

Artificial intelligence in rectal cancer

Citation for published version (APA):

van Griethuysen, J. (2021). *Artificial intelligence in rectal cancer*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20211029jg>

Document status and date:

Published: 01/01/2021

DOI:

[10.26481/dis.20211029jg](https://doi.org/10.26481/dis.20211029jg)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

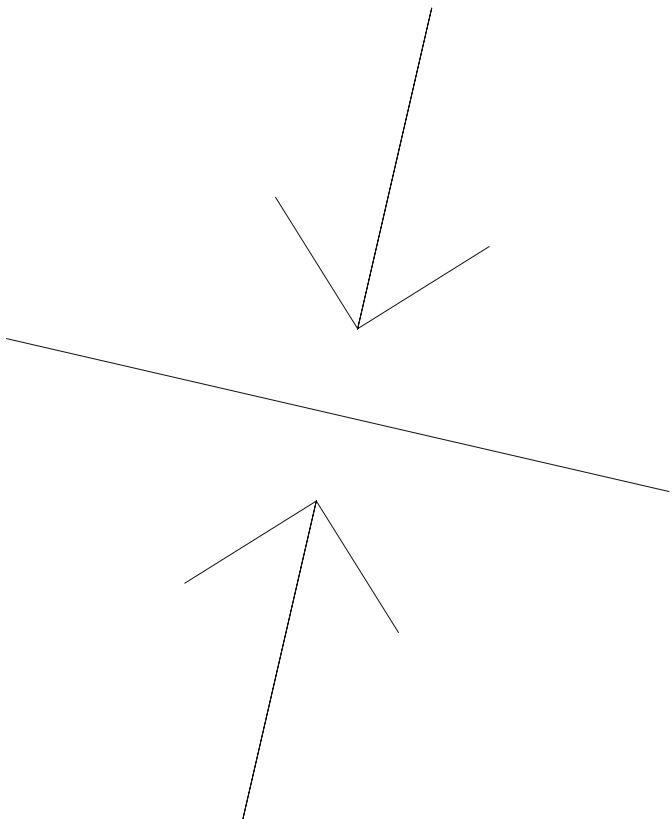
Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary / samenvatting



Summary

There are several key steps in the radiomics workflow, each associated with their own challenges that need to be overcome on the road towards implementing artificial intelligence tools in the clinical workflow of radiology. The goal of this thesis is to investigate these key steps and challenges, using the application of radiomics in rectal cancer as a clinical example.

Segmentation

Chapter 2 focuses on lesion segmentation, an important first step to allow extraction of quantitative features from a given tumor lesion on medical imaging. We investigated the potential of Deep Learning for fully-automated segmentation of rectal tumors on primary staging MRI, using a combination of T2-weighted imaging and Diffusion Weighted Imaging (DWI) sequences. This initial study used a fairly simple patch-based approach, where the network was trained to predict whether a voxel was part of the rectal tumor, based on a small surrounding region. By extracting and classifying patches for all voxels in the input image, a segmentation could be inferred. In the test set we achieved a high performance with Dice Similarity Coefficients (DSC; indicating the spatial overlap between two segmentations on a scale from 0 to 1) between network and manual segmentations of 0.68–0.70, though performance was lower than the agreement between two radiologists performing manual segmentations (0.83 DSC). Moreover, the patch-based approach was hampered by slower inference and low spatial awareness. This was also reflected by the results showing that the network struggled with larger Field-of-View (FOV) acquisitions, erroneously classifying structures near the edge of the FOV as tumor.

More advanced deep learning networks, such as the U-Net are able to process the entire image in a single inference, greatly increasing speed and spatial awareness. In Chapter 3, we optimized the network from Chapter 2 to address these limitations, using a U-Net with additional attention gating, which helps the network to focus on the area of interest. Furthermore, we explicitly investigated the influence of scan quality and tumor complexity on network performance and interreader agreement in a large multicenter dataset with a high variation of acquisition protocols and scan quality. Though performance of the AI model in the test set was still inferior to agreement between radiologists, with DSCs of 0.67 vs. 0.75, the difference was markedly smaller compared to the findings in Chapter 2. Tumor complexity had the largest influence on network performance, with more heterogeneous and irregular tumors resulting in lower DSC scores. Interestingly, tumor complexity had a similar negative effect on the agreement between two radiologists, indicating that these tumors are inherently difficult to segment for computers but also for experienced radiologists. The signal-to-noise on DWI also affected network performance, though to a lesser extent than tumor complexity.

Image Quality

As also highlighted in Chapter 3, image quality is an important prerequisite for Radiomics and AI research. In Chapter 4, we addressed this challenge focusing specifically on the acquisition of DWI of the rectum. Rectal DWI can suffer from significant artefacts caused by the presence of intraluminal gas in the rectum which can have a severe negative impact on diagnostic evaluations. We investigated the application of a preparatory micro-enema shortly prior to acquisition to reduce the amount of intraluminal gas. We observed a marked reduction of significant artefacts from 24% in scans acquired without a micro-enema to <4% in scans performed after application of the micro-enema, showing that even a fairly simple intervention may significantly improve DWI scan quality.

Feature extraction and modelling

In Chapter 5, we addressed the challenge of reproducibility in feature extraction. To this end, we developed PyRadiomics, an easy-to-use open source package for radiomics feature extraction. It is developed in Python, a popular programming language used by many researchers investigating the application of radiomics in medical imaging, coupled with extensions in C for high-performance feature extraction.

Finally, in Chapter 6, we put the radiomics workflow to the test, investigating the application of radiomics to predict the response to neoadjuvant chemoradiotherapy in rectal cancer patients, using only the pre-therapy staging MRI. We compared the performance of a radiomics-based prediction model to the performance of expert radiologists who predicted the response to chemoradiotherapy based on their visual interpretation of the images. The radiomics models achieved a promising performance, with AUCs of 0.69-0.79, which were comparable to the performance of the visual predictions by the expert radiologists. The radiomics model was built using different types of tumor segmentation, performed manually by either expert radiologists or non-expert readers, but also semi-automatically using a basic segmentation algorithm. Interestingly, radiomics performance was similar when using the expert or non-expert manual segmentations, but significantly poorer when using the semi-automatic segmentation algorithm. This highlights the need for better automatic segmentation support tools, such as those addressed in Chapters 2 and 3.

Samenvatting

In het proces van radiomics zijn enkele belangrijke stappen te onderscheiden, elk geassocieerd met uitdagingen die moeten worden overwonnen op de weg naar klinische implementatie van kunstmatige intelligentie (ook wel "artificiële intelligentie" of "AI") in de dagelijkse praktijk van de radiologie. Het doel van deze thesis is het onderzoeken van deze stappen en de bijbehorende uitdagingen, gebruik makende van de toepassing van kunstmatige intelligentie bij endeldarmkanker als klinisch voorbeeld.

Segmentatie

Om kwantitatieve analyse van bijvoorbeeld een tumor laesie mogelijk te maken, is segmentatie van de betreffende laesie op beeldvorming een eerste belangrijke noodzakelijke stap. In Hoofdstuk 2 werd de potentie van AI (Deep Learning) voor het volledig geautomatiseerd segmenteren van tumoren in de endeldarm onderzocht. Hiervoor werd gebruik gemaakt van een combinatie van T2-gewogen en diffusie gewogen imaging (DWI) sequenties die beide onderdeel uit maken het standaard MRI protocol voor endeldarmkanker. In deze initiële studie werd gebruik gemaakt van een vrij eenvoudig netwerk, waarbij per voxel binnen het MRI plaatje werd voorspeld of deze wel of niet tot het tumorgebied behoorde op basis van analyse van een kleine omliggende regio. Een volledige segmentatie van de tumor werd op deze manier verkregen door dit proces te herhalen voor elke voxel in de MRI scan. Toepassing van dit netwerk op een onafhankelijke test set toonde een goed resultaat, met een Dice Similarity Coefficient (DSC; een maat om de spatiale overlap tussen 2 segmentaties weer te geven op een schaal van 0 tot 1) tussen het AI-model en manuele segmentatie door een radioloog van 0.68-0.70. Dit resultaat was echter nog niet geheel vergelijkbaar met de overeenstemming tussen manueel verkregen segmentaties van twee ervaren radiologen (DSC 0.83). Bovendien werd het netwerk gelimiteerd door de regio-gebaseerde methode, wat leidde tot trage segmentatie en weinig ruimtelijk inzicht. Deze nadelen zijn ook herkenbaar in de resultaten, waarbij het netwerk bij beelden met een groter afgebeeld gebied (Field-Of-View, FOV) foutief structuren aan de rand van de afbeeldingen classificeerde als "tumor".

Meer geavanceerde Deep Learning AI netwerken, zoals het U-Net, zijn in staat een afbeelding in zijn geheel te verwerken en zijn zodoende in staat meer ruimtelijk inzicht te tonen. Om de limitaties van het netwerk uit Hoofdstuk 2 te overkomen hebben wij dan ook gebruik gemaakt van een dergelijk U-Net in Hoofdstuk 3, met de toevoeging van "attention-gating" bedoeld om het netwerk te helpen focussen op het belangrijkste doelgebied binnen de scan: de endeldarm en het omliggende vetweefsel. Naast een meer geavanceerde architectuur van het netwerk hebben we ook expliciet de invloed van beeldkwaliteit en tumor complexiteit op de accuratesse van de segmentaties onderzocht. Hoewel de accuratesse van het netwerk ook in dit onderzoek lager uitviel in vergelijking met de overeenkomst tussen ervaren radiologen onderling (DSC 0.67 versus 0.75), was het verschil duidelijk kleiner dan in Hoofdstuk 2. Tumor complexiteit had de grootste invloed op de segmentatie accuratesse, waarbij meer heterogene en grillig gevormde tumoren zowel voor het AI netwerk als voor de radiologen resulteerden in lagere DSCs. Dit leert ons dat deze meer complexe tumoren inherent lastiger zijn om te segmenteren. De beeldkwaliteit (signaal-ruis verhouding) van de DWI-beelden was ook van invloed op de accuratesse van het netwerk, al was dit effect minder uitgesproken in vergelijking met de tumor complexiteit.

Beeldkwaliteit

Zoals ook beschreven in Hoofdstuk 3, is goede beeldkwaliteit een belangrijke vereiste voor onderzoek op het gebied van Radiomics en kunstmatige intelligentie. In Hoofdstuk 4 hebben we deze uitdaging onderzocht in het specifieke geval van beeldartefacten op DWI sequenties van de endeldarm. Bij deze beelden kunnen ernstige susceptibiliteitsartefacten optreden die worden veroorzaakt door de aanwezigheid van lucht in de endeldarm, welke kunnen leiden tot substantieel verslechterde beoordeelbaarheid van deze beelden. In deze studie hebben wij onderzocht of het toepassen van een micro-klysma leidt tot een verminderde hoeveelheid intra-luminaal lucht en dus tot een afname in de aanwezigheid van artefacten. Wij hebben kunnen constateren dat deze toepassing inderdaad leidde tot een duidelijke vermindering van klinisch significante artefacten, met een reductie tot <4% na toediening van het micro-klysma versus 24% in scans vervaardigd zonder micro-klysma. Dit toont aan dat zelfs een vrij eenvoudige interventie kan leiden tot een significante verbetering van beeldkwaliteit.

Feature extractie en data modellering

Om de reproduceerbaarheid van radiomics-onderzoek te vergroten hebben wij een open-source software pakket ontwikkeld, genaamd *PyRadiomics*. In Hoofdstuk 5 introduceren wij dit pakket. Het is ontwikkeld in Python, een populaire programmeertaal bij onderzoekers in het veld van medische beeldvorming. Bovendien is het specifiek ontwikkeld met het oog op het gebruik door een gemeenschap van radiomics-onderzoekers, met extra nadruk op de leesbaarheid van de broncode en uitgebreide documentatie. Om een hoge doorvoersnelheid van de feature extractie te garanderen, zijn enkele stukken van het pakket in C-code geschreven.

Tot slot hebben we het proces van radiomics getest in Hoofdstuk 6 in een klinische studie naar endeldarmkanker, waarbij we de applicatie van radiomics voor het voorspellen van respons op neoadjuvante chemoradiotherapie hebben onderzocht met behulp van MRI scans vervaardigd voor start van de behandeling. We hebben de accuratesse van een radiomics-model vergeleken met de accuratesse van ervaren radiologen, welke hun voorspelling baseerden op een visuele interpretatie van dezelfde beelden. Het radiomics model toonde een veelbelovend resultaat met een AUC van 0.69-0.79, vergelijkbaar met de accuratesse van de ervaren radiologen. Het radiomics-model werd ontwikkeld met behulp van segmentaties die manueel werden vervaardigd door respectievelijk ervaren radiologen en meer onervaren lezers, maar ook een door een simpel semi-automatisch segmentatie algoritme. Hieruit bleek dat het radiomics-model een goed en vergelijkbaar resultaat gaf ongeacht het ervarings niveau van de handmatige intekeningen, maar duidelijk slechter presteerde wanneer het semi-automatische algoritme werd gebruikt. Dit benadrukt de noodzaak voor het ontwikkelen van betere segmentatie algoritmen zoals die onderzocht werden in Hoofdstukken 2 en 3.