SYSTEMATISCHE REVIEW: OPLOSSEN VAN HET SEGMENTATIEPROBLEEM IN MRI-MAMMOGRAFIE

TOEPASSING VAN MACHINE LEARNING EN DEEP LEARNING TEN BEHOEVE VAN AUTOMATISCHE BORST- EN TUMORWEEFSEL SEGMENTATIE IN MRI-MAMMOGRAFIE ONDERZOEKEN







SYSTEMATISCHE REVIEW: OPLOSSEN VAN HET SEG-MENTATIEPROBLEEM IN MRI-MAMMOGRAFIE

TOEPASSING VAN MACHINE LEARNING EN DEEP LEARNING TEN BEHOEVE VAN AUTOMATISCHE BORST- EN TU-MORWEEFSEL SEGMENTATIE IN MRI-MAMMOGRAFIE ONDERZOEKEN

Groningen, 8 juni, 2020

Auteurs:	Y.S. Kunnen (357814)
E-mail:	Y.s.kunnen@st.hanze.nl
	M. Chen (353028)
E-mail:	M.chen@st.hanze.nl
Opdrachtgever:	X. Jing, MSc Universitair Medisch Centrum Groningen Afdeling Data Science Center in Health
E-mail:	x.jing@umcg.nl
Opleiding:	Hanzehogeschool Groningen Academie voor Gezondheidsstudies Medisch Beeldvormende en Radiotherapeutische Technieken (MBRT)
Onderzoeksbegeleider:	H. Erenstein, MSc Hanzehogeschool Groningen Docent MBRT
E-mail:	h.erenstein@pl.hanze.nl
Telefoon:	050-5953810
Co-beoordelaar:	R. Visser, MSc, PhD Hanzehogeschool Groningen Docent MBRT
E-mail:	ru.visser@pl.hanze.nl
Telefoon:	050-5957439

Omslag foto genomen uit Piantidosi et al. (1).

VOORWOORD

Dit praktijkgerichte onderzoek (PO) is geschreven in opdracht van de opleiding Medisch Beeldvormende en Radiotherapeutische Technieken en het Data Science Center in Health. Het onderwerp van deze scriptie is het oplossen van het segmentatieprobleem in Magnetic Resonance Imaging (MRI)-mammografie. Wij zijn beide gefascineerd door de innovaties die plaatsvinden in de medische beeldvorming, waaronder kunstmatige intelligentie. Tijdens het schrijven van deze scriptie hebben wij waardevolle kennis over toepassingsmogelijkheden van kunstmatige intelligentie opgedaan. Wij hebben met genoegen aan dit praktijkgericht onderzoek gewerkt.

Graag maken wij van deze gelegenheid gebruik om de volgende mensen te bedanken voor de medewerking tijdens dit praktijkgerichte onderzoek:

- X. Jing, MSc (Opdrachtgever Data Science Center in Health)
 Voor de begeleiding tijdens het opstellen van de onderzoeksopzet en de onderzoeksverslaglegging.
 Wij hebben veel geleerd over internationale samenwerking en zijn dankbaar voor de geleverde deskundige feedback.
- L. Cornelissen, MSc, PhD (Postdoctoraat Onderzoeker in Universitair Medisch Centrum Groningen)
 Voor de begeleiding tijdens het opstellen van de onderzoeksopzet en leveren van deskundige feedback.
 Ook voor zijn beschikbaarheid om sporadische vragen te beantwoorden.
- Medewerkers van het Data Science Centre in Health (DASH)
 Voor het welkomen van ons met open armen in het DASH-team. Wij hebben met genoegen op locatie in het Universitair Medisch Centrum Groningen gewerkt, deelgenomen aan werkoverleggen en vonden de steun vanuit het DASH-team zeer fijn.
- H. Erenstein, MSc (MBRT-docent) en P. Plasman, MSc (MBRT-docent)
 Voor de gestructureerde begeleiding die zij tijdens werkcolleges hebben gegeven. Ook voor de waardevolle feedback die zij tijdens werkcolleges hebben gegeven op zowel de onderzoeksopzet als scriptieonderdelen.

Y.S. Kunnen, M. Chen Groningen, 8 juni, 2020

ABSTRACT (NL)

Inleiding: In dit onderzoek zijn recente ontwikkelingen in Machine Learning (ML) en Deep Learning (DL) met betrekking tot borst-, fibroglandulair (FG)- en tumorweefselsegmentatie in MRI-mammografie in kaart gebracht, de prestaties van deze segmentatiesystemen zijn vergeleken en als laatste worden toekomst toepassingsmogelijkheden van Machine Learning en Deep Learning in MRI-mammografie bediscussieerd.

Materiaal en methode: Een systematische zoekopdracht is uitgevoerd in Pubmed, Arxiv en IEEE Explore om artikelen te identificeren met betrekking tot ML- of DL-segmentatie van borst-, fibroglandulair FG- en tumor-weefsel in MRI-mammografie. Na identificatie en screening van artikelen is kwaliteitsbeoordeling en data-ex-tractie uitgevoerd. Per artikel zijn de auteur, tijdschrift, doel, dataset, methode, resultaten (ofwel prestatie-metrics) en conclusie in een evidence tabel opgenomen.

Resultaten: In totaal zijn 21 artikelen geïncludeerd (borstsegmentatie=7, FG-segmentatie=1, borst+FG-segmentatie=4 en tumorsegmentatie=9). Verschillende prestatie-metrics zijn in de artikelen benut: Dice Similarity Coefficient (DICE) (100%), Jaccard Index (JAC) (14%), sensitiviteit (47,7%) en specificiteit (29%). De gemiddelde DICE voor DL-borstweefselsegmentatie was 0,94 (range, 0,77-0,982), de gemiddelde DICE voor ML-borstsegmentatie was 0,90 (range, 0,846-0,94). De gemiddelde DICE voor DL FG-weefselsegmentatie was 0,86 (range, 0,813-0,925), de gemiddelde DICE voor ML FG-weefselsegmentatie was 0,78 (range, 0,619-0,917). Tenslotte was de gemiddelde DICE voor DL-tumorsegmentatie 0.735 (range, 0,59-0,836), de gemiddelde DICE voor ML-tumorsegmentatie was 0,698 (range, 0,608-0,786).

Conclusie: ML- en DL-systemen zijn in staat om borst-, FG- of tumorweefsel met precisie te segmenteren in MRImammografie onderzoeken, echter is de klinische toepassing nog niet in zicht. Nader extensief onderzoek dient verricht te worden naar validiteit en reproduceerbaarheid van ML- en/of DL-systemen.

ABSTRACT (EN)

Introduction: The purpose of this study was to review recent developments in Machine Learning (ML) and Deep Learning (DL) systems concerning breast-, fibroglandulair (FG) and tumor segmentation in MRI-mammography, to compare performances of the respective tasks per system and future applications of both systems concerning segmentation in MRI-mammography were reviewed.

Method: A systematic search for literature was performed in Pubmed, Arxiv en IEEE Explore to identify articles concerning ML and DL segmentation of breast, FG or tumor segmentation in MRI-mammography. After identification and screening, quality analysis and data extraction were performed. For each article, the surname of the first author, journal, aim, dataset, method, results (metrics) and conclusion were recorded in an evidence table.

Results: In total, 21 articles (breast segmentation=7, FG-segmentation=1, breast+FG-segmentation=4, tumor segmentation=9) were included. Several measures of performance were recorded: Dice Similarity Coefficient (DICE) (100%), Jaccard Index (JAC) (14%), sensitivity (47,7%) and specificity (29%). The mean DICE for DL breast segmentation was 0,94 (range, 0,77-0,982), the mean DICE for ML breast segmentation was 0,90 (range, 0,846-0,94). The mean DICE for DL FG-segmentation was 0,86 (range, 0,813-0,925), the mean DICE for ML FG-segmentation was 0,78 (range, 0,619-0,917). Lastly, the mean DICE for DL tumor segmentation was 0,786), the mean DICE for ML tumor segmentation was 0,698 (range, 0,608-0,786).

Conclusion: ML and DL models are capable of performing breast, FG and tumor segmentations in MRI-mammography, it is however not ready for clinical application. Further research must be performed on the validity and reproducibility of ML and DL segmentation models.

INHOUDSOPGAVE

INLEIDING	7
THEORETISCH KADER	
MRI-MAMMOGRAFIE KUNSTMATIGE INTELLIGENTIE AFREEL DINGSSEGMENTATIE	
METHODE	
ONDERZOEKSOPZET	
SELECTIE VAN ARTIKELEN VOOR REVIEW	
ZOEKSTRATEGIEEN	
BEOORDELING INHOUD & KWALITEIT	13
DATA-ANALYSE	14
RESULTATEN	15
UITKOMSTEN SELECTIEPROCEDURE	15
TOEPASBAARHEID VAN ARTIKELEN	
BORSTWEEFSELSEGMENTATIE	
FG-WEEFSELSEGMENTATIE	19
TUMORSEGMENTATIE	20
DISCUSSIE	21
BORSTWEEFSEL SEGMENTATIE	
FG-WEEFSEL SEGMENTATIE	
TUMORSEGMENTATIE	22
SYSTEMATISCHE REVIEW	23
CONCLUSIE	25
LITERATUURLIJST	26
BIJLAGE 1: DATABANK ZOEKSTRINGS	29
BIJLAGE 2: CHECKLIST TOEPASBAARHEID ARTIKELEN	30
BIJLAGE 3: INGEVULDE CHECKLIST	31
BIJLAGE 4: EVIDENCE TABLE	32
4.1 EVIDENCE TABLE BORSTWEEFSEL SEGMENTATIE	
4.2 EVIDENCE TABLE FG-WEEFSELSEGMENTATIE	
4.3 EVIDENCE TABLE TUMOR SEGMENTATIE	40

INLEIDING

Borstkanker is de op één na meest voorkomende kankersoort onder de gehele bevolking in Nederland. In 2019 werden naar schatting 14.994 mensen gediagnosticeerd met een invasieve vorm van borstkanker (14.862 vrouwen en 132 mannen)(2), waarbij geldt dat bij zowel vrouwen als mannen de incidentie stijgt met leeftijd (2). Onder vrouwen geldt dat borstkanker de meest voorkomende vorm van kanker is (2). Het stadium waarin borstkanker gediagnosticeerd wordt is bepalend voor de kans op overleving en mogelijke terugkeer van borstkanker (3). Röntgenmammografie wordt beschouwd als de gouden standaard screeningsmethode om borstkanker te detecteren (4). Bij vrouwen die een verhoogde kans hebben op borstkankervorming wordt doorgaans aanvullend Magnetic Resonance Imaging (MRI) onderzoek of Echografisch (US) onderzoek verricht (5).

Om een effectief behandeltraject op te stellen is het essentieel dat radiologen juiste diagnoses kunnen stellen bij uitgevoerde radiologische onderzoeken, zoals: MRI-scans, US-scans en röntgenfoto's. Afgelopen jaren zijn vele toepassingen ontwikkeld en geïmplementeerd om radiologen te ondersteunen bij het beoordelen van eerdergenoemde onderzoeken. Een voorbeeld hiervan in de radiologie is: op kunstmatige intelligentie (AI)(6) gebaseerde Computer Aided Detection (CAD)-systemen (7)(8). Binnen de radiologie zijn met name de AI-toepassingen Machine Learning (ML) en Deep Learning (DL) in opkomst. De opkomst van de voornoemde systemen biedt een grote ontwikkelingsmogelijkheid om de werkroutine van radiologen te verbeteren (9). Foutieve diagnoses ten gevolge van vermoeidheid, subjectiviteit van radiologen en het verschil tussen beoordelaars kunnen mogelijk verminderd worden door middel van AI-systemen (10). Zowel CAD- als AI-systemen kunnen details opmerken die radiologen mogelijk missen, tevens kunnen zij zelfstandig onderzoeksafbeeldingen interpreteren en beoordelen. CAD- en AI-systemen worden doorgaans als 'tweede opinie of 'hulp bij diagnosestelling' benut door radiologen (11). Volgens Shiraishi et al. (7) en Thrall et al. (6) bieden zowel CAD- als AI-systemen mogelijkheden om accuratere diagnoses te stellen ten opzichte van inzet van enkel radiologen. Hiernaast biedt de opkomst van CAD- en AI-systemen veel mogelijkheden om tijdsintensieve taken te automatiseren. Dit dient volgens Jha et al. per vakgebied bekeken te worden (10).

Eén van de taken die binnen het vakgebied MRI-mammografie geautomatiseerd kan worden door middel van Machine Learning en/of Deep Learning is het segmenteren van MRI-onderzoeksafbeeldingen (9). Segmenteren houdt in dat aanwezige anatomie wordt ingedeeld in categorieën (12), bij MRI-mammografieën zijn dit doorgaans borst- en tumorweefsel (13). De gesegmenteerde afbeeldingen kunnen benut worden voor het opstellen van behandelplannen, voor beoordeling tijdens follow-ups of gebruikt worden voor verdere analyse (12)(14). Handmatig segmenteren van MRI-scans is een tijdsintensieve klus wegens de hoeveelheid beschikbare data voor segmentatie. Hiernaast is handmatig segmenteren gevoelig voor intra- en interbeoordelaars verschillen (14). Afgelopen jaren zijn verschillende ML- en DL-toepassingen ontwikkeld om, na training met geprepareerde MRIdatasets, zelfstandig en automatisch MRI-scans te kunnen segmenteren (8)(15). Voorbeelden hiervan zijn toepassingen bij MRI-hersenscans (14). Vele toepassingen zijn ontwikkeld om segmentatie mogelijk te maken in MRI-mammografie onderzoeken. Echter, tot onze beste kennis, is geen vergelijkend literatuuronderzoek verricht naar de bestaande ML- en DL-toepassingen binnen de MRI-mammografie ten aanzien van segmentatie van borst- en tumorweefsel. Tevens is weinig bekend over de prestaties van verschillende ML- en DL-systemen ten aanzien van segmentatie van borst- en tumorweefsel in MRI-mammografie series. Onder prestaties wordt onder andere sensitiviteit, specificiteit en Dice Similarity Coefficient ten aanzien van segmentatie van borst- en tumorweefsel verstaan.

Dit onderzoek heeft drie doelen: het in kaart brengen van bestaande ML- en DL-systemen ten behoeve van segmentatie van borst- en tumorweefsel in MRI-mammografie onderzoeken; het vergelijken van de prestaties van de bestaande ML- en DL-systemen en als laatst wordt gespeculeerd over toekomstige toepassingsmogelijkheden van ML- en DL-segmentatiesystemen in MRI-mammografie onderzoeken. Aan de hand van de onderzoeksresultaten wordt een voorspelling gemaakt over in hoeverre ML- en DL-systemen in staat zijn om respectievelijk borst- of tumorweefsel zelfstandig te segmenteren in MRI-mammografie onderzoeken.

THEORETISCH KADER

MRI-MAMMOGRAFIE

ANATOMIE

Gemiddeld zitten tussen de 10 en 25 kegelvormige klierkwabben in een vrouwelijke borst. Deze kunnen zich naast of achter de papilla mammae (tepel) bevinden (16). Aan een kwab zitten veel lobuli glandulae mammariae (melkklieren) vast, in deze klieren wordt melk geproduceerd (lactatie). Vervolgens wordt in de ducti lacteri (afvoergangen) de melk via de papilla mammae naar buiten vervoerd (16). Aan de afvoergangen zitten melk reservoirs vast, hierin wordt de geproduceerde melk bewaard. Door ligamenten suspensoria mammaeriae (banden van Cooper) worden de klieren en afvoerwegen ondersteund en krijgt de borst haar vorm (16). Een opsomming van het klierweefsel in de mammae wordt ook wel fibroglandulair (FG) borstweefsel genoemd. In de borst bedekt vet de klieren. De grootte van borsten wordt bepaald door het volume vetweefsel (16). Nabij de mammae vinden zich rondom de thoraxwand ook andere weefsels en structuren. Onder de thoraxwand bevinden de pectoralis major en pectoralis minor, sternum (borstbeen), costae (ribben), subcutaan vet, fibroglandulair en fasciae (bindweefsel)(16).

PATHOLOGIE (MAMMACARCINOMEN)

Onder vrouwen komt borstkanker het vaakst voor, dit worden ook mammacarcinomen genoemd. Verschillende soorten van mammacarcinomen kunnen zich manifesteren in de borsten. Naast verschillende soorten carcinomen bestaan ook verschillende gradaties van borstkanker (17). Borstkanker wordt verdeeld in invasieve en noninvasieve typen. Non-invasieve borstkanker wordt genoemd als carcinoma in situ. Alleen in de ducti lacteri en/of de lobuli glandulae mammariae manifesteren maligne cellen. Bij carcinoma in situ groeien de maligne cellen niet uit de voorheen genoemde structuren. Bij invasieve borstkanker verspreiden maligne cellen vanuit de ductus en/of lobus naar omgevende borstweefsel (18). Om onderscheid te maken tussen maligne en benigne tumor kunnen morfologische karakteristieken van de laesies (afwijkend weefsel) gebruikt worden. Hierbij wordt gekeken naar de vorm, afgrenzing en aankleuringspatroon van de tumoren. Maligne tumoren zijn meestal onregelmatige van vorm en de afgrenzingen van maligne tumoren zijn meestal ook onregelmatig en gespikkeld. Benigne tumoren zijn over het algemeen rond en ovaal van vorm en hebben een duidelijk afgrenzing. Maligne en benigne laesies hebben een verschillend aankleuringspatroon in MRI-mammografie (19).

SCAN TECHNISCH

Bij een MRI-mammografie onderzoek wordt gebruikt gemaakt van gadolinium contrastmiddel. De patiënt ligt tijdens het onderzoek in buikligging waarbij de borsten vrij hangen. Bij het positioneren wordt nog gecontroleerd of de borsten vrij hangen en of geen huidplooien rond om de borsten zijn. De belangrijkste sequenties voor het MRI-mammografie onderzoek zijn de dynamische contrast weging (DCE), Diffusion Weighted Imaging (DWI) en vet suppressie. Bij de DCE wordt het aankleuringspatroon weergegeven van laesies. Dit zegt iets over wat voor soort laesie het is. DCE heeft een hoge sensitiviteit bij het detec-

teren van borstkanker (sensitiviteit gemiddeld 100%) (20). Ver- Tabel 1: UMCG MRI-protocol geleken met DCE heeft DWI een hoge specificiteit bij het differentiëren tussen benigne en maligne laesies (DWI-specificiteit 79%) (21). Vet suppressie wordt gebruikt om vetachtige structuren in de borsten te onderdrukken, dit wordt gedaan om laesies niet te missen. Ook is vet suppressie nodig en voordelig voor het maken van DCE-, DWI- en T2-wegingen (22). In tabel 1 is het protocol van Universitair Medisch Centrum Groningen (UMCG) te zien, MRI-scan protocollen kunnen verschillen per ziekenhuis.

Loca	Localizer				
2	T2-w				
3	DWI (b0, b500 en b1000)				
4	T1-w fat-sat pre contrast				
Con	Contrast toediening				
6	Twist (dynamic)				
7	T1-w fat-sat (4 post-contrast series)				

KUNSTMATIGE INTELLIGENTIE

Al-systemen zijn computersystemen die kunnen leren van ervaringen (data) en met nieuwe informatie zich kunnen aanpassen voor een gewenst resultaat. Onder het domein Al vallen Machine Learning en Deep Learning (23). Deep Learning is een deelverzameling van Machine Learning toepassingen, Machine Learning is een deelverzameling van Al-toepassingen. In figuur 1 (24) wordt de onderlinge relatie visueel tussen de Al-systemen weergeven.

MACHINE LEARNING

Machine learning behoort tot een subonderdeel van kunstmatige intelligentie. Bij Machine Learning zijn de systemen in staat om bepaalde patronen in een situatie te leren herkennen. Met andere woorden wordt een algoritmisch learning machine getraind om verbanden te zien zonder expliciete programmering (hardcodering). Dit wordt gedaan door soft gecodeerde algoritmes. Deze zor-



Figuur 1: Visualisatie onderlinge relaties Al-systemen. Illustratie genomen uit Amaratunga, 2016.

gen voor dat de codering automatisch gewijzigd wordt door training, zodat de algoritmen beter worden in het kunnen uitvoeren van de gewenste taak. Met de trainingsdata kan het algoritme zichzelf optimaal configureren, zodat naast het gewenste resultaat in de trainingsdata ook uit voorheen ongebruikte gegevens een gewenst resultaat gegeneraliseerd kan worden (25). Machine learning kan verdeeld worden in meerdere subtypen, de meest voorname subtypen zijn: Supervised Learning (SL), Unsupervised Learning (UL) en Semi-supervised Learning (SSL). Bij het subtype SL zijn de algoritmen geprogrammeerd om onderdelen in groepen in te delen. Dit type kan na de leerfase (input van trainingsdata) bij nieuwe data invoer zelfstandig resultaten produceren. Het Machine Learning model leert hierdoor verbanden te zien tussen input en de gewenste output (het resultaat). Het subtype kan nog verdeeld worden in classificatie of regressie type (26). Een nadeel van SL is dat alle labels handmatig ingevoerd moeten worden. Support Vector Machines (SVM), Random Forest (RF) en Descision trees zijn voorbeelden van SL-algoritmen. Bij het subtype UL wordt ongesorteerde informatie geordend in groepen op overeenkomsten, patronen en verschillen. Dit wordt gedaan zonder voorafgaande training van data. Het algoritme maakt zonder gebruik van labels een beslissing bij het sorteren van informatie. UL kan gesplitst worden in twee categorieën: clustering en associatie. K-means Clustering is een voorbeeld van UL met clustering type. Tenslotte is het derde type SSL-algoritmes. Bij SSL zijn de twee hierboven genoemde algoritmen types gecombineerd (23). Een deel van het algoritme is geprogrammeerd met gedefinieerde labels en het andere deel gebruikt geen labels. Hierdoor kan het gelabelde deel gebruikt worden om het niet gelabelde deel te assisteren (25).

DEEP LEARNING

Deep Learning is een specifieke toepassing van Machine Learning om complexe vraagstukken op te lossen. Complexe vraagstukken kunnen door middel van Deep Learning beantwoord worden door (1) computers in staat te stellen om te leren van ervaringen en (2) de wereld te laten begrijpen in een hiërarchie van concepten. Deep Learning systemen zijn in staat om zelfstandig, zonder input of hulp van mensen, kennis te verzamelen om zich-

zelf te trainen. Daarbovenop, zijn Deep Learning systemen in staat om hiërarchie van concepten te begrijpen. Hierdoor zijn zij in staat om complexe concepten te begrijpen door deze op te bouwen uit simpelere concepten (23). Deep Learning systemen functioneren net als neuronen in hersenen, waarbij informatie als 'input' door verschillende "diepe" verborgen lagen gaat om uiteindelijk een complex vraagstuk op te lossen (27). Afhankelijk van het Deep Learning ontwerp kan de input vele diepe lagen passeren, waarbij per laag abstracte (verborgen) onderdelen van de informatie wordt ontleed. Aan de hand van de ontlede informatie beogen Deep Learning systemen onderlinge relaties in de input te verklaren. Per laag worden simpele concepten bij elkaar gevoegd om een oplossing te vinden voor een complex vraagstuk (23). Een toepassing van Deep Learning modellen is Convolutional Neural Network (CNN). CNN-modellen zijn de meest populaire vorm van Neural Network t.a.v. afbeeldingssegmentatie en is effectief



Figuur 2: Illustratie van een deep learning model. Illustratie genomen uit Goodfellow et al., 2016.

bij herkenning en classificatie in beelden (dataprocessing). Het model bestaat uit meerdere convolutionele lagen. In elk laag worden de patronen in het beeld geaccentueerd (23). Een voorbeeld van de werking van een Deep Learning model is geïllustreerd in figuur 2 (23), waarbij een afbeelding als input wordt geleverd om een classificatie te maken van het object in kwestie.

KANSEN KUNSTMATIGE INTELLIGENTIE IN MEDISCHE BEELDVORMING

In radiologische toepassing wordt Machine Learning gebruikt bij het herkennen van patronen in afbeeldingen. Machine learning leert patronen herkennen en groepeert gelijke patronen. Hierdoor kan dit gebruikt worden bij het analyseren van scans. Met ML- en DL-systemen kunnen scans geclassificeerd worden in verdacht en niet verdacht weefsel. Ook kunnen beide systemen geprogrammeerd worden om de laesie te groeperen in type tumor, daarnaast kunnen ML- en DL-systemen met genoeg training bepaalde lichaamsdelen segmenteren. De respectievelijke systemen kunnen gemiste laesies detecteren, hierdoor kunnen vals-negatieven van radiologen verminderd worden (10). Een ander voorbeeld is een röntgenfoto van de humerus (bovenarmbot), hierbij kan het algoritme de humerus indelen in gebroken of niet gebroken (28)(25). Beiden systemen kunnen ook ingezet worden in röntgenapparatuur om bijvoorbeeld optimale beeldkwaliteit te waarborgen, zoals de signaal-ruisverhouding en automatische toepassing van de LUT-curve (29). In de radiotherapie kunnen ML- en DL-systemen ingezet worden voor het optimaliseren van radiotherapie planningen, respons van behandelingen en uitkomsten voorspelling. Op basis van Deep Learning en Machine Learning optimalisatie in radiotherapie kan een behandelingsplan gemaakt worden. Ook kunnen de systemen gebruikt worden bij de kwaliteitsborging van stra-lingsfysica en beeldkwaliteit (25).

AFBEELDINGSSEGMENTATIE

Segmentatie van afbeeldingen heeft de volgende definitie: het verdelen van een afbeelding in samengestelde onderdelen of objecten. Het segmenteren van afbeeldingen kan een waardevol onderdeel zijn om oplossingen te vinden voor complexe vraagstukken waarbij herkenning van onderdelen of objecten in een afbeelding essentieel is (30). In de medische beeldvorming is de toepassing hiervan: het segmenteren van aanwezige anatomie en pathologie in onderzoekafbeeldingen. Aan de hand van segmentaties kunnen soorten weefsels of laesies in medische afbeeldingen specifiek gelokaliseerd worden (12). Dit is met name nuttig voor het opstellen van behandelplannen, voor beoordeling tijdens follow-ups of voor gebruik in analyses (12)(14). Segmentatie van me

dische afbeeldingen is mogelijk middels Machine Learning en Deep Learning segmentatiemodellen, hiernaast zijn ook rekenkundige of geometrische methoden beschikbaar om afbeeldingsegmentatie uit te voeren (12). In figuur 3 (31) zijn aanwezige objecten en personen geannoteerd. Deze annotaties kunnen als input dienen voor Machine Learning of Deep Learning systemen om verder geanalyseerd te worden (12)(14). Alle eerdergenoemde methoden voor segmentatie hebben voordelen en limitaties. Een limitatie is dat rekenkundige en geometrische methoden moeilijk sterk wisselende structuren kunnen segmenteren. Dit kan mogelijk een negatieve invloed hebben op de prestaties van een dergelijk segmentatiemodel. Een voordeel van rekenkundige en geometrische methoden is dat deze minder rekenvermogen van computers vereisen ten opzichte van ML- of DLsegmentatiesystemen (12). ML- en DL-toepassingen bieden mogelijk robuuste oplossingen voor limitatie van de voornoemde methoden, omdat zij kunnen leren van voorgaande ervaringen.



Figuur 3: Een voorbeeld van annotaties binnen de PASCAL Visual Object Classes 2010 detectie challenge. Illustratie genomen uit Mottaghi et al., 2013.

Zowel Machine Learning als Deep Learning segmentatiemodellen kunnen rekenkundige of geometrische methoden benutten om intelligent en autonoom afbeeldingen te segmenteren (12). In figuur 4 (32) is een segmentatie van borst- en FG-weefsel afgebeeld in een T1-w MRI-mammografie afbeelding.



Figuur 4: MRI-mammografie T1-w afbeelding. Links, MRI. Rechts, handmatige segmentatie van borst (rode lijn) en fibroglandulair weefsel (groen gebied). Illustratie genomen uit Dalmis et al., 2017.

METRICS

Evaluatie van de gesegmenteerde onderzoeksafbeeldingen is een essentieel onderdeel om vast te stellen in hoeverre een Machine Learning- of Deep Learning-segmentatiemodel in staat is om accuraat de gewenste weefsels, laesies of organen te kunnen segmenteren (33). De prestatie van Machine Learning- en Deep Learning segmentatiemodellen worden uitgedrukt in uitkomstmaten (hierna metrics). Vele metrics zijn beschikbaar om de prestaties van dergelijke segmentatiemodellen te beoordelen. Gebruikelijke metrics zijn: Dice Coefficient Similarity-score (DICE), Jaccard Index (JAC), True positive rate (SEN), True negative rate (SPEC) en de Area under ROC curve (AUC)(33). Voor deze metrics geldt: hoe dichter de DICE, JAC, SEN of SPEC bij de 1 liggen, hoe preciezer het segmentatiemodel in staat is om afbeeldingen te segmenteren in referentie tot een handmatige segmentatie.

METHODE

ONDERZOEKSOPZET

In deze systematische review is literatuuronderzoek verricht en zijn bestaande ML- en DL-systemen ten behoeve van segmentatie van borst- en tumorweefsel in MRI-mammografie onderzoeken in kaart gebracht, de prestaties van de segmentatiesystemen zijn vergeleken en tot slot wordt geblikt op toekomstige toepassingsmogelijkheden van ML- en DL-segmentatiesystemen in MRI-mammografie onderzoeken. Onder borstweefsel segmentatie wordt segmentatie van de gehele borst of FG-weefsel segmentatie verstaan. Het systematisch literatuuronderzoek is uitgevoerd vanaf maart tot en met begin mei 2020.

SELECTIE VAN ARTIKELEN VOOR REVIEW

Artikelen zijn op basis van opgestelde in- en exclusiecriteria geselecteerd voor inclusie aan dit onderzoek. Een lijst is opgesteld met in- en exclusiecriteria, zie tabel 2. De in- en exclusiecriteria zijn dusdanig opgesteld dat de onderzoekers de onderzoeksvraag konden beantwoorden met de vergaarde literatuur. In het artikel dienden de prestaties van de respectievelijke segmentatiesystemen onderzocht te zijn, voor de lijst van beschikbare metrics wordt verwezen naar het artikel van Taha en Hanbury (33). De selectie van artikelen is gelimiteerd tot Engels en Nederlandstalige artikelen. Hiervoor is gekozen aangezien het merendeel van de publicaties met betrekking tot kunstmatige intelligentie geschiedt in het Engels (34). Hiernaast is gekozen om enkel artikelen te includeren die zijn gepubliceerd tussen 2015 en 2020. Hiervoor is gekozen om enkel de meest recente ontwikkelingen binnen Machine Learning en Deep Learning mee te nemen. Conferentie artikelen zijn in dit onderzoek geïncludeerd om zodoende de meest recente ontwikkelingen en toepassingen mee te nemen. Exclusie vond plaats indien toepassingen van Machine Learning en Deep Learning segmentatie zijn onderzocht buiten de scope van dit onderzoek. Reviews werden geëxcludeerd van inclusie in dit onderzoek. In een PRISMA-flowchart (35) zijn de stappen van de selectieprocedure visueel weergegeven. Hiermee hebben de onderzoekers getracht om de reproduceerbaarheid en betrouwbaarheid van dit onderzoek te verhogen.

Inclusiecriteria	Exclusiecriteria
Artikelen zijn Engels- of Nederlandstalig.	Artikelen die zijn uitgebracht vóór 2015.
Artikelen waarin ML of DL wordt toegepast voor segmentatie van borstweefsel, FG- weefsel of tumorweefsel in MRI-mammo- grafieën	Artikelen waar beide onderzoekers geen toegang tot hebben.
Artikelen waarin de prestaties van bij seg- mentatiesystemen van borstweefsel of tu- morweefsel wordt onderzocht.	Artikelen die gaan ML- of DL-systemen to passing alléén in CT-scans, conventionele röntgenscans of echoscans.
	Artikelen die over ML of DL toepassingen

Tabel 2 In- en exclusiecriteria hij selectie van artikelen

ZOEKSTRATEGIEEN

Allereerst zijn primaire zoekopdrachten uitgevoerd in PubMed, Arxiv en IEEE Xplore (alle databanken zijn doorzocht in mei 2020). PubMed bevat voornamelijk publicaties met betrekking tot de klinische toepassing van MLen DL-systemen, zowel Arxiv als IEEE Xplore bevatten voornamelijk publicaties met betrekking tot de technische toepassing van ML- en DL-systemen. Door zowel de klinische als technische toepassing af te wegen is een volledig beeld van de mogelijkheden van ML- en DL-segmentatiemodellen- gepresenteerd. Aparte zoekopdrachten zijn verricht om literatuur over zowel segmentatie van borst-, FG- als tumorweefsel te verkrijgen. Naast de primaire zoekopdracht in de eerdergenoemde drie databanken is ook een secundaire zoekopdracht uitgevoerd. Reviews resulterende uit de primaire zoekopdracht zijn geanalyseerd op relevante artikelen. De screening (ofwel selectie) van artikelen werden door beide onderzoekers onafhankelijk uitgevoerd op basis van titel en abstract aan de hand van voornoemde inclusie- en exclusiecriteria. Alle gescreende artikelen zijn in aparte Excelbestanden opgenomen, waarna duplicaten zijn verwijderd. De onderzoekers hebben na screening van alle artikelen

gaan bij lymfepathologie.

toe-

dezen besproken om consensus te krijgen over de definitief te includeren artikelen. De primaire zoekopdrachten zijn uitgevoerd met vooraf opgestelde zoekstrings (zie Bijlage 1).

TOEGEPAST FILTER

In dit onderzoek zijn zes zoekstrings benut voor de primaire zoekstrategie. Per databank zijn twee zoekstrings opgesteld: één omtrent borstweefselsegmentatie en één omtrent tumorweefselsegmentatie. Hiervoor is gekozen om per databank artikelen geordend en gestructureerd te vergaren. In de zoekstrings zijn de modaliteiten "MRI mammography" en "breast MRI" opgenomen. Verder zijn in de zoekstrings de termen "tumor tissue" en "breast tissue" opgenomen. Onder breast tissue valt zowel borst- als FG-weefsel. Hiervoor is gekozen aangezien dit onderzoek de segmentatie van deze weefsels betreft. Tevens gaat de systematische review over segmentatie, hierbij zijn in alle zoekstrings "segmentation" of een combinatie met "segmentation" gebruikt. In de zoekstring zijn de metrics niet expliciet opgenomen, dit gezien meer dan twintig metrics beschikbaar zijn voor evaluatie(s) van segmentatiesystemen (33). Hier is rekening mee gehouden in de inclusiecriteria. Termen als sensitiviteit en specificiteit zijn niet opgenomen in de zoekstrings, aangezien deze termen synoniemen hebben die afwisselend gebruikt worden in AI-artikelen. Tenslotte is voor Arxiv en IEEE Xplore gekozen om niet expliciet Machine Learning of Deep Learning in de zoekstring te benoemen. Hiervoor is gekozen aangezien beiden databanken voornamelijk publicaties bevatten over kunstmatige intelligentie, Machine Learning en Deep Learning. De zoekstrings zijn per databank opgesteld, hiervoor is gekozen aangezien de zoekmachines per databank sterk in structuur verschillen.

BEOORDELING INHOUD & KWALITEIT

Na selectie van artikelen, verwijdering van duplicaten en bespreking van de selectie is consensus gecreëerd over de te includeren artikelen ten behoeve van deze systematische review. De volledige tekst van de geselecteerde artikelen is door beide onderzoekers onafhankelijk gelezen en beoordeeld op inhoud en kwaliteit, hierna te noemen toepasbaarheid en betrouwbaarheid. Aan de hand van een opgestelde checklist (gebaseerd op Dassen et al. (36)) zijn de beoordelingen onafhankelijk door beiden onderzoekers uitgevoerd, zie bijlage 2 voor de checklist. De checklist bestaat uit twee onderdelen: toepasbaarheid en betrouwbaarheid.

In de categorie toepasbaarheid konden maximaal zes punten gescoord worden, verspreid over vijf criteria (zie tabel 3). Voor het criteria metrics konden twee punten verdiend worden, dit aangezien het essentieel is voor de resultaten van dit onderzoek. Exclusie vond plaats indien een artikel beoordeeld was met een toepasbaarheidsscore van twee punten of lager. In de categorie betrouwbaarheid konden maximaal twaalf punten gescoord worden, verspreid over negen criteria (zie tabel 3). Exclusie vond plaats indien een artikel met een score van vijf punten of lager werd beoordeeld.

Na onafhankelijke beoordeling door beide onderzoekers zijn de checklists besproken. Opeenvolgend is consensus bereikt over zowel de toepasbaarheid als betrouwbaarheid van geselecteerde artikelen. De uitkomsten van beide beoordelaars werden opgeteld en hieruit is een gemiddelde score voor de checklist berekend. Tevens is beschrijvende statistiek (mediaan en interkwartielafstand) uitgevoerd over de beoordelingen van de onderzoekers om de mate van overeenkomst in beoordeling te bepalen. Hiermee kan aangetoond worden of aantoonbaar verschil tussen beoordelaars aanwezig is. De geïncludeerde artikelen werden opgenomen in een evidence tabel (zie bijlage 4).

Criteria Toepasbaarheid	Punten	Betrouwbaarheid	Punten
ML- of DL-segmentatie model is on-	1	Auteurs: heeft meer artikelen gepubli-	1
derzocht.		ceerd betreft ML- of DL-segmentatie-	
		modellen in medische beeldvorming	
Metrics volgens Taha en Hanbury (33)	2	Artikel komt uit een ziekenhuis of in-	1
gebruikt.		stelling.	
Sensitiviteit van ML- of DL-segmenta-	1	Het artikel is vaker geciteerd.	1
tie is onderzocht.			
Specificiteit van ML- of DL-segmenta-	1	Bevindingen zijn helder beschreven	2
tie is onderzocht.		(vaststellen van gegevens)	

Tabel 3 Puntentoekenning voor checklist toepasbaarheid en betrouwbaarheid

ML- en/of DL-segmentatiemodellen	1	Onderzoekgroep is concreet beschre-	1
		Methodiek is stap voor stap beschreven	2
		(reproduceerbaarheid).	
		Discussie onderbouwd met literatuur	2
		en geeft meerwaarde op onderwerp	
		(interpretatie en limitaties).	
		Conclusie van het onderwerp is eendui-	1
		dig en beantwoord onderzoeksvraag.	
		Verwijzingen en literatuur zijn volledig	1
		genoteerd.	

DATA-ANALYSE

De geïncludeerde artikelen zijn opgenomen in een evidence tabel. Voor elk type segmentatie is een aparte resultatentabel opgesteld (tabel 7, 8 en 9). De drie soorten segmentatie zijn: borst, FG-, en tumorweefselsegmentatie. Hiervoor is gekozen om overzichtelijkheid in de resultaten te bieden. In de evidence tabel werden enkel bevindingen uit artikelen met betrekking tot ML- of DL-segmentatie van de voornoemde weefselsoorten opgenomen. Hiermee werd getracht de onderzoeksvraag te beantwoorden. In de evidence tabel is opgenomen: auteur en jaartal, tijdschrift en impactfactor, doel, dataset, methode, resultaten, conclusie en uitkomst checklist. Kenmerken van geïncludeerde artikelen zijn beschreven, tevens zijn de metrics per ML- of DL-segmentatiemodel beschreven. Aan de hand van beschrijvende statistiek zijn berekeningen gemaakt om de gemiddelde en range per ML- of DL-segmentatiecategorie te bepalen. Middels de gemiddelden en ranges wordt een vergelijking gemaakt tussen de prestaties van ML- en/of DL- segmentatiemodellen. Gebruikelijke metrics zijn: DICE, JAC, SEN, SPEC en AUC (33).

RESULTATEN

UITKOMSTEN SELECTIEPROCEDURE

In figuur 5 is een schematisch flowdiagram van de

selectieprocedure weergegeven. De primaire zoekstrategie heeft in totaal 606 artikelen opgeleverd. De secundaire zoekstrategie heeft 11 artikelen opgeleverd. Vanuit 4 reviews zijn de literatuurlijsten doorzocht op relevante literatuur, hier zijn 11 artikelen uit bemachtigd. Uit de 617 artikelen zijn 113 duplicaten verwijderd. Beide onderzoekers hebben onafhankelijk de titels en abstracts van de artikelen gescreend aan de hand van de in- en exclusiecriteria. Na consensus tussen beide onderzoekers zijn 471 artikelen geëxcludeerd. In totaal voldeden 33 artikelen op basis van titel en abstract aan zowel de in- als exclusiecriteria. Beide onderzoekers hebben onafhankelijk van elkaar de 33 artikelen beoordeeld op toepasbaarheid en betrouwbaarheid aan de hand van een gevalideerde checklist. Bij alle artikelen hadden de onderzoekers toegang tot de volledige artikelen. Nadat consensus tussen beide onderzoekers aan de hand van de ingevulde checklists was bereikt, zijn 12 artikelen geëxcludeerd.

De reden van exclusie is opgenomen in figuur 5. In totaal zijn 21 artikelen geïncludeerd in dit onderzoek. Deze artikelen zijn ingedeeld in acht categorieën, zie tabel 4.

GEBRUIKTE SEGMENTATIEMODELLEN

In totaal zijn 21 artikelen geïncludeerd in dit onderzoek, waarvan 5 artikelen omtrent ML-segmentatiemodellen. In de 5 artikelen is gebruik gemaakt van 7 verschillende ML-segmentatiemodellen. Dit waren: MCWT, K-means(+), GaMM, GMM, RF, EM en KDP. Naast ML zijn ook 16 artikelen omtrent DL-segmentatiemodellen gebruikt. In de 16 artikelen is gebruik gemaakt van 7 verschillende DL-segmentatiemodellen. Dit waren: U-Net, DLCNN, 2D Seg-Net, Supervised Cross modal Deep Representation Learning, GOCS-DLP, LOCS en andere CNN-modellen. U-Net is het meest gebruikt in de artikelen, in totaal is dit model 42 keer benut in de artikelen. U-Net is een type CNN-architectuur waarin data verwerkt wordt. In de artikelen van Verburg et al. (37) en Maicas et al. (38) wordt het type architectuur niet beschreven.



Figuur 5: PRISMA-flowdiagram selectieprocedure

Tabel 4: Categorieën geïncludeerde artikelen

Categorie	Aantal artikelen
ML borstsegmentatie	1
DL borstsegmentatie	5
ML borst en FG-weefsel-	1
segmentatie	T
DL borst en FG-weefsel-	2
segmentatie	5
ML FG-weefselsegmen-	1
tatie	T
DL FG-weefselsegmenta-	p.v.t
tie	11.v.t.
ML tumorsegmentatie	2
DL tumorsegmentatie	8
Totaal:	21

TOEPASBAARHEID VAN ARTIKELEN

In totaal zijn 33 artikelen door beide onderzoekers middels de opgestelde checklist beoordeeld op zowel toepasbaarheid als betrouwbaarheid, zie bijlage 3 voor een ingevulde checklist. In totaal zijn 12 artikelen geëxcludeerd wegens een betrouwbaarheidsscore \leq 5 en/of toepasbaarheid \leq 2. Zie tabel 5 voor de puntentoekenning van de artikelen. De scores zijn per artikel vermeld in de evidence tabel, zie hiervoor bijlage 4.

Score	Aantal artikelen			
Geëxcludeerd*	12 artikelen			
12 punten	2 artikelen			
13 punten	2 artikelen			
14 punten	5 artikelen			
15 punten	8 artikelen			
16 punten	2 artikelen			
17 punten	1 artikel			
18 punten	1 artikel			

Tabel 5 gemiddelde punten bij artikelen

* toepasbaarheidsscore ≤2 en/of betrouwbaarheidsscore ≤5.

Per beoordelaar zijn de uitkomsten van de checklist geëvalueerd met een mediaan en interkwartielafstand (IQR). De toepasbaarheid en betrouwbaarheid mediaan en IQR zijn berekend per beoordelaar over de 21 geïncludeerde artikelen, zie hiervoor tabel 6. De mediaan in toepasbaarheid en betrouwbaarheid waren bij beide beoordelaars gelijk. Een minimaal verschil is te zien in IQR in toepasbaarheids- (0,5) en betrouwbaarheidsscores (1) tussen beide beoordelaars.

Tabel 6 Mediaan en interkwartielafstand van beoordeling artikelen volgens gebruikte checklist

	Toepasbaarheid Mediaan (IQR)	Betrouwbaarheid Mediaan (IQR)
Beoordelaar 1	4 (1,5)	10 (2)
Beoordelaar 2	4 (1)	10 (3)

IQR= interkwartielafstand.

EVIDENCE TABEL

Om overzicht in de resultaten te bewaren en duidelijkheid te scheppen over de details van de geïncludeerde artikelen is gekozen om deze toe te lichten in een evidence tabel, zie bijlage 4. De evidence tabel is gebruikt voor analyse van de resultaten. De geselecteerde bevindingen van de geïncludeerde artikelen zijn opgenomen in tabellen 7, 8 en 9. In de tabellen worden respectievelijk de bevindingen van borstsegmentatie, FG-weefselsegmentatie en tumorsegmentatie samengevat. Onder de bevindingen vallen: De methode, dataset, MRI-wegingen en metric(s) van het ML- of DL- segmentatiemodel. In de tabellen bij DICE, JAC, SEN en SPE zijn de uitkomsten van de testset gebruikt. Tevens staan in de tabellen DL- en/of ML-segmentatiemodellen die in een artikel gebruikt werden als vergelijking voor het ontwikkelde model. In de tabellen wordt bij de dataset gesproken over aantal scans, hiermee wordt de gehele MRI serie bedoeld. Uitkomsten in de tabellen zijn afgerond op twee decimalen. In alle artikelen is een referentiestandaard (ground truth) gehanteerd om de respectievelijke metrics te berekenen. Voor de wijze van referentie wordt gerefereerd naar de evidence tabel, zie bijlage 4.

BORSTWEEFSELSEGMENTATIE

In totaal zijn elf artikelen met betrekking tot gehele borstsegmentatie geïncludeerd, waarvan twee middels Machine Learning methoden en negen middels Deep Learning methoden. Een samenvatting van de belangrijkste bevindingen is opgenomen in tabel 7, een uitgebreidere toelichting wordt gegeven in bijlage 4.1 (evidence tabel). In de Machine Learning artikelen (40, 41) zijn verschillende methoden toegepast om gehele borstsegmentatie mogelijk te maken. Dit middels een EM-algoritme of K-means Dynamic Programming. In deze twee artikelen zijn de prestaties enkel geëvalueerd middels DICE, hiervan was de gemiddelde DICE 0,90 (range, 0,846-0,94) (40, 41). In het artikel van Gubern-Merida et al. (40) is de hoogste DICE gemeten, dit was 0,94. De JAC, SPE en SEN ten behoeve van evaluatie van de systemen zijn in beide Deep Learning artikelen (40, 41) niet benut. In de Deep Learning artikelen zijn verschillende methoden toegepast om gehele borstsegmentatie mogelijk te maken. De meest voorkomende segmentatiemethode was toepassing van een U-Net model, welke is benut in acht van de negen Deep Learning artikelen (1, 32, 37, 39, 42-46). In het artikel van Verburg et al. (37) is een DCNN-model benut. De gemiddelde datasetgrootte van de Deep Learning artikelen was 230 scans (range, 40-1114) (1, 32, 37, 39, 42-46). In de acht Deep Learning artikelen zijn de prestaties geëvalueerd middels DICE (1,34-41), hiervan was de gemiddelde DICE 0,94 (range, 0,77-0,982). In het onderzoek van Verburg et al. (37) is de hoogste DICE gemeten, dit was 0,982. Enkel in Fashandi et al. (39) is naast DICE als aanvullende metric de JAC berekend, hiervan is de gemiddelde DICE 0,91 (range, 0,90-0,92). In de artikelen (1, 32, 39, 44) is naast de DICE ook SEN de berekend, hiervan is de gemiddelde DICE 0,96 (range, 0,929-0,998). In artikelen (1, 32, 44) is de SPE berekend, hiervan is de gemiddelde DICE 0,99 (range, 0,981-0,998). In het artikel van Piantidosi et al. (1) is de hoogste SPE gemeten, dit was 0,998.

Tabel 7: Uitkomsten metrics bij borstweefselsegmentatie ten aanzien van Machine Learning of Deep Learning segmentatiemodellen (vette tekst zijn de hoogste waarden in de categorie). De waarden zijn gemiddelden en ±standaarddeviatie.

Auteur	Methode	Dataset	Wegingen	DICE	JAC	SEN	SPE
Machine Learning	g segmentatiemodellen						
Gubern-Mérida et al., 2015 (40)	EM	N=27 scans	T1-w	0,94 (±0,03)	N.V.T.	N.V.T.	N.V.T.
Rosado-Toro et al., 2015 (41)	KDP met FSRM	N=14 scans	Vet en water beel- den (tracer 1)	0,922 (±0,047)	N.V.T.	N.V.T.	N.V.T.
	KDP met FSRM	N=14 scans	Tracer 2	0,910 (±0,052)	N.V.T.	N.V.T.	N.V.T.
	KDP met FSRM	N=14 scans	Tracer 3	0,908 (±0,067)	N.V.T.	N.V.T.	N.V.T.
	KDP met MSRM	N=14 scans	Tracer 1	0,897 (±0,114)	N.V.T.	N.V.T.	N.V.T.
	KDP met MSRM	N=14 scans	Tracer 2	0,846 (±0,123)	N.V.T.	N.V.T.	N.V.T.
	KDP met MSRM	N=14 scans	Tracer 3	0,899 (±0,107)	N.V.I.	N.V.I.	N.V.I.
Deep Learning se	gmentatiemodellen						
Dalmış et al., 2017 (32)	2C-U-Net	N=66 scans	T1-w non-fatsat	0,944 (±0,026)	N.V.T.	0,933	0,99
	3C-U-Net	N=66 scans	T1-w non-fatsat	0,933 (±0.028)	N.V.T.	0,976	0,981
Fashandi et al., 2019 (39)	2D U-Net, WOFS	N=70 scans	T1-w non-fatsat (WOFS)	0,96 (±0.03)	0,92 (±0,05)	0,95 (±0,03)	N.V.T.
	2D U-Net, FS	N=70 scans	T1-w fatsat (FS)	0,94 (±0.07)	0,90 (±0,09)	0,95 (±0,03)	N.V.T.
	2D U-Net, MIXED	N=70 scans	WOFS + FS	0,95 (±0.05)	0,91 (±0,08)	0,95 (±0,03)	N.V.T.
	2D U-Net, MUTLI	N=70 scans	WOFS + FS	0,95 (±0.04)	0,91 (±0,06)	0,95 (±0,03)	N.V.T.
	3D U-Net, WOFS	N=70 scans	T1-w non-fatsat (WOFS)	0,95 (±0.02)	0,91 (±0,04)	0,93 (±0,04)	N.V.T.
	3D U-Net, FS	N=70 scans	T1-w fatsat (FS)	0,95 (±0.04)	0,90 (±0,06)	0,94 (±0,03)	N.V.T.
	3D U-Net, MIXED	N=70 scans	WOFS + FS	0,95 (±0.02)	0,91 (±0,04)	0,94 (±0,03)	N.V.T.
	3D U-Net, MULTI	N=70 scans	WOFS + FS	0,96 (±0.02)	0,92 (±0,04)	0,96 (±0,03)	N.V.T.
Ha et al., 2019 (42)	3D U-Net CNN	N=1114 scans	contrast (post-c), T1 substractie beel- den	0,947	N.V.T.	N.V.T.	N.V.T.
Ivanovska et al., 2019 (43)	2D U-Net	N=40 scans	T1-weging zonder contrast	0,98 (±0.05)	N.V.T.	N.V.T.	N.V.T.
Piantadosi et al., 2018 (44) ^b	2D U-Net	N=420 scans	T1-w pre- en post- contrast (Axiaal)	0,958	N.V.T.	0,959	0,993
Piantadosi et al., 2020 (1)	U-Net CNN W. majo- rity voting	N=42 scans	3D DCE-MRI T1- w+T1-w post-c	0,966 (±0,003)ª	N.V.T.	0,969 (±0,005) ª	0,995 (±0,001)ª
	U-Net CNN majority voting	N=42 scans	3D DCE-MRI T1- w+T1-w post-c	0,966 (±0,003)ª	N.V.T.	0,969 (±0,005) ª	0,995 (±0,001)ª
	U-Net CNN Naïve Bayes	N=42 scans	3D DCE-MRI T1- w+T1-w post-c	0,961 (±0,004)ª	N.V.T.	0,981 (±0,004) ª	0,991 (±0,001) ª
	U-Net CNN Pixelwise AND	N=42 scans	3D DCE-MRI T1- w+T1-w post-c	0,958 (±0,004) ª	N.V.T.	0,937 (±0,008) ª	0,998 (±0,0004) ª
	U-Net CNN BKS	N=42 scans	3D DCE-MRI T1- w+T1-w post-c	0,960 (±0,005)ª	N.V.T.	0,960 (±0,006) ª	0,995 (±0,001)ª

	U-Net CNN Pixelwise OR	N=42 scans	3D DCE-MRI T1- w+T1-w post-c	0,957 (±0,008)ª	N.V.T.	0,998 (±0,003) ª	0,989 (±0,002) ª
	U-Net CNN W. majo- rity voting	N=67 scans	3D DCE-MRI T1- w+T1-w post-c	0,956 (±0,005)ª	N.V.T.	0,962 (±0,008) ª	0,996 (±0,001)ª
	U-Net CNN majority voting	N=67 scans	3D DCE-MRI T1- w+T1-w post-c	0,956 (±0,005)ª	N.V.T.	0,962 (±0,008) ª	0,996 (±0,001) ª
	U-Net CNN Naïve Bayes	N=67 scans	3D DCE-MRI T1- w+T1-w post-c	0,943 (±0,005)ª	N.V.T.	0,950 (±0,008) ª	0,994 (±0,002) ª
	U-Net CNN Pixelwise AND	N=67 scans	3D DCE-MRI T1- w+T1-w post-c	0,936 (±0,006)ª	N.V.T.	0,929 (±0,010) ª	0,996 (±0,001) ª
	U-Net CNN BKS	N=67 scans	3D DCE-MRI T1- w+T1-w post-c	0,932 (±0,006)ª	N.V.T.	0,964 (±0,008) ª	0,989 (±0,001) ª
	U-Net CNN Pixelwise OR	N=67 scans	3D DCE-MRI T1- w+T1-w post-c	0,929 (±0,006)ª	N.V.T.	0,967 (±0,008) ª	0,988 (±0,002) ª
Verburg et al., 2019ª (37)	DCNN	N=115 scans	T1-w	0,982 (±0,006)	N.V.T.	N.V.T.	N.V.T.
L. Zhang et al., 2020 (45)	2D U-Net	N=144 scans	DWI	0,85 (±0,07)	N.V.T.	N.V.T.	N.V.T.
	2D Seg-Net	N=144 scans	DWI	0,77 (±0,09)	N.V.T.	N.V.T.	N.V.T.
Y. Zhang et al., 2019 (46)	U-Net	N=224 scans	T1-w non-fatsat	0,86 (±0,05)	N.V.T.	N.V.T.	N.V.T.

²⁰¹⁹ (40) ³Gemiddelden + interkwartiel-range; ^bConferentie artikel. N.V.T. Niet van toepassing; DICE, Dice Similarity Coefficient; JAC, Jaccard Index; SEN, Sensitiviteit, SPE, Specificiteit. CNN, Convolutional neural network; DCNN, Dilated convolutional neural network; EM, Expection-maximization algoritme; KDP, K-means Dynamic Programming.

FG-WEEFSELSEGMENTATIE

In totaal zijn vijf artikelen met betrekking tot FG-weefselsegmentatie geïncludeerd, waarvan twee (40, 47) middels Machine Learning methoden en drie middels Deep Learning methoden (42, 43, 46). Een samenvatting van de belangrijkste bevindingen is opgenomen in tabel 8. Een uitgebreidere toelichting wordt gegeven in bijlage 4.2 (evidence tabel). In de Machine Learning artikelen zijn verschillende methoden toegepast om FG-weefselsegmentatie mogelijk te maken, middels een EM-algoritme, GaGMM en GMM (40, 47). In Gubern-Merida et al. (40) en Kumari et al. (47) zijn de prestaties geëvalueerd middels DICE, hiervan was de gemiddelde DICE 0,78 (range, 0,619-0,917). In het artikel van Kumari et al. (47) is de hoogste DICE gemeten, dit was 0,918. De SEN en SPE ten behoeve van evaluatie van de systemen zijn in beide Machine Learning niet benut. In Kumari et al. (47) is de JAC als aanvullende metric gebruikt ter evaluatie van twee segmentatiemethoden, hiervan is de gemiddelde DICE 0,72 (range, 0,549-0,882). Tevens wordt in hetzelfde artikel de hoogste JAC gemeten, dit was 0,882. In de Deep Learning artikelen zijn methoden toegepast om FG-weefselsegmentatie mogelijk te maken. De meest voorkomende segmentatiemethode was toepassing van een U-Net model, welke is benut in de drie Deep Learning artikelen (42, 43, 46). De gemiddelde datasetgrootte van de Deep Learning artikelen was 459 scans (range, 40-1114) (42, 43, 46). In de drie Deep Learning artikelen zijn de prestaties voornamelijk geëvalueerd middels DICE (42, 43, 46), hiervan was de gemiddelde DICE 0,86 (range, 0,813-0,925). In artikel Ivanovska et al. (43) is de hoogste DICE gemeten, dit was 0,925. Tevens wordt enkel in dit artikel als aanvullende metrics de SEN en SPE berekend, deze waren respectievelijk 0,933 en 0,999. Doordat enkel het artikel van Ivanovska et al. (43) de SEN en SPE heeft gerapporteerd konden hier geen gemiddelden en range over berekend worden. De JAC is niet benut voor evaluatie van segmentatieprestatie in de Deep Learning artikelen.

Tabel 8: Uitkomsten metrics bij FG-weefselsegmentatie ten aanzien van. Machine Learning of Deep Learning segmentatiemodellen (vette tekst zijn de
hoogste waarden in de categorie). De waarden zijn gemiddelden en ±standaarddeviatie.

Auteur	Methode	Dataset	Wegingen	DICE	JAC	SEN (%)	SPE (%)
Machine Learning	segmentatiemo	odellen					
Gubern-Mérida et al., 2015 (40)	EM	N=50 scans	T1-w non-fatsat	0,80 (±0,13)	N.V.T.	N.V.T.	N.V.T.
Kumari et al., 2017 (47) ^b	GaGMM	N=50 scans	T1-w non-fatsat	0,917	0,882	N.V.T.	N.V.T.
	GMM	N=50 scans	T1-w non-fatsat	0,619	0,549	N.V.T.	N.V.T.
Deep Learning seg	gmentatiemode	llen					
Ha et al., 2019 (42)	3D CNN U- net	N=1114 scans	T1-w pre- en post-con- trast, T1 substractie beelden	0,813	N.V.T.	N.V.T.	N.V.T.
Ivanovska et al., 2019 (43)	2D U-Net	N=40 scans	T1-w zonder contrast	0,925 (±0,011)	N.V.T.	0933 (±0,08)	0,999 (±0,00005)
Y. Zhang et al., 2019 (46)	U-Net	N=224 scans	T1-w non-fatsat	0,83 (±0,06)	N.V.T.	N.V.T.	N.V.T.

^aGemiddelden + interkwartiel-range; ^bConferentie artikel.

N.V.T. Niet van toepassing; DICE, Dice Similarity Coefficient; JAC, Jaccard Index; SEN, Sensitiviteit, SPE, Specificiteit; CNN, Convolutional neural network; EM, Expectation-Maximization algoritme; GaGMM, Gamma Gaussian Mixture Model; GMM, Gaussian Mixture Model.

TUMORSEGMENTATIE

In totaal zijn negen artikelen met betrekking tot tumorsegmentatie geïncludeerd, waarvan twee middels Machine Learning methoden (48, 49) en zeven middels Deep Learning methoden (38, 50-55). Een samenvatting van de belangrijkste bevindingen is opgenomen in tabel 9 (zie hieronder) en een uitgebreidere toelichting wordt gegeven in bijlage 4.3 (evidence tabel). In de Machine Learning artikelen zijn Vesal et al. (48) en Vogl et al. (49) zijn verschillende methoden toegepast om tumorsegmentatie mogelijk te maken, dit middels K-means, GMM, MCWT, RF GI en RF mRMR. De gemiddelde datasetgrootte over de Machine Learning artikelen Vesal et al. (48) en Vogl et al. (49) was niet berekend. In Vesal et al. (48) was de dataset 106 laesies, het totale aantal MRI-scans is niet benoemd. De dataset van Vogl et al. (48) bevatte 34 scans. In Vesal et al. (48) en Vogl et al. (49) zijn de prestaties enkel geëvalueerd middels DICE, hiervan was de gemiddelde DICE 0,70 (range, 0,608-0,786). De JAC, SEN en SPE ten behoeve van evaluatie van de methoden is enkel in het onderzoek van Vesal et al. benut, hiervan is geen gemiddelde of range berekend. De artikelen onder Deep Learning hebben verschillende methoden toegepast om gehele tumorsegmentatie mogelijk te maken. De meest voorkomende segmentatiemethode was toepassing van een U-Net model, welke is benut in de zeven Deep Learning artikelen (38, 50-55). De gemiddelde datasetgrootte van de Deep Learning artikelen (38, 51-55) was 173 scans (range, 35-313). In Spuhler et al. (50) is het aantal MRI-scans niet in het artikel vermeld. In de Deep Learning artikelen zijn de prestaties voornamelijk geëvalueerd middels DICE. Hiervan was de gemiddelde DICE: 0,74 (range, 0,59-0,836). Enkel in Li et al. (51), Piantadosi et al. (52) en J. Zhang et al. (53) is als aanvullende metric de SEN berekend, hiervan was de gemiddelde 0,77 (range, 0,68-0,844). Doordat enkel in Piantadosi et al. (52) de SPE is gerapporteerd konden geen gemiddelde en range berekend worden. De JAC is niet benut voor evaluatie van segmentatieprestatie in de Deep Learning artikelen.

Tabel 9: Uitkomsten metrics bij tumorweefsel segmentatie ten aanzien van Machine Learning of Deep Learning segmentatiemodellen (vette tekst zijn de hoogste waarden in de categorie), waarden zijn gemiddelden en ±standaarddeviatie.

Auteur	Methode	Dataset	Wegingen	DICE	JAC	SEN (%)	SPE (%)
Machine Lear	ning segmentatiemode	llen					
Vesal et al., 2018 (48)	K-means	N=106 lae- sies ^c	T1-w post-c	0,732 (±0,206)	0,612 (±0,209)	0,805 (±0,243)	0,702 (±0,204)
	GMM	N=106 lae- sies ^c	T1-w post-c	0,746 (±0,180)	0,623 (±0,193)	0,855 (±0,213)	0,697 (±0,195)
	MCWT	N=106 lae- sies ^c	T1-w post-c	0,786 (±0,172)	0,679 (±0,217)	0,866 (±0,199)	0,752 (±0,250)
Vogl et al., 2019 (49)	RF GI LOOCV	N=34 scans	MRI met DWI	0,608	N.V.T.	0,739	N.V.T.
	RF mRMR LOOCV	N=34 scans	MRI met DWI	0,618	N.V.T.	0,748	N.V.T.
Deep Learning	g segmentatiemodellen	1					
Li et al., 2019 (51)	Supervised Cross- modal Deep Repre- sentation Learning	N=313 scans	T2-w pre-C fat-sat + T1- w post-C fat-sat	0,776	N.V.T.	0,844	N.V.T.
	U-Net CNN	N=313 scans	T1 post-C	0,733	N.V.T.	0,795	N.V.T.
	U-Net CNN Vgg 16	N=67 scans	T1-w, T2-w, DWI + dy- namic	0,832	N.V.T.	N.V.T.	N.V.T.
Lu et al., 2019 (54) ^b	U-Net CNN ResNet- 50	N=67 scans	T1-w, T2-w, DWI + dy- namic	0,829	N.V.T.	N.V.T.	N.V.T.
	U-Net CNN Inception V3	N=67 scans	T1-w, T2-w, DWI + dy- namic	0,815	N.V.T.	N.V.T.	N.V.T.
	U-Net CNN DenseNet	N=67 scans	T1-w, T2-w, DWI + dy- namic	0,836	N.V.T.	N.V.T.	N.V.T.
Maicas et al., 2017 (38)	GOCS-DLP	N=117 scans	T1-w en T2-w	0,77 (±0,13)	N.V.T.	N.V.T.	N.V.T.
, í	CNN	N=117 scans	T1-w en T2-w	0,66 (±0,18)	N.V.T.	N.V.T.	N.V.T.
	LOCS	N=117 scans	T1-w en T2-w	0,59 (±0,17)	N.V.T.	N.V.T.	N.V.T.
Piantadosi et al., 2019 (52)	3TP U-Net CNN	N=35 scans	T1-w Pre-contrast + post contrast	0,61 (±0,12)	N.V.T.	0,68 (±0,09)	1
Spuhler et al., 2019 (50)	U-Net CNN	N=263 lae- sies ^b	DCE MRI (R1 vs R1)	0,71 (±0,16)	N.V.T.	N.V.T.	N.V.T.
· · · ·			DCE MRI (R3 vs R3)	0,67 (±0,20)	N.V.T.	N.V.T.	N.V.T.
	U-Net FCN	N=236 scans	DCE MRI	0,813	N.V.T.	N.V.T.	N.V.T.
Wang et al., 2019 (55)	U-Net FCN DS	N=236 scans	DCE MRI	0,698	N.V.T.	N.V.T.	N.V.T.
	U-Net FCN DM	N=236 scans	DCE MRI	0,700	N.V.T.	N.V.T.	N.V.T.
1 Thomast	U-Net FCN DM+BM	N=236 scans	DLE IVIRI	0,728	N.V.I.	N.V.I.	N.V.I.
J. Znang et al., 2019 (53)	3D U-Net MHL FCN	N=272 scans	post-contrast	0,72 (±0,24)	N.V.T.	0,75 (±0,23)	N.V.T.

^aGemiddelden+interkwartiel-range; ^bConferentie artikel; ^cAantal MRI-scans niet vermeld in artikel.

N.V.T. Niet van toepassing; DICE, Dice Similarity Coefficient; JAC, Jaccard Index; SEN, Sensitiviteit, SPE, Specificiteit; GOCS-DLP, Globally optimal interference on a continuous space using a DL shape prior; CNN, Convolutional neural network; LOCS, locally optimal inference on a continuous space.

DISCUSSIE

In dit literatuuronderzoek is onderzoek verricht naar in hoeverre ML- en DL-systemen in staat zijn om borstweefsel en/of tumorweefsel te segmenteren in MRI-mammografie onderzoeken. De huidige toepassingen van MLen DL-systemen ten aanzien van borst-, FG- en tumorweefsel segmentatie en haar prestaties hierin zijn in kaart gebracht. Op basis van de gehanteerde systematische aanpak kan gesteld worden dat vergelijkbare resultaten bij herhaling verkregen worden en dat daarmee de resultaten van deze studie valide zijn.

BORSTWEEFSEL SEGMENTATIE

Uit de uitkomsten bleek dat voor segmentatie van borstweefsel Deep Learning systemen (gemiddelde DICE 0,94; range, 0,77-0,982) beter presteerden dan Machine Learning systemen (gemiddelde DICE 0,90; range, 0,846-0,94). Kijkend naar de range kan gesteld worden dat zowel ML- als DL-systemen met hoge precisie borstweefsel kunnen segmenteren.

Noemenswaardig is de lage mate van spreiding in de DICE-range voor zowel ML- als DL-systemen. In het artikel van Verburg et al. (37) is een DICE van 0.94 gemeten bij segmentatie van extreem dense borsten, wat een zeer hoge mate van precisie voor segmentatie inhoudt. In dit artikel is echter enkel een interne dataset met extreem dense borsten gesegmenteerd, waardoor de prestaties niet doorgetrokken kunnen worden naar andere datasets. In het artikel van Dalmis et al. (32) is een toepassing van een 2C U-Net segmentatiemodel onderzocht, welke robuuste prestaties (DICE = 0,944) leverde. In dit onderzoek zijn verschillende borstdensiteiten middels verschillende MRI-protocollen gesegmenteerd om prestaties in een ziekenhuissetting na te bootsen. Hieruit bleek dat de 2C U-Net toepassing significant (p=0,0055) beter presteerde dan Atlas-based segmentatiemethoden in segmentatie van borstweefsel. Waar in Dalmis et al. (32) echter geen rekening mee is gehouden zijn mogelijke verschillen tussen verschillende merken MRI-scanners. Tevens is enkel gebruik gemaakt van non fat-sat MRI-scans. Hierdoor kan niet met volledigheid gezegd worden dat het 2C U-Net model in staat is om in klinische settingen, waar andere MRI-wegingen ook worden benut, borstsegmentatie uit te voeren. In L. Zhang et al. (45) zijn U-Net en SegNet segmentatiemethoden voor segmentatie van borstweefsel onderzocht in DWI MRI-scans. De prestaties van respectievelijk U-Net en SegNet waren een DICE van 0,85 en 0,77. Uit het onderzoek kan geconcludeerd worden dat het U-Net segmentatiemodel effectiever in staat is om borstweefsel te segmenteren ten opzichte van SegNet.

Opmerkelijk is de sterk wisselende datasetpopulatie per artikel. In de borstsegmentatie artikelen is enkel door Piantadosi et al. (1) een publieke database benut voor meting van de prestaties, in de andere artikelen zijn private datasets benut. Dit komt mogelijk doordat weinig grote kwalitatieve (gelabelde) MRI-datasets beschikbaar zijn voor onderzoeksdoeleinden. In alle artikelen zijn de DICE-waarden berekend door de overeenkomst van de ML- of DL-segmentatie te vergelijken met een referentie handmatige of semiautomatische segmentatie. Uit de resultaten kan geconcludeerd worden dat zowel ML- als DL-systemen, binnen hun toepassingskader, in staat zijn om met hoge precisie borstweefsel te segmenteren in MRI-mammografieën.

FG-WEEFSEL SEGMENTATIE

Uit de uitkomsten bleek dat voor segmentatie van FG-weefsel Deep Learning systemen (gemiddelde DICE: 0,86; range, 0,813-0,925) beter presteerden dan Machine Learning systemen (gemiddelde DICE: 0,78; range, 0,619-0,918). Kijkend naar de range kan gesteld worden dat DL-systemen met hoge precisie FG-weefsel kunnen segmenteren. ML-systemen toonden wisselende prestaties ten aanzien van FG-weefselsegmentatie.

Noemenswaardig is de spreiding in de DICE-range (0,619-0,917) met betrekking tot ML-systemen (40,47). In het artikel van Kumari et al. (47) zijn twee segmentatiemethoden onderzocht, waarvan de prestaties onderling zijn vergeleken. De DICE-gemiddelde en -range voor de ML-systemen is niet representatief voor de prestaties van de ML-systemen, gezien zowel het gemiddelde als de range gevoelig is voor uitschieters in waarden (56).

De grootte van de datasets in Gubern-Merida et al. en Kumari et al. (40,47) was gelijk, echter verschillen de kenmerken in dataset en methodologie per artikel sterk. In de dataset van Kumari et al. (47) was een gelijke spreiding van borstdichtheden aanwezig. Het GaGMM-segmentatiemodel was in staat om FG-weefsel te segmenteren met een DICE van 0,917. De GaGMM methode presteerde 48% beter dan de GM-methode. De GaGMM methode bleek aanzienlijk beter om te kunnen gaan met (a)symmetrische verdelingen van FG-weefselintensiteiten ten opzichte van de GM-methode. In Gubern-Merida et al. (40) is een gemiddelde DICE 0.80 gemeten. Dit duidt dat het ML-segmentatiemodel met hoge mate van precisie FG-weefsel kan segmenteren. Dit lag aan het feit dat het ATLAS-onderdeel van het segmentatiemodel niet goed afgestemd was op de

samenstelling van MRI-data. Hierin waren nog verbeteringen mogelijk. In dit artikel wordt aangegeven dat niet te veel nadruk op vergelijkingen met andere artikelen gemaakt dient te worden. Dit wegens het feit dat de prestaties over verschillende datasets worden gemeten met verschillende segmentatiemodellen (40).

Opvallend is dat, net als bij de ML- en DL-borstsegmentatie artikelen het doel van segmentatie per artikel sterk verschilt. Hiernaast zijn in zowel de borst- als FG-artikelen verschillende methodologieën zijn gebruikt om het FG-weefsel te segmenteren. In Ivanovska et al. en Gubern-Merida et al. (43,40) werden de segmentatieresultaten benut voor analyse van borstdichtheid, waarin het artikel op is toegespitst, waar bij de artikelen van Ha et al., Y. Zhang, et al. en Gubern-Merida et al. (42, 46, 40) voornamelijk segmentatie van FG-weefsel is onderzocht. Hiernaast heerst heterogeniteit in de datasets die zowel in ML- als DL-artikelen gebruikt zijn. Dit blijkt uit zowel de spreiding in de datasetgroottes als de gebruikte sequenties binnen deze datasets (zie tabel 8). Tevens zijn de borstdensiteitskenmerken binnen de datasets niet in elk artikel benoemd. Hierdoor kan niet geconcludeerd worden dat de ML- en DL-modellen in staat zijn om hoge prestaties te leveren bij wisselende borstdensiteiten. De hoeveelheid FG-weefsel in de borsten kan sterk per patiënt wisselen, hier bleek weinig rekening mee gehouden te zijn. Enkel in het artikel van Kumari et al. (47) zijn de borstdensiteitskenmerken binnen de dataset benoemd. De heterogene datasets hebben echter ook interessante resultaten opgeleverd voor dit onderzoek. In het Deep Learning artikel van Ivanovska et al. (43) zijn voor FG-weefsel segmentatie MRI-scans benut met sterke signaal inhomogeniteit artefacten, hieruit bleek een prestatie van DICE 0,925. Het onderzochte segmentatiemodel bleek robuuste prestaties te leveren onder deze omstandigheden (43). Dit komt mogelijk doordat het segmentatiemodel een borst-masker benutte om FG-weefsel binnen het borst-masker te segmenteren. Door het borst-masker was het niet-relevante weefsel gemaskeerd, waardoor het U-Net CNN-model zich kon richten op segmentatie van het aanwezige FG-weefsel (43).

Opmerkelijk is het lage aantal ML- en DL-artikelen betreffende FG-weefselsegmentatie. Dit wordt verweten aan het feit dat niet expliciet in de zoekstrings termen over FG-weefsel zijn opgenomen. De informatie die verkregen is over FG-weefsel is ontleed uit artikelen waarin zowel borst- als FG-weefselsegmentatie is uitgevoerd. De artikelen toonden veelbelovende resultaten en lieten ook een toepassing van Al in MRI-mammografie zien. Zodoende is gekozen om FG-weefselsegmentatie artikelen in dit onderzoek te includeren om een volledig beeld van Al-mogelijkheden binnen MRI-mammografie te schetsen. Wegens de lage hoeveelheid van zowel MLals DL-artikelen met betrekking tot FG-weefselsegmentatie is het resultaat mogelijk niet representatief voor bestaande ML- en DL-segmentatiesystemen. Mede hierdoor kan geen duidelijke uitspraak gedaan worden over in hoeverre segmentatiesystemen in het algemeen in staat zijn om FG-weefsel te segmenteren. In alle artikelen zijn de DICE-waarden berekend door de overeenkomst van de ML- of DL-segmentatie te vergelijken met een referentie handmatige of semiautomatische segmentatie. Uit de resultaten kan geconcludeerd worden dat de onderzochte ML- en DL-systemen, binnen hun toepassingskader, in staat zijn om met hoge precisie FG-weefsel te segmenteren in MRI-mammografieën.

TUMORSEGMENTATIE

Uit de uitkomsten bleek dat voor segmentatie van tumorweefsel Deep Learning systemen (gemiddelde DICE: 0,74; range, 0,59-0,836) beter presteerden dan Machine Learning systemen (gemiddelde DICE: 0,70; range, 0,608-0,786). DL- segmentatiemodellen presteerden gemiddeld beter in tumorweefselsegmentatie dan ML- segmentatiemodellen.

Net zoals bij FG-segmentatie toonden DL-systemen voor tumorsegmentatie een grote spreiding in de DICE-range. In deze spreiding geldt dat wederom sprake is uitschieters, waardoor de gemiddelde DICE de -range niet representatief (55) is voor de prestaties van de DL-systemen.

In Maicas et al. (38) is een combinatie van Machine Learning (RF)- en Deep Learning (U-Net CNN) toegepast om automatische tumorsegmentatie mogelijk te maken. Dit wordt ook wel een ensemble systeem genoemd. Aan de hand van vooraf gesegmenteerde (halfautomatisch) en niet-gesegmenteerde (automatisch) datasets zijn tumorsegmentaties met hoge precisie uitgevoerd. De halfautomatische (gemiddelde DICE: 0,77) en automatische toepassingen (gemiddelde DICE: 0,77) hebben gelijke prestaties geleverd. In het artikel wordt door de auteurs een nuance over de prestaties aangebracht. In volledig automatische segmentatie waren niet alle laesies gedetecteerd en deze zijn niet meegenomen in de segmentatiestappen. Deze laesies waren het lastigst om te segmenteren. Dit heeft invloed op de validiteit van de resultaten. Mede hierdoor kan niet overduidelijk geconcludeerd worden of de volledig automatische toepassing beter is ten opzichte van de halfautomatische toepassing. Binnen Maicas et al. (38) zijn tevens de laagste DICE-waarden gemeten (0,59). De reden hiervoor is niet beschreven in het artikel. In Lu et al. (54) is een DL-toepassing ontwikkeld waarbij de output van een classificatiemodel is gebruikt om tumorweefsel te segmenteren. Dit is een bijzondere toepassing, gezien normaliter segmentatie alvorens classificatie plaatsvindt. In deze toepassing is een gemiddelde DICE van 0.836 gemeten, wat een hoge mate van segmentatieprecisie aangeeft. Hierbij dient echter gemeld te worden dat een precies ML- of DL-classificatiesysteem een noodzakelijke randvoorwaarde is voor het mogelijk maken van precieze tumorsegmentaties (54).

De twee artikelen met betrekking tot Machine Learning tumorsegmentatiemodellen (48,49) hebben verschillende methoden gebruikt om haar doelen te bereiken. Opvallend is dat de prestaties tussen de artikelen sterk uiteenlopen (gemiddelde DICE: 0,7; range 0,608-0,786). Het voornaamste verschil tussen Vesal et al. (48) en Vogl et al. (49) is de wijze van segmentatie en de grootte van de datasets. In Vogl et al. (47) is segmentatie uitgevoerd middels twee methoden. Voor beide methoden geldt dat sprake was van oversegmentatie van tumorweefsel doordat dit moeilijk te onderscheiden was van regulier borstweefsel bij de contrastweging. Dit heeft de DICE negatief beïnvloed. In Vesal et al. (48) zijn drie methoden onderzocht, waarbij de MCWT het best presteerde (DICE: 0,786). De MCWT-methode kon beter het tumorweefsel verbinden ten opzichte van GMM en Kmeans. Deze twee laatstgenoemden methoden presteerden daardoor slechter bij tumoren die uitgebreid waren over een groter oppervlak (48).

In de tumorsegmentatie artikelen is enkel door Wang et al. (55) een publieke database gebruikt voor meting van de prestaties, in de andere artikelen zijn private datasets benut. Dit kan de reproduceerbaarheid van artikelen benadelen, aangezien andere onderzoekers doorgaans geen toegang hebben tot de private datasets. In alle artikelen zijn de DICE-waarden berekend door de overeenkomst van de ML- of DL-segmentatie te verge-lijken met een handmatige of semiautomatische segmentatie. Uit de resultaten kan geconcludeerd worden dat een deel van zowel ML- als DL-systemen in staat is om met hoge precisie tumorweefsel te segmenteren in MRI-mammografieën.

SYSTEMATISCHE REVIEW

Alhoewel deze systematische review identificatie van recente ML- en DL-ontwikkelingen ten aanzien van borst, FG- en tumorweefselsegmentatie mogelijk maakt, leidt dit wel tot algemene versimpelde beschrijvingen van de werking van de segmentatiesystemen in dit onderzoek. Hierin geldt dat ML- en DL-systemen hetzelfde segmentatieprobleem ervaren, maar lossen de respectievelijke systemen dit middels zeer verschillende manieren op. Daarnaast geldt dat de geïncludeerde studies die een zeer heterogene aard hebben, verschillen in onder andere; het doel van het artikel, dataset kenmerken, dataset grootte, wijze van training van het ML- of DL-segmentatiemodel. Voornoemde redenen bemoeilijken het maken van directe vergelijkingen tussen artikelen. Beschrijvende statistiek is uitgevoerd om in kaart te brengen in hoeverre ML- en DL-systemen in staat zijn om borst-, FGT- en tumorsegmentaties uit te voeren.

In deze systematische review zijn in de resultaten de eindprestaties van deze systemen weergeven, echter toont dit niet de architectuur en de werkingen van de segmentatiemodellen. Deze informatie is ook belangrijk voor een globaal beeld over in hoeverre segmentatiesystemen functioneren. Informatie over trainingsschema's (epochs of batches), of de input getransformeerd is en aanvullende informatie over het segmentatiemodel zijn opgenomen in de evidence tabel (zie bijlage 4). De onderzoekers hebben hiervoor gekozen aangezien het doel van dit onderzoek was om recente ontwikkelingen in kaart te brengen en haar prestaties ten aanzien van de segmentatie taak te vergelijken. Om dit vraagstuk te beantwoorden is de keuze gemaakt om voornamelijk de metrics in de resultaten op te nemen. De geïnteresseerde lezer kan de aanvullende informatie inzien in de evidence tabel.

Veelal komt in de artikelen naar voren dat ML- en DL-segmentatiesystemen belovende resultaten tonen, echter gelden deze prestaties enkel voor een gelimiteerd kader. Flexibiliteit tussen ziekenhuizen en verschillende MRI-scanners of MRI-sequenties is niet aan de orde, aangezien de ML- en DL-segmentatiesystemen hier mogelijk niet voor getraind zijn. Gezien de onderzochte segmentatiemodellen voornamelijk getraind worden met private datasets zijn de resultaten niet overdraagbaar naar andere instellingen. In een groot deel van de artikelen is niet onderzocht hoe segmentatiesystemen presteren op data van andere MRI-scanners of ziekenhuisinstellingen. Een kans voor vervolgonderzoek is dat nader onderzoek wordt uitgevoerd op publieke datasets met een spreiding van MRI-sequenties van verschillende MRI-scanners. Op deze wijze kan aangetoond worden dat segmentatiesystemen robuust om kunnen gaan met verschillende MRI-scanners. Enkel Piantidosi et al. (1) en Wang et al. (55) hebben prestaties gemeten over een publieke dataset. Hiernaast kan ook winst behaald worden in de reproduceerbaarheid en validiteit van artikelen door verduidelijking van de kenmerken van de benutte datasets. In de artikelen is gebleken dat veel informatie wordt verschaft over de MRI-sequenties en parameters hierin, over de leeftijden van de patiënten in de dataset. Echter blijkt het dat weinig informatie wordt verschaft over de borstdensiteit, typen tumoren of laesies die heersen binnen datasets. Hiermee wordt niet de indruk gewekt dat segmentatiesystemen robuust met onbekende uiteenlopende MRI-scans kunnen omgaan. Het is raadzaam dat in vervolg AI-artikelen duidelijk informatie hierover wordt verschaft, dit kan de reproduceerbaarheid en validiteit ten goede komen.

Bij de selectieprocedure van dit onderzoek zijn 471 artikelen geëxcludeerd op basis van de titel, abstract, in- en exclusiecriteria. Dit omdat zoektermen zoals tumorweefsel en borstweefsel gebruikt zijn. Aangezien in de zoekstring de term "tumorweefsel" gebruikt werd, resulterende dat in artikelen over allerlei tumoren, bijvoorbeeld: hersentumoren, levertumoren en nier tumoren. Tevens door de termen "tumorweefsel" en "borstweefsel" te gebruiken stonden in de resultaten veel artikelen over histopathologische segmentatie van borst- en/of tumorweefsel. Doordat in de zoekstrings niet expliciet stond dat andere modaliteiten niet in de zoekresultaten mochten staan, kwamen artikelen in de zoekresultaten naar voren over andere modaliteiten zoals CT, echo en x-mammografie bij de resultaten te staan. Verder zijn in dit onderzoek drie databanken gebruikt voor het vergaren van artikelen. Door enkel drie databanken te gebruiken zijn mogelijk relevante artikelen voor dit onderzoek gemist. De onderzoekers achten de kans hiervoor echter laag gezien het feit dat in de gekozen databanken op jaarbasis veel Al-publicaties worden uitgebracht. Hiernaast zijn de zoektermen algemeen gehouden, waardoor relevante literatuur niet werd gemist. In dit onderzoek zijn enkel Engelstalige en Nederlandstalige artikelen geïncludeerd. Het merendeel van de publicaties met betrekking tot kunstmatige intelligentie geschiedt in het Engels (34), de onderzoekers achtten hierdoor de kans laag in dat ten gevolge van de taallimitatie literatuur gemist werd.

Mogelijke selectiebias van artikelen is verminderd door een robuuste zoekstrategie. Beide onderzoekers hebben onafhankelijk van elkaar artikelen geïdentificeerd, gescreend en beoordeeld op kwaliteit. De geïncludeerde artikelen zijn getoetst met behulp van een checklist (zie bijlage 2). De checklist is gebaseerd op Dassen et al. (36) en bevat criteria om de toepasbaarheid en betrouwbaarheid van artikelen te beoordelen. Achteraf bleek dat in de checklist betrouwbaarheid stellingen verschillend geïnterpreteerd werden door beide onderzoekers. Dit heeft echter geen waarneembare gevolgen gehad, omdat beide onderzoekers doorgaans dezelfde punten aan artikelen hebben toegekend. De mediaan en IQR betreft wat beoordeling van toepasbaarheid en betrouwbaarheid kwamen bij beide beoordelaars relatief overeen. De mediaan waarden bij toepasbaarheid en betrouwbaarheid waren van beide beoordelaars gelijk, minimaal verschil heerste tussen beide beoordelaars in IQR. Hieruit is af te leiden dat beide onderzoekers een overeenkomst in kwaliteit bij de artikelen hebben ondervonden. Om mogelijke selectiebias verder te verlagen zijn de checklists van alle artikelen besproken en is een consensus omtrent inclusie van artikelen gezocht en bereikt.

In deze systematische review zijn conferentie artikelen geïncludeerd. Hiervoor is gekozen aangezien middels dit onderzoek de meest recente ontwikkelingen en toepassingen in kaart werden gebracht. De onderzoeken binnen kunstmatige intelligentie verlopen op hoog tempo, resultaten voortvloeiend uit experimentele onderzoeken worden doorgaans gepresenteerd op een conferentie. De conferentie artikelen zijn net als reguliere artikelen getoetst met de toepasbaarheids- en betrouwbaarheidschecklist. De checklist uitkomsten van de conferentie artikelen waren vergelijkbaar met reguliere artikelen.

Hiernaast mist bij een aantal artikelen informatie over de kenmerken van de gebruikte dataset, dit verlaagt de reproduceerbaarheid van de respectievelijke artikelen. Tevens is in de resultaten (zie tabellen 7,8 en 9) mogelijk sprake van selectiebias van de uitkomsten. Hierin zijn enkel uitkomsten van de testdata opgenomen. De onderzoekers vonden testdata uitkomsten het meest relevant aangezien dit weergeeft hoe het ML- of DL-segmentatiemodel uiteindelijk heeft gepresteerd voor het segmentatieprobleem. Hierdoor mist de lezer mogelijk context over de prestaties tijdens de ontwikkeling of trainingsfase van het segmentatiesysteem.

CONCLUSIE

Kunstmatige intelligentie heeft in andere vakgebieden aangetoond dat mogelijkheden bestaan om de segmentatietaken te automatiseren, echter is weinig onderzoek verricht naar de toepassingen in MRI-mammografieën. In dit onderzoek is onderzocht in hoeverre ML- en DL-modellen in staat zijn om borst-, FG-weefsel en tumorweefsel te segmenteren in MRI-mammografie onderzoeken. ML- en DL-modellen tonen veelbelovende resultaten voor segmentatie van borst-, FG- en tumorweefsel. Ondanks de veelbelovende resultaten zijn de respectievelijke systemen nog niet klaar voor klinische toepassing. Waar de ML- en DL-modellen in hun ingekaderde toepassingsgebied lovende resultaten tonen, zijn met name validiteit en reproduceerbaarheid obstakels die nog opgelost dienen te worden voor brede inzet van kunstmatige intelligentie. Als aanbeveling dient nader extensief onderzoek verricht te worden naar de prestaties van ML- en DL-modellen op grotere datasets met variërende MRI-sequenties van verschillende MRI-scanners. Hiernaast dient kritisch nagedacht te worden over de implementatie van ML- en DL-modellen en in hoeverre zij accuraat borst-, FG- of tumorweefsel dienen te segmenteren. Dit is bepalend voor in hoeverre het ML- of DL- segmentatiemodel robuust in haar prestaties moet zijn. Aanvullend onderzoek dient hiernaar verricht te worden. Definitief kan geconcludeerd worden: ML- en DL-systemen zijn in staat zijn om borst-, FG- of tumorweefsel te segmenteren in MRI-mammografie onderzoeken, maar in de huidige staat zijn zij nog niet klinisch inzetbaar in de medische beeldvorming.

LITERATUURLIJST

- 1. Piantadosi G, Sansone M, Fusco R, Sansone C. Multi-planar 3D breast segmentation in MRI via deep convolutional neural networks. Artif Intell Med. maart 2020;103:101781.
- 2. Nederlandse Kankerregistratie (NKR), IKNL. NKR Cijfers, Dataselectie, Incidentie, Borstkanker, 2019 [Internet]. IKNL; 2019. Beschikbaar op: iknl.nl/nkr-cijfers
- 3. Duncan, W., Kerr, G.R. The curability of breast cancer. Br Med J. 1976;2:781–3.
- 4. Oeffinger KC, Fontham ETH, Etzioni R, Herzig A, Michaelson JS, Shih Y-CT, e.a. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 20 oktober 2015;314(15):1599–614.
- Lee CH, Dershaw DD, Kopans D, Evans P, Monsees B, Monticciolo D, e.a. Breast Cancer Screening With Imaging: Recommendations From the Society of Breast Imaging and the ACR on the Use of Mammography, Breast MRI, Breast Ultrasound, and Other Technologies for the Detection of Clinically Occult Breast Cancer. J Am Coll Radiol. 1 januari 2010;7(1):18–27.
- 6. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, e.a. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success. J Am Coll Radiol JACR. 2018;15(3 Pt B):504–8.
- 7. Shiraishi J, Li Q, Appelbaum D, Doi K. Computer-Aided Diagnosis and Artificial Intelligence in Clinical Imaging. Semin Nucl Med. 1 november 2011;41(6):449–62.
- Behrens S, Laue H, Althaus M, Boehler T, Kuemmerlen B, Hahn HK, e.a. Computer assistance for MR based diagnosis of breast cancer: Present and future challenges. Comput Med Imaging Graph. 1 juni 2007;31(4):236–47.
- 9. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp. 24 oktober 2018;2(1):35.
- 10. Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. JAMA. 13 december 2016;316(22):2353–4.
- 11. Giger ML. Machine Learning in Medical Imaging. J Am Coll Radiol. 1 maart 2018;15(3, Part B):512–20.
- 12. Pham DL, Xu C, Prince JL. Current Methods in Medical Image Segmentation. Annu Rev Biomed Eng. 1 augustus 2000;2(1):315–37.
- 13. Codari M, Schiaffino S, Sardanelli F, Trimboli RM. Artificial Intelligence for Breast MRI in 2008–2018: A Systematic Mapping Review. Am J Roentgenol. februari 2019;212(2):280–92.
- 14. Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Trans Med Imaging. mei 2016;35(5):1240–51.
- 15. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol. 1 september 2017;10(3):257–73.
- 16. Martini FH, Bartholomew EF. Hoofdstuk 19 Het voortplantingstelsel, 19.3.6 De melkklieren. In: Anatomie en fysiologie. Pearson; 2016.
- 17. Hermanek P, Sobin LH. TNM Classification of Malignant Tumours. Springer Science & Business Media; 2012. 205 p.
- 18. Martini FH, Bartholomew EF. Hoofdstuk 19 Het voortplantingsstelsel, Klinische aantekening: Borstkanker. In: Anatomie en fysiologie. Pearson; 2016.
- 19. Macura KJ, Ouwerkerk R, Jacobs MA, Bluemke DA. Patterns of Enhancement on Breast MR Images: Interpretation and Imaging Pitfalls. Radiogr Rev Publ Radiol Soc N Am Inc. 2006;26(6):1719.
- Pinker K, Moy L, Sutton EJ, Mann RM, Weber M, Thakur SB, e.a. Diffusion-weighted Imaging with Apparent Diffusion Coefficient Mapping for Breast Cancer Detection as a Stand-Alone-Parameter: Comparison with Dynamic Contrast-enhanced and Multiparametric Magnetic Resonance Imaging. Invest Radiol. oktober 2018;53(10):587–95.

- 21. Chen X, Li W, Zhang Y, Wu Q, Guo Y, Bai Z. Meta-analysis of quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesions. BMC Cancer. 29 december 2010;10:693.
- 22. Lin C, Rogers CD, Majidi S. Fat suppression techniques in breast magnetic resonance imaging: a critical comparison and state of the art [Internet]. Vol. 8, Reports in Medical Imaging. Dove Press; 2015 [geciteerd 6 mei 2020]. p. 37–49. Beschikbaar op: https://www.dovepress.com/fat-suppression-techniques-in-breast-magnetic-resonance-imaging-a-crit-peer-reviewed-fulltext-article-RMI
- 23. Goodfellow I, Bengio Y, Courville A. Deep Learning [Internet]. MIT Press; 2016. Beschikbaar op: http://www.deeplearningbook.org
- 24. Amaratunga, T. Difference between Artificial Intelligence, Machine Learning, and Deep Learning [Internet]. 2016. Beschikbaar op: https://www.codesofinterest.com/2016/11/difference-artificial-intelligencemachine-learning-deep-learning.html
- 25. El Naqa I, Murphy MJ. What Is Machine Learning? In: El Naqa I, Li R, Murphy MJ, redacteuren. Machine Learning in Radiation Oncology [Internet]. Cham: Springer International Publishing; 2015 [geciteerd 29 april 2020]. p. 3–11. Beschikbaar op: http://link.springer.com/10.1007/978-3-319-18305-3_1
- 26. Garbade DMJ. Regression Versus Classification Machine Learning: What's the Difference? [Internet]. Medium. 2018 [geciteerd 29 april 2020]. Beschikbaar op: https://medium.com/quick-code/regression-versus-classification-machine-learning-whats-the-difference-345c56dd15f7
- 27. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. mei 2015;521(7553):436–44.
- 28. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. RadioGraphics. 17 februari 2017;37(2):505–15.
- 29. Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine Learning in Medical Imaging. IEEE Signal Process Mag. juli 2010;27(4):25–38.
- 30. Gonzalez RC, Woods RE. Digital image processing [Internet]. 3rd ed. Upper Saddle River, N.J: Prentice Hall; 2008. 954 p. Beschikbaar op: http://web.ipac.caltech.edu/staff/fmasci/home/astro_refs/Digital_Image_Processing_3rdEd_truncated.pdf
- 31. Mottaghi R, Chen X, Liu X, Cho N-G, Lee S-W, Fidler S, e.a. The Role of Context for Object Detection and Semantic Segmentation in the Wild. In 2013.
- 32. Dalmış MU, Litjens G, Holland K, Setio A, Mann R, Karssemeijer N, e.a. Using deep learning to segment breast and fibroglandular tissue in MRI volumes. Med Phys. 2017;44(2):533–46.
- 33. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging [Internet]. 12 augustus 2015 [geciteerd 23 april 2020];15. Beschikbaar op: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4533825/
- 34. Niu J, Tang W, Xu F, Zhou X, Song Y. Global Research on Artificial Intelligence from 1990–2014: Spatially-Explicit Bibliometric Analysis. ISPRS Int J Geo-Inf. mei 2016;5(5):66.
- 35. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Med. 21 juli 2009;6(7):e1000097.
- 36. Dassen ThWN, Keuning FM, Jansen WS, Jansen GJ. Lezen en beoordelen van onderzoekspublicaties. ThiemeMeulenhoff bv; 2007.
- 37. Verburg E, Wolterink JM, Waard SN de, Išgum I, Gils CH van, Veldhuis WB, e.a. Knowledge-based and deep learning-based automated chest wall segmentation in magnetic resonance images of extremely dense breasts. Med Phys. 2019;46(10):4405–16.
- Maicas G, Carneiro G, Bradley AP. Globally optimal breast mass segmentation from DCE-MRI using deep semantic segmentation as shape prior. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). 2017. p. 305–9.
- 39. Fashandi H, Kuling G, Lu Y, Wu H, Martel AL. An investigation of the effect of fat suppression and dimensionality on the accuracy of breast MRI segmentation using U-nets. Med Phys. 2019;46(3):1230–44.
- Gubern-Mérida A, Kallenberg M, Mann RM, Martí R, Karssemeijer N. Breast Segmentation and Density Estimation in Breast MRI: A Fully Automatic Framework. IEEE J Biomed Health Inform. januari 2015;19(1):349–57.

- 41. Rosado-Toro JA, Barr T, Galons J-P, Marron MT, Stopeck A, Thomson C, e.a. Automated Breast Segmentation of Fat and Water MR Images Using Dynamic Programming. Acad Radiol. 1 februari 2015;22(2):139–48.
- 42. Ha R, Chang P, Mema E, Mutasa S, Karcich J, Wynn RT, e.a. Fully Automated Convolutional Neural Network Method for Quantification of Breast MRI Fibroglandular Tissue and Background Parenchymal Enhancement. J Digit Imaging. februari 2019;32(1):141–7.
- 43. Ivanovska T, Jentschke TG, Daboul A, Hegenscheid K, Völzke H, Wörgötter F. A deep learning framework for efficient analysis of breast volume and fibroglandular tissue using MR data with strong artifacts. Int J Comput Assist Radiol Surg. 1 oktober 2019;14(10):1627–33.
- 44. Piantadosi G, Sansone M, Sansone C. Breast Segmentation in MRI via U-Net Deep Convolutional Neural Networks. In: 2018 24th International Conference on Pattern Recognition (ICPR). 2018. p. 3917–22.
- 45. Zhang L, Mohamed AA, Chai R, Guo Y, Zheng B, Wu S. Automated deep learning method for wholebreast segmentation in diffusion-weighted breast MRI. J Magn Reson Imaging JMRI. februari 2020;51(2):635–43.
- 46. Zhang Y, Chen J-H, Chang K-T, Park VY, Kim MJ, Chan S, e.a. Automatic Breast and Fibroglandular Tissue Segmentation in Breast MRI Using Deep Learning by a Fully-Convolutional Residual Neural Network U-Net. Acad Radiol. 1 november 2019;26(11):1526–35.
- 47. Kumari V, Sheoran G, Kanumuri T, Koul P. Gamma Gaussian Mixture Modeling for Fibroglandular Tissue Segmentation in MR Images. In: 2017 14th IEEE India Council International Conference (INDICON). 2017. p. 1–5.
- 48. Vesal S, Ravikumar N, Ellman S, Maier A. Comparative Analysis of Unsupervised Algorithms for Breast MRI Lesion Segmentation. ArXiv180208655 Cs [Internet]. 23 februari 2018 [geciteerd 3 mei 2020]; Beschikbaar op: http://arxiv.org/abs/1802.08655
- 49. Vogl W-D, Pinker K, Helbich TH, Bickel H, Grabner G, Bogner W, e.a. Automatic segmentation and classification of breast lesions through identification of informative multiparametric PET/MRI features. Eur Radiol Exp [Internet]. 27 april 2019 [geciteerd 1 mei 2020];3. Beschikbaar op: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6486931/
- Spuhler KD, Ding J, Liu C, Sun J, Serrano-Sosa M, Moriarty M, e.a. Task-based assessment of a convolutional neural network for segmenting breast lesions for radiomic analysis. Magn Reson Med. 2019;82(2):786–95.
- Li C, Sun H, Liu Z, Wang M, Zheng H, Wang S. Learning Cross-Modal Deep Representations for Multi-Modal MR Image Segmentation. ArXiv190801997 Cs Eess [Internet]. 6 augustus 2019 [geciteerd 3 mei 2020]; Beschikbaar op: http://arxiv.org/abs/1908.01997
- 52. Piantadosi G, Marrone S, Galli A, Sansone M, Sansone C. DCE-MRI Breast Lesions Segmentation with a 3TP U-Net Deep Convolutional Neural Network. In: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). 2019. p. 628–33.
- 53. Zhang J, Saha A, Zhu Z, Mazurowski MA. Hierarchical Convolutional Neural Networks for Segmentation of Breast Tumors in MRI With Application to Radiogenomics. IEEE Trans Med Imaging. februari 2019;38(2):435–47.
- Lu W, Wang Z, He Y, Yu H, Xiong N, Wei J. Breast Cancer Detection Based on Merging Four Modes MRI Using Convolutional Neural Networks. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019. p. 1035–9.
- 55. Wang L, Shen H, Zhang J, Zhu Y, Jiang C. A Clifford Analytic Signal-Based Breast Lesion Segmentation Method for 4D Spatial-Temporal DCE-MRI Sequences. IEEE Access. 2020;8:3901–10.
- 56. Fisher MJ, Marshall AP. Understanding descriptive statistics. Aust Crit Care. 1 mei 2009;22(2):93–7.

BIJLAGE 1: DATABANK ZOEKSTRINGS

ZOEKSTRING PUBMED

De zoekstring die benut werd voor het vergaren van artikelen met betrekking tot ML- en DL-segmentatie van borstweefsel was als volgt:

(breast segmentation) AND (breast tissue OR breast/exp OR breast) AND (MRI mammography OR MRI/exp OR breast MRI) AND (artificial intelligence/exp OR machine learning OR machine learning/exp OR deep learning OR deep learning/exp) AND ("2015"[Date - Publication] : "2020"[Date - Publication])

De zoekstring die benut werd voor het vergaren van artikelen met betrekking tot ML- en DL-segmentatie van tumorweefsel was als volgt:

(tumor segmentation/exp OR tumor segmentation) AND (MRI mammography OR MRI/exp OR breast MRI) AND (artificial intelligence/exp OR machine learning OR machine learning/exp OR deep learning OR deep learning/exp) AND ("2015"[Date - Publication]: "2020"[Date - Publication])

ZOEKSTRING ARXIV

De zoekstring die benut werd voor het vergaren van artikelen met betrekking tot ML- en DL-segmentatie van borstweefsel was als volgt:

Query: order: -announced_date_first; size: 50; date_range: from 2015-01-01 to 2020-12-31; include_cross_list: True; terms: AND all=breast segmentation; OR all=segmentation; AND all=MRI mammography; OR all=breast mri; AND all=breast tissue

De zoekstring die benut werd voor het vergaren van artikelen met betrekking tot ML- en DL-segmentatie van tumorweefsel was als volgt: Query: order: -announced_date_first; size: 50; date_range: from 2015-01-01 to 2020-12-31; include_cross_list: True; terms: AND all=tumor segmentation; OR all=segmentation; AND all=MRI mammography; OR all=breast MRI; AND all=tumor tissue

ZOEKSTRING IEEE XPLORE

De zoekstring die benut werd voor het vergaren van artikelen met betrekking tot ML- en DL-segmentatie van borstweefsel was als volgt:

((((("All Metadata":breast segmentation) OR "All Metadata":segmentation) AND "All Metadata":MRI mammography) OR "All Metadata":breast mri) AND "All Metadata":breast tissue)

Naast het gebruik van de bovenstaande zoekstring was een jaarfilter toegepast van 2015 tot en met 2020.

De zoekstring die benut werd voor het vergaren van artikelen met betrekking tot ML- en DL-segmentatie van tumorweefsel was als volgt:

((((("All Metadata":tumor segmentation) OR "All Metadata":segmentation) AND "All Metadata":MRI mammography) OR "All Metadata":breast mri) AND "All Metadata":tumor tissue)

Naast het gebruik van de bovenstaande zoekstring was een jaarfilter toegepast van 2015 tot en met 2020.

BIJLAGE 2: CHECKLIST TOEPASBAARHEID ARTIKELEN

Artikel:			Segmentatie van Borst-, FG- of tumor	weefsel
Auteur:			Beoordelaar:	
Toepasbaarheid	Ja	Nee	Opmerkingen	Score
In het artikel wordt het onderwerp ML- of DL- seg-				
mentatiemodel bij MRI-segmentatie van borst- of	:			
tumorweefsel onderzocht. (1 punt)				
In het artikel worden prestaties van een van de				
systemen van segmentatie borst- of tumorweef-				
sel beschreven. Zo ja, welke? (2 punten)				
In de resultaten wordt gesproken over sensitiviteit				
van het segmentatiemodel. (1 punten)				
In de resultaten wordt gesproken over specificiteit				
van het segmentatiemodel. (1 punten)				
Zijn binnen hetzelfde onderzoek ML- en/of DL-mo-				
dellen met elkaar vergeleken? (1 punt)				
Toepasbaarheid score (max 6 punten)	I			
Betrouwbaarheid	Ja	Nee	Opmerkingen	Score
Hebben de auteurs meerdere artikelen over toe-				
passingen van ML- of DL-segmentatiemodellen				
met betrekking tot Medische Beeldvorming ge-				
schreven? (1 punt)				
Het artikel komt uit een ziekenhuis of instel-				
ling. Schrijf in opmerkingen. (1 punt)				
Wordt het artikel geciteerd door anderen?				
Schrijf in opmerkingen. (1 punt)				
Ziin de bevindingen in het artikel helder beschre-				
ven (vaststellen van gegevens)? (2 punten)				
Zit het onderzoek degelijk in elkaar? (May 7 pup				
ten)				
- Onderzoeksgroep is concreet beschreven. (1				
punt)				
- Methodiek is stap voor stap beschreven (2 pun-				
ten)				
- Discussie onderbouwd met literatuur en geeft				
meerwaarde op het onderwerp. (2 punten)				
- Conclusie van het onderwerp is eenduidig en be-				
antwoordt onderzoeksvraag. (1 punt)				
- Verwijzingen en literatuurlijst is volledig geno-				
teerd. (1 punt)				
Wat is de impactfactor van de tijdschrift? Schrijf in	n.v.t.			n.v.t.
opmerkingen.				
In welk tijdschrift of publicatiewebsite is het arti-	n.v.t.			n.v.t.
kel gepubliceerd? Schrijf in opmerkingen.				
Betrouwbaarheid score (max. 12 punten)				
Totale score				

BIJLAGE 3: INGEVULDE CHECKLIST

Artikel: Fully Automated Convolutional Neural Network Method for Quanti-Segmentatie van fication of Breast MRI Fibroglandular Tissue and Background Parenchymal En-Borst en FGT-tissue hancement

Auteur: Ha et al., 2019			Beoordelaar:	
Toepasbaarheid	Ja	Nee	Opmerkingen	Score
In het artikel wordt het onderwerp ML- of DL-sys- teem bij MRI-segmentatie van borst- of tumorweefsel on- derzocht. (1 punt)	- Ja -		U-Net FGT segmentation	1
In het artikel worden prestaties van een van de sys- temen van segmentatie borst- of tumorweefsel beschre- ven. Zo ja, welke? (2 punten)	- Ja -		DICE	2
In de resultaten wordt gesproken over sensitiviteit van het segmentatiesysteem. (1 punten)	t	Nee		
In de resultaten wordt gesproken over specificiteit van het segmentatiesysteem. (1 punten)	t	Nee		
Worden ML- en/of DL-systeem met elkaar vergeleken? (1 punt)		Nee		
Toepasbaarheid score (max 6 punten)				3
Betrouwbaarheid	Ja	Nee	Opmerkingen	Score
Hebben de auteurs meerdere artikelen over toepassingen	ı Ja		Meerdere ML/DL toepassingen m.b.t.	.1
van ML- of DL-modellen m.b.t. Medische Beeldvorming ge- schreven? (1 punt)	-		medische beeldvorming	
Het artikel komt uit een ziekenhuis/instelling. Schrijf in op- merkingen. (1 punt)	- Ja		Department of Radiology, Columbia University Medical Center	1
Wordt het artikel geciteerd door anderen? Schrijf in opmer- kingen. (1 punt)	- Ja		8 cititaties (Scholar)	1
Zijn de bevindingen in het artikel helder beschreven (vast- stellen van gegevens)? (2 punten)	- Ja		Helder vastgesteld	2
Zit het onderzoek degelijk in elkaar? (max 7 punten) - Onderzoeksgroep is concreet beschreven. (1 punt) - Methodiek is stap voor stap beschreven (2 punten) - Discussie onderbouwd met literatuur en geeft meer- waarde op het onderwerp. (2 punten) - Conclusie van het onderwerp is eenduidig en beant- woordt onderzoeksvraag. (1 punt)			Onderzoeksgroep: Benoemd, echter niet volledig uitgelegd. Geen duidelijk- heid over kenmerken onderzoekspo- pulatie (mate van dense borsten) Methode is duidelijk! Sterke discussie!	6
 Verwijzingen en literatuurlijst is volledig genoteerd. (1 punt) 	1			
Wat is de impact factor van de Journal? Schrijf in opmerkin- gen.	-n.v.t.		2,99 (2019)	n.v.t.
In welk tijdschrift/publicatiewebsite is het artikel gepubli- ceerd? Schrijf in opmerkingen.	-n.v.t.		Journal of Digital Imaging	n.v.t.
Betrouwbaarheid score (max. 12 punten)				11
Totale score				14

BIJLAGE 4: EVIDENCE TABLE

4.1 EVIDENCE TABLE BORSTWEEFSEL SEGMENTATIE

Auteur, jaartal	Tijdschrift [Impact- factor]	Doel	Dataset	Methode	Resultaten	Conclusie	Score Checklist
Fas- handi et al., 2019 (38)	Medical Physics: The Inter- national Journal of Medical Physics Re- search and Practice [3.177]	Prestaties van aan- tal 2D U-net en 3D U-Net segmentatie- modellen bij gebruik vetsuppressie en non-fat suppressie beelden beoorde- len. Vergelijking tus- sen borstsegmenta- tie methoden bij t1- w weging met en zonder vetsuppres- sie.	N=70 T1-gewogen sequen- tie zonder en met vet- suppressie. Vier verschil- lende type MRI-data zijn gebruikt om de segmenta- tiemodellen te trainen: T1- w precontrast zonder vet- suppressie (WOFS), T1-w precontrast met vet- suppressie (FS), combinatie van WOFS en FS (MIXED) en 2-channel volume met gebruik van uitgelijnde WOFS en FS-volumes (MULTI).	2D U-Net en 3D U-Net zijn onderzocht voor het segmenteren van borstweef- sel. Dit wordt gedaan bij verschillende wegingen en combinatie van wegingen. Annotaties worden gemaakt door een RF-classificatie en door worden door drie auteurs gecontroleerd (ground truth).	DICE, JAC, Sensitiviteit: 2D WOFS 0.96 ± 0.03 , 0.92 ± 0.05 , 0.95 ± 0.03 2D FS 0.94 ± 0.07 , 0.90 ± 0.09 , 0.95 ± 0.03 2D MIXED 0.95 ± 0.05 , 0.91 ± 0.08 , 0.95 ± 0.03 en 2D MULTI 0.95 ± 0.04 , 0.91 ± 0.06 , 0.95 ± 0.03 3D WOFS 0.95 ± 0.02 , 0.91 ± 0.04 , 0.93 ± 0.04 3D FS 0.95 ± 0.04 , 0.90 ± 0.06 , 0.94 ± 0.03 3D MIXED 0.95 ± 0.02 , 0.91 ± 0.04 , 0.94 ± 0.03 en 3D MULTI 0.96 ± 0.02 , 0.92 ± 0.04 , 0.96 ± 0.02	Hoogste DICE is behaald bij de 3D MULTI-dataset. Alle methode geven een accurate segmentatie met een ge- middelde DICE-score van >94% bij ge- bruik van U-Net. Dit verkregen met de informatie van beide vet onderdrukt en niet vet onderdrukt volumes.	17
Zhang L et al., 2020 (43)	Journal of Magnetic Resonance Imaging [3.732]	In het onderzoek wordt een deep transfer learning methode ontwik- keld voor het seg- menteren van borst- weefsel bij DWI MRI-scans en getest met vier verschil- lende datasets.	N= 144 MRI-scans van 98 patiënten. In totaal 11035 slices. Vier verschillende datasets zijn gebruikt: 1. 4251 2D DCE MRI slices (precontrast wegingen), van 39 patiënten zonder mammaca. (D1) 2. 431 2D DCE MRI slices 20 van 20 patiënten. (D2) 3. 6343 2D DWI MRI slices van 75 MRI-scans (29	De datasets zijn individueel geanno- teerd door radiologen. Elk dataset is door een radioloog geannoteerd. Daarna zijn twee deep learning segmen- tatiemodellen gebruikt voor het seg- menteren van borstweefsel, namelijk: 2D U-Net en 2D SegNet. Segmentatie wordt slice-to-slice gedaan. D1 is ge- bruikt voor het pre-trainen van de mo- dellen. D2 is gebruikt voor toetsing en evaluatie van de segmentatiemodellen die getraind op D1. De DWI-data van D3 wordt daarna gebruikt bij het afstellen	Pre-training DCE MRI-dataset 1 U-Net 0.92 ± 0.07 SegNet 0.84 ± 0.11 Pre-training DCE MRI-dataset 2 U-Net 0.87 ± 0.06 SegNet 0.80 ± 0.06 Fine-tuned DWI-dataset 3 U-Net 0.85 ± 0.07 SegNet 0.77 ± 0.09 DWI Dataset 3 zonder pre-training U-Net 0.69 ± 0.12 SegNet 0.73 ± 0.18	Samenvattend zijn geautomatiseerde segmentatiemodellen onderzocht bij het segmenteren van borstweefsel in DWI MRI. De segmentatiemodellen en methode die gebruikt zijn, kan toege- past worden om kwantitatieve ana- lyse van DWI-scans te verbeteren.	15

			patiënten zonder mam-	van de pre-trained segmentatiemodellen	DWI Fine-tuned dataset 4		
			maca.). (D3)	en worden ook gebruikt bij training zon-	U-Net 0.72 ±0.16		
			4. 10 representatieve 2D	der pre-trained modellen. In de pre-trai-	SegNet 0.65 ±0.10		
			DWI MRI Slices van 10 MRI-	ning worden 30 epochs gebruikt. D4 is			
			scans van 10 patiënten met	gebruikt voor toetsing van de segmenta-			
			mammacs. (D4)	tiemodellen die getraind zijn op D4.			
Zhang Y	Academic	In het onderzoek	Trainingsdata N= 276 pati-	U-Net is gebruikt voor het segmenteren	Bij training dataset	Deep learning model: U-Net kan ge-	14
et al.,	Radiology	wordt een deep	enten.	van borstweefsel en daarna FG. Ground	Borst segmentatie:	bruikt worden om accurate resultaten	
2019	[2.267]	learning segmenta-	Testdata N= 224 scans.	truth was gedaan door template-based	DICE 0.83-0.98 (gemiddelden 0.95	te verkrijgen bij het segmenteren van	
(44)		tie methode ge-	224 scans komen van 28	automatische borst segmentatie me-	±0.02).	borstweefsel en FG bij MRI onder-	
. ,		maakt en getoetst	patiënten, die allemaal ge-	thode. De segmentatie is daarna geëva-	FG-weefselsegmentatie:	zoeken. Dit kan een efficiënt en be-	
		op het segmenteren	scand zijn door 4 verschil-	lueerd door een ervaren radioloog met	DICE 0.73-0.97 (gemiddelden 0.91	trouwbare methode zijn om een groot	
		van borstweefsel en	lende MRI-scanners Linker-	12 jaar ervaring.	±0.03).	aantal MRI onderzoeken te segmente-	
		fibroglandulair	en rechterborst wordt ge-	U-Net is daarna getraind en gevalideerd	Bij test dataset:	ren voor het kwantitatief analyseren	
		weefsel (FG) bij pre-	splitst	en daarna zijn de prestaties van het sys-	Borst segmentatie:	van borstdichtheid.	
		contrast T1-gewo-	opiliot	teem gemeten (DICE).	Gemiddelde DICE 0.86 ±0.05		
		gen beelden zonder			FG-weefselsegmentatie:		
		vetsuppressie.			Gemiddelde DICE 0.83 ±0.06		
Pianta-	Artificial In-	In deze onderzoek	N=109 DCE MRI series met	Annotaties van beide datasets zijn ge-	DSprivate42 ACC, SEN, SPE en DICE (%)	Voorgestelde methode laat zien dat	16
dosi et	telligence	wordt een methode	histopathologisch vastge-	daan door een ervaren radioloog. Twee	Coronaal: 99.04 (±0.11), 96.91 (± 0.70),	de methode met voting strategie in	
al.,	in Medi-	voorgesteld bij de	stelde laesies. N=42 patiën-	verschillende dataset zijn gebruikt om	99.40 (± 0.10) en 96.03 (± 0.32).	staat is om borstweefsel in een multi-	
2020	cine	automatisch extrac-	ten (DSprivate42) en N=67	de voorgestelde CNN-model: 2D U-Net	Sagittaal: 98.94 (±0.12), 96.19 (± 0.61),	protocol (3D volume) met hoge preci-	
(1)	[3.574]	tie van borsten in	patiënten (DSpublic88).	te testen op haar prestaties. Dataset is	99.46 (± 0.08) en 96.02 (± 0.56).	sie te segmenteren.	
		een CAD-systeem.	Twee verschillende data-	in 3D volume en wordt gesplitst in drie	Transversaal: 98.98 (±0.14), 96.92 (±	-	
		De methode wordt	sets zijn gebruikt. In beide	richtingen: axiaal, coronaal en sagittaal	0.59), 99.34 (± 0.12) en 95.83 (± 0.58)		
		getest of 3D data	series zijn pre- en post-con-	en ook getraind (3D naar 2D). De trai-	W. Majority Voting: 99.16 (± 0.13),		
		gesegmenteerd kan	trast wegingen geïnclu-	ning wordt slice-to-slice uitgevoerd. Na	96.85 (± 0.47), 99.49 (± 0.09) en 96.60		
		worden zonder een	deerd. Bij DSpublic88 is	de trainingen in de drie richtingen wor-	(± 0.30).		
		parameter explosie	vetsuppressie gebruikt bij	den de slices op weer elkaar gestapeld			
		te veroorzaken. On-	alle scans.	(2D weer naar 3D). Dit wordt gedaan	DSpublic88 ACC, SEN, SPE en DICE (%)		
		derzochte methode		door voting strategieën te gebruiken.	Coronaal: 98.79 (± 0.11), 95.33 (± 0.80),		
		gebruikt gemodifi-			99.44 (± 0.13) en 94.80 (± 0.51).		
		ceerde 2D U-Net.			Sagittaal: 98.41 (± 0.11), 96.31 (± 0.83),		
					98.89 (± 0.10) en 93.07 (± 0.60).		
					Transversaal: 98.42 (± 0.13), 96.37 (±		
					0.83), 98.91 (± 0.13) en 93.25 (± 0.60).		
					W. Majority Voting: 98.97 (± 0.13),		
					96.20 (± 0.82), 99.54 (± 0.10) en 95.78		
					(± 0.51).		

Daluata	Madi		N-CC MPL seems die von	De second worden hendrastig geographie	2011 Note Dovet componiatio (CC date	Canaludarand can Deep Learning LL	10
Dalmis et al., 2017 (31)	Medi- cal Physics The Inter- natio- nal Journal of Medical Physics Re- search and Practice [3.177]	Het doel van het on- derzoek is om met U-Net Deep Lear- ning segmentatiemo- dellen borstweefsel en FG te segmente- ren. Hierbij worden de prestaties van 2C-U-Net met een 3C-U-Net model vergeleken. Ook wordt het vergele- ken met traditionele technieken zoals at- las-based en tem- plating matching.	N=66 MRI-scans die ran- dom geselecteerd zijn van het wetenschappelijk ar- chief (leeftijdscategorie van 25-75). De patiënten die gebruikt zijn hebben geen voorgeschiedenis met borstkanker. Dataset is ver- deeld in vijf groepen (pro- tocollen), omdat verschil in parameters en apparaten zijn: Protocol 1 bestaat uit 27 scans in coronale vlak (1.5T); Protocol 2 bestaat uit 17 scans in coronale vlak (3T); Protocol 3 bestaat uit 4 scans in axiale vak (3T); Protocol 4 bestaat uit 13 scans in axiale vlak (3T); Protocol 5 bestaat uit 5 scans in axiale vlak (3T).	De scans worden handmatig gesegmen- teerd door een biomedisch engineer en zijn daarna gevalideerd door een borst- radioloog (9 jaar ervaring in MRI-mam- mografie). Annotaties zijn gemaakt van de borsten en fibroglandulair weefsel. Alle coronaire scans zijn voor de trai- ning gereconstrueerd naar axiale plak- ken. De MRI-scans zijn allemaal gesplitst in linker en rechterborst scans. Scans wordt getraind op twee methoden: 2C U-Net en 3C U-Net. Eerst wordt borst- volume gesegmenteerd en daarna FG. In 2C U-Net methode wordt eerst een 2- class U-Net gebruikt om de borsten te segmenteren en vervolgens wordt een tweede 2-class U-Net gebruikt om de FG te segmenteren. Bij 3C U-Net model worden FG en borstvolume in een keer gesegmenteerd (input scans, output segmentatie FG en borstvolume). Voor- gestelde methoden zijn vergeleken met twee bestaande segmentatiemodellen (atlas-based en sheetness-based).	2C U-Nets Borst segmentatie (66 data- set) Overal DICE 0.944 (±0.026); SEN 93.3%, SPE 99%. 3C U-Net borst segmentatie (66 data- set) Overal DICE 0.933 (±0.028); SEN 97.6%, SPE 98.1%.	Concluderend, een Deep Learning U- Net model is ontwikkeld om borst- en FG-weefsel te segmenteren in inho- mogene MRI-datasets. Deze dataset bevatte vele MRI-scans vanuit ver- schillende MRI-protocollen met vari- erende borstdichtheden. Het segmen- tatiemodel presteerde goed bij seg- mentatie van borst- en FG-weefsel.	18
Ivanovs ka et al., 2019 (41)	Internatio- nal Journal of Compu- ter As- sisted Radi- ology and Surgery [2,40]	Het doel van dit on- derzoek was om een deep learning me- thode te ontwikke- len en te gebruiken om de borstdicht- heidsschatting in MRI-mammografie data (met sterke ar- tefacten zoals inho- mogeniteitsartefac- ten). Deze methode wordt ook geëvalu- eerd tijdens het on- derzoek.	N=40 MRI-mammografie met T1-weging zonder con- trast. De borsten uit dit on- derzoek werd geclassifi- ceerd door een radioloog volgens BIRADS.	De scans zijn handmatig gesegmenteerd op fibroglandulair weefsel en borst- weefsel (ground truth). Eerst wordt een correctie van artefacten uitgevoerd bij de dataset. Daarna werd de gegevens van de data vergroot. Als derde stap wordt borstvolume (zonder tepels) ge- segmenteerd. Tenslotte wordt de fi- broglandulair weefsel gesegmenteerd. Voor het segmenteren wordt een 2D U- Net model gebruikt. Om variëteit in de trainingsdata te krijgen zijn simpele data-vergrotingstechnieken gebruikt. In een MRI-scan zijn de linker- en rechter- borst van elkaar gescheiden en wordt beiden gebruikt als trainingsdata.	FG-segmentatie met tepelcorrectie Ge- middelde DICE 0.925 ±0.011 Sensitiviteit 0.933 ±0.08 Specificiteit 0.9994 ±0.00005 Borst volume segmentatie DICE 0.98 ±0.05	Deze methode kan accuraat borst- en FG-weefsel segmenteren en is hoog efficiënt om te gebruiken bij een groot epidemiologische data met 1000 deel- nemers. Het model is getraind op een moderne Graphics Processing Unit (GPU) en is hierdoor een snel. De GPU heeft 8 seconden heeft de GPU nodig om een epoch (batch) te verwerken.	15

Pi- antidos i et al., 2018 (42)	Internatio- nal Confe- rence on Pattern Re- cognition (ICPR) [NB]	Doel van het onder- zoek is onderzoeken of het gebruik CNN bij het automatisch segmenteren bij borstweefsel in 3D MRI-data accuraat is.	N=42 DCE-MRI-series. Per studie zijn 10 series van FLASH 3D coronaal DCE T1- gewogen beelden (1 pre- contrast en 9 post-con- trast). Geen rekening ge- houden met soort patiën- tengroep.	3D borstsegmentatie wordt uitgevoerd door een 2D U-Net model. De 3D vo- lume van de borsten worden gesegmen- teerd door slice-to-slice methode in de axiale, coronaire en sagittale vlak. An- notaties worden uitgevoerd door een ervaren radioloog. Voor validatie wordt een 10-fold cross validatie uitgevoerd.	Uitkomsten testset Transversale vlak: DICE 95.81%, Sensitiviteit 95.94%, Spe- cificiteit 99.34%, Accuraatheid 98.93%. Sagittale vlak DICE 95.90%, Sensitiviteit 95.95%, Spe- cificiteit 99.35%, Accuraatheid 98.93%. Coronale vlak	Het CNN-model dat is gebruikt in dit onderzoek is in staat om borstweefsel te segmenteren. Het CNN-model pres- teert beter in DICE, specificiteit en ac- curaatheid dan andere methoden. Sensitiviteit van de CNN-model lag la- ger dan de Pixel Based methode. Dit kwam doordat de CNN-model had	15
					DICE 95.39%, Sensitiviteit 95.86%, Spe- cificiteit 99.28%, Accuraatheid 98.89%.	over-segmenteert.	
Ha et al., 2019 (40)		Het doel van het on- derzoek is om een volledig automati- sche convolutional neural network (CNN) methode te ontwikkelen voor het segmenteren van FG en back- ground parenchy- mal enhancement (BPE) op MRI-scans.	N= 1114 borstvolumes van 137 patiënten. Borstvolu- mes bestaat uit één borst. T1-weging pre- en post- contrast, T1-weging sub- stractie beelden.	Een 3D CNN-model is ontwikkeld uit een standaard 2D U-Net structuur om de gehele borst en FG op een voxel- wise te segmenteren. Manuale segmen- tatie is uitgevoerd middels een me- thode gebaseerd op eerder uitge- brachte artikelen. Daarna zijn de anno- taties gecontroleerd door een ervaren borst radioloog met acht jaar ervaring. Segmentaties van het voorgestelde mo- del wordt met de ground truth vergele- ken (DICE en Pearson) en hierbij is een five fold cross-validation gebruikt.	Uitkomsten testset Borst segmentatie CNN-model: DICE 0.947, Pearson correlation 0.998. FG-segmentatie CNN-model: DICE 0.813, Pearson correlation 0.975.	De voorgestelde methode: volledig automatische CNN-model kan ge- bruikt worden om FG en BPE te kwan- tificeren. Uit het onderzoek is geble- ken dat het onderzochte CNN-model met hoge precisie FG en BPE kan seg- menteren en kwantificeren.	14
Rosado -Toro et al., 2015 (40)	Academic Radiology [2.267]	In het onderzoek wordt een automa- tische algoritme die de borsten uitlijnt ontwikkeld en ge- test. Hierbij worden fat en water beel- den van MRI-mam- mografieën gebruikt als input van de al- goritme. Automati- sche algoritme is ge- baseerd op machine learning model K- means++ en dyna- mic programming (KDP).	N=266 MRI-slices in trans- versale richting. Scans van 14 patiënten zijn gebruikt (19 slices per patiënt). De 266 MRI-slices zijn beschik- baar in vet en water beel- den.	Met K-means++ worden in de eerste stap de water gewogen beelden geseg- menteerd op contrastverschil in FG, pectorale spier en vetweefsel. Daarna werd op de vet afbeeldingen de grens pectoralis spier gedefinieerd met dyna- mic programming. Op vet afbeeldingen is de pectoralis spier en vetweefsel dui- delijk afgrensbaar. Hierna zijn twee ver- schillende automatische methoden ge- bruikt om het startpunt aan te geven in het sternum (linker- en rechterborst mediale startpunt) en om het vetach- tige weefsel (niet borstweefsel) in de sternum regio te verwijderen, namelijk Fixed sternum removal (FSR) methode en Morphological sternum removal	Uitkomsten testset Automatische segmentatie van bor- sten: 1. DP met FSRM gemiddelde DICE (STD): Tracer #1 0.922 (0.047), Tracer #2 0.910 (0.052) en Tracer #3 0.908 (0.067) 2. KDP met MSRM gemiddelde DICE (STD): Tracer #1 0.897 (0.114), Tracer #2 0.878 (0.123) en Tracer #3 0.899 (0.107) 3. HSF met FSRM gemiddelde DICE (STD): Tracer #1 0.864 (0.079), Tracer #2 0.846 (0.082) en Tracer #3 0.860 (0.083) 4. HSF met MSRM gemiddelde DICE (STD): Tracer #1 0.847 (0.095), Tracer	Gebaseerd op de resultaten is KDP praktische optie voor het segmente- ren van vet en water beelden van de borsten in MRI-mammografieën. KDP presteert beter dan CLG en HSF en brengt bijna dezelfde resultaten als de manuele segmentatie door radiologen en getrainde ROI-tracer (van de bor- sten).	14

				(MSR) methode. FSR heeft een vast startpunt en verwijderd vetachtige weefsel in de 29 kolommen random het startpunt (in MRI-mammografie altijd zelfde positie met de borst RF spoel). MSR verwijderd vetweefsel in het ster- num door het kleinste element met morfologische opening te gebruiken. Tenslotte is nog een laatste borstseg- mentatie mask uitgevoerd. Annotaties zijn gedaan door twee ervaren borst ra- diologen en een persoon die getraind is om de borsten ROI (Region of Interest) te tracen (inter-observer toetsing wordt uitgevoerd). KDP wordt vergeleken met twee andere segmentatie methoden, namelijk: Clustering and local gradient analysis (CLG) en Hessian-based sheet- ness Filter (HSF).	 #2 0.827 (0.100) en Tracer #3 0.835 (0.092). 5. CLG met FSRM gemiddelde DICE (STD): Tracer #1 0.843 (0.107), Tracer #2 0.830 (0.109) en Tracer #3 0.835 (0.119) 6. CLG met FSRM gemiddelde DICE (STD): Tracer #1 0.833 (0.113), Tracer #2 0.817 (0.121) en Tracer #3 0.831 (0.119) 		
Gubern - Mérida et al., 2015 (39)	IEEE Jour- nal of Bio- medical and Health Informatics [5,18, 2018]	Het ontwikkelen van een framework dat de 3D-aard van MRI- mammografieën in acht neemt om au- tomatisch borst- en FG-weefsel te seg- menteren, met doel om borstdichtheid te berekenen.	N=50 MRI-volumes [Geen maligniteit na <2 jaar na ac- quisitie] (gemiddelde leef- tijd 45.84j, range [23-76j]) Trainingsset borstweefsel- segmentatie N=27. Testset FG-segmentatie N=50.	Allereerst is een sternum landmark sys- teem benut om een VOI (Volume of In- terest) om beide borsten te plaatsen, dit ter begeleiding van de borst- en FG- segmentatie. Hierna zijn is een normali- satie uitgevoerd om signaalintensiteiten te standaardiseren in de MRI-volumes. Borstsegmentatie is uitgevoerd met een atlas-based voxel classificatie algoritme. Vervolgens dient borstsegmentatie als input om overige niet-borst vetweefsels te verwijderen, ook worden de MRI-sli- ces opgesplitst in linker en rechter borsthelft. FG-segmentatie vindt plaats met de linker en rechter borsthelft mid- dels het EM-algoritme. Achtereenvol- gend worden skin-folds verwijderd in het gehele MRI-volume om met precisie FG-weefsel te segmenteren. Prestaties van het EM-algoritme zijn berekend door handmatige segmentaties (STAPLE	Uitkomsten testset Borstweefselsegmentatie respectieve- lijke DICE + total overlap (Gemiddelde + std): 0,94 \pm 0,03; 0,96 \pm 0,02 FG-segmentatie DICE + total overlap (Gemiddelde + std): Viewer 1: 0,84 \pm 0,11; 0,77 \pm 0,11 Viewer 2: 0,88 \pm 0,08; 0,93 \pm 0,11 Viewer 3: 0,80 \pm 0,15; 0,95 \pm 0,08 Viewer 4: 0,85 \pm 0,10; 0,77 \pm 0,17 Automatisch: 0,80 \pm 0,13; 0,85 \pm 0,11	Een volledig automatische methode is ontwikkeld om het volume dense borstweefsel te berekenen. Evaluatie laat zijn dat deze aanpak tamelijke re- sultaten oplevert bij segmentatie van borst- en FG-weefsel. Het framework kan als betrouwbaar hulpmiddel wor- den ingezet in borstdichtheidsstudies.	15

				gefuseerde annotaties van vier viewers) te vergelijken met de EM-segmentaties, hieruit is een DICE berekend.			
Ver- burg et al., 2019 (36)	Medical Physics [2,99]	Het vergelijken van een Knowledge-ba- sed system en een Deep-Learning be- nadering ten aan- zien van segmenta- tie van thoraxwand in MRI-series met extreem dense bor- sten.	N=115 T1-w MRI-volumes via de DENSE-trial [179 sli- ces per serie]. Trainingsset: 36 T1-w MRI-volumes. Testset: 79 T1-w MRI-vol- umes.	Handmatige segmentaties (ground truth) in alle MRI-volumes zijn uitge- voerd door een technische fysicus mid- dels annotatiesoftware onder supervisie van een borstradioloog (6 jaar ervaring). In alle MRI-volumes is automatisch een rechthoekige ROI om het borstweefsel gezet. Twee segmentatiemodellen (Knowledge Based Model (KBM) en Deep Learning Model (DLM)) zijn ont- wikkeld en geoptimaliseerd om borst- segmentatie uit te voeren. Evaluatie van de prestaties voor zowel 2D- als 3D-seg- mentatie is zowel kwantitatief (DICE, HD, HD95, FPF, FNF) als kwalitatief be- oordeeld. De kwalitatieve beoordeling is gemaakt door een radioloog welke fou- ten van segmentatiesystemen in vier ca- tegorieën had opgedeeld.	Gemiddelde ± IQR testset DLM: DICE: 0,982(0,006); HD(mm): 12,81(27,28) HD95(mm): 2,58(1,78); FPF: 0,030(0,009) FNF: 0,003(0,009) FNF: 0,003(0,003) KBM: DICE: 0,984(0,006); HD(mm): 24,14(20,69) HD95(mm): 3,37(2,11); FPF: 0,018(0,009) FNF: 0,009(0,011)	Twee automatische methoden zijn ontwikkeld zijn ontwikkeld van de tho- raxwand in MRI-afbeeldingen. Beide segmentatie methoden zijn geëvalu- eerd met 79 MRI-series, goede presta- ties zijn vertoond.	14
Afkorting	en: NB, Niet B	eschikbaar					

	4.2 EVIDENCE TABLET G-WELTSLESEGMENTATIE									
Auteur, jaartal	Tijdschrift (impact- factor)	Doel	Dataset	Methode	Resultaten	Conclusie	Score Check- list			
Ivanovs ka et al., 2019 (41)	Internatio- nal Journal of Compu- ter As- sisted Ra- diology and Sur- gery [NB]	Het doel van dit onderzoek was om een deep learning methode te ontwikkelen en te gebruiken om de borst- dichtheidsschatting in MRI- mammografie data (met sterke artefacten zoals in- homogeniteitsartefact. Deze methode wordt ook geëvalueerd tijdens het on- derzoek.	N=40 MRI-mammogra- fieën met T1-weging zon- der contrast. De borsten uit dit onderzoek werd geclassificeerd door een radioloog volgens BIRADS.	De scans zijn handmatig gesegmenteerd op fi- broglandulair weefsel en borstweefsel (ground truth). Eerst wordt een correctie van artefacten uit- gevoerd bij de dataset. Daarna werd de gegevens van de data vergroot. Als derde stap wordt het borstvo- lume (zonder tepels) gesegmenteerd. Tenslotte wordt de fibroglandulair weefsel gesegmenteerd. Voor het segmenteren wordt een 2D U-Net model gebruikt. Om variëteit in de trainingsdata te krijgen zijn simpele datavergrotingstechnieken gebruikt. In een MRI-scan zijn de linker- en rechterborst van el- kaar gescheiden en wordt beiden gebruikt als trai- ningsdata.	Uitkomsten testset FG-segmentatie met tepelcorrec- tie gemiddelden + std DICE 0.925 ±0.011 Sensitiviteit 0.933 ±0.08 Specificiteit 0.9994 ±0.00005 Borst volume segmentatie gemid- delden + std DICE 0.98 ±0.05	Deze methode kan accuraat segmenteren en is hoog ef- ficiënt om te gebruiken bij een groot epidemiologische data met 1000 deelnemers. Het model is getraind op een moderne GPU en is hierdoor een snel. 8 secon- den heeft de GPU nodig om een enoch te verwerken.	15			
Gubern- Mérida et al., 2015 (39)	IEEE Jour- nal of Bio- medical and Health Informa- tics [5,18, 2018]	Het ontwikkelen van een framework dat de 3D-aard van MRI-mammografieën in acht neemt om automa- tisch borst- en FG-weefsel te segmenteren, met doel om borstdichtheid te bere- kenen.	N=50 MRI-volumes [Geen maligniteit na <2 jaar na acquisitie] (gemiddelde leeftijd 45.84j, range [23-76j]) Trainingsset borstweef- sel-segmentatie N=27 Testset FG-segmentatie N=50	Allereerst is een sternum landmark systeem benut om een VOI om beide borsten te plaatsen, dit ter be- geleiding van de borst- en FG-segmentatie. Hierna zijn is een normalisatie uitgevoerd om signaalintensi- teiten te standaardiseren in de MRI-volumes. Borst- segmentatie is uitgevoerd met een atlas-based voxel classificatie algoritme. Vervolgens dient borstseg- mentatie als input om overige niet-borst vetweefsels te verwijderen, ook worden de MRI-slices opgesplitst in linker en rechter borsthelft. FG-segmentatie vindt plaats met de linker en rechter borsthelft middels het EM-algoritme. Achtereenvolgens worden skin- folds verwijderd in het gehele MRI-volume om met precisie FG-weefsel te segmenteren. Prestaties van het EM-algoritme zijn berekend door handmatige segmentaties (STAPLE gefuseerde annotaties van vier viewers) te vergelijken met de EM-segmentaties, hieruit is een DICE berekend.	Uitkomsten testset Borstweefselsegmentatie respec- tievelijke DICE + total overlap (Ge- middelde + std): $0,94 \pm 0,03$; $0,96 \pm 0,02$ FG-segmentatie DICE + total over- lap (Gemiddelde + std): Viewer 1: $0,84 \pm 0,11$; $0,77 \pm 0,11$ Viewer 2: $0,88 \pm 0,08$; $0,93 \pm 0,11$ Viewer 3: $0,80 \pm 0,15$; $0,95 \pm 0,08$ Viewer 4: $0,85 \pm 0,10$; $0,77 \pm 0,17$ Automatisch: $0,80 \pm 0,13$; $0,85 \pm 0,11$	Een volledig automatische methode is ontwikkeld om het volume dense borst- weefsel te berekenen. Eva- luatie laat zijn dat deze aan- pak tamelijke resultaten op- levert bij segmentatie van borst- en FG-weefsel. Het framework kan als be- trouwbaar hulpmiddel wor- den ingezet in borstdicht- heidsstudies.	15			
Kumari et al.,	IEEE India Council In- ternational	Het ontwikkelen van een automatische Machine Learning methode om met	N=50 MRI-volumes 160 slices per MRI-vo- lume, Transversale T1-w	Allereerst is de volgende processing uitgevoerd: ghost-artefacten zijn verwijderd middels morfolo- gische (erosie en dilatie) operaties; spikkelruis in de	Uitkomsten testset Respectievelijke DICE, JAC en DI per class [ACR] per MRI-volume	De voorgestelde methode (GaGMM) biedt een verstel- bare en snelle manier voor	13			

4.2 EVIDENCE TABLE FG-WEEFSELSEGMENTATIE

2017 (37)	Confe- rence (IN- DICON) [NB]	hoge precisie fibroglandu- lair borstweefsel te seg- menteren middels een Gamma Gaussian Mixed (GaGMM) model.	non-fatsat afbeeldingen met onder andere Class A, class B, class C en class D borstdensiteiten [ACR].	MRI-slices is geminimaliseerd d.m.v. anisotropische difussie; signaal intensiteitsinhomogeniteit in de MRI-slices zijn verwijderd middels een set-level me- thode. Na de preprocessing is automatische fi- broglandulair borstweefsel (FG) segmentatie middels GaGMM uitgevoerd over de 50 MRI-volumes. Presta- ties (DICE) van de FG-segmentatie zijn berekend door overlapberekening van de automatische segmenta- ties met een handmatige segmentaties (gouden stan- daard, geverifieerd door een radioloog).	(gemiddelde) van GaGMM Class A: 0,9381; 0,8834; 0,1166 Class B: 0,8344; 0,8739; 0,1261 Class C: 0,9873; 0,9113; 0,0887 Class D: 0,8481; 0,8525; 0,1475 Gemiddelde DICE, JAC en DI per onderzocht model Gamma Mixture Model: 0,6191; 0,5486; 0,4514 GaGMM (nieuw): 0,9177; 0,8819; 0,1181	het segmenteren van FG- weefsel in MRI-mammogra- fieën. Waar GaGMM een voordeel in biedt is de ruime informatie en toe- passingsmogelijkheid in verschillende borstklassen volgens ACR. Ook heeft het segmentatiemodel geen voorafgaande informatie nodig over borstanatomie voor de FG-segmentatie.	
Ha et al., 2019 (39)	Journal of Digital Imaging [NB]	Het doel van het onderzoek is om een volledig automa- tische convolutionial neural network (CNN) methode te ontwikkelen voor het seg- menteren van FG en back- ground parenchymal en- hancement (BPE) op MRI scans.	N= 1114 borstvolumes van 137 patiënten. T1-weging pre- en post- contrast, T1-weging sub- stractie beelden	Een 3D CNN-model is ontwikkeld uit een standaard 2D U-Net structuur om de gehele borst en FG voxel- wise te segmenteren. Manuale segmentatie is uitge- voerd middels een methode gebaseerd op eerder uit- gebrachte artikelen. Daarna zijn de annotaties gecon- troleerd door een ervaren borst radioloog met acht jaar ervaring. Segmentaties van het voorgestelde model wordt met de ground truth vergeleken (DICE en Pearson) en hierbij is een five fold cross-validation gebruikt.	Uitkomsten testset Borst segmentatie CNN-model: Gemiddelde DICE 0.947, Pearson correlation 0.998 FG-segmentatie CNN model: Gemiddelde DICE 0.813, Pearson correlation 0.975.	De voorgestelde methode: volledig automatische CNN- model kan gebruikt worden om FG en BPE te kwantifi- ceren. In het onderzoek is gebleken dat de CNN-mo- del hoge prestaties heeft bij het kwantificeren van FG en BPE. Met een grotere data- set kan het model zich aan- nemelijk verbeteren.	14
Zhang Y et al., 2019 (43)	Academic Radiology [2.267]	In het onderzoek wordt een deep learning segmentatie methode gemaakt en ge- toetst op het segmenteren van borstweefsel en fi- broglandulair weefsel (FG) bij pre-contrast T1-gewo- gen beelden zonder vet- suppressie.	Trainingsdata N= 276 pa- tiënten. Testset N= 224 scans. 224 scans komen van 28 patiënten, die allemaal gescand zijn door 4 ver- schillende MRI-scanners. Linker- en rechterborst wordt gesplitst	U-Net is gebruikt voor het segmenteren van borst- weefsel en daarna FG. Ground truth was gedaan door template-based automatische borst segmenta- tie methode. De segmentatie is daarna geëvalueerd door een ervaren radioloog met 12 jaar ervaring. U-Net is daarna getraind, gevalideerd en een presta- tiemeting is uitgevoerd (DICE).	 Bij training dataset Borst segmentatie: DICE 0.83-0.98 (gemiddelden 0.95 ±0.02). FG-weefselsegmentatie: DICE 0.73-0.97 (gemiddelden 0.91 ±0.03). Bij Test dataset: Borst segmentatie: Gemiddelde DICE 0.86 ±0.05 FG-weefselsegmentatie: Gemiddelde DICE 0.83 ±0.06 	Deep learning model: U-Net kan gebruikt worden om ac- curate resultaten te verkrij- gen bij het segmenteren van borstweefsel en FG bij MRI onderzoeken. Dit kan een efficiënt en betrouw- bare methode zijn om een groot aantal MRI onder- zoeken te segmenteren voor het kwantitatief analy- seren van borstdichtheid.	14
Afkortinge	en: NB. Niet Be	eschikbaar					

	4.3 EVIDENCE TABLE TUMOR SEGMENTATIE						
Auteur, jaartal	Tijd- schrift [impact- factor]	Doel	Dataset (N=)	Methode	Resultaten	Conclusie	Score Checklist
Wang et al., 2020 (53)	IEEEE ac- cess [NB]	Om 3D Clifford temporal image (CTI) construction mo- del samen met 3D Clifford analytic signal (CAS) te ge- bruiken bij 4D spatial-tempo- ral DCE-MRI-data. Hiermee kan informatie van DCE-MRI sequenties automatisch ge- combineerd worden. Daarna worden twee succesvolle Fully Convolutional Network (FCN) zijn gebruikt bij het ontwikkelen van een para- meter-vrije end-to-end fra- mework en deze wordt ge- test of het geschikt is voor laesie segmentatie.	N=236 MRI-mammografie onderzoeken. 184 voor training, 20 voor validatie en 32 (12 van QIN DCE) voor testdata TBD: Dataset van zieken- huis van 2014 tot 2018. (N= 224) QIN DCE: openbare dataset (N=12)	Voorgestelde methode: twee FCN-segmentatiemodel- len werden gebruikt om borst laesies te segmenteren. De eerste FCN model segmenteert borstweefsel en de tweede FCN wordt getraind om borstlaesies te seg- menteren. De voorgestelde methode wordt vergeleken met: DCE- MRI subtraction image based (DS) laesie segmentatie model (FCN model) en DeepMedic (DM), een CNN mo- del. DM wordt ook met whole breast mask getest (DM+BM). Voxel-wise laesie annotaties zijn gemaakt door een ervaren radioloog en gecontroleerd door een ander radioloog.	Voorgesteld methode FCN-tumor- segmentatie Gemiddelde DICE: 0.813 QIN DCE DICE: 0,718 TBD DICE: 0,870. DS Gemiddelde DICE: 0.698 QIN DCE DICE: 0.618 TBD DICE:0.745 DM Gemiddelde DICE: 0.700 QIN DCE DICE: 0.542 TBD DICE: 0.795 DM+BM Gemiddelde DICE: 0.728 QIN DCE DICE: 0,614 TBD DICE: 0,797. Voorgestelde methode was de DICE 11.5%, 11.3% en 8.5% hoger dan die van DS, DM, en DM+BM	Het gebruik van 3D Clifford Temporal Image (CTI) bij twee FCN-segmentatiemo- dellen heeft betere prestaties geleverd dan segmentatie- modellen die nu gebruikt worden voor het segmente- ren van laesies.	15
Vogl et al., 2020 (46)	Euro- pean Ra- diology Experi- mental van Eu- ropean Society of Radio- logy [3.962]	In het onderzoek wordt on- derzocht of een automa- tisch segmentatiemodel in staat is om laesies in de borst te segmenteren en classificeren. Dataset van PET- en MRI-scans is ge- bruikt om de prestaties van segmentatie en classificatie te onderzoeken.	N=34 patiënten waarvan 12 patiënten een benigne laesie hebben en 22 ma- ligne laesie hebben. Alle 34 patiënten onder- gingen PET, CT- en MRI- onderzoeken	Random Forest classifier (RF) model (ML) werd ge- bruikt om de laesie segmentatie uit te voeren. Hier- bij werd op voxels geclassificeerd op binaire labels: label 1 (laesie) en label 0 (non-laesie). Annotaties zijn manueel gedaan door een ervaren radioloog (3 jaar ervaring). Training werd gedaan op MRI met PET, MRI zonder PET, MRI zonder DWI en MRI zon- der PET en DWI. Daarna zijn twee verschillende evalutaies methoden gebruik voor het evalueren van de contributie van specificaties in de scans GI-methode: Hiermee wordt de gemiddelde hoe- veelheid informatie gemeten door gebruik te	Gemiddelde DICE, gemiddelde sensitiviteit van GI-methode MRI met PET, DWI 0.607, 0.661 MRI zonder PET 0.608, 0.739 MRI zonder DWI 0.573, 0.669 MRI zonder PET, DWI 0.577, 0.658 Gemiddelde DICE, gemiddelde sensitiviteit van mRMR methode: MRI met PET, DWI 0.665, 0.743 MRI zonder PET 0.618, 0.748 MRI zonder DWI 0.601, 0.686 MRI zonder PET, DWI 0.584, 0.613	De voorgestelde methode detecteert en segmenteert automatisch potentiele ma- ligne regios. De methode maakt de beoordeling op segmentatie en classificatie van verschillende modali- teiten mogelijk. Het model kan gebruikt worden als hulpmiddel voor het classi- ficatie en detectie van borstlaesies.	14

				maken van Gini index splitting criteria en rangschikt de bijdrage van elk feature als onderdeel van een multivariate patroon. mRMR methode: informatie wordt gerangschikt op basis van relevantie en overbodigheid van de featu- res.			
Maicas et al., 2017 (37)	2017 IEEE 14th In- ternatio- nal Sym- posium on Bio- medical Imaging (ISBI 2017) [NB]	In het onderzoek wordt een nieuw methode GOCS-DLP getest bij het segmenteren van tumorweefsel in bor- sten. In GOCS-DPL wordt een deep learning segmen- tatiemodel gebruikt om de segmentatie te doen. Dit wordt vergeleken met stan- daard CNN- segmentatie- model.	N=117 patiënten 58 patiënten voor training en 59 voor de uiteinde- lijke test. In de 117 scans zijn 141 laesies geanno- teerd: 46 benigne laesie en 95 maligne laesie.	Voorgestelde methode wordt getest met MRI se- ries (T1-gewogen en T2-gewogen). Alle laesies wer- den geannoteerd door een MBB'er (ground truth). De dataset is willekeurig verdeeld in training en test data. De segmentatiemodellen GOCS-DLP, CNN, LOCS en GOCS-MS werden getraind en gevali- deerd met de dataset. GOCS-DLP, CNN en LOCS zijn segmentatiemodellen met deep learning. Ook wordt de data getraind op manuele detectie (hand- matig invoeren) en automatische detectie (DL-de- tectiemodel).	Gemiddelde DICE trainingsdata, gemiddelde DICE test dataset bij GOCS-DLP: Met automatische detectie: 0.80 (± 0.11) , 0.77 (± 0.14) . Met manuele detectie: 0.79 (± 0.13) , 0.77 (± 0.13) . Gemiddelde DICE trainingsdata, gemiddelde DICE bij test dataset CNN: Met automatische detectie: 0.69 (± 0.16) , 0.68 (± 0.19) . Met manuele detectie: 0.66 (± 0.16) , 0.66 (± 0.18) . Gemiddelde DICE trainingsdata, gemiddelde DICE trainingsdata, gemiddelde DICE test dataset LOCS: Met automatische detective: 0.64 (± 0.16) , 0.62 (± 0.15) . Met manuele detectie: 0.61 (± 0.15) , 0.59 (± 0.17) .	De resultaten laat zien dat de onderzochte methode (GOCS-DLP) significant ac- curater is dan andere me- thode (standaard CNN) bij het volledig automatische segmentatie. In de toe- komst is het plan om de voorgestelde methode toe te passen in andere seg- mentatieproblemen bij me- dische beeldvorming.	15
Li et al., 2019 (48)	Medical Image Compu- ting and Compu- ter As- sisted In- terven- tion – MICCAI 2019 [NB]	In het onderzoek wordt een gefuseerde CNN geba- seerde netwerk voorge- steld bij het processing en segmentatie van tumor- weefsel in MRI-mammogra- fie onderzoeken. MRI- mammografie is gekozen, omdat MRI-mammografie onderzoeken lastig zijn om te segmenteren.	N= 313 patiënten In de dataset komen twee MRI-sequenties voor: pre- contrast T2-gewogen met vetsuppressie (T2w) en post-contrast T1-gewogen met vetsuppressie scans (T1c).	Annotaties werden gedaan door twee ervaren radi- ologen. De centrale slices waar de tumor het grootst waren, werden geannoteerd op axiale, co- ronale en sagittale vlak. Voorgestelde methode: Supervised Cross-Modal Deep Representation Learning (voorgesteld) werd vergeleken met verschillende deep learning seg- mentatiemodel: FuseOriginUnet, FuseNet, FuseUnet en U-Net. Alleen U-Net is getraind op T1c. De andere metho- den zijn getraind op T1c en T2w slices. Bij FuseNet en Unet is nog een Spatial attention (SA) blok aan het netwerk geplakt. De blok is	DICE, Sensitiviteit bij T1c-gewogen beelden testset U-Net 73.3%, 79.5% U-Net + SA 75.8%, 82.6% DICE, Sensitiviteit bij gefuseerde beelden (T1c+T2w) testset FuseNet 52.8%, 54.9% FuseOriginUNet 75%, 82.1% FuseUnet 74.9%, 82.2% FuseUnet+SA 76.3%, 83.4% voorgesteld 77.6%, 84.4%	De voorgestelde methode: Supervised Cross-Modal Deep Representation Lear- ning is effectief bij het seg- menteren van tumorweef- sel bij twee verschillende wegingen volgens de resul- taten van dit onderzoek. De methode kan worden ge- bruikt voor verschillende scenario's voor het seg- menteren van medische beelden, zoals combinatie	13

				ontworpen om informatie van de scans (T1c en bij FuseNet ook T2w) te extraheren.		van twee verschillende mo- daliteiten en zal in de toe- komst meer onderzocht worden voor de toepassing hierin.	
Vesal et al., 2018 (45)	Bild- verar- beitung für die Medizin 2018 [NB]	Onderzoek naar prestaties van automatische segmen- tatie van tumorweefsel in MRI-mammografieën d.m.v. Unsupervised Ma- chine Learning- segmenta- tiemodel. Namelijk: Gaus- sian mixture model (GMM) clustering; K-means cluste- ring en Marker-controlled Watershed Transformation (MCWT) en vergelijking van metrics.	N=106 laesies (42=be- nigne, 64=maligne) over 80 patiënten. Dataset MRI-slices onbe- kend.	De laesies zijn handmatig gesegmenteerd (ook in 2D) door twee radiologen. Een 2D-slice uit dataset is gekozen op basis van deze segmentatie, waarna een ROI om laesies is getekend. Als preprocessing is gebruik gemaakt van CLAHE om het afbeeldings- contrast te verhogen binnen de MRI-slices. Hierna is automatische segmentatie verricht middels GMM-clustering, K-means clustering en Marker- controller watershed transformation en zijn de prestaties (DICE, JAC, HD, PRE, RE) berekend.	Gemiddelde ± std resultaten test- set: Dice: K-means: 0,732±0,206; GMM: 0,746±0,180; MCWT: 0,786±0,172 JAC: K-means: 0,612±0,209; GMM: 0,623±0,193; MCWT: 0,679±0,21w7 HD (mm): K-means: 2,292±1,05; GMM: 2,275±1,08; MCWT: 2,265±1,24 PRE: K-means: 0,805±0,243; GMM: 0,855±0,213; MCWT: 0,866±0,199 RE: K-means: 0,702±0,204; GMM: 0,697±0,195; MCWT: 0,752±0,250	Laesie segmentatie is een belangrijke stap voor de ka- rakterisering van tumoren in MRI-mammografieën In totaal zijn drie Machine Learning segmentatiemo- dellen onderzocht ten aan- zien van tumorweefsel seg- mentatie, hieruit bleek MCWT het beste presteren.	15
Spuhler et al., 2019 (47)	Magne- tic Reso- nance in Medi- cine [NB]	Onderzoek is verricht naar een automatische CNN-be- nadering om tumorweefsel te segmenteren in MRI- mammografieën. Deze seg- mentatie diende als input voor een reeds gepubli- ceerd classificatiesysteem om de kans op Sentinel Node Metastasen te voor- spellen.	CNN: N=263 laesies Training set: n=263 lae- sies Testing set: n=54 laesies Dataset MRI-slices onbe- kend. MRI-afbeeldingen zonder contrast, en 2; 4; 6; 8 minuten na contrast toediening.	Laesies zijn handmatig gesegmenteerd op de eerste post-contrast MRI-scan door drie radiologen (R1, R2, R3). R1 en R3 hebben laesies onafhankelijk in Dit onderzoek gesegmenteerd (d.m.v. een ROI), R2 heeft de laesies in de testset gesegmenteerd. U-net architecture CNN-model is benut voor segmentatie van de laesies, hiervoor zijn de slices uit de MRI- pre- en postcontrast series voor benut. Training van het CNN-model is gedaan met de trainingsset in minibatches, ook werden deze at random geaug- menteerd om de testset te vergroten. Prestaties van het CNN-model zijn gemeten door de segmen- taties van de radioloog te vergelijken met de seg- mentatie van het CNN-model.	Dice (gemiddelde + std) over test- set CNN getraind met ROI's R1, dice in vergelijking met R1, R2, R3: 0,71 ± 0,16; 0,61 ± 0,17; 0,67 ± 0,18 CNN getraind met ROI's R3, dice in vergelijking met R1, R2, R3: 0,61 ± 0,18; 0,64 ± 0,22; 0,67 ± 0,20	De U-Net benadering ten aanzien van tumorweefsel- segmentatie bereikte ver- gelijkbare precisie in verge- lijking met de handmatige gesegmenteerde laesies door R1, R2 en R3. Dit des- ondanks een relatief kleine trainingsset, zijn vergelijk- bare resultaten behaald met reeds gepubliceerde artikelen.	15
Zhang J et al., 2019 (50)	IEEE Trans- actions on Medi- cal	Een mask-guided hiërar- chisch deep-learning seg- mentatiemodel is ontwik- keld om coarse-to-fine seg- mentation voor	N=272 MRI-volumes (Trai- ning set D1 n=224, Test set D2 n=48) Wegingen: T1-w FatSat, 3x T1-w post-contrast.	Zes radiologen ([6-20] jaar ervaring) hebben onaf- hankelijk alle aanwezige laesies in D1 handmatig gesegmenteerd (=ground truth). Om inter-rater study te beoordelen hebben vier radiologen ([0-22] jaar ervaring) laesie in de testset handmatig	[Gemiddelde + std testset] DICE: 0.72 ± 0.24 SEN:0.75 ± 0.23 PPV: 0.77 ± 0.21	In dit onderzoek is een Mask-guided hiërarchisch deep-learning- segmenta- tiemodel ontwikkeld voor borst tumorsegmentatie.	16

	Imaging [9,71]	borsttumoren uit te voe- ren. Dit met het doel om meerdere tumoren in MRI- mammografieën te seg- menteren, deze te lokalise- ren en moleculair te classi- ficeren.	Data vergaard tussen Sep- tember 2007 en Juni- 2009.	gesegmenteerd welke voor biopt zijn uitgesneden (=ground truth). De CNN-segmentatiemodellen zijn getraind met D1 (gesplitst in twee subsets middels een two-fold cross validation schema). Segmentatie van tumorweefsel is uitgevoerd aan de hand van twee FCN (U-Net) modellen. FCN-model-1 heeft een ruwe segmentatie van laesies gemaakt aan de hand van MRI-mammografie slices en een borst- ROI, FCN-model-2 heeft de ruwe segmentatie als input gebruikt om de segmentatie rondom de lae- sie te verfijnen. Prestaties van FCN-model-1 en FCN-model-2 zijn gemeten door de handige seg- mentaties te vergelijken met de automatische seg- mentaties, hieruit is een dice berekend.		Door middel van gegene- reerde borst maskers kan door middel van 2 FCN's ac- curaat borst tumorweefsel gesegmenteerd worden.	
Piantidosi et al., 2019 (49)	IEEE CBMS [NB]	Het exploiteren van Deep Learning om automatische borst tumorsegmentatie in DCE-MRI mogelijk te maken met een architectuur die leert van afbeeldingen op zeer specifieke tijdspunten. Dit aan de hand van het voorstel 3TP U-Net.	N=35 DCE-MRI 4D-volu- mes (gemiddelde leeftijd 40j, range [16-69j], laesies in deze volumes zijn his- tologisch bewezen). DCE-MRI-volumes be- staan uit 1 serie pre-CM, hierna 9 series post-CM.	Handmatige segmentaties zijn uitgevoerd door een radioloog om zo de ground truth te bepalen. Een breast mask wordt benut om overbodige weefsels buiten de borst te verwijderen. Hierna worden per MRI-volume slices geëxtraheerd uit de pre-contrast (t0) scan en respectievelijk twee (t1-w) en zes (t2- w) minuten na contrasttoediening. Deze slices wor- den gefuseerd in een drie channel (3TP) afbeelding welke als input wordt gevoerd aan een U-Net CNN- model voor automatische segmentatie van laesies. Prestaties van het 3TP U-Net model zijn berekend door de ground truth segmentaties met de auto- matische segmentaties te vergelijken, dit middels een 10-cross validatie schema.	Gemiddelde + std testset 3TP: DICE: 0,61 ± 0,12 ACC: 0,99 ± 0,01 SPE: 1,0 SEN: 0,68 ± 0,09 10TP DICE: 0,58 ± 0,16 ACC: 0,99 ± 0,01 SPE: 1,0 SEN: 0,54 ± 0,19	Aangetoond is dat een 3TP U-Net CNN-model effectief benut kan worden om het borst tumorweefsel seg- mentatie proces te verbete- ren, dit op een relatief sim- pele wijze. Prestaties op grootschalig vlak dienen met dit segmentatiemodel uitgevoerd te worden om de toepasbaarheid in kaart te brengen.	15
Lu et al., 2019 (51)	IEEE ICASSP [2,31]	Het ontwikkelen van een Deep Learning borstkanker detectiemodel. Dit model bestaat uit zowel verschil- lende CNN-architecturen [Vgg 16, ResNet-50, Incep- tion V3 & DenseNet] voor zowel classificatie als seg- mentatie van borst tumor- weefsel.	N=67 MRI-volumes (High- Riskpatiënten, in totaal 8,132 slices) MRI-volumes bevatten: T1-W, T2-W, DWI- en DYN-series. Alle slices zijn uitgelijnd op dezelfde lo- catie. 6.000 slices zijn benut voor classificatie. 1.800 slices zijn benut voor seg- mentatie.	Als preprocessing zijn de MRI-volumes genormali- seerd aan de hand van de gemiddelden signaalin- tensiteiten en standaarddeviatie. Na training van [Vgg 16, ResNet-50, Inception V3 & DenseNet], middels de trainingsset, van zowel de classificatie als segmentatiemodellen zijn deze toe- gepast op de verificatie set. MRI-slices uit de vier voornoemde series zijn als bij het respectievelijke CNN-classificatiemodel voor beoordeling op aan- wezigheid van een laesie. Indien een laesie aanwe- zig was werden de slices als input ingevoerd bij het respectievelijke U-Net CNN-segmentatiemodel,	<i>Gemiddelde DICE-waarden testset</i> DICE U-Net+Init+Rf: - Vgg16: 0,860; - ResNet-50: 0,847; - Inception V3: 0,839; - DenseNet: 0,865. DICE U-Net+Init: - Vgg16: 0,847; - ResNet-50: 0,833; - Inception V3: 0,821; - DenseNet: 0,849.	Een Deep Learning frame- work is ontwikkeld om radi- ologen te ondersteunen in het borstkanker diagnose proces. Dit geschiedt d.m.v. U-Net CNN-classificatie en segmentatiemodellen aan de hand van vier MRI-we- gingen welke gefuseerd worden voor analyse. Het vier-modes netwerk is in staat om met precisie	12

	Data is verdeeld in een	waarbij de laesie automatisch werd gesegmen-	DICE U-Net:	tumorlaesies kunnen seg-	
	Training set, verificatie	teerd. Prestaties van de CNN-segmentatiemodellen	- Vgg16: 0,832;	menteren en classificeren.	
	set en testing set in ver-	is gemeten door de automatische segmentaties te	- ResNet-50: 0,829;		
	houding 8:1:1.	vergelijken met handmatige segmentaties, dit	- Inception V3: 0,815;		
		d.m.v. de DICE.	- DenseNet: 0,836.		
Afkortingen: NB, Niet Beschikbaar					