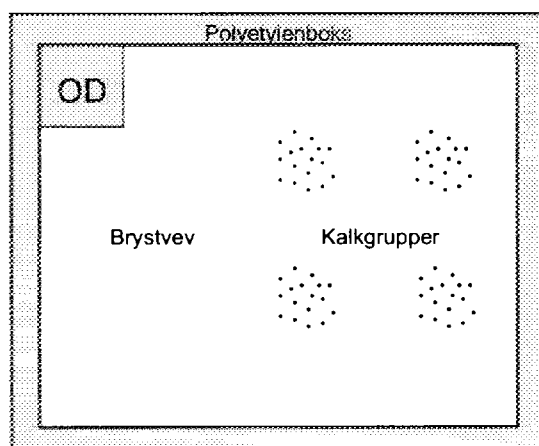
NRH-fantomet [2,3] er et antropomorft fantom (figur 4), som hovedsakelig består av formalinfiksert brystvev. Figur 9 viser en skisse og figur 10 et røntgenbilde av fantomet. Vevet er lagt inn i en hermetisk lukket polyetylenboks. På toppen av vevet er det lagt små korn som simulerer forkalkninger. Disse er fiksert mellom to lag med tynn plastfolie, og danner fire grupper. Den gruppen som inneholder de minste kornene (hovedsakelig korn mellom 0,2 og 0,3 mm) har vist seg å være spesielt nyttig ved vurdering av bildekvalitet. I det ene hjørnet av fantomet er vevet erstattet med polyetylen, slik at svertningen i bildene kan måles på en konsistent måte (området OD på figur 9).

| Type | Nummer | Dimensjon |
|---|--------|-----------|
| Fibrill | 1 | 1,56 mm |
| | 2 | 1,12 mm |
| | 3 | 0,89 mm |
| | 4 | 0,75 mm |
| | 5 | 0,54 mm |
| | 6 | 0,40 mm |
| Kalk (Al ₂ O ₃) | 7 | 0,54 mm |
| | 8 | 0,40 mm |
| | 9 | 0,32 mm |
| | 10 | 0,24 mm |
| | 11 | 0,16 mm |
| Fortetning | 12 | 2,00 mm |
| | 13 | 1,00 mm |
| | 14 | 0,75 mm |
| | 15 | 0,50 mm |
| | 16 | 0,25 mm |

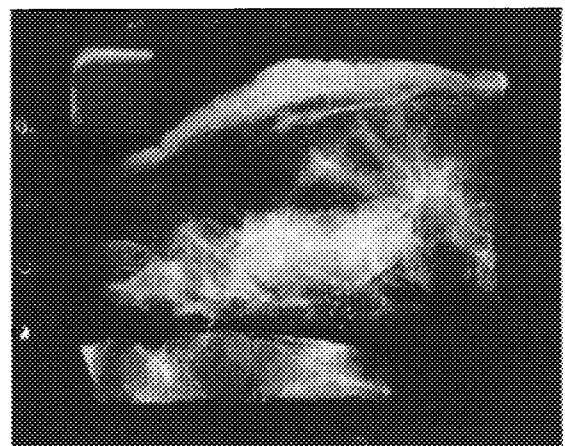
Tabell 3. Strukturene i RMI-fantomet.

Fantomet ble eksponert slik at filmen fikk en svertning mellom 1,10 og 1,40 OD. Bilder av dette fantomet kan kun vurderes subjektivt, men det har likevel flere fortrinn sammenlignet med de konvensjonelle, fysiske bildekvalitetsfantomene. En av fordelene er at man kan vurdere fremstillingen/visualiseringen av strukturer som radiologene er vant til å se, men også at man kan utnytte radiologenes erfaring i vurdering av bilder og bildekvalitet.

Den subjektive vurderingen av NRH-bildene kan gjøres på flere måter. I dette prosjektet ble først det beste bildet for hvert sted og for hver teknikk valgt ut. Dette ble så sammenlignet med et referansebilde av god kvalitet. Sammenligningen ble gjort to



Figur 9. Skjematisk fremstilling av NRH-fantomet.



Figur 10. Røntgenbilde av NRH-fantomet.

ganger; først med henblikk på kontrast (visualisering av bløtvevsstrukturer), og deretter med henblikk på skarphet (fremstilling av kalkkorn). For hver av disse egenskapene ble bildene kategorisert i 5 nivåer på grunnlag av kriteriene gitt i tabell 4. Bilder med score 5-3 representerer bilder omtrent som referansebildet, og forskjellen mellom disse score-verdiene kan være svært liten. Score 2 representerer bilder av betydelig dårligere kvalitet enn score 5-3. Bilder med score 1 er av så dårlig kvalitet at det kan ha betydning for den diagnostiske bruken av bildene. Sammenligningen ble gjort av en radiolog med lang erfaring i mammografi og i vurdering av bildekvalitet med utgangspunkt i bilder av dette spesielle fantomet. I tillegg til angitt score for kontrast og skarphet ble det anmerket eventuelle kommentarer til bildet. Dette kunne være ting som f.eks. dårlig kontrast, generelt uskarpt, manglende eller uskarpe kalkkorn, støv på foliene, rasterstriper og framkallerartefakter.

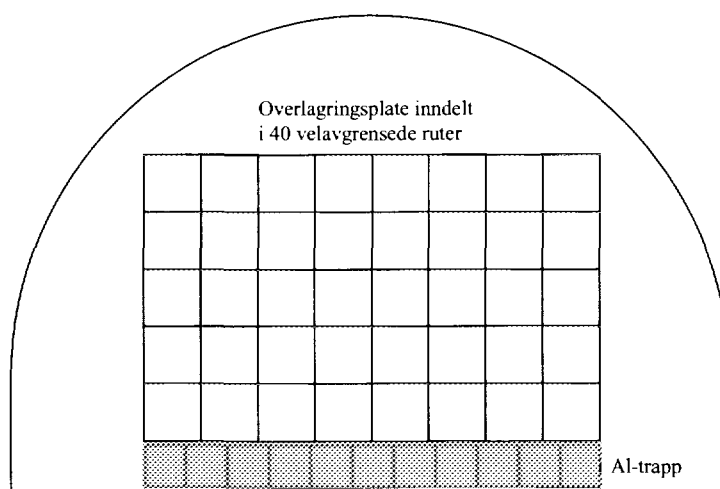
| Score | Bløtvev | Kalk |
|-------|--------------------------------|-------------------------------------|
| 5 | Visualiseres bedre | Fremstilles bedre |
| 4 | Visualiseres som på referansen | Fremstilles som på referansen |
| 3 | Visualiseres, men dårligere | Fremstilles mer diffus |
| 2 | Ikke alt visualiseres | Fremstilles dårligere (25 % mistes) |
| 1 | Dårlig visualisering | Fremstilles dårlig (> 50 % mistes) |

Tabell 4. Score-kriterier for NRH-fantomet ved sammenligning med et referansebilde av god kvalitet.

3.5 NRH-ROC-fantomet

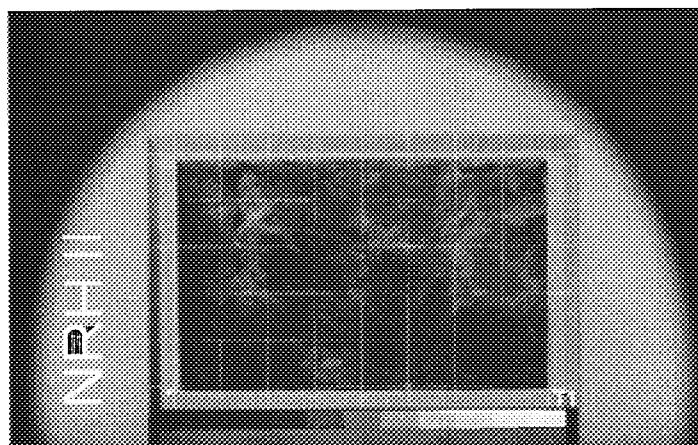
NRH-ROC-fantomet (figur 4) [4,5] er et fantom utviklet for vurdering av totalytelse ved mammografivirksomhet. Figur 11 viser en skisse av fantomet, som består av en perspekkloss formet som et komprimert bryst (45 mm tykt). I fantomet er det et rom med plass for en boks inneholdende fiksert brystvev og en såkalt overlageringsplate. Hver overlageringsplate er delt inn i 40 ruter, og i ca. halvparten av disse er det støpt inn strukturer som simulerer forkalkninger. Det er utviklet en teknikk for å lage "kalkkorn" med diameter helt ned til 0,1 mm og med eksakt den ønskede dempningsevne for røntgenstrålingen.

I dette eksperimentet ble to bokser med brystvev og to overlageringsplater benyttet. Det ble tatt to bilder ved anvendt teknikk og to bilder ved forstørrelsesteknikk. Figur 12 viser et røntgen-bilde av fantomet for den ene av kombinasjonene brystvev og overlageringsplate, og figur 13 viser et røntgenbilde av de to overlageringsplatene uten brystvevet superponert.



Figur 11. Skjematisk fremstilling av NRH-ROC-fantomet.

Bildene ble “beskrevet” av en radiolog tilknyttet laboratoriet. For hver rute i bildet skulle radiologen sette en score med følgende betydning: 1=“kalk” sikkert ikke tilstede (N!), 2=“kalk” sannsynligvis ikke tilstede (N?), 3=“kalk” sannsynligvis tilstede (P?) og 4=“kalk” sikkert tilstede (P!).

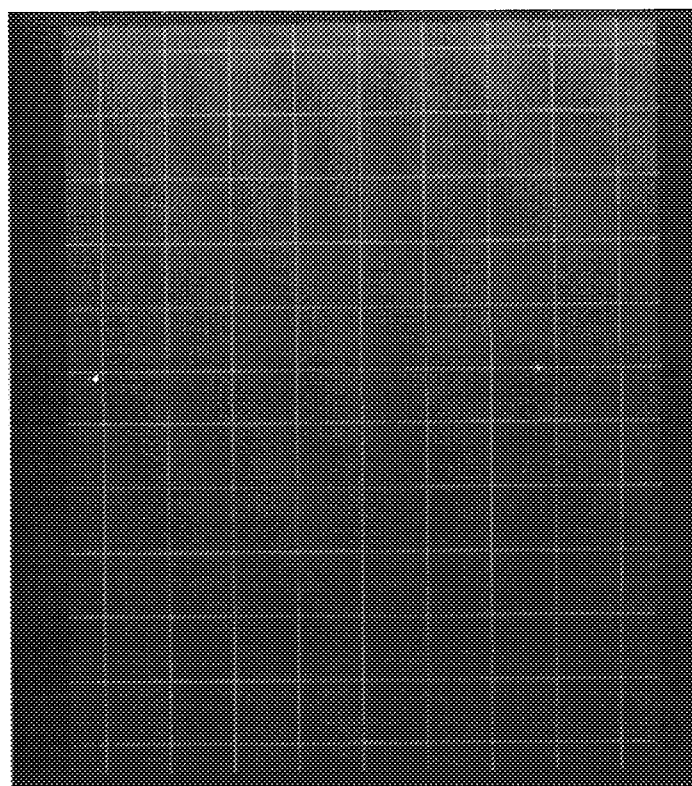


Figur 12. Röntgenbilde av NRH-ROC-fantomet med den ene av kombinasjonene brystvev og overlagringsplate.

Ved å sammenholde radiologens “funn” med en fasit, kan man bestemme ROC-kurven, som viser sammenhengen mellom falske positive funn og sanne positive funn ved ulike konfidensnivåer. Ved stadig å innbefatte flere ruter (først kun 4’erne, deretter 4’erne og 3’erne og til slutt også 2’erne), flytter man konfidensterskelen mot venstre. Denne forskyvningen av konfidensterskelen fører som oftest til flere sanne positive funn, men samtidig også til et økende antall falske positive funn.

I litteraturen er resultatene fra ROC-eksperimenter fremstilt på ulike måter. Det finnes også flere tilgjengelige programmer for tilpasning av ROC-kurver og for beregning av arealet under kurven [12,13,14,15]. Disse har ikke blitt benyttet i dette prosjektet, da det også var en målsetning for dette prosjektet å utvikle verktøy for tilpasning av ROC-kurver, og for grafisk presentasjon av disse.

Programmer for tilpasning av ROC-kurvene til de eksperimentelle dataene har blitt skrevet i IDL (Interactive Data Language; Research Systems, Boulder, CO). Dette er et interaktivt programmeringsspråk som inneholder funksjoner for matematikk og statistikk samt for grafisk presentasjon av resultatene. Flere metoder for tilpasning av dataene har blitt forsøkt, og de tre metodene som har blitt benyttet er beskrevet i



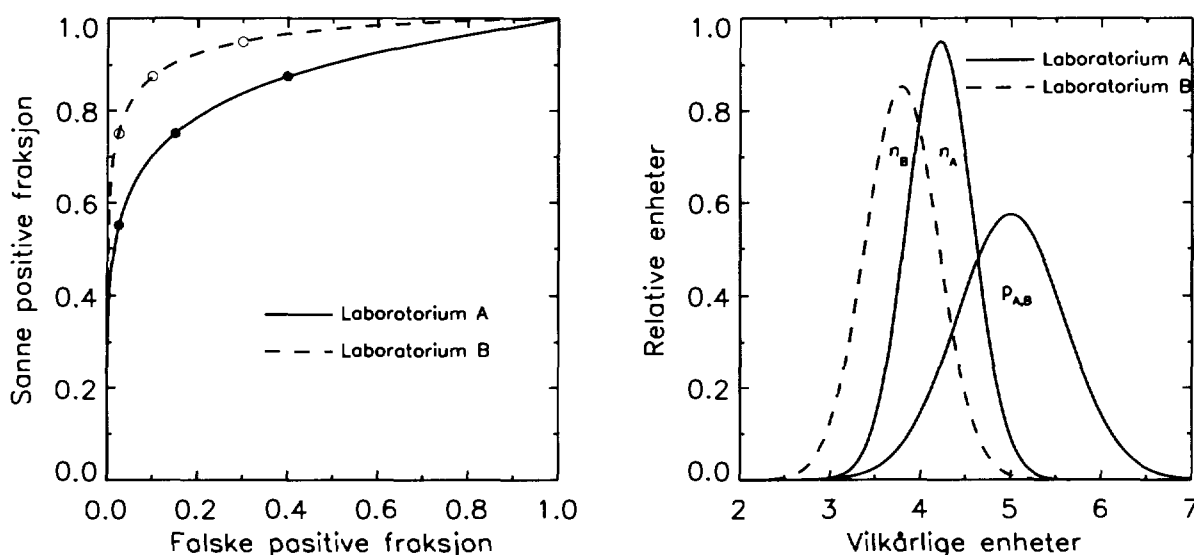
Figur 13. Röntgenbilde av de to overlagringsplatene som er benyttet i prosjektet uten brystvevet superponert.

avsnittene under. I utgangspunktet ble en parametriske modell valgt, da dette er det vanligste. Denne tar utgangspunkt i to fordelingsfunksjoner, der den ene utgjør den negative fordelingen og andre den positive fordelingen (tilsvarende beskrivelsen i kapittel 2). Det er vist at en slik modell er uavhengig av valg av fordelingsfunksjoner [16], og det er derfor naturlig å velge to normalfordelinger.

Det ble først utviklet et program hvor de to normalfordelingene og tersklene ble bestemt ved modelltilpasning (avsnitt 3.5.1). I litteraturen [6,17] er en alternativ tilpasning basert på samme modell (to normalfordelinger) gjort ved å legge en rett linje gjennom målepunktene i normalavvikskoordinater. Denne metoden ble også implementert pga. dens utbredelse og popularitet (avsnitt 3.5.2). Det kan forventes at resultatene fra denne metoden samsvarer med resultatene fra metoden beskrevet i avsnitt 3.5.1, når forutsetningen om normalfordelte fordelingsfunksjoner er tilstede. Det viste seg at begge disse metodene feilet i enkelte situasjoner. Det ble derfor også inkludert en såkalt ikke-parametriske metode (Wilcoxon statistikk), hvor det ikke tas utgangspunkt i noen underliggende fordelingsfunksjon, (avsnitt 3.5.3). Disse metodene diskuteres i de neste avsnittene, og ytterligere detaljer finnes i vedlegg II.

3.5.1 TILPASNING TIL TO NORMALFORDELINGER

Modellen tar utgangspunkt i to normalfordelinger, en for de “tomme” rutene og en for rutene med “kalkkorn”. Hver av fordelingene er beskrevet av en middelværdi og et standardavvik. Innebygd i modellen er også konfidensersklene (tre stykker i dette eksperimentet – tilsvarende konfidensnivåene 4, 3 og 2). Middelværdien og standardavviket for den positive fordelingen ble vilkårlig valgt til 5,0 og 0,6. I selve tilpasningen

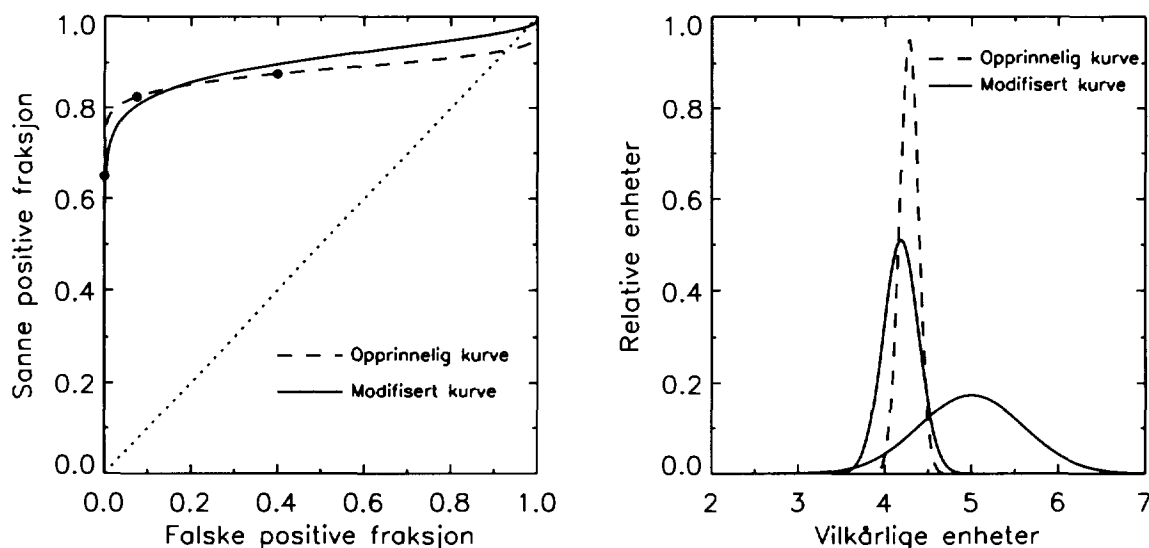


Figur 14. ROC-kurvene for to laboratorier, A og B. Til høyre vises fordelingsfunksjonene. Den positive fordelingen er vilkårlig valgt i modellen, og er den samme for alle laboratorier. Ved tilpasningen blir middelværdien og standardavviket for den negative fordelingen bestemt. ROC-kurven for laboratorium B ligger høyere enn kurven for laboratorium A. Dette illustreres også av at den negative fordelingen ligger lenger mot venstre for laboratorium B (bedre separasjon av de to fordelingene).

bestemmes middelveien og standardavviket for den negative fordelingen og posisjonen til konfidensersklene slik at arealene til høyre for hver terskel (for hver fordeling) blir nærmest mulig de observerte sanne og falske positive fraksjonene for hvert av konfidensnivåene. Figur 14 viser ROC-kurvene for to laboratorier med de tilhørende normalfordelingene.

Denne metoden viste seg å underestimere arealet under ROC-kurven i enkelte tilfeller. Figur 15 viser et eksempel på dette. ROC-kurvens stigningskoeffisient avtar etterhvert som kurven stiger fra nedre, venstre hjørne, for deretter å øke igjen når kurven nærmer seg øvre, høyre hjørne. Denne oppoverkrumningen (“kroken”) i øvre, høyre hjørne kan bli svært markant når antallet bilder er få. Tilpasningen kan i mange slike tilfeller synes god, da ROC-kurvens forløp stemmer bra overens med målepunktene. Målepunktene ligger imidlertid ofte langt mot venstre (lave falske positive verdier). “Kroken” ligger dermed utenfor det området som dekkes av målepunktene, og kan ses på som en ekstrapolasjonsartefakt. Når kurven er eksakt tilpasset målepunktene og samtidig danner en markert “krok” i øvre, høyre hjørne sier man at de eksperimentelle dataene er degenererte [7]. Arealet under kurven blir i slike tilfeller for lite. Dette illustrerer også et av ankepunktene mot å bruke arealet under kurven som et mål for totalytelse, da man vektlegger konfidensnivåer hvor man ikke ville operert i mange kliniske situasjoner (stor andel falske positive).

I situasjoner som beskrevet ovenfor, vil arealet under kurven bli for lite, og ofte også mindre enn arealet under en kurve gitt som rette linjer mellom målepunktene. Det ble derfor gjort en modifikasjon i programmet for å unngå disse spesielle kurveforløpene. Kriteriet som ble valgt var at ROC-kurven ikke skulle kunne krysse den positive diagonalen (prikket linje i figur 15). Dette oppnås ved å kreve at ROC-kurvens stigningskoeffisient ikke skal overstige 1,0 når den falske positive fraksjonen går mot 1,0. Overført til modellen med de to normalfordelingene vil en “krok” på ROC-kurven



Figur 15. Til høyre vises hvordan den negative fordelingen endres (større standardavvik) som følge av kriteriet om at ROC-kurven ikke skal krysse den positive diagonalen.

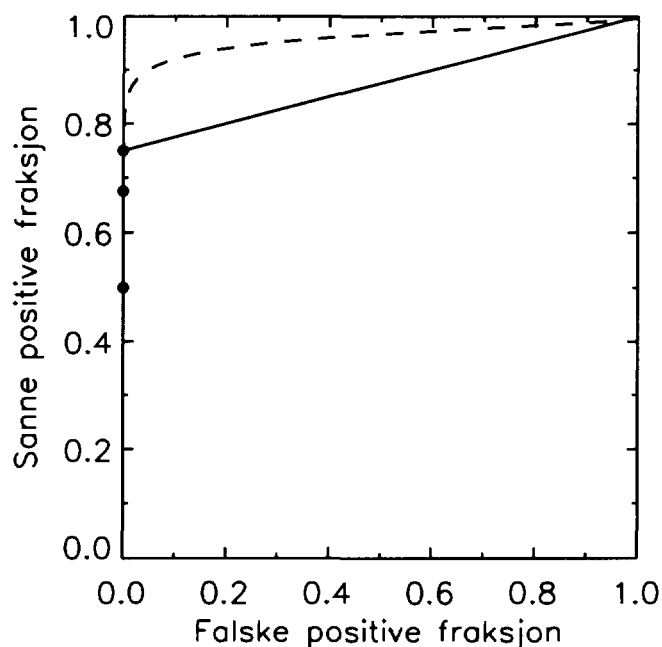
kjennetegnes av et lite standardavvik for den negative fordelingen. I tilfeller hvor stigningskoeffisienten overstiger 1,0, vil programmet iterativt øke standardavviket for den negative fordelingen inntil stigningskoeffisienten blir mindre enn 1,0. Figur 15 viser hvordan programmet på denne måten “tvinger” ROC-kurven ned ved lave falske positive fraksjoner og opp ved høye falske positive fraksjoner. Det er også vist hvordan den negative fordelingen endres for å oppnå dette (større standardavvik og større avstand mellom de to fordelingene).

En annen situasjon oppstår når radiologen ikke har utnyttet de tilgjengelige konfidensnivåene i tilstrekkelig grad. Dette kan resultere i at alle punktene blir liggende på ordinataksen, svarende til en falsk positiv fraksjon lik 0,0 (figur 16). I slike tilfeller er det ikke mulig å gjøre en fornuftig tilpasning av dataene til en ROC-kurve. Det ble i slike situasjoner valgt å bruke den kurven som dannes ved å trekke rette linjer mellom målepunktene (stiplet linje på figuren). Arealet under kurven som fås ved å trekke rette linjer mellom målepunktene er identisk med det som fås ved bruk av Wilcoxon statistikk (avsnitt 3.5.3).

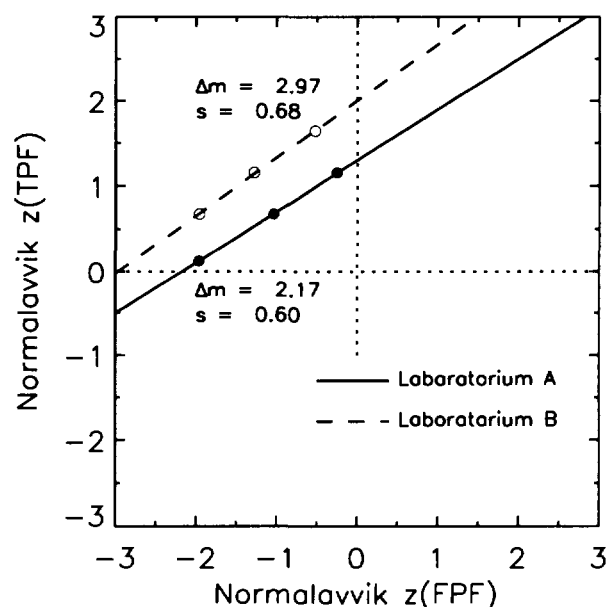
3.5.2 TILPASNING TIL RETT LINJE I NORMALAVVIKSKOORDINATER

I litteraturen er det beskrevet en metode for tilpasning som i prinsippet kan gjøres for hånd [6,17]. Denne metoden tar også utgangspunkt i en modell bestående av to normalfordelinger. I stedet for å plote dataene i et kartesisk koordinatsystem, plottes målepunktene i såkalte normalavvikskordinater (z-kordinater). Under forutsetning om at fordelingsfunksjonene er normalfordelte, vil målepunktene ligge langs en rett linje.

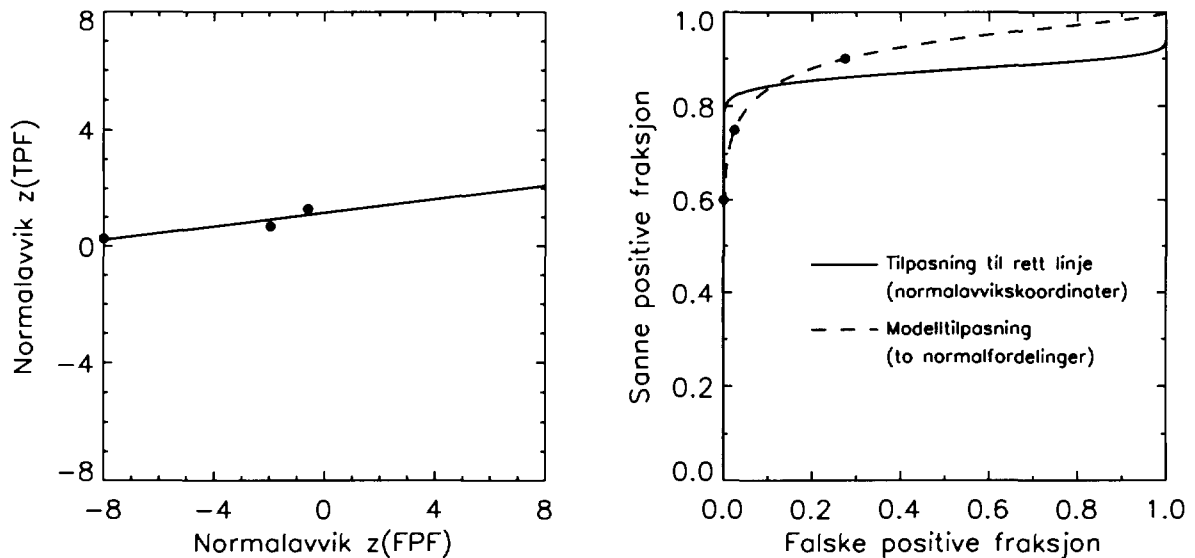
IDL inneholder en funksjon som kan benyttes for transformasjon av målepunktene til normalavvikskordinater, og



Figur 16. I dette tilfellet har ikke radiologen utnyttet konfidensnivåene i tilstrekkelig grad. ROC-kurven går gjennom alle målepunktene, men følger ordinataksen langt forbi det siste målepunktet. Arealet under kurven blir dermed betydelig større enn arealet som fås ved å trekke rette linjer mellom målepunktene.



Figur 17. ROC-kurver i normalavvikskordinater (samme datasett som i figur 14).



Figur 18. I situasjoner hvor målepunktene ikke blir liggende langs en rett linje i normalavviks-koordinater, vil ROC-kurven ikke gå gjennom målepunktene. Ved transformasjon tilbake til kartesiske koordinater vil ROC-kurven gå utenom målepunktene, og resultatet blir forskjellig fra det som fås ved tilpasning til to normalfordelinger.

også en funksjon for å finne den rette linjen som best stemmer overens med de tre målepunktene. Figur 17 viser en slik tilpasning for de to samme datasettene som i figur 14. Størrelsen Δ_m er absoluttverdien av avstanden mellom punktet $(0,0)$ og den rette linjens skjæringspunkt med linjen $z(\text{TPF}) = 0$ (prikket linje på figuren). Parameteren s er den rette linjens stigningskoeffisient. På grunnlag av Δ_m og s , kan arealet under ROC-kurven, A_z , beregnes.

Denne metoden hadde imidlertid en svakhet sammenlignet med metoden beskrevet i avsnitt 3.5.1. Figur 18 (til venstre) viser en ROC-kurve i normalavvikskoordinater. I dette tilfellet ligger ikke målepunktene langs en rett linje, og ROC-kurven går derfor heller ikke gjennom alle målepunktene. Ved transformasjon av ROC-kurven tilbake til kartesiske koordinater vil ROC-kurven gå utenom målepunktene også i dette koordinat-systemet (figur 18 til høyre, heltrukken linje), og arealet under kurven blir til dels betydelig underestimert. Den stiplede linjen viser resultatet fra tilpasningen beskrevet i avsnitt 3.5.1.

Problemet beskrevet i avsnittet over er spesielt fremtredende når antallet datapunkter er lite, noe som er tilfellet i denne undersøkelsen. Dette medfører for mange steder at resultatene blir forskjellig for de ulike metodene. I tillegg til problemet nevnt ovenfor, gjelder også de samme problemene som for metoden beskrevet i avsnitt 3.5.1.

3.5.3 IKKE-PARAMETRISK MODELL – WILCOXON STATISTIKK

For å kunne håndtere alle situasjoner, ble også en ikke-parametrisk modell anvendt. Denne baserer seg på Wilcoxon statistikk, som gir en parameter W som tilsvarer sannsynligheten for rett ranking i et 2-AFC eksperiment (2-AFC – two-alternative forced choice) [18,19,20]. Denne sannsynligheten er tilnærmet lik arealet under en ROC-kurve. Parameteren W , eller sannsynligheten for rett ranking i et 2-AFC eksperiment, tilsvarer

arealet under en kurve som består av rette linjer mellom målepunktene. Med få konfidensnivåer vil W vanligvis bli noe mindre enn arealet under ROC-kurven tilpasset ved de to andre metodene beskrevet ovenfor, og forskjellen avhenger bl.a. av plasseringen av det enkelte punkt. Denne metoden gir også muligheter for å estimere standardfeilen og dermed også konfidensintervallet for W . Dette kan brukes som et konservativt estimat også for de andre metodene diskutert i de foregående avsnittene. Hvis man ønsker å teste om en forskjell i areal under to bestemte ROC-kurver er signifikant, bør man ta hensyn til at bildene er basert på eksponeringer av identiske fantomkonfigurasjoner (brystvev og overlagringsplate), og at arealene til en viss grad derfor er korrelerte [21].

4 Resultater

Alle resultater finnes i vedleggene III, IV, V og VI. Det enkelte laboratorium er tildelt en to-bokstavers kode, og ved hjelp av denne kan det enkelte sted finne sine resultater og sammenligne med de øvrige laboratorier. Vedleggene inneholder også histogrammer som, der det er relevant, viser hvordan resultatene for de enkelte parametre fordelte seg. I avsnittene nedenunder er de viktigste resultatene for hvert av de fire fantomene omtalt.

4.1 Resultater for NRT-fantomet

Resultatene for NRT-fantomet finnes i vedlegg III. Tabellene III.1-3 inneholder alle resultatene for hvert sted for henholdsvis anvendt teknikk, referanseteknikk og forstørrelsesteknikk. Størrelsene C_{Al} , C_{Cu} og S/P er gitt som gjennomsnittet for 3 filmer. R_1 og R_2 er gitt som medianverdien av 3 eller 6 filmer for kontaktteteknikk (anvendt og referanseteknikk slått sammen) og som medianverdien av 3 filmer for forstørrelsesteknikk. CD-score representerer kontrast-detalj-oppløsningen, og verdien er gitt som gjennomsnittsverdien med tilhørende standardavvik for 3 filmer lest av 2 observatører. Figur III.1 inneholder histogrammer som viser fordelingen av de enkelte parametrene for hver av teknikkene.

Kontrastindeksene C_{Al} og C_{Cu} varierte med en faktor 3,5 og 2,0 mellom høyeste og laveste verdi. De 7 stedene som ikke benyttet raster hadde gjennomgående lav kontrast, med gjennomsnittsverdiene 0,060 for C_{Al} og 0,24 for C_{Cu} ved anvendt teknikk. Ved å sammenligne kun de stedene som benyttet rasterteknikk, finner man fortsatt betydelige variasjoner, og forholdet mellom høyeste og laveste verdi var henholdsvis 2,0 og 1,5. Kontrastindeksene er noe lavere ved forstørrelsesteknikk enn ved kontaktteteknikk, 0,085 mot 0,095 i gjennomsnitt for C_{Al} og 0,30 mot 0,33 for C_{Cu} for de laboratoriene som benyttet forstørrelsesteknikk.

I SIS Råd 1992:1 [22] er det et overordnet krav at oppløsningsevnen målt med linjetestraster midt i et 4,5 cm tykt perspeksfantom (som f.eks. NRT-fantomet) skal være minst 15 linjepar per millimeter (lp/mm). I europeiske retningslinjer [23] anses 12 lp/mm målt i midten av et 4 cm tykt fantom for å være akseptabelt, og 15 lp/mm for ønskelig (målt i et plan vinkelrett på røntgenrørets referanseakse). NRT-fantomet inneholder grupper med oppløsning 14,3 og 16,6 lp/mm, og vi har derfor valgt 14,3 lp/mm som nedre grenseverdi. Oppløsningsevnen var stort sett tilfredsstillende ved kontaktteteknikk (resultatene for anvendt teknikk og referanseteknikk sett under ett), mens mange steder (69 %) hadde noe lav oppløsningsevne ved forstørrelsesteknikk.

Forholdet mellom spredt stråling og primærstråling var vesentlig høyere ved anvendt teknikk og ved referanseteknikk for de syv laboratoriene som ikke benyttet rasterteknikk.

For kontrast-detalj-oppløsningen var det en faktor 3,0-3,5 mellom beste og dårligste verdi. Også for denne parameteren har de stedene som ikke benyttet raster lavere verdier enn de øvrige, og variasjonene er fortsatt store når kun de stedene som benyttet rasterteknikk sammenlignes.

| <i>Parameter</i> | <i>Anvendt</i> | <i>Referanse</i> | <i>Forstørrelse</i> |
|------------------|----------------|------------------|---------------------|
| C_{Al} | 0,084 | 0,084 | 0,073 |
| C_{Cu} | 0,31 | 0,32 | 0,28 |
| CD-score | 10,4 | 10,4 | |

Tabell 5. Nedre kvartil for enkelte av resultatene for NRT-fantomet.

For anvendt og referanseteknikk er verdiene beregnet kun for de stedene som benyttet rasterteknikk.

Tabell 5 viser den nedre kvartilen for kontrastindeksene C_{Al} og C_{Cu} , samt for kontrast-detalj-oppløsningen CD-score. Disse verdiene kan ses på som anbefalte minimumsverdier eller tiltaksgrenser. For de stedene som ligger under disse minimumsverdiene, bør det vurderes tiltak for å forbedre bildekvaliteten.

4.2 Resultater for RMI-fantomet

Resultatene for RMI-fantomet finnes i vedlegg IV. Tabell IV.1 inneholder resultatene for anvendt teknikk, mens tabell IV.2 inneholder resultatene for referanseteknikk. Figur IV.1 inneholder histogrammer som viser hvordan resultatene fordelte seg for hver type struktur i fantomet.

Hvert enkelt bilde er bedømt som enten godkjent eller ikke godkjent med utgangspunkt i de amerikanske kravene [11]. For fibrillene ble så mange som 35 % av bildene underkjent, mens det for kalk (10 %) og fortetninger (5 %) var vesentlig bedre.

I tillegg til dårligste og beste verdi for de tre filmene ved hver teknikk for hvert laboratorium, er det i vedlegget angitt om i) alle (+), ii) minst ett (?) eller iii) ingen (-) av bildene oppfylte de amerikanske kravene. Dette finnes oppsummert i tabell 6. Ved anvendt teknikk tilfredsstilte ingen av de tre bildene de amerikanske kravene for fibrillene ved 6 steder (13 %), mens ytterligere 21 laboratorier (47 %) hadde minst ett bilde som ikke ble bedømt som godkjent. For kalkgruppene var det bare ett sted hvor ingen av bildene tilfredsstilte kravene, og for samtlige steder var kravene for fortetningene oppfylt for minst ett av bildene. Alle steder hvor minst ett av bildene ikke ble godkjent i henhold til de amerikanske kravene for en av strukturene bør vurdere tiltak, og dette gjelder så mange som 28 laboratorier (62 %).

| <i>Type struktur</i> | <i>Anvendt teknikk</i> | | | <i>Referanseteknikk</i> | | |
|----------------------|------------------------|----------|----------|-------------------------|----------|----------|
| | <i>+</i> | <i>?</i> | <i>-</i> | <i>+</i> | <i>?</i> | <i>-</i> |
| <i>Fibriller</i> | 18 | 21 | 6 | 15 | 22 | 8 |
| <i>Kalk</i> | 39 | 5 | 1 | 35 | 7 | 3 |
| <i>Fortetninger</i> | 40 | 5 | 0 | 41 | 4 | 0 |

Tabell 6. Oppsummering av resultatene for RMI-fantomet.

4.3 Resultater for NRH-fantomet

Resultatene for NRH-fantomet finnes i vedlegg V. Tabellene V.1-3 inneholder alle resultatene for anvendt teknikk, referanseteknikk og forstørrelseteknikk. Tabellene

inneholder både resultatet for hvert enkelt sted sammenlignet med et utvalgt referanse-bilde og eventuelle anmerkninger til det enkelte bildet. Figur V.1 inneholder histogrammer som viser hvordan resultatene av vurderingen av kontrast og skarphet fordelte seg for hver av teknikkene.

Kontrasten ble ved anvendt teknikk vurdert som betydelig dårligere enn referansebildet (score 1 eller 2) for 4 laboratorier (9 %), og skarpheten ble vurdert som betydelig dårligere enn referansebildet for 8 laboratorier (18 %). Ved forstørrelsesteknikk gjelder tilsvarende for 5 laboratorier (14 %) for både kontrasten og skarpheten. Ved kontaktteknikk er det hovedsakelig de laboratoriene som ikke benyttet rasterteknikk som har fått score 1 eller 2.

4.4 Resultater for NRH-ROC-fantomet

En av forutsetningene for at dette ROC-eksperimentet skulle kunne lykkes, var at de strukturene som simulerte kalk varierte i størrelse og tetthet fra det alle kunne detektere med suboptimal bildekvalitet til det ingen kunne detektere selv med optimal bildekvalitet. Detekterbarheten kan defineres som laboratorienes evne til å oppdage (ved å sette score 3 eller 4) at det forelå en gruppe "kalkkorn" i en bestemt rute. Tabell 8 viser detekterbarheten som funksjon av "kalkkornenes" diameter og relative tetthet ved kontaktteknikk, mens tabell 7 viser de tilsvarende tallene for forstørrelsesteknikk.

Resultatene for NRH-ROC-fantomet finnes i vedlegg VI, og tabell VI.1 inneholder antallet sanne og falske positive for hvert av de fire konfidensnivåene, både for anvendt teknikk og for forstørrelsesteknikk. Tabellene VI.2 og VI.3 inneholder resultatene av tilpasningene til ROC-kurver for alle de tre metodene beskrevet under avsnitt 3.5. For tilpasning til to normalfordelinger er parametrene μ_N og σ_N oppgitt sammen med arealet A under den resulterende ROC-kurven. For den ikke-parametriske metoden (Wilcoxon) er størrelsene W og SE(W) oppgitt. For tilpasningen til rett linje i normalavviks-

| "Kalkkornenes" diameter [mm] | "Kalkkornenes" relative tetthet | | | | | | | | | | | | | |
|---------------------------------|---------------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 24 | 32 | 40 | |
| 0,1 | | | | | | | | | | | | 24 | 36 | 39 |
| 0,2 | | | 8 | 7 | 21 | 30 | 38 | 40 | 40 | 44 | | | | |
| 0,3 | 7 | 9 | 43 | 43 | 35 | 45 | 45 | 45 | | | | | | |
| 0,4 | 15 | 33 | 43 | 45 | 44 | 45 | | | | | | | | |

Tabell 8. Detekterbarheten, definert som antallet laboratorier som satte score 3 eller 4, ved anvendt teknikk (antall laboratorier = 45).

| "Kalkkornenes" diameter [mm] | "Kalkkornenes" relative tetthet | | | | | | | | | | | | | |
|---------------------------------|---------------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 24 | 32 | 40 | |
| 0,1 | | | | | | | | | | | | 32 | 35 | 34 |
| 0,2 | | | 21 | 11 | 33 | 33 | 34 | 35 | 34 | 35 | | | | |
| 0,3 | 8 | 10 | 35 | 35 | 31 | 35 | 35 | 35 | | | | | | |
| 0,4 | 10 | 30 | 35 | 35 | 35 | 35 | | | | | | | | |

Tabell 7. Detekterbarheten, definert som antallet laboratorier som satte score 3 eller 4, ved forstørrelsesteknikk (antall laboratorier = 35).

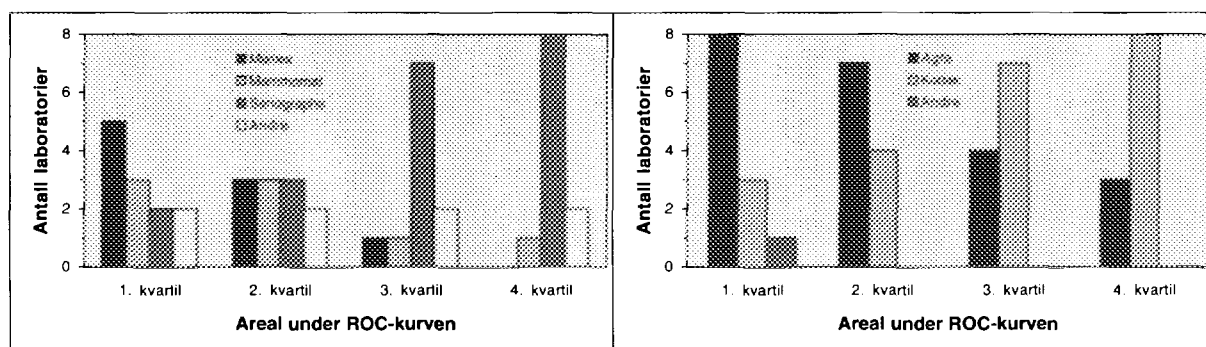
koordinater er parametrene Δm og s oppgitt sammen med arealet A_z . I tillegg er svertningen og to verdier for kontrasten i bildene oppgitt. Figurene VI.1-15 inneholder ROC-kurvene for alle laboratoriene for både anvendt og forstørrelsesteknikk.

Resultatene viser store variasjoner mellom de ulike laboratoriene. Ved enkelte laboratorier hadde ikke radiologene benyttet de ulike konfidensnivåene i tilstrekkelig grad. Dette resulterte i noen tilfeller i degenererte datasett, som ikke lot seg tilpasse til en "meningsfull" ROC-kurve. For disse stedene ble i stedet parameteren W benyttet som et mål for arealet under kurven.

Tilpasningen til to normalfordelinger resulterte i uforholdsmessig store areal under de resulterende ROC-kurvene for laboratoriene BD og FC ved kontaktteknikk og laboratorium BD ved forstørrelsesteknikk på grunn av degenererte datasett (ingen falske positive for noen av konfidensnivåene), og parameteren W fra den ikke-parametriske metoden må i stedet benyttes som mål for arealet under kurven. Det samme gjelder for laboratoriene DD, EA, EC og FA ved kontaktteknikk og laboratorium DC ved forstørrelsesteknikk, hvor tilpasningen til to normalfordelinger ikke lot seg gjennomføre overhodet. For de øvrige laboratoriene resulterte tilpasningen til to normalfordelinger i meningsfulle ROC-kurver. Imidlertid kan det synes som om enkelte laboratorier får "for stort" areal under ROC-kurven på bakgrunn av kurvens forløp i forhold til de tre målepunktene. Dette gjelder bl.a. laboratorium FF ved kontaktteknikk og laboratoriene AC og DD ved forstørrelsesteknikk. Dette ses også ved at arealet under kurven blir betydelig større enn parameteren W . Det finnes foreløpig ingen kriterier som beskriver godheten eller konfidensintervallet for tilpasningen til to normalfordelinger, og som kunne indikert om det resulterende arealet under ROC-kurven er rimelig eller ikke.

Nesten samtlige laboratorier oppnådde et betydelig større areal under ROC-kurven ved forstørrelsesteknikk sammenlignet med kontaktteknikk. For 23 av de 35 laboratoriene som benyttet forstørrelsesteknikk var forbedringen signifikant ($p < 0,05$). Det er også interessant å merke seg at både sensitiviteten og spesifisiteten bedres. Dette betyr at man oppdager flere "kalkgrupper", men også at man blir sikrere i "diagnostiseringen".

Arealet under ROC-kurven varierte betydelig for de ulike apparattypene og film-folie-kombinasjonene som ble omfattet av undersøkelsen. Figur 19 til venstre viser resultatet (arealet under ROC-kurven) for de ulike apparatene, og det ses at blant Mamex'ene og



Figur 19. Histogrammer som viser hvordan arealet under ROC-kurven for de ulike apparattypene og filmleverandørene fordelte seg i de fire kvartilene.

Mammomat'ene ligger størsteparten i den nedre halvdel, mens Senographe'ene hovedsakelig ligger i den øvre halvdel. Figur 19 til høyre viser tilsvarende for de ulike filmleverandørene, og det ses at Agfas film-folie-kombinasjoner kommer vesentlig dårligere ut enn Kodaks.

5 Diskusjon

Hovedmålet med dette prosjektet var å teste ut en metode for vurdering av totalytelse, og å benytte denne i en sammenlignende undersøkelse av norske mammografi-laboratorier. Det ble samtidig gjort en kartlegging av bildekvaliteten til det enkelte laboratorium. Diskusjonen består av tre hovedelementer: i) kartleggingen av bildekvalitet, ii) metoden for vurdering av totalytelse ved mammografivirksomhet og til slutt iii) resultatene av den sammenlignende undersøkelsen av totalytelse.

5.1 Kartlegging av bildekvalitet

Det ble funnet store variasjoner i bildekvaliteten mellom de forskjellige laboratoriene, uavhengig av hvilket fantom som ble benyttet. Det finnes kun unntaksvis anbefalinger om grenseverdier. For NRT-fantomet kan det kun knyttes en anbefalt grenseverdi til oppløsningsevnen. Vi har derfor foreslått nedre kvartil som tiltaksgrense for de øvrige parametrene. Denne kvartilen er imidlertid beregnet kun på grunnlag av de laboratorier som benyttet rasterteknikk. For denne typen fantomer forekommer det alltid små variasjoner fra fantom til fantom. Dette betyr at de foreslåtte tiltaksgrensene for NRT-fantomet kun gjelder for det bestemte eksemplaret som ble benyttet i dette prosjektet.

For kontrastindeksene C_{Al} og C_{Cu} , samt for lavkontrastoppløsningen CD-score, er det store forskjeller mellom de steder som benyttet raster kontra de som ikke benyttet raster. Man kan vurdere å innføre ulike tiltaksgrenser for rasterteknikk og for de som ikke benytter raster [24], noe som kan forsvares med at også dosebelastningen vil være svært forskjellig for de to situasjonene [1]. Vi har imidlertid valgt å sette kun en tiltaksgrense, og denne blir dermed gjeldende for alle uansett valg av teknikk. Begrunnelsen for dette er at de laboratorier som ikke benytter rasterteknikk, gjør dette av tekniske årsaker. Dette kan være at apparatet ikke kan utstyres med raster, eller at apparatet ikke har høyt nok stråleutbytte til at raster skal kunne brukes. Da tekniske årsaker, og ikke reduksjon av stråledosene er årsaken til at raster ikke benyttes, kan vi forsvare bruken av kun en grenseverdi. Det kan også stilles spørsmålsteget om bildekvaliteten er tilfredsstillende for de steder som ikke benytter rasterteknikk. Bildene av NRH-fantomet fikk score 1 eller 2 (viktig diagnostisk informasjon mistet) for seks av de syv stedene som ikke benyttet raster, mens kun to av de resterende 38 laboratoriene fikk tilsvarende lav score. De mange likhetstrekkene mellom bildene av NRH-fantomet og kliniske bilder, og det at en erfaren radiolog stod for vurderingen av bildene, gjør at resultatene for NRH-fantomet tillegges betydelig vekt. Denne vurderingen viser helt tydelig at rasterteknikk må benyttes i mammografi for at tilfredsstillende bildekvalitet skal kunne oppnås.

For RMI-fantomet finnes det faste grenseverdier fra USA, der dette fantomet inngår i godkjenningen av mammografilaboratoriene. Grenseverdiene er fremkommet ved korrelasjon mellom fremstillingen av de enkelte strukturer i fantomet og radiologers subjektive vurdering av bildekvaliteten basert på et stort antall kliniske bilder, og må ses på som et absolutt minimumskrav. Når da kun 17 laboratorier (38 %) får alle tre bildene bedømt som godkjent etter dette opplegget, kan man sette et spørsmålsteget ved bildekvaliteten ved norske mammografilaboratorier. Til sammenligning kan det nevnes at 11 % av de amerikanske mammografilaboratoriene har fått underkjent disse fantombildene [11].

5.2 Metode for vurdering av totalytelse

Resultatene for NRH-ROC-fantomet viser at eksperimentene stort sett har vært vellykket. Størsteparten av radiologene benyttet de fire konfidensnivåene tilfredsstillende, og det var kun unntaksvis problemer med å tilpasse dataene til den valgte modellen. Tabell 8 viser hvor mange observatører som detekterte de enkelte “kalkgruppene” ved anvendt teknikk, og for hver diameter (rad i tabellen) ser man at samtlige eller nesten samtlige detekterte (score 3 eller 4) “kalkgruppene” med de høyeste tetthetene (kolonnene mot høyre i tabellen). Man ser videre at det var betydelig færre som detekterte “kalkgruppene” med de laveste tetthetene. Arealet under kurven var gjennomgående høyt (0,879 i gjennomsnitt for kontaktteknikk), noe som indikerer at flertallet av “kalkgruppene” var for lette å detektere sett fra et metodesynspunkt.

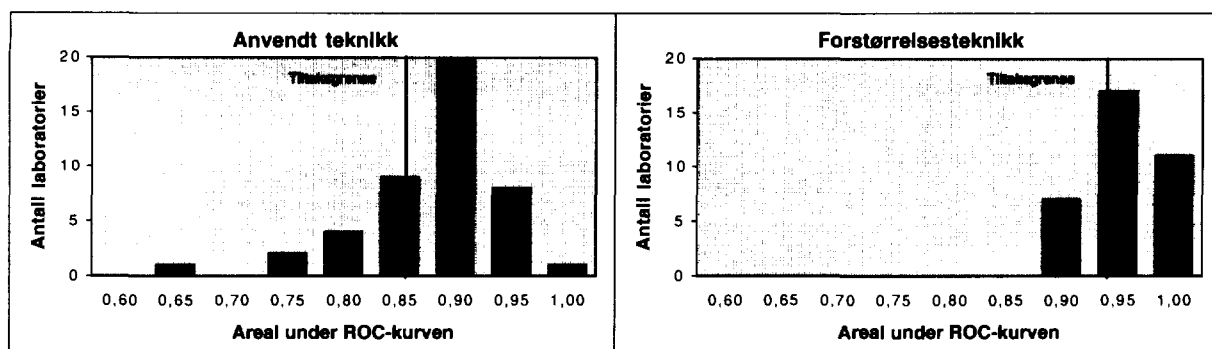
Det finnes ingen endelige optimale verdier for hvor stort arealet under en ROC-kurve bør være, men litteraturen antyder et tyngdepunkt rundt 0,75-0,80 [25]. Spesielt for “kalkkornene” med diameter 0,3 og 0,4 mm ble svært mange av gruppene detektert av størstedelen av laboratoriene. Dette gjør metoden mindre følsom for variasjoner mellom laboratoriene. I denne undersøkelsen lå arealet under kurven mellom 0,658 og 0,978 ved kontaktteknikk. Hvis man bruker SE(W) som et anslag for standardfeilen, og har som nullhypotese at arealene under ROC-kurvene skal være like, er det svært få av forskjellene som er signifikante. En riktigere måte for å vurdere signifikans fordrer et mer nøyaktig estimat av standardfeilen, ved bl.a. å ta hensyn til at det finnes en viss korrelasjon mellom datasettene.

Den forbedringen i detekterbarheten av “forkalkninger” som finner sted ved forstørrelsesteknikk sammenlignet med kontaktteknikk, viser også at funnene ved bruk av metoden stemmer overens med klinisk erfaring. Det kan forventes at forskjellene ville vært enda større hvis flere av “kalkgruppene” hadde vært vanskeligere å detektere.

5.3 Sammenlignende undersøkelse av totalytelse

Det var store forskjeller i totalytelse mellom de ulike laboratoriene (figur 20), med 0,658 og 0,978 som minste og største verdi for arealet under ROC-kurven ved anvendt teknikk. Forskjellene var vesentlig mindre ved forstørrelsesteknikk, noe som virker rimelig med tanke på de relativt høye verdiene for arealet under ROC-kurven.

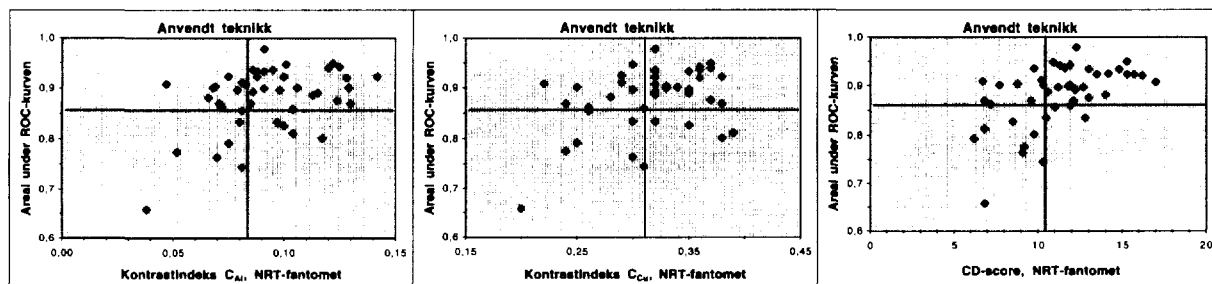
Vedlegg VII oppsummerer alle resultatene for vurderingene av bildekvalitet og totalytelse. For parametrene C_{Al} , C_{Cu} og CD-score (NRT-fantomet) er laboratorier med verdier under tiltaksgrensene (nedre kvartil for de laboratorier som benyttet rasterteknikk) angitt med et “÷”-tegn. For R_1 og R_2 er 14,3 lp/mm brukt som nedre grenseverdi. For RMI-fantomet er laboratorier hvor ingen av tre bilder oppfylte de amerikanske kravene angitt med et “÷”-tegn. For NRH-fantomet er laboratorier med score 1 og 2 angitt. For NRH-ROC-fantomet er nedre kvartil for arealet under ROC-kurven brukt som tiltaksgrense (0,859 ved anvendt teknikk og 0,939 ved forstørrelsesteknikk), og det er angitt hvilke laboratorier som oppnådde et areal under kurven som var mindre enn disse verdiene. Fra tabellene VII.1 og VII.2 ses det at laboratorier med dårlig totalytelse også gjennomgående hadde dårlig bildekvalitet. I de fleste tilfeller kan derfor dårlig totalytelse forklares av dårlig bildekvalitet. I noen få tilfeller kan den dårlige



Figur 20. Histogrammer som viser hvordan arealet under ROC-kurven fordelte seg for anvendt og forstørrelsesteknikk. Tiltaksgrensene er definert som den nedre kvartilen.

totalytelsen kun forklares av granskningsforholdene eller radiologen som deltok i undersøkelsen. Figur 21 viser arealet under ROC-kurven som funksjon av NRT-parametrene C_{Al} , C_{Cu} og CD-score ved anvendt teknikk. De horisontale og vertikale linjene representerer tiltaksgrensene for de enkelte parametrene. Det ses en klar sammenheng mellom areal under ROC-kurven og NRT-parametrene, og at de fleste laboratorier med dårlig totalytelse også har dårlig kontrast eller kontrast-detalj-oppløsning. Enkelte laboratorier med dårlig kontrast eller kontrast-detalj-oppløsning har likevel tilfredsstillende totalytelse. Dette kan skyldes at radiologen har vent seg til den dårlige bildekvaliteten, og at dette kompenseres ved bedre granskningsforhold eller granskningsteknikk (f.eks. bruk av Matsson-kikkert) eller høyere filmsvertning.

Det fremgår av figur 20 hvordan arealet under ROC-kurven øker ved bruk av forstørrelsesteknikk. Ved anvendt teknikk var arealet under kurven 0,879 i gjennomsnitt, mens den tilsvarende verdien ved forstørrelsesteknikk var 0,956. Den økte sensitiviteten i deteksjonen av "kalkgruppene" ved bruk av forstørrelsesteknikk, og kanskje enda viktigere den økte spesifisiteten, tilsier at det er viktig å benytte forstørrelsesteknikk ved utredning av suspekte funn. Det er derfor betenkelig at forstørrelsesteknikk ikke var en del av den kliniske undersøkelsen ved alle sykehus, og at to sykehus totalutredet pasienter uten disse viktige tillegg bildene. Dette kan nok skyldes varierende kompetansenivå blant radiologene, noe uttalelsen "Vi trenger ikke forstørrelsesbilder, vi bruker forstørrelsesglass i stedet." viser.



Figur 21. Figurene viser arealet under ROC-kurven som funksjon av kontrastindeksene C_{Al} og C_{Cu} og kontrast-detalj-oppløsningen CD-score. De horisontale og vertikale linjene viser nedre kvartil for hver av parametrene. Man ser at laboratorier med lite areal under ROC-kurven gjennomgående har dårlig kontrast og kontrast-detalj-oppløsning.

Det var relativt mange tilfeller hvor antallet sanne eller falske positive for et eller flere av konfidensnivåene var lik null. Dette kan skyldes det relativt lave antallet “kalkgrupper” (40), eller at det var for få “kalkgrupper” som representerte noen reell utfordring i deteksjonsprosessen (omlag halvparten av “kalkgruppene” ble detektert i minst 90 % av tilfellene). Det kan også synes som om antallet konfidensnivåer (4) ikke var optimalt, og det ville kanskje vært enklere for observatørene å gradere sikkerheten/usikkerheten i deteksjonsprosessen med 5 nivåer. Flere konfidensnivåer gir også et bedre grunnlag for å bestemme ROC-kurven, og modelltilpasningen ville kanskje fungert enda bedre. Det vil derfor være naturlig å benytte flere overlageringsplater i nye eksperimenter, og samtidig benytte minst 5 konfidensnivåer. Det er også viktig at den generelle vanskelighetsgraden øker, slik at observatørene vil oppleve flere av rutene som usikre.

Både svertningen og kontrasten for bildene av NRH-ROC-fantomet er oppgitt i vedlegg VI, men disse størrelsene vil ikke bli diskutert i denne rapporten.

6 Konklusjoner

Det ble funnet store variasjoner i bildekvaliteten, og en uforholdsmessig stor andel av laboratoriene (62 %) hadde varierende eller for dårlig bildekvalitet vurdert etter de amerikanske kriteriene.

De laboratoriene som ikke benyttet rasterteknikk oppnådde betydelig dårligere resultater enn de som brukte raster, og rasterteknikk må derfor anses som nødvendig for å kunne oppnå tilfredsstillende bildekvalitet i mammografi.

Fantomet utviklet spesielt for å kunne gjennomføre ROC-eksperimenter viste seg godt egnet til oppgaven. Antallet bilder (antallet ruter med og uten "kalkgrupper") bør imidlertid økes. En større andel av "kalkgruppene" i nye overlageringsplater bør ha vanskelighetsgrad (diameter og tetthet) slik at observatørene opplever disse som usikre "funn".

Antallet konfidensnivåer (4) var for lite, og i fremtidige eksperimenter bør det brukes minst 5 ulike nivåer. Dette samsvarer bedre med det som er vanlig i slike eksperimenter, og vil sannsynligvis også styrke modelltilpasningen.

Programmet utviklet for modelltilpasning av dataene og for beregning av ROC-kurvene fungerte tilfredsstillende. Det må imidlertid utvikles en metodikk for å estimere standardavviket og konfidensintervallet for arealet under ROC-kurven.

Det ble funnet store forskjeller i totalytelse (areal under ROC-kurven), og dårlig totalytelse samsvarte ofte med utilfredsstillende bildekvalitet. Det forekom imidlertid også tilfeller hvor totalytelsen åpenbart var avhengig av betrakningsforholdene eller observatøren, da observatøren til tross for tilfredsstillende bildekvalitet ikke klarte å oppdage tilstrekkelig mange "kalkgrupper".

Det var en betydelig forbedring i totalytelse med bruk av forstørrelsesteknikk sammenlignet med kontaktteknikk. Forstørrelsesteknikk må derfor være en mulig opsjon ved klinisk mammografi, noe som dessverre ikke var tilfelle for to av laboratoriene som deltok i denne undersøkelsen.

Referanser

1. Widmark A, Olsen JB. Mammografivirkosomhet i Norge: tekniske resultater. StrålevernRapport 1995:5. Østerås: Statens strålevern, 1995.
2. White DR. The specifications of 25 phantoms presented at the Würzburg workshop (June, 1992). Radiation Protection Dosimetry 1993; 49: 371-386.
3. Olsen JB, Sager EM. Subjective evaluation of image quality based on images obtained with a breast tissue phantom: comparison with a conventional image quality phantom. British Journal of Radiology 1995; 68: 160-164.
4. Olsen JB, Sager EM, Skretting A. Applications of a mammary tissue phantom for image quality assessment in mammography. (Abstract of scientific paper 0867C at 78th Scientific Assembly and Annual Meeting of Radiological Society of North America, RSNA) Radiology 1992; 185 (P) Supplement: 250.
5. Olsen JB, Sager EM, Skretting A. Applications of a mammary tissue phantom for image quality assessment in mammography. (Scientific exhibit, SPACE 10-036 (poster) at 78th Scientific Assembly and Annual Meeting of Radiological Society of North America, RSNA) Radiology 1992; 185 (P) Supplement: 395.
6. Metz CE. ROC methodology in radiologic imaging. Investigative Radiology 1986; 21: 720-733.
7. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Investigative Radiology 1989; 24: 234-245.
8. Vining DJ, Gladish GW. Receiver operating characteristic curves: a basic understanding. RadioGraphics 1992; 12: 1147-1154.
9. Nielsen B. Image-quality optimization in mammography using the CGR test phantom. I: Optimization of image quality and patient exposure in diagnostic radiology. Moores BM, Wall BF, Eriskat H, Schibilla H, red. BIR Report 20. London: British Institute of Radiology, 1989: 60-63.
10. Barnes GT, Hendrick RE. Mammography accreditation and equipment performance. RadioGraphics 1994; 14: 129-138.
11. Conway BJ, Suleiman OH, Rueter FG, Antonsen RG, Slayton RJ. National survey of mammographic facilities in 1985, 1988, and 1992. Radiology 1994; 191: 323-330.
12. Dorfman DD, Alf E. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals - rating-method data. Journal of Mathematical Psychology 1969; 6: 487-496.
13. Swets JA, Pickett RM. Evaluation of diagnostic systems: Methods from signal detection theory. New York: Academic Press, 1982.
14. Metz CE. FORTRAN-programmene ROCFIT, CORROC og ROCPWR. Tilgjengelig fra C Metz, Department of Radiology, University of Chicago, Chicago, IL.
15. Vida S. A computer program for non-parametric receiver operating characteristic analysis. Computer Methods and Programs in Biomedicine 1993; 40: 95-101.
16. Hanley JA. The robustness of the "binormal" assumptions used in fitting ROC curves. Medical Decision Making 1988; 8: 197-203.
17. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. Investigative Radiology 1979;14: 109-121.

18. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* 1975; 12: 387-415.
19. Burgess AE. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. *Medical Physics* 1995; 22: 643-655.
20. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29-36.
21. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148: 839-843.
22. Statens strålevern. Kvalitetssikring i mammografi. SIS Råd 1992:1. Østerås: Statens strålevern, 1992.
23. Thijssen MAO, Young KC, van Woudenberg S. The european protocol for the quality control of the physical and technical aspects of mammography screening. I: European guidelines for quality assurance in mammography screening. 2nd edition. de Wolf CMJ, Perry NM, red. Luxembourg: Office for Official Publications of the European Communities, 1996: vedlegg II-D.
24. Olerud HM, Olsen JB, Widmark A, Fosmark H. A Norwegian survey of image quality, doses and film processing in mammography, with reference to two technical phantoms. *Radiation Protection Dosimetry* 1996; 67: 199-210.
25. Metz CE. Current problems in ROC analysis. I: Proceedings of the Chest Imaging Conference 1987. Peppler WW, Alter AA, red. Madison: Medical Physics Publications, 1988: 315-336.

Vedlegg I. Sykehus/institutter som deltok i undersøkelsen

| <i>Sykehus/institutt</i> | <i>Mammografiapparat</i> | <i>Film</i> |
|--------------------------------------|----------------------------|---------------------|
| Aust-Agder Sentralsykehus | Siemens Mammomat 2 | Agfa MR 6 |
| Beenfeldts Røntgeninstitut | Siemens Mammomat 2 | Kodak Min-RE |
| Brystinstituttet A/S, Gjøvik | General Medical Merate MSM | Kodak Min-RE |
| Buskerud Sentralsykehus | GE/CGR Senographe DMR | Kodak Min-RE |
| Det Norske Radiumhospital | GE/CGR Senographe 600 T | Kodak Min-RH |
| Det Norske Radiumhospital | GE/CGR Senographe DMR | Kodak Min-RH |
| Dr. Smith's Røntgeninstitut | Soredex Mamex DC S | Agfa MR 3-II |
| Dr. Willes Legesenter | Xerox 116 | Kodak Min-RH |
| Drammen Røntgeninstitut A/S | Siemens Mammomat 2 | Agfa MR 3-II |
| Drammen Røntgeninstitut A/S | Siemens Mammomat 2 | Agfa MR 5 |
| Fredrikstad Røntgeninstitut | GE/CGR Senographe 500 T | Agfa MR 3-II |
| Fylkessjukehuset i Florø | Soredex Mamex DC Mag | Agfa MR 3-II |
| Fylkessjukehuset i Molde | GE/CGR Senographe DMR | Agfa MR 5 |
| Halden Sykehus | GE/CGR Senographe 600 T | Kodak Min-RH |
| Harstad Sykehus | Soredex Mamex DC Mag | Agfa MR 6 |
| Haukeland Sykehus | Siemens Mammomat 2 | Kodak Min-RE |
| Hedmark Sentralsjukehus | GE/CGR Senographe 600 T | Kodak Min-RE |
| Medi 3 | Philips Mammo Diagnost UM | Agfa MR 6 |
| Namdal Sykehus | Soredex Mamex DC Mag | Agfa MR 3-II |
| Nordland Sentralsykehus | GE/CGR Senographe 600 T | Kodak Min-RE Impr |
| Norsk Institutt for Brystdiagnostikk | General Medical Merate MSM | Kodak Min-RH |
| Nye Bergen Røntgeninstitut | Siemens Mammomat B | Agfa MR 3-II |
| Orkdal Sanitetsforenings Sykehus | Soredex Mamex DC Mag | Kodak Min-RE |
| Oslo Akutten | Siemens Mammomat B | Agfa MR 5 |
| Regionsykehuset i Tromsø | GE/CGR Senographe 600 T | Agfa MR 3-II |
| Regionsykehuset i Trondheim | Soredex Mamex DC Mag | Agfa MR 6 |
| Rjukan Sykehus | GE/CGR Senographe 600 T | Agfa MR 3-II |
| Røntgensenterer i Kristiansand | GE/CGR Senographe DMR | Kodak Min-RH |
| Røntgensenteret i Kristiansand | Xerox 116 | Kodak Min-RH |
| Sandnessjøen Sykehus | GE/CGR Senographe DMR | Kodak Min-RH |
| Sentralsjukehuset i Møre og Romsdal | Philips Mammo Diagnost UC | Kodak Min-RH |
| Sentralsjukehuset i Rogaland | GE/CGR Senographe DMR | Kodak Min-RH |
| Sentralsjukehuset i Sogn og Fjordane | GE/CGR Senographe 600 T | Agfa MR 5 |
| Sentrum Røntgeninstitut | GE/CGR Senographe DMR | Agfa MR 3-II |
| Sentrum Røntgeninstitut | Siemens Mammomat 300 | Agfa MR 6 |
| Sentrum Røntgeninstitut | Soredex Mamex DC | Agfa MR 3-II |
| Sentrum Røntgeninstitut | Soredex Mamex DC Ami | Agfa MR 3-II |
| Spesialistgruppen | GE/CGR Senographe DMR | Kodak Min-RH |
| Telemark Sentralsykehus, Porsgrunn | GE/CGR Senographe 600 T | Kodak Min-RH |
| Tønsberg Røntgeninstitut | GE/CGR Senographe 500 T | Agfa MR 6 |
| Ullevål Sykehus | GE/CGR Senographe 600 T | Kodak Min-RH |
| Vest-Agder Sentralsykehus | Soredex Mamex DC Mag | Agfa MR 6 |
| Volvat Mammografen | General Medical Merate MSM | Du Pont MicroVision |
| Volvat Medisinske Senter | GE/CGR Senographe DMR | Kodak Min-RH |
| Volvat Medisinske Senter | General Medical Merate MSM | Kodak Min-RH |

Vedlegg II. Matematisk tilpasning av ROC-kurvene

II.1 Tilpasning til to normalfordelinger

En normalfordeling er definert som

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, (\sigma > 0). \quad (1)$$

Dette er en symmetrisk fordeling med middelværdi μ og standardavvik σ . Den kumulative normalfordelingen er gitt ved

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv. \quad (2)$$

I tabeller og numeriske beregninger tar man gjerne utgangspunkt i en normalfordeling med middelværdi lik 0 og standardavvik lik 1:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du. \quad (3)$$

Ved å innføre $u = \frac{v-\mu}{\sigma}$, får vi at $dv = \sigma du$. Den øvre integrasjonsgrensen blir da $z = \frac{x-\mu}{\sigma}$. Den kumulative normalfordelingen for en fordeling med vilkårlig middelværdi og standardavvik kan da skrives som

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{(x-\mu)/\sigma}{} e^{-\frac{u^2}{2}} du = \Phi\left(\frac{x-\mu}{\sigma}\right). \quad (4)$$

Verdien av $F(x)$ for vilkårlige verdier av μ og σ kan da enkelt beregnes eller finnes fra tabeller.

Modelltilpasning

For ROC-eksperimentene kan vi anta at fordelingen av de positive (P) og de negative (N) beskrives ved to normalfordelinger, den positive fordelingen $f_P(x)$ med middelværdi $\mu = \mu_P$ og standardavvik $\sigma = \sigma_P$, og den negative fordelingen $f_N(x)$ med middelværdi $\mu = \mu_N$ og standardavvik $\sigma = \sigma_N$. Størrelsene TPF_i og FPF_i representerer henholdsvis den sanne positive fraksjon og den falske positive fraksjon for scores større eller lik konfidensnivået i . Etter definisjonen er da

$$TPF_i^{\text{modell}}(x_i, \mu_P, \sigma_P) = 1,0 - \Phi\left(\frac{x_i - \mu_P}{\sigma_P}\right) \text{ og} \quad (5)$$

$$FPF_i^{\text{modell}}(x_i, \mu_N, \sigma_N) = 1,0 - \Phi\left(\frac{x_i - \mu_N}{\sigma_N}\right), \quad (6)$$

der x_i er de ukjente konfidenstersklene som er knyttet opp mot hvert konfidensnivå. I dette eksperimentet ble det benyttet fire konfidensnivåer, og som beskrevet i kapittel 2 gir ikke l'erne noen informasjon om kurvens forløp (representerer punktet med sann og falsk positiv fraksjon lik 1,0). Det er derfor kun konfidensnivåene 4, 3 og 2 som bestemmer kurven, som alltid starter i punktet (0, 0) og ender i punktet (1, 1). Skalaen for abscisseaksen (figur 1) er ikke relatert til noen absolutt størrelse, og det er dermed mulig å velge to av verdiene (bestemme relativ målestokk). Vi har vilkårlig valgt $\mu_P = 5,0$ og $\sigma_P = 0,6$, og valget av disse har ingen betydning for resultatene. Modellen inneholder dermed fem ukjente verdier: x_4, x_3, x_2, μ_N og σ_N . Samtidig kjenner vi seks verdier, de sanne positive fraksjonene $TPF_{4,3,2}^{\text{observert}}$ og de falske positive fraksjonene $FPF_{4,3,2}^{\text{observert}}$. De observerte verdiene for konfidensnivået x_4 får vi ved først å telle opp antallet riktig plasserte 4'ere og antallet feil plasserte 4'ere. Neste sett fås ved å telle opp antallet riktig

plasserte 4'ere og 3'ere og antallet feil plasserte 4'ere og 3'ere. Det siste settet fås ved å telle opp antallet riktig plasserte 4'ere, 3'ere og 2'ere og antallet feil plasserte 4'ere, 3'ere og 2'ere. Deretter deles alle disse tallene på henholdsvis det totale antallet ruter med "kalk" og på det totale antallet ruter uten "kalk" for å finne fraksjonene $TPF_i^{\text{observert}}$ og $FPP_i^{\text{observert}}$.

Tilpasningen går ut på å bestemme de verdiene av for x_4, x_3, x_2, μ_N og σ_N som minimaliserer kvadratavviket mellom de observerte sanne og falske positive fraksjonene og de tilsvarende verdiene fra modellen:

$$\sum_{i=2}^4 [TPF_i^{\text{observert}} - TPF_i^{\text{modell}}]^2 + \sum_{i=2}^4 [(FPP_i^{\text{observert}} - FPP_i^{\text{modell}})]^2 = \text{minimum}. \quad (7)$$

IDL inneholder en rutine for å gjøre slik tilpasning med bruk av en ikke-lineær metode (Marquart-metoden), som krever at modellens funksjonsverdier og deres partiellderiverte med hensyn på modellparametrene er kjent. Funksjonsverdiene er allerede gitt ovenfor (ligning 5 og 6), og da både middelverdien μ_p og standardavviket σ_p for den positive fordelingen er valgt avhenger TPF_i^{modell} kun av konfidensstersklene x_i . Tabellen under viser hvilke funksjonsverdier som må partiellderivere med hensyn på hver av modellparametrene.

| Modellparameter | TPF_1^{modell} | TPF_2^{modell} | TPF_3^{modell} | FPP_1^{modell} | FPP_2^{modell} | FPP_3^{modell} |
|-----------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| x_1 | x | | | x | | |
| x_2 | | x | | | x | |
| x_3 | | | x | | | x |
| μ_N | | | | x | x | x |
| σ_N | | | | x | x | x |

Tabell II.1 Tabellen viser hvilke funksjonsverdier som må partiellderivere med hensyn på de ulike modellparametrene.

For en vilkårlig funksjon $f(x,y,z)$ kan den partiellderiverte med hensyn på størrelsen x tilnærmet skrives som $\frac{\partial f(x,y,z)}{\partial x} = \frac{f(x+\Delta x,y,z) - f(x,y,z)}{\Delta x}$. I programmet er de partiellderiverte med hensyn på konfidensstersklene x_i gitt av uttrykkene

$$\frac{\partial TPF_i^{\text{modell}}(x_i)}{\partial x_i} = \frac{1}{0.01} \cdot \left\{ 1 - \Phi\left(\frac{x_i + 0.01 - \mu_p}{\sigma_p}\right) \right\} - \left[1 - \Phi\left(\frac{x_i - \mu_p}{\sigma_p}\right) \right] \quad \text{og} \quad (8)$$

$$\frac{\partial FPP_i^{\text{modell}}(x_i, \mu_N, \sigma_N)}{\partial x_i} = \frac{1}{0.01} \cdot \left\{ 1 - \Phi\left(\frac{x_i + 0.01 - \mu_N}{\sigma_N}\right) \right\} - \left[1 - \Phi\left(\frac{x_i - \mu_N}{\sigma_N}\right) \right]. \quad (9)$$

De partiellderiverte med hensyn på middelverdien og standardavviket for den negative fordelingen er gitt ved

$$\frac{\partial FPP_i^{\text{modell}}(x_i, \mu_N, \sigma_N)}{\partial \mu_N} = \frac{1}{0.01} \cdot \left\{ 1 - \Phi\left(\frac{x_i - (\mu_N + 0.01)}{\sigma_N}\right) \right\} - \left[1 - \Phi\left(\frac{x_i - \mu_N}{\sigma_N}\right) \right] \quad \text{og} \quad (10)$$

$$\frac{\partial FPP_i^{\text{modell}}(x_i, \mu_N, \sigma_N)}{\partial \sigma_N} = \frac{1}{0.01} \cdot \left\{ 1 - \Phi\left(\frac{x_i - \mu_N}{\sigma_N + 0.01}\right) \right\} - \left[1 - \Phi\left(\frac{x_i - \mu_N}{\sigma_N}\right) \right]. \quad (11)$$

Beregning av ROC-kurven og arealet under denne

Modelltilpasningen gir ikke ROC-kurven direkte, men på grunnlag av de valgte parametrene for den positive fordelingen (μ_p og σ_p) og de som ble bestemt for den negative fordelingen (μ_N og σ_N) er det enkelt å beregne ROC-kurven. ROC-kurven er en fremstilling av den sanne positive fraksjon som funksjon av den falske positive fraksjon. Basert på modellen bestående av to normalfordelinger vil et punkt på kurven være gitt av ligningene 5 og 6 for en bestemt x -verdi. For å få en tilnærmet kontinuerlig ROC-kurve,

er funksjonsverdiene $TPF^{\text{modell}}(x, \mu_p, \sigma_p)$ og $FPF^{\text{modell}}(x, \mu_N, \sigma_N)$ beregnet for 1000 x-verdier mellom 0 og 10. De resulterende vektorene med disse funksjonsverdiene kan kalles for $TPF^{\text{kurve}}(k)$ og $FPF^{\text{kurve}}(k)$, der k har verdier mellom 0 og 999.

Arealet A under ROC-kurven utgjøres av søyler med høyde $TPF^{\text{kurve}}(k)$ og med bredde $FPF^{\text{kurve}}(k) - FPF^{\text{kurve}}(k-1)$:

$$A = \sum_{k=1}^{999} \{ TPF^{\text{kurve}}(k) \cdot [FPF^{\text{kurve}}(k) - FPF^{\text{kurve}}(k-1)] \}. \quad (12)$$

II.2 Tilpasning til rett linje i normalavvikskoordinater

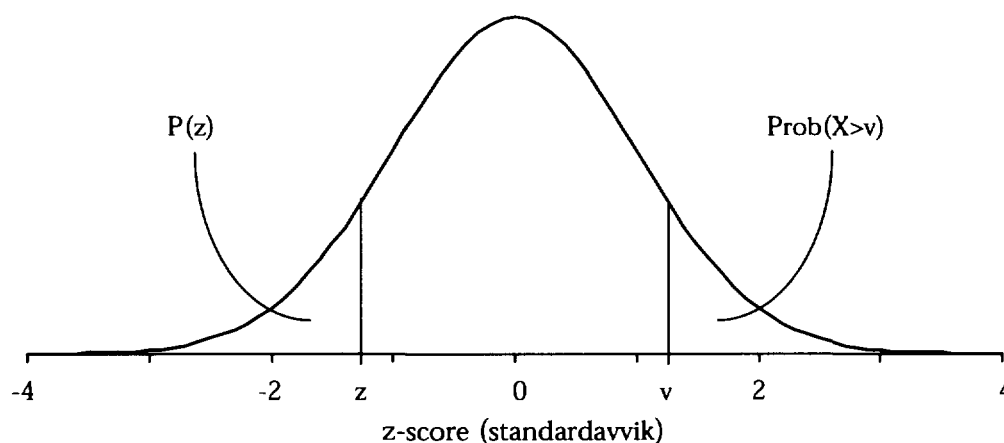
Sammenhengen mellom en gitt proporsjon P (i dette tilfellet den sanne eller falske positive fraksjon), og den tilhørende verdien i normalavvikskoordinater eller z-koordinater (z-score) er gitt ved integralet av den kumulative normalfordelingen,

$$P(z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du. \quad (13)$$

Den inverse relasjonen er gitt ved $z(P) = \Phi^{-1}(P)$. Disse relasjonene kan beskrives ved algebraiske approksimasjoner, men finnes også i tabeller. I IDL finnes funksjonen *gauss_cvf(p)* som beregner terskelverdien v slik at sannsynligheten for at X er større enn v blir lik verdien p, $\text{Pr ob}(X > v) = p(v)$, der X er en tilfeldig variabel fra normalfordelingen med middelerdi 0,0 og standardavvik 1,0. Figur II.1 viser sammenhengen mellom de tilhørende størrelsene P(z) og z og de tilhørende størrelsene p(v) og v, og man ser at hvis $P(z) = \text{Prob}(X > v)$ så er $|z| = |v|$ eller $z = -v$. z-score for en sann eller falsk positive fraksjon blir dermed $z(P) = -\text{gauss_cvf}(P)$.

Modelltilpasning

Ved tilpasning til en rett linje i normalavvikskoordinater beregnes først z-score for hver av de observerte verdiene $TPF_{4.3.2}^{\text{observert}}$ og $FPF_{4.3.2}^{\text{observert}}$ som beskrevet ovenfor. Deretter gjøres det en minste kvadraters tilpasning til en rett linje gjennom de tre målepunktene (IDL-funksjonen *regress()*).



Figur II.1. Sammenhengen mellom de observerte verdiene i kartesiske koordinater og i normalavvikskoordinater. En sann eller falsk positiv fraksjon kan utgjøre arealet P(z), og den samme fraksjonen i normalavvikskoordinater blir da verdien z. IDL-funksjonen *gauss_cvf(p)* beregner verdien v slik at sannsynligheten for at en tilfeldig variabel X er større enn v blir lik verdien p. På grunn av symmetrien blir dermed $z = -v$.

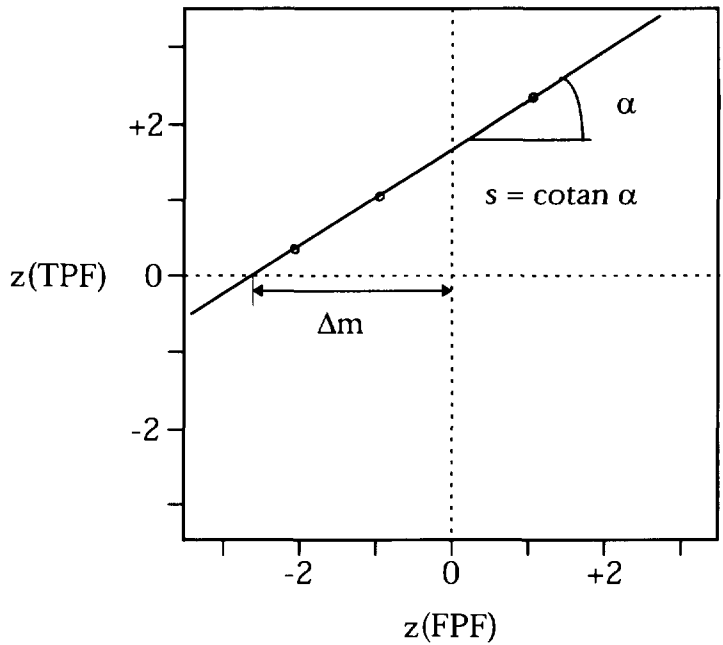
For å kunne vise den resulterende kurven i kartesiske koordinater, må den rette linjen transformeres punkt for punkt med bruk av ligning 13. I vårt program har vi benyttet IDL-funksjonen *gaussint()*.

Beregning av arealet under ROC-kurven

For å beregne arealet A_z under ROC-kurven, må størrelsene Δm og s bestemmes (figur II.2). Δm er absoluttverdien av avstanden mellom punktet $(0,0)$ og skjæringspunktet mellom linjen $z(\text{TPF})=0$ og den rette linjen, mens s er stigningskoeffisienten til den rette linjen. z -verdien av arealet under kurven er gitt som

$$z(A) = \frac{\Delta m \cdot s}{1+s^2}. \quad \text{Arealet } A_z \text{ i}$$

kartesiske koordinater bestemmes ved bruk av ligning 13 eller med bruk av IDL-funksjonen *gaussint()*.



Figur II.2. Størrelsen Δm er absoluttavstanden mellom punktet $(0,0)$ og skjæringspunktet mellom linjen $z(\text{TPF})=0$ og den rette linjen (ROC-kurven). Størrelsen s er stigningskoeffisienten til den rette linjen.

II.3 Ikke-parametrisk modell - Wilcoxon statistikk

TP_i og FP_i er antallet sanne og falske positive for konfidensnivået i . Størrelsen n_p er det totale antallet positive (ruter med "kalk") og n_N det totale antallet negative ("tomme" ruter). W gitt av uttrykket

$$W = \frac{1}{n_N \cdot n_p} \cdot \sum_{i=1}^4 \left\{ FP_i \cdot \left[\sum_{j=i+1}^4 TP_j + \frac{1}{2} \cdot TP_i \right] \right\}, \quad (14)$$

beskriver sannsynligheten for rett beslutning i et såkalt two-alternative forced choice (2-AFC) eksperiment. Det er også vist at denne størrelsen er lik arealet under ROC-kurven, når antallet konfidensnivåer blir mange. Når antallet konfidensnivåer er begrenset, vil størrelsen W bli noe mindre enn arealet under ROC-kurven.

For beregning av standardfeilen inngår, i tillegg til W , også størrelsene Q_1 og Q_2 :

$$Q_1 = \frac{1}{n_N \cdot n_p^2} \cdot \sum_{i=1}^4 \left\{ FP_i \cdot \left[\left(\sum_{j=i+1}^4 TP_j \right)^2 + \left(\sum_{j=i+1}^4 TP_j \right) \cdot TP_i + \frac{1}{3} \cdot (TP_i)^2 \right] \right\} \text{ og} \quad (15)$$

$$Q_2 = \frac{1}{n_N^2 \cdot n_p} \cdot \sum_{i=1}^4 \left\{ TP_i \cdot \left[\left(\sum_{j=1}^{i-1} FP_j \right)^2 + \left(\sum_{j=1}^{i-1} FP_j \right) \cdot FP_i + \frac{1}{3} \cdot (FP_i)^2 \right] \right\}. \quad (16)$$

Standardfeilen for W , $SE(W)$, er da gitt som

$$SE(W) = \sqrt{\frac{W \cdot (1-W) + (n_p - 1) \cdot (Q_1 - W^2) + (n_N - 1) \cdot (Q_2 - W^2)}{n_N \cdot n_p}}. \quad (17)$$

Beregning av $SE(W)$ som ovenfor overestimerer standardfeilen, og konfidensintervallet vil i realiteten være mindre enn det som fremgår av den oppgitte $SE(W)$.

Da alle bildene er eksponeringer av det samme fantomet med de samme kombinasjoner av vev og overlagringsplater, er alle datasettene (antallet sanne og falske positive for hvert konfidensnivå) korrelerte. Dette kan brukes til å styrke statistikken i de oppnådde resultatene, bl.a. for sammenligning mellom anvendt og forstørrelsesteknikk. Først må da Kendall's tau beregnes for både de sanne positive og de falske positive. Størrelsen Kendall's tau uttrykker korrelasjonen mellom antallet sanne eller falske positive ved de to metodene, og beregnes som:

$$r_{\text{tau}} = \sum_{(i,j):i<j} \frac{\text{sign}(S_i - S_j)\text{sign}(R_i - R_j)}{\frac{1}{2}n(n-1)}, \quad (18)$$

der $\text{sign}(A-B) = +1$ hvis $A > B$ og -1 hvis $A < B$. R_i og S_i representerer antallet sanne eller falske positive for konfidensnivået i , og n er antallet konfidensnivåer. For beregning av korrelasjonen mellom anvendt og forstørrelsesteknikk, kan S 'ene representere antallet sanne eller falske positive ved anvendt teknikk, mens R 'ene representerer de tilsvarende størrelsene ved forstørrelsesteknikk.

Vi har brukt IDL-funksjonen $r_correlate()$ til å beregne Kendall's tau for korrelasjonen mellom de sanne positive, r_p , og for korrelasjonen mellom de falske positive, r_N . Ved hjelp av størrelsene $\frac{1}{2}(r_p + r_N)$ og $\frac{1}{2}(A_1 + A_2)$, finnes den endelige korrelasjonskoeffisienten r fra en tabell som finnes i referanse 21. Denne inngår i uttrykket for det kritiske ratio z , gitt som:

$$z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2 \cdot r \cdot SE_1 \cdot SE_2}}. \quad (19)$$

Ved sammenligning av anvendt og forstørrelsesteknikk er det rimelig å anta at arealet under ROC-kurven skal være størst ved forstørrelsesteknikk. Det kan da gjøres en ensidig test, og z -verdier større enn 1,645 indikerer at forskjellen mellom de to teknikkene er signifikant ved 5 %-nivået.

Vedlegg III. Resultatene for NRT-fantomet

Tabell III.1 Resultatene for NRT-fantomet ved anvendt teknikk.

| Lab kode | C_{10} | C_{20} | R_1 | R_2 | S/P | CD-score |
|-----------------|----------|----------|-------|-------|------|----------|
| AA | 0,100 | 0,32 | 20,0 | 11,1 | 0,02 | 15,7±0,5 |
| AB | 0,106 | 0,33 | 20,0 | 11,1 | 0,03 | 11,9±2,9 |
| AC | 0,088 | 0,35 | 20,0 | 20,0 | 0,02 | 14,8±1,6 |
| AD | 0,142 | 0,38 | 16,6 | 16,6 | 0,00 | 15,3±0,5 |
| AE | 0,098 | 0,35 | 16,6 | 12,5 | 0,03 | 12,7±1,0 |
| AF | 0,117 | 0,38 | 16,6 | 14,3 | 0,02 | 9,8±0,4 |
| AG | 0,120 | 0,37 | 16,6 | 16,6 | 0,04 | 11,3±0,4 |
| BA | 0,081 | 0,29 | 16,6 | 20,0 | 0,00 | 10,2±0,8 |
| BB | 0,080 | 0,32 | 16,6 | 11,1 | 0,04 | 12,8±1,5 |
| BC | 0,068 | 0,25 | 16,6 | 14,3 | 0,28 | 7,7±0,5 |
| BD | 0,124 | 0,37 | 20,0 | 20,0 | 0,00 | 13,0±1,4 |
| BE | 0,129 | 0,34 | 20,0 | 16,6 | 0,02 | 10,3±1,5 |
| BF | 0,086 | 0,32 | 14,3 | 12,5 | 0,02 | 12,2±0,8 |
| BG | 0,125 | 0,36 | 16,6 | 20,0 | 0,02 | 11,9±1,7 |
| CA | 0,091 | 0,32 | 20,0 | 16,6 | 0,02 | 12,3±1,5 |
| CB | 0,075 | 0,25 | 16,6 | 16,6 | 0,27 | 6,2±1,3 |
| CC | 0,095 | 0,36 | 16,6 | 16,6 | 0,05 | 9,8±1,6 |
| CD | 0,086 | 0,33 | 20,0 | 20,0 | 0,02 | 11,6±1,3 |
| CE [†] | 0,047 | 0,22 | 14,3 | 14,3 | 0,32 | 6,7±0,5 |
| CF | 0,079 | 0,30 | 14,3 | 16,6 | 0,02 | 11,2±1,0 |
| CG | 0,113 | 0,32 | 16,6 | 16,6 | 0,05 | 10,6±1,3 |
| DA | 0,130 | 0,38 | 20,0 | 20,0 | 0,02 | 9,6±1,6 |
| DB | 0,088 | 0,32 | 20,0 | 16,6 | 0,02 | 13,5±1,7 |
| DC | 0,075 | 0,30 | 20,0 | 20,0 | 0,02 | 14,2±2,6 |
| DD | 0,100 | 0,35 | 16,6 | 14,3 | 0,05 | 8,5±0,8 |
| DE | 0,072 | 0,26 | 14,3 | 12,5 | 0,22 | 7,2±1,9 |
| DF | 0,069 | 0,33 | 20,0 | 16,6 | 0,03 | 8,8±1,2 |
| DG | 0,071 | 0,24 | 16,6 | 14,3 | 0,35 | 6,8±2,0 |
| EA | 0,081 | 0,26 | 20,0 | 16,6 | 0,05 | 11,0±2,2 |
| EB | 0,091 | 0,32 | 20,0 | 16,6 | 0,03 | 13,0±1,9 |
| EC | 0,052 | 0,24 | 16,6 | 14,3 | 0,35 | 9,2±0,8 |
| ED | 0,115 | 0,35 | 16,6 | 14,3 | 0,08 | 12,2±0,8 |
| EE | 0,104 | 0,31 | 16,6 | 14,3 | 0,03 | 11,9±1,8 |
| EF | 0,083 | 0,32 | 20,0 | 16,6 | 0,06 | 17,0±0,9 |
| EG | 0,085 | 0,38 | 12,5 | 10,0 | 0,03 | 12,1±1,4 |
| FA | 0,038 | 0,20 | 16,6 | 14,3 | 0,31 | 6,8±1,2 |
| FB | 0,066 | 0,28 | 20,0 | 20,0 | 0,05 | 14,0±1,4 |
| FC | 0,104 | 0,39 | 16,6 | 14,3 | 0,02 | 6,8±1,8 |
| FD | 0,070 | 0,30 | 14,3 | 12,5 | 0,05 | 9,1±1,0 |
| FE | 0,128 | 0,36 | 20,0 | 20,0 | 0,03 | 16,2±1,0 |
| FF | 0,122 | 0,37 | 14,3 | 14,3 | 0,05 | 15,3±1,0 |
| FG | 0,081 | 0,31 | 16,6 | 16,6 | 0,03 | 10,3±0,9 |
| GA | 0,097 | 0,30 | 16,6 | 14,3 | 0,03 | 10,5±1,2 |
| GB | 0,101 | 0,30 | 20,0 | 16,6 | 0,05 | 10,9±1,4 |
| GC | 0,091 | 0,33 | 14,3 | 14,3 | 0,05 | 11,8±0,8 |

[†]Rasterteknikk ble ikke benyttet.

Tabell III.2 Resultatene for NRT-fantomet ved referanseteknikk.

| Lab. kode | C_{100} | C_{150} | R_1 | R_2 | SP | CD _{norm} |
|-----------------|-----------|-----------|-------|-------|------|--------------------|
| AA | 0,100 | 0,32 | 20,0 | 11,1 | 0,02 | 15,7±0,4 |
| AB | 0,106 | 0,33 | 20,0 | 11,1 | 0,03 | 11,9±2,8 |
| AC | 0,080 | 0,33 | 20,0 | 20,0 | 0,00 | 12,0±2,3 |
| AD | 0,136 | 0,37 | 16,6 | 16,6 | 0,05 | 14,3±1,0 |
| AE | 0,095 | 0,32 | 16,6 | 12,5 | 0,05 | 10,5±1,3 |
| AF | 0,109 | 0,37 | 16,6 | 14,3 | 0,03 | 9,1±1,0 |
| AG | 0,120 | 0,37 | 16,6 | 16,6 | 0,04 | 11,5±0,5 |
| BA | 0,083 | 0,31 | 20,0 | 20,0 | 0,00 | 11,8±1,6 |
| BB | 0,080 | 0,32 | 16,6 | 11,1 | 0,04 | 12,8±1,5 |
| BC | 0,055 | 0,22 | 16,6 | 14,3 | 0,27 | 5,0±1,5 |
| BD | 0,109 | 0,35 | 20,0 | 20,0 | 0,02 | 12,8±1,0 |
| BE | 0,129 | 0,34 | 20,0 | 16,6 | 0,02 | 10,3±1,2 |
| BF | 0,086 | 0,32 | 14,3 | 12,5 | 0,02 | 12,2±0,8 |
| BG | 0,121 | 0,35 | 16,6 | 20,0 | 0,03 | 10,6±1,1 |
| CA | 0,091 | 0,35 | 20,0 | 14,3 | 0,02 | 13,3±0,8 |
| CB | 0,065 | 0,23 | 16,6 | 16,6 | 0,28 | 5,5±1,0 |
| CC | 0,112 | 0,37 | 16,6 | 16,6 | 0,03 | 10,1±0,7 |
| CD | 0,099 | 0,34 | 20,0 | 20,0 | 0,02 | 12,3±1,5 |
| CE [†] | 0,050 | 0,20 | 16,6 | 12,5 | 0,31 | 5,8±1,0 |
| CF | 0,092 | 0,33 | 14,3 | 16,6 | 0,05 | 10,7±0,8 |
| CG | 0,113 | 0,32 | 16,6 | 16,6 | 0,05 | 10,6±1,3 |
| DA | 0,126 | 0,40 | 20,0 | 20,0 | 0,00 | 10,9±0,7 |
| DB | 0,088 | 0,32 | 20,0 | 16,6 | 0,02 | 13,5±1,4 |
| DC | 0,075 | 0,30 | 20,0 | 20,0 | 0,02 | 14,2±2,6 |
| DD | 0,096 | 0,35 | 16,6 | 14,3 | 0,02 | 9,3±0,5 |
| DE [†] | 0,061 | 0,24 | 14,3 | 14,3 | 0,23 | 5,8±0,9 |
| DF | 0,069 | 0,33 | 20,0 | 16,6 | 0,03 | 8,8±1,2 |
| DG [†] | 0,065 | 0,19 | 14,3 | 14,3 | 0,33 | 4,5±0,8 |
| EA | 0,089 | 0,28 | 20,0 | 16,6 | 0,06 | 9,0±0,6 |
| EB | 0,091 | 0,32 | 20,0 | 16,6 | 0,03 | 13,0±1,9 |
| EC [†] | 0,058 | 0,23 | 16,6 | 14,3 | 0,37 | 8,8±1,2 |
| ED | 0,107 | 0,34 | 14,3 | 12,5 | 0,03 | 11,4±1,3 |
| EE | 0,104 | 0,31 | 16,6 | 14,3 | 0,03 | 11,9±1,9 |
| EF | 0,083 | 0,32 | 20,0 | 16,6 | 0,06 | 17,0±0,9 |
| EG | 0,061 | 0,35 | 12,5 | 10,0 | 0,02 | 10,4±0,8 |
| FA [†] | 0,039 | 0,21 | 16,6 | 14,3 | 0,34 | 7,3±0,9 |
| FB | 0,066 | 0,28 | 20,0 | 20,0 | 0,05 | 14,0±1,5 |
| FC | 0,104 | 0,39 | 16,6 | 14,3 | 0,02 | 6,8±1,8 |
| FD | 0,070 | 0,30 | 14,3 | 12,5 | 0,05 | 9,1±1,0 |
| FE | 0,101 | 0,34 | 16,6 | 20,0 | 0,00 | 13,3±1,7 |
| FF | 0,114 | 0,35 | 14,3 | 14,3 | 0,05 | 11,8±1,4 |
| FG | 0,081 | 0,31 | 16,6 | 16,6 | 0,03 | 10,2±0,8 |
| GA | 0,088 | 0,30 | 16,6 | 14,3 | 0,05 | 11,8±0,9 |
| GB | 0,107 | 0,31 | 16,6 | 16,6 | 0,05 | 12,2±2,2 |
| GC | 0,105 | 0,34 | 14,3 | 14,3 | 0,02 | 11,3±1,8 |

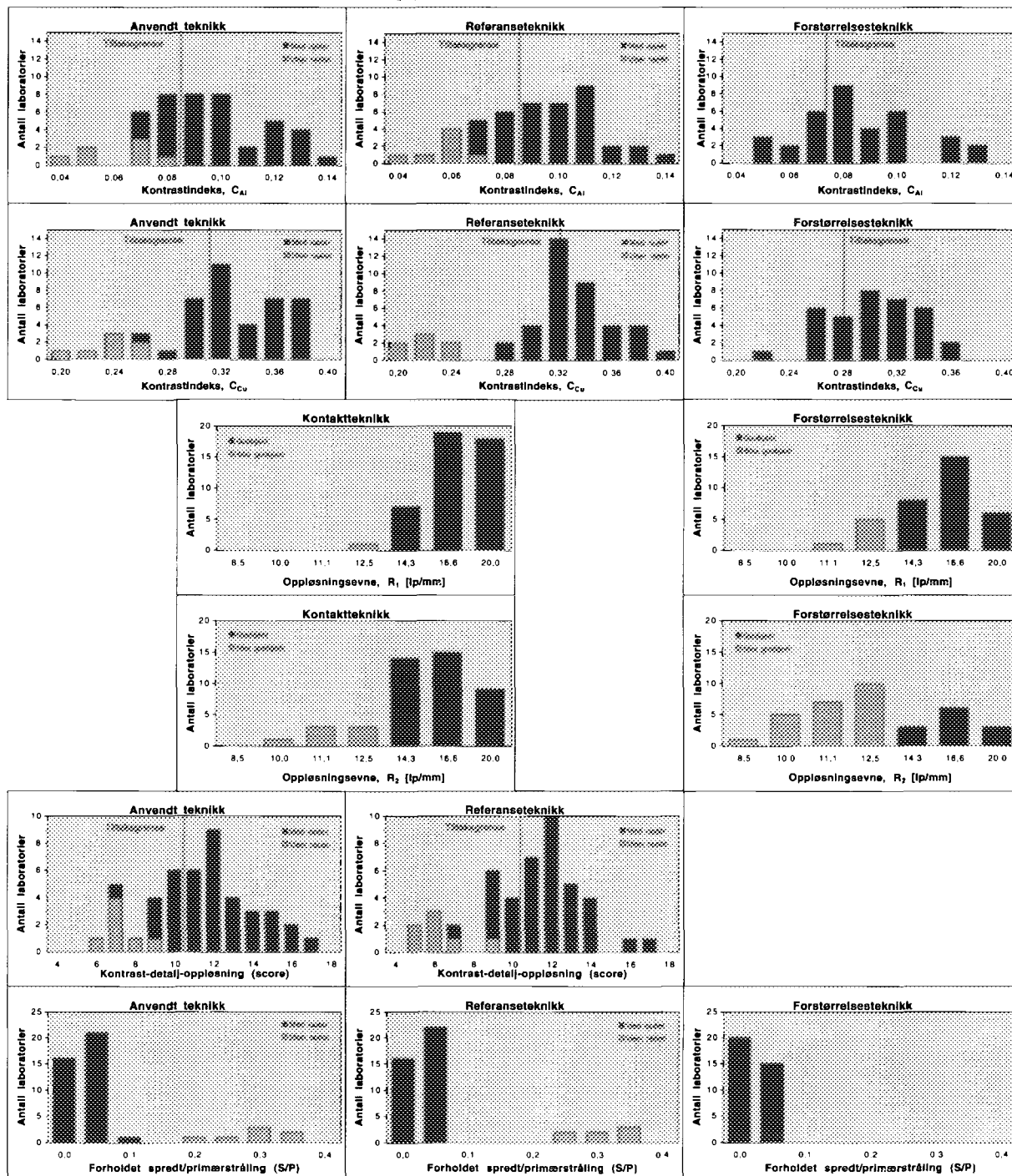
[†]Rasterteknikk ble ikke benyttet.

Tabell III.3 Resultatene for NRT-fantomet ved forstørrelsesteknikk.

| <i>Lab. kode</i> | C_{12} | C_{13} | R_1 | R_2 | S/P |
|------------------|----------|----------|-------|-------|-------|
| AA | 0,087 | 0,30 | 16,6 | 20,0 | 0,04 |
| AB | 0,083 | 0,31 | 14,3 | 11,1 | 0,00 |
| AC | 0,088 | 0,30 | 20,0 | 11,1 | 0,00 |
| AD | 0,096 | 0,32 | 14,3 | 8,5 | 0,00 |
| AE | 0,071 | 0,31 | 16,6 | 16,6 | 0,01 |
| AF | 0,080 | 0,34 | 16,6 | 16,6 | 0,03 |
| AG | 0,120 | 0,37 | 16,6 | 10,0 | 0,00 |
| BA | 0,067 | 0,26 | 20,0 | 16,6 | 0,00 |
| BB | 0,075 | 0,31 | 14,3 | 20,0 | 0,02 |
| BC | | | | | |
| BD | 0,098 | 0,30 | 16,6 | 11,1 | 0,03 |
| BE | 0,079 | 0,32 | 12,5 | 11,1 | 0,00 |
| BF | 0,081 | 0,26 | 14,3 | 12,5 | 0,03 |
| BG | 0,089 | 0,28 | 16,6 | 14,3 | 0,01 |
| CA | | | | | |
| CB | 0,074 | 0,27 | 16,6 | 12,5 | 0,05 |
| CC | 0,121 | 0,33 | 20,0 | 16,6 | 0,00 |
| CD | 0,091 | 0,29 | 16,6 | 14,3 | 0,02 |
| CE | 0,057 | 0,25 | 12,5 | 12,5 | 0,04 |
| CF | 0,048 | 0,26 | 14,3 | 12,5 | 0,01 |
| CG | | | | | |
| DA | 0,129 | 0,36 | 16,6 | 14,3 | 0,03 |
| DB | 0,081 | 0,32 | 16,6 | 12,5 | 0,02 |
| DC | 0,100 | 0,30 | 20,0 | 10,0 | 0,02 |
| DD | 0,053 | 0,28 | 20,0 | 16,6 | 0,03 |
| DE | 0,073 | 0,28 | 12,5 | 12,5 | 0,01 |
| DF | 0,119 | 0,34 | 20,0 | 12,5 | 0,03 |
| DG | | | | | |
| EA | 0,056 | 0,22 | 16,6 | 10,0 | 0,03 |
| EB | 0,072 | 0,32 | 14,3 | 12,5 | 0,02 |
| EC | | | | | |
| ED | 0,078 | 0,31 | 14,3 | 16,6 | 0,03 |
| EE | | | | | |
| EF | 0,098 | 0,34 | 16,6 | 10,0 | 0,00 |
| EG | 0,101 | 0,34 | 14,3 | 11,1 | 0,03 |
| FA | | | | | |
| FB | 0,079 | 0,25 | 16,6 | 12,5 | 0,03 |
| FC | | | | | |
| FD [†] | | | | | |
| FE | 0,095 | 0,29 | 16,6 | 12,5 | 0,02 |
| FF | | | | | |
| FG | 0,045 | 0,29 | 11,1 | 20,0 | 0,04 |
| GA | 0,082 | 0,28 | 16,6 | 10,0 | 0,00 |
| GB | 0,070 | 0,30 | 12,5 | 11,1 | 0,00 |
| GC | 0,126 | 0,35 | 12,5 | 11,1 | 0,03 |

[†]Forstørrelsesteknikk ble ikke benyttet.

Figur III.1 Histogrammer som oppsummerer resultatene for NRT-fantomet.



Vedlegg IV. Resultatene for RMI-fantomet

Tabell IV.1 Resultatene for RMI-fantomet ved anvendt teknikk.

| Lab. kode | Fibriller | | Kalk | | Fortetninger | |
|-----------------|--------------|------------------|--------------|------------------|--------------|------------------|
| | Dårlyst/Best | ACR ¹ | Dårlyst/Best | ACR ¹ | Dårlyst/Best | ACR ¹ |
| AA | 0,75/0,54 | + | 0,32 | + | 0,75/0,50 | + |
| AB | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| AC | 0,75 | + | 0,32 | + | 0,75 | + |
| AD | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| AE | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| AF | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| AG | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| BA | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| BB | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| BC ² | 1,12/0,75 | ? | 0,40/0,32 | ? | 0,75 | + |
| BD | 0,75/0,54 | + | 0,32/0,24 | + | 0,50 | + |
| BE | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| BF | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| BG | 1,12/0,75 | ? | 0,32 | + | 0,75 | + |
| CA | 0,75 | + | 0,32 | + | 1,00/0,75 | ? |
| CB ² | 0,89/0,75 | ? | 0,40/0,32 | ? | 0,75 | + |
| CC | 0,89/0,75 | ? | 0,32 | + | 1,00/0,75 | ? |
| CD | 0,75/0,54 | + | 0,32 | + | 0,75/0,50 | + |
| CE ² | 1,12/0,75 | ? | 0,32 | + | 1,00/0,75 | ? |
| CF | 1,12/0,89 | - | 0,32 | + | 1,00/0,75 | ? |
| CG | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| DA | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| DB | 0,75 | + | 0,32 | + | 0,50 | + |
| DC | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| DD | 0,89 | - | 0,40/0,32 | ? | 0,50 | + |
| DE ² | 1,12/0,75 | ? | 0,40/0,32 | ? | 0,75/0,50 | + |
| DF | 1,12/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| DG ² | 1,12 | - | 0,40 | - | 0,75/0,50 | + |
| EA | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| EB | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| EC ² | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| ED | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| EE | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| EF | 0,75 | + | 0,32/0,24 | + | 0,75/0,50 | + |
| EG | 0,75 | + | 0,32 | + | 0,50 | + |
| FA | 1,12/0,89 | - | 0,40/0,32 | ? | 1,00/0,75 | ? |
| FB | 0,75/0,54 | + | 0,32 | + | 0,50 | + |
| FC | 0,89 | - | 0,32 | + | 0,75/0,50 | + |
| FD | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| FE | 0,75/0,54 | + | 0,32/0,24 | + | 0,50 | + |
| FF | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| FG | 0,89 | - | 0,32 | + | 0,75/0,50 | + |
| GA | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| GB | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| GC | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |

¹Rasterteknikk ble ikke benyttet.

²Alle (+), minst ett (?) eller ingen (-) av tre bilder oppfylte de amerikanske kravene.

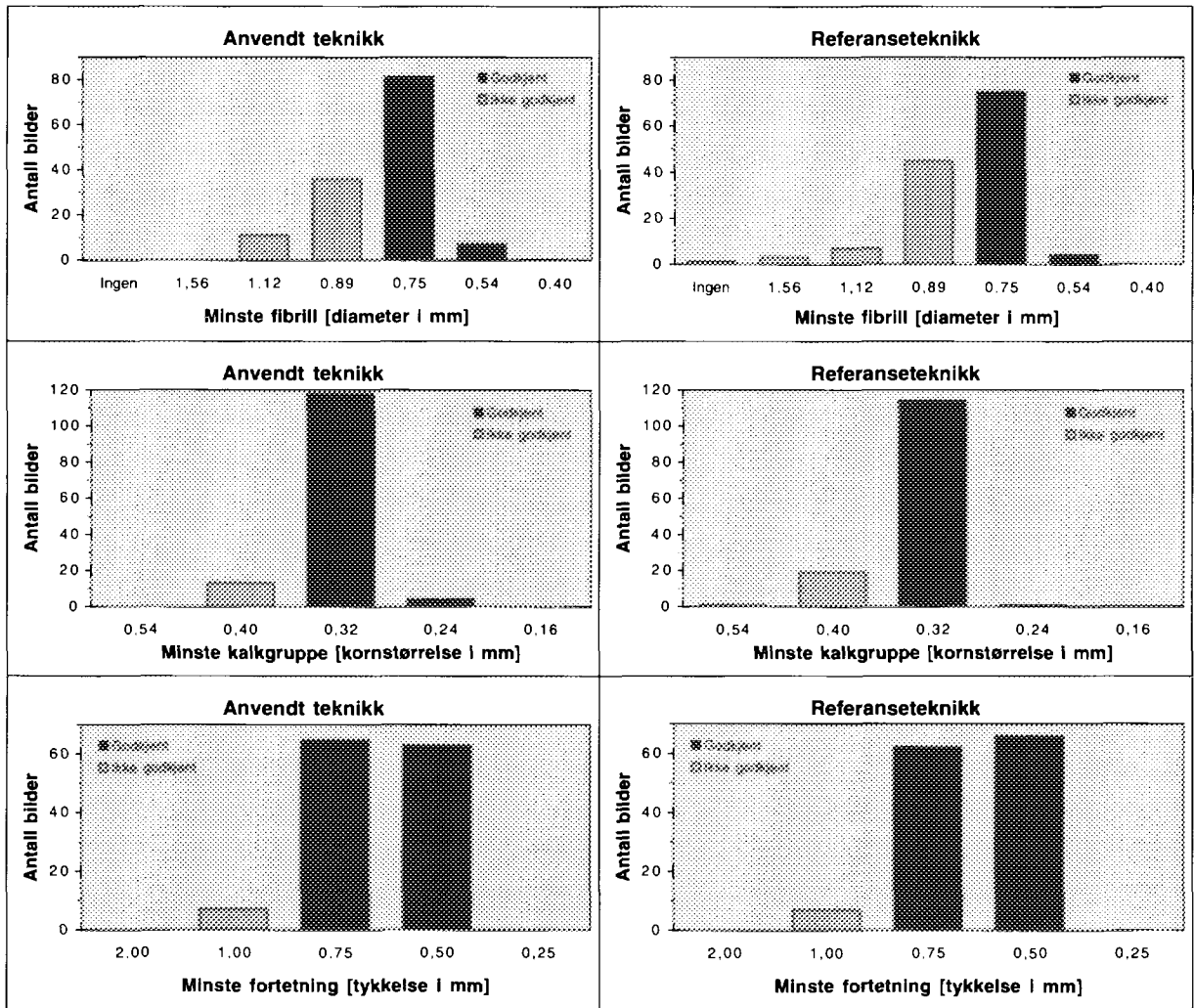
Tabell IV.2 Resultatene for RMI-fantomet ved referanseteknikk.

| Lab kode | Fibriller | | Kalk | | Fortetninger | |
|----------|---------------|------------------|---------------|------------------|---------------|------------------|
| | Dårligst/Best | ACR ¹ | Dårligst/Best | ACR ¹ | Dårligst/Best | ACR ¹ |
| AA | 0,75/0,54 | + | 0,32 | + | 0,75/0,50 | + |
| AB | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| AC | 0,75 | + | 0,32 | + | 0,75 | + |
| AD | 0,75 | + | 0,32 | + | 0,75 | + |
| AE | 0,75 | + | 0,40/0,32 | ? | 0,75/0,50 | + |
| AF | 0,89/0,75 | ? | 0,40/0,32 | ? | 0,75 | + |
| AG | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| BA | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| BB | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| BC | 1,12/0,89 | - | 0,40/0,32 | ? | 0,75/0,50 | + |
| BD | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| BE | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| BF | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| BG | 0,89/0,75 | ? | 0,32 | + | 0,75 | + |
| CA | 0,89/0,75 | ? | 0,32 | + | 1,00/0,75 | ? |
| CB | 1,56/0,89 | - | 0,40 | - | 1,00/0,75 | ? |
| CC | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| CD | 0,75/0,54 | + | 0,32 | + | 0,75 | + |
| CE | 1,12/0,75 | ? | 0,40/0,32 | ? | 1,00/0,75 | ? |
| CF | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| CG | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| DA | 0,75 | + | 0,32 | + | 0,50 | + |
| DB | 0,75 | + | 0,32 | + | 0,50 | + |
| DC | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| DD | 0,89 | - | 0,40/0,32 | ? | 0,50 | + |
| DE | 1,12/0,75 | ? | 0,40 | - | 0,75 | + |
| DF | 1,12/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| DG | Ingen/1,56 | - | 0,54/0,40 | - | 1,00/0,75 | ? |
| EA | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| EB | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| EC | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| ED | 0,75/0,54 | + | 0,32 | + | 0,50 | + |
| EE | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| EF | 0,75 | + | 0,32/0,24 | + | 0,75/0,50 | + |
| EG | 0,89/0,75 | ? | 0,40/0,32 | ? | 0,75/0,50 | + |
| FA | 0,89 | - | 0,40/0,32 | ? | 0,75/0,50 | + |
| FB | 0,75/0,54 | + | 0,32 | + | 0,50 | + |
| FC | 0,89 | - | 0,32 | + | 0,75/0,50 | + |
| FD | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |
| FE | 0,75 | + | 0,32 | + | 0,75/0,50 | + |
| FF | 0,89 | - | 0,32 | + | 0,75 | + |
| FG | 0,89 | - | 0,32 | + | 0,75/0,50 | + |
| GA | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| GB | 0,89/0,75 | ? | 0,32 | + | 0,75/0,50 | + |
| GC | 0,89/0,75 | ? | 0,32 | + | 0,50 | + |

¹Rastereteknikk ble ikke benyttet.

²Alle (+), minst ett (?) eller ingen (-) av tre bilder oppfylte de amerikanske kravene.

Figur IV.1 Histogrammer som oppsummerer resultatene for RMI-fantomet.



Vedlegg V. Resultatene for NRH-fantomet

Tabell V.1 Resultatene for NRH-fantomet ved anvendt teknikk.

| <i>Lab. kode</i> | <i>Score</i> | | <i>Kommentarer</i> |
|----------------------|-----------------|-----------------|--|
| | <i>Kontrast</i> | <i>Skarphet</i> | |
| AA | 4 | 4 | |
| AB | 5 | 5 | |
| AC | 4 | 4 | |
| AD | 4 | 3 | |
| AE | 3 | 3 | |
| AF | 4 | 4 | |
| AG | 4 | 4 | |
| BA | 4 | 4 | |
| BB | 4 | 4 | Støv på foliene, framkallerartefakter |
| BC [†] | 3 | 2 | Dårlig kontrast, uskarpt, kalk uskarp |
| BD | 5 | 5 | |
| BE | 4 | 4 | |
| BF | 4 | 4 | |
| BG | 4 | 4 | |
| CA | 4 | 4 | |
| CB [†] | 2 | 2 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| CC | 4 | 4 | Rasterstriper |
| CD | 4 | 4 | |
| CE [†] | 2 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| CF | 3 | 3 | |
| CG | 4 | 3 | |
| DA | 4 | 4 | |
| DB | 4 | 4 | |
| DC | 4 | 4 | |
| DD | 3 | 2 | Rasterstriper |
| DE [†] | 2 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| DF | 4 | 4 | |
| DG [†] | 4 | 3 | |
| EA | 4 | 3 | |
| EB | 4 | 4 | |
| EC [†] | 3 | 2 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| ED | 3 | 3 | |
| EE | 4 | 4 | |
| EF | 3 | 3 | |
| EG | 3 | 2 | Dårlig kontrast, uskarpt, kalk uskarp |
| FA [†] | 1 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| FB | 3 | 4 | |
| FC | 3 | 3 | Rasterstriper |
| FD | 4 | 4 | |
| FE | 4 | 4 | |
| FF | 3 | 3 | |
| FG | 4 | 4 | |
| GA | 3 | 4 | |
| GB | 3 | 4 | |
| GC | 3 | 4 | |

[†]Rasterteknikk ble ikke benyttet.

Tabell V.2 Resultatene for NRH-fantomet ved referanseteknikk.

| Lab. kode | Score | | Kommentarer |
|--------------|----------|----------|--|
| | Kontrast | Skarphet | |
| AA | 4 | 4 | |
| AB | 5 | 5 | |
| AC | 4 | 4 | |
| AD | 3 | 3 | |
| AE | 3 | 3 | |
| AF | 4 | 4 | |
| AG | 4 | 4 | |
| BA | 4 | 4 | |
| BB | 4 | 4 | |
| BC | 3 | 2 | Dårlig kontrast, uskarpt, kalk uskarp |
| BD | 5 | 4 | |
| BE | 4 | 4 | |
| BF | 4 | 4 | |
| BG | 4 | 4 | |
| CA | 4 | 4 | |
| CB | 2 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler, framkallerartefakter |
| CC | 5 | 5 | |
| CD | 4 | 4 | |
| CE | 1 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| CF | 3 | 3 | |
| CG | 4 | 3 | |
| DA | 4 | 4 | |
| DB | 4 | 4 | |
| DC | 4 | 4 | |
| DD | 2 | 2 | Rasterstriper |
| DE | 2 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| DF | 4 | 4 | |
| DG | 3 | 2 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| EA | 3 | 3 | |
| EB | 4 | 4 | |
| EC | 2 | 2 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| ED | 4 | 4 | |
| EE | 4 | 4 | |
| EF | 3 | 3 | |
| EG | 2 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler, rasterstriper |
| FA | 1 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| FB | 3 | 4 | |
| FC | 3 | 3 | |
| FD | 4 | 4 | |
| FE | 4 | 4 | Rasterstriper |
| FF | 3 | 3 | |
| FG | 4 | 4 | |
| GA | 3 | 3 | |
| GB | 3 | 4 | |
| GC | 3 | 3 | |

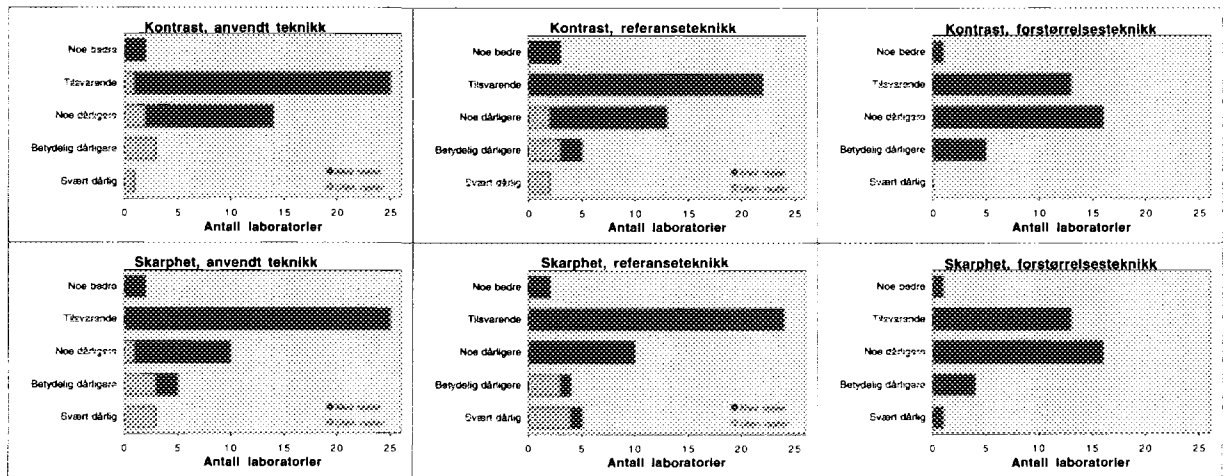
[†]Rasterteknikk ble ikke benyttet.

Tabell V.3 Resultatene for NRH-fantomet ved forstørrelsesteknikk.

| Lab. kode | Score | | Kommentarer |
|-----------------|----------|----------|--|
| | Kontrast | Skarphet | |
| AA | 3 | 3 | |
| AB | 4 | 3 | |
| AC | 4 | 4 | |
| AD | 4 | 4 | |
| AE | 3 | 3 | |
| AF | 3 | 3 | |
| AG | 5 | 5 | |
| BA | 4 | 4 | |
| BB | 3 | 3 | Støv på foliene |
| BC | | | |
| BD | 4 | 4 | |
| BE | 4 | 4 | |
| BF | 3 | 3 | |
| BG | 4 | 3 | |
| CA ⁻ | | | |
| CB | 3 | 3 | |
| CC | 3 | 3 | |
| CD | 3 | 3 | |
| CE | 2 | 2 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| CF | 2 | 2 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| CG ⁺ | | | |
| DA | 4 | 4 | |
| DB | 4 | 4 | |
| DC | 3 | 3 | |
| DD | 2 | 2 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| DE | 2 | 1 | Dårlig kontrast, uskarpt, kalk uskarp og mangler |
| DF | 3 | 3 | |
| DG ⁻ | | | |
| EA | 2 | 2 | Uskarpt |
| EB | 4 | 4 | |
| EC ⁺ | | | |
| ED | 3 | 4 | |
| EE ⁻ | | | |
| EF | 4 | 3 | |
| EG | 3 | 4 | |
| FA ⁻ | | | |
| FB | 3 | 3 | |
| FC ⁻ | | | |
| FD ⁺ | | | |
| FE | 4 | 4 | |
| FF ⁺ | | | |
| FG | 3 | 3 | |
| GA | 3 | 3 | |
| GB | 3 | 4 | |
| GC | 4 | 4 | |

⁻Forstørrelsesteknikk ble ikke benyttet.

Figur V.1 Histogrammer som oppsummerer resultatene for NRH-fantomet.



Vedlegg VI. Resultatene for NRH-ROC-fantomet

Tabell VI.1 Resultatene for NRH-ROC-fantomet ved anvendt og forstørrelsesteknikk.

| Lab. kode | Anvendt teknikk | | | | | | | | Forstørrelsesteknikk | | | | | | | |
|------------------|-----------------|----|----|----|-----------------|----|----|----|----------------------|----|---|---|-----------------|----|----|----|
| | Sanne positive | | | | Falske positive | | | | Sanne positive | | | | Falske positive | | | |
| | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 |
| AA | 24 | 7 | 7 | 2 | 0 | 2 | 20 | 18 | 33 | 3 | 3 | 1 | 0 | 3 | 8 | 29 |
| AB | 18 | 14 | 4 | 4 | 0 | 2 | 19 | 19 | 26 | 8 | 5 | 1 | 0 | 2 | 20 | 18 |
| AC | 13 | 13 | 12 | 2 | 0 | 2 | 10 | 28 | 28 | 7 | 5 | 0 | 0 | 0 | 10 | 30 |
| AD | 24 | 6 | 6 | 4 | 0 | 1 | 10 | 29 | 33 | 2 | 3 | 2 | 1 | 3 | 3 | 33 |
| AE | 30 | 2 | 3 | 5 | 2 | 4 | 5 | 29 | 35 | 1 | 0 | 4 | 1 | 1 | 1 | 37 |
| AF | 18 | 10 | 9 | 3 | 3 | 7 | 16 | 14 | 29 | 2 | 6 | 3 | 0 | 2 | 7 | 31 |
| AG | 16 | 14 | 7 | 3 | 0 | 0 | 11 | 29 | 29 | 4 | 5 | 2 | 0 | 0 | 5 | 35 |
| BA | 33 | 3 | 2 | 2 | 1 | 22 | 11 | 6 | 34 | 4 | 0 | 2 | 1 | 7 | 15 | 17 |
| BB | 20 | 7 | 6 | 7 | 0 | 4 | 12 | 24 | 27 | 10 | 3 | 0 | 0 | 4 | 6 | 30 |
| BC [†] | 11 | 12 | 11 | 6 | 0 | 0 | 7 | 33 | | | | | | | | |
| BD | 22 | 5 | 3 | 10 | 0 | 0 | 0 | 40 | 22 | 9 | 1 | 8 | 0 | 0 | 0 | 40 |
| BE | 25 | 9 | 4 | 2 | 1 | 8 | 15 | 16 | 32 | 2 | 3 | 3 | 0 | 5 | 13 | 22 |
| BF | 24 | 8 | 7 | 1 | 0 | 8 | 17 | 15 | 33 | 4 | 3 | 0 | 4 | 12 | 13 | 11 |
| BG | 9 | 15 | 11 | 5 | 0 | 0 | 3 | 37 | 15 | 14 | 5 | 6 | 0 | 0 | 5 | 35 |
| CA [‡] | 21 | 12 | 7 | 0 | 0 | 1 | 26 | 13 | | | | | | | | |
| CB [‡] | 19 | 11 | 1 | 9 | 3 | 7 | 8 | 22 | 30 | 5 | 1 | 4 | 0 | 6 | 14 | 20 |
| CC | 29 | 8 | 3 | 0 | 1 | 11 | 22 | 6 | 37 | 2 | 1 | 0 | 3 | 13 | 18 | 6 |
| CD | 30 | 7 | 2 | 1 | 1 | 11 | 15 | 13 | 33 | 4 | 2 | 1 | 3 | 13 | 8 | 16 |
| CE [†] | 16 | 12 | 8 | 4 | 0 | 3 | 9 | 28 | 32 | 4 | 4 | 0 | 0 | 1 | 10 | 29 |
| CF | 14 | 7 | 12 | 7 | 0 | 2 | 6 | 32 | 21 | 10 | 5 | 4 | 0 | 1 | 4 | 35 |
| CG [‡] | 27 | 6 | 4 | 3 | 1 | 8 | 13 | 18 | | | | | | | | |
| DA | 25 | 4 | 1 | 10 | 0 | 3 | 1 | 36 | 30 | 1 | 2 | 7 | 0 | 2 | 2 | 36 |
| DB | 28 | 7 | 2 | 3 | 1 | 7 | 7 | 25 | 34 | 2 | 2 | 2 | 3 | 4 | 6 | 27 |
| DC | 24 | 11 | 1 | 4 | 0 | 4 | 6 | 30 | 35 | 3 | 2 | 0 | 1 | 0 | 4 | 35 |
| DD | 18 | 7 | 4 | 11 | 0 | 2 | 4 | 34 | 19 | 7 | 5 | 9 | 0 | 1 | 1 | 38 |
| DE [†] | 22 | 7 | 9 | 2 | 0 | 7 | 18 | 15 | 32 | 5 | 2 | 1 | 1 | 4 | 17 | 18 |
| DF | 18 | 11 | 6 | 5 | 0 | 1 | 10 | 29 | 30 | 4 | 5 | 1 | 0 | 1 | 5 | 34 |
| DG ^{†‡} | 23 | 10 | 5 | 2 | 0 | 13 | 13 | 14 | | | | | | | | |
| EA | 18 | 9 | 5 | 8 | 0 | 1 | 9 | 30 | 30 | 4 | 5 | 1 | 0 | 0 | 1 | 39 |
| EB | 21 | 9 | 7 | 3 | 0 | 1 | 12 | 27 | 31 | 6 | 2 | 1 | 0 | 1 | 15 | 24 |
| EC [†] | 5 | 12 | 9 | 14 | 0 | 1 | 5 | 34 | | | | | | | | |
| ED | 19 | 11 | 5 | 5 | 0 | 2 | 11 | 27 | 24 | 10 | 5 | 1 | 0 | 0 | 2 | 38 |
| EE [†] | 28 | 9 | 0 | 3 | 7 | 8 | 4 | 21 | | | | | | | | |
| EF | 32 | 2 | 1 | 5 | 2 | 3 | 5 | 30 | 39 | 0 | 1 | 0 | 2 | 3 | 4 | 31 |
| EG | 26 | 5 | 4 | 5 | 0 | 6 | 13 | 21 | 32 | 3 | 0 | 5 | 0 | 5 | 6 | 29 |
| FA ^{†‡} | 9 | 4 | 1 | 26 | 0 | 0 | 2 | 38 | | | | | | | | |
| FB | 22 | 9 | 4 | 5 | 0 | 6 | 8 | 26 | 30 | 7 | 1 | 2 | 0 | 4 | 9 | 27 |
| FC [‡] | 20 | 4 | 1 | 15 | 0 | 0 | 0 | 40 | | | | | | | | |
| FD [‡] | 33 | 4 | 3 | 0 | 19 | 8 | 13 | 0 | | | | | | | | |
| FE | 26 | 8 | 2 | 4 | 0 | 2 | 13 | 26 | 31 | 4 | 3 | 2 | 0 | 1 | 11 | 28 |
| FF [†] | 18 | 6 | 10 | 6 | 0 | 2 | 2 | 36 | | | | | | | | |
| FG | 17 | 12 | 8 | 3 | 6 | 8 | 15 | 11 | 24 | 9 | 6 | 1 | 2 | 4 | 21 | 13 |
| GA | 24 | 7 | 4 | 5 | 0 | 12 | 12 | 16 | 31 | 6 | 2 | 1 | 0 | 2 | 9 | 29 |
| GB | 27 | 4 | 6 | 3 | 0 | 3 | 5 | 32 | 31 | 4 | 2 | 3 | 0 | 0 | 3 | 37 |
| GC | 29 | 4 | 3 | 4 | 0 | 4 | 19 | 17 | 37 | 3 | 0 | 0 | 1 | 9 | 16 | 14 |

[†]Rasterteknikk ble ikke benyttet. [‡]Forstørrelsesteknikk ble ikke benyttet.

Tabell VI.2 Resultatene for NRH-ROC-fantomet ved anvendt teknikk.

| Lab.kode | Tilpasning til 2 normalfordelinger | | | Wilcoxon statistikk | | Tilpasning til rett linje i normalavviksk. | | | Sværtning og kontrast | | |
|----------|------------------------------------|---------------|--------------------|---------------------|---------|--|---------------|---------------|-----------------------|-----------|-----------|
| | μ_0 | σ_0 | A | W | $SE(W)$ | μ_0 | σ | A_0 | OD | $G_{0.5}$ | $G_{0.9}$ |
| AA | 4,05 | 0,30 | 0,922 | 0,904 | 0,035 | 9,20 | 0,147 | 0,909 | 1,06 | 1,04 | 2,93 |
| AB | 4,19 | 0,20 | 0,901 | 0,886 | 0,041 | 7,15 | 0,168 | 0,882 | 1,07 | 0,99 | 3,00 |
| AC | 3,68 | 0,65 | 0,933 | 0,907 | 0,033 | 5,67 | 0,230 | 0,898 | 0,99 | 1,38 | 3,06 |
| AD | 4,07 | 0,27 | 0,922 | 0,912 | 0,035 | 9,76 | 0,118 | 0,874 | 0,86 | 1,01 | 3,62 |
| AE | 4,17 | 0,27 | 0,896 | 0,881 | 0,041 | 3,10 | 0,443 | 0,895 | 0,96 | 1,22 | 3,13 |
| AF | 4,34 | 0,51 | 0,801 | 0,779 | 0,052 | 1,29 | 0,858 | 0,800 | * | * | * |
| AG | 4,04 | 0,16 | 0,940 | 0,928 | 0,031 | 9,27 | 0,166 | 0,935 | 1,02 | 0,88 | 3,05 |
| BA | 4,16 | 0,18 | 0,911 | 0,885 | 0,043 | 6,04 | 0,222 | 0,905 | 0,98 | 1,27 | 2,71 |
| BB | 4,36 | 0,29 | 0,834 | 0,831 | 0,047 | 7,74 | 0,102 | 0,783 | 1,17 | 1,14 | 2,99 |
| BC | 4,19 | 0,20 | 0,901 | 0,888 | 0,039 | 6,84 | 0,176 | 0,881 | 0,80 | 1,17 | 2,92 |
| BD | 4,02 | 0,15 | 0,944 ¹ | 0,875 | 0,043 | ²⁾ | ²⁾ | ²⁾ | 0,84 | 1,54 | 2,79 |
| BE | 4,11 | 0,36 | 0,901 | 0,883 | 0,039 | 2,49 | 0,599 | 0,900 | 1,10 | 0,98 | 2,65 |
| BF | 4,12 | 0,38 | 0,893 | 0,887 | 0,036 | 9,13 | 0,162 | 0,928 | 1,23 | 1,14 | 2,96 |
| BG | 4,01 | 0,21 | 0,942 | 0,923 | 0,033 | 6,82 | 0,214 | 0,923 | 0,97 | 1,02 | 3,35 |
| CA | 3,40 | 0,53 | 0,978 | 0,935 | 0,027 | 7,15 | 0,520 | 1,000 | 1,23 | 0,93 | 2,99 |
| CB | 4,43 | 0,37 | 0,792 | 0,766 | 0,055 | 1,45 | 0,644 | 0,784 | 0,91 | 0,82 | 3,44 |
| CC | 3,93 | 0,37 | 0,936 | 0,915 | 0,032 | 2,04 | 1,56 | 0,957 | 1,14 | 1,19 | 3,11 |
| CD | 3,97 | 0,32 | 0,936 | 0,917 | 0,033 | 3,23 | 0,533 | 0,935 | 0,95 | 0,98 | 2,75 |
| CE | 3,98 | 0,48 | 0,908 | 0,886 | 0,038 | 6,31 | 0,173 | 0,859 | 1,22 | 0,67 | 2,67 |
| CF | 3,96 | 0,58 | 0,896 | 0,853 | 0,044 | 4,94 | 0,141 | 0,755 | 1,81 | 0,71 | 2,28 |
| CG | 4,20 | 0,28 | 0,886 | 0,876 | 0,041 | 2,88 | 0,468 | 0,889 | 1,25 | 0,98 | 2,37 |
| DA | 4,28 | 0,25 | 0,869 | 0,857 | 0,045 | 14,6 | 0,048 | 0,758 | 1,10 | 1,20 | 3,10 |
| DB | 4,03 | 0,33 | 0,922 | 0,906 | 0,036 | 2,90 | 0,558 | 0,921 | 1,12 | 1,05 | 2,51 |
| DC | 4,11 | 0,16 | 0,924 | 0,919 | 0,034 | 9,83 | 0,138 | 0,910 | 1,33 | 0,89 | 2,56 |
| DD | ¹⁾ | ¹⁾ | ¹⁾ | 0,827 | 0,048 | 6,49 | 0,090 | 0,720 | 1,11 | 1,04 | 3,04 |
| DE | 4,23 | 0,38 | 0,862 | 0,854 | 0,042 | 8,38 | 0,143 | 0,883 | 1,23 | 0,94 | 3,12 |
| DF | 4,16 | 0,24 | 0,903 | 0,894 | 0,038 | 6,98 | 0,156 | 0,860 | 1,22 | 0,92 | 3,23 |
| DG | 4,25 | 0,31 | 0,869 | 0,857 | 0,042 | 9,10 | 0,144 | 0,903 | 0,89 | 1,01 | 2,84 |
| EA | ¹⁾ | ¹⁾ | ¹⁾ | 0,855 | 0,045 | 6,77 | 0,121 | 0,792 | 1,20 | 0,67 | 2,30 |
| EB | 4,00 | 0,30 | 0,933 | 0,917 | 0,033 | 8,03 | 0,159 | 0,897 | 1,06 | 1,15 | 2,70 |
| EC | ¹⁾ | ¹⁾ | ¹⁾ | 0,775 | 0,054 | 2,06 | 0,198 | 0,656 | 0,93 | 0,93 | 3,26 |
| ED | 4,22 | 0,22 | 0,890 | 0,887 | 0,039 | 7,39 | 0,146 | 0,857 | 0,93 | 1,33 | 3,28 |
| EE | 3,99 | 0,73 | 0,859 | 0,822 | 0,049 | 1,44 | 1,13 | 0,860 | 1,42 | 0,94 | 2,97 |
| EF | 4,17 | 0,19 | 0,907 | 0,893 | 0,040 | 4,32 | 0,319 | 0,905 | 1,29 | 0,94 | 2,79 |
| EG | 4,28 | 0,25 | 0,869 | 0,867 | 0,043 | 12,5 | 0,080 | 0,842 | 0,70 | 1,33 | 4,53 |
| FA | ¹⁾ | ¹⁾ | ¹⁾ | 0,658 | 0,062 | ²⁾ | ²⁾ | ²⁾ | 1,22 | 0,77 | 2,41 |
| FB | 4,21 | 0,31 | 0,881 | 0,874 | 0,041 | 8,96 | 0,116 | 0,850 | 1,04 | 0,92 | 2,54 |
| FC | 2,99 | 0,52 | 0,995 ³ | 0,812 | 0,051 | ²⁾ | ²⁾ | ²⁾ | 1,17 | 1,18 | 3,12 |
| FD | 4,40 | 0,59 | 0,764 | 0,684 | 0,060 | 1,28 | 0,804 | 0,788 | 0,87 | 1,06 | 3,11 |
| FE | 4,13 | 0,17 | 0,920 | 0,918 | 0,035 | 11,3 | 0,113 | 0,898 | 1,05 | 1,16 | 2,67 |
| FF | 3,32 | 0,84 | 0,949 | 0,895 | 0,038 | 6,81 | 0,123 | 0,796 | 0,92 | 1,11 | 5,30 |
| FG | 4,44 | 0,62 | 0,744 | 0,721 | 0,057 | 0,90 | 0,985 | 0,736 | 0,96 | 0,95 | 3,20 |
| GA | 4,36 | 0,29 | 0,834 | 0,829 | 0,048 | 10,6 | 0,092 | 0,834 | 1,42 | 0,92 | 2,69 |
| GB | 3,75 | 0,50 | 0,947 | 0,931 | 0,030 | 12,3 | 0,100 | 0,889 | 0,97 | 0,78 | 2,67 |
| GC | 4,19 | 0,21 | 0,899 | 0,891 | 0,041 | 15,8 | 0,073 | 0,876 | 1,21 | 0,80 | 3,62 |

¹Rasterteknikk ble ikke benyttet. ²Uforholdsmessig stort areal pga. degenerert datasett.

³Tilpasning til 2 normalfordelinger feilet. ²⁾Tilpasning til rett linje feilet. ³⁾Filmene ikke tilgjengelige.

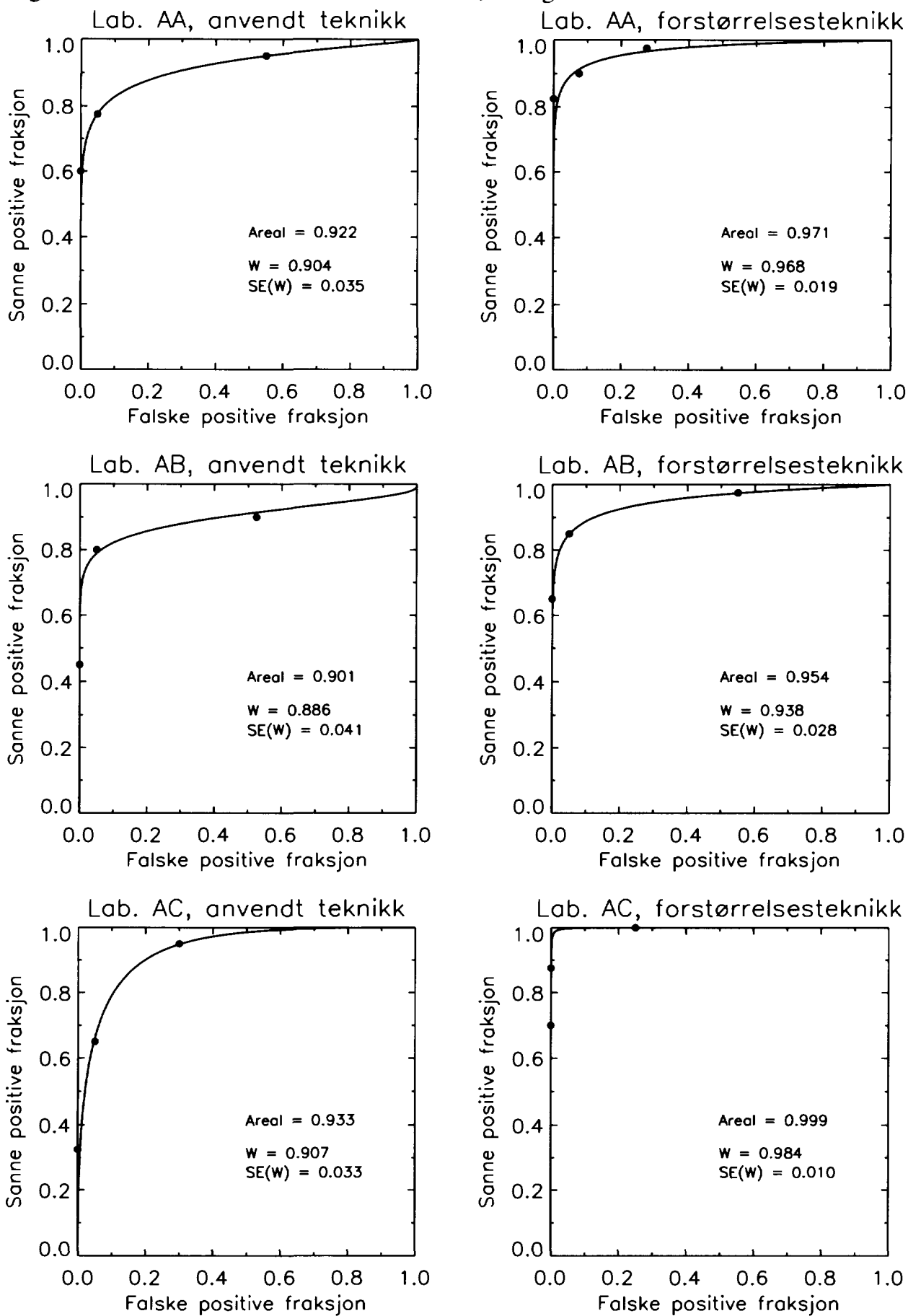
Tabell VI.3 Resultatene for NRH-ROC-fantomet ved forstørrelsesteknikk.

| Lab.kode | Tilpasning til 2 normalfordelinger | | | Wilcoxon statistikk | | Tilpasning til rett linje i normalavvik | | | Svartning og kontrast | | |
|----------|------------------------------------|---------------|---------------|---------------------|---------|---|---------------|---------------|-----------------------|---------------|---------------|
| | μ_0 | σ_0 | A | W | $SE(W)$ | A_{00} | r | A_{10} | OD | G_{00} | G_{10} |
| AA | 3,71 | 0,33 | 0,971 | 0,968 | 0,019 | 16,5 | 0,106 | 0,959 | 1,16 | 0,92 | 2,53 |
| AB | 3,87 | 0,31 | 0,954 | 0,938 | 0,028 | 9,82 | 0,169 | 0,950 | 1,42 | 1,06 | 2,73 |
| AC | 2,37 | 0,56 | 0,999 | 0,984 | 0,010 | 9,40 | 0,598 | 1,000 | 1,52 | 1,06 | 2,46 |
| AD | 3,67 | 0,44 | 0,964 | 0,947 | 0,027 | 3,33 | 0,641 | 0,964 | 1,18 | 0,82 | 2,79 |
| AE | 3,98 | 0,16 | 0,950 | 0,934 | 0,032 | 6,34 | 0,266 | 0,948 | 1,08 | 1,04 | 2,73 |
| AF | 3,87 | 0,37 | 0,946 | 0,932 | 0,030 | 14,8 | 0,082 | 0,887 | ¹⁾ | ²⁾ | ³⁾ |
| AG | 3,81 | 0,18 | 0,972 | 0,964 | 0,022 | 14,0 | 0,128 | 0,962 | 0,95 | 1,01 | 3,14 |
| BA | 3,92 | 0,21 | 0,955 | 0,939 | 0,031 | 5,90 | 0,287 | 0,948 | 1,41 | 0,89 | 2,12 |
| BB | 3,24 | 0,66 | 0,976 | 0,974 | 0,013 | 8,68 | 0,442 | 1,000 | 1,12 | 1,11 | 2,72 |
| BC | | | | | | | | | | | |
| BD | 3,99 | 0,14 | 0,950 | 0,900 | 0,039 | ²⁾ | ²⁾ | ²⁾ | 1,38 | 1,03 | 2,02 |
| BE | 4,14 | 0,17 | 0,917 | 0,921 | 0,034 | 21,9 | 0,059 | 0,901 | 1,55 | 0,99 | 2,36 |
| BF | 3,98 | 0,37 | 0,927 | 0,892 | 0,038 | 1,45 | 2,22 | 0,907 | 1,02 | 0,89 | 2,42 |
| BG | 4,16 | 0,19 | 0,910 | 0,908 | 0,037 | 9,07 | 0,131 | 0,880 | 1,34 | 0,69 | 2,54 |
| CA | | | | | | | | | | | |
| CB | 4,16 | 0,18 | 0,912 | 0,907 | 0,038 | 17,2 | 0,073 | 0,895 | 1,20 | 0,76 | 2,53 |
| CC | 3,73 | 0,29 | 0,973 | 0,938 | 0,030 | 2,08 | 1,54 | 0,960 | 1,32 | 0,97 | 2,73 |
| CD | 4,01 | 0,30 | 0,931 | 0,900 | 0,037 | 3,00 | 0,574 | 0,932 | 1,40 | 0,84 | 2,28 |
| CE | 3,05 | 0,58 | 0,990 | 0,984 | 0,010 | 9,16 | 0,434 | 1,000 | 1,52 | 0,70 | 2,32 |
| CF | 3,79 | 0,39 | 0,956 | 0,931 | 0,031 | 8,24 | 0,154 | 0,895 | 2,43 | | 1,91 |
| CG | | | | | | | | | | | |
| DA | 4,18 | 0,20 | 0,903 | 0,899 | 0,039 | 32,7 | 0,027 | 0,812 | 1,58 | 1,06 | 2,45 |
| DB | 3,88 | 0,36 | 0,947 | 0,916 | 0,035 | 3,09 | 0,616 | 0,947 | 1,22 | 1,09 | 2,36 |
| DC | ¹⁾ | ¹⁾ | ¹⁾ | 0,983 | 0,016 | 2,26 | 4,72 | 0,986 | 1,62 | 0,98 | 2,25 |
| DD | 3,65 | 0,55 | 0,952 | 0,875 | 0,042 | 7,32 | 0,104 | 0,776 | 1,38 | 0,95 | 2,69 |
| DE | 3,75 | 0,36 | 0,964 | 0,944 | 0,028 | 3,69 | 0,525 | 0,957 | 1,99 | 0,85 | 2,24 |
| DF | 3,28 | 0,54 | 0,984 | 0,973 | 0,017 | 12,4 | 0,140 | 0,957 | 1,72 | 1,12 | 2,85 |
| DG | | | | | | | | | | | |
| EA | 3,10 | 0,37 | 0,997 | 0,986 | 0,015 | 12,7 | 0,183 | 0,989 | 1,70 | | 2,00 |
| EB | 3,78 | 0,18 | 0,975 | 0,970 | 0,020 | 13,0 | 0,144 | 0,968 | 1,29 | 1,23 | 2,41 |
| EC | | | | | | | | | | | |
| ED | 3,49 | 0,21 | 0,992 | 0,984 | 0,015 | 11,1 | 0,207 | 0,988 | 0,92 | 1,12 | 2,88 |
| EE | | | | | | | | | | | |
| EF | 3,34 | 0,31 | 0,993 | 0,971 | 0,021 | 2,05 | 3,51 | 0,976 | 1,66 | 1,07 | 2,56 |
| EG | 4,14 | 0,17 | 0,918 | 0,916 | 0,036 | 27,5 | 0,043 | 0,882 | 0,86 | 1,04 | 3,87 |
| FA | | | | | | | | | | | |
| FB | 3,95 | 0,15 | 0,957 | 0,953 | 0,026 | 13,4 | 0,123 | 0,949 | 1,07 | 0,87 | 2,24 |
| FC | | | | | | | | | | | |
| FD | | | | | | | | | | | |
| FE | 3,91 | 0,21 | 0,958 | 0,954 | 0,025 | 15,0 | 0,103 | 0,938 | 1,35 | 0,84 | 2,10 |
| FF | | | | | | | | | | | |
| FG | 3,89 | 0,55 | 0,914 | 0,880 | 0,040 | 2,07 | 0,790 | 0,900 | 1,07 | 0,89 | 2,82 |
| GA | 3,65 | 0,29 | 0,979 | 0,972 | 0,018 | 13,1 | 0,143 | 0,968 | 1,53 | 0,82 | 2,49 |
| GB | 4,00 | 0,10 | 0,951 | 0,958 | 0,026 | 20,9 | 0,074 | 0,938 | 1,27 | 0,76 | 2,36 |
| GC | 2,15 | 1,01 | 0,976 | 0,978 | 0,016 | 3,14 | 1,66 | 0,996 | 1,38 | 0,79 | 3,24 |

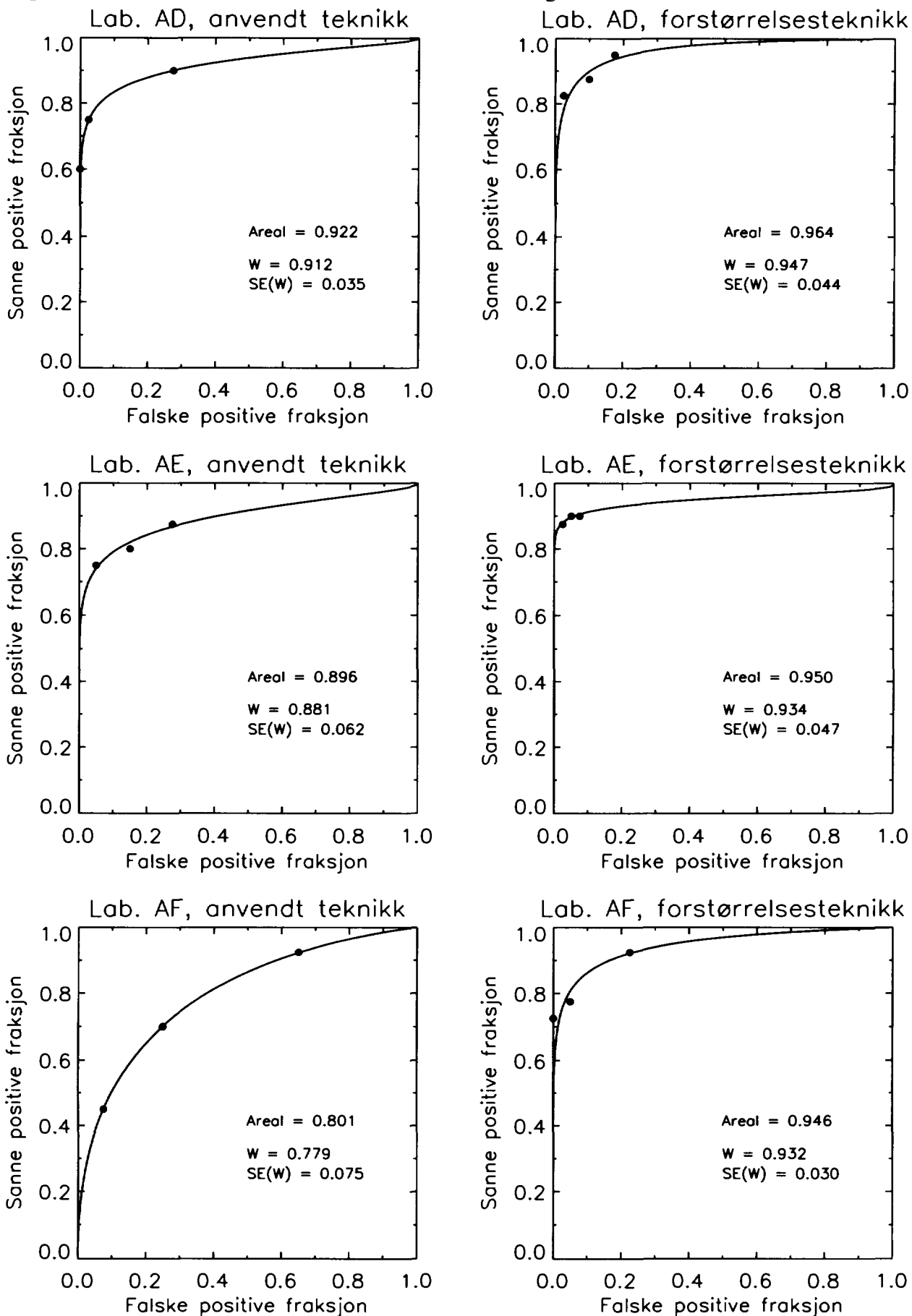
Forstørrelsesteknikk ble ikke benyttet. ¹⁾Uforholdsmessig stort areal pga. degenerert datasett.

²⁾Tilpasning til 2 normalfordelinger feilet. ³⁾Tilpasning til rett linje feilet. ³⁾Filmene ikke tilgjengelige.

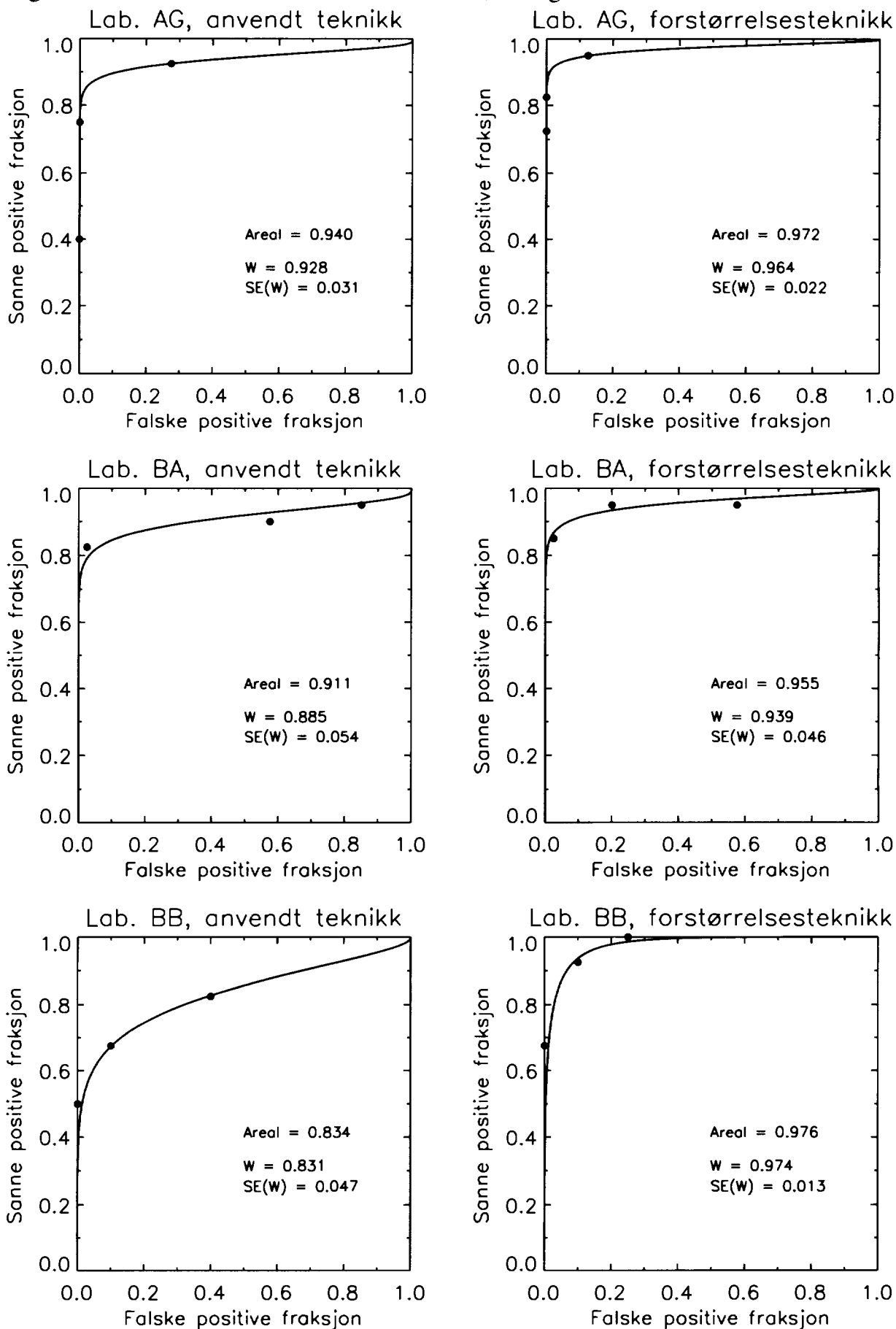
Figur VI.1 ROC-kurvene for laboratoriene AA, AB og AC.



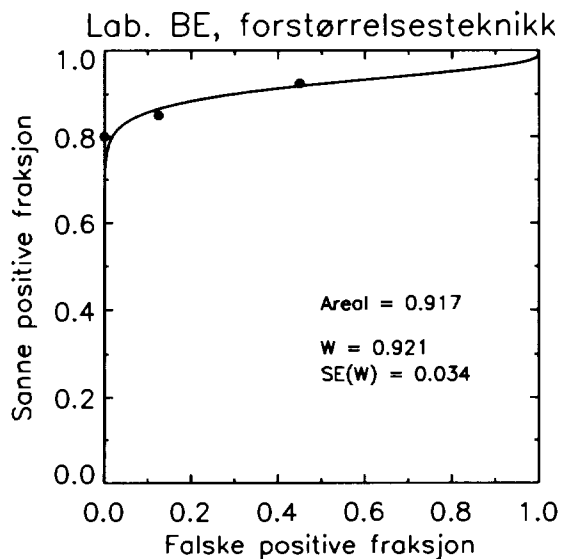
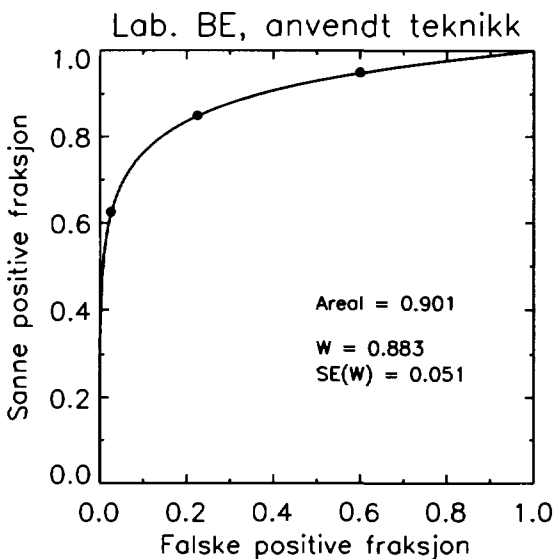
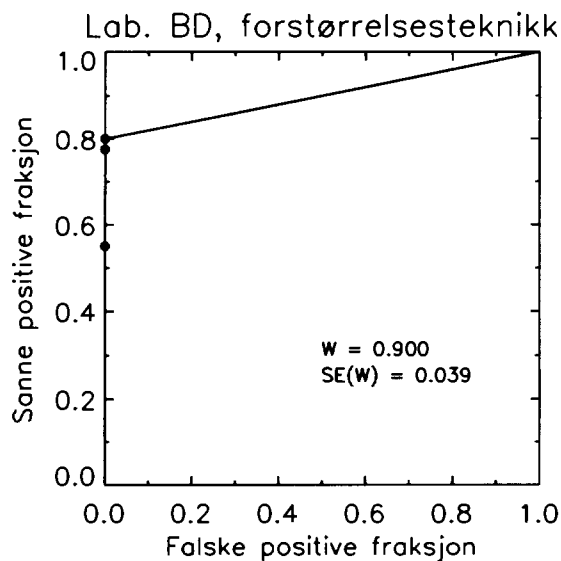
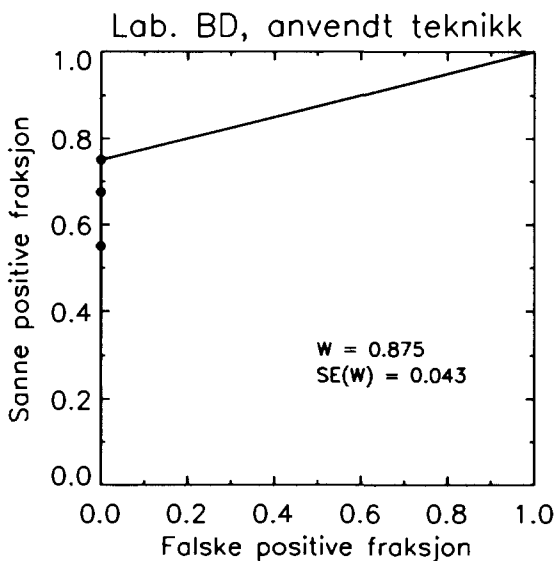
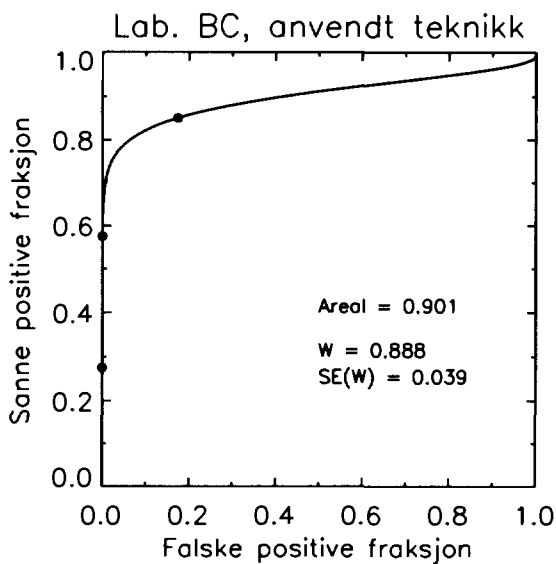
Figur VI.2 ROC-kurvene for laboratoriene AD, AE og AF.



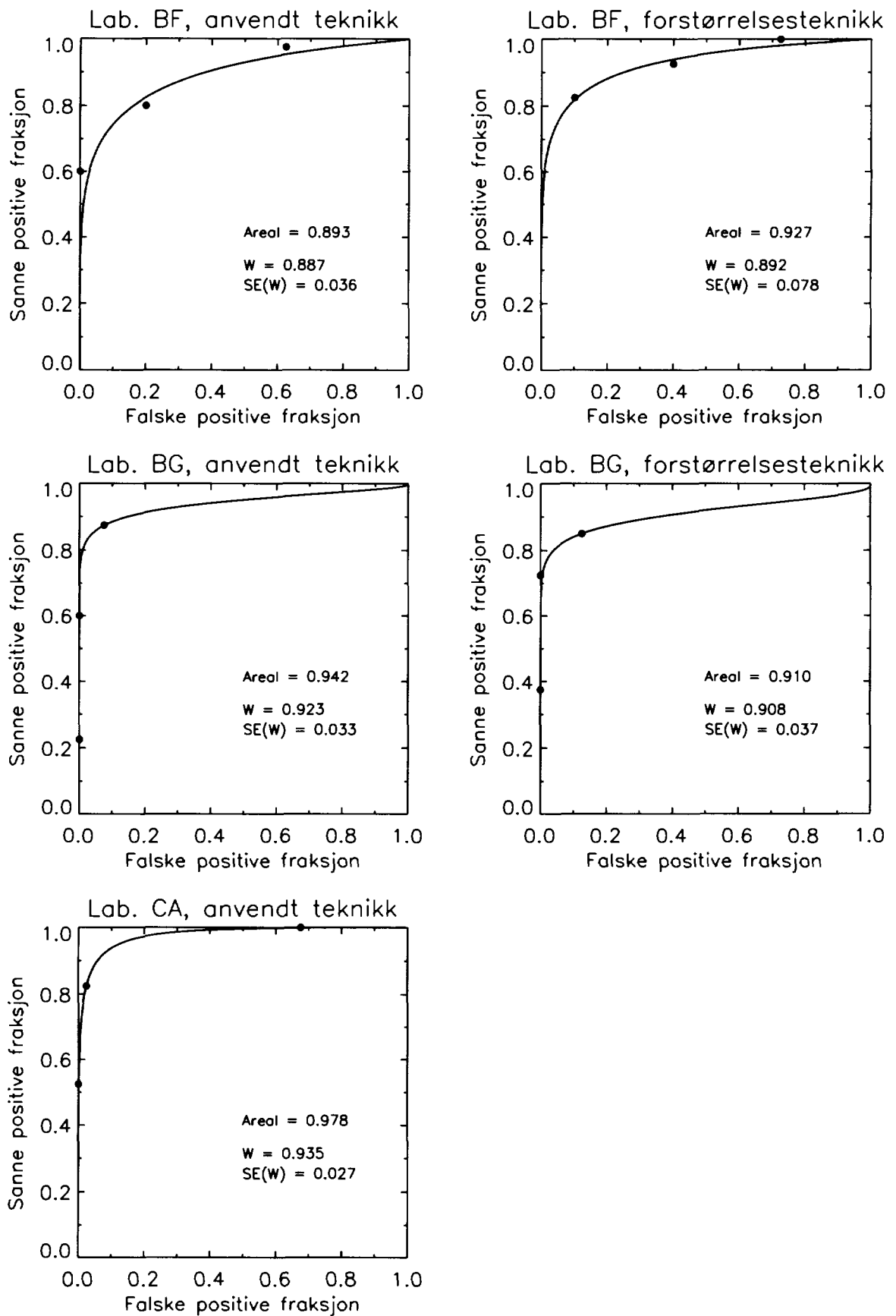
Figur VI.3 ROC-kurvene for laboratoriene AG, BA og BB.



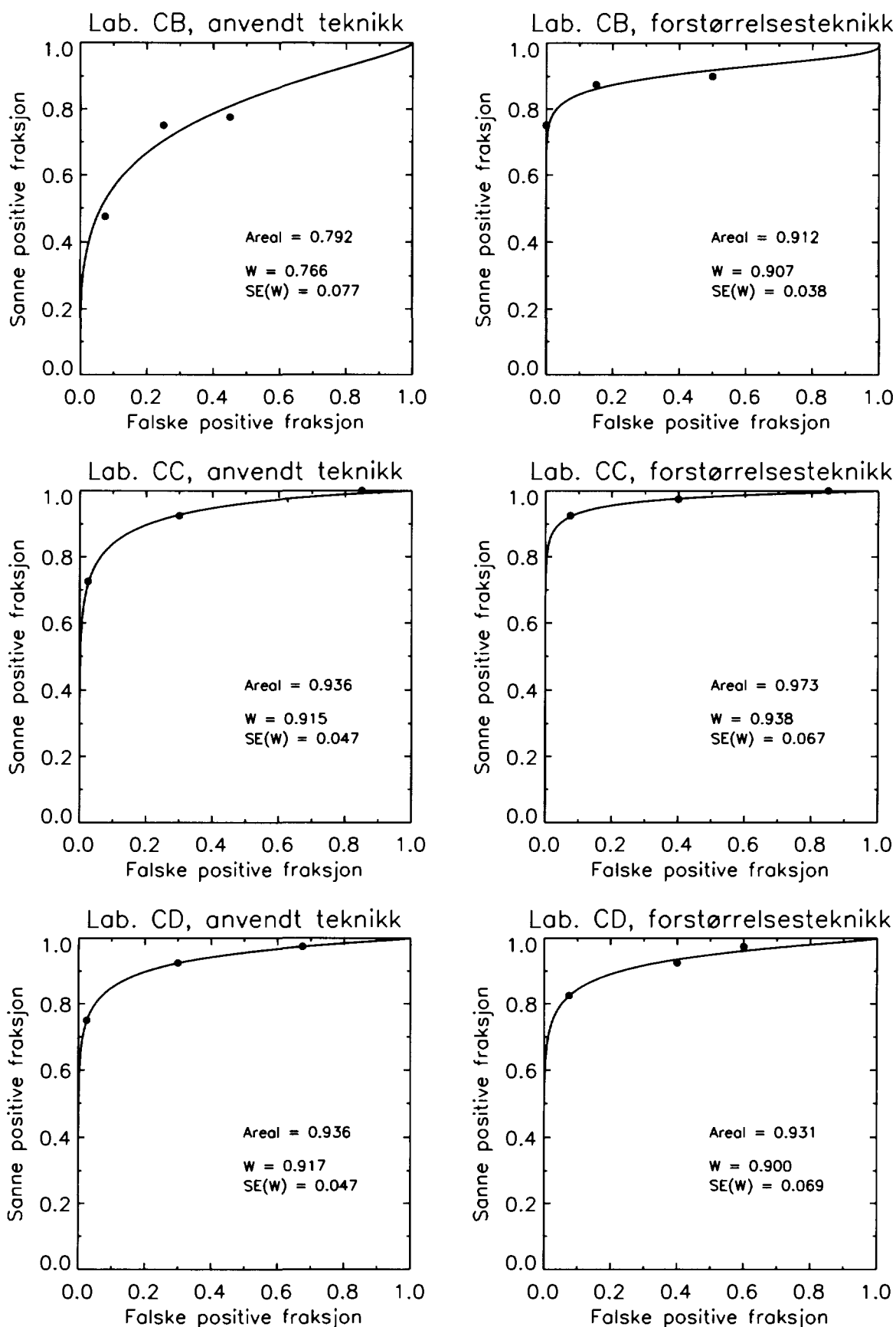
Figur VI.4 ROC-kurvene for laboratoriene BC, BD og BE.



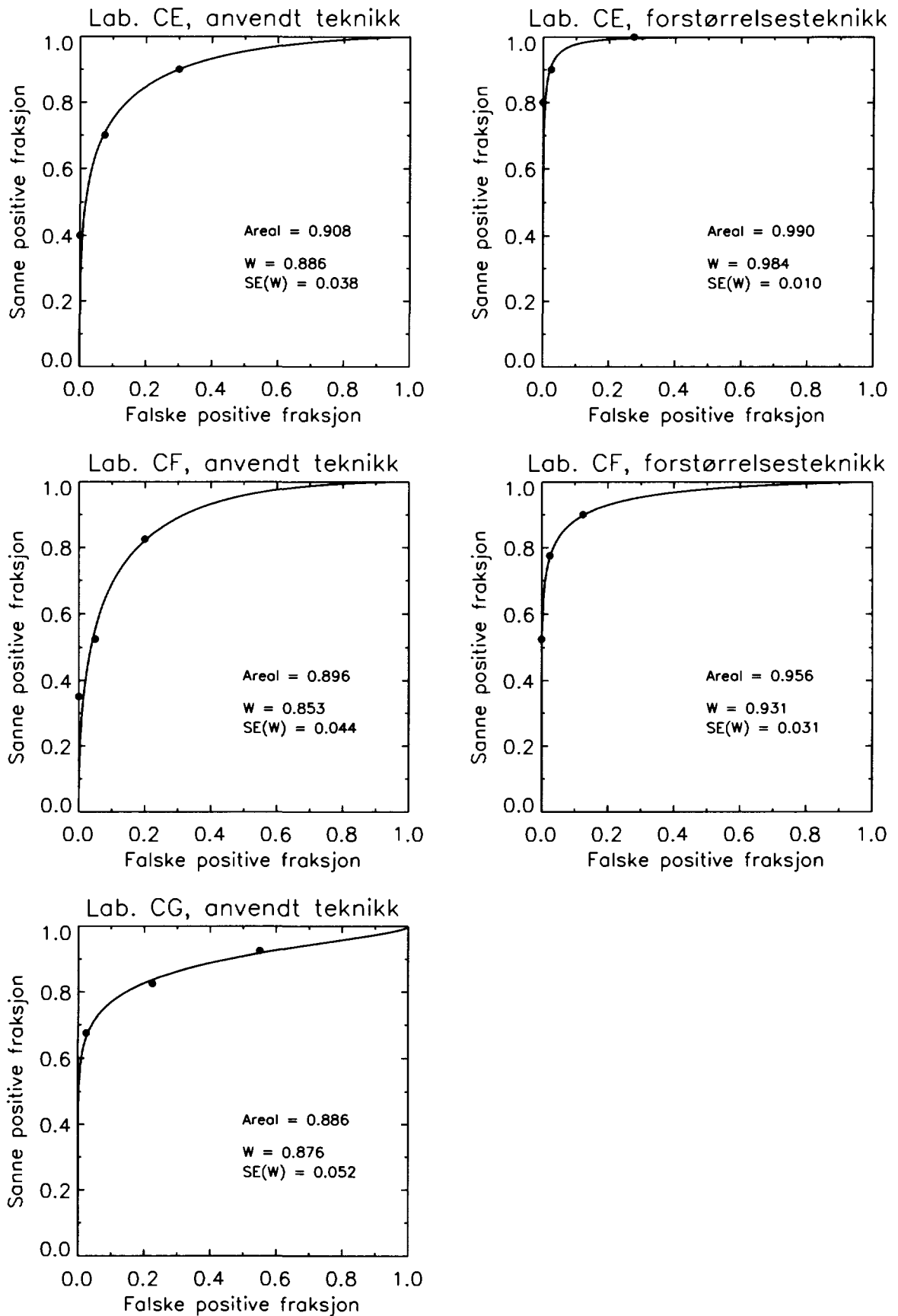
Figur VI.5 ROC-kurvene for laboratoriene BF, BG og CA.



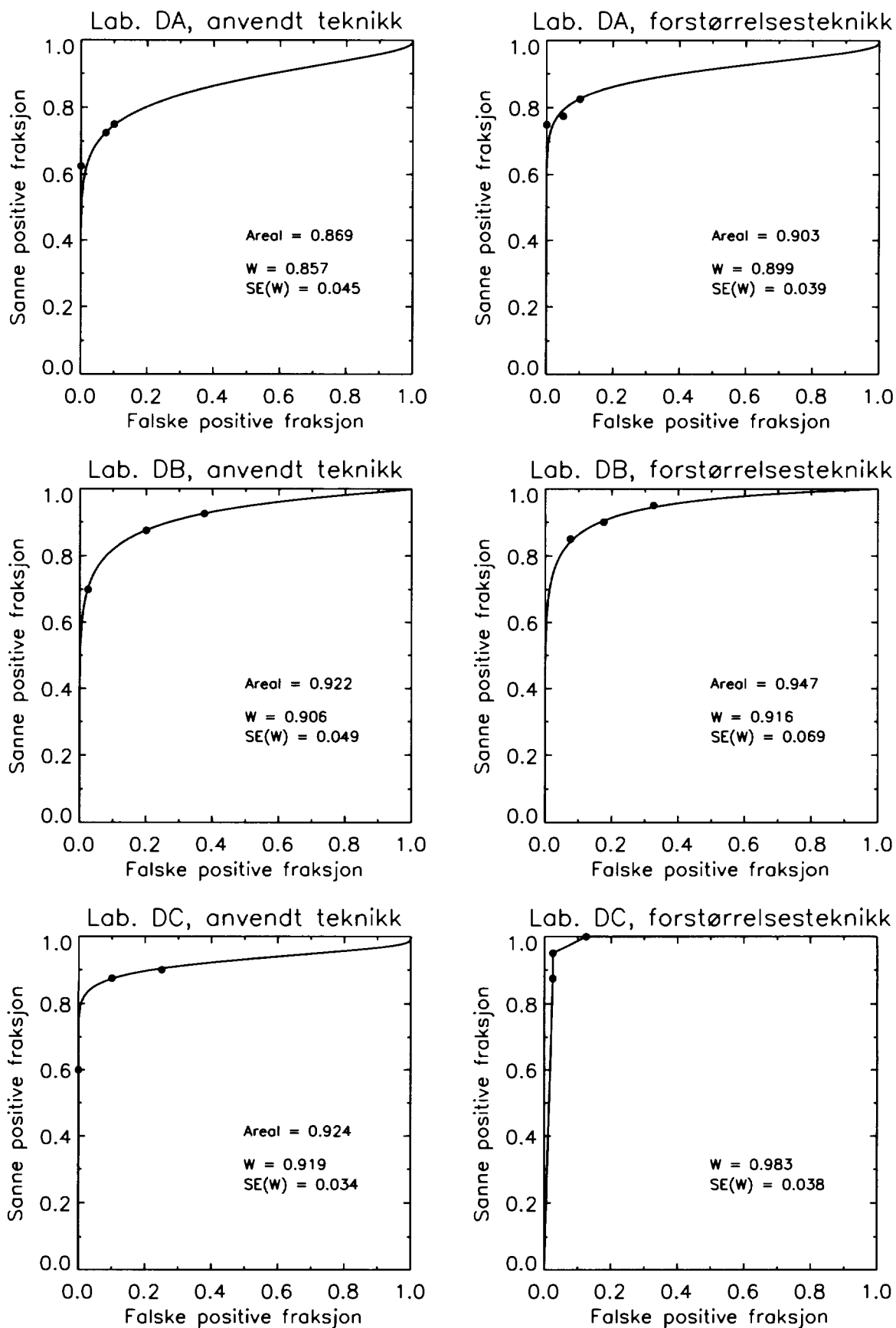
Figur VI.6 ROC-kurvene for laboratoriene CB, CC og CD.



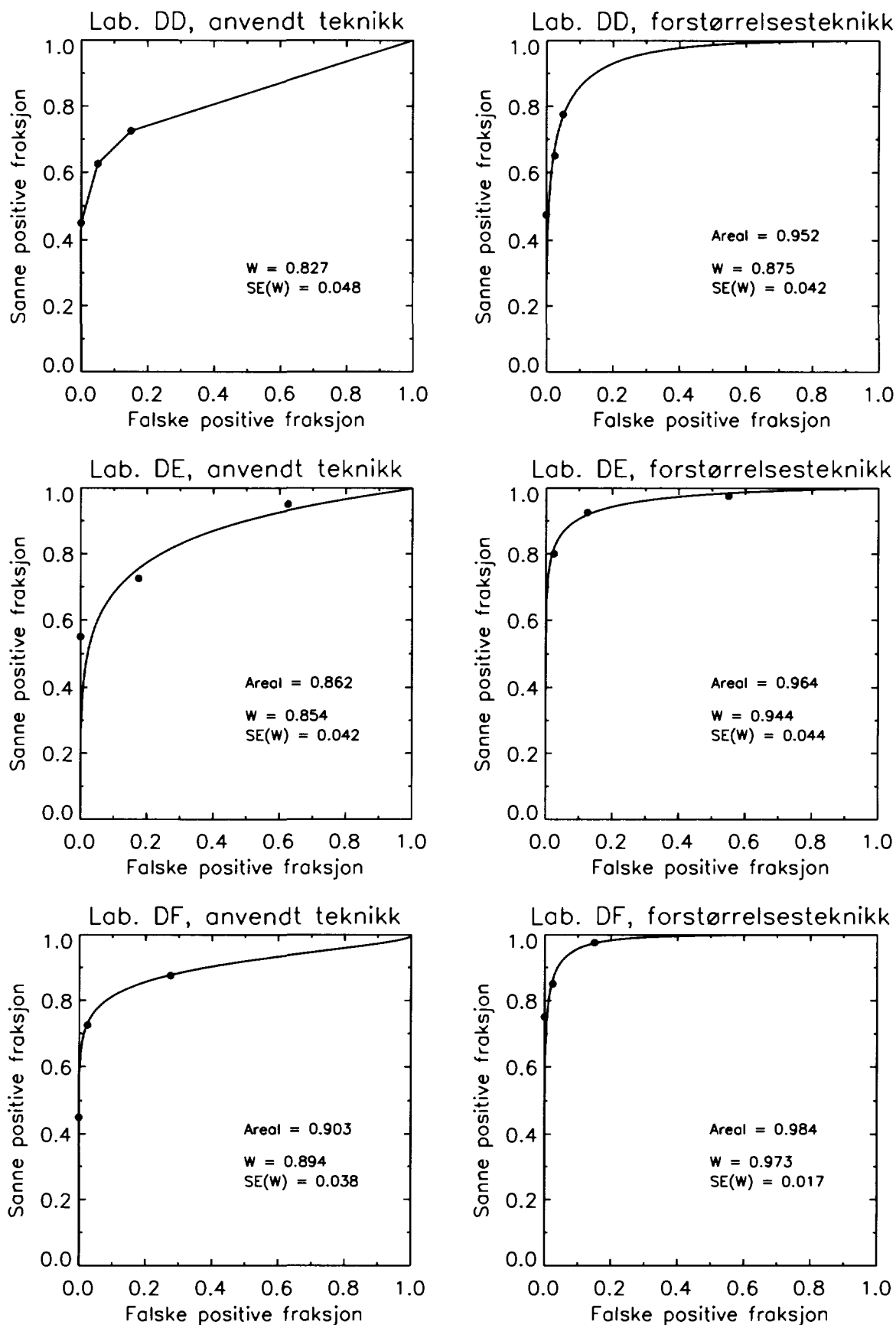
Figur VI.7 ROC-kurvene for laboratoriene CE, CF og CG.



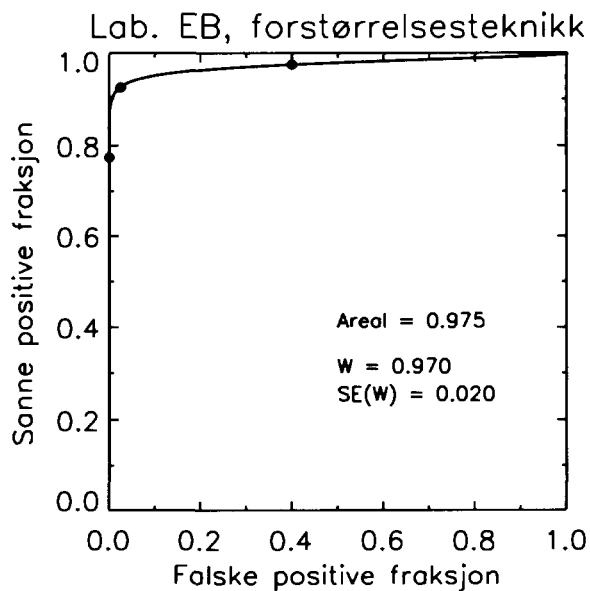
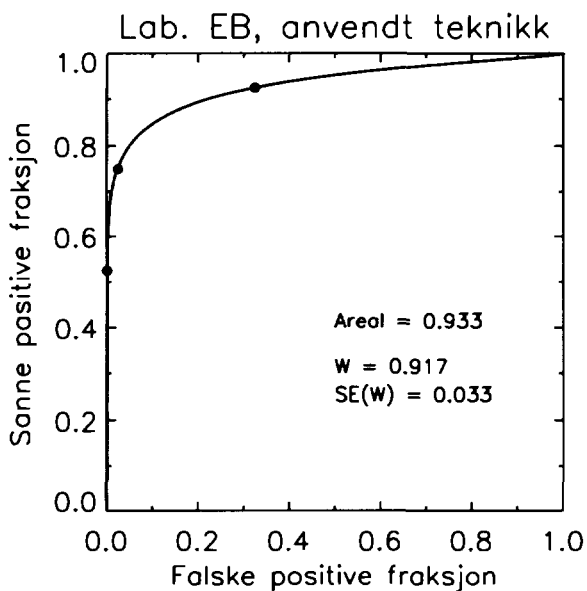
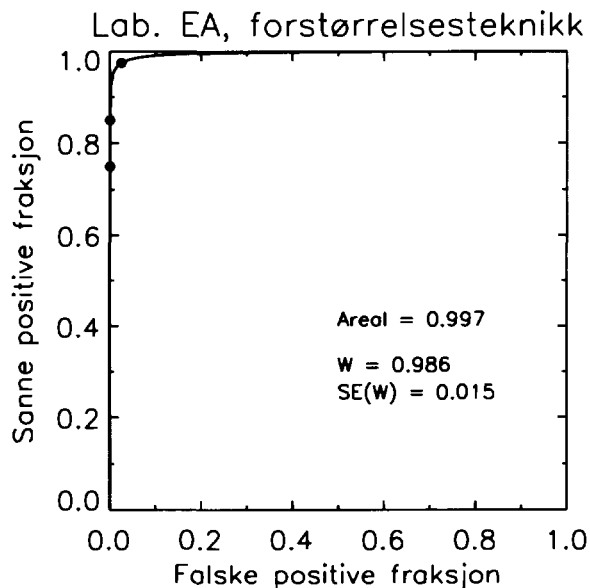
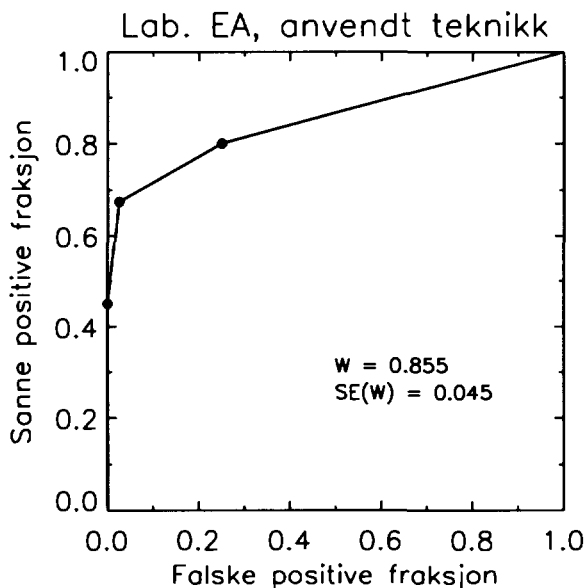
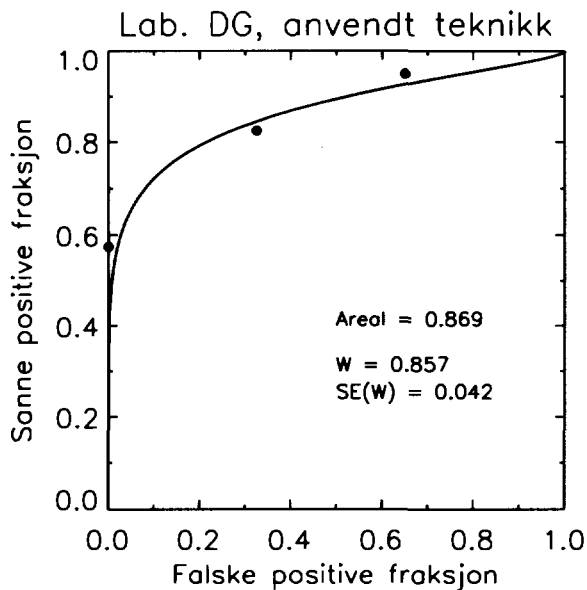
Figur VI.8 ROC-kurvene for laboratoriene DA, DB og DC.



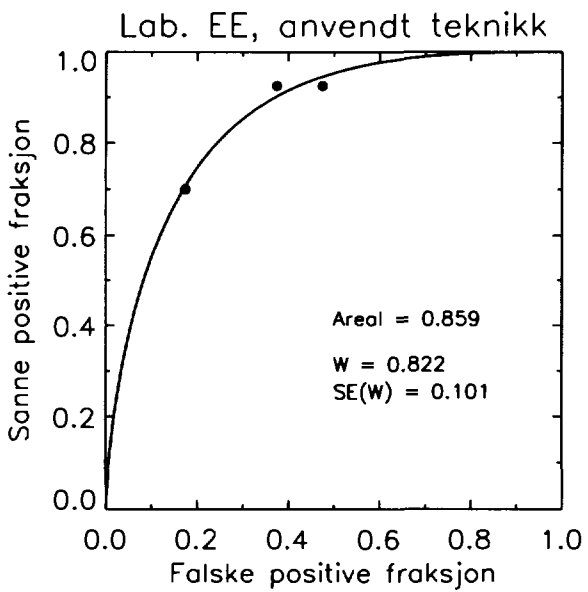
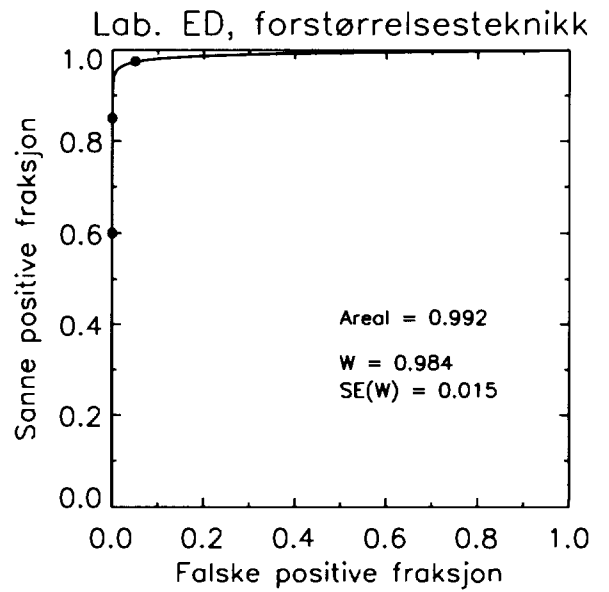
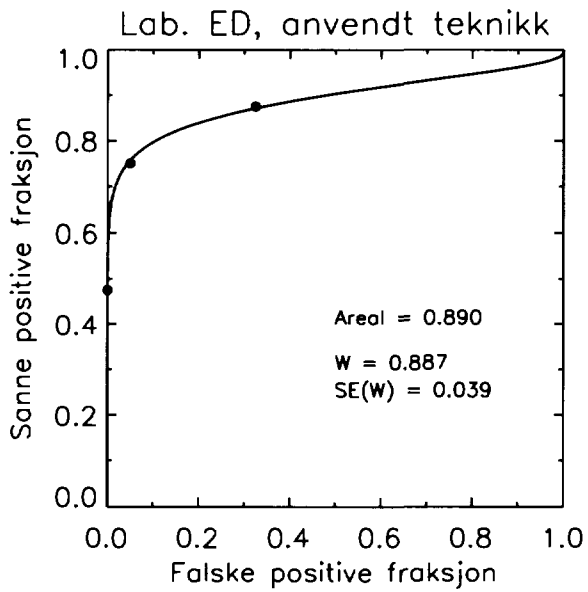
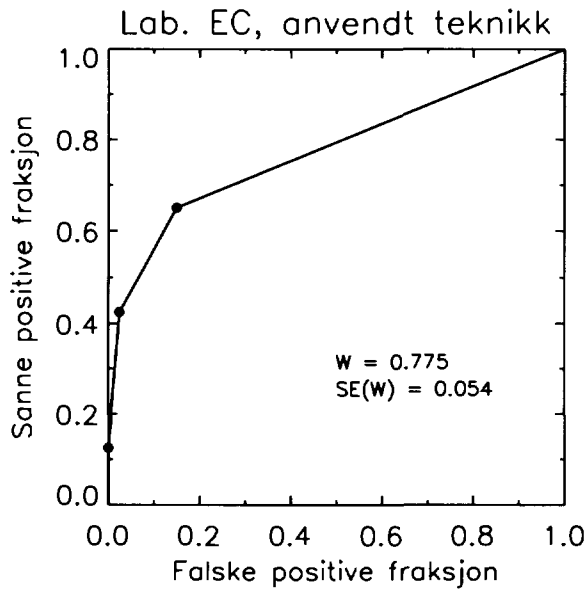
Figur VI.9 ROC-kurvene for laboratoriene DD, DE og DF.



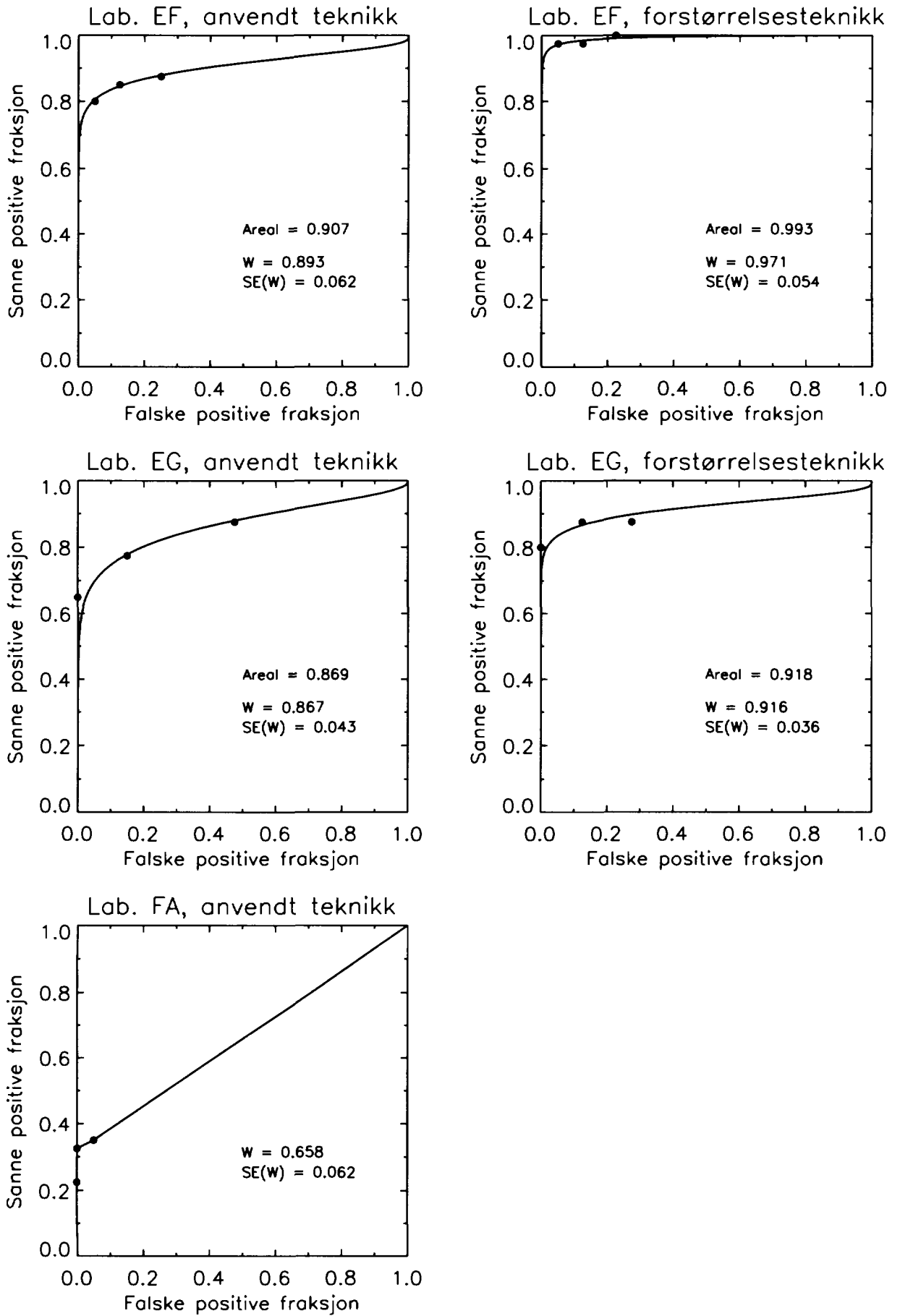
Figur VI.10 ROC-kurvene for laboratoriene DG, EA og EB.



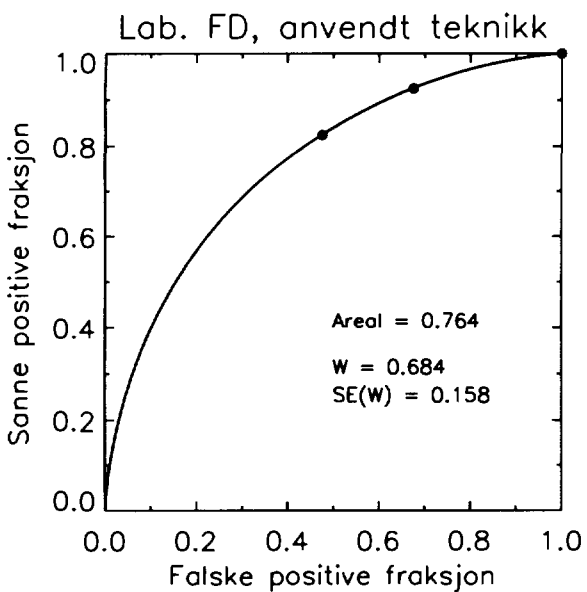
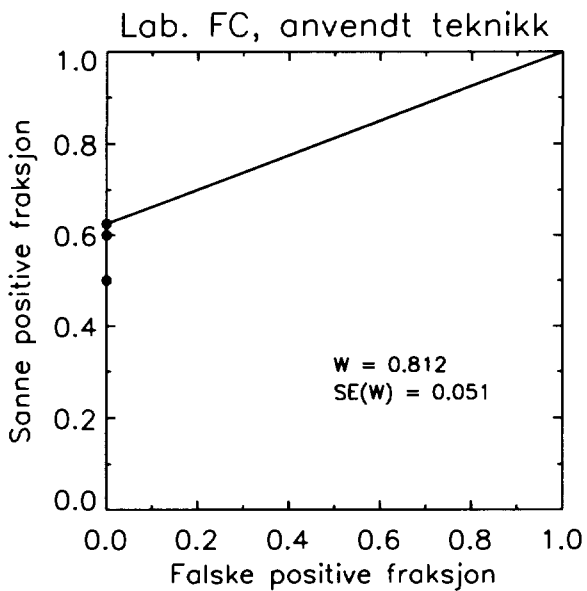
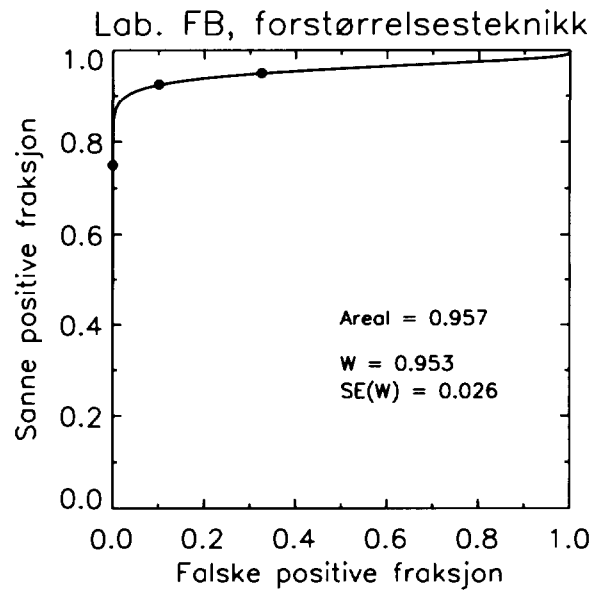
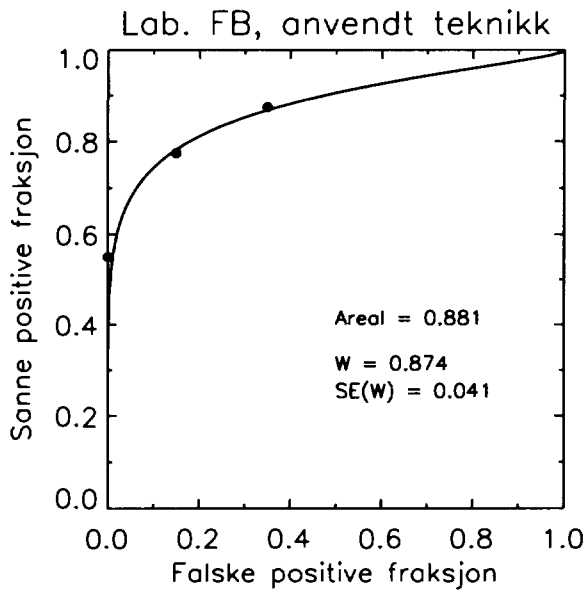
Figur VI.11 ROC-kurvene for laboratoriene EC, ED og EE.



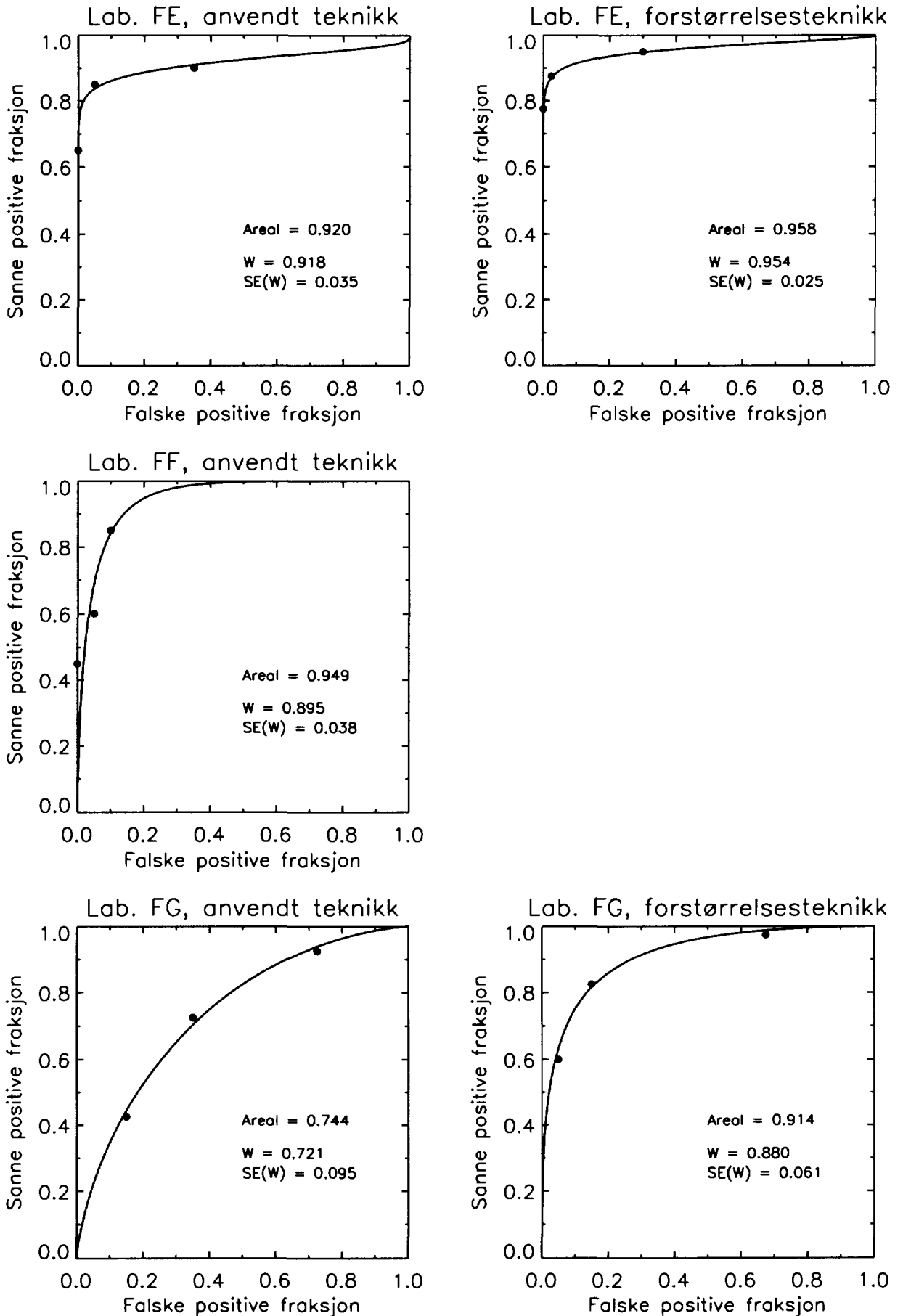
Figur VI.12 ROC-kurvene for laboratoriene EF, EG og FA.



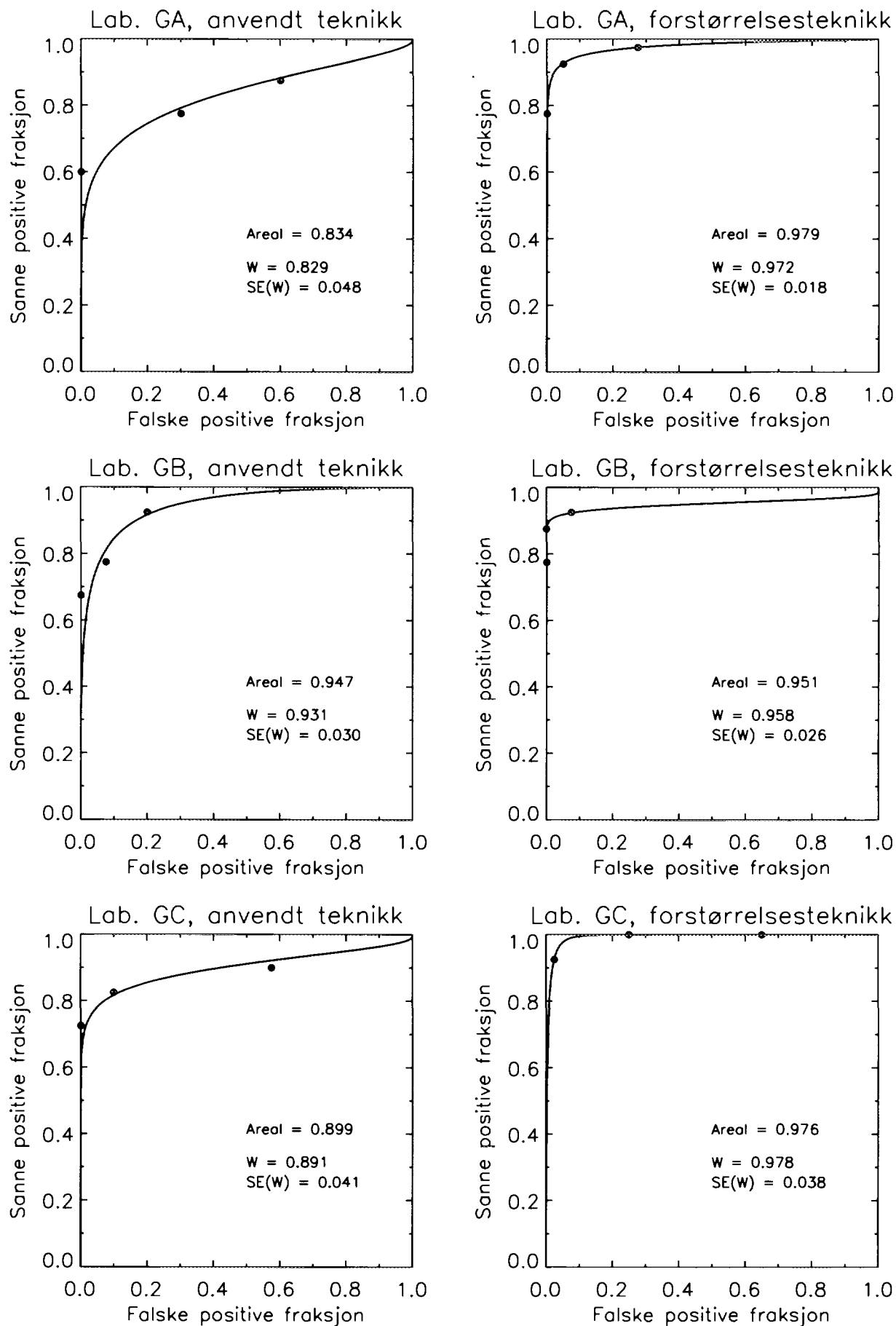
Figur VI.13 ROC-kurvene for laboratoriene FB, FC og FD.



Figur VI.14 ROC-kurvene for laboratoriene FE, FF og FG.



Figur VI.15 ROC-kurvene for laboratoriene GA, GB og GC.



Vedlegg VII. Oppsummering av resultatene

Tabell VII.1 Oppsummering av resultatene ved anvendt teknikk.

| Lab kode | NRT | | | | | RMI | | | NRH | | NRH-ROC |
|-----------------|----------------|----------------|----------------|----------------|----|-------|------|---------|--------|---------|---------|
| | C ₁ | C ₂ | R ₁ | R ₂ | CD | Fibr. | Kalk | Fortem. | Kontr. | Skarph. | Areal |
| AA | | | | | ÷ | | | | | | |
| AB | | | | | ÷ | | | | | | |
| AC | | | | | | | | | | | |
| AD | | | | | | | | | | | |
| AE | | | | | ÷ | | | | | | |
| AF | | | | | | | | | | | ÷ |
| AG | | | | | | | | | | | |
| BA | ÷ | ÷ | | | ÷ | | | | | | |
| BB | ÷ | | | | ÷ | | | | | | ÷ |
| BC | ÷ | ÷ | | | ÷ | | | | | ÷ | |
| BD | | | | | | | | | | | |
| BE | | | | | ÷ | | | | | | |
| BF | | | | | ÷ | | | | | | |
| BG | | | | | | | | | | | |
| CA | | | | | | | | | | | |
| CB | ÷ | ÷ | | | ÷ | | | | ÷ | ÷ | ÷ |
| CC | | | | | ÷ | | | | | | |
| CD | | | | | | | | | | | |
| CE [†] | ÷ | ÷ | | | ÷ | | | | ÷ | ÷ | |
| CF | ÷ | ÷ | | | | | ÷ | | | | |
| CG | | | | | | | | | | | |
| DA | | | | | ÷ | | | | | | |
| DB | | | | | | | | | | | |
| DC | ÷ | ÷ | | | | | | | | | |
| DD | | | | | ÷ | ÷ | | | | ÷ | ÷ |
| DE | ÷ | ÷ | | ÷ | ÷ | | | | ÷ | ÷ | |
| DF | ÷ | | | | ÷ | | | | | | |
| DG | ÷ | ÷ | | | ÷ | ÷ | ÷ | | | | |
| EA | ÷ | ÷ | | | | | | | | | ÷ |
| EB | | | | | | | | | | | |
| EC [†] | ÷ | ÷ | | | ÷ | | | | | ÷ | ÷ |
| ED | | | | | | | | | | | |
| EE | | | | | | | | | | | |
| EF | ÷ | | | | | | | | | | |
| EG | | | ÷ | ÷ | | | | | | ÷ | |
| FA | ÷ | ÷ | | | ÷ | ÷ | | | ÷ | ÷ | ÷ |
| FB | ÷ | ÷ | | | | | | | | | |
| FC | | | | | ÷ | ÷ | | | | | ÷ |
| FD | ÷ | ÷ | | ÷ | ÷ | | | | | | ÷ |
| FE | | | | | | | | | | | |
| FF | | | | | | | | | | | |
| FG | ÷ | | | | ÷ | ÷ | | | | | ÷ |
| GA | | ÷ | | | | | | | | | ÷ |
| GB | | ÷ | | | | | | | | | |
| GC | | | | | | | | | | | |

[†]Rasterteknikk ble ikke benyttet.

Tabell VII.2 Oppsummering av resultatene ved forstørrelsesteknikk.

| Lab. kode | NRT | | | | NRH | | NRH-ROC |
|-----------------|----------|----------|-------|-------|--------|---------|---------|
| | C_{24} | C_{30} | R_1 | R_2 | Kontv. | Skarpb. | Areal |
| AA | | | | | | | |
| AB | | | | + | | | |
| AC | | | | ÷ | | | |
| AD | | | | + | | | |
| AE | ÷ | | | | | | |
| AF | | | | | | | |
| AG | | | | ÷ | | | |
| BA | + | + | | | | | |
| BB | | | | | | | |
| BC [†] | | | | | | | |
| BD | | | | ÷ | | | ÷ |
| BE | | | + | + | | | ÷ |
| BF | | ÷ | | ÷ | | | ÷ |
| BG | | | | | | | ÷ |
| CA [†] | | | | | | | |
| CB | | + | | + | | | + |
| CC | | | | | | | |
| CD | | | | | | | ÷ |
| CE | ÷ | ÷ | ÷ | ÷ | ÷ | ÷ | |
| CF | + | + | | + | + | + | |
| CG [†] | | | | | | | |
| DA | | | | | | | + |
| DB | | | | ÷ | | | |
| DC | | | | + | | | |
| DD | ÷ | | | | ÷ | ÷ | |
| DE | | | + | + | + | + | |
| DF | | | | ÷ | | | |
| DG [†] | | | | | | | |
| EA | ÷ | ÷ | | ÷ | ÷ | ÷ | |
| EB | + | | | + | | | |
| EC [†] | | | | | | | |
| ED | | | | | | | |
| EE [†] | | | | | | | |
| EF | | | | + | | | |
| EG | | | | ÷ | | | ÷ |
| FA [†] | | | | | | | |
| FB | | ÷ | | ÷ | | | |
| FC [†] | | | | | | | |
| FD [†] | | | | | | | |
| FE | | | | + | | | |
| FF [†] | | | | | | | |
| FG | + | | + | | | | + |
| GA | | | | ÷ | | | |
| GB | + | | + | + | | | |
| GC | | | ÷ | ÷ | | | |

[†]Forstørrelsesteknikk ble ikke benyttet.