

Aus dem Institut für Radiologie
der Medizinischen Fakultät Charité - Universitätsmedizin Berlin

DISSERTATION

Untersuchung des Einflusses der Brustdichte und des
Parenchymvolumens auf die Ergebnisse von
CAD-Untersuchungen in der digitalen Mammographie

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité - Universitätsmedizin Berlin

von

Raphael Witlandt
geb. Bremme
aus Düsseldorf

Gutachter/in: 1. Herr Prof. Dr. med. U. Bick
 2. Frau Priv.-Doz. Dr. med. U. Kettritz
 3. Herr Prof. Dr. med. J. Hierholzer

Datum der Promotion: 22.6.2014

1. Einleitung	5
1.1 Mammakarzinom	5
1.1.1 Definition	5
1.1.2 Epidemiologie	5
1.1.3 Pathogenese und Pathologie	5
1.1.4 Klinische Symptomatik	6
1.1.5 Diagnostik	6
1.1.6 Therapie	7
1.2 Mammographie.....	7
1.2.1 Technische Grundlagen	7
1.2.2 Digitale Mammographie	7
1.2.3 Befundung.....	8
1.2.4 Mammographie-Screening.....	9
1.2.5 Mammographische Brustdichte.....	10
1.3 CAD-Systeme.....	11
2. Literaturbericht und Fragestellung	14
2.1 Beeinträchtigung der Sensitivität.....	14
2.2 Beeinträchtigung der Spezifität.....	15
2.3 Fragestellung	16
3. Material und Methoden	17
3.1 Auswahlkriterien	17
3.2 Verwendete Geräte und Systeme.....	18
3.2.1 CAD-Algorithmus	18
3.2.2 Algorithmus zur Analyse der Gewebeeigenschaften	21
3.2.3 Mammographiegerät	22
3.2.4 Weitere Software	23
3.3 Studiendurchführung	23
3.4 Statistische Methoden	24

4. Ergebnisse	26
4.1 Studienkollektiv.....	26
4.2 Ergebnisse der Computeralgorithmen.....	27
4.2.1 Ergebnisse des Algorithmus zur Analyse der Gewebeeigenschaften	27
4.2.2 Ergebnisse des CAD-Algorithmus	30
4.3 Statistische Analyse der Ergebnisse	32
4.3.1 Spezifität des CAD-Algorithmus	32
4.3.2 Kategorisierung nach Dichte und Parenchymvolumen	32
4.3.3 Kategorisierung nach Anzahl der Marker	37
4.3.4 Korrelationsanalyse.....	41
4.3.5 Regressionsanalyse	45
5. Diskussion	46
5.1 Einordnung der Ergebnisse	46
5.1.1 Einfluss der Brustdichte.....	46
5.1.2 Einfluss des Parenchymvolumens	48
5.2 Limitationen und Schwächen der Arbeit	49
5.2.1 Zusammenstellung der Studienpopulation	49
5.2.2 Auswahl der Untersuchungs- und Analysemethoden.....	50
5.2.3 Einschränkungen der Arbeit	51
5.2.4 Verzerrung der Ergebnisse durch die CAD-Ausgabe.....	52
5.3 Konsequenzen der Arbeit	52
6. Zusammenfassung	54
Literaturverzeichnis	56
Abbildungsverzeichnis	60
Tabellenverzeichnis	62
Abkürzungsverzeichnis	64
Erklärung	65
Lebenslauf	66
Danksagung	67

1. Einleitung

1.1 Mammakarzinom

1.1.1 Definition

Unterschiedliche maligne Befunde der Mamma werden unter der Bezeichnung „Mammakarzinom“ subsummiert. Die Erkrankung wird als Systemerkrankung mit lokoregionaler Komponente betrachtet¹. Dieses Modell hat die Vorstellung einer rein lokalen Erkrankung ersetzt. Moderne Therapieverfahren gehen daher über den lokalen Aspekt hinaus und behandeln die Erkrankung interdisziplinär.

1.1.2 Epidemiologie

Das Mammakarzinom ist mit einem lebenslangen Erkrankungsrisiko von etwa 12% das Karzinom mit der größten Prävalenz unter den Krebserkrankungen in der weiblichen Bevölkerung der westlichen Industrienationen². Nur 0,6% der Karzinome treten bei Männern auf. In Deutschland starben im Jahr 2010 17.466 Frauen und 107 Männer an den Folgen der Erkrankung³. Die Mortalität ist seit den 1990er Jahren rückläufig, wohingegen die Inzidenz seit 1970 stetig ansteigt¹. Der wichtigste Risikofaktor für das Auftreten eines Mammakarzinoms ist, nach dem Geschlecht, das Alter.

In der Bevölkerung der USA, in England und Deutschland ist die Inzidenz höher als z.B. in Japan. Als Auslöser für die epidemiologischen Unterschiede werden unter anderem genetische Faktoren, Ernährung und Fertilität diskutiert⁴.

1.1.3 Pathogenese und Pathologie

Die exakte Pathogenese ist bis heute unbekannt. Das Karzinom entsteht wahrscheinlich als Folge von Veränderungen auf mehreren Ebenen der Zellproliferationsregulation. Man unterscheidet spontan entstehende, „sporadische“ Karzinome von „hereditären“ Karzinomen, denen eine Genmutation zugrunde liegt. Sporadische Karzinome machen dabei mit ca. 90% die deutliche Mehrheit unter den Neuerkrankungen aus⁵.

Histopathologisch handelt es sich beim Mammakarzinom um eine sehr heterogene Erkrankung. Es wird zwischen invasiven und nicht-invasiven Karzinomen unterschieden. Um die Lokalisation eines Tumors zu beschreiben, wird die Brust in vier Quadranten und eine zentrale, retromammillär gelegene Zone unterteilt. Etwa die Hälfte

der Mammakarzinome treten im oberen äußeren Quadranten auf. Im unteren inneren Quadranten finden sich die wenigsten Erkrankungen.

1.1.4 Klinische Symptomatik

Eine Frühsymptomatik des Mammakarzinoms gibt es nicht. Fällt der Tumor nicht in einer Früherkennungs- oder Screeninguntersuchung auf, so bildet sich als Leitsymptom ein Knoten, der oft ab einem Durchmesser von etwa 1-2 cm palpabel ist.

1.1.5 Diagnostik

Die klinische Untersuchung umfasst die vergleichende Inspektion und Palpation beider Mammae, der Axillen und der Supraklavikularregion. Bei suspekten Befunden oder in der Früherkennung steht außerdem apparative Diagnostik zur Verfügung.

Das zur Zeit insbesondere wegen des Einsatzes im primären Screening wohl wichtigste Werkzeug in der Mammadiagnostik ist die Mammographie. Das Verfahren wird trotz geringer Sensitivität insbesondere bei dichten Drüsengewebe prinzipiell zur Primärdiagnostik eingesetzt. Eine Untersuchung umfasst Aufnahmen beider Mammae in jeweils mindestens zwei Ebenen. Die Sensitivität liegt in fettreichem Gewebe bei 80-90% und nimmt in dichtem Gewebe auf 50-60% ab⁶.

Die Sonographie ist weniger abhängig von der Brustdichte und ist daher als Zusatzuntersuchung vor allem bei dichterem Drüsengewebe hilfreich⁷. Sie wird außerdem zur differentialdiagnostischen Abklärung verwendet. Allerdings ist diese Methode stark von der Fähigkeit des Untersuchers abhängig und Mikrokalk lässt sich nicht ohne Weiteres darstellen.

Die Magnetresonanztomographie hat in den letzten Jahren an Bedeutung in der Mammadiagnostik gewonnen. Die Vorteile liegen vor allem in der hohen Kontrastauflösung und in der überlagerungsfreien Darstellung der Befunde. Nach i.v.-Kontrastmittelgabe ist die Sensitivität auch für kleine Läsionen sehr hoch. Problematisch sind aber unter Umständen der große Aufwand und die hohen Kosten, die diese Untersuchungen verursachen⁸.

Die endgültige Diagnosestellung erfolgt meistens über die Gewinnung einer Gewebeprobe zur histologischen Klassifizierung der Läsion. Dies kann mittels einer Stanzbiopsie unter sonographischer Kontrolle oder stereotaktisch unter Mammographiekontrolle erfolgen.

1.1.6 Therapie

Da das Mammakarzinom heute auch ohne den Nachweis von Fernmetastasen als Systemerkrankung angesehen wird, umfasst die Therapie je nach Risikoprofil sowohl lokale als auch systemische Behandlungsstrategien.

Die lokoregionäre Therapie umfasst die Operation der betroffenen Mamma, sowie gegebenenfalls eine Strahlentherapie. Eine Operation wird entweder als brusterhaltende Therapie, oder als modifizierte radikale Mastektomie durchgeführt.

Die systemische Therapie erfolgt entweder postoperativ als adjuvante Systemtherapie oder präoperativ als neoadjuvante Systemtherapie. Sie kann aus Chemotherapie, endokriner Therapie, Antikörpertherapie oder einer Kombination daraus bestehen.

1.2 Mammographie

1.2.1 Technische Grundlagen

Die Mammographie ist ein wichtiges Instrument der Diagnostik von Mammakarzinomen. Dabei besteht die Schwierigkeit, dass die Dichteunterschiede der verschiedenen Gewebe der Brust verhältnismäßig gering sind, was eine radiologische Bildgebung erschwert. Um einen ausreichenden Kontrast zu gewährleisten, wird die Weichstrahltechnik mit 25-30 kV Spannung angewandt. Mit neueren Geräten kann dabei die Strahlenexposition relativ gering gehalten werden. Die normale Parenchymdosis liegt bei ≤ 3 mGy für zwei Aufnahmen⁹.

Um Überlagerungen des Gewebes zu minimieren, Bewegung zu vermeiden und die Abbildungsschärfe zu verbessern, wird die Brust während der Aufnahme komprimiert. Dies trägt außerdem zur Verringerung der Strahlendosis bei.

1.2.2 Digitale Mammographie

Seit dem Ende der 1990er Jahre werden zunehmend digitale Mammographietechniken eingesetzt¹⁰. Dabei vollzog sich die schrittweise Abkehr von analogen Methoden im Vergleich zu anderen Bereichen der medizinischen Bildgebung relativ spät. Zunächst wurde oft eine zu geringe Auflösung der Sensoren bemängelt. Heute ist die Digitalmammographie aber als dem analogen Verfahren mindestens gleichwertig, oft sogar als überlegen, allgemein anerkannt¹¹.

Aus diesen Neuerungen ergeben sich einige Vorteile. Exemplarisch sei hier die wesentliche Vereinfachung des Managements großer Mammographiebestände

genannt, was vor allem in größeren Kliniken und beim Screening eine enorme Erleichterung darstellt. Außerdem ergeben sich aus der Tatsache, dass die Bilder digital vorliegen, noch weitere Möglichkeiten. So kann jetzt eine digitale Tomosynthese angefertigt werden, die überlagerungsfreie Schnittbilder der Brust liefert und es ist besser möglich, die Aufnahmen von Computeralgorithmen analysieren zu lassen.

1.2.3 Befundung

Da die Mammographie eine zweidimensionale Untersuchung ist, sind mindestens zwei Aufnahmen in unterschiedlichen Ebenen nötig, um räumliche Befunde beurteilen zu können. Üblicherweise werden die Aufnahmen mit cranio-caudalem sowie mediolateral-obliquem Strahlengang angefertigt.

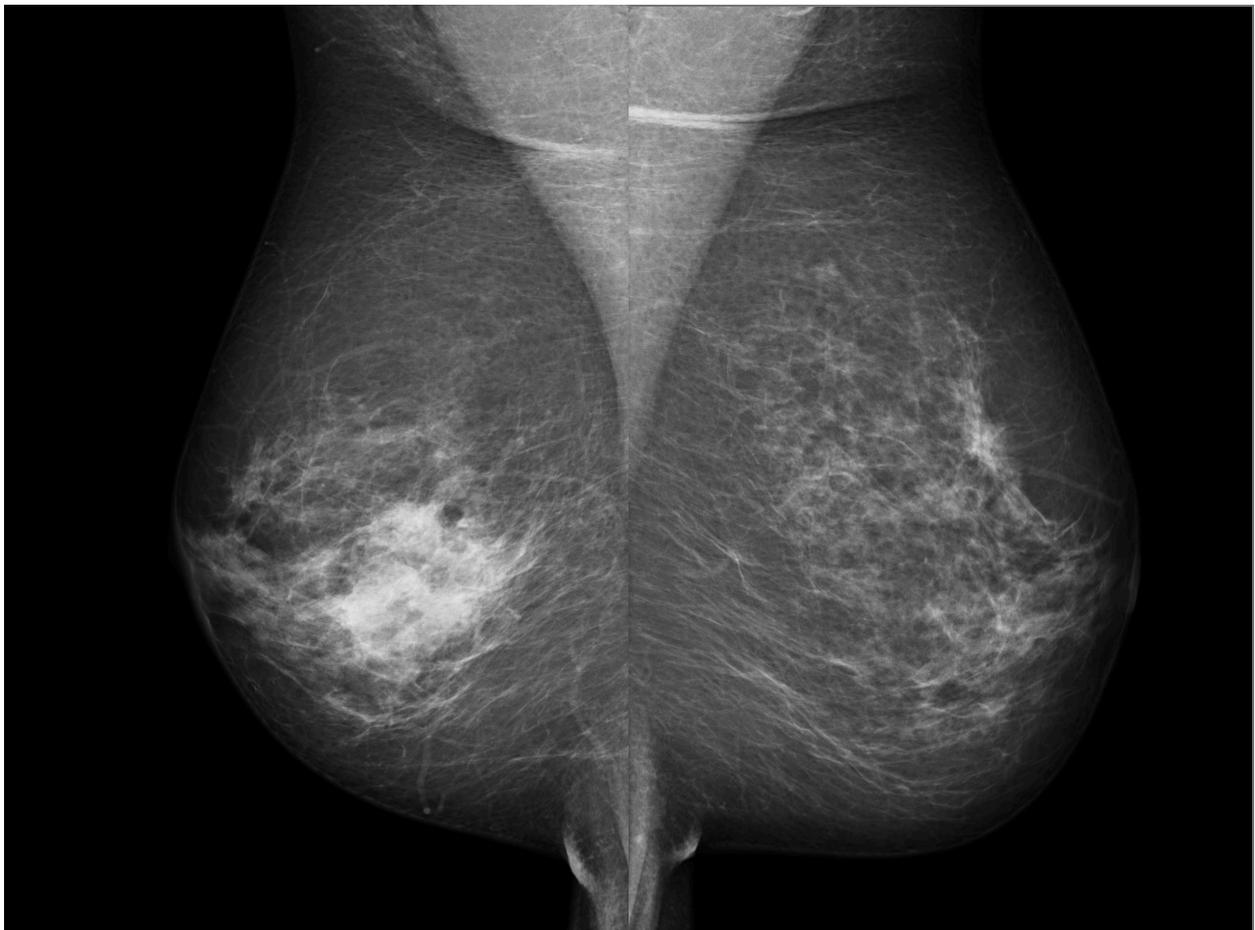


Abbildung 1: Mammographie in mediolateral-obliquem Strahlengang mit suspekter Verdichtung rechts

Häufigste mammographische Hinweise auf Malignome sind Verdichtungen, Herdbefunde und Mikroverkalkungen. Dazu kommen weitere, oft diskrete Veränderungen wie umschriebene Störungen der Parenchymarchitektur, Asymmetrie

des Drüsenkörpers, Seitenasymmetrie und Erweiterungen eines oder mehrerer Milchgänge. Liegen Voraufnahmen vor, können außerdem eine zunehmende Verdichtung oder Vergrößerung suspekter Herde ein Hinweis auf möglicherweise pathologische Vorgänge sein.

Mikroverkalkungen sind ein weiterer wichtiger Anhaltspunkt bei der Befundung von Mammographien. Insgesamt bilden etwa 50% der Mammakarzinome Verkalkungen. Sie entstehen entweder durch aktive Calciumsekretion oder durch passive Calciumbindung¹². Bei Malignomen stellen sich Mikrokalzifikationen meist polymorph sowie unterschiedlich groß und dicht dar. Außerdem stehen sie meist gruppiert. Je größer die Anzahl suspekter Verkalkungen, desto größer ist die Wahrscheinlichkeit einer malignen Entartung.

Um die Ergebnisse der Mammographiebefundung zu standardisieren, wurde die Klassifikation des Breast Imaging Reporting and Data System (BI-RADS Klassifikation) des American College of Radiology (ACR) festgelegt.

BI-RADS	Diagnose	Weiteres Vorgehen
0	bildgebende Diagnostik noch nicht abgeschlossen	weitere Bildgebung
I	unauffällig	normale Vorsorge
II	benigne	normale Vorsorge
III	unklar, wahrscheinlich benigne	Kontrolle in 6 Monaten
IV	suspekt	histologische Sicherung
V	malignitätsverdächtig	histologische Sicherung
VI	histologisch gesichertes Mammakarzinom	

Tabelle 1: BI-RADS Klassifikation des American College of Radiology (ACR)¹³

1.2.4 Mammographie-Screening

Neben dem Einsatz im klinischen Alltag zur Abklärung klinischer Auffälligkeiten wie z.B. suspekter Tastbefunde, der so genannten „kurativen Mammographie“, wird die Mammographie seit einigen Jahren auch zum Screening, d.h. zur breiten Früherkennung, eingesetzt. In Deutschland wurde 2008 ein flächendeckendes Screening-Programm gestartet, das alle symptomlosen Frauen zwischen 50 und 69 Jahren einbezieht. Das Untersuchungsintervall beträgt 2 Jahre¹⁴.

1.2.5 Mammographische Brustdichte

Das Drüsenparenchym und das umgebende Binde-Stütz-Gewebe weisen eine höhere Röntgendichte auf als das Fettgewebe der Brust. Der Anteil dieser röntgendichteren Strukturen am Gesamtvolumen der Brust variiert bei der Untersuchung verschiedener Patientinnen teilweise stark. Die Röntgendichte hat aber erheblichen Einfluss auf das Erscheinungsbild der Mammographie.

Verschiedene Studien kommen zu dem Schluss, dass eine erhöhte Brustdichte die Beurteilung einer Mammographie erschwert. Ein negativer Effekt wird sowohl auf die Sensitivität^{6, 15-17}, als auch auf die Spezifität¹⁵ der Mammographie beschrieben.

Da die Brustdichte einen Einfluss auf die Befundung von Mammographien hat, wird dieser Parameter häufig gemeinsam mit den Untersuchungsergebnissen angegeben. Dafür kann die Untersuchung einem Dichte-Typ nach der Empfehlung des American College of Radiology (ACR) zugeordnet werden¹⁸ (s. Tabelle 2).

ACR Typ	Anteil Drüsengewebe	Beschreibung
I	< 25%	überwiegend lipomatös
II	25 - 50%	fibroglandulär
III	51 - 75%	inhomogen dicht
IV	> 75%	extrem dicht

Tabelle 2: Einteilung der Brustdichte nach Empfehlung des American College of Radiology (ACR)¹⁸

Die Schwächen dieser Einteilung liegen vor allem in der Beurteilung von sehr ungleich verteiltem Parenchymgewebe. Außerdem wird diskutiert, ob die Etablierung eines Systems mit höherer Reliabilität erforderlich ist.

Eine etablierte Methode, Reliabilität in einer Studie abzubilden, ist die Berechnung des κ -Wertes (kappa-Wert). Ein κ -Wert von 0 steht für eine ausschließlich auf Zufall basierende Übereinstimmung zweier Stichproben. Stimmen alle Werte überein, ist der κ -Wert 1.

Zur Reliabilität der Brustdichte-Einschätzung nach der Empfehlung des ACR sind einige Studien veröffentlicht, die den κ -Wert berechnen. Für die Inter-Rater-Reliabilität werden κ -Werte von 0,43²⁰; 0,54²¹; 0,77²²; 0,63-0,81²³ angegeben. Die Intra-Rater-Reliabilität wird mit κ gleich 0,71²¹ beschrieben.

1.3 CAD-Systeme

Die Abkürzung „CAD“ steht für „computerassistierte Detektion“ oder für „computerassistierte Diagnose“. Sie beschreibt ein Verfahren, bei dem Computerprogramme digitale oder digitalisierte Mammographien analysieren, um malignitätsverdächtige Veränderungen des Brustgewebes zu erkennen. Dazu werden die Rohdaten der Mammographie an einen Server übermittelt, auf dem das CAD-Programm installiert ist. Als Rohdaten werden die unveränderten Ausgabedaten des Detektors bezeichnet. Sie entsprechen nicht den Daten, die einem menschlichen Befunder zur Betrachtung angezeigt werden, da die Aufnahmen hierfür zuerst optimiert werden.

Areale, die der Computeralgorithmus als suspekt einstuft, werden markiert. Üblicherweise wird mit unterschiedlichen Markierungen zwischen Herdbefunden und Mikrokalzifikationen unterschieden. Diese CAD-Marker sollen den Radiologen bei der Befundung unterstützen. Die Arbeitsweise eines CAD-Algorithmus wird in Kapitel 3.2.1 genauer beschrieben.

In den letzten Jahren wurden etliche Studien veröffentlicht, die den potenziellen Nutzen von CAD-Systemen für die Befundung von Mammographien untersuchen. Vor allem über den Einsatz im Screening wird intensiv diskutiert. Hier könnten CAD-Systeme eventuell die Rolle eines Zweitbefunders übernehmen. Tabelle 3 bietet eine Übersicht über die Ergebnisse einiger Studien mit großer Fallzahl, die in den letzten Jahren veröffentlicht wurden.

Autor und Fallzahl	Datum	Design	Sens. [%]		Spez. [%]		RR [%]		CDR [%]	
Gromet²⁴ 231.221	2008	SR vs.	81,4				10,2		4,1	
		DR vs. SR +CAD	88,0	90,4			11,9	10,6	4,5	4,2
Die Sensitivität ist bei SR+CAD höher als bei DR (keine Signifikanz) bei geringerer RR. Die Leistung von SR+CAD ist größer als von SR.										
Gilbert²⁵ 31.057	2008	DR vs. SR +CAD	87,7	87,2	97,4	96,9	3,4	3,9		
		SR+CAD könnte eine Alternative für das SR im Screening sein. CAD könnte die CDR im Screening erhöhen.								
Fenton²⁶ 429.345	2007	kein CAD vs. CAD	80,4	84	90,2	87,2	11,2	13,2		
		Nach Hinzunahme von CAD steigt die RR, das führt nicht zu einer Verbesserung der Sens. Die Leistung des Screenings wird reduziert.								
Gilbert²⁷ 10.267	2006	DR vs. SR +CAD	42,6	49,1			5,7	7,7		
		Die Leistung von SR+CAD ist mit der Leistung des DR vergleichbar. Eine randomisierte Studie ist notwendig und ethisch vertretbar.								
Dean²⁸ 9.520	2006	SR vs. SR +CAD					6,2	7,8		10,9
		CAD erhöht die RR, ohne die PPV zu beeinträchtigen. Die zusätzlich gefundenen Malignome sind kleiner.								
Morton²⁹ 21.349	2006	SR vs. SR +CAD					9,8	10,8	4,9	5,3
		Die Hinzunahme von CAD erhöht die CDR und verursacht einen akzeptablen Anstieg der RR.								
Gur³⁰ 115.571	2004	kein CAD vs. CAD					11,4	11,4	3,5	3,6
		Die Hinzunahme von CAD führt zu keiner signifikanten Veränderung der RR oder der CDR.								
Cupples³¹ 27.274	2004	SR vs. SR +CAD					7,7	8,3	3,7	4,1
		CAD verbessert Screening, da die CDR steigt, das Alter bei Diagnosestellung sinkt und das Tumorstadium geringer ist.								
Freer³² 12.860	2001	SR vs. SR +CAD					6,5	7,7	3,2	3,8
		CAD verbessert die CDR, ohne die RR oder die PPV übermäßig zu beeinträchtigen.								

Sens. = Sensitivität, Spez. = Spezifität, RR = recall rate, CDR = cancer detection rate, SR = single reading, DR = double reading, PPV = positive predictive value

Tabelle 3: Zusammenfassung einiger Studien zum Einsatz von CAD in der Mammographiebefundung

Die Aussagekraft der CAD-Untersuchung einer Mammographie wird, wie bei jedem anderen medizinische Test, sowohl von ihrer Sensitivität, als auch von ihrer Spezifität bestimmt. Das Verfahren ist nur dann sinnvoll im medizinischen Alltag einzusetzen, wenn für beide Parameter ausreichend gute Werte erreicht werden können. Eine zu geringe Sensitivität schwächt das Vertrauen in die Untersuchung, während eine zu geringe Spezifität dazu führt, dass richtige Marker vom Befunder übersehen werden.

Die Sensitivität der heute verfügbaren CAD-Systeme ist relativ hoch. Dabei werden Mikrokalzifikationen häufiger (98%) als Herdbefunde (84%) erkannt³³. Ein Hauptgrund dafür, dass CAD-Systeme in Deutschland bis heute nicht flächendeckend eingesetzt werden, liegt in ihrer vergleichsweise schwachen Spezifität. Je Bild werden durchschnittlich 0,5 Mikrokalk und 1 Herdbefund falsch positiv markiert¹⁰. Die Spezifität des in dieser Studie eingesetzten CAD-Algorithmus liegt nach Herstellerangaben abhängig von der Einstellung zwischen 24,5% und 48,3%³⁴.

2. Literaturbericht und Fragestellung

Um einen breiteren Einsatz von CAD-Systemen in der Mammographiebefundung zu ermöglichen, ist es notwendig, die Leistung der CAD-Algorithmen zu verbessern. Um eine Verbesserung der Spezifität zu erreichen, ohne eine Verringerung der Sensitivität in Kauf nehmen zu müssen, werden Faktoren gesucht, die die Leistung der verfügbaren Algorithmen beeinflussen. Da ein Einfluss der Brustdichte auf menschliche Befunder in der Mammographie bekannt ist^{6, 15-17}, wird ein solcher Einfluss auch auf CAD-Systeme vermutet.

2.1 Beeinträchtigung der Sensitivität

In den letzten Jahren wurden einige Studien veröffentlicht, die untersuchen, ob eine erhöhte Brustdichte die Sensitivität von CAD-Systemen beeinträchtigt. W. Ho und P. Lam untersuchen in ihrer 2003 veröffentlichten Arbeit 108 digitalisierte Mammographien, die eine maligne Veränderung zeigen, mit dem CAD-System SecondLook (SecondLook, Version 1.1, CADx Medical Systems, Beavercreek, OH, USA)³⁵. Die Untersuchungen werden in vier Dichtekategorien eingeteilt und die Sensitivität in den verschiedenen Gruppen verglichen. Dabei zeigt sich in den höheren Kategorien eine signifikant schlechtere Sensitivität, als in den niedrigeren Kategorien (64,3% vs. 93,3%).

2006 werden zwei Arbeiten veröffentlicht. L. Li et al. evaluieren die Dichte bei 100 falsch-negativ bewerteten digitalisierten Mammographien und schließen so auf den Einfluss der Brustdichte auf einen selbst entwickelten CAD-Algorithmus³⁶. Sie kommen zu dem Ergebnis, dass, vor allem in einem frühen Stadium, Läsionen in dichterem Gewebe mit größerer Wahrscheinlichkeit nicht markiert werden.

S. Obenauer et al. nutzen in Ihrer Arbeit den CAD-Algorithmus ImageChecker (ImageChecker, Version 2.3, R2 Technology, Los Altos, CA, USA)³⁷. Es werden 412 digitale Mammographien, die eine maligne Veränderung zeigen, untersucht. Zur Differenzierung der Brustdichte wird die ACR-Dichteeinteilung eingesetzt. Auch hier zeigt sich ein Abfall der Sensitivität bei größerer Brustdichte (Kategorie 1: 84,85%; Kategorie 4: 69,70%), wenngleich die Ergebnisse dieser Studie keine statistische Signifikanz aufweisen.

Einige Arbeiten geben getrennte Ergebnisse für Herdbefunde und Mikrokalzifikationen an. 2005 werden zwei solcher Arbeiten veröffentlicht. R. Brem et al. untersuchen 906

digitale Mammographien mit dem SecondLook CAD-System (SecondLook, Version 4.0, i-CAD Inc., Fairborn, OH, USA) und teilen die Aufnahmen in zwei Dichtegruppen³⁸. Für Herdbefunde zeigt sich in der dichteren Gruppe eine signifikant schlechtere Sensitivität (83% vs. 89%), während für Mikrokalzifikationen kein Unterschied nachweisbar ist.

Eine neuere Version dieses CAD-Algorithmus wird von A. Malich et al. untersucht (SecondLook, Version 5.0, i-CAD Inc., Fairborn, OH, USA), die zu vergleichbaren Ergebnissen kommen³⁹. Eine Studienpopulation mit 127 Läsionen wird in zwei Dichtekategorien unterteilt. Die Sensitivität für Herdbefunde in der höheren Kategorie ist signifikant schlechter als in der niedrigeren Kategorie (77,42% vs. 95,71%), während bei Mikrokalzifikationen keine signifikanten Unterschiede auftreten.

C. Romero Castellano et al. veröffentlichten 2011 eine Arbeit, in der die Sensitivität in vier Dichteklassen verglichen wird⁴⁰. Dabei kommt der CAD-Algorithmus ImageChecker zum Einsatz (ImageChecker, Version 5.4, R2 Technology, Los Altos, CA, USA), der insgesamt 57 maligne Läsionen analysiert. Auch hier zeigt sich eine Verminderung der Sensitivität bei größerer Dichte nur für Herdbefunde (Kategorie 1: 96,9%; Kategorie 4: 45%).

Zusammenfassend lässt sich sagen, dass viele Studien, die zu dieser Fragestellung veröffentlicht wurden, eine Beeinträchtigung der Sensitivität von CAD-Algorithmen bei höherer Brustdichte aufzeigen. Wird der Einfluss für Herdbefunde und Mikrokalzifikationen getrennt ermittelt, lässt sich der Effekt auf die Sensitivität bei Herdbefunden einschränken.

2.2 Beeinträchtigung der Spezifität

Drei der oben aufgeführten Studien untersuchen neben der Sensitivität auch die Spezifität der CAD-Untersuchung. Die Studie von W. Ho und P. Lam untersucht 156 Mammographien, die keine Läsion abbilden³⁵. Die Spezifität wird als Anzahl der falsch-positiven Marker pro Bild angegeben. Ein vermehrtes Auftreten falsch-positiver Marker bei zunehmender Brustdichte zeigt sich nicht.

R. Brem et al. untersuchen 147 Mammographien, die als Normalbefund bewertet wurden³⁸. Für Herdbefunde zeigt sich ein signifikanter Anstieg der falsch-positiven Marker bei dichtem Drüsengewebe (2,09 vs. 2,60), während sich ein solcher Anstieg bei Mikrokalzifikationen nicht nachweisen lässt.

200 Mammographien ohne maligne Veränderung werden von A. Malich et al. untersucht³⁹. Ein Einfluss der Brustdichte auf die Anzahl der falsch-positiven Kalkmarker wird nicht festgestellt. In der ersten Dichtekategorie finden sich signifikant weniger Herdmarker, als in den anderen Kategorien, während ansonsten keine signifikanten Unterschiede festgestellt werden können.

Zum Einfluss der Brustdichte auf die Spezifität von CAD-Algorithmen existiert bisher keine eindeutige Studienlage. Wenige Studien mit relativ geringer Fallzahl untersuchen diesen Zusammenhang und kommen zu heterogenen Ergebnissen. Für Mikrokalzifikationen erscheint ein Einfluss der Brustdichte auf die Spezifität bisher unwahrscheinlich.

2.3 Fragestellung

Nach Durchsicht der aktuellen Literatur erscheint eine Beeinträchtigung der Sensitivität durch höhere Brustdichte wahrscheinlich. Allerdings weisen aktuelle CAD-Systeme vor allem in ihrer Spezifität Schwächen auf. Hier sind Einflussfaktoren bisher wenig untersucht.

Die vorliegende Studie hat sich daher zum Ziel gesetzt, Gewebeparameter auf ihre Wirksamkeit auf die Spezifität eines CAD-Algorithmus hin zu untersuchen. Dazu wurde zuerst die Brustdichte ausgewählt, da ein Einfluss auf die Sensitivität wahrscheinlich ist und die Spezifität der Mammographiebefundung durch Radiologen ebenfalls von der Brustdichte beeinträchtigt werden kann¹⁵.

Die Markteinführung des R2 Quantra-Algorithmus (R2 Quantra®, Hologic, Bedford, MA, USA) hat einen weiteren Gewebeparameter verfügbar gemacht: das Volumen des Parenchymgewebes. Ein Einfluss dieses Parameters auf die Spezifität der Mammographie-Diagnostik erscheint grundsätzlich plausibel, da vor allem Parenchymgewebe und das umliegende Binde-Stütz-Gewebe mammographisch malignomähnliche Strukturen bilden können. Ob ein solcher Einfluss tatsächlich existiert, soll deshalb ebenfalls eine Fragestellung dieser Studie sein.

3. Material und Methoden

3.1 Auswahlkriterien

Die vorliegende Studie war retrospektiv angelegt. Von jeder Patientin wurde vor der Untersuchung eine Einverständniserklärung unterschrieben, die besagt, dass persönliche Daten und Bilddaten in der Forschung und Lehre der Charité Universitätsmedizin Berlin verwendet werden dürfen. Die Zustimmung der Ethikkommission der Charité zur retrospektiven Analyse der Untersuchungsdaten lag vor (Auftragsnummer EA1/227/11).

Alle in dieser Arbeit verwendeten Mammographieuntersuchungen wurden im Zeitraum zwischen dem 1.6.2002 und dem 31.8.2006 am Radiologischen Institut der Charité durchgeführt. Es wurden nur die Aufnahmen eines Mammographiegerätes verwendet (GE Senographe 2000D, General Electric Healthcare, Wauwatosa, WI, USA).

Die Mammographie durfte zumindest auf einer Seite keinerlei pathologische Veränderungen zeigen. Um dies zu gewährleisten, war ein unauffälliges Follow-up über mindestens 22 Monate Voraussetzung. Eine Untersuchung wurde als unauffällig gewertet, wenn sie weder eine malignitätsverdächtige Mikrokalzifizierung, noch einen suspekten Herdbefund oder eine suspekte Architekturstörung zeigte.

Einschlusskriterien:

- weibliche Patientin
- Mammographie in mindestens 2 Ebenen, CC und MLO
- Aufnahme mit dem GE Senographe 2000D
- unauffälliger Mammographiebefund
- unauffälliges Follow-up über mindestens 22 Monate

Ausschlusskriterien:

- Z.n. Operation oder Vakuum-Stanzbiopsie auf der ausgewählten Seite
- technische Defizite der Aufnahme

Die infrage kommenden Mammographieuntersuchungen wurden von einem Radiologen mit fünfjähriger Erfahrung in der Befundung von Mammographien überprüft, um sicherzustellen, dass die Ein- und Ausschlusskriterien korrekt erfüllt waren.

3.2 Verwendete Geräte und Systeme

3.2.1 CAD-Algorithmus³⁴

Als CAD-Algorithmus kam das System R2 ImageChecker der Firma Hologic zum Einsatz (R2 ImageChecker®, Version 9.3, Hologic, Bedford, MA, USA). Dieses ist eine Software, die suspekte Areale in Mammographien erkennen und markieren soll. Diese werden dem Radiologen vorgestellt, nachdem eine erste Befundung abgeschlossen ist. Dadurch soll die Rate der falsch-negativ befundeten Mammographien gesenkt, also die Sensitivität erhöht werden.

Das CAD-System sucht nach suspekten Mikrokalzifikationen sowie Herdbefunden, Verdichtungen oder Architekturstörungen und markiert diese unterschiedlich:

- ▲ Region mit suspekten Mikrokalzifikationen
- * Region mit suspektem Herdbefund oder suspekter Architekturstörung
- + Region mit suspektem Befund aus beiden o.g. Kategorien

Detektion von Mikrokalzifikationen

Zum Auffinden von Mikrokalzifikationen werden die Rohdaten der Mammographie zunächst von zwei Filtern analysiert, die aus künstlichen neuronalen Netzen bestehen. Der erste Filter identifiziert Kalzifikationen, die dann im nächsten Schritt weiter analysiert werden. Dabei werden etliche Eigenschaften überprüft, wie zum Beispiel Kontrast, Form und Größe. Außerdem wird die Brust in Segmente unterteilt und die Lokalisation der suspekten Areale bestimmt.

Besteht eine Studie aus mehr als einer Aufnahme, was bei allen in dieser Studie verwendeten Mammographien der Fall war, werden nach der Analyse der einzelnen Aufnahmen auch die Ergebnisse der Bilder untereinander verglichen. So sucht der Algorithmus nach ähnlichen Befunden in mehreren Aufnahmen derselben Brust (z.B. RMLO und RCC), beziehungsweise nach Asymmetrien zu der jeweils kontralateralen Aufnahme (z.B. RMLO und LMLO).

Detektion von Herdbefunden, Verdichtungen und Architekturstörungen

Das Programm sucht nach Herdbefunden, Verdichtungen und Architekturstörungen, indem es unterschiedliche Charakteristika suspekter Areale ermittelt. Dabei werden vor allem deren Dichte und Form beurteilt. Außerdem wird nach eventuell vorhandenen

Spiculae gesucht. Je dichter die Läsion ist und je mehr Spiculae vorhanden sind, desto eher wird ein Areal als malignitätsverdächtig eingestuft.

Wie bei den Kalzifikationen, so findet auch für Herdbefunde ein Abgleich mit anderen Bildern derselben Studie statt. Es wird die Position des Befundes in der Brust ermittelt und nach Ähnlichkeiten in anderen Aufnahmen derselben Seite (z.B. RMLO und RCC), beziehungsweise Asymmetrien zu der kontralateralen Seite (z.B. RMLO und LMLO) gesucht.

Erstellen der Marker

Die Ergebnisse der unterschiedlichen Analysen werden kombiniert und verarbeitet. Abschließend werden die suspekten Befunde mit einer Trainingsdatenbank abgeglichen und bezüglich ihrer Wahrscheinlichkeit auf Malignität in ein Ranking eingeteilt. Dieses Ranking ermöglicht es dem Programm, verschiedene Werte zu definieren, ab der eine Läsion für den Befunder als tumorverdächtig markiert wird.

Die Werte lassen sich vom Anwender in Form so genannter „Operating Points“ festlegen und können für Kalzifikationen und Herdbefunde jeweils getrennt ausgewählt werden.

Operating Point 0	möglichst wenige Markierungen	möglichst hohe Spezifität, dafür eingeschränkte Sensitivität
Operating Point 1	mittlere Einstellung	ausgeglichene Einstellung für Sensitivität und Spezifität
Operating Point 2	möglichst alle Markierungen	möglichst hohe Sensitivität, dafür eingeschränkte Spezifität

Tabelle 4: Übersicht der Einstellungsmöglichkeiten des CAD-Algorithmus

Um den Befunder nicht mit zu vielen Markierungen zu verwirren, ist deren Anzahl begrenzt. Pro Bild werden höchstens 4 Marker für Kalzifikationen, 2 Marker für Herdbefunde und 2 Marker für Mischbefunde eingetragen. Besteht eine Studie aus mehreren Bildern, werden insgesamt höchstens 8 Kalzifikationen, 4 Herdbefunde und 4 Mischbefunde markiert. Würden diese Grenzen nach der Analyse und unter Berücksichtigung des jeweiligen Operating Point überschritten, werden diejenigen

Areale nicht markiert, die der Algorithmus als am wenigsten malignitätsverdächtig eingestuft hat.

Ausgabe der Ergebnisse

Das Programm unterstützt unterschiedliche Wege der Datenausgabe. In dieser Arbeit wurde die Ausgabe als „Mammography CAD SC“ gewählt. SC steht für „secondary capture“ und ist ein DICOM-Format (digital imaging and communications in medicine). Die vier Standardaufnahmen der Studie (RMLO, RCC, LMLO, LCC) werden, soweit vorhanden, komprimiert und in einem neu erstellten Bild nach einem festgelegten Standard angeordnet. In dieses Bild werden nun die CAD-Markierungen eingetragen. Außerdem werden die Operating Points für Kalzifikationen und Herdbefunde vermerkt. Alle Ergebnisse der CAD-Analyse lassen sich so ablesen und können leicht in einem bestehenden PACS-System verwaltet werden.

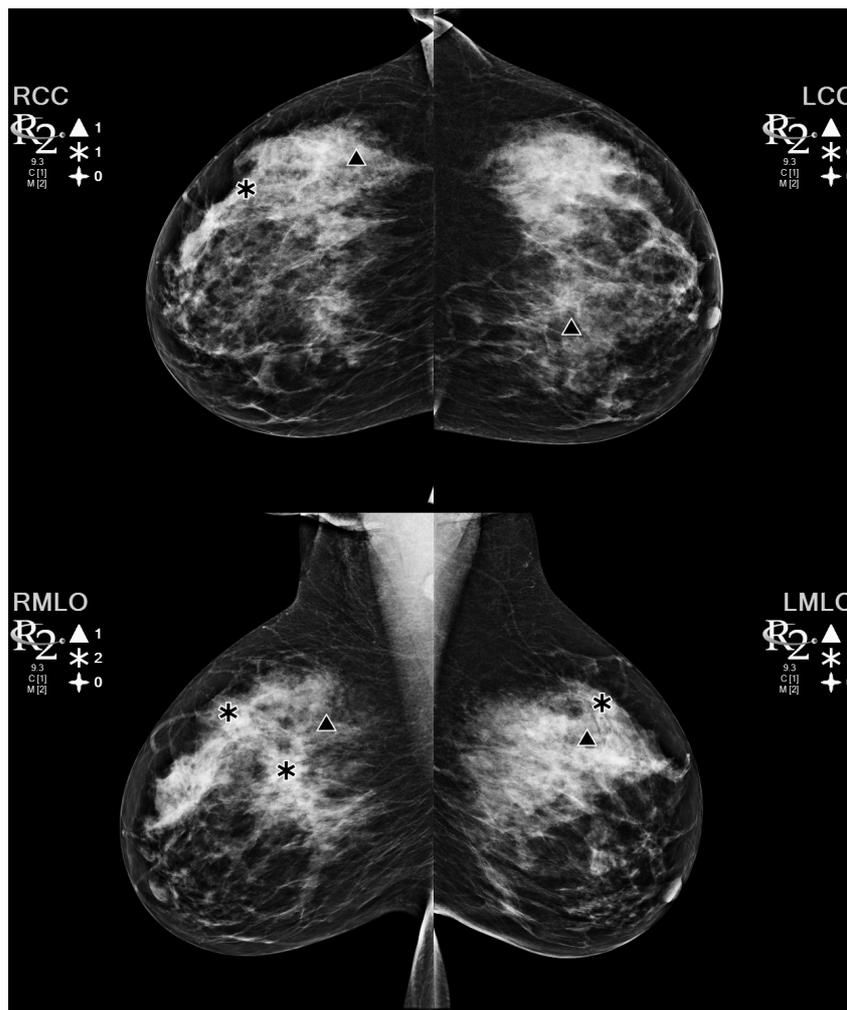


Abbildung 2: Ausgabe des CAD-Algorithmus mit Markern für Herdbefunde und Mikrokalzifikationen

3.2.2 Algorithmus zur Analyse der Gewebeeigenschaften⁴¹

In dieser Studie sollte der Einfluss der Faktoren Parenchymvolumen und Brustdichte auf die Anzahl der falsch-positiven Marker eines CAD-Algorithmus untersucht werden. Um diese Kenngrößen möglichst genau und reproduzierbar zu ermitteln, wurde der Bildanalyse-Algorithmus R2 Quantra der Firma Hologic eingesetzt (R2 Quantra®, Version 1.3, Hologic, Bedford, MA, USA).

R2 Quantra ist ein Computerprogramm, das für den Einsatz mit digitalen Mammographiegeräten entwickelt wurde. Es analysiert die Bildinformationen der Rohdaten der Mammographie und ermittelt die Brustgröße und das Volumen des Brustdrüsengewebes. Durch Division wird ein prozentualer Wert für die Brustdichte errechnet.

Das Parenchymvolumen wird ermittelt, indem für jeden Pixel die Intensität der eintreffenden Röntgenstrahlung ausgelesen wird. Auf der Grundlage bekannter Werte für den Dämpfungskoeffizienten des Gewebes⁴² und des Röntgenspektrums der Anode⁴³, sowie der verwendeten Spannung und Stromstärke und der Kompressionsdicke lässt sich berechnen, wie viel Parenchym die Strahlung durchdrungen hat. Dieser Wert entspricht der Höhe eines Parenchymzylinders über dem entsprechenden Pixel. Durch Addition der Werte aller Pixel ergibt sich das Gesamtparenchymvolumen.

Mit einer ähnlichen Methode wird das Gesamtvolumen der Brust berechnet, wobei der Algorithmus eventuelle Fehlmessungen durch nicht vollständig komprimierte Bereiche der Brust kompensiert. Die Brustdichte ergibt sich als Quotient aus Parenchymvolumen und Brustvolumen und wird als prozentualer Wert angegeben.

2/22/2010		
Factor	R	L
Volume of fibroglandular tissue (cm3)	146	154
Volume of breast (cm3)	633	593
Percentage of fibroglandular tissue (%)	23	26

HOLOGIC | **R₂** Earlier. Smarter. Better.™

Abbildung 3: Tabelle mit den Ergebnissen des Quantra-Algorithmus

Enthält eine Studie mehr als eine Aufnahme derselben Brust (z.B. LMLO und LCC), was bei allen in dieser Arbeit untersuchten Studien der Fall war, so werden die Ergebnisse für beide Aufnahmen verglichen und anschließend zu einem Ergebnis für jede Brust zusammengefasst.

Der Algorithmus verfügt über eine eigene Qualitätskontrolle, deren Bewertung mit den Ergebnissen angezeigt wird. Ist die Qualität eingeschränkt, wird den Werten ein „ ~ “ vorangestellt. Stuft die Kontrolle die Ergebnisse als unzureichend ein, werden die Daten nicht angezeigt („ -- “).

3.2.3 Mammographiegerät

Alle Mammographien wurden mit demselben digitalen Mammographiegerät, einem GE Senograph 2000D (GE Healthcare, Wauwatosa, WI, USA) aufgenommen. Dieses Mammographiegerät wird seit 1999 an der Charité eingesetzt⁴⁴.

Die Strahlung wird aus der Röntgenröhre von einer Doppelfokus-Bimetallanode aus Molybdän (Mo) und Rhodium (Rh) emittiert, der Filter besteht aus denselben Materialien. Abhängig von der Brustdicke wird als Anode/Filter-Kombination entweder Mo/Mo, Mo/Rh oder Rh/Rh gewählt. Die mittlere Energie der Photonen nimmt in dieser Reihenfolge zu, was bei einer dickeren Brust eine bessere Penetration des Gewebes ermöglicht⁴⁵.

Im Detektor wird die Röntgenstrahlung zunächst über einen Szintillator aus Cäsiumjodid (CsI) in Licht umgewandelt. Dieses wird von Photodioden in einer nachgelagerten Ebene gemessen und so ein analoges elektrisches Signal erzeugt. Der elektrische Strom entspricht dabei der Menge der einfallenden Strahlen, die wiederum abhängig von der Röntgendichte des darüber liegenden Gewebes ist. Diese Signalunterschiede werden mit 14 bit Tiefe gespeichert. Das entspricht mehr als 16000 Graustufen. Daher können die Mammographien in verschiedenen Fensterungen befundet werden.

Die Ortsauflösung ist bei diesem Gerät in der Regel durch die Dichte der Photodioden begrenzt. Jede Diode liefert ein Pixel der Mammographie. Das aktive Detektorfeld des Senograph 2000D ist 19 x 23 cm groß und besitzt eine Auflösung von 1900 x 2304 Bildpunkten. Dies entspricht einer Pixelgröße von 100 x 100 µm.

3.2.4 Weitere Software

Die Verwaltung der Patientendaten sowie der Ergebnisse der Computeranalyse erfolgte in einer Excel-Datei (Microsoft Excel®, Microsoft Corporation, Redmond, WA, USA). In diesem Dokument wurden auch einfache Berechnungen und statistische Analysen durchgeführt. Für weitergehende Statistiken kam das Programm SPSS Statistics (IBM SPSS Statistics®, Version 19.0.0, Armonk, NY, USA) zum Einsatz.

Diagramme wurden zum Teil direkt mit SPSS Statistics erstellt, für weitere Diagramme wurde Numbers der Firma Apple verwendet (Numbers '09®, Version 2.1, Apple Inc., Cupertino, CA, USA).

3.3 Studiendurchführung

Die Grundlage der vorliegenden Arbeit bildeten die Mammographien der eingeschlossenen Patientinnen. Zunächst wurden die Aufnahmen aus dem digitalen Archiv der Charité geladen und auf einem Rechner des Campus Mitte gespeichert. Hier erfolgte die Verwaltung und Bearbeitung mit Hilfe des Programms RA600 der Firma General Electric Healthcare (Centricity Radiology RA 600, General Electric Healthcare, Wauwatosa, WI, USA).

Die Patientendaten wurden in einer Excel-Tabelle verwaltet. Von jeder Patientin wurde zunächst die Patientennummer, der Name, der Vorname und das Geburtsdatum, sowie das Datum der Mammographieuntersuchung, die in diese Arbeit eingeschlossen wurde, eingetragen. Dazu kam das Datum der letzten Untersuchung, die keine pathologische Veränderung zeigte, um dadurch im weiteren Verlauf den genauen Zeitraum des unauffälligen follow-ups berechnen zu können.

Um Verknüpfungen der untersuchten Daten untereinander zu vermeiden, wurden nur die Mammographien einer Brust jeder Patientin in die Studie einbezogen. Entsprachen beide Seiten den Einschlusskriterien der Studie, wurde mit Hilfe des Programms Excel eine Seite zufällig ausgewählt.

Die Rohdaten der Mammographien wurden nun zur Analyse an den CAD-Server gesendet, wo suspekte Areale mit Markern versehen wurden. Dabei unterschied das Programm zwischen potentiellen Mikrokalzifikationen und suspekten Herdbefunden, Verdichtungen oder Architekturstörungen. Um für alle drei Einstellungen des R2 CAD-Algorithmus Ergebnisse zu erhalten, wurden die Rohdaten insgesamt drei Mal, jeweils mit einer anderen Einstellung, an den Server gesendet.

Zur besseren Übersicht und Prozessierung der Ergebnisse wurden die Ausgabedaten nun in die Excel-Tabelle übertragen. Zunächst wurde für jede Patientin die Ausgabe des R2 Quantra-Algorithmus, also das Volumen der Brust, das Volumen des Drüsenparenchyms und die daraus resultierende prozentuale Dichte, übertragen. Anschließend wurden die Marker protokolliert. Für jede Patientin wurde dabei gezählt, wie viele Marker das CAD-System bei den einzelnen Einstellungen geliefert hatte. Dabei wurde unterschieden, ob es sich um Marker für Mikrokalk oder für Herdbefunde handelte und in welcher Ebene diese Marker verzeichnet waren (CC oder MLO). Keine der Mammographien, die in diese Studie einbezogen wurden, zeigte eine pathologische Veränderung. So konnten alle Marker als falsch-positive Ergebnisse des CAD-Systems interpretiert werden. Mit Hilfe verschiedener statistischer Methoden wurde nun untersucht, ob das Volumen des Drüsenparenchyms oder die prozentuale Dichte der Brust einen Einfluss auf die Anzahl der falsch-positiven Ergebnisse des CAD-Systems haben.

3.4 Statistische Methoden

Zur Analyse der Ergebnisse wurden verschiedene statistische Untersuchungen durchgeführt. Einfache Kenngrößen wie Mittelwert, Minimum und Maximum einer Stichprobe wurden direkt in Excel berechnet. Für aufwändigere Analysen wurde die Statistiksoftware SPSS Statistics verwendet (SPSS Statistics®, Version 19.0.0, IBM Corporation, Armonk, NY, USA).

Die Spezifität ist der Quotient der Anzahl der richtig-negativen Untersuchungen durch die Gesamtzahl der Untersuchungen. Eine Untersuchung ist richtig-negativ, wenn der CAD-Algorithmus keinen Marker gesetzt hat. Alle anderen Untersuchungen sind falsch-positiv. Die Anzahl der Marker auf einer solchen Untersuchung geht bei der Angabe der Ergebnisse als Spezifität verloren. Aus diesem Grund wird in den meisten Berechnungen statt der Spezifität die Anzahl der falsch-positiven Marker angegeben.

Um den Einfluss der Brustdichte und des Parenchymvolumens auf die Anzahl der falsch-positiven Marker zu ermitteln, wurden die Untersuchungen nach diesen drei Parametern kategorisiert. Anschließend wurden die Kategorien auf statistisch signifikante Unterschiede untersucht. Dazu wurde der Mann-Whitney-U-Test, ein Test für unabhängige Stichproben mit nicht normalverteilten Werten, eingesetzt. Das Signifikanzniveau wurde auf $\alpha=0,05$ festgelegt.

Um einen Zusammenhang zwischen diesen Parametern näher zu beschreiben, wurde eine Korrelationsanalyse durchgeführt. Eine Korrelationsanalyse untersucht, ob ein Zusammenhang zwischen zwei metrischen Variablen besteht, sowie dessen Größe und Richtung.

Um zu ermitteln, wie groß der Einfluss der Brustdichte und des Parenchymvolumens auf die Anzahl der falsch-positiven Marker ist, wurde eine Regressionsanalyse nach Poisson durchgeführt.

4. Ergebnisse

4.1 Studienkollektiv

Es konnten 231 Patientinnen gefunden werden, die den Einschlusskriterien entsprachen. In 112 Fällen entsprachen die Mammographien beider Seiten den Anforderungen der Studie. Hier wurde eine Seite zufällig ausgewählt.

In 9 Fällen ließ die Selbstkontrolle des R2 Quantra - Algorithmus eine Analyse der Gewebeeigenschaften der Brust nicht zu. Diese Mammographien wurden von der Studie ausgeschlossen. Insgesamt wurden also 222 Mammographien in die Studie eingeschlossen.

Das Alter der Patientinnen betrug bei der ersten Untersuchung im Mittel 54,3 Jahre, bei einer Standardabweichung von $\pm 9,8$ Jahren. Die jüngste Patientin war 30 Jahre, die älteste 79 Jahre alt.

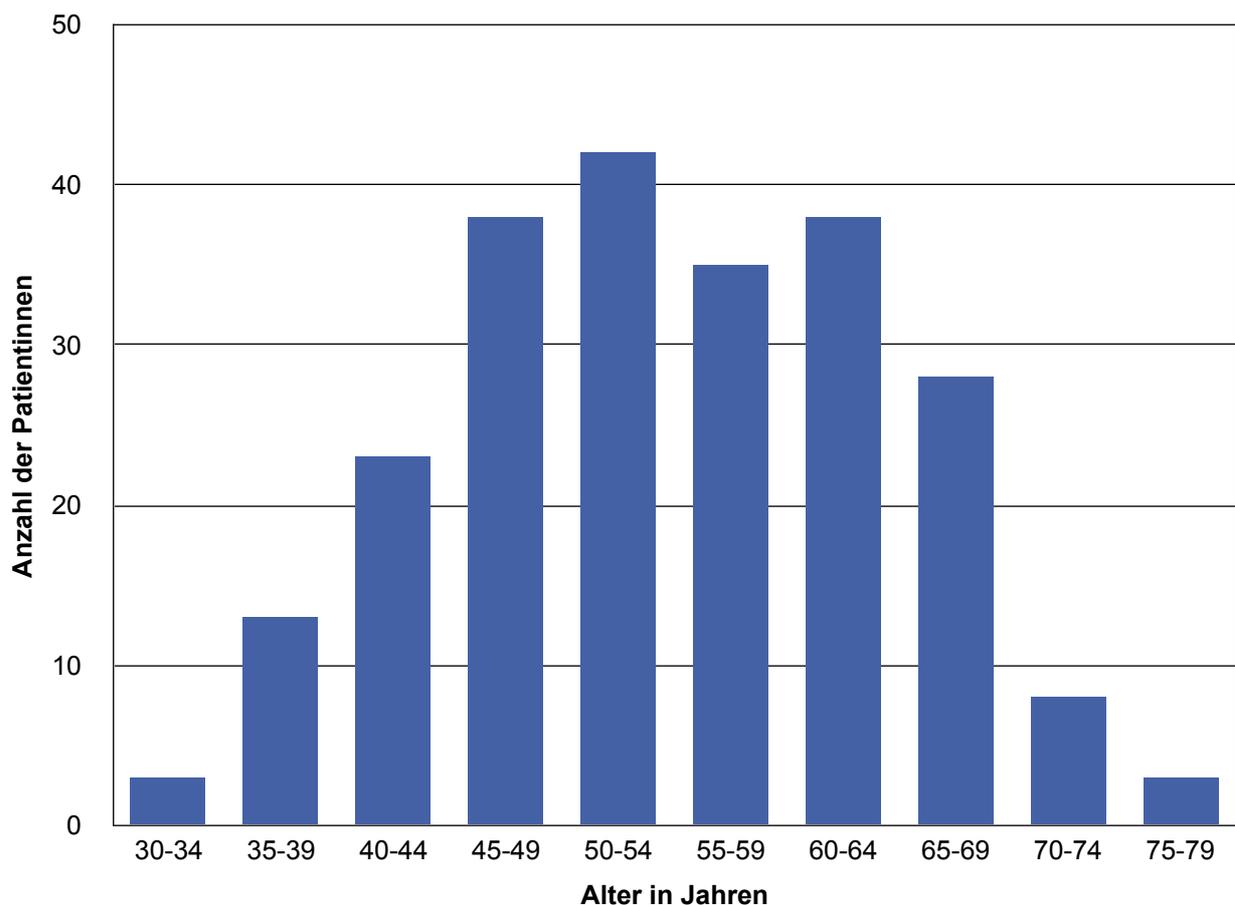


Abbildung 4: Alter der Patientinnen bei der ersten Untersuchung

4.2 Ergebnisse der Computeralgorithmen

4.2.1 Ergebnisse des Algorithmus zur Analyse der Gewebeeigenschaften

Im Folgenden werden die Ergebnisse des Algorithmus R2 Quantra zusammenfassend wiedergegeben. Zur besseren Übersicht wurden die Ergebnisse in Gruppen zusammengefasst und in Diagrammen abgebildet.

Das Volumen der untersuchten Brust betrug durchschnittlich 454,9 ml bei einer Standardabweichung von ± 242 ml. Die kleinste untersuchte Brust war 63 ml groß, die größte wies ein Volumen von 1564 ml auf.

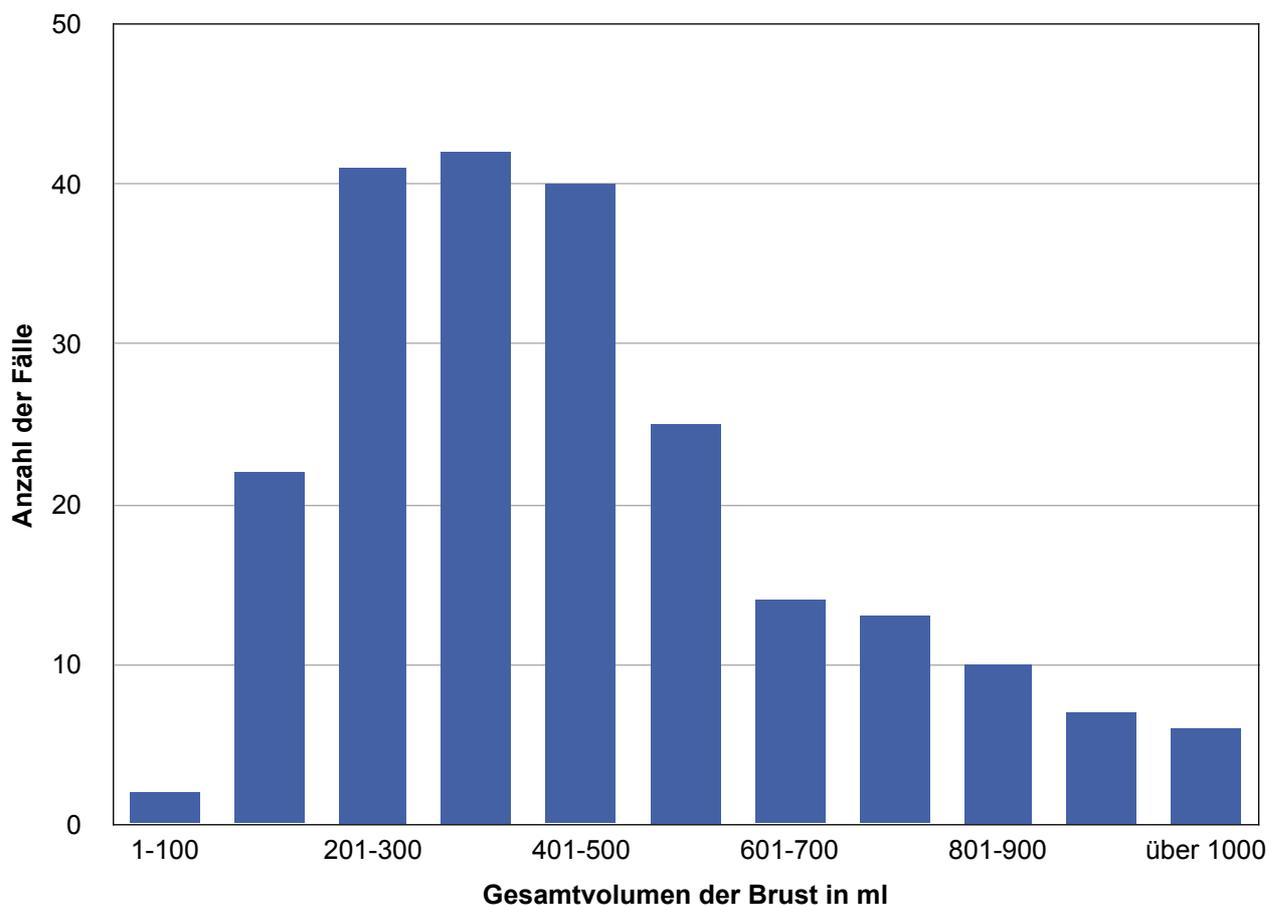


Abbildung 5: Brustvolumen (einseitig) in Milliliter

Das Volumen des Drüsengewebes betrug im Mittel 98,5 ml bei einer Standardabweichung von ± 58 ml. Die geringste Menge Parenchym wies ein Fall mit 22 ml Parenchymvolumen auf, die größte Menge wurde mit 379 ml bestimmt.

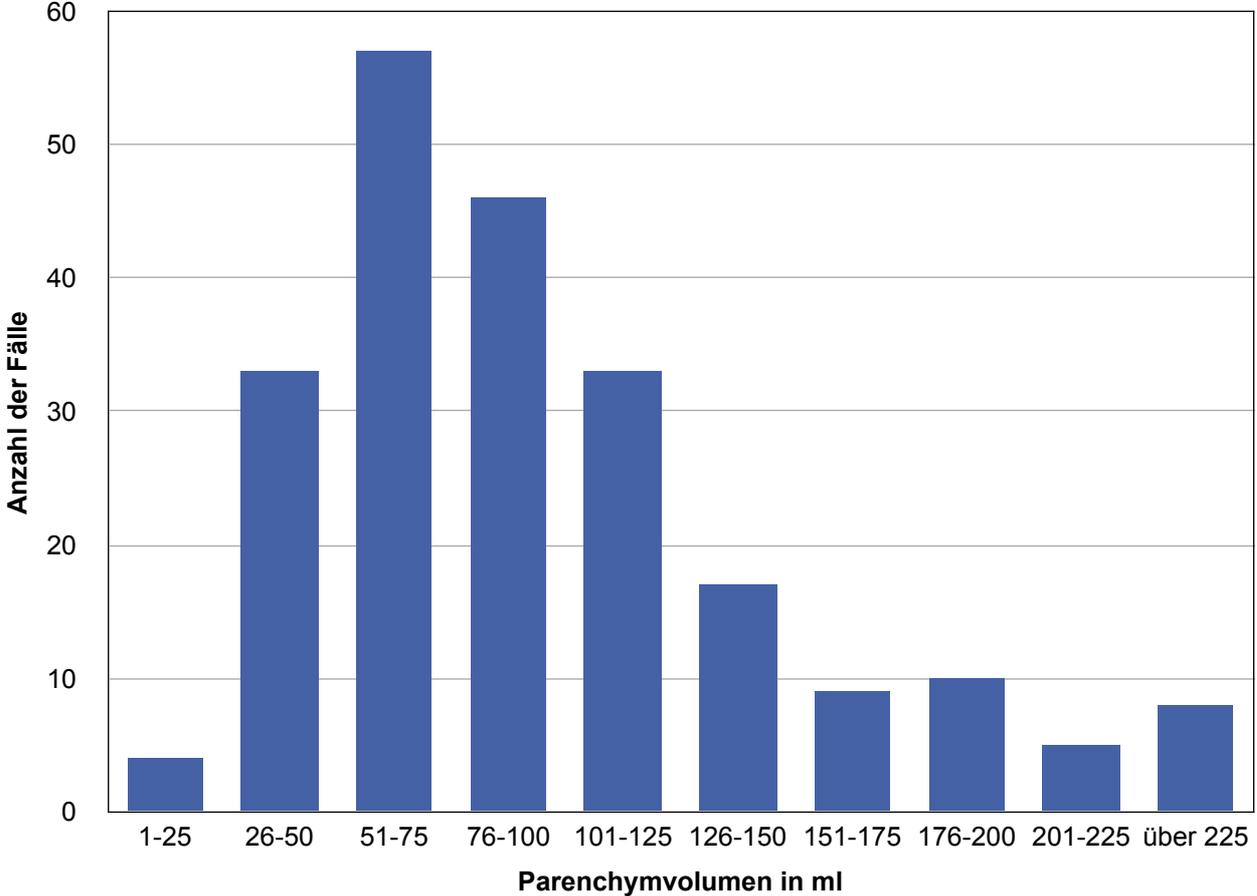


Abbildung 6: Drüsenparenchymvolumen je Brust in Milliliter

Aus der Division des Parenchymvolumens durch das Brustvolumen errechnet sich die Brustdichte in Prozent. Sie betrug im Mittel 23,7% bei einer Standardabweichung von $\pm 11\%$. Die Brust mit dem geringsten Anteil des Parenchymvolumens am Gesamtvolumen wies eine Dichte von 8% auf, für die dichteste Brust betrug der Wert 61%.

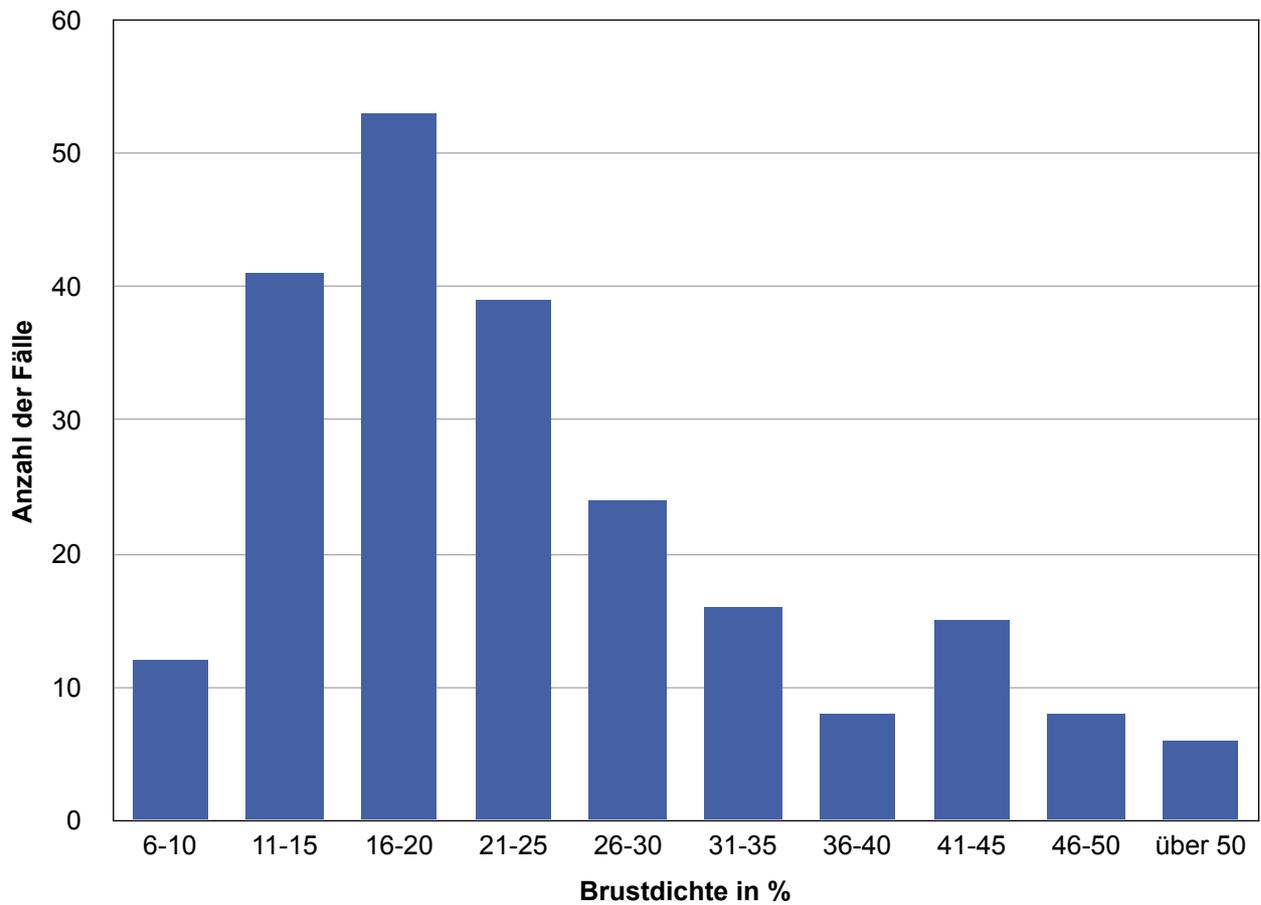


Abbildung 7: Brustdichte in Prozent

4.2.2 Ergebnisse des CAD-Algorithmus

Die 222 untersuchten Mammographie-Studien enthielten jeweils eine cranio-caudale Aufnahme (CC) und eine mediolateral-oblique Aufnahme (MLO) für die ausgewählte Seite. Der CAD-Algorithmus untersuchte also insgesamt 444 Bilder. Sowohl für Mikrokalzifikationen als auch für Herdbefunde und Architekturstörungen wurden alle drei Einstellungen des Operating Point (OP) getestet. Die Anzahl der Marker, die der CAD-Algorithmus ausgegeben hat, sind in den Tabellen 5 und 6 zusammengefasst.

		Summe	MW	s
Operating Point 0	gesamt	93	0,42	0,70
	CC	41	0,18	0,43
	MLO	52	0,23	0,51
Operating Point 1	gesamt	116	0,52	0,82
	CC	53	0,24	0,51
	MLO	63	0,28	0,56
Operating Point 2	gesamt	154	0,69	0,93
	CC	72	0,32	0,60
	MLO	82	0,37	0,65

Tabelle 5: Zusammenfassung der Marker zu Herdbefunden bei den drei Einstellungen des Operating Point mit Mittelwert (MW) und Standardabweichung (s)

		Summe	MW	s
Operating Point 0	gesamt	53	0,24	0,66
	CC	27	0,12	0,37
	MLO	26	0,12	0,40
Operating Point 1	gesamt	95	0,43	0,90
	CC	48	0,22	0,50
	MLO	47	0,21	0,52
Operating Point 2	gesamt	125	0,56	1,11
	CC	61	0,27	0,62
	MLO	64	0,29	0,62

Tabelle 6: Zusammenfassung der Marker zu Mikrocalc bei den drei Einstellungen des Operating Point mit Mittelwert (MW) und Standardabweichung (s)

Das Diagramm 8 bildet die Gesamtanzahl der Marker für alle drei Einstellungen des Operating Point ab.

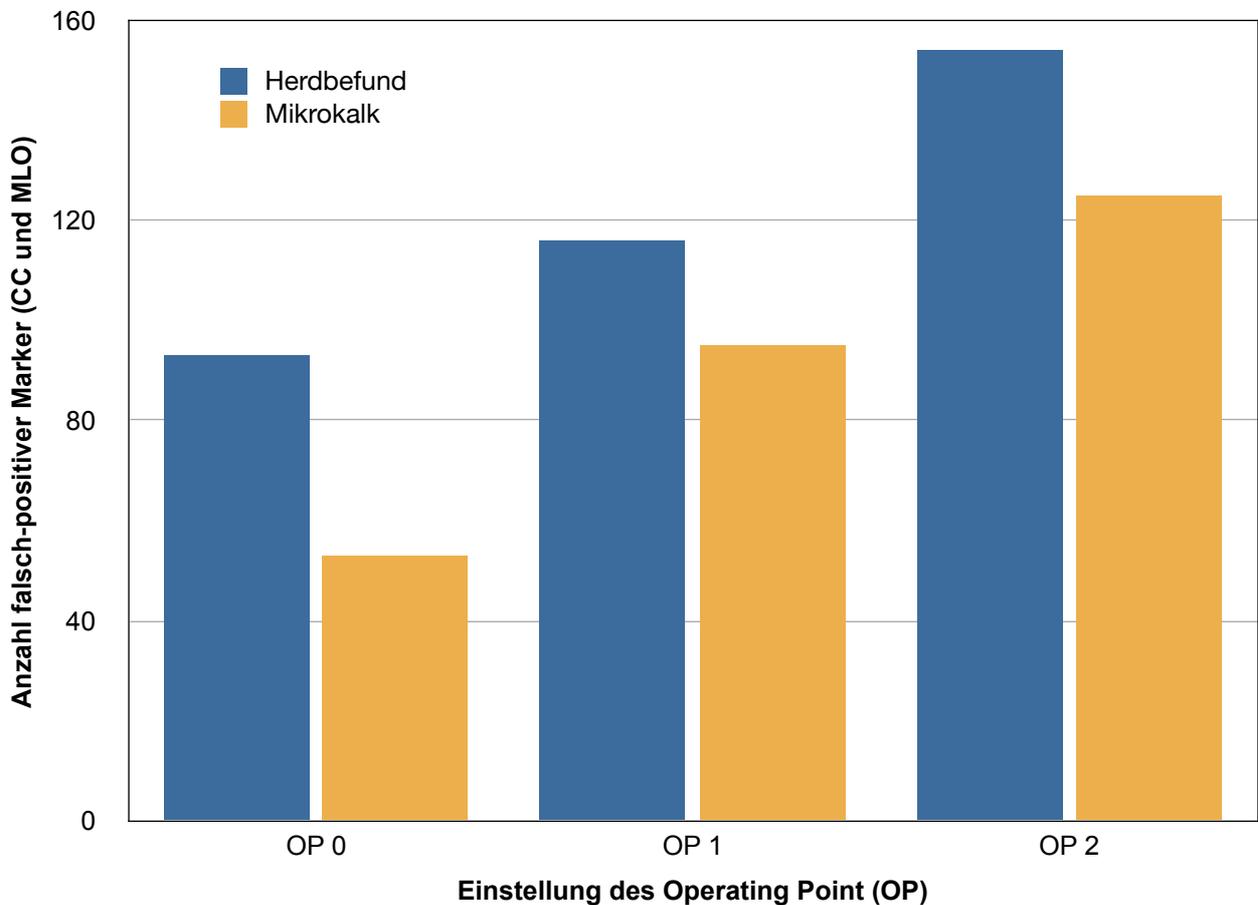


Abbildung 8: Summe der falsch-positiven Marker in CC und MLO je Einstellung des Operating Point (OP)

Der CAD-Algorithmus begrenzt die Marker pro Bild und Studie (s. 3.2.1, *Erstellen der Marker*). Die Begrenzungen pro Studie lassen sich in dieser Arbeit nicht nachvollziehen, da nur die Aufnahmen einer Seite ausgewertet wurden. Die Grenzwerte pro Bild wurden bei den Markern für Mikrokalk nicht erreicht. Die Anzahl der Untersuchungen, bei denen die Grenzwerte für Herdbefund-Marker pro Bild erreicht wurden, sind in Tabelle 7 zusammengefasst.

OP 0	OP 1	OP 2
13	20	36

Tabelle 7: Anzahl der Untersuchungen, bei denen die Grenzwerte der Marker zu Herdbefunden pro Bild erreicht wurden; je Einstellung des Operating Point (OP)

4.3 Statistische Analyse der Ergebnisse

4.3.1 Spezifität des CAD-Algorithmus

Einfache und zugleich aussagekräftige Maßzahlen für die Güte einer Untersuchung sind Sensitivität und Spezifität. Da in diese Studie nur Mammographien ohne pathologischen Befund eingeschlossen waren, lassen sich nur Aussagen über die Spezifität treffen.

Die Spezifität ist der Anteil an negativen Fällen, die der Test richtig als „negativ“ ermittelt hat. In dieser Studie ist die Spezifität also die Anzahl der Studien ohne Marker geteilt durch die Anzahl der untersuchten Studien (222). Sie lässt sich entweder für Herdbefunde und Mikrokalzifikationen getrennt, oder für beides gemeinsam berechnen. In Tabelle 8 sind die Werte für die Spezifität bei den drei Operating Points zusammengestellt.

	OP 0	OP 1	OP2
Herdbefund	68,9 %	64,4 %	56,3 %
Mikrokalk	84,7 %	75,7 %	70,7 %
gemeinsam	61,3 %	51,8 %	43,7 %

Tabelle 8: Spezifität des CAD-Algorithmus bei den drei Einstellungen des Operating Point (OP)

4.3.2 Kategorisierung nach Dichte und Parenchymvolumen

Um einen Überblick über die erfassten Daten zu gewinnen und gleichzeitig einen ersten Eindruck von den Einflüssen der Brustdichte und der Menge des Parenchymvolumens auf die falsch-positiv-Rate des CAD-Algorithmus zu erhalten, wurden die untersuchten Studien (bestehend aus je einer CC- und einer MLO-Aufnahme) in Kategorien unterteilt. Die Studien wurden nach Dichte bzw. Parenchymvolumen sortiert und in drei möglichst gleich große Gruppen unterteilt: „gering“, „mittel“ und „hoch“. Anschließend konnte für jede Kategorie der Mittelwert der falsch-positiven Marker pro Studie berechnet werden. Die Ergebnisse dieser Kategorisierung sind in den Tabellen 9 und 10 zusammengefasst.

Dichte-Kategorie	gering			mittel			hoch		
Grenzen [%]	8 - 17			17 - 26			26 - 61		
Anzahl Fälle	75			73			74		
Operating Point	0	1	2	0	1	2	0	1	2
Herd-Marker	27	34	49	34	39	55	32	42	53
Herd-Marker MW	0,36	0,45	0,65	0,47	0,53	0,75	0,43	0,57	0,72
Kalk-Marker	19	28	39	14	31	49	20	30	35
Kalk-Marker MW	0,25	0,37	0,52	0,19	0,42	0,67	0,27	0,41	0,47

Tabelle 9: Kategorisierung der untersuchten Studien (CC und MLO) nach der Dichte der untersuchten Brust. Bei den drei Einstellungen des Operating Point wurde in jeder Kategorie die Anzahl der falsch-positiven Marker gezählt und der Mittelwert (MW) pro Studie berechnet.

Volumen-Kategorie	gering			mittel			hoch		
Grenzen [ml]	22 - 70			71 - 105			106 - 379		
Anzahl Fälle	77			71			74		
Operating Point	0	1	2	0	1	2	0	1	2
Herd-Marker	27	31	45	29	32	46	37	52	66
Herd-Marker MW	0,35	0,40	0,58	0,41	0,45	0,65	0,50	0,70	0,89
Kalk-Marker	19	30	41	18	32	38	16	27	44
Kalk-Marker MW	0,25	0,39	0,53	0,25	0,45	0,54	0,22	0,36	0,59

Tabelle 10: Kategorisierung der untersuchten Studien (CC und MLO) nach dem Parenchymvolumen der untersuchten Brust. Bei den drei Einstellungen des Operating Point wurde in jeder Kategorie die Anzahl der falsch-positiven Marker gezählt und der Mittelwert (MW) pro Studie berechnet.

Sollte die Brustdichte oder die Menge des Parenchymvolumens einen Einfluss auf die Anzahl der falsch-positiven Marker haben, wäre zu erwarten, dass der Mittelwert der Marker pro Studie in den größeren Kategorien steigt.

Ein solcher Anstieg lässt sich bei der Kategorisierung nach der Brustdichte weder für die Anzahl der Herd-Marker (OP 0: 0,36/0,47/0,43; OP 1: 0,45/0,53/0,57; OP 2: 0,65/0,75/0,72), noch für die Anzahl der Kalk-Marker (OP 0: 0,25/0,19/0,27; OP 1: 0,37/0,42/0,41; OP 2: 0,52/0,67/0,47) durchgehend feststellen.

Bei der Kategorisierung nach Parenchymvolumen besteht augenscheinlich eine Tendenz, dass der CAD-Algorithmus bei Aufnahmen aus der Kategorie „mittel“ mehr falsch-positive Herd-Marker ausgibt als bei Aufnahmen aus der Kategorie „gering“. In der Kategorie „hoch“ ist der Mittelwert noch größer. Dieser Effekt zeigt sich bei allen Einstellungen des Operating Point (OP 0: 0,35/0,41/0,50; OP 1: 0,40/0,45/0,70; OP 2: 0,58/0,65/0,89).

Bei den Mittelwerten für die Anzahl der Kalk-Marker zeigt sich diese Tendenz allerdings auch bei der Kategorisierung nach dem Parenchymvolumen nicht durchgehend (OP 0: 0,25/0,25/0,22; OP 1: 0,39/0,45/0,36; OP 2: 0,53/0,54/0,59).

Vergleicht man die Gruppen mit Hilfe des Mann-Whitney-U-Test, zeigt sich ein statistisch signifikanter Anstieg des Mittelwertes der falsch-positiven Treffer tatsächlich nur bei der Kategorisierung nach Parenchymvolumen. Vergleicht man den Mittelwert der untersten („gering“) mit dem der obersten („hoch“) Kategorie, so findet sich bei den Einstellungen OP 1 ($p=0,042$) und OP 2 ($p=0,021$) ein signifikanter Anstieg. Bei der Einstellung OP 0 ist der Anstieg hingegen nicht signifikant ($p=0,164$).

	OP	gering / mittel	mittel / hoch	gering / hoch
Herd	0	0,192	0,222	0,469
	1	0,396	0,445	0,337
	2	0,426	0,439	0,467
Kalk	0	0,234	0,126	0,361
	1	0,219	0,438	0,157
	2	0,207	0,207	0,488

Tabelle 11: Mann-Whitney-U-Test, Signifikanz des Anstiegs der Trefferwahrscheinlichkeit in höheren Kategorien bei Kategorisierung nach Dichte

	OP	gering / mittel	mittel / hoch	gering / hoch
Herd	0	0,275	0,365	0,164
	1	0,293	0,125	0,042 *
	2	0,281	0,076	0,021 *
Kalk	0	0,358	0,337	0,205
	1	0,399	0,168	0,243
	2	0,471	0,358	0,336

Tabelle 12: Mann-Whitney-U-Test, Signifikanz des Anstiegs der Trefferwahrscheinlichkeit in höheren Kategorien bei Kategorisierung nach Parenchymvolumen

Zur besseren Übersicht sind die Mittelwerte der falsch-positiven Marker pro Studie im Folgenden in Diagrammen abgebildet.

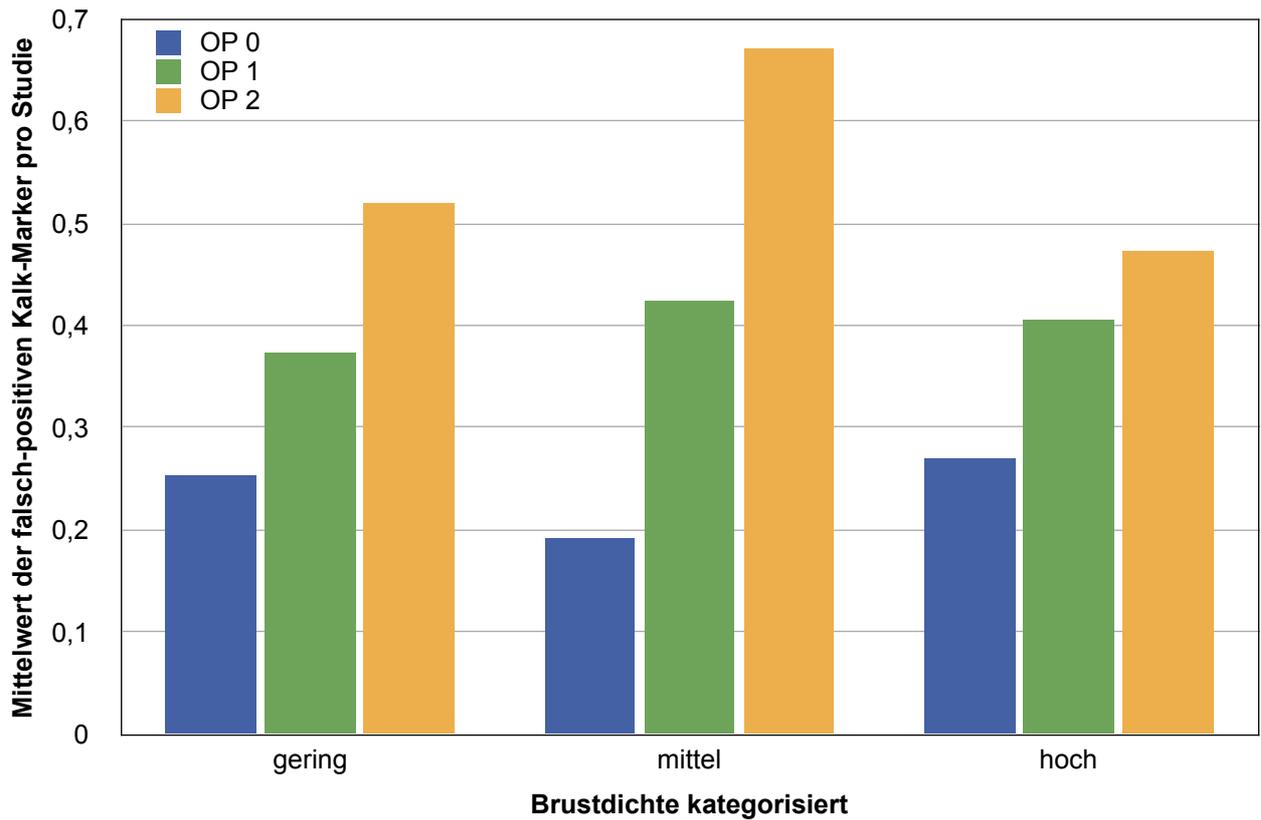


Abbildung 9: Mittelwerte der falsch-positiven Kalk-Marker pro Studie (CC und MLO) je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Brustdichte.

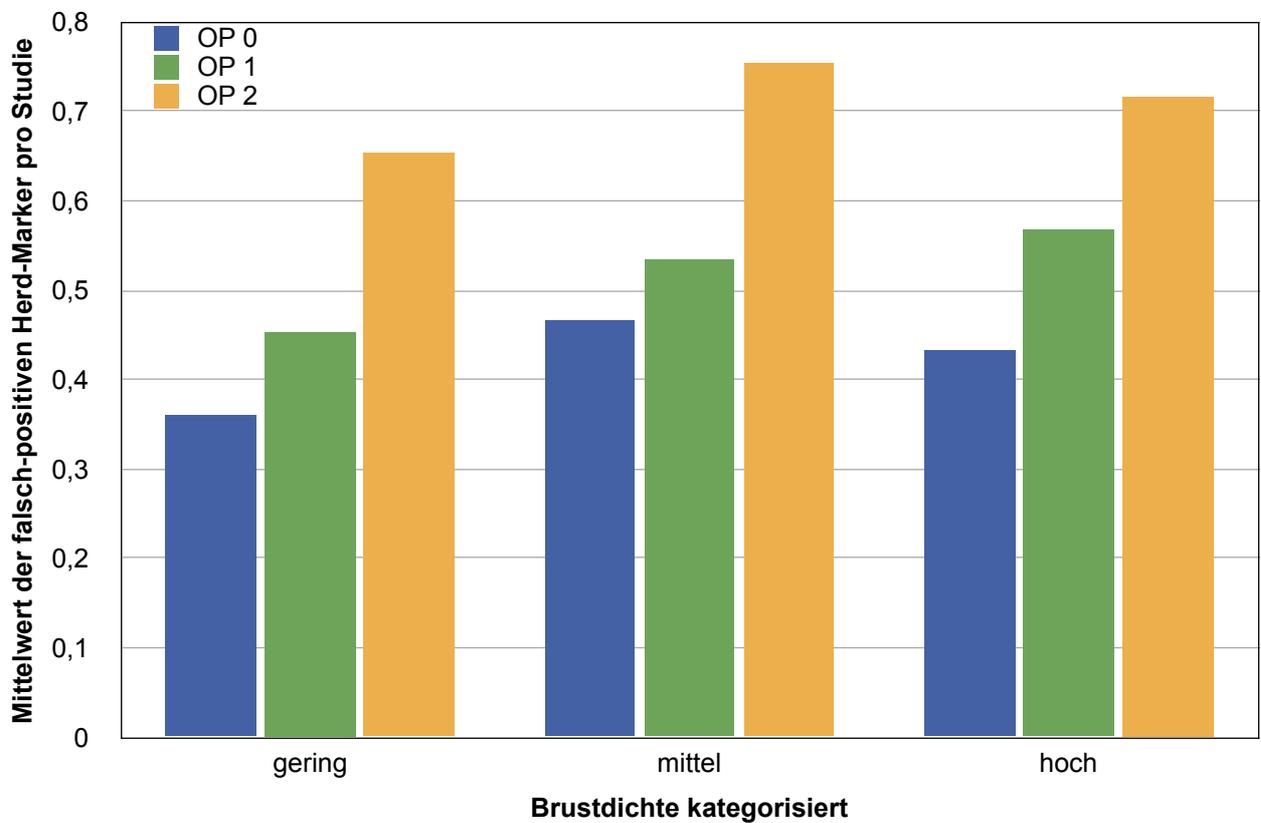


Abbildung 10: Mittelwerte der falsch-positiven Herd-Marker pro Studie (CC und MLO) je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Brustdichte.

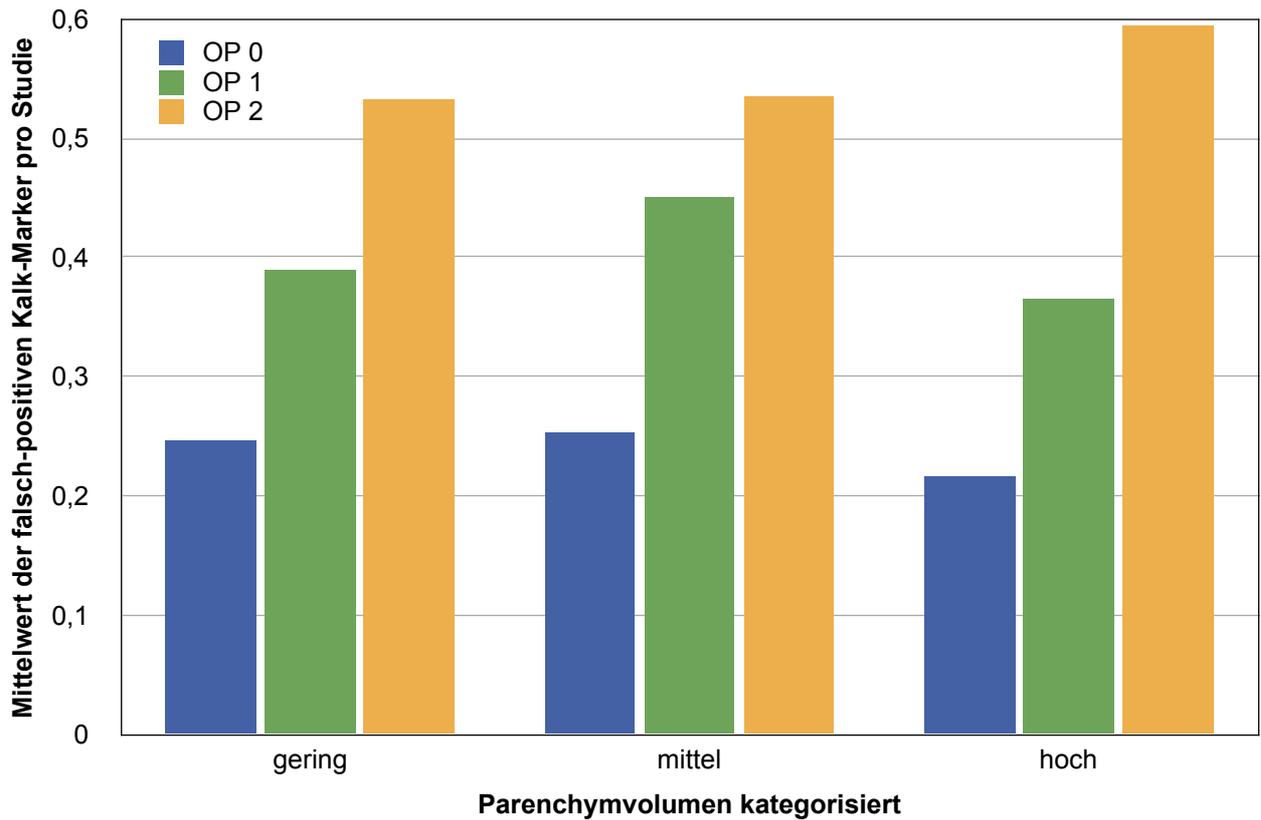


Abbildung 11: Mittelwerte der falsch-positiven Kalk-Marker pro Studie (CC und MLO) je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Parenchymvolumen.

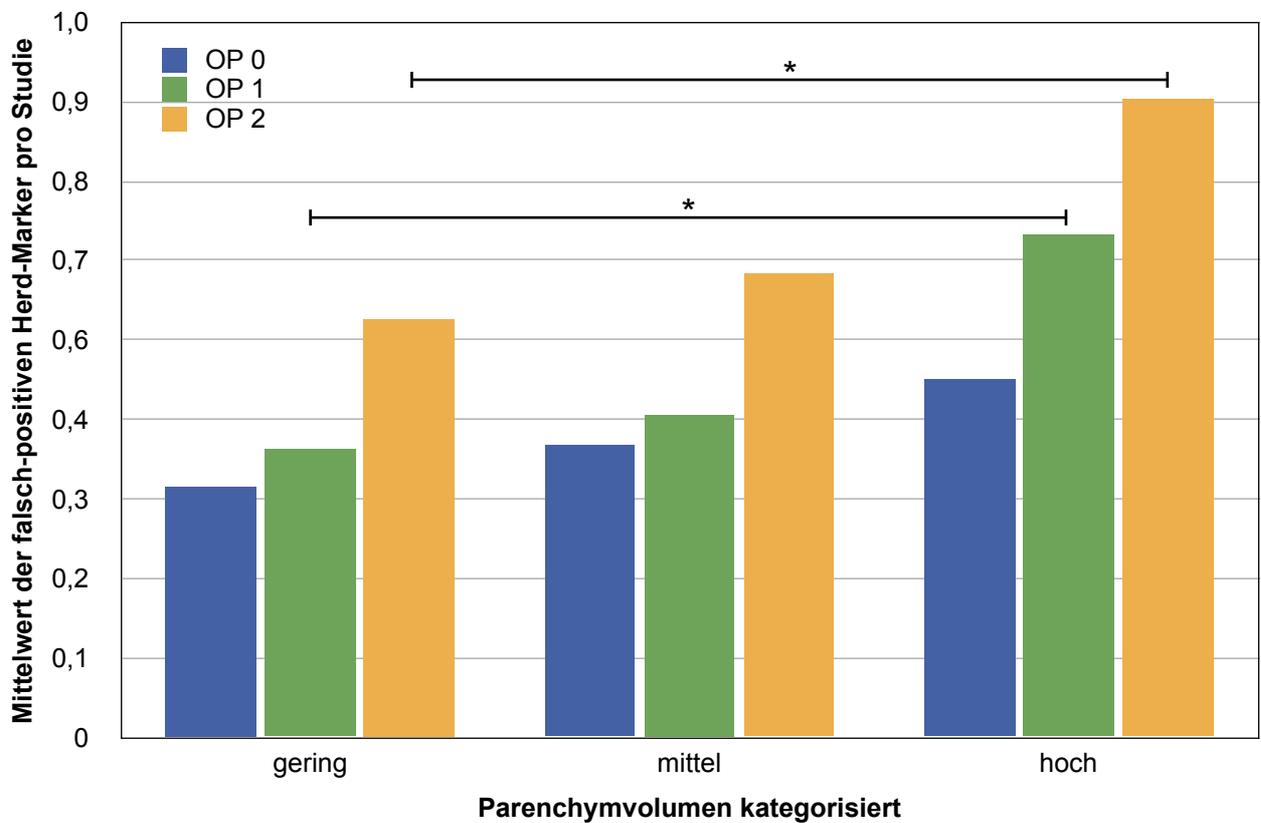


Abbildung 12: Mittelwerte der falsch-positiven Herd-Marker pro Studie (CC und MLO) je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Parenchymvolumen. Signifikante Unterschiede sind markiert (*).

4.3.3 Kategorisierung nach Anzahl der Marker

Neben der Kategorisierung nach Dichte oder Parenchymvolumen bietet sich auch die Kategorisierung der Ergebnisse nach der Anzahl der falsch-positiven Marker pro Studie an. Sollte die Brustdichte oder das Parenchymvolumen einen Einfluss auf die Anzahl haben, so wäre zu erwarten, dass der Mittelwert der Dichte bzw. des Parenchymvolumens bei Aufnahmen mit mehreren Markern größer ist als bei Aufnahmen ohne Marker.

Um diese Fragestellung zu beantworten, wurden die untersuchten Mammographien erneut kategorisiert. Eine Gruppe bilden so die Aufnahmen ohne falsch-positive Marker, eine zweite Gruppe besteht aus den Aufnahmen mit mehreren (2 oder mehr) Markern. Die Aufteilung wurde so gewählt, um eine deutliche Trennung der Gruppen zu ermöglichen.

In den Tabellen 13 und 14 sind die Ergebnisse dieser Untersuchung zusammengefasst.

Anzahl Kalk-Marker	0			≥2		
	0	1	2	0	1	2
Operating Point						
Anzahl Fälle	188	170	157	12	23	31
Dichte (MW ± s)	23,4 ± 11,19	23,2 ± 11,25	23,6 ± 11,48	24,9 ± 12,41	23,6 ± 12,13	22,7 ± 11,06
Parenchymvolumen (MW ± s)	100,8 ± 60,05	99,4 ± 55,94	99,6 ± 56,64	101,9 ± 51,82	102,1 ± 76,16	106,6 ± 71,74

Tabelle 13: Kategorisierung der untersuchten Studien (CC und MLO) nach der Anzahl der Kalk-Marker. Bei den drei Einstellungen des Operating Point sind jeweils die Anzahl der Fälle angegeben sowie Mittelwert (MW) und Standardabweichung (s) der Dichte und des Parenchymvolumens.

Anzahl Herd-Marker	0			≥2		
	0	1	2	0	1	2
Operating Point						
Anzahl Fälle	153	143	123	21	28	44
Dichte (MW ± s)	24,1 ± 12,03	23,8 ± 11,88	23,9 ± 12,03	26,2 ± 10,89	25,2 ± 10,54	24,2 ± 9,91
Parenchymvolumen (MW ± s)	95,1 ± 54,07	92,9 ± 53,21	90,3 ± 51,59	120,8 ± 91,98	122,6 ± 83,54	113,9 ± 74,02

Tabelle 14: Kategorisierung der untersuchten Studien (CC und MLO) nach der Anzahl der Herd-Marker. Bei den drei Einstellungen des Operating Point sind jeweils die Anzahl der Fälle angegeben sowie Mittelwert (MW) und Standardabweichung (s) der Dichte und des Parenchymvolumens.

Ein deutlicher Anstieg des Mittelwertes einer Testvariablen findet sich nur für das Parenchymvolumen bei der Trennung nach Anzahl der Herd-Marker (OP 0:

95,1/120,8ml; OP 1: 92,9/122,6ml; OP 2: 90,3/113,9ml). Bei der Trennung nach Anzahl der Kalk-Marker findet sich ein minimaler Anstieg (OP 0: 100,8/101,9ml; OP 1: 99,4/102,1ml; 99,6/106,6ml).

Bei der Untersuchung des Mittelwertes der Brustdichte findet sich für die Kalk-Marker kein durchgängiger Anstieg (OP 0: 23,4/24,9%; OP 1: 23,2/23,6%, OP 2: 23,6/22,7%) für die Herdmarker findet sich ein minimaler Anstieg (OP 0: 24,1/26,2%; OP 1: 23,8/25,2%; 23,9/24,2%).

Um Aussagen über die Signifikanz treffen zu können, wurden die Ergebnisse mit Hilfe der Mann-Whitney-U-Test untersucht. Die p-Werte sind in den Tabellen 15 und 16 zusammengestellt.

	OP	0 / ≥2
Dichte	0	0,359
	1	0,489
	2	0,377
Parenchymvolumen	0	0,395
	1	0,346
	2	0,432

Tabelle 15: Mann-Whitney-U-Test, Signifikanz des Anstiegs des Mittelwertes der Testvariable bei größerer Anzahl falsch-positiver Kalk-Marker

	OP	0 / ≥2
Dichte	0	0,135
	1	0,160
	2	0,262
Parenchymvolumen	0	0,199
	1	0,046 *
	2	0,034 *

Tabelle 16: Mann-Whitney-U-Test, Signifikanz des Anstiegs des Mittelwertes der Testvariable bei größerer Anzahl falsch-positiver Herd-Marker

Ein signifikanter Anstieg findet sich nur bei der Untersuchung des Parenchymvolumens und der Aufteilung nach Anzahl der Herd-Marker. Der Anstieg ist bei den Einstellungen OP 1 (p=0,046) und OP 2 (p=0,034) signifikant.

Um einen besseren Überblick zu ermöglichen, sind die Ergebnisse im Folgenden in Balkendiagrammen zusammengefasst.

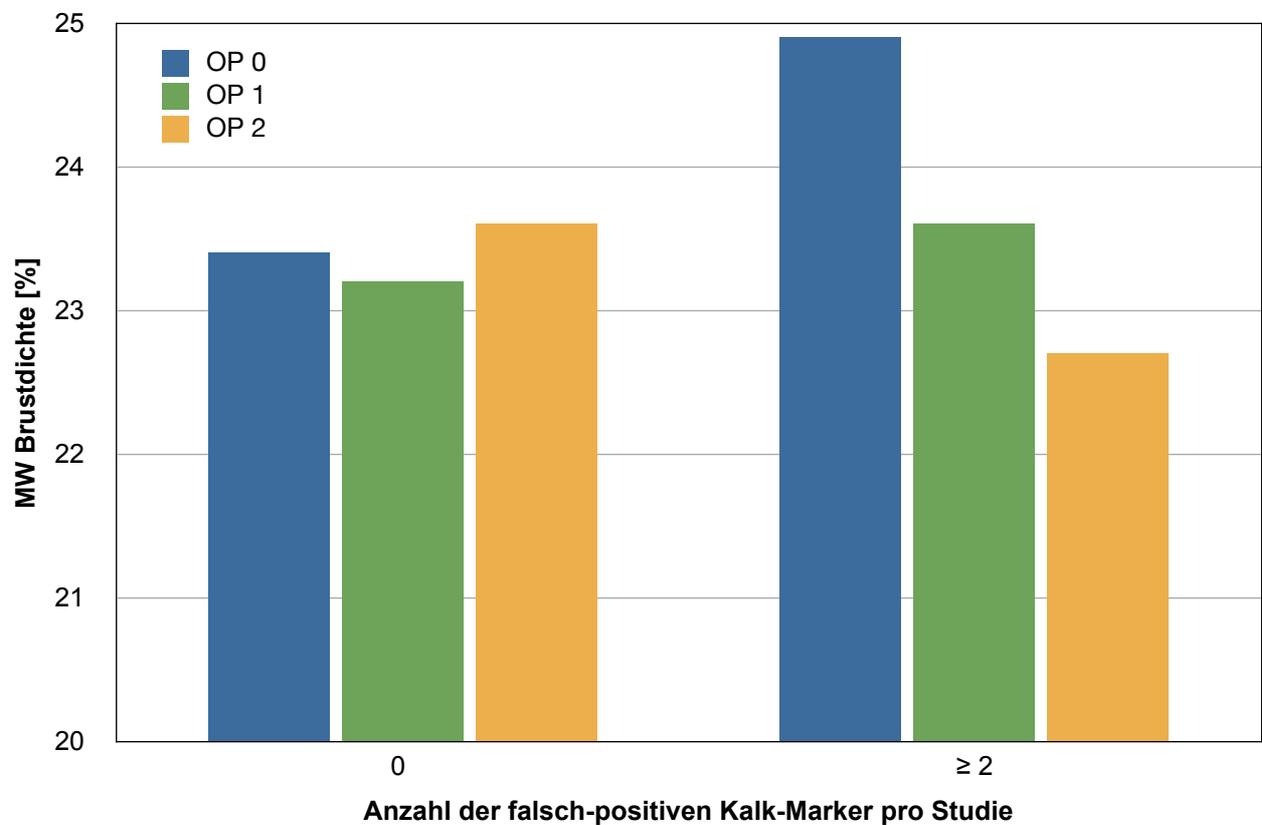


Abbildung 13: Mittelwert (MW) der Brustdichte in Prozent je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Anzahl der falsch-positiven Kalk-Marker pro Studie (CC und MLO).

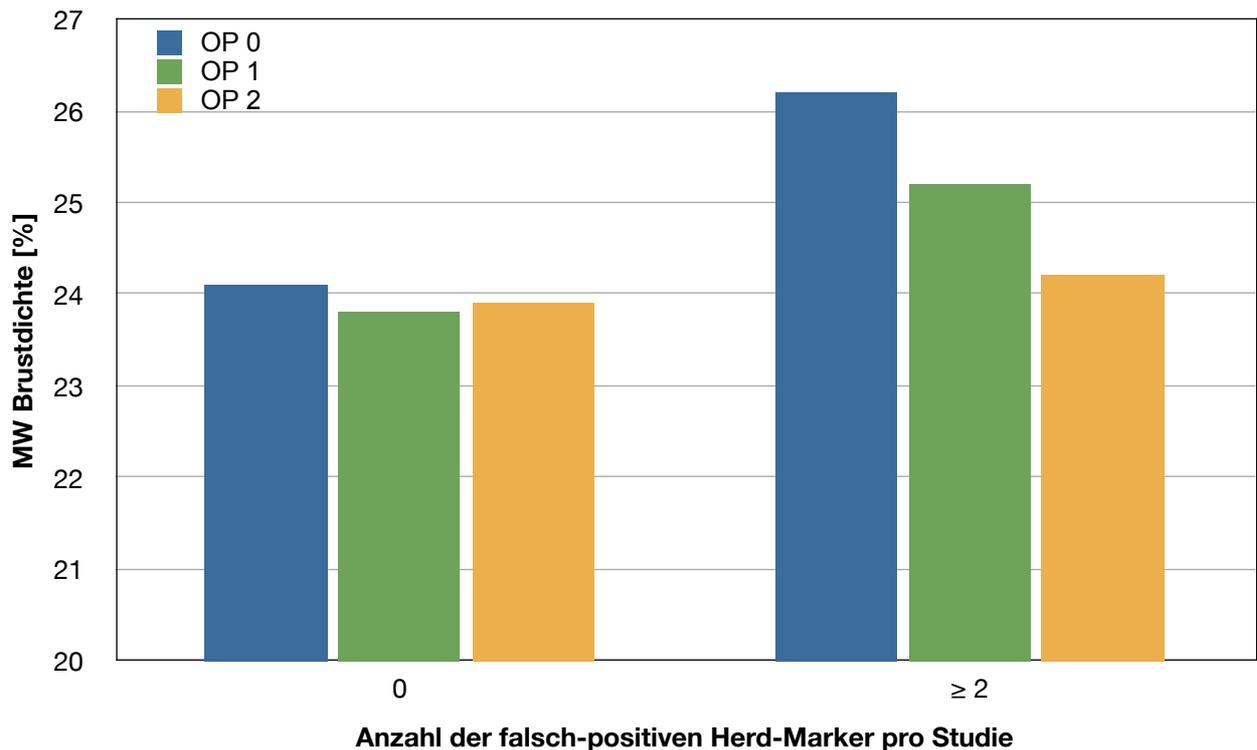


Abbildung 14: Mittelwert (MW) der Brustdichte in Prozent je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Anzahl der falsch-positiven Herd-Marker pro Studie (CC und MLO).

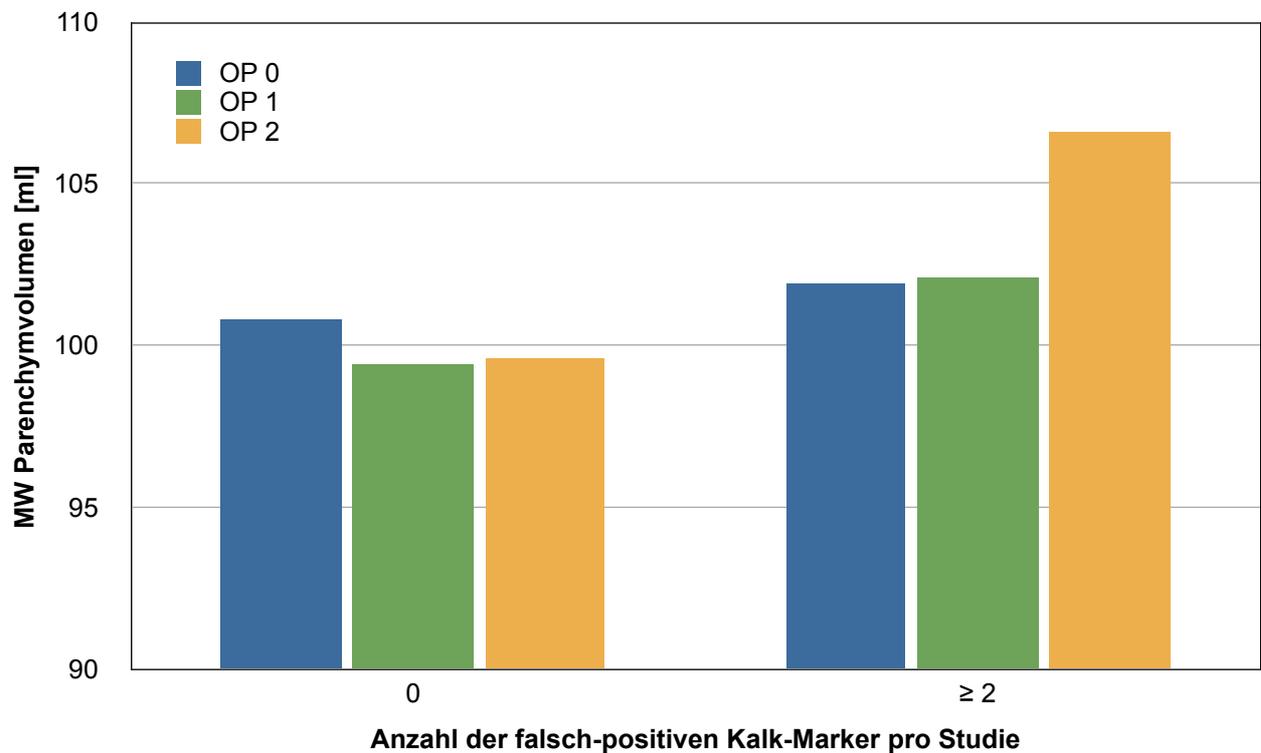


Abbildung 15: Mittelwert (MW) des Parenchymvolumens in ml je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Anzahl der falsch-positiven Kalk-Marker pro Studie (CC und MLO).

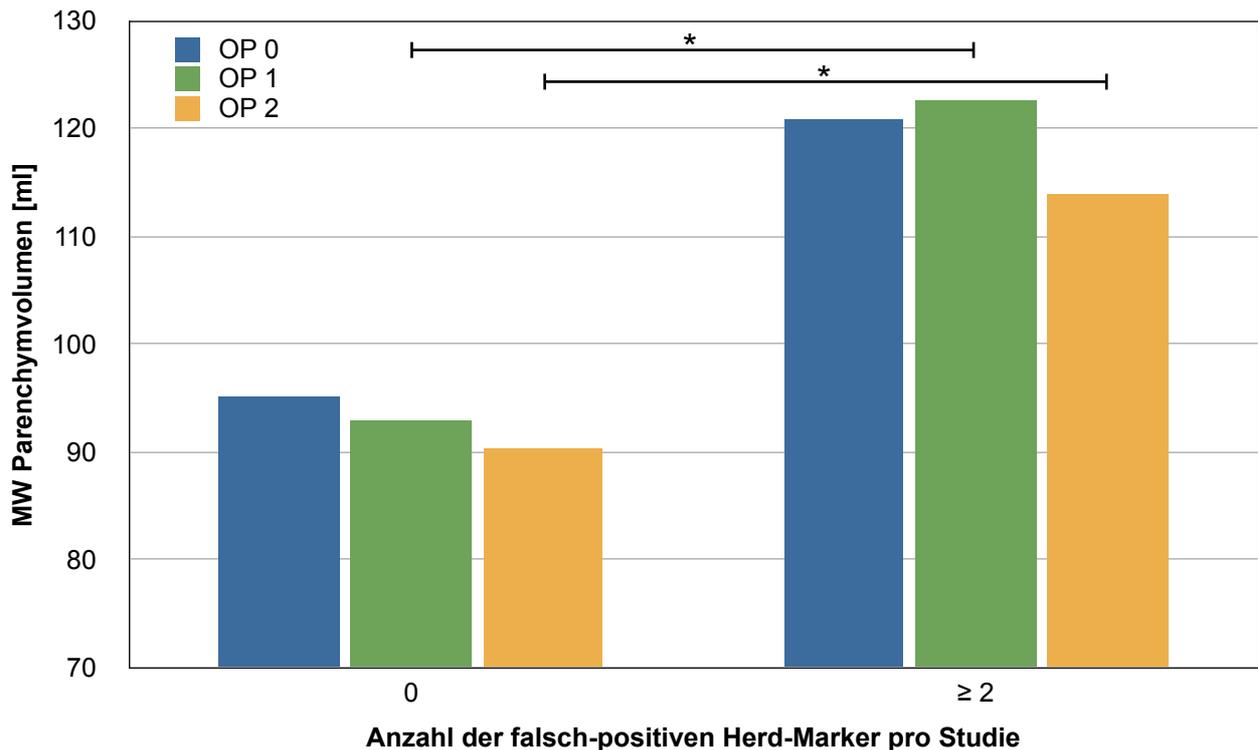


Abbildung 16: Mittelwert (MW) des Parenchymvolumens in ml je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Anzahl der falsch-positiven Kalk-Marker pro Studie (CC und MLO). Signifikante Unterschiede sind markiert (*).

4.3.4 Korrelationsanalyse

Um den Einfluss der beiden Variablen Brustdichte und Parenchymvolumen auf die Anzahl der falsch-positiven Marker möglichst exakt zu bestimmen, wurde eine Korrelations- und eine Regressionsanalyse durchgeführt. Dazu war es zunächst erforderlich, Punktwolken zu erstellen. Der Einfluss von Brustdichte und Parenchymvolumen wurde jeweils für Kalk- und Herd-Marker getestet. Da jeweils drei Einstellungen des Operating Point untersucht wurden, wurden insgesamt zwölf Parameterkombinationen analysiert.

Dichte	Kalk-Marker	Operating Point: 0, 1, 2
	Herd-Marker	Operating Point: 0, 1, 2
Parenchymvolumen	Kalk-Marker	Operating Point: 0, 1, 2
	Herd-Marker	Operating Point: 0, 1, 2

Tabelle 17: Zusammenstellung der Parameterkombinationen zur Korrelations- und Regressionsanalyse

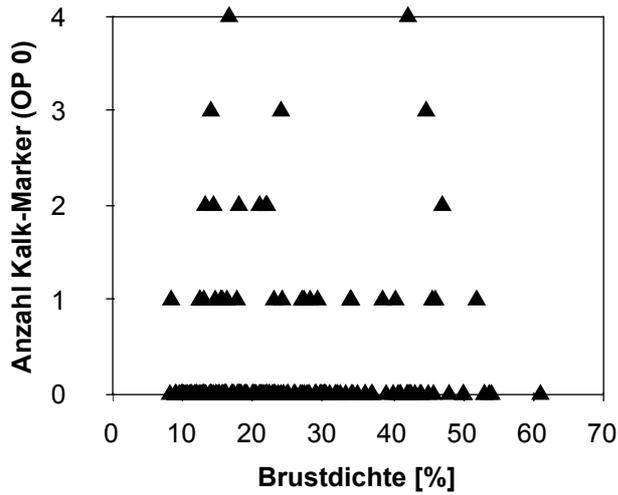


Abbildung 17: Anzahl der Kalk-Marker über der Brustdichte in Prozent, Operating Point 0 (OP 0)

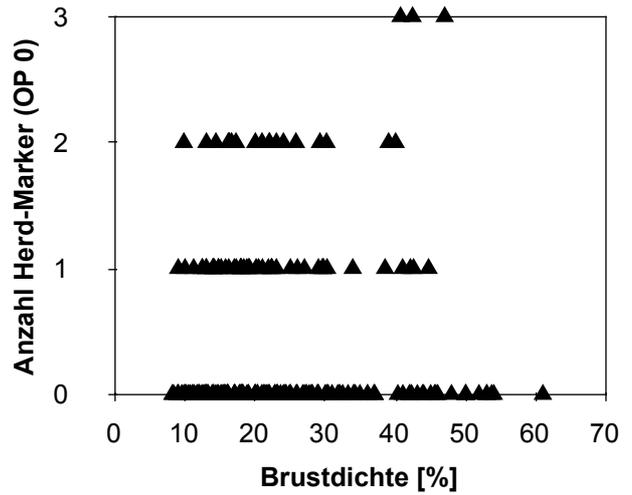


Abbildung 18: Anzahl der Herd-Marker über der Brustdichte in Prozent, Operating Point 0 (OP 0)

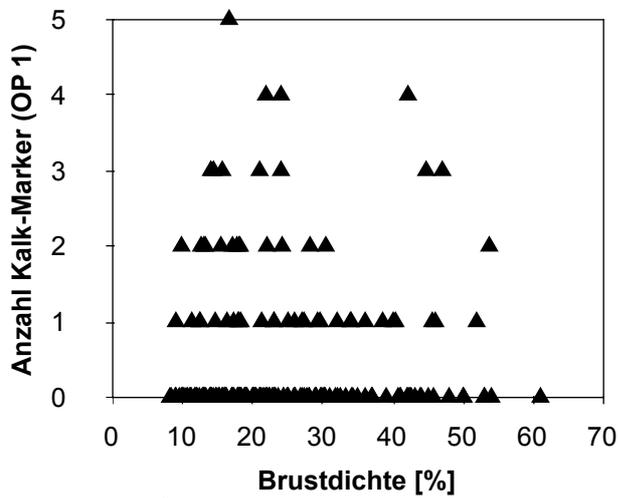


Abbildung 19: Anzahl der Kalk-Marker über der Brustdichte in Prozent, Operating Point 1 (OP 1)

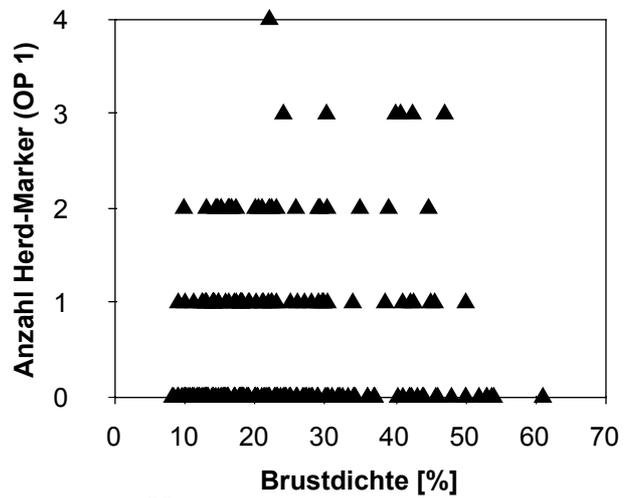


Abbildung 20: Anzahl der Herd-Marker über der Brustdichte in Prozent, Operating Point 1 (OP 1)

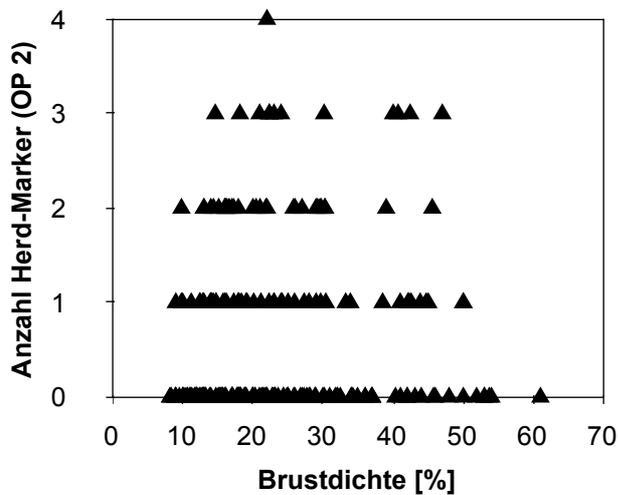


Abbildung 21: Anzahl der Kalk-Marker über der Brustdichte in Prozent, Operating Point 2 (OP 2)

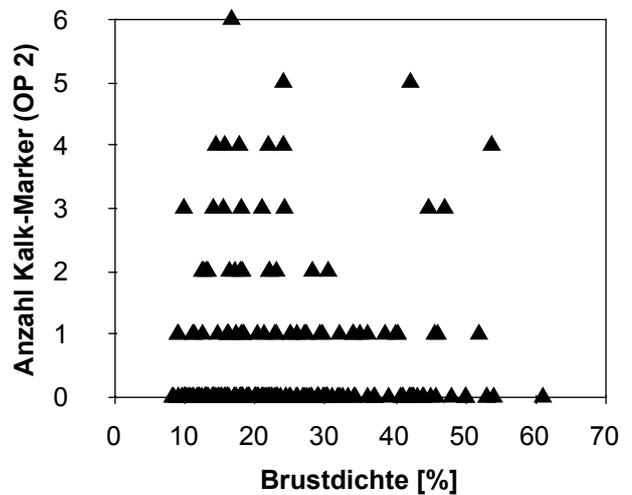


Abbildung 22: Anzahl der Herd-Marker über der Brustdichte in Prozent, Operating Point 2 (OP 2)

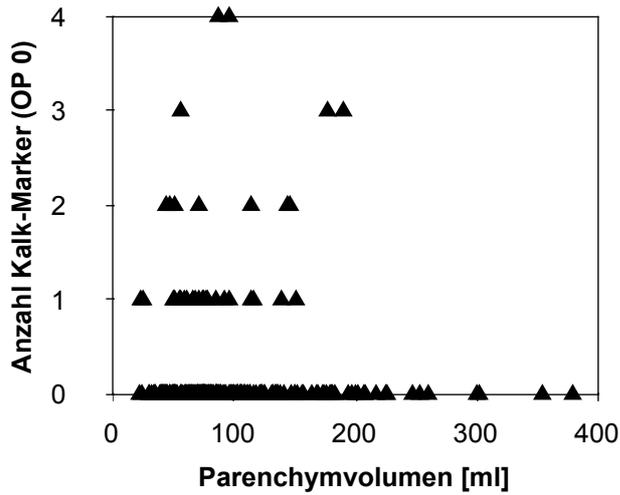


Abbildung 23: Anzahl der Kalk-Marker über dem Parenchymvolumen in ml, Operating Point 0 (OP 0)

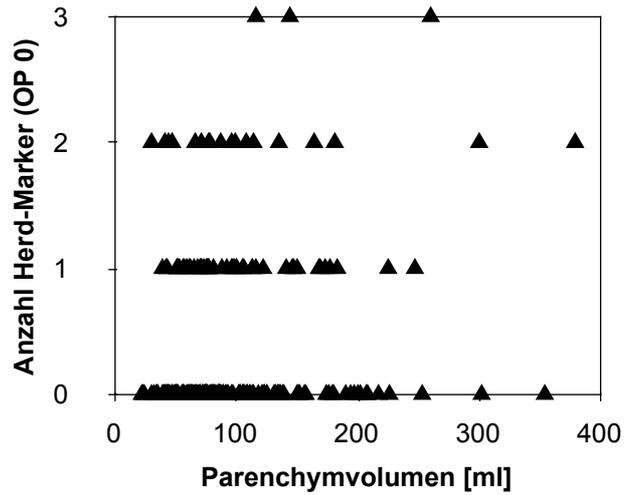


Abbildung 24: Anzahl der Herd-Marker über dem Parenchymvolumen in ml, Operating Point 0 (OP 0)

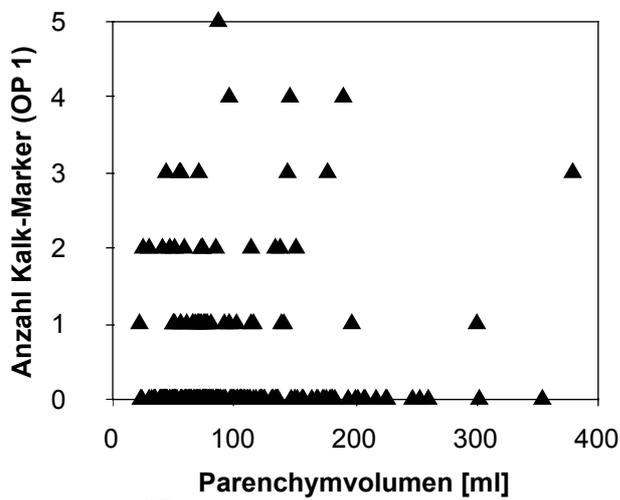


Abbildung 25: Anzahl der Kalk-Marker über dem Parenchymvolumen in ml, Operating Point 1 (OP 1)

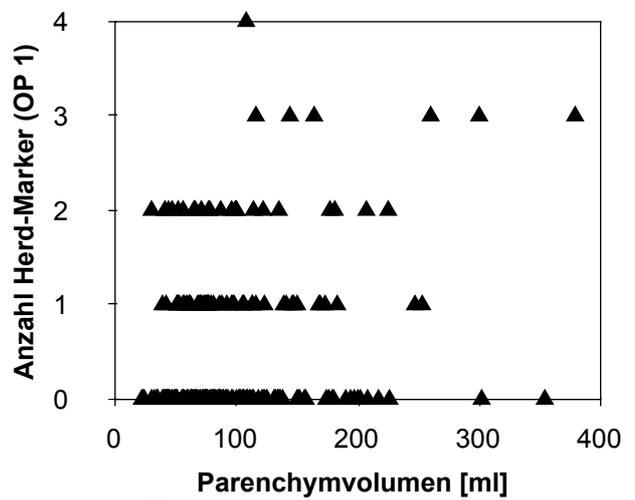


Abbildung 26: Anzahl der Herd-Marker über dem Parenchymvolumen in ml, Operating Point 1 (OP 1)

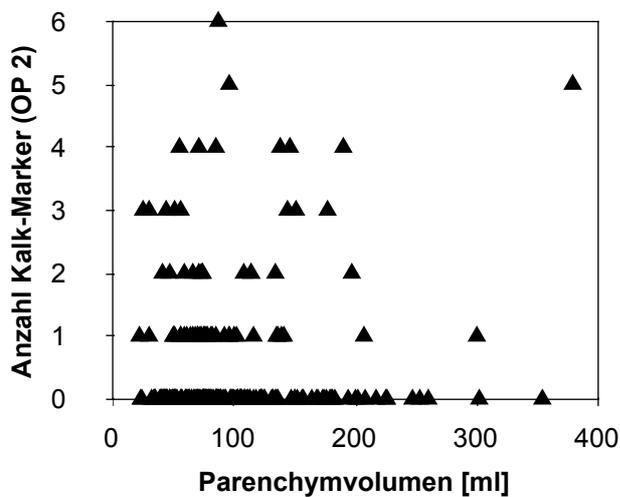


Abbildung 27: Anzahl der Kalk-Marker über dem Parenchymvolumen in ml, Operating Point 2 (OP 2)

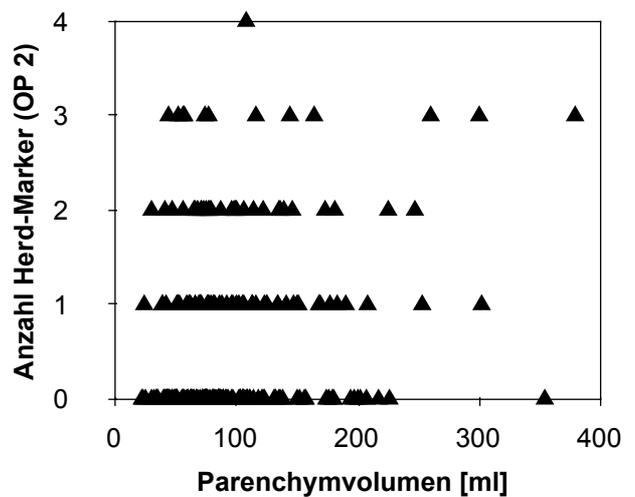


Abbildung 28: Anzahl der Herd-Marker über dem Parenchymvolumen in ml, Operating Point 2 (OP 2)

Um zu ermitteln, wie groß ein eventuell vorhandener Einfluss der Variablen Parenchymvolumen und Brustdichte auf die Anzahl der Marker ist, wurde der Korrelationskoeffizient nach Pearson (r) berechnet. Ist das Ergebnis signifikant, weist ein positiver Korrelationskoeffizient auf eine gleichsinnige Beziehung hin, während ein negatives Ergebnis auf eine gegensinnige Beziehung hinweist. Das Quadrat des Korrelationskoeffizienten wird als Bestimmtheitsmaß (r^2) bezeichnet und gibt in Prozent wieder, wie groß der Einfluss der untersuchten Variable in Bezug auf alle Einflussfaktoren ist.

Die Ergebnisse der Korrelationsanalyse nach Pearson sind in Tabelle 18 zusammengefasst.

		Parenchymvolumen			Dichte		
		r	r ²	p	r	r ²	p
Herd-Marker	OP 0	0,136	0,0185	0,021 *	0,035	0,0012	0,304
	OP 1	0,219	0,0480	0,001 *	0,058	0,0034	0,196
	OP 2	0,188	0,0353	0,002 *	0,024	0,0006	0,363
Kalk-Marker	OP 0	-0,036	0,0013	0,298	0,069	0,0048	0,153
	OP 1	0,027	0,0007	0,342	0,052	0,0027	0,219
	OP 2	0,072	0,0052	0,141	0,021	0,0004	0,378

Tabelle 18: Zusammenfassung der Ergebnisse der Korrelationsanalyse nach Pearson. Angegeben sind der Korrelationskoeffizient (r), das Bestimmtheitsmaß (r^2), und der p-Wert (p). Aufteilung nach Operating Point (OP).

Ein signifikantes Ergebnis wird nur für den Einfluss der Menge des Parenchymvolumens auf die Anzahl der Herd-Marker ermittelt. Hier ist der p-Wert für alle drei Einstellungen des Operating Point kleiner 0,05. Der positive Korrelationskoeffizient weist auf eine gleichsinnige Beziehung hin, d.h. mehr Parenchymvolumen führt zu einer Zunahme falsch-positiver Herd-Marker.

Der Anteil des Einflusses der untersuchten Variable am gesamten Einfluss auf die Marker-Anzahl wird durch das Bestimmtheitsmaß als 1,85% (OP 0), 4,80% (OP 1), bzw. 3,53% (OP 2) beschrieben. Ein Einfluss der Dichte auf die Anzahl der Herdmarker oder ein Einfluss einer Variable auf die Anzahl der Kalkmarker lässt sich nicht nachweisen.

4.3.5 Regressionsanalyse

Um zu bestimmen, wie sich eine Zunahme des Parenchymvolumens oder der Dichte auf die Anzahl falsch-positiver Marker quantitativ auswirkt, wurde eine Regressionsanalyse nach Poisson durchgeführt. Ein positiver Regressionskoeffizient weist bei größerem Parenchymvolumen beziehungsweise größerer Dichte auf eine Zunahme der Testvariable hin, während ein negatives Ergebnis auf eine Abnahme der Testvariable schließen lässt. Bei einem signifikanten Zusammenhang errechnet sich das relative Risiko (RR) für falsch-positive Marker aus Regressionskoeffizient (β) und Differenz der Testvariable (Δa) wie folgt:

$$RR = (1 + \beta)^{\Delta a}$$

Die der Regressionsanalyse zugrunde liegenden Punktwolken entsprechen denen der Korrelationsanalyse (s. 3.3.4). Die Ergebnisse der Regressionsanalyse sind in Tabelle 19 zusammengefasst.

		Parenchymvolumen			Dichte		
		β	RR (+100ml)	p	β	RR (+20%)	p
Herd-Marker	OP 0	0,0036	1,43	0,034 *	-0,0036	-1,07	0,719
	OP 1	0,0049	1,63	0,000 *	-0,0046	-1,10	0,606
	OP 2	0,0041	1,50	0,001 *	-0,0072	-1,16	0,352
Kalk-Marker	OP 0	-0,0042	-1,51	0,162	0,0237	1,60	0,057
	OP 1	0,0002	1,02	0,914	0,0090	1,20	0,364
	OP 2	0,0023	1,26	0,141	-0,0018	-1,04	0,832

Tabelle 19: Zusammenfassung der Ergebnisse der Regressionsanalyse nach Poisson. Angegeben sind der Regressionskoeffizient (β), das relative Risiko (RR) und der p-Wert (p). Aufteilung nach Operating Point (OP).

Signifikante Ergebnisse zeigen sich bei der Prüfung des Einflusses des Parenchymvolumens auf die Anzahl der falsch-positiven Herd-Marker. Der p-Wert ist für alle drei Einstellungen des Operating Point kleiner als 0,05. In einer Untersuchung mit einem 100 ml größerem Parenchymvolumen ist das relative Risiko für falsch positive Herdmarker also 1,43 (OP0), 1,63 (OP1), bzw. 1,50 (OP2). Ein Einfluss der Dichte auf die Anzahl der Herdmarker oder ein Einfluss einer Variable auf die Anzahl der Kalkmarker lässt sich auch in der Regressionsanalyse nicht nachweisen.

5. Diskussion

5.1 Einordnung der Ergebnisse

Der CAD-Algorithmus weist nach den Ergebnissen dieser Studie, je nach Einstellung des Operating Point, eine Spezifität von 61,3% (OP0), 51,8% (OP1), bzw. 43,7% (OP2) auf. Diese Werte beziehen sich auf die Untersuchung einer Seite. Bei einer Untersuchung beider Seiten beziffert der Hersteller die Spezifität auf 48,3% (OP0), 35,1% (OP1) bzw. 24,5% (OP2)³⁴.

Ziel dieser Studie ist es, Faktoren zu identifizieren, welche die Spezifität des eingesetzten CAD-Algorithmus beeinflussen. Dadurch soll eine weitere Optimierung der CAD-Analyse unterstützt werden, welche es erlaubt, deren Spezifität zu verbessern.

5.1.1 Einfluss der Brustdichte

Um einen eventuellen Einfluss der Brustdichte oder des Parenchymvolumens auf die Anzahl der falsch-positiven CAD-Marker zu ermitteln, wurden die Mammographien nach diesen Parametern kategorisiert und in jeder Kategorie der Mittelwert der falsch-positiven Marker pro Fall berechnet. Anschließend wurden die Kategorien mit Hilfe des Mann-Whitney-U-Tests verglichen, um die Signifikanz eines gegebenenfalls bestehenden Unterschiedes zu ermitteln. Bei der Kategorisierung der Mammographien nach Brustdichte lassen sich keine signifikanten Unterschiede der Kategorien bezüglich der Markeranzahl feststellen. Weder für Herd- noch für Kalkmarker lässt sich ein vermehrtes Auftreten in den höheren Kategorien gegenüber den niedrigeren Kategorien nachweisen.

In einem zweiten Schritt wurden die Mammographien nach der Markeranzahl kategorisiert. Dabei weisen weder die Aufnahmen mit mehreren Herdmarkern, noch die Aufnahmen mit mehreren Kalkmarkern eine signifikant höhere Brustdichte auf als die Aufnahmen ohne Herd- beziehungsweise Kalkmarker.

Die Einteilung der Ergebnisse in Kategorien bietet einen groben Überblick über die erhobenen Daten. Um eine genauere Beurteilung der Ergebnisse zu ermöglichen, wurde eine Korrelationsanalyse sowie eine Regressionsanalyse der Daten durchgeführt. Beide Analysen zeigen keinen signifikanten Zusammenhang zwischen der Brustdichte und der Anzahl der falsch-positiven Marker.

Die Ergebnisse dieser Studie lassen insgesamt den Schluss zu, dass die Brustdichte keinen Einfluss auf die Rate der falsch-positiven Marker hat. Unterschiede der

Brustdichte beeinflussen offensichtlich weder die Wahrscheinlichkeit des Auftretens von Herdmarkern, noch von Kalkmarkern.

In der Literatur finden sich mehrere Arbeiten, die sich mit dem Einfluss der Brustdichte auf die Spezifität von CAD-Systemen auseinandersetzen. Eine Arbeit von W. Ho und P. Lam analysiert 264 nachträglich digitalisierte Mammographien³⁵. Die Autoren kommen zu dem Ergebnis, dass eine erhöhte Brustdichte zwar die Sensitivität des CAD-Algorithmus (SecondLook, Version 1.1, CADx Medical Systems, Beavercreek, OH, USA) einschränkt, die Spezifität aber nicht verändert.

R. Brem et. al. untersuchen 147 unauffällige Mammographien, die sie manuell in zwei Dichtekategorien einteilen (SecondLook, Version 4.0, i-CAD Inc., Fairborn, OH, USA)³⁸. Hier ist der Mittelwert der falsch-positiven Herdmarker in der Gruppe der Aufnahmen mit höherer Brustdichte größer als in der anderen Gruppe. Dieser Unterschied ist statistisch signifikant ($p=0,04$). Bezüglich der Kalkmarker ergibt sich kein signifikanter Unterschied zwischen den Gruppen.

Zu ähnlichen Ergebnissen kommen auch A. Malich et. al. in ihrer Studie, die unter anderem die Spezifität eines CAD-Algorithmus (SecondLook, Version 5.0, i-CAD Inc., Fairborn, OH, USA) in einer Studie mit 200 unauffälligen Mammographien untersucht³⁹. Die Aufnahmen werden in vier Dichtekategorien geteilt. Ein vermehrtes Auftreten von falsch-positiven Markern in höheren Dichtekategorien ist für Herdmarker signifikant.

Die unterschiedlichen Ergebnisse dieser Studien, auch im Vergleich mit der vorliegenden Arbeit, lassen sich zum Teil mit den Unterschieden des Studiendesigns erklären. Sie unterscheiden sich vor allem in der Fallzahl (Ho, W.: 264; Brem, R.: 147; Malich, A.: 200; diese Arbeit: 222) der Einschätzung der Brustdichte (4 Kategorien; 2 Kategorien; 4 Kategorien; 3 Kategorien und linear), der Aufnahmetechnik (digitalisierte Film-Mammographie; sonst Digitalmammographie) und bezüglich des eingesetzten CAD-Systems (SecondLook von CADx Medical Systems/i-CAD Inc., Version 1.1; Version 4.0; Version 5.0; R2 ImageChecker von Hologic, Version 9.3).

Die Frage, ob die Brustdichte die Spezifität von CAD-Untersuchungen beeinflusst, wird unterschiedlich beantwortet. Die hier vorliegende Arbeit kommt zu dem Schluss, dass kein signifikanter Zusammenhang besteht. In der Literatur finden sich Arbeiten, die einen solchen Zusammenhang aufzeigen, allerdings nutzen diese Untersuchungen ein anderes CAD-System.

Es ist möglich, dass die Spezifität des R2 ImageChecker von der Brustdichte unabhängig ist, während die Ergebnisse des SecondLook von der Brustdichte beeinflusst werden. Die einzige Studie, die hier ebenfalls keinen Zusammenhang erkennt, ist die älteste der hier berücksichtigten Untersuchungen und arbeitet mit digitalisierten Film-Mammographien. Insgesamt finden sich zur Beantwortung dieser Frage nur wenige Untersuchungen mit relativ geringer Fallzahl.

5.1.2 Einfluss des Parenchymvolumens

Anders als bei der Brustdichte lassen sich bei einer Kategorisierung der untersuchten Fälle nach dem Parenchymvolumen signifikante Unterschiede der einzelnen Gruppen bezüglich der Markeranzahl nachweisen. Ist der Operating Point „1“ oder „2“, ist die Markeranzahl in der höchsten Kategorie signifikant größer als in der niedrigsten Kategorie (OP1: $p=0,042$; OP2: $p=0,021$).

Ähnliches lässt sich auch bei der Kategorisierung nach der Anzahl der Herdmarker finden. Die Aufnahmen mit mehreren Herdmarkern weisen bei den Einstellungen „1“ und „2“ des Operating Point signifikant mehr Parenchymvolumen auf als die Aufnahmen ohne Herdmarker (OP1: $p=0,046$; OP2: $p=0,034$).

Ist der Operating Point auf „0“ eingestellt, findet sich in beiden Untersuchungen kein signifikanter Unterschied der Kategorien. Sollte allerdings das Parenchymvolumen die Spezifität des CAD-Algorithmus beeinflussen, so ist es unwahrscheinlich, dass dieser Einfluss nur bei bestimmten Einstellungen des Operating Point auftritt. Wahrscheinlicher ist, dass ein existierender Zusammenhang in diesen Analysen nicht korrekt dargestellt werden konnte. Dies ist gerade für die möglichst spezifische Einstellung des Operating Point (OP0) denkbar, weil dadurch die Fallzahlen in der jeweils kleinsten Kategorie besonders niedrig sind.

Diese Vermutung wird auch dadurch gestützt, dass sowohl die Korrelations- als auch die Regressionsanalyse, für die dieser Effekt jeweils keine Rolle spielt, bei allen drei Einstellungen des Operating Point einen Einfluss des Parenchymvolumens zeigen. Beide Analysen weisen auf einen signifikanten Zusammenhang zwischen dem Parenchymvolumen und der Anzahl der falsch-positiven Herdmarker hin und sind dabei in der Lage, diesen weiter zu beschreiben.

In der Korrelationsanalyse wird das Bestimmtheitsmaß mit 1,85% (OP0), 4,80% (OP1) und 3,53% (OP2) angegeben. Der p-Wert liegt unter dem geforderten Signifikanzniveau

von 0,05 (OP0: $p=0,021$; OP1: $p=0,001$; OP2: $p=0,002$). Die Menge des Parenchymvolumens scheint also einen Einfluss auf die Anzahl der Herdmarker zu haben, auch wenn dieser im Vergleich zu anderen Einflussfaktoren offensichtlich gering ist.

Auch die Regressionsanalyse weist auf einen signifikanten Zusammenhang zwischen dem Parenchymvolumen und der Anzahl der Herdmarker hin (OP0: $p=0,034$; OP1: $p=0,000$; OP2: $p=0,001$). Der Regressionskoeffizient ist dabei 0,0036 (OP0), 0,0049 (OP1) und 0,0041 (OP2). Daraus ergibt sich, dass bei einem um 100 ml größeren Parenchymvolumen das relative Risiko für falsch-positive Herdmarker 1,43 (OP0), 1,63 (OP1), bzw. 1,50 (OP2) beträgt.

In diese Studie waren Untersuchungen mit 22 ml bis 379 ml Parenchymvolumen eingeschlossen, die Standardabweichung betrug bei diesem Parameter ± 58 ml. Eine Schwankung um 100 ml oder mehr ist also auch in der Praxis realistisch.

Insgesamt lassen die Ergebnisse dieser Studie auf einen Einfluss des Parenchymvolumens auf die Spezifität des CAD-Algorithmus bezüglich der Herdmarker schließen. Läsionen, die der CAD-Algorithmus als Herd markiert, sind Parenchymgewebe bildmorphologisch ähnlich. Ergibt sich ein falsch-positiver Marker aus einer Fehleinschätzung von Parenchymgewebe als maligne Läsion, so ist es nachvollziehbar, dass eine Zunahme von Gewebe zu einem vermehrten Auftreten solcher Marker führt.

Das erklärt auch, warum die Anzahl der Kalk-Marker nicht vom Parenchymvolumen beeinträchtigt wird. Mikrokalzifikationen sind in der Mammographie gut von anderem Gewebe zu unterscheiden.

Bisher sind keine weiteren Studien veröffentlicht, die das Parenchymvolumen als Einflussfaktor auf die Spezifität von CAD-Analysen der Mammographie untersuchen. Die Ergebnisse der hier vorgelegten Arbeit weisen darauf hin, dass ein solcher Einfluss existiert. Vor allem auf Grund der relativ geringen Fallzahl sollten diese Ergebnisse in weiteren Studien nachvollzogen werden.

5.2 Limitationen und Schwächen der Arbeit

5.2.1 Zusammenstellung der Studienpopulation

Die Zusammenstellung der Studienpopulation stellt eine erste Schwäche dieser Arbeit dar. So wurden nicht alle Untersuchungen berücksichtigt, die den Einschlusskriterien

entsprachen, sondern nur ein Teil einbezogen. Dabei besteht die Gefahr, dass eine zufällige Selektion der Ausgangsdaten die Ergebnisse der Studie beeinträchtigt. Außerdem erhöht sich bei einer geringeren Fallzahl die Wahrscheinlichkeit, einen tatsächlich bestehenden Zusammenhang zu übersehen, weil der p-Wert die vorgegebene Grenze überschreitet.

Es sollte eine Studienpopulation zusammengestellt werden, deren Größe es einerseits ermöglicht, den Einfluss der untersuchten Faktoren ausreichend genau einzuschätzen. Auf der anderen Seite erschien es erforderlich, die Fallzahl zu limitieren, um die Durchführung der Studie im Rahmen einer Doktorarbeit zu ermöglichen. Um hier einen angemessenen Mittelweg zu finden, wurde ein Statistiker zu Rate gezogen, der die erforderliche Fallzahl vor Beginn der Untersuchungen abschätzte.

5.2.2 Auswahl der Untersuchungs- und Analysemethoden

Zur CAD-Analyse und zur Analyse der Gewebeeigenschaften wurden Algorithmen des selben Herstellers eingesetzt. Dieses stellt eine potentielle Fehlerquelle dar. Ein Informationsaustausch zwischen den beiden Programmen würde zu einer Verknüpfung der untersuchten Variablen führen, was die Untersuchungsergebnisse verfälschen könnte. Ein Datenaustausch zwischen den beiden Algorithmen findet allerdings nach Aussage des Herstellers nicht statt.

Die Aufnahmetechnik der analysierten Mammographien war teilweise unterschiedlich. Das Mammographiegerät nutzt unterschiedliche Anode/Filter-Kombinationen zur Generierung der Röntgenstrahlung (Mo/Mo, Mo/Rh, Rh/Rh). Unterschiede bestehen außerdem in der Dosis, der Spannung, der Kompressionskraft und der Kompressionsdicke. Eine Beschränkung auf eine bestimmte Aufnahmetechnik hätte allerdings die Fallzahl verringert. Zudem wären die Ergebnisse zunächst nur auf Aufnahmen generalisierbar gewesen, die mit dieser einen Technik aufgenommen wurden.

Zur statistischen Analyse wurden die Ergebnisse nach mehreren Parametern kategorisiert. Dabei wurden drei Kategorien gebildet, wobei zum Beispiel im ACR-Dichteindex vier Kategorien üblich sind. Das war in dieser Untersuchung notwendig, um eine ausreichende Fallzahl in jeder definierten Kategorie zu gewährleisten. Eine weitere Unterteilung der Ergebnisse hätte zu einer geringeren Fallzahl je Kategorie geführt, was eine saubere statistische Analyse erschwert hätte.

Ein Ziel dieser Arbeit ist es, den Einfluss der Brustdichte auf die CAD-Analyse zu untersuchen. Dieser Parameter wurde von einem Algorithmus als Quotient aus Parenchymvolumen und Brustvolumen errechnet. Im klinischen Alltag sind allerdings andere Verfahren zur Einschätzung der Dichte üblich: Der Befunder klassifiziert die Aufnahme nach dem ACR-Dichteindex. Abweichungen zwischen diesen beiden Methoden sind vor allem dann zu erwarten, wenn Parenchymgewebe besonders ungleichmäßig verteilt ist. Allerdings existieren mehrere Arbeiten über die eingeschränkte Reliabilität der Brustdichteanalyse durch den Befunder^{20-22, 46}. Daher wurde das hier angewandte Verfahren bevorzugt.

5.2.3 Einschränkungen der Arbeit

In der vorliegenden Arbeit wurde der Einfluss von Brustdichte und Parenchymvolumen ausschließlich auf die Spezifität untersucht, über den Einfluss auf die Sensitivität lässt sich keine Aussage treffen. Diese Limitierung ergibt sich aus der Wahl der Studienpopulation, die keine malignen Befunde einschließt. Die Beeinflussung der Sensitivität ist allerdings Inhalt einiger bereits veröffentlichter Studien^{16, 35-40, 47, 48}.

Ebenfalls nicht untersucht wurde der Einfluss des Brustvolumens auf die Ergebnisse der CAD-Analyse. Dieser Parameter wird zwar von dem hier eingesetzten Quantra-Algorithmus zusammen mit dem Parenchymvolumen und der Brustdichte ausgegeben, ein Zusammenhang zwischen Brustvolumen und Spezifität erscheint aber unwahrscheinlich. Eine Untersuchung dieses Zusammenhangs könnte jedoch Inhalt zukünftiger Studien sein.

Die Reliabilität der beiden eingesetzten Computeralgorithmen wurde in dieser Arbeit nicht überprüft. So wurden die Mammographien nicht mehrmals an den CAD-Server gesendet, um die jeweiligen Ergebnisse zu vergleichen.

Alle Mammographien dieser Studie wurden mit demselben Mammographiegerät aufgenommen. Das war nötig, um ungewollte Effektgrößen zu minimieren und die Studie möglichst sauber durchzuführen. Alle Ergebnisse sind somit zunächst nur auf Aufnahmen des GE Senographe 2000D anzuwenden. Obwohl es wahrscheinlich erscheint, dass die Aussagen der Diskussion auch auf andere Systeme übertragbar sind, sollte dies in zukünftigen Arbeiten untersucht werden.

In dieser Studie kam nur ein CAD-Algorithmus zum Einsatz (R2 ImageChecker®, Version 9.3, Hologic, Bedford, MA, USA). Alle Ergebnisse lassen sich also nur auf

Untersuchungen mit diesem System anwenden. Ob und in welchem Maße die untersuchten Parameter andere CAD-Algorithmen beeinflussen, lässt sich aus den hier gewonnenen Daten nicht beantworten.

5.2.4 Verzerrung der Ergebnisse durch die CAD-Ausgabe

In dieser Studie sollte der Einfluss verschiedener Gewebeparameter auf die Ergebnisse eines CAD-Algorithmus untersucht werden. Allerdings entsprechen die ausgewerteten Marker nicht den ersten Ergebnissen des Algorithmus. Die Marker werden erstellt, indem suspekta Befunde in ein Ranking eingeteilt werden und anschließend eine Grenze definiert wird, ab der ein Befund als malignitätsverdächtig markiert wird³⁴. Dieses Sortieren der Ergebnisse in „markiert“ oder „nicht markiert“ entspricht einer groben Rundung der ursprünglichen Ergebnisse.

Eine weitere Manipulation des ersten Rankings findet statt, indem eine Obergrenze für die Anzahl der Marker pro Bild und Studie festgelegt wird. Wird diese Anzahl überschritten, werden die am niedrigsten bewerteten Befunde nicht markiert. Dieses Vorgehen ist in der vorliegenden Studie deshalb von besonderem Interesse, weil gezeigt wurde, dass parenchymreichere Mammographien ein erhöhtes Risiko für falsch-positive Marker aufweisen. Nun ist es denkbar, dass der Einfluss auf den Algorithmus tatsächlich größer ausfällt als die Ergebnisse zeigen, weil bei besonders parenchymreichen Mammographien ein Teil der suspekten Befunde nicht markiert wurde, obwohl ihr Ranking eine Markierung erfordern würde.

Die vorliegende Studie beschränkt sich auf die Untersuchung der endgültigen Marker, wodurch ein möglichst enger Praxisbezug gewährleistet ist. Zur genaueren Analyse und zur weiteren Optimierung des CAD-Algorithmus kann es aber sinnvoll sein, eine ähnliche Studie mit den ursprünglichen Ausgabedaten durchzuführen. Da diese dem Befunder aber nicht dargestellt werden, ist hierfür die Unterstützung des Herstellers erforderlich.

5.3 Konsequenzen der Arbeit

Während einige Studien zeigen, dass die Sensitivität von CAD-Untersuchungen von der Brustdicke beeinflusst wird^{35, 36, 38-40}, existieren zumindest für den R2 ImageChecker keine Daten, die einen Einfluss auf die Spezifität vermuten lassen. Die Daten dieser

Studie deuten vielmehr darauf hin, dass die Menge des Parenchymvolumens die Spezifität beeinflusst.

Dass hier Unterschiede vorliegen, erscheint plausibel: Je dichter sich das abgebildete Gewebe darstellt, desto eher werden maligne Läsionen der Brust maskiert, was bei steigender Brustdichte zu einer Verschlechterung der Sensitivität führt. Im Unterschied dazu führt ein vermehrtes Parenchymgewebe zu einer geringeren Spezifität, weil mehr Strukturen analysiert werden, die als maligne Läsionen fehlinterpretiert werden können. Allerdings wurden bisher keine Studien veröffentlicht, die den Einfluss des Parenchymvolumens auf die Sensitivität untersuchen. Es ist also möglich, dass ein solcher Einfluss existiert, bisher aber noch nicht nachgewiesen wurde. Dieser Frage könnte in zukünftigen Studien nachgegangen werden.

Nach aktueller Studienlage ist es gut möglich, dass die Sensitivität und die Spezifität des R2 ImageChecker von unterschiedlichen Faktoren beeinflusst werden. Dieses könnte in der Weiterentwicklung des Algorithmus genutzt werden. Schon jetzt ist es möglich, durch eine individuelle Festlegung des Operating Point auf das Verhältnis zwischen Sensitivität und Spezifität Einfluss zu nehmen. Dabei führt eine Verschiebung zu Gunsten des einen Parameters stets zu einer Verschlechterung des jeweils anderen Wertes. Diese Beziehung ist aber nicht linear. Ist zum Beispiel die Sensitivität besonders hoch, führt eine Verschiebung des Operating Point, die eine geringe Abnahme der Sensitivität zur Folge hat, zu einer verhältnismäßig großen Verbesserung der Spezifität.

Diesen Effekt könnte man bei Mammographien besonders großer, parenchymreicher Mammae nutzen, in denen die Sensitivität gleichbleibend hoch, die Spezifität aber erniedrigt ist, indem man den Operating Point geringfügig in Richtung einer besseren Spezifität kalibriert. Bei Mammographien kleiner Mammae, die eine besonders hohe Dichte aufweisen, könnte der Operating Point hingegen etwas zu Gunsten einer verbesserten Sensitivität verschoben werden, da diese durch die erhöhte Brustdichte eingeschränkt ist, während die Spezifität hiervon nicht beeinflusst wird. Ein solches Vorgehen könnte helfen, Sensitivität und Spezifität bei unterschiedlichen Gewebeeigenschaften möglichst konstant zu halten.

6. Zusammenfassung

1. Einleitung: Das Mammakarzinom ist die häufigste maligne Erkrankung der Frauen in Deutschland. Ein wichtiges Werkzeug zur Diagnostik ist die Mammographie. Computer-assistierte-Diagnose-Systeme (CAD-Systeme) können die Sensitivität der Mammographie erhöhen, wobei die hohe Anzahl falsch-positiver Marker einen Einsatz im Klinikalltag bisher behindert.

2. Literaturbericht und Fragestellung: Während einige veröffentlichte Studien Faktoren untersuchen, die die Sensitivität von CAD-Systemen beeinflussen, existieren bisher nur wenige Daten zur Beeinflussung der Spezifität. Zum Einfluss des Parenchymvolumens auf die Spezifität konnte keine Veröffentlichung gefunden werden.

In der vorliegenden Arbeit soll untersucht werden, ob bestimmte Parameter die Spezifität eines CAD-Algorithmus beeinflussen. Als erster Parameter wurde die Brustdichte ausgewählt, da ein Einfluss auf menschliche Befunder in der Mammographie bekannt ist. Als zweiter Parameter wurde das Parenchymvolumen untersucht, da vor allem das Drüsenparenchym und das umliegende Binde-Stütz-Gewebe mammographisch malignomähnliche Strukturen bilden können und ein Einfluss daher plausibel erscheint.

3. Material und Methoden: In diese Studie wurden Mammographieuntersuchungen eingeschlossen, die im Zeitraum zwischen dem 1.6.2002 und dem 31.8.2006 erstellt wurden und keinerlei pathologische Veränderung zeigten. Ein unauffälliges Follow-up über mindestens 22 Monate war erforderlich.

Die Standardansichten (craniocaudal und mediolateral-oblique) einer Seite jeder Untersuchung wurden von einem CAD-Algorithmus mit drei Sensitivitäts-Einstellungen (spezifisch, mittel, sensibel) analysiert (R2 ImageChecker®, Version 9.3, Hologic, Bedford, MA, USA) und die Anzahl der falsch-positiven Marker ermittelt. Ein weiterer Algorithmus analysierte die Gewebeeigenschaften der untersuchten Brust (R2 Quantra®, Version 1.3, Hologic, Bedford, MA, USA) und ermittelte das Brustvolumen, das Parenchymvolumen sowie die prozentuale Brustdichte. Zur Analyse der Daten wurden verschiedene statistische Untersuchungen (Kategorisierung und Analyse durch Mann-Whitney-U-Test, Korrelationsanalyse, Regressionsanalyse) durchgeführt.

4. Ergebnisse: 222 Mammographien konnten in die Untersuchung eingeschlossen werden. Je Aufnahme erstellte der CAD-Algorithmus im Mittel 0,42 (spezifisch), 0,52 (mittel), bzw. 0,69 (sensibel) falsch-positive Marker zu Herdbefunden, sowie 0,24 (spezifisch), 0,43 (mittel), bzw. 0,56 (sensibel) falsch-positive Marker zu Mikrokalz.

In den Gruppen mit geringem, mittlerem und großem Parenchymvolumen war die durchschnittliche Anzahl falsch-positiver Herdmarker 0,35, 0,41 und 0,50 (spezifisch); 0,40, 0,45 und 0,70 (mittel); bzw. 0,58, 0,65 und 0,89 (sensibel). Einen statistisch signifikanten Unterschied der Anzahl der Herdmarker ergab der Vergleich der Gruppen mit geringem bzw. großem Parenchymvolumen bei den CAD-Einstellungen ‚mittel‘ und ‚sensibel‘ ($p < 0,05$). Ein Unterschied in der Anzahl der Marker für Mikrokalz konnte nicht festgestellt werden. Wurden die Mammographien nach der Brustdichte kategorisiert, war kein signifikanter Unterschied der Gruppen nachweisbar.

Die Regressionsanalyse zeigte bei einer Zunahme des Parenchymvolumens um 100ml ein relatives Risiko für Herd-Marker von 1,43 (spezifisch, $p < 0,05$), 1,63 (mittel, $p < 0,001$), bzw. 1,50 (sensibel, $p < 0,01$). Ein Einfluss auf das relative Risiko für Kalz-Marker war nicht festzustellen. Die Regressionsanalyse zeigte keinen signifikanten Einfluss der Brustdichte auf das relative Risiko.

5. Diskussion: Verschiedene Veröffentlichungen der letzten Jahre zeigen auf, dass die Sensitivität von CAD-Untersuchungen von der Brustdichte beeinflusst wird, während diese Studie nahelegt, dass deren Spezifität von der Menge des Parenchymgewebes abhängig ist. Eine Anpassung des Operating Points unter Berücksichtigung von Brustdichte und Drüsenparenchymvolumen kann so dazu beitragen, die Sensitivität und Spezifität von CAD-Systemen zu optimieren.

Literaturverzeichnis

1. Kaufmann M, Pfeleiderer A. Maligne Tumoren der Mamma. In: Breckwoldt M, Kaufmann M, Pfeleiderer A, eds. Gynäkologie und Geburtshilfe 5. ed. Stuttgart: Georg Thieme Verlag 2008:241-260.
2. Hortobagyi GN, de la Garza Salazar J, Pritchard K, et al. The global breast cancer burden: variations in epidemiology and survival. *Clin Breast Cancer* 2005;6:391-401.
3. Statistisches Bundesamt. Gesundheit - Todesursachen in Deutschland 2010. Wiesbaden: Statistisches Bundesamt; 2011.
4. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: An international comparison of screening, incidence, survival and mortality. *Cancer Epidemiol* 2012;36:237-248.
5. Weyerstahl T, Genz T. Mammakarzinom. In: Stauber M, Weyerstahl T, eds. Gynäkologie und Geburtshilfe. 3. ed. Stuttgart: Georg Thieme Verlag; 2007:371-396.
6. Kavanagh AM, Cawson J, Byrnes GB, et al. Hormone replacement therapy, percent mammographic density, and sensitivity of mammography. *Cancer Epidemiol Biomarkers Prev* 2005;14:1060-1064.
7. Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. *AJR Am J Roentgenol* 2003;180:1675-1679.
8. Robertson C, Arcot Ragupathy SK, Boachie C, et al. The clinical effectiveness and cost-effectiveness of different surveillance mammography regimens after the treatment for primary breast cancer: systematic reviews registry database analyses and economic evaluation. *Health Technol Assess* 2011;15:v-vi, 1-322.
9. Klütsch D, Kotsianos D. Mammographie. In: Reiser M, Kuhn F-P, Debus J, eds. *Duale Reihe Radiologie*. Stuttgart: Georg Thieme Verlag; 2004:544-547.
10. Schulz-Wendtland R, Hermann KP, Wacker T, Bautz W. Current situation and future perspectives of digital mammography. *Radiologe* 2008;48:324-334.
11. Lewin JM, D'Orsi CJ, Hendrick RE, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. *AJR Am J Roentgenol* 2002;179:671-677.
12. Teifke A. Mamma. In: Bücheler E, Lackner K, Thelen M, eds. *Einführung in die Radiologie*. 11. ed. Stuttgart: Georg Thieme Verlag; 2006:612-626.

13. Kaufmann M, Minckwitz G. Mammakarzinom. In: Kaufmann M, Costa SD, Scharl A, eds. Die Gynäkologie. 2. ed. Heidelberg: Springer Medizin Verlag; 2006:477-533.
14. Schulz-Wendtland R, Becker N, Bock K, Anders K, Bautz W. Mammography screening. *Radiologe* 2007;47:359-369; quiz 370.
15. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 2003;138:168-175.
16. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology* 1992;184:613-617.
17. Harvey JA, Fajardo LL, Innis CA. Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation. 1993 ARRS President's Award. *AJR Am J Roentgenol* 1993;161:1167-1172.
18. American College of Radiology. Breast Imaging Reporting and Data System Atlas. Reston, VA: American College of Radiology; 2003.
19. Diebold T, Vogl T. Bildgebende und Interventionelle Diagnostik der Brustdrüse mittels Röntgen-Mammographie, Mamma-Sonographie und MR-Mammographie. In: Vogl T, Reith W, Rummeny E, eds. Diagnostische und Interventionelle Radiologie. 1. ed. Berlin: Springer Verlag; 2011:613-678.
20. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol* 2000;174:1769-1777.
21. Ciatto S, Houssami N, Apruzzese A, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. *Breast* 2005;14:269-275.
22. Ooms EA, Zonderland HM, Eijkemans MJ, et al. Mammography: interobserver variability in breast density assessment. *Breast* 2007;16:568-576.
23. Bernardi D, Pellegrini M, Di Michele S, et al. Interobserver agreement in breast radiological density attribution according to BI-RADS quantitative classification. *Radiol Med* 2012.
24. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *AJR Am J Roentgenol* 2008;190:854-859.

25. Gilbert FJ, Astley SM, Gillan MG, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008;359:1675-1684.
26. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007;356:1399-1409.
27. Gilbert FJ, Astley SM, McGee MA, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. *Radiology* 2006;241:47-53.
28. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. *AJR Am J Roentgenol* 2006;187:20-28.
29. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: interpretation with computer-aided detection--prospective evaluation. *Radiology* 2006;239:375-383.
30. Gur D, Sumkin JH, Rockette HE, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *J Natl Cancer Inst* 2004;96:185-190.
31. Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. *AJR Am J Roentgenol* 2005;185:944-950.
32. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001;220:781-786.
33. Brem RF, Rapelyea JA, Zisman G, Hoffmeister JW, Desimio MP. Evaluation of breast cancer with a computer-aided detection system by mammographic appearance and histopathology. *Cancer* 2005;104:931-935.
34. Ziegler R. Understanding R2 ImageChecker CAD 9.3. *Hologic Manual* 2009;1223:1-39.
35. Ho WT, Lam PW. Clinical performance of computer-assisted detection (CAD system in detecting carcinoma in breasts of different densities. *Clin Radiol* 2003;58:133-136.
36. Li L, Wu Z, Salem A, et al. Computerized analysis of tissue density effect on missed cancer detection in digital mammography. *Comput Med Imaging Graph* 2006;30:291-297.

37. Obenauer S, Sohns C, Werner C, Grabbe E. Impact of breast density on computer-aided detection in full-field digital mammography. *J Digit Imaging* 2006;19:258-263.
38. Brem RF, Hoffmeister JW, Rapelyea JA, et al. Impact of breast density on computer-aided detection for breast cancer. *AJR Am J Roentgenol* 2005;184:439-444.
39. Malich A, Fischer DR, Facius M, et al. Effect of breast density on computer aided detection. *J Digit Imaging* 2005;18:227-233.
40. Romero Castellano C, Varela Nunez C, Cuenca Boy R, Almenar Gil A, Pinto Varela JM, Botella Lopez M. [Impact of mammographic breast density on computer-assisted detection (CAD) in a breast imaging department]. *Radiologia* 2011;53:456-461.
41. Ziegler R. Understanding R2 Quantra 1.3. *Hologic Manual* 2009;1224:1-18.
42. Johns PC, Yaffe MJ. X-ray characterisation of normal and neoplastic breast tissues. *Phys Med Biol* 1987;32:675-695.
43. Boone JM, Fewell TR, Jennings RJ. Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography. *Med Phys* 1997;24:1863-1874.
44. Grebe S, Diekmann F, Bick U, Paepke S, Winzer KJ, Hamm B. [Initial clinical experiences with digital full-field mammography]. *Zentralbl Gynakol* 2000;122:589-594.
45. Young KC, Oduko JM, Bosmans H, Nijs K, Martinez L. Optimal beam quality selection in digital mammography. *Br J Radiol* 2006;79:981-990.
46. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst* 1998;90:1801-1809.
47. Brem RF, Baum J, Lechner M, et al. Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial. *AJR Am J Roentgenol* 2003;181:687-693.
48. Noble M, Bruening W, Uhl S, Schoelles K. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. *Arch Gynecol Obstet* 2009;279:881-890.

Abbildungsverzeichnis

- Abbildung 1 Mammographie in mediolateral-obliquem Strahlengang mit suspekter Verdichtung rechts
- Abbildung 2 Ausgabe des CAD-Algorithmus mit Markern für Herdbefunde und Mikrokalzifikationen
- Abbildung 3 Tabelle mit den Ergebnissen des Quantra-Algorithmus
- Abbildung 4 Alter der Patientinnen bei der ersten Untersuchung
- Abbildung 5 Brustvolumen (einseitig) in Milliliter
- Abbildung 6 Drüsenparenchymvolumen je Brust in Milliliter
- Abbildung 7 Brustdichte in Prozent
- Abbildung 8 Summe der falsch-positiven Marker in CC und MLO je Einstellung des Operating Point (OP)
- Abbildung 9 Mittelwerte der falsch-positiven Kalk-Marker pro Studie (CC und MLO) je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Brustdichte.
- Abbildung 10 Mittelwerte der falsch-positiven Herd-Marker pro Studie (CC und MLO) je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Brustdichte.
- Abbildung 11 Mittelwerte der falsch-positiven Kalk-Marker pro Studie (CC und MLO) je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Parenchymvolumen.
- Abbildung 12 Mittelwerte der falsch-positiven Herd-Marker pro Studie (CC und MLO) je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Parenchymvolumen. Signifikante Unterschiede sind markiert (*).
- Abbildung 13 Mittelwert (MW) der Brustdichte in Prozent je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Anzahl der falsch-positiven Kalk-Marker pro Studie (CC und MLO).
- Abbildung 14 Mittelwert (MW) der Brustdichte in Prozent je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Anzahl der falsch-positiven Herd-Marker pro Studie (CC und MLO).
- Abbildung 15 Mittelwert (MW) des Parenchymvolumens in ml je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Anzahl der falsch-positiven Kalk-Marker pro Studie (CC und MLO).
- Abbildung 16 Mittelwert (MW) des Parenchymvolumens in ml je Einstellung des Operating Point (OP). Aufnahmen kategorisiert nach Anzahl der falsch-positiven Kalk-Marker pro Studie (CC und MLO). Signifikante Unterschiede sind markiert (*).
- Abbildung 17 Anzahl der Kalk-Marker über der Brustdichte in Prozent, Operating Point 0 (OP 0)
- Abbildung 18 Anzahl der Herd-Marker über der Brustdichte in Prozent, Operating Point 0 (OP 0)
- Abbildung 19 Anzahl der Kalk-Marker über der Brustdichte in Prozent, Operating Point 1 (OP 1)
- Abbildung 20 Anzahl der Herd-Marker über der Brustdichte in Prozent, Operating Point 1 (OP 1)

- Abbildung 21 Anzahl der Kalk-Marker über der Brustdichte in Prozent, Operating Point 2 (OP 2)
- Abbildung 22 Anzahl der Herd-Marker über der Brustdichte in Prozent, Operating Point 2 (OP 2)
- Abbildung 23 Anzahl der Kalk-Marker über dem Parenchymvolumen in Prozent, Operating Point 0 (OP 0)
- Abbildung 24 Anzahl der Herd-Marker über dem Parenchymvolumen in Prozent, Operating Point 0 (OP 0)
- Abbildung 25 Anzahl der Kalk-Marker über dem Parenchymvolumen in Prozent, Operating Point 1 (OP 1)
- Abbildung 26 Anzahl der Herd-Marker über dem Parenchymvolumen in Prozent, Operating Point 1 (OP 1)
- Abbildung 27 Anzahl der Kalk-Marker über dem Parenchymvolumen in Prozent, Operating Point 2 (OP 2)
- Abbildung 28 Anzahl der Herd-Marker über dem Parenchymvolumen in Prozent, Operating Point 2 (OP 2)

Tabellenverzeichnis

- Tabelle 1 BI-RADS Klassifikation des American College of Radiology (ACR)
- Tabelle 2 Einteilung der Brustdichte nach Empfehlung des American College of Radiology (ACR)
- Tabelle 3 Zusammenfassung einiger Studien zum Einsatz von CAD in der Mammographiebefundung
- Tabelle 4 Übersicht der Einstellungsmöglichkeiten des CAD-Algorithmus
- Tabelle 5 Zusammenfassung der Marker zu Herdbefunden bei den drei Einstellungen des Operating Point mit Mittelwert (MW) und Standardabweichung (s)
- Tabelle 6 Zusammenfassung der Marker zu Mikrokalk bei den drei Einstellungen des Operating Point mit Mittelwert (MW) und Standardabweichung (s)
- Tabelle 7 Anzahl der Untersuchungen, bei denen die Grenzwerte der Marker zu Herdbefunden pro Bild erreicht wurden; je Einstellung des Operating Point (OP)
- Tabelle 8 Spezifität des CAD-Algorithmus bei den drei Einstellungen des Operating Point (OP)
- Tabelle 9 Kategorisierung der untersuchten Studien (CC und MLO) nach der Dichte der untersuchten Brust. Bei den drei Einstellungen des Operating Point wurde in jeder Kategorie die Anzahl der falsch-positiven Marker gezählt und der Mittelwert (MW) pro Studie berechnet.
- Tabelle 10 Kategorisierung der untersuchten Studien (CC und MLO) nach dem Parenchymvolumen der untersuchten Brust. Bei den drei Einstellungen des Operating Point wurde in jeder Kategorie die Anzahl der falsch-positiven Marker gezählt und der Mittelwert (MW) pro Studie berechnet.
- Tabelle 11 Mann-Whitney-U-Test, Signifikanz des Anstiegs der Trefferwahrscheinlichkeit in höheren Kategorien bei Kategorisierung nach Dichte
- Tabelle 12 Mann-Whitney-U-Test, Signifikanz des Anstiegs der Trefferwahrscheinlichkeit in höheren Kategorien bei Kategorisierung nach Parenchymvolumen
- Tabelle 13 Kategorisierung der untersuchten Studien (CC und MLO) nach der Anzahl der Kalk-Marker. Bei den drei Einstellungen des Operating Point sind jeweils die Anzahl der Fälle angegeben sowie Mittelwert (MW) und Standardabweichung (s) der Dichte und des Parenchymvolumens.
- Tabelle 14 Kategorisierung der untersuchten Studien (CC und MLO) nach der Anzahl der Herd-Marker. Bei den drei Einstellungen des Operating Point sind jeweils die Anzahl der Fälle angegeben sowie Mittelwert (MW) und Standardabweichung (s) der Dichte und des Parenchymvolumens.
- Tabelle 15 Mann-Whitney-U-Test, Signifikanz des Anstiegs des Mittelwertes der Testvariable bei größerer Anzahl falsch-positiver Kalk-Marker
- Tabelle 16 Mann-Whitney-U-Test, Signifikanz des Anstiegs des Mittelwertes der Testvariable bei größerer Anzahl falsch-positiver Herd-Marker
- Tabelle 17 Zusammenstellung der Parameterkombinationen zur Korrelations- und Regressionsanalyse

Tabelle 18 Zusammenfassung der Ergebnisse der Korrelationsanalyse nach Pearson. Angegeben sind der Korrelationskoeffizient (r), das Bestimmtheitsmaß (r^2), und der p-Wert (p). Aufteilung nach Operating Point (OP).

Tabelle 19 Zusammenfassung der Ergebnisse der Regressionsanalyse nach Poisson. Angegeben sind der Regressionskoeffizient (β), das relative Risiko (RR) und der p-Wert (p). Aufteilung nach Operating Point (OP).

Abkürzungsverzeichnis

ACR	American College of Radiology
BI-RADS	breast imaging reporting and data system
CAD	computerassistierte Detektion
CAD	computerassistierte Diagnose
CC	craniocaudal
CDR	cancer detection rate
CsI	Cäsiumjodid
DICOM	digital imaging and communications in medicine
DR	double reading
i.v.	intravenös
kV	Kilovolt
LCC	links craniocaudal
LMLO	links mediolateral-oblique
mGy	Milligray
MLO	mediolateral-oblique
Mo	Molybdän
MW	Mittelwert
OP	Operating Point
PACS	picture archiving and communication system
PPV	positive predictive value
RCC	rechts craniocaudal
Rh	Rhodium
RMLO	rechts mediolateral-oblique
RR	recall rate
RR	relatives Risiko
SC	secondary capture
Sens.	Sensitivität
Spez.	Spezifität
SR	single reading

Erklärung

Ich, Raphael Witlandt, erkläre, dass ich die vorgelegte Dissertation mit dem Thema: „Untersuchung des Einflusses der Brustdichte und des Parenchymvolumens auf die Ergebnisse von CAD-Untersuchungen in der digitalen Mammographie“ selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, ohne die (unzulässige) Hilfe Dritter verfasst und auch in Teilen keine Kopien anderer Arbeiten dargestellt habe.

Ort, Datum

Unterschrift

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Danksagung

Beim Verfassen dieser Arbeit haben mich einige Menschen unterstützt und begleitet, ohne deren Hilfe ich keinesfalls zu dem vorliegenden Ergebnis gekommen wäre.

Zunächst ist mein Doktorvater Herr PD Dr. Felix Diekmann zu nennen. Es ist üblich, sich an dieser Stelle für die Überlassung des Themas zu bedanken und natürlich möchte auch ich dies tun. Mein Dank geht aber deutlich weiter. Seine Art, mich in den letzten Jahren zu begleiten war stets aufmerksam und engagiert, dabei aber immer äußerst freundlich und motivierend. Für mich hätte es keinen besseren Doktorvater geben können.

Dank gilt auch meinem Betreuer Dr. Florian Engelken. Unzählige Mails und Treffen, ausführliche Korrekturen und Anmerkungen zeugen von seinem Engagement für meine Arbeit. Florian hat mich behutsam in das komplexe Feld des wissenschaftlichen Arbeitens eingearbeitet und stand dabei für Fragen jederzeit zur Verfügung. Viele Klippen und Fallstricke, die auf den unbedarften Doktoranden lauern, konnte ich so weiträumig umschiffen oder zumindest letztendlich auflösen.

Ich hatte das Glück, in meinen Eltern, Ulla und Michael Bremme, und meiner Patentante, Maria Meesmann, aufmerksame Korrektoren und Unterstützer zu haben. Ich danke herzlich für die unzähligen hilfreichen Hinweise in etlichen Korrekturen zu den einzelnen Kapiteln. Auch wenn ich bemüht war, einen ordentlichen Text zu verfassen, so tauchten doch immer wieder viele kleine Dinge auf, die ich noch besser machen konnte. Neben der fachlichen Hilfe bin ich besonders für die seelische Unterstützung dankbar, die mich durch die vielen kleinen und größeren Krisen mit der Doktorarbeit getragen hat.

Zu guter Letzt danke ich meiner Frau Veronika Witlandt, die mir zur Seite steht und den zusätzlichen Stress eines solch aufwändigen Projekts immer wieder aufgefangen hat.