

Image features for the future in stage IV non-small cell lung cancer

Citation for published version (APA):

de Jong, E. (2019). *Image features for the future in stage IV non-small cell lung cancer*. Ipskamp Printing BV. <https://doi.org/10.26481/dis.20190206ej>

Document status and date:

Published: 01/01/2019

DOI:

[10.26481/dis.20190206ej](https://doi.org/10.26481/dis.20190206ej)

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

S

Background and aim of the thesis

In the Netherlands, each year more than 100,000 people are diagnosed with cancer, and although many of them are cured, still about 40,000 patients yearly die. Of those 40,000 yearly cancer deaths, 24 % is due to lung cancer. The overall survival rate in lung cancer depends on the stage of the disease. Patients with distant metastases are classified as stage IV, independent of the size of the tumor and the number of involved lymph nodes. The 5-years survival rate in stage IV non-small cell lung cancer (NSCLC) is < 10 %. Systemic therapy is advised in all stage IV patients with a good performance status. Patients with driver mutations, as an epidermal growth factor receptor (EGFR) mutated tumor or an anaplastic lymphoma kinase (ALK) rearranged tumor, can be treated with targeted therapy as tyrosine kinase inhibitors.

Prognostic as well as predictive models can be used to tailor treatment decisions by individually optimizing the balance between side effects, prognosis and quality of life of patients. Clinical variables used in prognostic models in lung cancer are for example age, gender, smoking history, performance status, histological grading and mutation status. Prognostic models can also contain biological markers that for example are measured in the blood, so called biomarkers. Image characteristics, like standardized uptake value (SUV) metrics, but also computed tomography (CT) intensity values can potentially improve prognostic and predictive models. Image biomarkers can be qualitative, so called semantic features, or quantitative, for example radiomic features. Using radiomics thousands of quantitative features can be extracted from standard-of-care medical images, like CT, PET and MRI. Radiomic features, typically divided into intensity features, shape features, texture feature and filtered features offer information about cancer phenotype.

In this thesis, the use of image features of positron emission tomography (PET) and CT imaging of stage IV NSCLC patients was investigated for prognosis assessment. Simple PET metrics were used for response assessment, CT radiomic features were used in prognostic models for overall survival and cachexia and associations between semantic features and radiomic features for prediction of the mutation status were explored. Next to that, the quality of PET imaging, the use of contrast-enhanced CT imaging and magnetic resonance (MR) imaging and the influence of image acquisition and reconstruction on radiomic features was studied. Because image biomarkers can only improve prognostic models when the imaging is acquired and reconstructed on a standardized way.

Methodological challenges in image feature studies

The first part of this thesis reviewed the image quality of contrast-enhanced (CE)-CT, magnetic resonance (MR) and PET imaging in multicenter trials. When using data from multiple centers a large variability in image acquisition and reconstruction parameters exist. For image biomarker studies, it is important to use standardized imaging or to correct for variation in image parameters. The influence of image acquisition and reconstruction on radiomic feature values and stability is investigated and methods to correct this variability are proposed. The slice thickness had the largest influence on the radiomic features, but this could be reduced by resampling the images before extracting the features.

In **chapter 3**, image acquisition and reconstruction parameters from ^{18}F -FDG PET scans of 223 stage IV NSCLC patients included in a phase II multicenter trial were compared to recommended image parameters of the European Association of Nuclear Medicine (EANM). Only 19 % of the scans had an uptake time that fulfilled the recommendations, and those scans had a smaller deviation in the mean standardized uptake value of the liver than scans with non-compliant uptake times. SUV_{mean} of the liver can be used a reference organ due to its stable uptake over time and is therefore suitable for assessment of scan parameters. We showed that despite the agreement on a standardization protocol there was a large variability in image parameters. For future multicenter studies including PET imaging, a prospective central quality

review during patient inclusion is needed to improve guideline compliance.

A multicenter trial, in which the added value of gadolinium MRI to dedicated CE-CT for detecting asymptomatic brain metastases was investigated in stage III NSCLC patients, showed that not only in the image acquisition and reconstruction phase differences develop but also in the starting phase of a trial (**chapter 4**). One hospital did not have a MRI scanner that met the required technical standard and two other hospital had to update their protocols to be able to participate in the trial. In 26 % of the patients, the CE-CT of the brain did not meet the criteria for a dedicated brain scan, which was remarkably not reported. Next to all technical guidelines, we recommend that radiologists need be informed about the indication for the requested scan to determine the optimal imaging protocol and that non-imaging clinicians have the knowledge to evaluate whether the applied protocol is adequate for the requested purpose. These recommendations are expected to lead to a more uniform clinical interpretation.

Although there is a large variability in image acquisition and reconstruction parameters between different centers no effect of exposure on radiomic features was shown in **chapter 5**. Slice thickness did influence radiomic feature values but this could be reduced by resampling the CT-images before feature extraction. Optimization of gray level discretization to potentially improve prognostic value can be performed without compromising feature stability.

Image features for prognosis

In the second part of this thesis the use of image features from CT and PET imaging for response assessment, prognosis, mutation status and cachexia is investigated.

In **chapter 6** tumor response defined as a 30 % decrease in CT and PET parameters after three weeks was compared to the response criteria in solid tumors (RECIST) response at week 6. A 30 % decrease in the peak SUV value identified more patients as responders compared to a 30 % decrease in CT diameter, however this was not correlated to overall survival. Although the phase II trial in which patients were randomized between paclitaxel-carboplatin-bevacizumab with or without nitroglycerin showed no added value of nitroglycerin on progression free survival, there was a significant larger decrease in CT diameter in the control group compared to the intervention group, which even shows, although not significant, negative effect of nitroglycerin. No significant difference was seen in the PET parameters between the two treatment arms. This might be due the decreased perfusion caused by bevacizumab in combination with the timing of the PET scan.

In 2014, our group published a prognostic radiomic signature for predicting overall survival. This signature was trained and validated in a cohort of stage I-III NSCLC patients. In **chapter 7** we aimed to validate this prognostic radiomic signature in stage IV NSCLC patients undergoing chemotherapy. We showed that patients having a prognostic index lower than the signature median had a significantly better overall survival than patients with a prognostic index higher than that median (Hazard Ratio 1.445, 95 % confidence interval 1.07-1.95, $p=0.02$). Although the radiomic signature derived from daily practice CT scans, has prognostic value for stage IV, the signature performs less than for stage I-III.

Qualitative image characteristics, such as cavitation or air bronchogram are prognostic for treatment outcome in NSCLC. In **chapter 8** we showed that in radiomic features of the primary tumor could predict the semantic features air bronchogram, necrosis and pleural effusion (area under the curve 0.71, 0.68, 0.65, respectively) and that radiomic features of the ipsilateral lung are associated to pleural thickening, satellite nodules and pleural effusion (AUC 0.62, 0.64 and 0.87, respectively).

Skeletal muscle loss is associated with low quality of life, diminished treatment efficacy and poor survival rates in cancer patients. Early prediction of patients that will suffer from cachexia may lead to timely treatment and prevention. In **chapter 9** we assessed if baseline skeletal muscle radiomic features predicts future muscle loss. Although baseline radiomic features of the skeletal muscle were not predictive for muscle loss (AUC 0.49), differences in intensity and texture radiomic

features over time were seen between patients with and without muscle loss (AUC 0.68, 95 % CI 0.51-0.84).

Future perspectives

In **chapter 10** we discussed the results presented in this thesis and the future perspectives of image features. We hypothesize that standardization of image acquisition, reconstruction and analysis will lead to less variation in image parameters and therefore more reliable models. In combination with initiatives as distributed learning in which data does not leave the hospital anymore, deep learning and machine learning prognostic models will in the future include image characteristics as well as other clinical variables that will improve patient care by prolonged survival and a better quality of life.

Samenvatting

Achtergrond en doel van het proefschrift

In Nederland krijgen jaarlijks 100.000 mensen de diagnose kanker, hoewel velen van hen zullen genezen, overlijden er nog steeds 40.000 patiënten. Van deze 40.000 doden door kanker is 24 % te wijten aan longkanker. Het overlevingspercentage bij longkanker is afhankelijk van het stadium van de ziekte. Patiënten met uitzaaiingen op afstand worden onafhankelijk van de grootte van de tumor en het aantal aangedane lymfeklieren geïnclassificeerd als stadium IV. Het 5-jaarsoverlevingspercentage in stadium IV niet-kleincellig longcarcinoom patiënten is minder dan 10 %. Systeemtherapie wordt geadviseerd voor patiënten met stadium IV ziekte met een goede conditie. Patiënten met een zogenaamde “driver mutatie” zoals epidermale groei factor receptor gemuteerde tumor of een tumor met een translocatie in het anaplastic lymphoma kinase gen kunnen worden behandeld met gerichte therapie bestaande uit tyrosine kinase remmers.

Prognostische en predictieve modellen kunnen worden gebruikt om behandelkeuzes te sturen waardoor de balans tussen bijwerkingen, prognose en kwaliteit van leven geïndividualiseerd en geoptimaliseerd wordt. Klinische variabelen die gebruikt worden in prognostische modellen in longkanker zijn bijvoorbeeld leeftijd, geslacht, rookgeschiedenis, conditie, histologie en mutatiestatus. Prognostische modellen kunnen ook biologische merkers bevatten die bijvoorbeeld in het bloed gemeten kunnen worden, de zogenaamde bloed-biomerkers. Beeldeigenschappen, zoals de activiteit (SUV) van een positronemissietomografie (PET) scan en de intensiteit (HU) van een computertomografie (CT), kunnen potentieel prognostische en predictieve modellen verbeteren. Beeldmerkers kunnen zowel kwalitatief, zogenaamde semantische beeldeigenschappen, als kwantitatief, zogenaamde radiomics beeldeigenschappen, zijn. Door het gebruik van radiomics kunnen duizenden beeldeigenschappen berekend worden uit standaard medische beelden, zoals CT, PET en magnetische resonantie (MR) beelden. Radiomics beeldeigenschappen, te verdelen in intensiteitsbeeldeigenschappen, vormbeeldeigenschappen, textuurbeeldeigenschappen en gefilterde beeldeigenschappen, bevatten informatie over het kanker fenotype.

In dit proefschrift is het gebruik van beeldeigenschappen van PET- en CT-beelden van stadium IV niet-kleincellig longcarcinoom patiënten onderzocht met als doel het bepalen van de prognose van patiënten. Simpele maten voor PET activiteit zijn gebruikt voor responsbepaling op chemotherapie en CT radiomics beeldeigenschappen zijn gebruikt in prognostische modellen voor het voorspellen van overleving en cachexie, ongewild gewichtsverlies. Daarnaast zijn relaties tussen semantische beeldeigenschappen en radiomics beeldeigenschappen voor het voorspellen van de mutatiestatus onderzocht. Van belang voor deze vraagstelling is de kwaliteit van contrast-versterkte CT, MR, en PET. Dit is onderzocht in een multicentrische studie waar gekeken is naar de invloed van beeldacquisitie- en reconstructieparameters op radiomics beeldeigenschappen. Beeldeigenschappen kunnen prognostische modellen namelijk alleen verbeteren als er gebruik wordt gemaakt van gestandaardiseerde beeldacquisitie en reconstructie.

Methodologische uitdagingen in beeldeigenschappen studies

In het eerste deel van dit proefschrift is er gekeken naar de beeldkwaliteit van contrast-versterkte CT-, MR- en PET-beelden in multicentrische trials. Wanneer er data gebruikt wordt uit verschillende centra zal er een variatie in beeldacquisitie- en reconstructieparameters zijn. Voor beeldmarker studies is het belangrijk dat gestandaardiseerde beelden gebruikt worden dan wel dat er gecorrigeerd wordt voor de variatie in beeldparameters. De invloed van beeldacquisitie- en reconstructieparameters op radiomics beeldeigenschappen en hun stabiliteit is onderzocht en er is een voorstel gedaan voor correctiemethoden voor deze variatie. De plakdikte bleek de grootste invloed te hebben op de radiomics beeldeigenschappen, maar dit effect kan verminderd worden door de beelden te herschalen.

In **hoofdstuk 3** zijn beeldacquisitie- en reconstructieparameters van PET scans van 223 stadium IV niet-kleincellig longcarcinoom patiënten, die geïncludeerd waren in een fase II multicentrische studie, vergeleken met de door de Europese associatie van nucleaire geneeskunde (EANM) aanbevolen beeldparameters. Slechts 19 % van de beelden had een opname tijd die voldeed aan de aanbevelingen, en deze scans hadden een kleinere variatie in de gemiddelde activiteit (SUV_{mean}) in de lever dan de scans met een opname tijd die niet-conform de aanbevelingen was. De gemiddelde activiteit in de lever kan worden gebruikt als referentie doordat de activiteit stabiel is over de tijd en daardoor gebruikt kan worden om de beeldparameters te controleren. We hebben laten zien dat ondanks dat men het eens was over het beeldacquisitie- en reconstructieprotocol er in dagelijkse praktijk een grote variatie in beeldparameters is. Voor toekomstige multicentrische studies, die gebruik maken van PET-beelden, is het belangrijk dat vooraf en tijdens patiëntinclusie, een centrale kwaliteitscheck gedaan wordt om de variatie in beeldacquisitie en reconstructie te minimaliseren.

In een multicentrische studie, waar de toegevoegde waarde van gadolinium MR ten opzichte van contrast-versterkte CT in de detectie van asymptomatische hersenmetastasen werd onderzocht in stadium III niet-kleincellig longcarcinoom patiënten, is gebleken dat er niet alleen verschillen ontstaan in de beeldacquisitie- en reconstructie-fase, maar ook tijdens de opstartfase van een studie (**hoofdstuk 4**). Een van de ziekenhuizen had geen MR-scanner die voldeed aan de benodigde technische specificaties en twee andere ziekenhuizen moesten hun protocollen herzien voor ze konden deelnemen aan de studie. Bij 26 % van de patiënten voldeed de contrast-versterkte CT niet aan de criteria voor een speciale hersenscan, wat opmerkelijk genoeg ook niet vermeld werd in het verslag. Naast alle technische richtlijnen raden we aan dat radiologen geïnformeerd worden over de indicatie waar de scan voor gemaakt moet worden, zodat zij kunnen meedenken over het meest optimale scan protocol. Daarnaast moeten andere specialisten inzicht hebben in de verschillende protocollen zodat ze mede kunnen evalueren of het toegepaste scan protocol adequaat is voor het aangevraagde doel. Deze aanbevelingen zouden moeten leiden tot een meer uniforme klinische interpretatie.

Hoewel er een grote variatie in beeldacquisitie- en reconstructieparameters tussen verschillende centra is, is er in **hoofdstuk 5** geen effect gezien van de expositie op radiomics beeldeigenschapswaarden. De plakdikte beïnvloedde de radiomics beeldeigenschapswaarden wel, maar dit kan verminderd worden door de CT-beelden eerst te herschalen voordat de beeldeigenschappen uitgerekend worden. Optimalisatie van grijswaardendiscretisatie, om potentieel de prognostische waarde van radiomics beeldeigenschappen te verbeteren, kan gedaan worden zonder de stabiliteit te veranderen.

Beeldeigenschappen voor prognose

In het tweede deel van dit proefschrift wordt het gebruik van beeldeigenschappen van CT- en PET-beelden voor response beoordeling, prognose, mutatiestatus en cachexie beschreven.

In **hoofdstuk 6** is tumor respons, gedefinieerd als een afname van 30 % in de CT- en PET-parameters na drie weken, vergeleken met de respons criteria in solide tumoren (RECIST) na zes weken. Een afname van 30 % in de piek activiteit (SUV_{peak}) identificeerde meer patiënten als responders dan een 30 % afname in de CT-diameter, alhoewel dit niet gecorreleerd is aan overleving. Hoewel de fase II studie, waar patiënten gerandomiseerd waren tussen paclitaxel-carboplatin-bevacizumab met of zonder nitroglycerine, geen toegevoegde waarde liet zien van nitroglycerine op de progressie vrije overleving, was er een significant grotere afname in CT-diameter in de controlegroep. Er was geen significant verschil in PET-parameters tussen beide groepen. Dit zou kunnen komen door de verstoorde perfusie, veroorzaakt door bevacizumab, in combinatie met de timing van de scan.

In 2014 heeft onze groep een prognostisch radiomics profiel voor overleving gepubliceerd. Dit profiel is getraind en gevalideerd in een cohort van stadium I-III niet-kleincellig longcarcinoom patiënten. In **hoofdstuk 7** hebben we geprobeerd dit prognostisch radiomics

profiel te valideren in stadium IV niet-kleincellig longcarcinoom patiënten die behandeld werden met chemotherapie. We hebben laten zien dat patiënten met een prognostische index lager dan de gemiddelde prognostische index een significant betere overleving hadden dan patiënten met een hoge prognostische index (Hazard Ratio 1.445, 95 % betrouwbaarheidsinterval 1.07-1.95, $p=0.02$). Hoewel het radiomics profiel afgeleid van standaard CT-beelden ook prognostische waarde heeft in stadium IV, voorspelt het profiel beter in stadium I-III.

Kwalitatieve beeldeigenschappen, zoals cavitatie of luchtbronchogram, zijn prognostisch voor behandeluitkomsten bij het niet-kleincellig longcarcinoom. In **hoofdstuk 8** hebben we laten zien dat radiomics beeldeigenschappen van de primaire tumor de semantische beeldeigenschappen luchtbronchogram, necrose en pleurale effusie kunnen voorspellen (AUC 0.71, 0.68, 0.65, respectievelijk) en dat radiomics beeldeigenschappen van de ipsilaterale long gerelateerd zijn aan pleurale verdikking, satelliet laesies en pleurale effusie (AUC 0.62, 0.64 en 0.87, respectievelijk).

Skeletspierverlies is gerelateerd met een lagere kwaliteit van leven, een slechtere behandeluitkomst en een slechtere prognose in kankerpatiënten. Een vroege voorspelling van cachexie kan leiden tot preventie en eerdere behandeling. In **hoofdstuk 9** hebben we geïnventariseerd of radiomics beeldeigenschappen van de skeletspier op de CT, gemaakt voor de start van behandeling, toekomstig spierverlies kan voorspellen. Hoewel deze radiomics beeldeigenschappen niet voorspellend waren voor spierverlies (AUC 0.49), waren er verschillen in radiomics intensiteitseigenschappen en textuureigenschappen over de tijd te zien tussen patiënten met en zonder spierverlies (AUC 0.68, 95 % betrouwbaarheidsinterval 0.51-0.84).

Toekomstperspectieven

In **hoofdstuk 10** hebben we resultaten en toekomstperspectieven van de studies over beeldeigenschappen in dit proefschrift besproken. We denken dat gestandaardiseerde beeldacquisitie, reconstructie, analyse en initiatieven als “distributed learning”, waarbij data het ziekenhuis niet meer hoeft te verlaten, in combinatie met “deep learning” en “machine learning” in de toekomst zal leiden tot prognostische modellen die naast klinische variabelen ook beeldeigenschappen bevatten de patiëntzorg zal verbeteren door langere overleving en een betere kwaliteit van leven.

