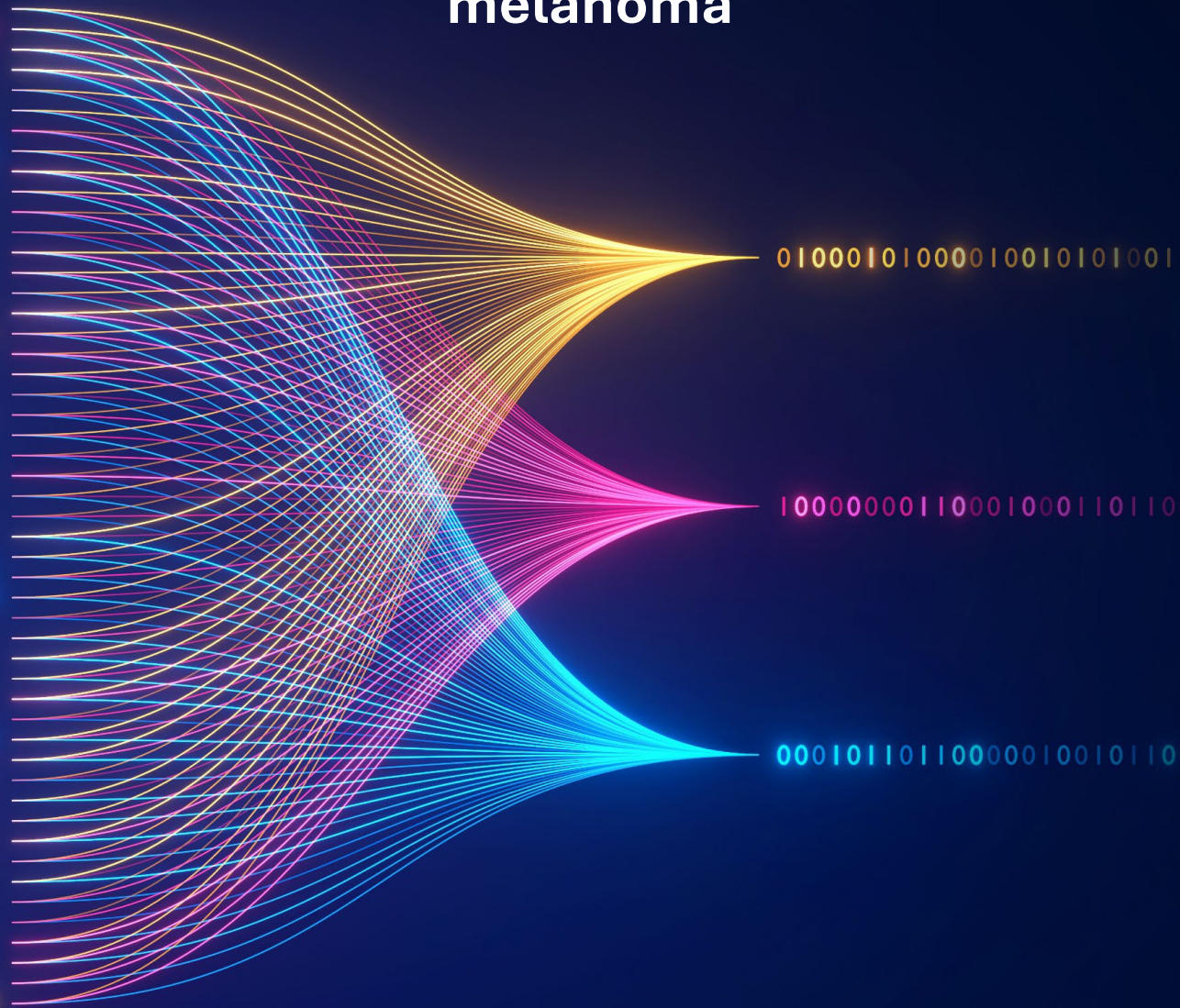


# Imaging to predict checkpoint inhibitor outcomes in advanced melanoma





# Imaging to predict checkpoint inhibitor outcomes in advanced melanoma

Behandeluitkomsten van checkpointremming bij uitgezaaid  
melanoom voorspellen door middel van beeldvorming  
(met een samenvatting in het Nederlands)

## Proefschrift

ter verkrijging van de graad van doctor aan de  
Universiteit Utrecht  
op gezag van de  
rector magnificus, prof. dr. H.R.B.M. Kummeling,  
ingevolge het besluit van het College voor Promoties  
in het openbaar te verdedigen op  
woensdag 27 november 2024 des middags te 12.15 uur

door

**Laurens Sebastian ter Maat**

geboren op 26 april 1995  
te Woerden

**Promotoren:**

Prof. dr. J.P.W. Pluim

Prof. dr. P.J. van Diest

Prof. dr. K.P.M. Suijkerbuijk

**Copromotor:**

Dr. M. Veta

**Beoordelingscommissie:**

Dr. K.G.A. Gilhuijs

Prof. dr. G.A.P. Hospers

Prof. dr. J. van der Laak

Prof. dr. M.G.E.H. Lam

Prof. dr. ir. J. de Ridder (voorzitter)

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van ZonMw, Koninklijk Philips N.V. en Stichting Hanarth Fonds.

*Voor Kobe*



# Contents

<b>1 Introduction</b>	1
1.1 Overview	1
1.2 An introduction to checkpoint inhibitor treatment	2
1.3 Drawbacks of checkpoint inhibitor treatment	3
1.4 Predicting individual treatment outcomes	4
1.5 Research question and thesis outline	6
1.6 References	8
<b>2 Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.</b>	11
2.1 Abstract	12
2.1.1 Introduction	12
2.1.2 Methods	12
2.1.3 Results	12
2.1.4 Conclusions	13
2.2 Introduction	13
2.3 Methods	15
2.3.1 Adherence to quality standards	15
2.3.2 Selection of studies	15
2.3.3 Screening process	16
2.3.4 Critical appraisal	16
2.3.5 Data extraction	17
2.3.6 Synthesis	18
2.4 Results	18

## Contents

---

2.4.1	General characteristics	18
2.4.2	Tumor burden	19
2.4.3	Body composition	20
2.4.4	Metastasis location	21
2.4.5	FDG-PET features	22
2.4.6	Other PET radioactive tracers	22
2.4.7	Radiomics	23
2.5	Discussion	25
2.5.1	Overview	25
2.5.2	Future research	27
2.5.3	Limitations	28
2.5.4	Conclusion	28
2.6	Supplementary Materials	29
2.7	References	29
<b>3</b>	<b>CT radiomics compared to a clinical model for predicting check-point inhibitor treatment outcomes in patients with advanced melanoma</b>	<b>39</b>
3.1	Abstract	40
3.1.1	Introduction	40
3.1.2	Methods	40
3.1.3	Results	40
3.1.4	Discussion	41
3.2	Introduction	41
3.3	Methods	43
3.3.1	Patient selection	43
3.3.2	Lesion selection and segmentation	43
3.3.3	Feature extraction	44
3.3.4	Outcome definition	44
3.3.5	Evaluated models	44
3.3.6	Statistical analysis	45
3.3.7	Adherence to quality standards	45
3.4	Results	46
3.4.1	Patient characteristics	46



3.4.2 Interobserver variability . . . . .	46
3.4.3 Treatment outcome prediction . . . . .	47
3.4.4 Comparison of radiomics and clinical model . . . . .	48
3.5 Discussion . . . . .	48
3.6 Supplementary Materials . . . . .	54
3.7 References . . . . .	54
<b>4 Deep learning on CT scans to predict checkpoint inhibitor treatment outcomes in advanced melanoma</b>	<b>57</b>
4.1 Abstract . . . . .	58
4.1.1 Introduction . . . . .	58
4.1.2 Methods . . . . .	58
4.1.3 Results . . . . .	58
4.1.4 Discussion . . . . .	58
4.2 Introduction . . . . .	59
4.3 Methods . . . . .	61
4.3.1 Patient selection . . . . .	61
4.3.2 ROI selection and preprocessing . . . . .	61
4.3.3 Outcome definition . . . . .	62
4.3.4 Model selection and hyperparameter selection . . . . .	62
4.3.5 Cross validation, model training and evaluation . . . . .	63
4.3.6 Statistical analysis . . . . .	63
4.3.7 Adherence to quality standards . . . . .	64
4.4 Results . . . . .	65
4.4.1 Patient characteristics . . . . .	65
4.4.2 Hyperparameter selection . . . . .	65
4.4.3 Treatment outcome prediction . . . . .	65
4.4.4 Interpretability analysis . . . . .	67
4.5 Discussion . . . . .	67
4.6 Supplementary materials . . . . .	69
4.7 References . . . . .	69
<b>5 Body composition and checkpoint inhibitor treatment outcomes in advanced melanoma: a multicenter cohort study</b>	<b>77</b>
5.1 Abstract . . . . .	78

## Contents

---

5.1.1 Introduction	78
5.1.2 Methods	78
5.1.3 Results	78
5.1.4 Discussion	79
5.2 Introduction	79
5.3 Methods	80
5.3.1 Patient selection	80
5.3.2 BMI and clinical predictors	81
5.3.3 CT body composition metrics extraction	81
5.3.4 Outcome definition	82
5.3.5 Statistical analysis	83
5.4 Results	83
5.4.1 Patient characteristics	83
5.4.2 Body mass index	84
5.4.3 CT derived body composition metrics	85
5.5 Discussion	86
5.6 Supplementary Materials	89
5.7 References	89
<b>6 Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study</b>	<b>93</b>
6.1 Abstract	94
6.1.1 Introduction	94
6.1.2 Methods	94
6.1.3 Results	94
6.1.4 Discussion	94
6.2 Introduction	95
6.3 Methods	97
6.3.1 Patient selection	97
6.3.2 Data collection	98
6.3.3 Outcome definition	98
6.3.4 Deep learning model	98
6.3.5 Hyperparameter exploration	100
6.3.6 Model evaluation	100

<b>6.4 Results</b>	101
6.4.1 Patient characteristics	101
6.4.2 Hyperparameter tuning	102
6.4.3 Treatment outcome prediction	102
6.4.4 Other prediction targets	103
<b>6.5 Discussion</b>	103
<b>6.6 Supplementary Materials</b>	107
<b>6.7 References</b>	107
<b>7 General discussion</b>	<b>111</b>
7.1 Overview	111
7.2 Summary of findings	112
7.3 Strengths and limitations	114
7.4 Interpretation	116
7.5 Recommendations	118
7.6 References	119
<b>8 Summary</b>	<b>123</b>
<b>9 Dutch Summary (Nederlandse Samenvatting)</b>	<b>127</b>
<b>10 Acknowledgements (Dankwoord)</b>	<b>131</b>
<b>11 Curriculum Vitae</b>	<b>135</b>

## Contents

---

# Chapter 1

## Introduction

### 1.1 Overview

This thesis covers the predictive value of computed tomography- and histopathology-based predictors for checkpoint inhibitor treatment outcomes in patients with advanced melanoma. Accurate predictors in this setting are lacking but would be very valuable in clinical practice, since checkpoint inhibitor treatment is associated with high costs and serious toxicity. These concerns will become even more relevant in the coming decades due to an aging population. Predictors that are covered in this thesis include the location and size of metastases, body composition metrics, and machine and deep learning models based on pretreatment computed tomography (CT) scans and histopathology images. Special emphasis is placed on the added value over known predictors, as this determines the eventual clinical value.

This introduction will first outline what checkpoint inhibitor treatment is, and the place it has in the treatment of advanced melanoma. Second, it will discuss the drawbacks of this treatment, which form the motivation for accurate predictions. Third, it will provide the requirements for a clinically useful predictor and categorize previous efforts to this end. Lastly, it will outline the structure of the rest of this thesis, the research questions that will be addressed and the corresponding hypotheses.

# 1.2 An introduction to checkpoint inhibitor treatment

Checkpoint inhibitor treatment has revolutionized the treatment of advanced cancer, in particular advanced melanoma. Melanoma is an aggressive form of cancer, arising from the pigmented cells called melanocytes, mostly those from the skin. When a melanoma metastasizes to such a degree that it can no longer be curatively treated through surgery, it is defined as advanced melanoma [1]. Prior to the introduction of checkpoint inhibitors, the prognosis of patients diagnosed with advanced melanoma was very poor, with most patients succumbing within one year of diagnosis [2]. With current treatment, overall survival rates at 6.5 years of 49% are being reached for patients treated in clinical trials [3]. Even more strikingly, survival curves level off after several years, which suggests that this treatment may even be curative in these cases. Checkpoint inhibitor treatment therefore represents one of the most significant advances in oncological care of the past decades.

Checkpoint inhibitor treatment works by mobilizing the body's immune system against the tumor. In addition to fighting infections, the immune system is tasked with identifying and neutralizing cells that have accumulated (epi-)genetic changes that may eventually lead to the development of cancer. This concept is called 'immune surveillance' [4]. In cases where cancer develops, the tumor has developed mechanisms to resist the efforts of the immune system [5]. A mechanism for doing so is through the expression of certain proteins, called immune checkpoints, that block the immune response. These immune checkpoints are used in a healthy situation to maintain the careful balance between too much immune activity (resulting in auto-inflammatory diseases) or too little (potentially resulting in infections or cancer formation). Checkpoint inhibitor treatment blocks the immune checkpoint pathway and thereby reactivates the body's immune response to the tumor [6].

There are currently two mostly used types of checkpoint inhibitor treatment for advanced melanoma. The first is anti-CTLA4 treatment, which targets the CTLA-4 protein, a crucial regulator of T-cell activation. Anti-CTLA4 drugs, such as ipilimumab, work by inhibiting the CTLA-4 protein, thereby enhancing the immune system's ability to recognize and attack tumor cells. This type of treatment,

introduced in 2011 [7], was among the first to show significant success in treating advanced melanoma. The second type, which became available in 2014 [8], is anti-PD1 treatment. These inhibitors, including pembrolizumab and nivolumab, target the PD-1 surface protein on T-cells. By blocking this protein, anti-PD1 treatment prevents the tumor from evading immune surveillance. Currently, checkpoint inhibitor therapy for melanoma in The Netherlands is given mainly as anti-PD1 monotherapy, or as a combination therapy of anti-PD1 and anti-CTLA4 [9].

### **1.3 Drawbacks of checkpoint inhibitor treatment**

Although for advanced melanoma checkpoint inhibition is a clear improvement over historical treatments such as chemotherapy, a significant fraction of patients does not respond. This difference in derived benefit often becomes apparent early during treatment. Real-world data on the status of patients treated with anti-PD1 in The Netherlands at three months showed a complete response in 1%, partial response in 33%, stable disease in 29%, progressive disease or death in 38% of patients. Of the patients with partial response or stable disease, 80% and 57%, respectively, were alive two years from the three-month landmark. In contrast, of patients with progressive disease at three months, only 15% was alive after two years [10].

Furthermore, checkpoint inhibitor treatment is associated with potentially severe toxicity. These side effects, often termed 'immune-related adverse events' (irAEs), can range from mild to life-threatening [11] and are a consequence of the immune system's heightened activity. Common irAEs include dermatological conditions like rash and pruritus, gastrointestinal issues such as colitis and diarrhea, and endocrine disorders including thyroiditis, hypophysitis and adrenal insufficiency. Other toxicities involve the lungs, presenting as pneumonitis, or the liver, manifesting as hepatitis [12]. Immune-related adverse events more commonly occur in patients treated with combination therapy [3], can be chronic and may in some cases have fatal outcomes [13].

In addition to the risk of toxicity, checkpoint inhibitor treatment is a very costly therapy. In fact, anti-PD1 drugs were by far the most expensive medication given in hospitals in The Netherlands in 2021, representing a total cost of 327 million EUR [14]. Advanced melanoma is second largest indication for this treatment,

## 1.4. Predicting individual treatment outcomes

---

after non-small cell lung carcinoma [15,16]. Estimates of additional costs per gained quality-adjusted life year range from 25,000 to 81,000 USD [17,18].

The issue of medication costs will become even more relevant in the coming decades for several reasons. First, an aging population means that the health care system will require more resources to maintain the current standard level of care [19]. Oncological diagnoses in particular are expected to increase, which is one of the primary drivers of medication costs. Second, expensive medications have recently been developed at a much faster rate than the expiration of patents. If this trend continues, this will further increase the share of health care spending on expensive medication [20].

## 1.4 Predicting individual treatment outcomes

Accurate predictors of individual treatment would be extremely valuable. Such a predictor, by identifying patients unlikely to respond, could prevent both unnecessary toxicity, thus benefiting patients, and unnecessary costs, thus benefiting society. It would thereby directly address the rising problem of medication costs. Furthermore, accurate predictors could be used to guide individual initial treatment choices, so that no time is lost with ineffective treatments. For example, patients with BRAF-mutant disease may be directed towards upfront BRAF/MEK-inhibition therapy. In addition, patients may be selected earlier in their disease process for clinical trials, accelerating the development of new treatment options for these patients. For these reasons, research into potential predictors has attracted much attention in the past decade [21,22].

A successful predictor especially needs to identify patients who will not respond with high reliability. This is because the potential benefit of treatment for most patients with advanced melanoma is enormous: a durable remission of disease, versus near-certain death within a year from diagnosis. Furthermore, proven alternatives are limited in patients without targetable mutations (e.g. BRAF-V600) [1]. A predictor that wrongfully advises against treatment would therefore cause significant harm. Clearly, a high negative predictive value is of paramount importance to ensure that a predictor benefits both society and the individual.

Furthermore, a successful predictor must ideally be available before the start of treatment. The reason for this is that most irAEs occur in the first three months of



treatment, which is typically the first moment of follow-up [23,24]. Furthermore, especially when combination treatment with anti-PD1+anti-CTLA4 is given, most costs are also incurred in this period. This makes pretreatment predictors much more impactful than those that become available at a later moment.

Research aimed to identify potential predictors falls into one of three categories. The first is that of patient characteristics: features that are readily available to the treating physician through history, physical examination, or routine diagnostics. The main driver of this research is the practical clinical experience of physician-researchers. The second category is research into biology-inspired predictors. The hypotheses for this research come from a mechanistic model of checkpoint inhibitors mode of action. The third category comprises data-driven research, which is driven by developments in modelling techniques for high-dimensional data, such as images and genomics.

Several clinical characteristics have been shown to be associated with treatment outcomes. These include stage of disease, serum markers (e.g. lactate dehydrogenase) and clinical condition (e.g. ECOG/WHO performance status) [25]. Another example of predictor of worse outcomes is the presence of symptomatic brain metastases [10]. Together, these predictors reflect a patient's physical reserves and the aggressiveness of disease. It should be noted that these factors were also identified as prognostic in melanoma before the introduction of checkpoint inhibitor treatment [26,27]. Although these factors may also reflect a patient's ability to mount an effective immune response following therapy, their prognostic value is therefore, at least to some extent, not specific to checkpoint inhibitor treatment. They may nonetheless be valuable for guiding treatment decisions by providing insight into the absolute survival that may be attained.

Biology-driven research has identified several potential predictors. A prime example is the level of PD-L1 protein expression on tumor cells. The PD-L1 protein blocks the immune response by binding to and thereby activating the PD-1 receptor on T-cells. Anti-PD(L)1 treatment stops this mechanism by blocking the PD-1 receptor on T-cells. According to this understanding, tumors with higher PD-L1 expression should respond more to anti-PD1 therapy. Indeed, a positive correlation was observed between PD-L1 expression and response rates. Nevertheless, response was still observed in 38% of patients in whom PD-L1 was absent [28]. Another example is the number of mutations present across the tumor genome, also

## 1.5. Research question and thesis outline

---

known as tumor mutational burden. As tumors with a higher mutational burden evoke a stronger immune response since the mutated genes express neoantigens, it has been shown that this measure also correlates with response to checkpoint inhibitor treatment [29].

Examples of data-driven predictors are pathomics, genomics and radiomics. These predictors are based on a type of high-dimensional data, such as histopathology imaging [30,31], gene expression data [32] or radiological imaging [33]. Developments in computing hardware and in the field of machine and deep learning have provided the tools necessary to analyze these complex types of data. Instead of investigating a specific hypothesis, machine learning aims to capture the variation in the dataset by extracting an extensive number of features. A wide range of techniques have been developed to subsequently analyze these extracted features by means of dimensionality reduction (e.g. principal component analysis [34]), regression (e.g. lasso regression [35]) and classification (e.g. random forest [36], support vector classifier [37]). Deep learning differs from machine learning by building a model directly on the input data, thereby skipping the step of extracting hand-crafted features. It leverages neural networks, a very powerful type of model due to its flexibility in the relationships that it can learn. The most important types of neural networks are the convolutional neural network [38] (for image data) and transformer [39] (for data including sequences of variable length). Despite previous efforts, current predictors have insufficient negative predictive value. Clinical predictors can stratify patients into categories with markedly different outcomes. However, patients in the category with the worst predicted outcome still have a significant probability of achieving a durable remission [25]. These predictors therefore do not reach the high negative predictive value that is required to change decisions in the majority of patients that are motivated for treatment.

## 1.5 Research question and thesis outline

The research described in this thesis is the result of the PREMIUM project [40]. This project was a combined effort of 11 melanoma treatment centers in The Netherlands, coordinated by the UMC Utrecht and TU Eindhoven. The goal of the project was to predict first-line checkpoint inhibitor treatment outcomes in patients with irresectable stage III or stage IV melanoma. For this purpose,

eligible patients were identified through the high-quality registry data of the Dutch Melanoma Treatment Registry [41]. This has resulted in a unique cohort of almost 2000 patients, with available clinical characteristics, follow-up data, pre-treatment computed tomography (CT) scans and diagnostic histopathology images.

This thesis focuses on predictors that are based on pretreatment computed tomography (CT) scans and histopathology images. This choice of scope was made for a practical reason: since this data is already routinely collected, models based on these modalities do not require a modification of the standard-of-care. This maximizes the potential impact of the developed predictors.

The main research questions are as follows: what is the predictive value of CT- and histopathology-based features for checkpoint inhibitor outcomes in advanced melanoma and what is their added value over known clinical predictors? Specific emphasis is placed on the added value of potential predictors, for obvious reasons: a predictive model will not contribute if it only adds information that overlaps with known predictors. Previous works, however, typically do not evaluate this, as will be shown in chapter 2. This thesis aims to address this shortcoming by comparing the proposed predictors to a baseline model of clinical characteristics.

Chapter 2 provides a systematic review of previous literature on radiological imaging derived predictors. This category consists of patient characteristics that are directly visible on scans, such as the number and location of metastases. This chapter thereby aims to answer the question: what is the current state of research on radiological image-derived predictors for checkpoint inhibitor outcomes? Research on all cancers, not just melanoma, is covered in this chapter, since results in one malignancy may as well transfer to another. The categories of predictors that are covered in this chapter are tumor burden, location, tracer avidity, radiomics and body composition metrics.

Chapter 3 assesses the predictive value of radiomics for predicting checkpoint inhibitor treatment outcomes in advanced melanoma. Radiomics are a large number of image analysis-based features that are calculated based on a segmented metastasis on CT, which are subsequently analyzed using machine learning techniques. These features together aim to encompass the variation in presentation of metastases on CT. The main hypothesis here is that the phenotype of the tumor visible on CT images reflects its biology, which in turn determines the tumor's response to treatment. The added value over known predictors is assessed by

## 1.6. References

---

comparing predictive models with- and without a radiomics-based predictor.

Chapter 4 investigates the predictive value of deep learning on CT scans. Here, a model is fitted to directly predict outcomes on the raw CT volume of metastases. Features do not have to be manually defined, as they are learned during the training phase. The resulting advantage is that the model is not limited by the choice of these extracted features. The hypothesis of this chapter is therefore that deep learning, providing more flexibility, can improve over a radiomics-based method.

Chapter 5 covers CT derived body composition metrics, such as the amount of skeletal muscle and adipose tissue. The hypothesis of this chapter is that these body composition metrics provide additional prognostic value over known predictors. They are evaluated alongside body mass index (BMI), a traditional measure of body composition that is not dependent on extraction of metrics from CT imaging.

Chapter 6 discusses deep learning based on hematoxylin & eosin-stained whole slide histopathology images. These histopathology images include primary and metastatic samples. The hypothesis investigated in this chapter is that these samples contain information that is correlated with checkpoint inhibitor response, and that deep learning models can extract this information. Examples of features that could be informative are the aspect and growth pattern of tumor cells, abundance and location of tumor cells and immune cells and presence of fibrosis.

Chapter 7 provides the general discussion and conclusion of this thesis. The findings of previous chapters will be summarized and placed into perspective. Next, the limitations of the presented research will be outlined. Lastly, recommendations for clinical practice and future research will be presented.

## 1.6 References

- [1] Keilholz U, Ascierto PA, Dummer R, Robert C, Lorigan P, Akkooi A van, et al. ESMO consensus conference recommendations on the management of metastatic melanoma: under the auspices of the ESMO Guidelines Committee. *Ann Oncol* 2020;31:1435–48. <https://doi.org/10.1016/j.annonc.2020.07.004>
- [2] Korn EL, Liu P-Y, Lee SJ, Chapman J-AW, Niedzwiecki D, Suman VJ, et al. Meta-Analysis of Phase II Cooperative Group Trials in Metastatic Stage IV Melanoma to Determine Progression-Free and Overall Survival Benchmarks for Future Phase II Trials. *J Clin Oncol* 2008;26:527–34. <https://doi.org/10.1200/JCO.2007.12.7837>
- [3] Wolchok JD, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Long-Term Outcomes With Nivolumab Plus Ipilimumab or Nivolumab Alone Versus Ipilimumab in Patients With Advanced Melanoma. *J Clin Oncol* 2022;40:127–37. <https://doi.org/10.1200/jco.21.02229>

- [4] Swann JB, Smyth MJ. Immune surveillance of tumors. *J Clin Invest* 2007;117:1137–46. <https://doi.org/10.1172/JCI31405>
- [5] Kim R. Chapter 2 - Cancer Immunoediting: From Immune Surveillance to Immune Escape. In: Prendergast GC, Jaffee EM, editors. *Cancer Immunother.*, Burlington: Academic Press; 2007, p. 9–27. <https://doi.org/10.1016/B978-012372551-6/50066-3>
- [6] Sharma P, Allison JP. Dissecting the mechanisms of immune checkpoint therapy. *Nat Rev Immunol* 2020;20:75–6. <https://doi.org/10.1038/s41577-020-0275-8>
- [7] Lipson EJ, Drake CG. Ipilimumab: An Anti-CTLA-4 Antibody for Metastatic Melanoma. *Clin Cancer Res* 2011;17:6958–62. <https://doi.org/10.1158/1078-0432.CCR-11-1595>
- [8] Barone A, Hazarika M, Theoret MR, Mishra-Kalyani P, Chen H, He K, et al. FDA Approval Summary: Pembrolizumab for the Treatment of Patients with Unresectable or Metastatic Melanoma. *Clin Cancer Res* 2017;23:5661–5. <https://doi.org/10.1158/1078-0432.CCR-16-0664>
- [9] Carlino MS, Larkin J, Long GV. Immune checkpoint inhibitors in melanoma. *The Lancet* 2021;398:1002–14. [https://doi.org/10.1016/S0140-6736\(21\)01206-X](https://doi.org/10.1016/S0140-6736(21)01206-X)
- [10] van Zeijl MCT, Haanen JBAG, Wouters MWJM, de Wreede LC, Jochems A, Aarts MJB, et al. Real-world Outcomes of First-line Anti-PD-1 Therapy for Advanced Melanoma: A Nationwide Population-based Study. *J Immunother* 2020;43:256–64. <https://doi.org/10.1097/CJI.0000000000000334>
- [11] Almutairi AR, McBride A, Slack M, Erstad BL, Abraham I. Potential Immune-Related Adverse Events Associated With Monotherapy and Combination Therapy of Ipilimumab, Nivolumab, and Pembrolizumab for Advanced Melanoma: A Systematic Review and Meta-Analysis. *Front Oncol* 2020;10.
- [12] Ramos-Casals M, Brahmer JR, Callahan MK, Flores-Chávez A, Keegan N, Khamashta MA, et al. Immune-related adverse events of checkpoint inhibitors. *Nat Rev Dis Primer* 2020;6:1–21. <https://doi.org/10.1038/s41572-020-0160-6>
- [13] Jiang Y, Zhang N, Pang H, Gao X, Zhang H. Risk and incidence of fatal adverse events associated with immune checkpoint inhibitors: a systematic review and meta-analysis. *Ther Clin Risk Manag* 2019;15:293–302. <https://doi.org/10.2147/TCRM.S191022>
- [14] Zaken M van A. Bijlage 2: Uitgaven per geneesmiddel - Publicatie - Rijksoverheid.nl 2023. <https://www.rijksoverheid.nl/documenten/publicaties/2023/03/28/bijlage-2-uitgaven-2021-per-geneesmiddel> (accessed October 20, 2023).
- [15] Nederland Z. Horizonscan geneesmiddelen - Pembrolizumab n.d. <https://www.horizonscangeneesmiddelen.nl/geneesmiddelen/pembrolizumab-oncologie-en-hematologie-longkanker/versie1?lang=nl> (accessed October 20, 2023).
- [16] Nederland Z. Pembrolizumab n.d. [https://www.horizonscangeneesmiddelen.nl/geneesmiddelen/pembrolizumab-oncologie-huidkanker/versie4\(accessedJanuary23,2024\)](https://www.horizonscangeneesmiddelen.nl/geneesmiddelen/pembrolizumab-oncologie-huidkanker/versie4(accessedJanuary23,2024))
- [17] Leeneman B, Uyl-de Groot CA, Aarts MJB, van Akkooi ACJ, van den Berkmortel FWPJ, van den Eertwegh AJM, et al. Healthcare Costs of Metastatic Cutaneous Melanoma in the Era of Immunotherapeutic and Targeted Drugs. *Cancers* 2020;12:E1003. <https://doi.org/10.3390/cancers12041003>
- [18] Verma V, Sprave T, Haque W, Simone CB, Chang JY, Welsh JW, et al. A systematic review of the cost and cost-effectiveness studies of immune checkpoint inhibitors. *J Immunother Cancer* 2018;6:128. [://doi.org/10.1186/s40425-018-0442-7](https://doi.org/10.1186/s40425-018-0442-7).
- [19] Ministerie van Volksgezondheid W en S. Kiezen voor houdbare zorg - Mensen, middelen en maatschappelijk draagvlak - Rapport - Rijksoverheid.nl 2021 [https://www.rijksoverheid.nl/documenten/rapporten/2022/05/03/wrr-rapport-kiezen-voor-houdbare-zorg\(accessedOctober20,2023\)](https://www.rijksoverheid.nl/documenten/rapporten/2022/05/03/wrr-rapport-kiezen-voor-houdbare-zorg(accessedOctober20,2023))
- [20] Uitgaven geneesmiddelen kanker nemen fors toe, overleving stijgt niet overal n.d. [https://iknl.nl/nieuws/2023/uitgaven-geneesmiddelen\(accessedJanuary23,2024\)](https://iknl.nl/nieuws/2023/uitgaven-geneesmiddelen(accessedJanuary23,2024))
- [21] Duffy MJ, Crown J. Biomarkers for Predicting Response to Immunotherapy with Immune Checkpoint Inhibitors in Cancer Patients. *Clin Chem* 2019;65:1228–38. <https://doi.org/10.1373/clinchem.2019.303644>
- [22] Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol* 2016;17:e542–51. [https://doi.org/10.1016/S1470-2045\(16\)30406-5](https://doi.org/10.1016/S1470-2045(16)30406-5)

## 1.6. References

---

- [23 ] Larkin J, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med* 2019;381:1535–46. <https://doi.org/10.1056/NEJMoa1910836>
- [24 ] Robert C, Long GV, Brady B, Dutriaux C, Giacomo AMD, Mortier L, et al. Five-Year Outcomes With Nivolumab in Patients With Wild-Type BRAF Advanced Melanoma. *J Clin Oncol* 2020. <https://doi.org/10.1200/JCO.20.00995>
- [25 ] Silva IP da, Ahmed T, McQuade JL, Nebhan CA, Park JJ, Versluis JM, et al. Clinical Models to Define Response and Survival With Anti-PD-1 Antibodies Alone or Combined With Ipilimumab in Metastatic Melanoma. *J Clin Oncol* 2022. <https://doi.org/10.1200/JCO.21.01701>.
- [26 ] Raizer JJ, Hwu W-J, Panageas KS, Wilton A, Baldwin DE, Bailey E, et al. Brain and leptomeningeal metastases from cutaneous melanoma: Survival outcomes based on clinical features. *Neuro-Oncol* 2008;10:199–207. <https://doi.org/10.1215/15228517-2007-058>
- [27 ] Sirott MN, Bajorin DF, Wong GYC, Tao Y, Chapman PB, Templeton MA, et al. Prognostic factors in patients with metastatic malignant melanoma: A multivariate analysis. *Cancer* 1993;72:3091–8. [https://doi.org/10.1002/1097-0142\(19931115\)72:10<3091::AID-CNCR2820721034>3.0.CO;2-V](https://doi.org/10.1002/1097-0142(19931115)72:10<3091::AID-CNCR2820721034>3.0.CO;2-V)
- [28 ] Morrison C, Pabla S, Conroy JM, Nesline MK, Glenn ST, Dressman D, et al. Predicting response to checkpoint inhibitors in melanoma beyond PD-L1 and mutational burden. *J Immunother Cancer* 2018;6:32. <https://doi.org/10.1186/s40425-018-0344-8>
- [29 ] Ning B, Liu Y, Wang M, Li Y, Xu T, Wei Y. The Predictive Value of Tumor Mutation Burden on Clinical Efficacy of Immune Checkpoint Inhibitors in Melanoma: A Systematic Review and Meta-Analysis. *Front Pharmacol* 2022;13.
- [30 ] Saltz J, Almeida J, Gao Y, Sharma A, Bremer E, DiPrima T, et al. Towards Generation, Management, and Exploration of Combined Radiomics and Pathomics Datasets for Cancer Research. *AMIA Summits Transl Sci Proc* 2017;2017:85–94.
- [31 ] Gupta R, Kurc T, Sharma A, Almeida JS, Saltz J. The Emergence of Pathomics. *Curr Pathobiol Rep* 2019;7:73–84. <https://doi.org/10.1007/s40139-019-00200-x>.
- [32 ] Keenan TE, Burke KP, Van Allen EM. Genomic correlates of response to immune checkpoint blockade. *Nat Med* 2019;25:389–402. <https://doi.org/10.1038/s41591-019-0382-x>
- [33 ] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278:563–77. <https://doi.org/10.1148/radiol.2015151169>
- [34 ] Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemom Intell Lab Syst* 1987;2:37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- [35 ] Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B Methodol* 1996;58:267–88. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [36 ] Breiman L. Random Forests. *Mach Learn* 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
- [37 ] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. *Proc. Fifth Annu. Workshop Comput. Learn. Theory*, New York, NY, USA: Association for Computing Machinery; 1992, p. 144–52. <https://doi.org/10.1145/130385.130401>
- [38 ] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86:2278–324. <https://doi.org/10.1109/5.726791>
- [39 ] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. *Adv. Neural Inf. Process. Syst.*, vol. 30, Curran Associates, Inc.; 2017.
- [40 ] Predicting Response of metastatic Melanoma to Immunotherapy Using Machine learning (PREMIUM). ZonMw Proj n.d. <https://projecten.zonmw.nl/nl/project/predicting-response-metastatic-melanoma-immunotherapy-using-machine-learning-premium>(accessedOctober25, 2023)
- [41 ] Jochems A, Schouwenburg MG, Leeneman B, Franken MG, van den Eertwegh AJM, Haanen JBAG, et al. Dutch Melanoma Treatment Registry: Quality assurance in the care of patients with metastatic melanoma in the Netherlands. *Eur J Cancer* 2017;72:156–65. <https://doi.org/10.1016/j.ejca.2016.11.021>

## Chapter 2

# Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

L.S. ter Maat MD, I.A.J. van Duin MD, S.G. Elias MD PhD, P.J. van Diest MD PhD, J.P.W. Pluim PhD, J.J.C. Verhoeff MD PhD, P.A. de Jong MD PhD, T. Leiner MD PhD, M. Veta PhD, K.P.M. Suijkerbuijk MD PhD

*Published in European Journal of Cancer 175, 60-76*

## 2.1. Abstract

---

## 2.1 Abstract

### 2.1.1 Introduction

Checkpoint inhibition has radically improved the perspective for metastatic cancer patients, but predicting who will not respond with high certainty remains difficult. Imaging derived biomarkers may be able to provide additional insights into the heterogeneity in tumor response between patients. In this systematic review, we aimed to summarize and qualitatively assess the current evidence on imaging biomarkers that predict response and survival in patients treated with checkpoint inhibitors in all cancer types.

### 2.1.2 Methods

PubMed and Embase were searched from database inception to November 29th, 2021. Articles eligible for inclusion described baseline imaging predictive factors, radiomics and/or imaging machine learning models for predicting response and survival in patients with any kind of malignancy treated with checkpoint inhibitors. Risk of bias was assessed using the QUIPS and PROBAST tools and data was extracted.

### 2.1.3 Results

In total, 119 studies including 15580 patients were selected. Of these studies, 73 investigated simple imaging factors. 45 studies investigated radiomic features or deep learning models. Predictors of worse survival were (i) higher tumor burden, (ii) presence of liver metastases, (iii) less subcutaneous adipose tissue, (iv) less dense muscle and (v) presence of symptomatic brain metastases. Hazard rate ratios did not exceed 2.00 for any predictor in the larger and higher quality studies. The added value of baseline FDG-PET parameters in predicting response to treatment was limited. Pilot studies of radioactive drug tracer imaging showed promising results. Reports on radiomics were almost unanimously positive, but numerous methodological concerns exist.



### 2.1.4 Conclusions

There is well-supported evidence for several imaging biomarkers that can be used in clinical decision making. Further research, however, is needed into biomarkers that can more accurately identify which patients who will not benefit from checkpoint inhibition. Radiomics and radioactive drug labeling appear to be promising approaches for this purpose.

## 2.2 Introduction

The introduction of immune checkpoint inhibitors has greatly improved survival for patients in advanced stages of several cancer types. Since the approval of checkpoint inhibitors for metastatic melanoma and non-small cell lung carcinoma (NSCLC) in 2011 and 2015[1,2], respectively, 5-year survival rates have increased from less than 10% to more than 50% and 30%, respectively.[3-6] Checkpoint inhibitors have subsequently been approved for a range of malignancies with similar improvements in survival.[7]

However, the effect of checkpoint inhibitors varies significantly from patient to patient. Patients who reach complete or partial remission under therapy have a fair chance of long-term survival or even cure from metastatic disease. In melanoma patients who responded to a combination of checkpoint inhibitors, median overall survival was 6 years.[5] Non-responding patients, however, experience little-to-no benefit from treatment and have limited survival. For example, only 4% of NSCLC patients who were alive but showed progression at 6 months were still alive after 4.5 years.[7,8]

Prediction of response to treatment is a relevant topic. If non-responding patients can be identified before treatment is started, this can prevent severe and even life-threatening adverse events.[9] These severe events are especially common in patients treated with both anti-PD1 and CTLA-4 inhibitors, occurring in over 30% and 5% of NSCLC and melanoma patients, respectively.[9,10] Furthermore, accurate patient selection can reduce the high costs associated with checkpoint inhibitor therapy, which typically approach 100,000 USD per quality-adjusted life year gained .[11] Lastly prediction of non-response is relevant as these patients can, without delay, be treated with other treatments such as targeted therapy[12],

## 2.2. Introduction

---

or be enrolled in clinical trials investigating novel therapeutic approaches.

To guide treatment decisions, a biomarker must be able to identify non-responding patients with high specificity. If high specificity is not ensured, use of this biomarker alone would mean that potentially benefitting patients will not receive treatment. A potential biomarker should therefore demonstrate the ability to stratify patients into groups with a marked difference in survival and/or response.

Accurate prediction of response has proven to be a challenge, however, as we do not fully understand why this variation in response exists. Checkpoint inhibition work by blocking proteins (e.g. PD-1, PD-L1 or CTLA-4) that inhibit the body's immune response to tumors.[13] Several crucial factors in anti-tumor response have been explored as predictive markers, such as PD-L1 expression, presence of tumor infiltrating lymphocytes and tumor mutational burden.[14,15] Clinical biomarkers, for example stage of disease, WHO performance status, neutrophil-to-lymphocyte ratio and level of lactate dehydrogenase have been examined as well. None have, however, proven to be accurate enough to select patients who should not be treated with checkpoint inhibition.[16] NSCLC patients may, for instance, respond to anti-PD1 treatment even though PD-L1 expression is absent.[17]

Imaging may be able to provide additional insights into the heterogeneity in tumor response between patients. The underlying rationale for this hypothesis is that different tumor genotypes will be expressed as different imaging phenotypes. Readily available baseline imaging may therefore provide potentially valuable information about tumor size, tumor/metastasis location and, if acquired, FDG-PET parameters. Furthermore, measurements of lesion shape, intensity and texture on imaging can potentially capture information about the tumor phenotype. These measurements, collectively known as radiomics, may then subsequently be correlated to clinical outcomes.[18] Lastly, radioactive labeling of checkpoint inhibitor molecules can provide insight into the drug uptake throughout the body including in the tumor.[19]

To our knowledge, no comprehensive review has been published on the entire spectrum of prognosis research in imaging biomarkers and outcome to checkpoint inhibitors across malignancies. Earlier publications were dedicated to either a single modality (e.g. PET-imaging or radiomics) or a single malignancy.[20–23] This limits a complete overview, as advancements in one disease may very well be applicable in another. Furthermore, the predictive value of more sophisticated

## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

---

modalities (e.g. radiomics) should be compared to that of simple markers (e.g. tumor burden) to see if they add value. With this comprehensive review, we aim to fill this gap and facilitate future research.

In this work, we aimed to systematically review the ability of different imaging modalities to predict response to checkpoint inhibitors. The population of interest consists of patients treated with any checkpoint inhibitor for any malignancy. Investigated predictors are any individual biomarkers derived from imaging modalities and models including these. Outcomes of interest are response (according to RECIST[24] or iRECIST[25] criteria), overall and progression-free survival. Both prognostic and predictive factors are examined. A prognostic factor provides information about a future outcome irrespective of therapy (e.g. tumor stage for overall survival). In contrast, a predictive factor forecasts the effect of a specific treatment (e.g. estrogen receptor status for tamoxifen in breast cancer patients).[26] Despite this difference, prognostic factors are still important in guiding treatment decisions: preventing unnecessary side-effects and costs in a patient due to a very poor prognosis is no less valuable than doing so based on a pure predictive factor. For this reason, both types of factors were investigated.

## 2.3 Methods

### 2.3.1 Adherence to quality standards

This systematic review was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[27] Details of the protocol for this study were registered on PROSPERO and can be accessed at [www.crd.york.ac.uk/PROSPERO/display\\_record.asp?ID=CRD42020186199](http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42020186199).

### 2.3.2 Selection of studies

On 29 November 2021, the PubMed and Embase databases were searched for relevant studies. Other data sources were publications found from references of selected articles. Also, to ensure sensitivity of the search strategy and to identify additional relevant studies, Scopus was used. No date restrictions were applied on the systematic searches and included articles published through 29 November 2021.

## 2.3. Methods

---

Inclusion criteria for eligible articles were original full-text research articles describing baseline imaging prognostic factors and radiomics and/or imaging prediction models (e.g. using machine learning) for response and survival in patients treated with anti-PD1 checkpoint inhibition with any kind of malignancy above 18 years of age.

The literature search used the following terms (with synonyms, MeSH terms, and closely related words): “immunotherapy” or “immune checkpoint inhibitor” combined with “radiological”, “baseline factors” and “predictive”, or combined with “radiomics” or “machine learning”. We specifically adopted a broad search to include all articles related to imaging and predictive factors and to radiomics and machine learning studies. Duplicates were removed using EndNote. The complete search strategy is listed in Supplementary file 1.

All articles were screened for relevance. Studies only reported as conference abstracts without published full-text reports were not included owing to the inability to completely assess validity and methodologies. Other exclusion criteria were case reports, reviews and meta-analyses. The search was restricted to studies in human participants and papers written in English. Furthermore, studies only reporting predictive factors, radiomics or machine learning models based on on-treatment imaging (instead of pre-treatment imaging) were excluded.

### 2.3.3 Screening process

Titles and abstracts were screened for relevance by two reviewers (ID and RM) using the Rayyan QCRI web application.[28] Articles were excluded if they did not meet the inclusion criteria. Next, the selected full-text articles were assessed for eligibility by the same reviewers. Subsequently, the final selection of studies was made (Figure 1).

### 2.3.4 Critical appraisal

Two tools were used to evaluate risk of bias: the QUIPS tool[29] was used to assess studies reporting on individual prognostic or predictive factors; the PROBAST tool[30] was used to assess studies constructing models that make predictions for individual patients.

The QUIPS tool is specifically designed to assess risk of bias in prognostic

## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

---

factor studies and does so by judging the quality of a prognostic factor study on six key domains: ‘study participation’, ‘study attrition’, ‘prognostic factor measurement’, ‘outcome measurement’, ‘study confounding’ and ‘statistical analysis and reporting’. The domain ‘study attrition’ was not evaluated, as almost all studies were retrospective cohort studies that did not report on loss to follow-up during the data collection period. This domain could therefore not be accurately assessed and was consequently not used. Adaptation of the QUIPS tool for specific purposes is encouraged by the developers in the accompanying article.[29]

The PROBAST tool is designed to judge risk of bias in studies on models that make predictions for individual patients. As the PROBAST tool was developed for the appraisal of regression-type models, the authors recommend the use of additional signaling questions when evaluating studies on machine learning models.[30] The statistical analysis domain of the PROBAST tool was therefore augmented with the following three questions: (i) “Is all data from a single patient reserved to only a single data partition (e.g. training, testing or tuning)?”, (ii) “Is the optimal model selected and are hyperparameters tuned?” and (iii) “Is only the best model evaluated on the independent validation set?” (see Supplementary file 2). The remaining domains (‘Participants’, ‘Predictors’ and ‘Outcome’) were not altered.

In addition to the risk of bias assessment, the Radiomics Quality Score was used to evaluate study quality in all studies reporting on quantitative imaging derived features (radiomics). All quality assessments of the included studies were done by two independent reviewers (ID and RM). Any disagreement was resolved through discussion.

### 2.3.5 Data extraction

The following details were extracted from the studies: total number of patients investigated, cancer type, study treatment and design, imaging modality (CT, MRI or PET/CT), results and corresponding significance, and outcome. Both response to therapy (odds ratio or comparison between groups resulting in a p-value) and survival parameters (hazard ratio for progression free survival and overall survival) were obtained for the individual predictor studies. In the prediction model studies, an Area Under the Curve (AUC) or sensitivity and specificity of the model was stated, this information was also collected. For radiomics and machine learning

## 2.4. Results

---

studies, the size of the training- and validation cohorts were extracted as well.

### 2.3.6 Synthesis

The investigated prognostic factors and prediction models were grouped into six categories: tumor burden, body composition, location, FDG-PET features, other radioactive tracer imaging and radiomics. Extracted characteristics and results from all studies were grouped according to category, marker and disease. A quantitative meta-analysis was not considered feasible due to heterogeneity in population, predictor definitions and reported outcomes. The available evidence was therefore summarized based on (in order of importance) study quality, consistency of the results across studies and sample size.

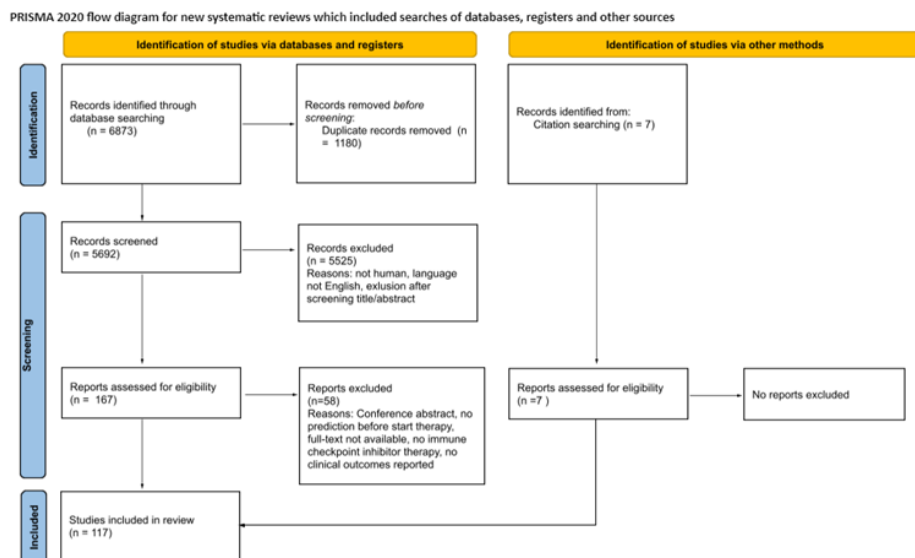
## 2.4 Results

### 2.4.1 General characteristics

The search yielded 6873 records from databases and 9 through reference screening. A total of 119 studies remained after title/abstract screening (Figure 2.1). These studies are listed in Supplementary table S1. The studies included a total of 15580 patients, with a median sample size of 74 (range 8-1461). The most studied malignancy was NSCLC (42 studies), followed by melanoma (33 studies) and urothelial carcinoma (seven studies). All but one study investigated patients with metastatic disease.

The predictive value of tumor burden was investigated by 19 papers; body composition by 24 papers; metastasis location by 18; FDG-PET features by 21; other traces by 8; radiomics by 45 papers; models other than radiomics by two (Supplementary table S1). All studies reporting on factors in the first five categories and nine radiomics studies investigated individual predictive factors. These studies were therefore assessed for risk of bias using the QUIPS tool (Supplementary table S2). The remaining studies reported performance of predictive models and were assessed for risk of bias using the PROBAST tool (Supplementary table S3). One study reported both on individual predictive factors and on a model and was assessed using both tools. The results of the RQS screening are shown in Supplementary table S4. Data extraction results are given per category in

## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.



**Figure 2.1:** PRISMA flow chart of article screening and selection

Supplementary table S5-S11. A summary table of all results is provided in Table 2.1 (available through <https://www.sciencedirect.com/science/article/pii/S0959804922004798#sec3>). A discussion of the two papers describing predictive models without the use of radiomics is provided in Supplementary file 3. For the other categories, an overview of the results is provided below.

### 2.4.2 Tumor burden

Measures of tumor burden (defined as the total amount of cancer in the body) were grouped into two categories: measures of total tumor volume (e.g. sum of largest diameters, sum of volumes) and tumor count (either number of metastases or number of affected organs). Although volume and tumor count are expected to be correlated in patients, these measures may diverge in patients with many small metastases. As this specific pattern of metastases may indicate a different tumor biology, count and volume were considered separately.

Measures of tumor volume were investigated in 15 studies.[31–44] Nine studies indicated that higher tumor volume was associated with worse survival across tumor types.[32,34–37,41–43,45] These included the three studies with the largest

## 2.4. Results

---

sample size (n=1461, n=583 and n=303) and a low risk of bias.[32,42,45] Hopkins et al. (n=1461) reported a hazard rate ratio (HR) of 1.64 for overall survival per decimeter increase of the sum of diameters of target lesions in NSCLC patients.[45] Similarly, Joseph et al. (n=583) reported a HR of 1.64 for overall survival in melanoma patients with a sum of diameters above the median.[42]

Six studies reported on the number of metastases as a prognostic factor.[32,41, 46–49] In univariate analysis, this factor was a significant prognostic factor for survival in three studies[32,46,48] with a trend towards significance in a fourth.[49] In multivariate analysis, this effect remained significant only in one paper.[46]

### 2.4.3 Body composition

Metrics of body composition were divided into four categories, namely visceral adipose tissue, subcutaneous adipose tissue, skeletal muscle quantity and skeletal muscle density.

The eight papers reporting on metrics of visceral adipose tissue showed conflicting findings: three papers demonstrated improved survival[50–52], whereas one paper reported worse survival in melanoma patients with more visceral adipose tissue.[53] The remaining papers reported no significant association with survival.[54–57] Furthermore, there were considerable methodological concerns: one paper[50] was at low, one[57] at moderate, five[52,52,53,55,56] at high and one[54] at an unclear risk of bias.

Seven papers investigated the predictive value of subcutaneous adipose tissue. The results indicated either better (4 papers)[50–52,58] or equal (3 papers)[55–57] survival in patients with higher amounts of subcutaneous fat, with hazard rate ratios for OS ranging from 0.2 to 1 at varying thresholds. Five papers[51,52,55,56, 58] were at high risk of bias, primarily due to the use of data driven optimized thresholds without validation. The risk of bias of the remaining two papers was low[50] and moderate[57].

Seventeen papers reported on various measures of skeletal muscle quantity. Eight demonstrated that higher skeletal muscle quantity was associated with better survival[59–66]; the remaining nine papers reported no significant correlation[39,53–55, 67–71]. Reported hazard rate ratios for overall survival ranged from 0.75 to 2.99. Risk of bias was low in 3[61,64,67], high in 10[39,53,55,59,63,65,



## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

---

66,68,70,71] and unclear in 4 papers[54,60,62,69], data driven thresholds again being the most common concern.

The influence of skeletal muscle density was investigated by 11 papers. Five papers indicated that higher skeletal muscle density was associated with better survival[52,53,57,70,72]; six papers reported non-significant findings[51,54,58,66,67,71]. One paper[67] was at low, one[57] at moderate, eight[51–53,58,66,70–72] were at high and one paper[54] at unclear risk of bias.

### 2.4.4 Metastasis location

In 14 papers, the presence of liver metastases was investigated.[32,33,41,42,47,48,73–80] These papers indicated that liver lesions were associated with worse survival across all tumor types, with hazard rate ratios between 1.6 and 1.9 for progression free survival in the three highest quality studies[47,48,74]. Additionally, radiological response to treatment appeared to be lower in melanoma patients with liver metastases (odds ratios between 0.3 and 0.6).[33,42,73,74] Results describing the correlation with response in other tumor types were not provided or showed no significant findings. Overall study quality varied: five studies[42,47,48,73,74] were at low risk, one[77] at high risk and eight[32,33,41,75,76,78–80] at unclear risk of bias.

Thirteen of the included studies investigated the presence of brain metastases.[32,33,41,47,48,74–76,78–82] The presence of brain metastases was not found to be a significant predictor of inferior outcomes in most studies. A notable exception was the largest and only real-world study on this topic by Van Zeyl et al. (n=583) in advanced melanoma, which showed that brain metastases in the presence of symptoms were associated with worse overall survival (HR 1.91).[47] The quality of included studies was reasonable: three studies[47,48,74] were at low risk, one[82] at high risk and nine[32,33,41,75,76,78–81] at unclear risk of bias.

Other investigated tumor locations were bone[32,33,41,48,76,78,79], lung[32,33,42,44,48,74,76,78,79], pleural effusion[48,75,76], lymph node[32,33,48,78], soft tissue[32,33], gastrointestinal[33], adrenal[33,76] and spleen[33]. None of these locations appeared to be a consistent and independent predictor of response or survival.

## 2.4. Results

---

### 2.4.5 FDG-PET features

Several FDG-PET features were investigated as potential predictors. The most reported features were standardized uptake value (SUV) (15 studies), (total) metabolic tumor volume (16 studies) and total lesion glycolysis (10 studies).

Sixteen studies examined SUV<sub>max</sub> and SUV<sub>mean</sub> of the primary lesion and metastases as prognostic factors.[31,38,40,44,49,77,81,83–91] The findings of the included studies indicated that neither SUV<sub>max</sub> nor SUV<sub>mean</sub> were robust predictors of survival: reported significant findings were sparse and conflicting. Furthermore, risk of bias was substantial: one study[38] was at low, nine studies[31,40,49,77,83,85–87,89] were at high and six studies[44,81,84,88,90,91] at unclear risk of bias.

Sixteen studies investigated total metabolic tumor volume.[38,40,46,49,50,77,77,81,83,85–90,92] Of these, eight studies demonstrated significantly worse survival in patients with higher metabolic tumor volume[38,46,49,50,77,85,92,93]. This included the largest study by Awada et al. (n=112), which was at a low risk of bias and reported a HR for OS of 1.004 per mL.[46] Considerable methodological concerns existed in the remaining studies: risk of bias was low in three studies[38,46,50], high in ten studies[40,49,77,83,85–87,89,90,92,93] and unclear in three[81,88,90]. Furthermore, two of the studies had at least a partial overlap in study population.[85,93]

Total lesion glycolysis, which is the product of SUV and metabolic tumor volume, was investigated by 11 studies.[31,38,40,77,81,83,85–87,90,93] It combines volumetric and metabolic information, and therefore presumably contains more information on the tumor than SUV and MTV. Four articles reported a significant association of total lesion glycolysis with survival[38,85,86,93], three of which studied melanoma patients.[85,86,93] Findings were not significant in the other studies. Overall risk of bias was similar to the previous markers: one study[38] was at a low risk of bias, 8 studies[31,40,77,83,85–87,93] were at high risk and two studies[81,90] at unclear risk of bias.

### 2.4.6 Other PET radioactive tracers

Other investigated tracers included sodium fluoride, F-fluorothymidine, and Zirconium labeled to different anti-PD1 antibodies, namely atezolizumab, pembrolizumab

and durvalumab.

Lim et al. investigated total lesion fluoride in genitourinary tumors and found this feature to be a significant prognostic factor for overall survival (HR 2.64).[94] Scarpelli et al. investigated the relation between tumor SUVmean and SUVtotal in F-fluorothymidine PET-CT. Neither feature was significant in the multivariate Cox-regression.[95] Furthermore, both studies were judged to be at a high risk of bias due to inadequate correction for known predictors.

Bensch et al. prospectively investigated the predictive value of Zirconium-labeled atezolizumab in various tumor types.[96] They found that increased uptake of labeled atezolizumab corresponded to better response to atezolizumab at first assessment and better overall and progression free survival (HR 6.3 and HR 11.7, respectively).

Zirconium was also used to label pembrolizumab[97,98] and durvalumab[99]. Similar results were found in these studies: increased uptake to labeled anti-PD1 corresponded with higher response and survival.

An interesting approach was performed by Van de Donk et al. Interleukin-2 was labeled to fluorine-18, in order to visualize T-cell activity by tumor infiltrating T-cells who express the high-affinity interleukin-2 receptor.[100] The tracer was safe, however no correlation with response to therapy could be found possibly due to including only 13 patients.

Another way to visualize mechanisms of PD1 inhibitors on a cellular level was carried out by Nienhuis et al.[101] They performed PET imaging in eight metastatic melanoma patients with a tracer that visualizes PD-L1 expression on the tumor. This pilot study indicated that baseline tracer uptake was associated with change in lesion size at follow up when normalized for tracer availability in the blood pool (Pearson's  $r = -0.43$ ).

### 2.4.7 Radiomics

Studies investigating radiomics were grouped according to their methodology: nine studies investigated the value of individual radiomic features; 30 studies constructed a (machine learning) model based on extracted radiomic features, and six studies trained a deep learning model.

The quality of the nine studies[34,102–109] investigating individual radiomic features was judged to be poor, as reflected in both the QUIPS rating and RQS

## 2.5. Results

---

score. Primary concerns were use of optimal thresholds, lack of independent validation and absence of correction for known predictors. Furthermore, all studies reported a significant finding, although none of the radiomic features were so far reproduced or validated in an independent study. Thus, no solid evidence exists for the predictive value of any single radiomics marker.

Similarly, all but two[110,111] of the 30 studies[110–139] that constructed a radiomics model reported a positive finding. The median reported area under the curve (AUC) for predicting response was 0.787 (range 0.52-0.963). However, numerous methodological concerns exist for these studies as well. First, a significant fraction of studies was at high (n=15) or unclear (n=6) overall risk of bias. The most common flaws were lack of correction for overfitting (ten studies) and a lack of transparency regarding model selection and tuning (11 studies). These weaknesses were affirmed by the low overall RQS, with a median score of ten out of a maximum of 36. Second, most studies had a limited sample size (median n=68). Third, the three studies with the highest RQS (RQS=24, 18 and 14) and largest sample size (n=289, 210 and 332) appeared to have a significant overlap in patient population.[116,117,123] These studies can therefore not be considered independent. Lastly, the predictions of the only radiomics model[138] that has been validated in subsequent studies[135,139] correspond closely with the presence of liver metastasis, which is a known predictor of worse outcome. As the authors did not correct for this predictor, the added value of this model is unclear and needs to be further investigated.

Six studies investigated deep learning radiomics models. Three studies were judged to be at a high risk of bias and had only small validation cohorts (41, 12 and 29 patients).[140–142] In the three remaining studies, risk of bias was judged to be low, size of the validation set was adequate (n=123, 187 and 94) and the RQS was at or above the median (13, 15 and 10).[143–145] Two of these studies appeared to have an overlap in study population.[143,144] Notably, all three studies reported on a deep learning model that was trained to predict an intermediate variable (PD-L1 expression, tumor mutational burden or EGFR mutation); patients could subsequently be stratified into risk groups with a HR for PFS of respectively 1.78 and 2.57, and OR for response of 2.03.

## 2.5 Discussion

### 2.5.1 Overview

The objective of this review was to identify imaging biomarkers in prognosis research in all cancer patients treated with checkpoint inhibitors. Based on the findings of the included studies, several groups of predictors were identified with varying strength and quality of evidence.

Higher tumor burden is very likely to be predictive of worse survival. This finding is consistently supported across tumor types by the highest quality studies on this topic. It also corresponds to our knowledge in other oncological populations undergoing other types of treatment.[146–148] Furthermore, there is a reasonable biological basis. First, higher tumor burden leads to sicker patients, and they are therefore more likely to succumb before they experience benefit from treatment. Second, hypoxia plays a bigger role in larger necrotic masses. Hypoxia is associated with immune escape and therefore worse response.[149–151] However, despite the correlation between tumor burden and survival, the reported effect sizes indicate that this marker is not strong enough to guide treatment decisions by itself and there is also insufficient evidence that tumor count adds predictive value to tumor volume.

Higher amounts of subcutaneous adipose tissue may be associated with better survival. Although the findings on visceral adipose tissue are conflicting, the results on subcutaneous adipose tissue are consistently in accordance with the so-called ‘obesity paradox’, in which a high body mass index (BMI) appears to be a protective factor in cancer patients.[152–155] It must be noted, however, that the reported results may be an overestimation of the true effect, as reflected in the risk of bias assessment. Furthermore, it is unknown whether the value of this predictor is independent from simple clinical metrics, such as BMI. It is therefore deemed unlikely that this marker will further impact clinical decision making in the near future.

More and denser muscle may be predictive of better survival. The findings of the included studies on this topic are supported by similar observations in other oncological populations.[156–158] Again, however, there is a risk that the observed effect is an overestimate due to biased analysis. Furthermore, the reported effect sizes appear to be smaller in the larger studies, indicating that publication bias

## 2.5. Discussion

---

may play a role. In conclusion, the association of muscle density and quantity with survival is plausible as they indicate fitter patients with more reserve, but currently investigated parameters may be only of limited predictive value.

Presence of liver metastases is shown to be a marker of worse survival across cancer types. This marker, too, is an indicator of more advanced disease with spread to the visceral organs. Interestingly, several large, high-quality studies in melanoma patients show that presence of liver metastases also predicts worse response compared to metastasis in other organs. Whether this is due to liver metastases being less responsive, or to patients with liver metastases being innately different, is the topic of an emerging field of research. In pre-clinical models, several hepatic cell types have shown to modulate T-cells in the liver and create a systemic immune desert.[159] Furthermore, systemic T-cell loss and diminished immunotherapy efficacy has been observed in patients with liver metastases.[159]

Symptomatic brain metastases may be associated with worse survival in melanoma. No significant impact of the presence of asymptomatic brain metastases was observed in most of the included studies. However, almost all included studies on this topic investigated trial populations, in which patients with brain metastases were excluded. The study conducted by Van Zeyl et al., however, examined real-world data and demonstrated that symptomatic brain metastases were associated with worse survival in melanoma patients.[47] As previous studies have shown that checkpoint inhibitors are effective against brain metastases, this difference in survival is likely to be caused by more frequent neurological complications.[160]

The added value of baseline FDG-PET features in predicting response to treatment seems to be limited. Of the investigated PET-features, only higher total MTV was consistently shown to be associated with worse survival. However, since metabolic and morphological tumor volume (MTV) are at least partly associated and none of the included studies corrected for morphological tumor burden, it is unclear if MTV is of added predictive value. Significant findings about other FDG-PET derived metrics (SUVmax, SUVmean and TLG) are scarcer and were often at a high risk of bias.

Radioactive drug labeling appears promising, although current evidence is very preliminary. The hypothesis that uptake corresponds to response has a very strong biological basis. Furthermore, the reported results from small pilot cohorts are promising. However, it remains to be investigated if the positive results will gen-

eralize to larger sample sizes and if they will be independent of known predictors.

The value of radiomics remains unknown due to the lack of high-quality evidence. Although the results of the included papers on radiomics are almost exclusively positive, the reported findings are likely to be over-optimistic for several reasons. First, methodological flaws may have led to an overestimation of the predictive value of the described models. Second, the aggregated results are likely to be additionally affected by publication bias. Arguably, studies into radiomics are at an even higher risk of publication bias: while negative findings about traditional markers may be informative, a negative finding about a radiomics model can be viewed as ‘a complex machine that does not work’. This, in combination with limited sample sizes in included papers and repeated publications on very similar datasets, may have considerably skewed the aggregate results. Third, many radiomics features are sensitive to variation in scanner type and protocol between centers.[161] This variation may therefore reduce the predictive value of the proposed models to only a fraction of what is shown. In conclusion, the positive findings of the few high-quality papers are promising, especially those that use an intermediate endpoint for training. These findings, however, remain to be confirmed through external and prospective validation.

### **2.5.2 Future research**

The predictive value of imaging biomarkers may improve through future developments. Specifically, we believe that subsequent research should focus on three key areas. First, imaging biomarkers should be integrated with predictors from other modalities. As no single biomarker has yet been proven to be sufficient for effectively guiding treatment decisions, we must investigate combinations of multiple - uncorrelated - predictors. Concretely, this can be envisioned as a multivariate prediction model combining imaging biomarkers with clinical, histological, biochemical and genetic predictors, among others. Second, the added value of radioactive drug labeling should be explored in larger studies. These studies should also particularly report on the added value of this biomarker over known predictors. In addition, negative results about these markers would also be very beneficial in advancing the field of research, due to the efforts and costs needed to produce these tracers. Third, new studies should more closely adhere to methodological guidelines and should confirm previous findings through rigorous validation. This

## 2.5. Discussion

---

is especially the case for radiomics studies, of which the impact is currently limited by methodological shortcomings. If, however, radiomics are proven to be independent predictors, they would be able to provide us with valuable information at no additional cost or harm to the patient.

### 2.5.3 Limitations

The first main limitation of this review is the lack of a universally agreed upon tool to assess risk of bias in machine learning studies. We used a combination of the PROBAST tool and RQS to assess the quality of the radiomics studies. Both tools, however, have limitations for this purpose. The PROBAST tool addresses most domains that put a machine learning model at a risk of bias, but not all. The PROBAST-AI tool is currently under development to meet this need.[162] Furthermore, the RQS provides excellent guidance in the design of a good radiomics study, but is not intended for scrutinizing papers to detect a possible risk of bias.

The second main limitation is the lack of a quantitative meta-analysis, due to the differences in definition of predictor or outcome in the included studies. Significant variation regarding predictors exists, often caused by dichotomizing continuous values at various thresholds. This, in combination with the fragmentation of evidence across different diseases and treatments, makes a quantitative analysis essentially impossible. We were therefore unable to quantify the predictive power of the investigated markers. We do, however, think that there is enough ground for the conclusion that no individual imaging-based biomarker is proven to be sufficient.

### 2.5.4 Conclusion

In conclusion, there is well-supported evidence for several imaging biomarkers of response to checkpoint inhibitors. Especially higher tumor burden and the presence of liver metastases are demonstrated to be predictors of worse outcomes across malignancies and drugs. However, none of these single predictors seem strong enough to reliably identify patients that will not derive benefit from treatment. A high degree of accuracy is required for this purpose, as falsely designating a patient as a non-responder would deny a patient access to long-term ICI. Radiomics and radioactive drug labeling appear to be very promising, although reported findings on



## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

---

these approaches should be regarded as preliminary at this moment. In addition to further validation of these methods, future research should focus on integrating imaging biomarkers with predictors from other modalities in high-quality and sufficiently large independent cohorts.

## 2.6 Supplementary Materials

Supplementary File 1:

<https://ars.els-cdn.com/content/image/1-s2.0-S0959804922004798-mmc1.docx>

Supplementary File 2:

<https://ars.els-cdn.com/content/image/1-s2.0-S0959804922004798-mmc2.xlsx>

## 2.7 References

- [1 ] Fellner C. Ipilimumab (Yervoy) Prolongs Survival In Advanced Melanoma. *Pharm Ther* 2012;37:503–30.
- [2 ] Hargadon KM, Johnson CE, Williams CJ. Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors. *Int Immunopharmacol* 2018;62:29–39. <https://doi.org/10.1016/j.intimp.2018.06.001>
- [3 ] Weiss SA, Wolchok JD, Sznol M. Immunotherapy of Melanoma: Facts and Hopes. *Clin Cancer Res* 2019;25:5191–201. <https://doi.org/10.1158/1078-0432.CCR-18-1550>
- [4 ] Garon EB, Rizvi NA, Hui R, Leigh N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the Treatment of Non–Small-Cell Lung Cancer. *N Engl J Med* 2015;372:2018–28. <https://doi.org/10.1056/NEJMoa1501824>.
- [5 ] Wolchok JD, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Long-Term Outcomes With Nivolumab Plus Ipilimumab or Nivolumab Alone Versus Ipilimumab in Patients With Advanced Melanoma. *J Clin Oncol* 2022;40:127–37. <https://doi.org/10.1200/jco.21.02229>
- [6 ] Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Five-Year Outcomes With Pembrolizumab Versus Chemotherapy for Metastatic Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score 50. *J Clin Oncol Off J Am Soc Clin Oncol* 2021;39:2339–49. <https://doi.org/10.1200/JCO.21.00174>
- [7 ] Vaddepally RK, Kharel P, Pandey R, Garje R, Chandra AB. Review of Indications of FDA-Approved Immune Checkpoint Inhibitors per NCCN Guidelines with the Level of Evidence. *Cancers* 2020;12:738. <https://doi.org/10.3390/cancers12030738>
- [8 ] Antonia SJ, Borghaei H, Ramalingam SS, Horn L, De Castro Carpeño J, Pluzanski A, et al. Four-year survival with nivolumab in patients with previously treated advanced non-small-cell lung cancer: a pooled analysis. *Lancet Oncol* 2019;20:1395–408. [https://doi.org/10.1016/S1470-2045\(19\)30407-3](https://doi.org/10.1016/S1470-2045(19)30407-3)
- [9 ] Postow MA, Sidlow R, Hellmann MD. Immune-Related Adverse Events Associated with Immune Checkpoint Blockade. *N Engl J Med* 2018;378:158–68. <https://doi.org/10.1056/NEJMr1703481>.
- [10 ] Hellmann MD, Paz-Ares L, Bernabe Caro R, Zurawski B, Kim S-W, Carcereny Costa E, et al. Nivolumab plus Ipilimumab in Advanced Non–Small-Cell Lung Cancer. *N Engl J Med* 2019;381:2020–31. <https://doi.org/10.1056/NEJNoa1910231>

## 2.7. References

---

- [11] Verma V, Sprave T, Haque W, Simone CB, Chang JY, Welsh JW, et al. A systematic review of the cost and cost-effectiveness studies of immune checkpoint inhibitors. *J Immunother Cancer* 2018;6:128. <https://doi.org/10.1186/s40425-018-0442-7>
- [12] Luke JJ, Flaherty KT, Ribas A, Long GV. Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol* 2017;14:463–82. <https://doi.org/10.1038/nrclinonc.2017.43>
- [13] Granier C, De Guillebon E, Blanc C, Roussel H, Badoual C, Colin E, et al. Mechanisms of action and rationale for the use of checkpoint inhibitors in cancer. *ESMO Open* 2017;2:e000213. <https://doi.org/10.1136/esmoopen-2017-000213>
- [14] Morrison C, Pabla S, Conroy JM, Nesline MK, Glenn ST, Dressman D, et al. Predicting response to checkpoint inhibitors in melanoma beyond PD-L1 and mutational burden. *J Immunother Cancer* 2018;6:32. <https://doi.org/10.1186/s40425-018-0344-8>
- [15] Blank CU, Haanen JB, Ribas A, Schumacher TN. The “cancer immunogram.” *Science* 2016;352:658–60. <https://doi.org/10.1126/science.aaf2834>
- [16] Jessurun CAC, Vos JAM, Limpens J, Luiten RM. Biomarkers for Response of Melanoma Patients to Immune Checkpoint Inhibitors: A Systematic Review. *Front Oncol* 2017;7:233. <https://doi.org/10.3389/fonc.2017.00233>
- [17] Conroy JM, Pabla S, Nesline MK, Glenn ST, Papanicolaou-Sengos A, Burgher B, et al. Next generation sequencing of PD-L1 for predicting response to immune checkpoint inhibitors. *J Immunother Cancer* 2019;7:18. <https://doi.org/10.1186/s40425-018-0489-5>
- [18] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278:563–77. <https://doi.org/10.1148/radiol.2015151169>
- [19] Rudin M, Weissleder R. Molecular imaging in drug discovery and development. *Nat Rev Drug Discov* 2003;2:123–31. <https://doi.org/10.1038/nrd1007>
- [20] Zhang C, de A. F. Fonseca L, Shi Z, Zhu C, Dekker A, Bermejo I, et al. Systematic review of radiomic biomarkers for predicting immune checkpoint inhibitor treatment outcomes. *Methods* 2021;188:61–72. <https://doi.org/10.1016/j.ymeth.2020.11.005>
- [21] Ayati N, Sadeghi R, Kiamanesh Z, Lee ST, Zakavi SR, Scott AM. The value of 18F-FDG PET/CT for predicting or monitoring immunotherapy response in patients with metastatic melanoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging* 2021;48:428–48. <https://doi.org/10.1007/s00259-020-04967-9>
- [22] Chen Q, Zhang L, Mo X, You J, Chen L, Fang J, et al. Current status and quality of radiomic studies for predicting immunotherapy response and outcome in patients with non-small cell lung cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging* 2021;49:345–60. <https://doi.org/10.1007/s00259-021-05509-7>
- [23] Guerrisi A, Loi E, Ungania S, Russillo M, Bruzzaniti V, Elia F, et al. Novel cancer therapies for advanced cutaneous melanoma: The added value of radiomics in the decision making process—A systematic review. *Cancer Med* 2020;9:1603–12. <https://doi.org/10.1002/cam4.2709>
- [24] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47. <https://doi.org/10.1016/j.ejca.2008.10.026>
- [25] Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* 2017;18:e143–52. [https://doi.org/10.1016/S1473-2045\(17\)30074-8](https://doi.org/10.1016/S1473-2045(17)30074-8)
- [26] Riley RD, Windt D van der, Croft P, Moons KGM. *Prognosis Research in Healthcare: Concepts, Methods, and Impact*. Oxford University Press; 2019.
- [27] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int J Surg* 2021;88:105906. <https://doi.org/10.1016/j.ijsu.2021.105906>
- [28] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210. <https://doi.org/10.1186/s13643-016-0384-4>
- [29] Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158:280–6. <https://doi.org/10.7326/0003-4819-158-4-201302190-00009>

## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

---

- [30 ] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170:51–8. <https://doi.org/10.7326/M18-1376>.
- [31 ] Kudura K, Dimitriou F, Basler L, Förster R, Mihic-Probst D, Kutzker T, et al. Prediction of Early Response to Immune Checkpoint Inhibition Using FDG-PET/CT in Melanoma Patients. *Cancers* 2021;13:3830. <https://doi.org/10.3390/cancers13153830>
- [32 ] Davis EJ, Perez MC, Ayoubi N, Zhao S, Ye F, Wang DY, et al. Clinical correlates of response to anti-PD-1-based therapy in patients with metastatic melanoma. *J Immunother Hagerstown Md* 1997 2019;42:221–7. <https://doi.org/10.1097/CJI.0000000000000258>
- [33 ] Pires da Silva I, Lo S, Quek C, Gonzalez M, Carlino MS, Long GV, et al. Site-specific response patterns, pseudoprogression, and acquired resistance in patients with melanoma treated with ipilimumab combined with anti-PD-1 therapy. *Cancer* 2020;126:86–97. <https://doi.org/10.1002/cncr.32522>.
- [34 ] Schraag A, Klumpp B, Afat S, Gatidis S, Nikolaou K, Eigentler TK, et al. Baseline clinical and imaging predictors of treatment response and overall survival of patients with metastatic melanoma undergoing immunotherapy. *Eur J Radiol* 2019;121:108688. <https://doi.org/10.1016/j.ejrad.2019.108688>
- [35 ] Nishino M, Giobbie-Hurder A, Ramaiya NH, Hodi FS. Response assessment in metastatic melanoma treated with ipilimumab and bevacizumab: CT tumor size and density as markers for response and outcome. *J Immunother Cancer* 2014;2:40. <https://doi.org/10.1186/s40425-014-0040-2>.
- [36 ] Sakata Y, Kawamura K, Ichikado K, Shingu N, Yasuda Y, Eguchi Y, et al. Comparisons between tumor burden and other prognostic factors that influence survival of patients with non-small cell lung cancer treated with immune checkpoint inhibitors. *Thorac Cancer* 2019;10:2259–66. <https://doi.org/10.1111/1759-7714.13214>
- [37 ] Bureau M, Chatellier T, Perennec T, Goronflot T, Greilsamer C, Chene A-L, et al. Baseline tumour size is an independent prognostic factor for overall survival in PD-L150% non-small cell lung cancer patients treated with first-line pembrolizumab. *Cancer Immunol Immunother* 2021;71:1747–56. <https://doi.org/10.1007/s00262-021-03108-x>
- [38 ] Hashimoto K, Kaira K, Yamaguchi O, Mouri A, Shiono A, Miura Y, et al. Potential of FDG-PET as Prognostic Significance after anti-PD-1 Antibody against Patients with Previously Treated Non-Small Cell Lung Cancer. *J Clin Med* 2020;9:725. <https://doi.org/10.3390/jcm9030725>.
- [39 ] Dercle L, Ammari S, Champiat S, Massard C, Ferté C, Taihi L, et al. Rapid and objective CT scan prognostic scoring identifies metastatic patients with long-term clinical benefit on anti-PD-1/-L1 therapy. *Eur J Cancer* 2016;65:33–42. <https://doi.org/10.1016/j.ejca.2016.05.031>
- [40 ] Jreige M, Letovanec I, Chaba K, Renaud S, Rusakiewicz S, Cristina V, et al. 18F-FDG PET metabolic-to-morphological volume ratio predicts PD-L1 tumour expression and response to PD-1 blockade in non-small-cell lung cancer. *Eur J Nucl Med Mol Imaging* 2019;46:1859–68. <https://doi.org/10.1007/s00259-019-04348-x>
- [41 ] Katsurada M, Nagano T, Tachihara M, Kiriu T, Furukawa K, Koyama K, et al. Baseline Tumor Size as a Predictive and Prognostic Factor of Immune Checkpoint Inhibitor Therapy for Non-small Cell Lung Cancer. *Anticancer Res* 2019;39:815–25. <https://doi.org/10.21873/anticancer.13180>.
- [42 ] Joseph RW, Elassaiss-Schaap J, Kefford R, Hwu W-J, Wolchok JD, Joshua AM, et al. Baseline Tumor Size Is an Independent Prognostic Factor for Overall Survival in Patients with Melanoma Treated with Pembrolizumab. *Clin Cancer Res* 2018;24:4960–7. <https://doi.org/10.1158/1078-0432.CCR-17-2386>
- [43 ] Inoue H, Yokota T, Hamauchi S, Onozawa Y, Kawakami T, Shirasu H, et al. Pre-treatment tumor size impacts on response to nivolumab in head and neck squamous cell carcinoma. *Auris Nasus Larynx* 2020;47:650–7. <https://doi.org/10.1016/j.anl.2020.01.003>.
- [44 ] Tabei T, Nakaigawa N, Kaneta T, Ikeda I, Kondo K, Makiyama K, et al. Early assessment with 18F-2-fluoro-2-deoxyglucose positron emission tomography/computed tomography to predict short-term outcome in clear cell renal carcinoma treated with nivolumab. *BMC Cancer* 2019;19:298. <https://doi.org/10.1186/s12885-019-5510-y>.
- [45 ] Hopkins AM, Kichenadasse G, McKinnon RA, Rowland A, Soric MJ. Baseline tumor size and survival outcomes in lung cancer patients treated with immune checkpoint inhibitors. *Semin Oncol* 2019;46:380–4. <https://doi.org/10.1053/j.seminoncol.2019.10.002>
- [46 ] Awada G, Jansen Y, Schwarze JK, Tijtgat J, Hellinckx L, Gondry O, et al. A Comprehensive Analysis of Baseline Clinical Characteristics and Biomarkers Associated with Outcome in Advanced Melanoma Patients Treated with Pembrolizumab. *Cancers* 2021;13:168. <https://doi.org/10.3390/cancers13020168>.

## 2.7. References

---

- [47] van Zeijl MCT, Haanen JBAG, Wouters MWJM, de Wreede LC, Jochems A, Aarts MJB, et al. Real-world Outcomes of First-line Anti-PD-1 Therapy for Advanced Melanoma: A Nationwide Population-based Study. *J Immunother* 2020;43:256–64. <https://doi.org/10.1097/CJI.0000000000000334>
- [48] Tamiya M, Tamiya A, Inoue T, Kimura M, Kunimasa K, Nakahama K, et al. Metastatic site as a predictor of nivolumab efficacy in patients with advanced non-small cell lung cancer: A retrospective multicenter trial. *PLOS ONE* 2018;13:e0192227. <https://doi.org/10.1371/journal.pone.0192227>
- [49] Seban R-D, Mezquita L, Berenbaum A, Derclé L, Botticella A, Le Pechoux C, et al. Baseline metabolic tumor burden on FDG PET/CT scans predicts outcome in advanced NSCLC patients treated with immune checkpoint inhibitors. *Eur J Nucl Med Mol Imaging* 2020;47:1147–57. <https://doi.org/10.1007/s00259-019-04615-x>.
- [50] Popinat G, Cousse S, Raitière O, Gardin I, Vera P, Guisier F, et al. *Med Nucleaire* 2019;43:216–7. <https://doi.org/10.1016/j.mednuc.2019.01.123>
- [51] Martini DJ, Olsen TA, Goyal S, Liu Y, Evans ST, Magod B, et al. Body Composition Variables as Radiographic Biomarkers of Clinical Outcomes in Metastatic Renal Cell Carcinoma Patients Receiving Immune Checkpoint Inhibitors. *Front Oncol* 2021;11:707050. <https://doi.org/10.3389/fonc.2021.707050>.
- [52] Martini DJ, Shabto JM, Goyal S, Liu Y, Olsen TA, Evans ST, et al. Body Composition as an Independent Predictive and Prognostic Biomarker in Advanced Urothelial Carcinoma Patients Treated with Immune Checkpoint Inhibitors. *The Oncologist* 2021;26:1017–25. <https://doi.org/10.1002/onco.13922>
- [53] Sabel MS, Lee J, Wang A, Lao C, Holcombe S, Wang S. Morphomics predicts response to ipilimumab in patients with stage IV melanoma. *J Surg Oncol* 2015;112:333–7. <https://doi.org/10.1002/jso.24003>.
- [54] Minami S, Ihara S, Tanaka T, Komuta K. Sarcopenia and Visceral Adiposity Did Not Affect Efficacy of Immune-Checkpoint Inhibitor Monotherapy for Pretreated Patients With Advanced Non-Small Cell Lung Cancer. *World J Oncol* 2020;11:9–22. <https://doi.org/10.14740/wjon1225>
- [55] Crombé A, Kind M, Toulmonde M, Italiano A, Cousin S. Impact of CT-based body composition parameters at baseline, their early changes and response in metastatic cancer patients treated with immune checkpoint inhibitors. *Eur J Radiol* 2020;133:109340. <https://doi.org/10.1016/j.ejrad.2020.109340>
- [56] Esposito A, Marra A, Bagnardi V, Frassoni S, Morganti S, Viale G, et al. Body mass index, adiposity and tumour infiltrating lymphocytes as prognostic biomarkers in patients treated with immunotherapy: A multi-parametric analysis. *Eur J Cancer* 2021;145:197–209. <https://doi.org/10.1016/j.ejca.2020.12.028>.
- [57] Deike-Hofmann K, Gutzweiler L, Reuter J, Paech D, Hassel JC, Sedlacek O, et al. Macroangiopathy is a positive predictive factor for response to immunotherapy. *Sci Rep* 2019;9:9728. <https://doi.org/10.1038/s41598-019-46189-6>
- [58] Martini DJ, Kline MR, Liu Y, Shabto JM, Williams MA, Khan AI, et al. Adiposity may predict survival in patients with advanced stage cancer treated with immunotherapy in phase 1 clinical trials. *Cancer* 2020;126:575–82. <https://doi.org/10.1002/cncr.32576>.
- [59] Araki T, Kitaguchi Y, Suzuki Y, Komatsu M, Sonehara K, Wada Y, et al. Prognostic implication of erector spinae muscles in non-small-cell lung cancer patients treated with immuno-oncology combinatorial chemotherapy. *Thorac Cancer* 2021;12:2857–64. <https://doi.org/10.1111/1759-7714.14142>
- [60] Takada K, Yoneshima Y, Tanaka K, Okamoto I, Shimokawa M, Wakasu S, et al. Clinical impact of skeletal muscle area in patients with non-small cell lung cancer treated with anti-PD-1 inhibitors. *J Cancer Res Clin Oncol* 2020;146:1217–25. <https://doi.org/10.1007/s00432-020-03146-5>
- [61] Shiroyama T, Nagatomo I, Koyama S, Hirata H, Nishida S, Miyake K, et al. Impact of sarcopenia in patients with advanced non–small cell lung cancer treated with PD-1 inhibitors: A preliminary retrospective study. *Sci Rep* 2019;9:2447. <https://doi.org/10.1038/s41598-019-39120-6>
- [62] Arribas L, Plana M, Taberna M, Sospedra M, Vilariño N, Oliva M, et al. Predictive Value of Skeletal Muscle Mass in Recurrent/Metastatic Head and Neck Squamous Cell Carcinoma Patients Treated With Immune Checkpoint Inhibitors. *Front Oncol* 2021;11.
- [63] Fukushima H, Fukuda S, Moriyama S, Uehara S, Yasuda Y, Tanaka H, et al. Impact of sarcopenia on the efficacy of pembrolizumab in patients with advanced urothelial carcinoma: a preliminary report. *Anticancer Drugs* 2020;31:866–71. <https://doi.org/10.1097/CAD.0000000000000982>

## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

---

- [64 ] Shimizu T, Miyake M, Hori S, Ichikawa K, Omori C, Iemura Y, et al. Clinical Impact of Sarcopenia and Inflammatory/Nutritional Markers in Patients with Unresectable Metastatic Urothelial Carcinoma Treated with Pembrolizumab. *Diagnostics* 2020;10:310. <https://doi.org/10.3390/diagnostics10050310>.
- [65 ] Kim Y-Y, Lee J, Jeong WK, Kim ST, Kim J-H, Hong JY, et al. Prognostic significance of sarcopenia in microsatellite-stable gastric cancer patients treated with programmed death-1 inhibitors. *Gastric Cancer* 2021;24:457–66. <https://doi.org/10.1007/s10120-020-01124-x>
- [66 ] Cortellini A, Bozzetti F, Palumbo P, Brocco D, Di Marino P, Tinari N, et al. Weighing the role of skeletal muscle mass and muscle density in cancer patients receiving PD-1/PD-L1 checkpoint inhibitors: a multicenter real-life study. *Sci Rep* 2020;10:1456. <https://doi.org/10.1038/s41598-020-58498-2>.
- [67 ] Young AC, Quach HT, Song H, Davis EJ, Moslehi JJ, Ye F, et al. Impact of body composition on outcomes from anti-PD1 +/- anti-CTLA-4 treatment in melanoma. *J Immunother Cancer* 2020;8:e000821. <https://doi.org/10.1136/jitc-2020-000821>.
- [68 ] Cortellini A, Verna L, Porzio G, Bozzetti F, Palumbo P, Masciocchi C, et al. Predictive value of skeletal muscle mass for immunotherapy with nivolumab in non-small cell lung cancer patients: A “hypothesis-generator” preliminary report. *Thorac Cancer* 2019;10:347–51. <https://doi.org/10.1111/1759-7714.12965>
- [69 ] Magri V, Gottfried T, Di Segni M, Urban D, Peled M, Daher S, et al. Correlation of body composition by computerized tomography and metabolic parameters with survival of nivolumab-treated lung cancer patients. *Cancer Manag Res* 2019;11:8201–7. <https://doi.org/10.2147/CMAR.S210958>.
- [70 ] Nishioka N, Naito T, Notsu A, Mori K, Kodama H, Miyawaki E, et al. Unfavorable impact of decreased muscle quality on the efficacy of immunotherapy for advanced non-small cell lung cancer. *Cancer Med* 2021;10:247–56. <https://doi.org/10.1002/cam4.3631>
- [71 ] Loosen SH, van den Bosch V, Gorgulho J, Schulze-Hagen M, Kandler J, Jördens MS, et al. Progressive Sarcopenia Correlates with Poor Response and Outcome to Immune Checkpoint Inhibitor Therapy. *J Clin Med* 2021;10:1361. <https://doi.org/10.3390/jcm10071361>
- [72 ] Youn S, Reif R, Chu MP, Smylie M, Walker J, Eurich DT, et al. Myosteotosis is prognostic in metastatic melanoma treated with nivolumab. *Clin Nutr ESPEN* 2021;42:348–53. <https://doi.org/10.1016/j.clnesp.2021.01.009>
- [73 ] Nosrati A, Tsai KK, Goldinger SM, Tumeh P, Grimes B, Loo K, et al. Evaluation of clinicopathological factors in PD-1 response: derivation and validation of a prediction scale for response to PD-1 monotherapy. *Br J Cancer* 2017;116:1141–7. <https://doi.org/10.1038/bjc.2017.70>.
- [74 ] Tumeh PC, Hellmann MD, Hamid O, Tsai KK, Loo KL, Gubens MA, et al. Liver Metastasis and Treatment Outcome with Anti-PD-1 Monoclonal Antibody in Patients with Melanoma and NSCLC. *Cancer Immunol Res* 2017;5:417–24. <https://doi.org/10.1158/2326-6066.CCR-16-0325>
- [75 ] Adachi Y, Tamiya A, Taniguchi Y, Enomoto T, Azuma K, Kouno S, et al. Predictive factors for progression-free survival in non-small cell lung cancer patients receiving nivolumab based on performance status. *Cancer Med* 2020;9:1383–91. <https://doi.org/10.1002/cam4.2807>.
- [76 ] Kawachi H, Tamiya M, Tamiya A, Ishii S, Hirano K, Matsumoto H, et al. Association between metastatic sites and first-line pembrolizumab treatment outcome for advanced non-small cell lung cancer with high PD-L1 expression: a retrospective multicenter cohort study. *Invest New Drugs* 2020;38:211–8. <https://doi.org/10.1007/s10637-019-00882-5>
- [77 ] Seban R-D, Assie J-B, Giroux-Leprieur E, Massiani M-A, Soussan M, Bonardel G, et al. FDG-PET biomarkers associated with long-term benefit from first-line immunotherapy in patients with advanced non-small cell lung cancer. *Ann Nucl Med* 2020;34:968–74. <https://doi.org/10.1007/s12149-020-01539-7>
- [78 ] Bilen MA, Shabto JM, Martini DJ, Liu Y, Lewis C, Collins H, et al. Sites of metastasis and association with clinical outcome in advanced stage cancer patients treated with immunotherapy. *BMC Cancer* 2019;19:857. <https://doi.org/10.1186/s12885-019-6073-7>
- [79 ] Sen S, Hess K, Hong DS, Naing A, Piha-Paul S, Janku F, et al. Development of a prognostic scoring system for patients with advanced cancer enrolled in immune checkpoint inhibitor phase 1 clinical trials. *Br J Cancer* 2018;118:763–9. <https://doi.org/10.1038/bjc.2017.480>
- [80 ] Botticelli A, Cirillo A, Scagnoli S, Cerbelli B, Strigari L, Cortellini A, et al. The Agnostic Role of Site of Metastasis in Predicting Outcomes in Cancer Patients Treated with Immunotherapy. *Vaccines* 2020;8:203. <https://doi.org/10.3390/vaccines8020203>

## 2.7. References

---

- [81] Nakamoto R, Zaba LC, Liang T, Reddy SA, Davidzon G, Aparici CM, et al. Prognostic Value of Bone Marrow Metabolism on Pretreatment 18F-FDG PET/CT in Patients with Metastatic Melanoma Treated with Anti-PD-1 Therapy. *J Nucl Med* 2021;62:1380–3. <https://doi.org/10.2967/jnumed.120.254482>
- [82] Cowey CL, Liu FX, Black-Shinn J, Stevinson K, Boyd M, Frytak JR, et al. Pembrolizumab Utilization and Outcomes for Advanced Melanoma in US Community Oncology Practices. *J Immunother* 2018;41:86–95. <https://doi.org/10.1097/CJI.0000000000000204>
- [83] Nobashi T, Baratto L, Reddy SA, Srinivas S, Torihara A, Hatami N, et al. Predicting Response to Immunotherapy by Evaluating Tumors, Lymphoid Cell-Rich Organs, and Immune-Related Adverse Events Using FDG-PET/CT. *Clin Nucl Med* 2019;44:e272. <https://doi.org/10.1097/RLU.0000000000002453>
- [84] Takada K, Toyokawa G, Yoneshima Y, Tanaka K, Okamoto I, Shimokawa M, et al. 18F-FDG uptake in PET/CT is a potential predictive biomarker of response to anti-PD-1 antibody therapy in non-small cell lung cancer. *Sci Rep* 2019;9:13362. <https://doi.org/10.1038/s41598-019-50079-2>
- [85] Seban R-D, Nemer JS, Marabelle A, Yeh R, Deutsch E, Ammari S, et al. Prognostic and theranostic 18F-FDG PET biomarkers for anti-PD1 immunotherapy in metastatic melanoma: association with outcome and transcriptomics. *Eur J Nucl Med Mol Imaging* 2019;46:2298–310. <https://doi.org/10.1007/s00259-019-04411-7>
- [86] Sanli Y, Leake J, Odu A, Xi Y, Subramaniam RM. Tumor Heterogeneity on FDG PET/CT and Immunotherapy: An Imaging Biomarker for Predicting Treatment Response in Patients With Metastatic Melanoma. *Am J Roentgenol* 2019;212:1318–26. <https://doi.org/10.2214/AJR.18.19796>
- [87] Evangelista L, Cuppari L, Menis J, Bonanno L, Reccia P, Frega S, et al. 18F-FDG PET/CT in non-small-cell lung cancer patients: a potential predictive biomarker of response to immunotherapy. *Nucl Med Commun* 2019;40:802–7. <https://doi.org/10.1097/MMN.0000000000001025>
- [88] Wong A, Callahan J, Keyaerts M, Neyns B, Mangana J, Aberle S, et al. 18F-FDG PET/CT based spleen to liver ratio associates with clinical outcome to ipilimumab in patients with metastatic melanoma. *Cancer Imaging* 2020;20:36. <https://doi.org/10.1186/s40644-020-00313-2>
- [89] Zhang S, Zhang R, Gong W, Wang C, Zeng C, Zhai Y, et al. Positron Emission Tomography-Computed Tomography Parameters Predict Efficacy of Immunotherapy in Head and Neck Squamous Cell Carcinomas. *Front Oncol* 2021;11:728040. <https://doi.org/10.3389/fonc.2021.728040>
- [90] Vekens K, Everaert H, Neyns B, Ilsen B, Decoster L. The Value of 18F-FDG PET/CT in Predicting the Response to PD-1 Blocking Immunotherapy in Advanced NSCLC Patients with High-Level PD-L1 Expression. *Clin Lung Cancer* 2021;22:432–40. <https://doi.org/10.1016/j.clcc.2021.03.001>
- [91] Ichiki Y, Taira A, Chikaishi Y, Matsumiya H, Mori M, Kanayama M, et al. Prognostic factors of advanced or postoperative recurrent non-small cell lung cancer targeted with immune check point inhibitors. *J Thorac Dis* 2019;11:1117–23. <https://doi.org/10.21037/jtd.2019.04.41>
- [92] Flaus A, Habouzit V, De Leiris N, Vuillez JP, Leccia MT, Perrot JL, et al. FDG PET biomarkers for prediction of survival in metastatic melanoma prior to anti-PD1 immunotherapy. *Sci Rep* 2021;11:18795. <https://doi.org/10.1038/s41598-021-98310-3>
- [93] Seban R-D, Moya-Plana A, Antonios L, Yeh R, Marabelle A, Deutsch E, et al. Prognostic 18F-FDG PET biomarkers in metastatic mucosal and cutaneous melanoma treated with immune checkpoint inhibitors targeting PD-1 and CTLA-4. *Eur J Nucl Med Mol Imaging* 2020;47:2301–12. <https://doi.org/10.1007/s00259-020-04757-3>
- [94] Lim I, Lindenberg ML, Mena E, Verdini N, Shih JH, Mayfield C, et al. 18F-Sodium fluoride PET/CT predicts overall survival in patients with advanced genitourinary malignancies treated with cabozantinib and nivolumab with or without ipilimumab. *Eur J Nucl Med Mol Imaging* 2020;47:178–84. <https://doi.org/10.1007/s00259-019-04483-5>
- [95] Scarpelli M, Zahm C, Perlman S, McNeel DG, Jeraj R, Liu G. FLT PET/CT imaging of metastatic prostate cancer patients treated with pTVG-HP DNA vaccine and pembrolizumab. *J Immunother Cancer* 2019;7:23. <https://doi.org/10.1186/s40425-019-0516-1>
- [96] Bensch F, van der Veen EL, Lub-de Hooge MN, Jorritsma-Smit A, Boellaard R, Kok IC, et al. 89Zr-atezolizumab imaging as a non-invasive approach to assess clinical response to PD-L1 blockade in cancer. *Nat Med* 2018;24:1852–8. <https://doi.org/10.1038/s41591-018-0255-8>
- [97] Kok IC, Hooiveld JS, van de Donk PP, Giesen D, van der Veen EL, Lub-de Hooge MN, et al. 89Zr-pembrolizumab imaging as a non-invasive approach to assess clinical response to PD-1 blockade in cancer. *Ann Oncol* 2022;33:80–8. <https://doi.org/10.1016/j.annonc.2021.10.213>

## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

---

- [98 ] Niemeijer A-LN, Lager DEO, Huisman MC, Hoekstra OS, Boellaard R, Veen B van de, et al. First-in-human study of 89Zr-pembrolizumab PET/CT in patients with advanced stage non-small-cell lung cancer. *J Nucl Med* 2021;56:927–32. <https://doi.org/10.2967/jnumed.121.261926>.
- [99 ] Smit J, Borm FJ, Niemeijer A-LN, Huisman MC, Hoekstra OS, Boellaard R, et al. PD-L1 PET/CT imaging with radiolabeled durvalumab in patients with advanced stage non-small cell lung cancer. *J Nucl Med* 2021;56:927–32. <https://doi.org/10.2967/jnumed.121.262473>
- [100 ] van de Donk PP, Wind TT, Hooiveld-Noeken JS, van der Veen EL, Glaudemans AWJM, Diepstra A, et al. Interleukin-2 PET imaging in patients with metastatic melanoma before and during immune checkpoint inhibitor therapy. *Eur J Nucl Med Mol Imaging* 2021;48:4369–76. <https://doi.org/10.1007/s00259-021-05407-y>
- [101 ] Nienhuis PH, Antunes IF, Glaudemans AWJM, Jalving M, Leung D, Noordzij W, et al. 18F-BMS986192 PET imaging of PD-L1 in metastatic melanoma patients with brain metastases treated with immune checkpoint inhibitors. A pilot study. *J Nucl Med* 2021;63:899–905. <https://doi.org/10.2967/jnumed.121.262368>.
- [102 ] Bhatia A, Birger M, Veeraghavan H, Um H, Tixier F, McKenney AS, et al. MRI radiomic features are associated with survival in melanoma brain metastases treated with immune checkpoint inhibitors. *Neuro-Oncol* 2019;21:1578–86. <https://doi.org/10.1093/neuonc/noz141>
- [103 ] Durot C, Mulé S, Soyer P, Marchal A, Grange F, Hoeffel C. Metastatic melanoma: pretreatment contrast-enhanced CT texture parameters as predictive biomarkers of survival in patients treated with pembrolizumab. *Eur Radiol* 2019;29:3183–91. <https://doi.org/10.1007/s00330-018-5933-x>
- [104 ] Aoude LG, Wong BZY, Bonazzi VF, Brosda S, Walters SB, Koufariotis LT, et al. Radiomics Biomarkers Correlate with CD8 Expression and Predict Immune Signatures in Melanoma Patients. *Mol Cancer Res MCR* 2021;19:950–6. <https://doi.org/10.1158/1541-7786.MCR-20-1038>
- [105 ] Bonnin A, Durot C, Barat M, Djelouah M, Grange F, Mulé S, et al. CT texture analysis as a predictor of favorable response to anti-PD1 monoclonal antibodies in metastatic skin melanoma: CT texture analysis of metastatic skin melanoma. *Diagn Interv Imaging* 2021;103:97–102. <https://doi.org/10.1016/j.diii.2021.09.009>.
- [106 ] Ravanelli M, Agazzi GM, Milanese G, Roca E, Silva M, Tiseo M, et al. Prognostic and predictive value of histogram analysis in patients with non-small cell lung cancer refractory to platinum treated by nivolumab: A multicentre retrospective study. *Eur J Radiol* 2019;118:251–6. <https://doi.org/10.1016/j.ejrad.2019.07.019>
- [107 ] Zerunian M, Caruso D, Zucchelli A, Polici M, Capalbo C, Filetti M, et al. CT based radiomic approach on first line pembrolizumab in lung cancer. *Sci Rep* 2021;11:6633. <https://doi.org/10.1038/s41598-021-86113-5>.
- [108 ] Dercle L, Fronheiser M, Lu L, Du S, Hayes W, Leung DK, et al. Identification of Non-Small Cell Lung Cancer Sensitive to Systemic Cancer Therapies Using Radiomics. *Clin Cancer Res* 2020;26:2151–62. <https://doi.org/10.1158/1078-0432.CCR-19-2942>
- [109 ] Khene Z-E, Kokorian R, Peyronnet B, Edeline J, Crouzet L, Laguerre B, et al. Metastatic clear-cell renal cell carcinoma: computed tomography texture analysis as predictive biomarkers of survival in patients treated with nivolumab: Nivomics 01-study. *J Urol* 2020;203:e242–3.
- [110 ] Chen X, Zhou M, Wang Z, Lu S, Chang S, Zhou Z. Immunotherapy treatment outcome prediction in metastatic melanoma through an automated multi-objective delta-radiomics model. *Comput Biol Med* 2021;138. <https://doi.org/10.1016/j.cmbione.2021.104916>
- [111 ] Liu Y, Wu M, Zhang Y, Luo Y, He S, Wang Y, et al. Imaging Biomarkers to Predict and Evaluate the Effectiveness of Immunotherapy in Advanced Non-Small-Cell Lung Cancer. *Front Oncol* 2021;11:657615. <https://doi.org/10.3389/fonc.2021.657615>
- [112 ] Brendlin AS, Peisen F, Almansour H, Afat S, Eigentler T, Amaral T, et al. A Machine learning model trained on dual-energy CT radiomics significantly improves immunotherapy response prediction for patients with stage IV melanoma. *J Immunother Cancer* 2021;9. <https://doi.org/10.1136/jitc-2021-003261>
- [113 ] Wang Z-L, Mao L-L, Zhou Z-G, Si L, Zhu H-T, Chen X, et al. Pilot Study of CT-Based Radiomics Model for Early Evaluation of Response to Immunotherapy in Patients With Metastatic Melanoma. *Front Oncol* 2020;10:1524. <https://doi.org/10.3389/fonc.2020.01524>.
- [114 ] Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol Off J Eur Soc Med Oncol* 2019;30:998–1004. <https://doi.org/10.1093/annonc/mdz108>

## 2.7. References

---

- [115 ] Liu C, Gong J, Yu H, Liu Q, Wang S, Wang J. A CT-Based Radiomics Approach to Predict Nivolumab Response in Advanced Non-Small-Cell Lung Cancer. *Front Oncol* 2021;11:544339. <https://doi.org/10.3389/fonc.2021.544339>.
- [116 ] Mu W, Katsoulakis E, Whelan CJ, Gage KL, Schabath MB, Gillies RJ. Radiomics predicts risk of cachexia in advanced NSCLC patients treated with immune checkpoint inhibitors. *Br J Cancer* 2021;125:229–39. <https://doi.org/10.1038/s41416-021-01375-0>.
- [117 ] Mu W, Tunali I, Qi J, Schabath MB, Gillies RJ. Radiomics of 18F fluorodeoxyglucose PET/CT images predicts severe immune-related adverse events in patients with NSCLC. *Radiol Artif Intell* 2020;2. <https://doi.org/10.1148/ryai.2019190063>.
- [118 ] Nardone V, Tini P, Pastina P, Botta C, Reginelli A, Carbone SF, et al. Radiomics predicts survival of patients with advanced non-small cell lung cancer undergoing PD-1 blockade using Nivolumab. *Oncol Lett* 2020;19:1559–66. <https://doi.org/10.3892/ol.2019.11220>.
- [119 ] Shen L, Fu H, Tao G, Liu X, Yuan Z, Ye X. Pre-Immunotherapy Contrast-Enhanced CT Texture-Based Classification: A Useful Approach to Non-Small Cell Lung Cancer Immunotherapy Efficacy Prediction. *Front Oncol* 2021;11. <https://doi.org/10.3389/fonc.2021.591106>.
- [120 ] Yang B, Zhou L, Zhong J, Lv T, Li A, Ma L, et al. Combination of computed tomography imaging-based radiomics and clinicopathological characteristics for predicting the clinical benefits of immune checkpoint inhibitors in lung cancer. *Respir Res* 2021;22. <https://doi.org/10.1186/s12931-021-01780-2>.
- [121 ] Del Re M, Cucchiara F, Rofi E, Fontanelli L, Petrini I, Gri N, et al. A multiparametric approach to improve the prediction of response to immunotherapy in patients with metastatic NSCLC. *Cancer Immunol Immunother* 2020;70:1667–78. <https://doi.org/10.1007/s00262-020-02810-6>.
- [122 ] Ladwa R, Roberts KE, O’Leary C, Maggacis N, O’Byrne KJ, Miles K. Computed tomography texture analysis of response to second-line nivolumab in metastatic non-small cell lung cancer. *Lung Cancer Manag* 2020;9:LMT38. <https://doi.org/10.2217/lmt-2020-0002>.
- [123 ] Tunali I, Tan Y, Gray JE, Katsoulakis E, Eschrich SA, Saller J, et al. Hypoxia-related radiomics predict checkpoint blockade immunotherapy response of nonsmall cell lung cancer patients. *Cancer Res* 2020;80. <https://doi.org/10.1158/1538-7445.AM2020-5806>.
- [124 ] Valentinuzzi D, Vrankar M, Boc N, Ahac V, Zupancic Z, Unk M, et al. [18F]FDG PET immunotherapy radiomics signature (iRADIOMICS) predicts response of non-small-cell lung cancer patients treated with pembrolizumab. *Radiol Oncol* 2020;54:285–94. <https://doi.org/10.2478/raon-2020-0042>.
- [125 ] Granata V, Fusco R, Costa M, Picone C, Cozzi D, Moroni C, et al. Preliminary report on computed tomography radiomics features as biomarkers to immunotherapy selection in lung adenocarcinoma patients. *Cancers* 2021;13. <https://doi.org/10.3390/cancers13163992>.
- [126 ] Corino VDA, Bologna M, Calareso G, Licitra L, Ghi M, Rinaldi G, et al. A CT-Based Radiomic Signature Can Be Prognostic for 10-Months Overall Survival in Metastatic Tumors Treated with Nivolumab: An Exploratory Study. *Diagn Basel Switz* 2021;11. <https://doi.org/10.3390/diagnostics11060979>.
- [127 ] Hellwig K, Ellmann S, Eckstein M, Wiesmueller M, Rutzner S, Semrau S, et al. Predictive Value of Multiparametric MRI for Response to Single-Cycle Induction Chemo-Immunotherapy in Locally Advanced Head and Neck Squamous Cell Carcinoma. *Front Oncol* 2021;11:734872. <https://doi.org/10.3389/fonc.2021.734872>.
- [128 ] Park KJ, Lee J-L, Yoon S-K, Heo C, Park BW, Kim JK. Radiomics-based prediction model for outcomes of PD-1/PD-L1 immunotherapy in metastatic urothelial carcinoma. *Eur Radiol* 2020;30:5392–403. <https://doi.org/10.1007/s00330-020-06847-0>.
- [129 ] Alessandrino F, Gujrathi R, Nassar AH, Alzaghala A, Ravi A, McGregor B, et al. Predictive Role of Computed Tomography Texture Analysis in Patients with Metastatic Urothelial Cancer Treated with Programmed Death-1 and Programmed Death-ligand 1 Inhibitors. *Eur Urol Oncol* 2020;3:680–6. <https://doi.org/10.1016/j.euo.2019.02.002>.
- [130 ] Khene Z-E, Mathieu R, Peyronnet B, Kokorian R, Gasmii A, Khene F, et al. Radiomics can predict tumour response in patients treated with Nivolumab for a metastatic renal cell carcinoma: an artificial intelligence concept. *World J Urol* 2020;39:3707–9. <https://doi.org/10.1007/s00345-020-03334-5>.
- [131 ] Ji Z, Cui Y, Peng Z, Gong J, Zhu H-T, Zhang X, et al. Use of Radiomics to Predict Response to Immunotherapy of Malignant Tumors of the Digestive System. *Med Sci Monit Int Med J Exp Clin Res* 2020;26:e924671. <https://doi.org/10.12659/MSM.924671>.



## Chapter 2. Imaging to predict checkpoint inhibitor outcomes in cancer: A systematic review.

---

- [132] Himoto Y, Veeraraghavan H, Zheng J, Zamarin D, Snyder A, Capanu M, et al. Computed Tomography-Derived Radiomic Metrics Can Identify Responders to Immunotherapy in Ovarian Cancer. *JCO Precis Oncol* 2019;3. <https://doi.org/10.1200/P0.19.00038>
- [133] Zhu Y, Yao W, Xu B-C, Lei Y-Y, Guo Q-K, Liu L-Z, et al. Predicting response to immunotherapy plus chemotherapy in patients with esophageal squamous cell carcinoma using non-invasive Radiomic biomarkers. *BMC Cancer* 2021;21:1167. <https://doi.org/10.1186/s12885-021-08899-x>
- [134] Yuan G, Song Y, Li Q, Hu X, Zang M, Dai W, et al. Development and Validation of a Contrast-Enhanced CT-Based Radiomics Nomogram for Prediction of Therapeutic Efficacy of Anti-PD-1 Antibodies in Advanced HCC Patients. *Front Immunol* 2020;11:613946. <https://doi.org/10.3389/fimmu.2020.613946>
- [135] Korpics MC, Polley M-Y, Bhawe SR, Redler G, Pitroda SP, Luke JJ, et al. A Validated T Cell Radiomics Score Is Associated With Clinical Outcomes Following Multisite SBRT and Pembrolizumab. *Int J Radiat Oncol Biol Phys* 2020;108:189–95. <https://doi.org/10.1016/j.ijrobp.2020.06.026>
- [136] Colen RR, Fujii T, Bilen MA, Kotrotsou A, Abrol S, Hess KR, et al. Radiomics to predict immunotherapy-induced pneumonitis: proof of concept. *Invest New Drugs* 2018;36:601–7. <https://doi.org/10.1007/s10637-017-0524-2>
- [137] Ligerio M, Garcia-Ruiz A, Viaplana C, Villacampa G, Raciti MV, Landa J, et al. A CT-based Radiomics Signature Is Associated with Response to Immune Checkpoint Inhibitors in Advanced Solid Tumors. *Radiology* 2021;299:109–19. <https://doi.org/10.1148/radiol.2021200928>
- [138] Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018;19:1180–91. [https://doi.org/10.1016/S1470-2045\(18\)30413-3](https://doi.org/10.1016/S1470-2045(18)30413-3)
- [139] Sun R, Sundahl N, Hecht M, Putz F, Lancia A, Rouyar A, et al. Radiomics to predict outcomes and abscopal response of patients with cancer treated with immunotherapy combined with radiotherapy using a validated signature of CD8 cells. *J Immunother Cancer* 2020;8. <https://doi.org/10.1136/jitc-2020-001429>
- [140] Rundo F, Banna GL, Prezzavento L, Trenta F, Conoci S, Battiato S. 3D Non-Local Neural Network: A Non-Invasive Biomarker for Immunotherapy Treatment Outcome Prediction. Case-Study: Metastatic Urothelial Carcinoma. *J Imaging* 2020;6. <https://doi.org/10.3390/jimaging6120133>
- [141] Rundo F, Bersanelli M, Urzia V, Friedlaender A, Cantale O, Calcara G, et al. Three-Dimensional Deep Noninvasive Radiomics for the Prediction of Disease Control in Patients With Metastatic Urothelial Carcinoma treated With Immunotherapy. *Clin Genitourin Cancer* 2021;19:396–404. <https://doi.org/10.1016/j.clgc.2021.03.012>
- [142] Park C, Na KJ, Choi H, Ock C-Y, Ha S, Kim M, et al. Tumor immune profiles noninvasively estimated by FDG PET with deep learning correlate with immunotherapy response in lung adenocarcinoma. *Theranostics* 2020;10:10838–48. <https://doi.org/10.7150/thno.50283>
- [143] He B, Dong D, She Y, Zhou C, Fang M, Zhu Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. *J Immunother Cancer* 2020;8. <https://doi.org/10.1136/jitc-2020-000550>
- [144] Mu W, Jiang L, Zhang JY, Shi Y, Gray JE, Tunali I, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun* 2020;11. <https://doi.org/10.1038/s41467-020-19116-x>
- [145] Tian P, He B, Mu W, Liu K, Liu L, Zeng H, et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. *Theranostics* 2021;11:2098–107. <https://doi.org/10.7150/thno.48027>
- [146] Sasaki K, Morioka D, Conci S, Margonis GA, Sawada Y, Ruzzenente A, et al. The Tumor Burden Score: A New “Metro-ticket” Prognostic Tool For Colorectal Liver Metastases Based on Tumor Size and Number of Tumors. *Ann Surg* 2018;267:132–41. <https://doi.org/10.1097/SLA.0000000000002064>
- [147] Specht L, Nordentoft AM, Cold S, Clausen NT, Nissen NI, For the Danish National Hodgkin Study Group. Tumor burden as the most important prognostic factor in early stage Hodgkin's disease. Relations to other prognostic factors and implications for choice of treatment. *Cancer* 1988;61:1719–27. [https://doi.org/10.1002/1097-0142\(19880415\)61:8<1719::AID-CNCR2820610834>3.0.CO;2-A](https://doi.org/10.1002/1097-0142(19880415)61:8<1719::AID-CNCR2820610834>3.0.CO;2-A)
- [148] Gobbi PG, Ghirardelli ML, Solcia M, Di Giulio G, Merli F, Tavecchia L, et al. Image-Aided Estimate of Tumor Burden in Hodgkin's Disease: Evidence of Its Primary Prognostic Importance. *J Clin Oncol* 2001;19:1388–94. <https://doi.org/10.1200/JCO.2001.19.5.1388>

## 2.7. References

---

- [149 ] Vito A, El-Sayes N, Mossman K. Hypoxia-Driven Immune Escape in the Tumor Microenvironment. *Cells* 2020;9:992. <https://doi.org/10.3390/cells9040992>
- [150 ] Pietrobon V, Marincola FM. Hypoxia and the phenomenon of immune exclusion. *J Transl Med* 2021;19:9. <https://doi.org/10.1186/s12967-020-02667-4>
- [151 ] Wang B, Zhao Q, Zhang Y, Liu Z, Zheng Z, Liu S, et al. Targeting hypoxia in the tumor microenvironment: a potential strategy to improve cancer immunotherapy. *J Exp Clin Cancer Res* 2021;40:24. <https://doi.org/10.1186/s13046-020-01820-7>
- [152 ] Kichenadasse G, Miners JO, Mangoni AA, Rowland A, Hopkins AM, Sorich MJ. Association Between Body Mass Index and Overall Survival With Immune Checkpoint Inhibitor Therapy for Advanced Non-Small Cell Lung Cancer. *JAMA Oncol* 2020;6:512–8. <https://doi.org/10.1001/jamaoncol.2019.5241>
- [153 ] Sanchez A, Furberg H, Kuo F, Vuong L, Ged Y, Patil S, et al. Transcriptomic signatures related to the obesity paradox in patients with clear cell renal cell carcinoma: a cohort study. *Lancet Oncol* 2020;21:283–93. [https://doi.org/10.1016/S1470-2045\(19\)30797-1](https://doi.org/10.1016/S1470-2045(19)30797-1)
- [154 ] McQuade JL, Daniel CR, Hess KR, Mak C, Wang DY, Rai RR, et al. Association of body-mass index and outcomes in patients with metastatic melanoma treated with targeted therapy, immunotherapy, or chemotherapy: a retrospective, multicohort analysis. *Lancet Oncol* 2018;19:310–22. [https://doi.org/10.1016/S1470-2045\(18\)30078-0](https://doi.org/10.1016/S1470-2045(18)30078-0)
- [155 ] Turner DC, Kondic AG, Anderson KM, Robinson AG, Garon EB, Riess JW, et al. Pembrolizumab Exposure–Response Assessments Challenged by Association of Cancer Cachexia and Catabolic Clearance. *Clin Cancer Res* 2018;24:5841–9. <https://doi.org/10.1158/1078-0432.CCR-18-0415>
- [156 ] Joglekar S, Nau PN, Mezhir JJ. The impact of sarcopenia on survival and complications in surgical oncology: A review of the current literature. *J Surg Oncol* 2015;112:503–9. <https://doi.org/10.1002/jso.24025>
- [157 ] Su H, Ruan J, Chen T, Lin E, Shi L. CT-assessed sarcopenia is a predictive factor for both long-term and short-term outcomes in gastrointestinal oncology patients: a systematic review and meta-analysis. *Cancer Imaging* 2019;19:82. <https://doi.org/10.1186/s40644-019-0270-0>
- [158 ] Ní Bhuachalla ÉB, Daly LE, Power DG, Cushen SJ, MacEneaney P, Ryan AM. Computed tomography diagnosed cachexia and sarcopenia in 725 oncology patients: is nutritional screening capturing hidden malnutrition? *J Cachexia Sarcopenia Muscle* 2018;9:295–305. <https://doi.org/10.1002/jcsm.12258>
- [159 ] Yu J, Green MD, Li S, Sun Y, Journey SN, Choi JE, et al. Liver metastasis restrains immunotherapy efficacy via macrophage-mediated T cell elimination. *Nat Med* 2021;27:152–64. <https://doi.org/10.1038/s41591-020-1131-x>
- [160 ] Tawbi HA, Forsyth PA, Algazi A, Hamid O, Hodi FS, Moschos SJ, et al. Combined Nivolumab and Ipilimumab in Melanoma Metastatic to the Brain. *N Engl J Med* 2018;379:722–30. <https://doi.org/10.1056/NEJMoa1805453>
- [161 ] Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring CT scanner variability of radiomics features. *Invest Radiol* 2015;50:757–65. <https://doi.org/10.1097/RLI.0000000000000180>
- [162 ] Collins GS, Dhiman P, Navarro CLA, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021;11:e048008. <https://doi.org/10.1136/bmjopen-2020-048008>

## Chapter 3

# CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma

L.S. ter Maat MD, I.A.J. van Duin MD, S.G. Elias MD PhD, T. Leiner MD PhD, J.J.C. Verhoeff MD PhD, E.R.A.N. Arntz BSc, M.F. Troenokarso BSc, W.A.M. Blokk MD PhD, I. Isgum, PhD, G.A. de Wit PhD, F.W.P.J. van den Berkmortel MD PhD, M.J. Boers-Sonderen MD PhD, M.F. Boomsma MD PhD, A.J.M. van den Eertwegh MD PhD, J.W.B. de Groot MD PhD, D. Piersma MD PhD, G. Vreugdenhil MD PhD, H.M Westgeest MD PhD, E. Kapiteijn MD PhD, P.J. van Diest MD PhD, J.P.W. Pluim, PhD, P.A. de Jong MD PhD, K.P.M. Suijkerbuijk MD PhD, M. Veta PhD

*Published in European Journal of Cancer 185, 167-177*

## 3.1 Abstract

### 3.1.1 Introduction

Predicting checkpoint inhibitors treatment outcomes in melanoma is a relevant task, due to the unpredictable and potentially fatal toxicity and high costs for society. However, accurate biomarkers for treatment outcomes are lacking. Radiomics are a technique to quantitatively capture tumor characteristics on readily available computed tomography (CT) imaging. The purpose of this study was to investigate the added value of radiomics for predicting clinical benefit from checkpoint inhibitors in melanoma in a large, multicenter cohort.

### 3.1.2 Methods

Patients who received first-line anti-PD1 ± anti-CTLA4 treatment for advanced cutaneous melanoma were retrospectively identified from nine participating hospitals. For every patient, up to five representative lesions were segmented on baseline CT and radiomics features were extracted. A machine learning pipeline was trained on the radiomics features to predict clinical benefit, defined as stable disease for more than six months or response per RECIST 1.1 criteria. This approach was evaluated using a leave-one-center-out cross validation and compared to a model based on previously discovered clinical predictors. Lastly, a combination model was built on the radiomics and clinical model.

### 3.1.3 Results

A total of 620 patients were included, of which 59.2% experienced clinical benefit. The radiomics model achieved an area under the receiver operator characteristic curve (AUROC) of 0.607 [95%CI 0.562-0.652], lower than that of the clinical model (AUROC=0.646 [95%CI 0.600-0.692]). The combination model yielded no improvement over the clinical model in terms of discrimination (AUROC=0.636 [95%CI 0.592-0.680]) or calibration. The output of the radiomics model was significantly correlated with three out of five input variables of the clinical model ( $p < 0.001$ ).

## Chapter 3. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma

### 3.1.4 Discussion

---

The radiomics model achieved a moderate predictive value of clinical benefit, which was statistically significant. However, a radiomics approach was unable to add value to a simpler clinical model, most likely due to the overlap in predictive information learned by both models. Future research should focus on the application of deep learning, spectral CT derived radiomics and a multimodal approach for accurately predicting benefit to checkpoint inhibitor treatment in advanced melanoma.

## 3.2 Introduction

Survival of patients with advanced melanoma has improved dramatically after the introduction of immunotherapy. The survival of patients with unresectable stage III and stage IV melanoma has historically been very poor with a 1-year overall survival of 25% in phase II trials up to 2007 [1]. This changed with the introduction of anti-CTLA4 therapy in 2011 [2] and anti-PD1 therapy in 2014 [3,4]. In patients treated with anti-PD1 antibodies, real-world 1-year overall survival is now 67%, with 40% of patients achieving remissions of several years [5]. For patients treated with anti-PD1 plus anti-CTLA4 therapy, 5-year overall survival is reported to be as high as 52% [6].

However, not all patients benefit from checkpoint inhibitors. At 6 months after start of anti-PD1 treatment, 43% of patients experience progression or death. Furthermore, overall survival of patients with progression at 6 months was shown to be only 16% at 30 months. This is in contrast to a 30-month overall survival of 60%, 79% and 96% for patients with stable disease, partial response and complete response at 6 months of follow-up, respectively, in real-world data [5]. Similar results were reported in patients treated with anti-PD1 plus anti-CTLA4 therapy [7].

Accurate prediction of treatment benefit is an important topic for several reasons. First, treatment with checkpoint inhibitors is associated with severe and potentially fatal or irreversible toxicity. Severe toxicity occurs in 10-15% of patients treated with anti-PD1 monotherapy [5,8–10], and in as much as 60% of patients treated with anti-PD1 plus anti-CTLA4 combination therapy [11]. Second,

### 3.2. Introduction

---

checkpoint inhibition therapy is very costly. Depending on country and setting, estimates of additional costs per gained quality adjusted life year range from 25,000 to 81,000 United States Dollars [12,13]. Lastly, if patients who will not benefit are identified before start of treatment, alternative or experimental therapies can be started without delay.

Previously identified predictors for treatment outcomes are not yet sufficient to guide clinical decisions. Known clinical predictors of poor outcome include high tumor load, presence of liver metastases and symptomatic brain metastases, increased lactate dehydrogenase (LDH) and worse Eastern Cooperative Oncology Group (ECOG) performance status [14]. In addition, other biomarkers have been explored, such as PD-L1 expression, tumor mutational burden and histopathology features. Thus far, however, these predictors are not strong enough to predict treatment outcomes with high certainty [15], or the results remain to be validated in future studies [16].

Radiomics are by now an established modality for diagnosis, prognosis and prediction. Radiomics capture information about shape, intensity and texture of lesions in imaging and thereby form a reflection of tumor characteristics, such as necrosis or vascularization. These extracted features can subsequently be correlated to a clinical outcome [17]. This makes radiomics a cheap and non-invasive modality to, for example, discern benign from malignant lung nodules [18], estimate prognosis in non-small cell lung cancer (NSCLC) patients [19] and assess mutation status in glioblastoma [20]. Regarding prediction of checkpoint inhibitor treatment outcomes, promising findings have been published, particularly in NSCLC patients [21].

The added value of CT radiomics for predicting clinical benefit to checkpoint inhibitors in melanoma remains to be determined in large multicenter studies. Three previous smaller studies have investigated radiomics for this purpose, with conflicting findings. The studies by Trebeschi et al. [22] and Peisen et al. [23] report a significant discriminative value of radiomics for treatment outcomes (AUROC=0.78 on a dataset of 80 patients, and AUROC=0.64 on a dataset of 262 patients, respectively). In contrast, Brendlin et al. [24] reported a non-discriminative performance, despite using a very similar methodology (AUROC=0.50 in 140 patients). These differences in results highlight the importance of a large dataset to determine the value of radiomics. Furthermore, only the study by Peisen et al.

## **Chapter 3. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma**

investigated the added value over a simpler clinical model, with varying results across different outcomes. Lastly, none of the previous studies evaluated their model on data from other centers, although variability in scanner protocol may add significant noise [25]. In this study, we aimed to address these limitations and determine the added value of radiomics for predicting checkpoint inhibitor outcomes in a multicenter study in advanced melanoma.

### **3.3 Methods**

#### **3.3.1 Patient selection**

Eligible patients were retrospectively identified from high-quality registry data [26] from nine participating centers in The Netherlands (Amphia Ziekenhuis, Isala Zwolle, Leids Universitair Medisch Centrum, Máxima MC, Medisch Spectrum Twente, Radboudumc, UMC Utrecht, Amsterdam UMC, Zuyderland MC). Patients over the age of 18 were included if they received first-line treatment with anti-PD1  $\pm$  anti-CTLA4 checkpoint inhibition for irresectable stage IIIC or stage IV cutaneous melanoma after 01-01-2016. Exclusion criteria were (i) unavailability of baseline contrast enhanced CT imaging (CE-CT), (ii) lack of eligible target lesions, and (iii) less than six months of follow-up. Clinical characteristics were collected for the included patients and compared to those of the excluded patients. CT acquisition characteristics were extracted for included patients.

#### **3.3.2 Lesion selection and segmentation**

For every patient, one to five lesions were selected on baseline CT imaging and segmented. We aimed to make this selection of lesions as informative and representative as possible by using the following protocol: first, the five largest lesions were selected with a maximum of two per organ. If more lesions remained after segmenting a maximum of two per organ, the largest remaining lesions were segmented up to a total of five. Lesion selections were made without knowledge of the outcome. Lesions were excluded if they were not well-demarcated, affected by imaging artifacts or if the maximum diameter was less than 5mm. Segmentations were performed in 3D Slicer [27] on the series with the lowest slice thickness by

### 3.3. Methods

---

authors LSM and IAJD, under supervision of board-certified radiologists with 17 and 18 years of experience (PJ and TL, respectively).

#### 3.3.3 Feature extraction

Features were extracted from the segmented volumes using PyRadiomics [28]. For every volume, 1874 features were extracted at five different levels of detail, resulting in a total of 9370 features. An overview of the extracted features is given in the Supplementary Methods. Interobserver agreement of segmentations and features was calculated using Dice scores and intraclass correlation coefficient (ICC), respectively, based on 16 scans segmented by both observers (LSM, IAJD).

#### 3.3.4 Outcome definition

The primary outcome was clinical benefit, defined as a best overall response of partial or complete response, or stable disease for a minimum of six months after start of treatment; response was determined by the treating physician in line with RECIST 1.1 criteria [29]. The secondary outcome was objective response, defined as a best overall response of partial or complete response. clinical benefit was used as the primary outcome, as the intended use of the model was to identify patients who would quickly progress despite treatment and therefore not derive any benefit from treatment. Individual lesion response was assessed using maximum diameter recordings at baseline and at three, six and nine months, or until treatment was changed. Given the possibility of pseudo-progression, the last available follow-up was used to determine lesion outcomes. If the maximum diameter at the last follow-up was less than 120% of the baseline diameter, the lesion was labeled as ‘does benefit’, and ‘does not benefit’ otherwise. In parallel, the lesion was labeled as ‘responsive’ if the maximum diameter was less than 70% of the baseline diameter at the last follow-up, and ‘non-responsive’ otherwise. These lesion-level cut-offs were chosen in correspondence with the patient-level cut-offs used in RECIST 1.1 [29].

#### 3.3.5 Evaluated models

Three predictive models were compared: a model based on radiomics, a model based on baseline clinical characteristics and an ensemble model that combined



### **Chapter 3. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma**

the predictions of these models. The radiomics model consisted of a machine learning pipeline that automatically selected optimal components and hyperparameters for feature selection, dimensionality reduction and classification (Figure 1). This pipeline was trained to predict outcomes per lesion; these outputs per lesion were then aggregated to a patient level prediction. The clinical model used the same machine learning pipeline, which was fitted on five clinical variables that were consistently shown to be predictive of checkpoint inhibitor treatment outcomes in previous literature [5,14,30,31]. These predictors were (i) ECOG performances status, (ii) LDH level, presence of (iii) brain and (iv) liver metastases, and (v) number of affected organs. All variables were one-hot encoded; missing values were encoded as a separate label. The ensemble model consisted of a logistic regression fitted on the output of the radiomics and clinical model. All three models were evaluated using a nested cross validation. The inner loop was used for optimal model selection and hyperparameter tuning; the outer loop was used to evaluate predictive performance on unseen data and was conducted in a leave-one-center-out manner. Further details are supplied in the Supplementary Methods.

#### **3.3.6 Statistical analysis**

The discriminative performance of the models was evaluated using the area under the receiver operator characteristic curve (AUROC) and corresponding 95% confidence interval. The cross validated AUROC and confidence interval were calculated using the cvAUC R package [32]. Methods for comparing cross validated AUROCs between models are detailed in the Supplementary Methods. Subgroup analyses were conducted for patients treated with anti-PD1 therapy and anti-PD1 plus anti-CTLA4 therapy by evaluating the fitted model only on patients from the respective groups. The output of the radiomics model for predicting clinical benefit was correlated to the input variables of the clinical model to determine if the radiomics model learned features that were already represented in the baseline clinical model.

#### **3.3.7 Adherence to quality standards**

The TRIPOD checklist [33] was completed and is available in Supplementary Table 1. The study design was reviewed by the Medical Ethics Committee and not

### 3.4. Results

---

considered subject to the Medical Research Involving Human Subjects Act in compliance with Dutch regulations; informed consent was waived.

## 3.4 Results

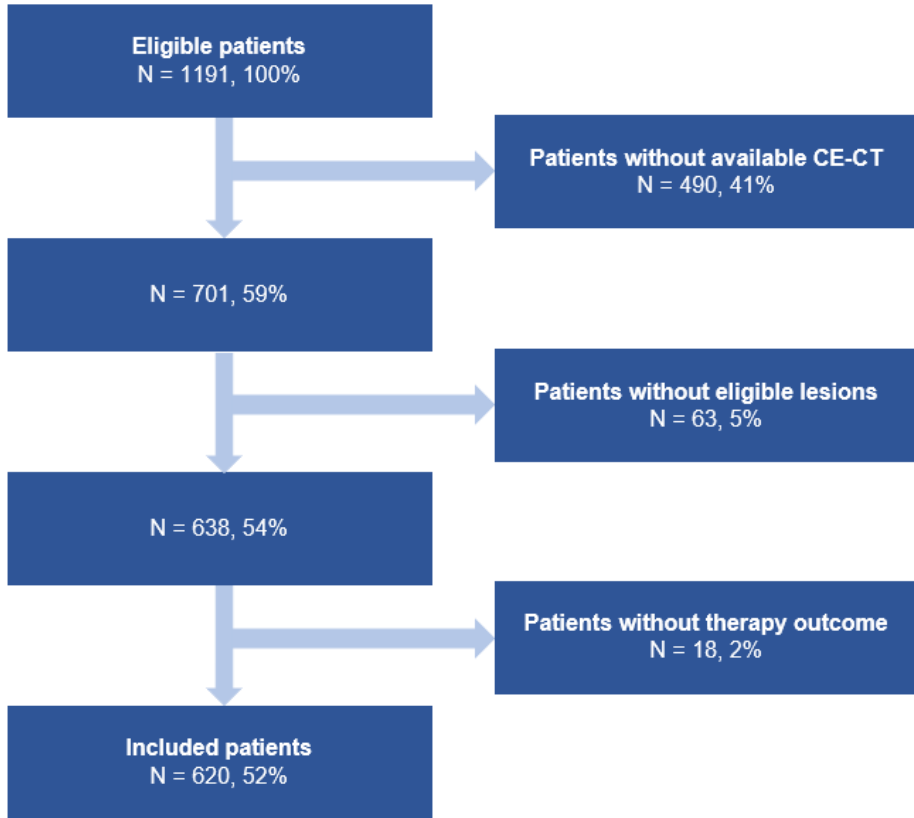
### 3.4.1 Patient characteristics

Out of 1191 eligible patients, 620 patients with a total of 2352 lesions were included. A flowchart of the selection process is shown in Figure 3.1. The rate of clinical benefit was 59.2% (367 patients); the objective response rate was 51.3% (318 patients); Lesion level outcomes were available for 75.2% of lesions. Lesion level outcomes could not be recorded for patients from the Radboudumc (327 lesions, 13.9%) due to local regulations. In addition, follow-up imaging was unavailable due to patient death or clinical progression before the first follow-up moment for 185 lesions (7.9%), due to the lesion falling outside the field of view in 27 lesions (1.1%) and due to technical issues in 44 lesions (1.9%). Rate of benefit was 79.4% among lesions with available labels, whereas response rate was 54.8% (Supplementary Table 5). Of all eligible patients, 490 patients were excluded because of the unavailability of contrast-enhanced pre-treatment CT. In most of these cases, an 18-fluorodeoxyglucose positron emission tomography (FDG-PET) with low-dose CT was made. Characteristics for the included patients are shown in Table 3.1 and compared to those of excluded patients in Supplementary Table 2. The subgroups of patient treated with anti-PD1 and combination therapy consisted of 370 and 250 patients, respectively. Supplementary Tables 3 and 4 show patient characteristics per center, and for the subgroups treated with monotherapy and combination therapy, respectively. CT acquisition characteristics per center are displayed in Supplementary Table 6.

### 3.4.2 Interobserver variability

52 lesions in 16 scans were segmented by two observers. Segmentations corresponded with a median Dice score of 0.88 (IQI 0.82-0.92). For the extracted features, the median intraclass correlation coefficient was 0.97 (IQI 0.92-0.99).

### Chapter 3. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma



**Figure 3.1:** Flowchart of patient selection

#### 3.4.3 Treatment outcome prediction

For predicting clinical benefit, the radiomics model achieved an AUROC of 0.607 [95% CI 0.562-0.652], the clinical model an AUROC of 0.646 [95% CI 0.600-0.692], and the ensemble model an AUROC of 0.636 [95% CI 0.592-0.680]. The difference in AUROC between the ensemble and clinical model was not statistically significant (Supplementary Figure 1). Calibration curves showed adequate calibration of the three models with no evidence of poor fit (Hosmer-Lemeshow  $p > 0.07$ ). The range of predicted probabilities was comparable between models (IQI 0.56-0.65, 0.53-0.67 and 0.52-0.69 for the radiomics, clinical and ensemble model, respectively). Results were similar for predicting objective response (Supplementary

### 3.5. Discussion

---

Figure 2-3). Predictive performance for both outcomes was comparable in subgroups of patients treated with monotherapy and combination therapy, with a trend of better discrimination in the subgroup of patients treated with combination therapy (Supplementary Figures 4-7). Details of the selected models and hyperparameters per fold are shown in Supplementary Table 7.

#### 3.4.4 Comparison of radiomics and clinical model

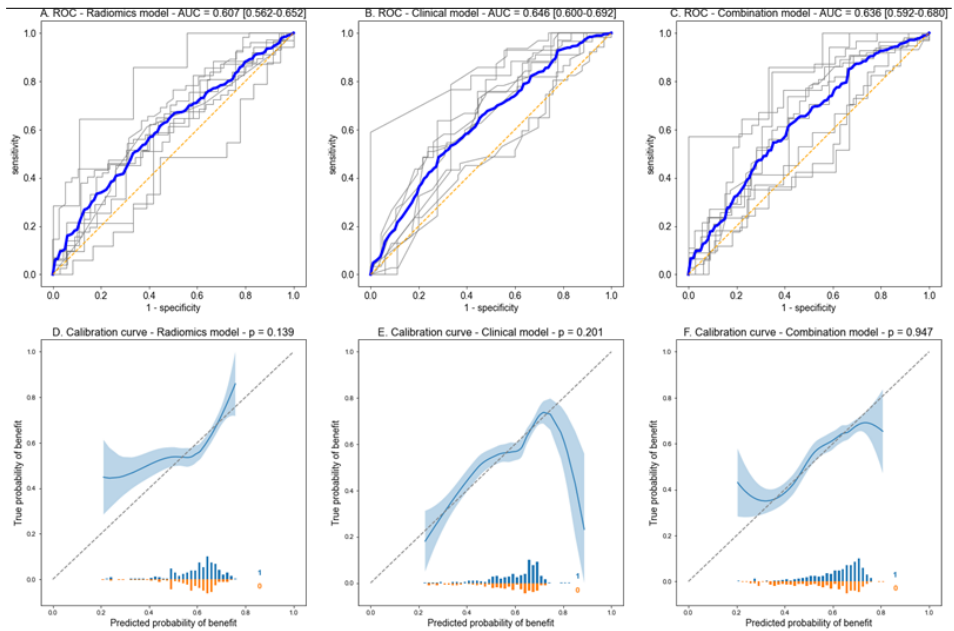
The predicted probability of clinical benefit by the radiomics model was significantly lower in patients in whom liver metastases were absent (Mann-Whitney U,  $p < 0.001$ , Figure 3.3A), in patients with higher LDH (Kruskal-Wallis  $p < 0.001$ , Figure 4D) and who had more affected organs (Mann-Whitney U  $p < 0.001$ , Figure 4E). The output of radiomics model was not significantly different in patients with and without brain metastases), and for different categories of ECOG performance status (Figure 4B-C). The output of radiomics and clinical models were significantly and positively correlated (Spearman's correlation coefficient = 0.369,  $p < 0.001$ , Figure 4F).

### 3.5 Discussion

The present work shows that radiomics are moderately predictive of checkpoint inhibitor treatment outcomes in patients with advanced melanoma. The results were consistent for both clinical benefit and objective response rate, and are most in line with the findings of the earlier study by Peisen et al. A recent work by Dercle et al. allows for comparison to a model that also incorporates radiomics from on-treatment CT scans [34]. This model reached an AUROC of 0.92 for predicting overall survival at six months, indicating that on-treatment radiomics are strongly predictive. However, most toxicity occurs in the first three months and long-term outcomes can already be accurately predicted using on-treatment information without the use of radiomics [6,35]. Predicting response using the 3-month on-treatment scan therefore appears to be of limited clinical relevance.

Addition of radiomics to known clinical predictors, however, did not yield improvement in predictive value. The combination model was not superior to the clinical model in either discrimination or calibration. This lack of improvement

### Chapter 3. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma



**Figure 3.2:** (A-C) Receiver operator characteristic (ROC) curves for predicting durable clinical benefit in patients with melanoma treated with anti-PD1 ± anti-CTLA4 checkpoint inhibition for the radiomics model (A), clinical model (B) and combination model (C). Gray curves correspond to results per fold; blue curves are the weighted average of the results per fold. The area under the curve (AUC) with corresponding 95% confidence intervals are displayed. (D-F) LOESS fitted calibration curves for predicting durable clinical benefit in the radiomics model (D), clinical model (E) and combination model (F); the shaded area corresponds to ±1 standard deviation. Histograms of the predictions for positive (blue) and negative (orange) samples are provided below the curves, the x-axis displays the predicted values for these histograms. P-values of the Hosmer-Lemeshow goodness-of-fit test are shown in the plot titles.

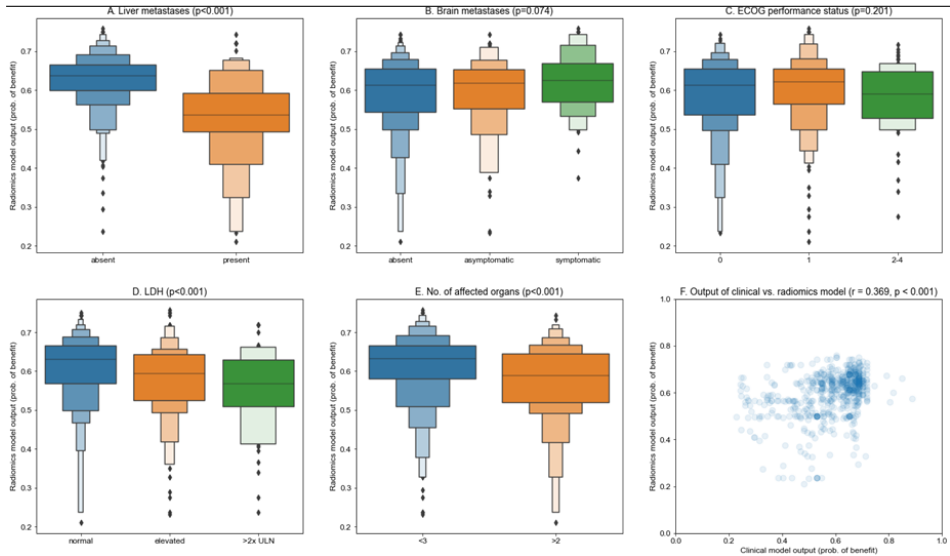
### 3.5. Discussion

---

		Missing	Overall
n			620
Age, median [Q1, Q3]		0	67.5 [58.0,75.0]
Sex, n (%)	Female	0	239 (38.5)
	Male		381 (61.5)
Stage, n (%)	IIIC	4	25 (4.1)
	M1a		49 (8.0)
	M1b		94 (15.3)
	M1c		296 (48.1)
	M1d		152 (24.7)
ECOG performance status, n (%)	0	26	287 (48.3)
	1		247 (41.6)
	2-4		60 (10.1)
Primary tumor location, n (%)	Acral	10	15 (2.5)
	Extremity		167 (27.4)
	Head, neck		66 (10.8)
	Trunk		247 (40.5)
	Unknown		115 (18.9)
Brain metastases, n (%)	Absent	45	423 (73.6)
	Asymptomatic		76 (13.2)
	Symptomatic		76 (13.2)
Liver metastases, n (%)	Absent	30	398 (67.5)
	Present		192 (32.5)
No. of affected organs, n (%)	<3	0	338 (54.5)
	>2		282 (45.5)
LDH, n (%)	Normal	9	381 (62.4)
	1-2x upper limit of normal		177 (29.0)
	>2x upper limit of normal		53 (8.7)
Clinical benefit, n (%)	No benefit	0	253 (40.8)
	Benefit		367 (59.2)
Objective response, n (%)	No response	0	302 (48.7)
	Response		318 (51.3)
Therapy, n (%)	Anti-PD1	0	370 (59.7)
	Ipilimumab & Nivolumab		250 (40.3)

**Table 3.1:** Characteristics of included patients

### Chapter 3. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma



**Figure 3.3:** Graphical overview of correspondence between the output of the radiomics and clinical models. (A-E) Boxenplots of the output of the radiomics model, compared across different values for clinical predictors. (A) The output of the radiomics model is significantly lower in patients with liver metastases than in patients without (Mann-Whitney U  $p < 0.001$ ). (B) No statistical difference was found in the output of the radiomics model between patients without or with asymptomatic or symptomatic brain metastases (Kruskal-Wallis  $p = 0.074$ ) and ECOG performance status (Kruskal-Wallis  $p = 0.201$ ). (D) The output of the radiomics model is significantly lower in patients with higher levels of LDH (Kruskal-Wallis  $p < 0.001$ ) and with more affected organs (Mann-Whitney U  $p < 0.001$ ). (F) The outputs of the clinical and radiomics models (predicted probability of response) are positively correlated (Spearman's rank correlation coefficient = 0.369,  $p < 0.001$ ).

### 3.5. Discussion

---

can be explained due to an overlap in the information learned by the radiomics model, and the information that is already represented in clinical variables. As demonstrated, the radiomics model indirectly learns to detect the presence of liver metastases and the amount of tumor burden, as reflected in LDH and number of affected organs. The fact that this information is indeed learned as expected is a strong argument for the validity of the present work. Furthermore, this indicates that such overlap is likely to be present in any radiomics model which is investigated for clinical purposes.

Studies on radiomics should assess the added value over simpler predictors. Many smaller exploratory studies have been conducted into the predictive value of radiomics for checkpoint inhibitor outcomes across different malignancies [21]. Their findings are almost exclusively positive, but the added value over clinical predictors was seldomly assessed. The present work demonstrates that clinical predictors can be captured by radiomics, and that the added value of radiomics should therefore always be investigated, even in exploratory studies.

Future works should aim to improve on radiomics through deep learning or spectral CT derived radiomics. Deep learning has a significant advantage over handcrafted radiomics, as this method is not limited by predefined features in what information can be captured. Instead, a deep learning approach is given the raw data as input and learns informative features on the fly [36]. Furthermore, spectral CT derived radiomics were shown to be superior over single energy radiomics for predicting response to checkpoint inhibition in patients with melanoma by Brendlin et al. [24]. As spectral CT scanners become increasingly available, this approach may be tested more thoroughly in future research.

Lastly, the multimodal approach should be extended with other data sources. Accurately predicting checkpoint inhibitor treatment outcomes in melanoma remains challenging. It is possible that individual biomarkers are insufficient to guide clinical decisions. An approach that combines different data sources may therefore prove to be superior. A possible modalities that may be explored for this purpose is histopathology imaging [16], which will be investigated in this cohort in a future work.

The strengths of this work are the large sample size, the multicenter design and extensive hyperparameter optimization. This is the largest work published on radiomics for prediction of checkpoint inhibitor treatment outcomes in any



### **Chapter 3. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma**

malignancy [21]. This large size adds to the weight of the presented conclusion. Furthermore, the dataset in this work includes patients from nine different centers. As stability of radiomics features across scanner types and protocols is far from certain, external validation is essential for determining the practical value of a radiomics approach. Lastly, the proposed pipeline systematically explores design choices, from the extraction of radiomics to the final prediction. This approach should maximize potential performance by avoiding arbitrary and therefore possibly suboptimal design choices.

A potential limitation is the exclusion of a large fraction of patients due to unavailability of CE-CT imaging. Comparison of patient characteristics between the included and excluded groups showed minor differences overall, with a trend towards more progressed disease in the included patients. Our hypothesis for this is that patients with more progressed disease are more likely to directly present to medical oncology, instead of being referred after having undergone imaging by a different specialty where FDG-PET is the preferred modality. Although this selection may theoretically have influenced the presented results, this risk is arguably limited as the characteristics of in- and excluded patients are overall very comparable.

Furthermore, patients with stable disease for a minimum of six months were labeled as having clinical benefit. This group could therefore theoretically include patients with indolent tumor progression (less than 120% of original diameters in six months), without effect from checkpoint inhibition therapy. However, given the consistent results across outcomes and small proportion of patients for which this may be the case, the impact on eventual results is likely limited.

In conclusion, radiomics are predictive of checkpoint inhibition treatment outcomes in patients with advanced melanoma, but did not improve predictive value over a simpler clinical model. A radiomics model can predict both clinical benefit and response from checkpoint inhibitor therapy with moderate discriminative performance. However, the predictive value of this radiomics model overlaps with that of a clinical model, which is evident from the lack of improvement of a combined model. The added value of a radiomics approach therefore appears to be limited. Future research should focus on related techniques, such as deep learning or radiomics on dual energy CT images. In addition, an approach that combines radiomics and clinical data with other modalities may provide a next step towards

### 3.7. Supplementary Materials

---

accurate prediction of checkpoint inhibitor treatment outcomes in melanoma.

## 3.6 Supplementary Materials

Supplementary materials are available through:

<https://ars.els-cdn.com/content/image/1-s2.0-S0959804923001090-mmc1.docx>

## 3.7 References

- [1] Korn EL, Liu P-Y, Lee SJ, Chapman J-AW, Niedzwiecki D, Suman VJ, et al. Meta-Analysis of Phase II Cooperative Group Trials in Metastatic Stage IV Melanoma to Determine Progression-Free and Overall Survival Benchmarks for Future Phase II Trials. *J Clin Oncol* 2008;26:527–34. <https://doi.org/10.1200/JCO.2007.12.7837>
- [2] Sondak VK, Smalley KSM, Kudchadkar R, Gripton S, Kirkpatrick P. Ipilimumab. *Nat Rev Drug Discov* 2011;10:411–3.
- [3] Barone A, Hazarika M, Theoret MR, Mishra-Kalyani P, Chen H, He K, et al. FDA Approval Summary: Pembrolizumab for the Treatment of Patients with Unresectable or Metastatic Melanoma. *Clin Cancer Res* 2017;23:5661–5. <https://doi.org/10.1158/1078-0432.CCR-16-0664>
- [4] Beaver JA, Theoret MR, Mushti S, He K, Libeg M, Goldberg K, et al. FDA Approval of Nivolumab for the First-Line Treatment of Patients with BRAFV600 Wild-Type Unresectable or Metastatic Melanoma. *Clin Cancer Res* 2017;23:3479–83. <https://doi.org/10.1158/1078-0432.CCR-16-0714>
- [5] van Zeijl MCT, Haanen JBAG, Wouters MWJM, de Wreede LC, Jochems A, Aarts MJB, et al. Real-world Outcomes of First-line Anti-PD-1 Therapy for Advanced Melanoma: A Nationwide Population-based Study. *J Immunother* 2020;43:256–64. <https://doi.org/10.1097/CJI.0000000000000334>
- [6] Larkin J, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med* 2019;381:1535–46. <https://doi.org/10.1056/NEJMoa1910836>
- [7] Asher N, Ben-Betzalel G, Lev-Ari S, Shapira-Frommer R, Steinberg-Silman Y, Gochman N, et al. Real World Outcomes of Ipilimumab and Nivolumab in Patients with Metastatic Melanoma. *Cancers* 2020;12:2329. <https://doi.org/10.3390/cancers12082329>
- [8] Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, et al. Nivolumab in Previously Untreated Melanoma without BRAF Mutation. *N Engl J Med* 2015;372:320–30. <https://doi.org/10.1056/NEJMoa1412082>
- [9] Robert C, Schachter J, Long GV, Arance A, Grob JJ, Mortier L, et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N Engl J Med* 2015;372:2521–32. <https://doi.org/10.1056/NEJMoa1503093>
- [10] Verheijden RJ, May AM, Blank CU, van der Veldt AAM, Boers-Sonderen MJ, Aarts MJB, et al. Lower risk of severe checkpoint inhibitor toxicity in more advanced disease. *ESMO Open* 2020;5:e000945. <https://doi.org/10.1136/esmoopen-2020-000945>
- [11] Wolchok JD, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Long-Term Outcomes With Nivolumab Plus Ipilimumab or Nivolumab Alone Versus Ipilimumab in Patients With Advanced Melanoma. *J Clin Oncol* 2022;40:127–37. <https://doi.org/10.1200/jco.21.02229>
- [12] Leeneman B, Uyl-de Groot CA, Aarts MJB, van Akkooi ACJ, van den Berkmortel FWPJ, van den Eertwegh AJM, et al. Healthcare Costs of Metastatic Cutaneous Melanoma in the Era of Immunotherapeutic and Targeted Drugs. *Cancers* 2020;12:E1003. <https://doi.org/10.3390/cancers12041003>
- [13] Verma V, Sprave T, Haque W, Simone CB, Chang JY, Welsh JW, et al. A systematic review of the cost and cost-effectiveness studies of immune checkpoint inhibitors. *J Immunother Cancer* 2018;6:128. <https://doi.org/10.1186/s40425-018-0442-7>

### Chapter 3. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma

- [14] Silva IP da, Ahmed T, McQuade JL, Nebhan CA, Park JJ, Versluis JM, et al. Clinical Models to Define Response and Survival With Anti-PD-1 Antibodies Alone or Combined With Ipilimumab in Metastatic Melanoma. *J Clin Oncol* 2022. <https://doi.org/10.1200/JCO.21.01701>
- [15] Morrison C, Pabla S, Conroy JM, Nesline MK, Glenn ST, Dressman D, et al. Predicting response to checkpoint inhibitors in melanoma beyond PD-L1 and mutational burden. *J Immunother Cancer* 2018;6:32. <https://doi.org/10.1186/s40425-018-0344-8>
- [16] Johannet P, Coudray N, Donnelly DM, Jour G, Illa-Bochaca I, Xia Y, et al. Using Machine Learning Algorithms to Predict Immunotherapy Response in Patients with Advanced Melanoma. *Clin Cancer Res Off J Am Assoc Cancer Res* 2021;27:131–40. <https://doi.org/10.1158/1078-0432.CCR-20-2415>
- [17] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278:563–77. <https://doi.org/10.1148/radiol.2015151169>
- [18] Way TW, Hadjiiski LM, Sahiner B, Chan H-P, Cascade PN, Kazerooni EA, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours. *Med Phys* 2006;33:2323–37. <https://doi.org/10.1118/1.2207129>
- [19] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006. <https://doi.org/10.1038/ncomms5006>
- [20] Gutman DA, Dunn WD, Grossmann P, Cooper LAD, Holder CA, Ligon KL, et al. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. *Neuroradiology* 2015;57:1227–37. <https://doi.org/10.1007/s00234-015-1576-7>
- [21] ter Maat LS, van Duin IAJ, Elias SG, van Diest PJ, Pluim JPW, Verhoeff JJC, et al. Imaging to predict checkpoint inhibitor outcomes in cancer. A systematic review. *Eur J Cancer* 2022;175:60–76. <https://doi.org/10.1016/j.ejca.2022.07.034>
- [22] Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol Off J Eur Soc Med Oncol* 2019;30:998–1004. <https://doi.org/10.1093/annonc/mdz108>
- [23] Peisen F, Hensch A, Hering A, Brendlin AS, Afat S, Nikolaou K, et al. Combination of Whole-Body Baseline CT Radiomics and Clinical Parameters to Predict Response and Survival in a Stage-IV Melanoma Cohort Undergoing Immunotherapy. *Cancers* 2022;14:2992. <https://doi.org/10.3390/cancers14122992>
- [24] Brendlin AS, Peisen F, Almansour H, Afat S, Eigentler T, Amaral T, et al. A Machine learning model trained on dual-energy CT radiomics significantly improves immunotherapy response prediction for patients with stage IV melanoma. *J Immunother Cancer* 2021;9. <https://doi.org/10.1136/jitc-2021-003261>
- [25] Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring CT scanner variability of radiomics features. *Invest Radiol* 2015;50:757–65. <https://doi.org/10.1097/RLI.0000000000000180>
- [26] Jochems A, Schouwenburg MG, Leeneman B, Franken MG, van den Eertwegh AJM, Haanen JBAG, et al. Dutch Melanoma Treatment Registry: Quality assurance in the care of patients with metastatic melanoma in the Netherlands. *Eur J Cancer* 2017;72:156–65. <https://doi.org/10.1016/j.ejca.2016.11.021>
- [27] Pieper S, Halle M, Kikinis R. 3D Slicer. 2004 2nd IEEE Int. Symp. Biomed. Imaging Nano Macro IEEE Cat No 04EX821, 2004, p. 632-635 Vol. 1. <https://doi.org/10.1109/ISBI.2004.1398617>
- [28] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017;77:e104–7. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
- [29] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47. <https://doi.org/10.1016/j.ejca.2008.10.026>
- [30] van Not OJ, de Meza MM, van den Eertwegh AJM, Haanen JB, Blank CU, Aarts MJB, et al. Response to immune checkpoint inhibitors in acral melanoma: A nationwide cohort study. *Eur J Cancer* 2022;167:70–80. <https://doi.org/10.1016/j.ejca.2022.02.026>
- [31] van Zeijl MCT, de Wreede LC, van den Eertwegh AJM, Wouters MWJM, Jochems A, Schouwenburg MG, et al. Survival outcomes of patients with advanced melanoma from 2013 to 2017: Results of a nationwide population-based registry. *Eur J Cancer* 2021;144:242–51. <https://doi.org/10.1016/j.ejca.2020.11.028>

### 3.7. References

---

- [32 ] LeDell E, Petersen M, van der Laan M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron J Stat* 2015;9:1583–607. <https://doi.org/10.1214/15-EJS1035>
- [33 ] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *Br J Surg* 2015;102:148–58. <https://doi.org/10.1002/bjs.9736>
- [34 ] Dercele L, Zhao B, Gönen M, Moskowitz CS, Firas A, Beylergil V, et al. Early Readout on Overall Survival of Patients With Melanoma Treated With Immunotherapy Using a Novel Imaging Analysis. *JAMA Oncol* 2022;8:385–92. <https://doi.org/10.1001/jamaoncol.2021.6818>
- [35 ] Robert C, Long GV, Brady B, Dutriaux C, Giacomo AMD, Mortier L, et al. Five-Year Outcomes With Nivolumab in Patients With Wild-Type BRAF Advanced Melanoma. *J Clin Oncol* 2020. <https://doi.org/10.1200/JCO.20.00995>
- [36 ] Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities. *IEEE Signal Process Mag* 2019;36:132–60. <https://doi.org/10.1109/MSP.2019.2900993>

## Chapter 4

# Deep learning on CT scans to predict checkpoint inhibitor treatment outcomes in advanced melanoma

Laurens S. Ter Maat MD, Rob A.J. De Mooij MSc, Isabella A.J. Van Duin MD, Joost J.C. Verhoeff MD PhD, Sjoerd G. Elias MD PhD, Tim Leiner MD PhD, Wouter A.C. van Amsterdam MD PhD, Max F. Troenokarso BSc, Eran R.A.N. Arntz BSc, Franchette W.P.J. Van den Berkmortel MD PhD, Marye J. Boers-Sonderen MD PhD, Martijn F. Boomsma MD PhD, Fons J.M. Van den Eertwegh MD PhD, Jan Willem de Groot MD PhD, Geke A.P. Hospers MD PhD, Djura Piersma MD PhD, Art Vreugdenhil MD PhD, Hans M. Westgeest MD PhD, Ellen Kapiteijn MD PhD, Ardine A. De Wit PhD, Willeke A.M. Blokx MD PhD, Paul J. Van Diest MD PhD, Pim A. De Jong MD PhD, Josien P.W. Pluim PhD, Karijn P.M. Suijkerbuijk MD PhD, Mitko Veta PhD

*Submitted*

# 4.1 Abstract

## 4.1.1 Introduction

Immune checkpoint inhibitor (ICI) treatment has proven successful for advanced melanoma, but is associated with potentially severe toxicity and high costs. Accurate biomarkers for response are lacking. The present work is the first to investigate the value of deep learning on CT imaging of metastatic lesions for predicting ICI treatment outcomes in advanced melanoma.

## 4.1.2 Methods

Adult patients that were treated with ICI for advanced melanoma were retrospectively identified from ten participating centers. A deep learning model (DLM) was trained on volumes of lesions on baseline CT to predict clinical benefit. The DLM was compared to and combined with a model of known clinical predictors (presence of liver and brain metastasis, level of lactate dehydrogenase, performance status and number of affected organs).

## 4.1.3 Results

A total of 730 eligible patients with 2722 lesions were included. The DLM reached an area under the receiver operating characteristic (AUROC) of 0.607 [9%CI 0.565–0.648]. In comparison, a model of clinical predictors reached an AUROC of 0.635 [95%CI 0.59–0.678]. The combination model reached an AUROC of 0.635 [95 CI 0.595–0.676]. Differences in AUROC were not statistically significant. The output of the DLM was significantly correlated with four of the five input variables of the clinical model.

## 4.1.4 Discussion

The DLM reached a statistically significant discriminative value, but was unable to improve over known clinical predictors. The present work shows that the assessment over known clinical predictors is an essential step for imaging-based prediction and brings important nuance to the almost exclusively positive findings in this field.

## 4.2 Introduction

Checkpoint inhibitors have revolutionized the treatment of advanced melanoma. The real-world 1-year overall survival of patients treated with anti-PD1 therapy is 67% [1], which is in stark contrast to the 1-year overall survival of 25% in phase II trials up to 2007 [2].

However, still a significant fraction of patients does not respond to this treatment, that is also associated with potentially severe toxicity and high costs. Approximately 40-50% of patients experience disease progression despite treatment, and subsequently derive little benefit in terms of survival [1,3]. Furthermore, checkpoint inhibition treatment is expensive, with estimates of additional costs of up to 81,000 US dollars per quality adjusted life year [4,5]. Lastly, severe and partly irreversible toxicity occurs in as much as 60% of patients treated with anti-PD1 + anti-CTLA4 combination therapy [6].

Therefore, accurate prediction at baseline of treatment outcomes is necessary. If non-responders can be identified with high certainty before start of treatment, alternative therapies can be started without delay in these patients. Furthermore, needless costs and toxic effects can be prevented.

However, current biomarkers are not accurate enough to guide treatment decisions. Previous research has identified several significant predictors of treatment outcomes, such as levels of lactate dehydrogenase, presence of liver and brain metastases, performance status and level of tumoral PD-L1 expression [7,8]. These biomarkers, however, have not reached the degree of accuracy that is necessary to adequately guide treatment decisions. Patients without PD-L1 expression, for instance, may still respond to therapy, even though this protein is the very target of anti-PD1 therapy [8]. This underlines the need for further research into accurate predictive biomarkers.

CT imaging of tumor lesions may be used as a biomarker in two ways: through handcrafted radiomics and through deep learning. In a handcrafted radiomics approach, predefined features that reflect shape and texture are calculated on a volume of interest. These features are subsequently used to train a model that can classify the lesion as, for instance, having a certain mutation or responding to a treatment [9]. In contrast, a deep learning approach skips the step of extracting manually predefined features and trains a model directly on the raw image as an

### 4.3. Introduction

---

input [10]. This approach has the advantage that it is not limited by the chosen features in what it can learn; instead, relevant features are learned during training in such a way that the predictive performance of the model is optimized. A potential downside is that, usually, a larger dataset is needed for adequate performance compared with a handcrafted radiomics approach. For both methods, the underlying hypothesis is that features visible on imaging reflect the tumor’s phenotype and may therefore also correlate to clinically relevant characteristics and biological behavior of the tumor.

Thus far, deep learning on CT imaging of lesions has not been investigated for predicting checkpoint inhibitor treatment outcomes in melanoma patients. Previous studies have investigated the use of deep learning on CT imaging for this purpose in other malignancies, namely non-small cell lung carcinoma (NSCLC) [11–14] and urothelial carcinoma [15,16], with positive findings. For melanoma, only handcrafted radiomics have been investigated thus far [17–20]. Initial findings by other smaller, single-center studies were promising, but our recent study of 620 patients from nine different centers showed different results: although the radiomics model had some value in predicting ICI treatment outcomes, it did not outperform a model based on clinical characteristics [20]. Deep learning may improve the performance over handcrafted radiomics as it is not limited by the choice of predefined features. This hypothesis remains to be experimentally verified, as studies comparing handcrafted radiomics to deep learning for other tasks show conflicting results [21–24].

The aim of this work was to determine the added value of deep learning on baseline CT imaging of lesions over clinical predictors for predicting first-line checkpoint inhibitor treatment outcomes in patients with advanced cutaneous melanoma. We have collected and curated a multi-center dataset of baseline CT imaging of these patients specifically for this purpose. With a sample size of 716 patients and 2722 lesions, this dataset is currently the largest of its kind in melanoma, and among the largest in all cancer types for imaging-based prediction of checkpoint inhibitor treatment outcomes [25].



## 4.3 Methods

### 4.3.1 Patient selection

Eligible patients were retrospectively identified from 10 participating centers (Amphia Ziekenhuis, Isala Zwolle, LUMC, Máxima MC, Medisch Spectrum Twente, Radboudumc, UMC Groningen, UMC Utrecht, Amsterdam UMC, Zuyderland MC) using prospectively collected high-quality registry data. With the exception of the UMC Groningen, this is the same population as in a previous work, which investigated handcrafted radiomics for the same purpose [20]. Patients were eligible if they were (i) treated for unresectable stage IIIC or IV cutaneous melanoma (ii) using first-line anti-PD1  $\pm$  anti-CTLA4 checkpoint inhibition (iii) on or after 1-1-2016 and (iv) were over 18 years of age at the start of treatment. Exclusion criteria were (i) unavailability of baseline contrast-enhanced CT imaging and (ii) absence of eligible lesions on CT.

### 4.3.2 ROI selection and preprocessing

Up to five lesions per patient were selected and manually segmented by authors LSM and IAJD under supervision of board-certified radiologists with 17 and 18 years of experience (PJ and TL, respectively). First, the five largest lesions were segmented with a maximum of two lesions per organ. Then, if fewer than five lesions had been segmented but more lesions remained, the largest remaining lesions were segmented up to a maximum of five. For example: in a patient with five large lung lesions and one small liver lesion, the two largest lung lesions and single liver lesion are segmented first. Then, the two largest remaining lung lesions are segmented, resulting in a total of five segmented lesions. Regions of interest (ROI) were extracted as isotropic cubes centered on the centroid of the segmentation; different methods were used to determine the size of the cube (Supplementary Materials, page 1). Pixel intensities were clipped to the range of -1024 to 3000 Hounsfield Units. During training and validation steps, the data was augmented through random rotation around all spatial axes and addition of Gaussian noise.

## 4.3. Methods

---

### 4.3.3 Outcome definition

Outcomes, as determined by the treating physician in line with RECIST 1.1 criteria [26], were extracted from prospectively collected high-quality registry data. The primary outcome was clinical benefit, defined as a best overall response of ‘stable disease’ for a minimum of six months, or ‘partial response’ or ‘complete response’. The secondary outcome was objective response, defined as a best overall response of ‘partial response’ or ‘complete response’. In addition, lesion outcomes were determined based on maximum diameter measurements at baseline, 3, 6 and 9 months. If the maximum diameter at the last available measurement exceeded 120% of the original maximum diameter, the lesion was labeled as ‘no benefit’, and otherwise as ‘benefit’. Similarly, lesions were labeled as ‘response’ or ‘no response’ using a 70% cut-off. Both cut-offs were chosen in line with the RECIST 1.1 criteria for determining patient response.

### 4.3.4 Model selection and hyperparameter selection

To arrive at a well-optimized model, a range of options for certain design choices (so-called ‘hyperparameters’) were systematically explored. These hyperparameters included, among others, model architecture, learning rate and parameters to control for overfitting, namely model size, weight decay and dropout. For model architectures, considered options were ResNet [27], Squeeze-Excitation ResNet [28], EfficientNet [29], ResNeXt [30] and Vision Transformer [31]. A full list of all hyperparameters along with possible values is supplied in Supplementary Table 1. To efficiently explore the vast space of possible hyperparameter combinations, an iterative process was used. In every iteration, a small number of hyperparameters were investigated using a random search strategy and a randomly chosen fixed train-validation split. The values with the highest validation area under the receiver operating characteristic (AUROC) for predicting patient level outcomes were subsequently fixed. This process was repeated until optimal values were selected for all hyperparameters. An iteration was continued for a maximum of 100 epochs, with early stopping after 10 epochs of no improvement of the patient level area under the curve (AUC) on the validation set.

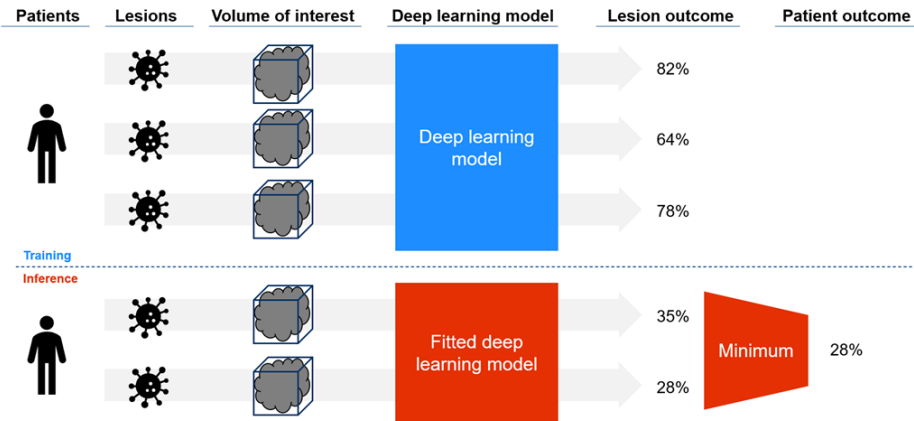
### 4.3.5 Cross validation, model training and evaluation

The selected configuration of model and hyperparameters was evaluated using a nested cross validation. The inner loop was conducted in a 5-fold cross validation. In every fold, 80% of the patients made up the training data; the remaining 20% was used as a validation set for monitoring training and early stopping. Repeating this process in all five folds resulted in five trained models, which were used in an ensemble: a combined model that averages the predictions of the five models per lesion. The outer loop was conducted in a leave-one-center-out manner and was used to evaluate the performance of the ensemble on an independent test set. During training, the model was optimized to predict the lesion level outcome based on the ROI of the corresponding lesion. During inference, these lesion level predictions were aggregated to a patient level by taking the minimum, mean or maximum of all predictions for a single patient. The choice for minimum, mean or maximum was also considered a hyperparameter. For predicting patient clinical benefit, lesion benefit was used as the lesion level label; for predicting patient objective response, lesion response was used as the lesion level label. Lesions with unavailable lesion level outcomes could not be used during training; these lesions were used during inference, however, as only patient level outcomes were necessary at this stage. The model was compared to a previously published clinical model [20] and a combination model of both the deep learning model and clinical model. The clinical model was a logistic regression based on four variables which were previously shown to be significant predictors of checkpoint inhibitor treatment outcomes in patients with advanced melanoma [1,7,33]. These predictors were presence of (i) liver and (ii) (a)symptomatic brain metastases, (iii) Eastern Cooperative Oncology Group (ECOG) performance status and (iv) levels of lactate dehydrogenase (LDH). Further details of the clinical and combination model are available in the Supplementary Methods.

### 4.3.6 Statistical analysis

Model calibration was assessed using calibration curves and Hosmer-Lemeshow test. Model discrimination was assessed using the receiver-operator characteristics (ROC) curve and corresponding AUC; 95% confidence intervals were calculated using the cvAUC R package [34]. Methods for comparing cross validated AUCs are

## 4.4. Methods



**Figure 4.1:** From left to right: for every eligible patient, up to five representative lesions are selected. A 3D volume of interest on the pretreatment CT scan is used as input for the deep learning model. During training (above the dotted line), the deep learning model is optimized to predict the probability of benefit from checkpoint inhibition for every individual lesion. During inference (below the dotted line), the fitted deep learning model is used to make lesion level predictions. These lesion level predictions are subsequently aggregated to a patient level prediction. Several options were explored for how to aggregate lesion level predictions, namely by taking the maximum, mean or minimum of predictions. After hyperparameter tuning, the ‘minimum’ function was selected.

described in the Supplementary Methods. The learned representation of the deep learning model was visualized using a two-dimensional t-distributed stochastic neighbor embedding (t-SNE).

### 4.3.7 Adherence to quality standards

After review by the Medical Ethics Committee (NedMec), this study was deemed not subject to the Medical Research Involving Human Subjects Act in accordance with Dutch regulations. Informed consent was waived.

## 4.4 Results

### 4.4.1 Patient characteristics

Out of 1347 eligible patients, 617 patients were excluded, resulting in 730 included patients with 2722 lesions; most exclusions were due to the availability of only a low-dose CT from a combined FDG-PET scan, instead of a diagnostic CT scan. A flowchart of the inclusion process is shown in Figure 4.2. 59.6% of 730 patients had clinical benefit (435 patients); the objective response rate was 51.1% (373 patients). Outcomes for individual lesions were available in 2128 lesions (78.2%); 21.8% of lesion outcomes were unavailable due to local regulations in one hospital (12.0%, 327 lesions), due to death or clinical progression before the first follow-up moment (7.4%, 202 lesions), due to the lesion falling outside the field-of-view of the scan (0.8%, 21 lesions) or due to technical problems (1.6%, 44 lesions); availability of lesion outcomes at 3, 6 and 9 months is shown in Supplementary Table 2. Among lesions with available outcomes, the rate of benefit was 79.7%; the lesion response rate was 55.2%. Characteristics of included and excluded patients are displayed in Table 4.1 and Supplementary Table 3. Included patients had on average more advanced disease than excluded patients. Scanner characteristics, acquisition parameters and patient characteristics per center and subgroup are shown in Supplementary Tables 4-6.

### 4.4.2 Hyperparameter selection

Ten iterations of preliminary experiments were performed; the results are available online through Supplementary Table 7. Based on these experiments, the model architecture was set to the Squeeze-Excitation ResNet50 [28] model with 3-dimensional input and random initial parameters; the function for aggregating predictions of all lesions belonging to one patient was selected to be ‘minimum’. The Adam optimizer was used with a cosine annealing learning rate scheduler. Other hyperparameters are listed in Supplementary Table 1.

### 4.4.3 Treatment outcome prediction

The deep learning model achieved a leave-one-center-out cross-validated AUROC of 0.607 [95% CI 0.565-0.648] for predicting clinical benefit. In comparison, the

#### 4.4. Results

n		730
Age, median [Q1,Q3]		68.0 [58.0,75.0]
Sex, n (%)	Female	285 (39.0)
	Male	445 (61.0)
Therapy, n (%)	Anti-PD1	458 (62.7)
	Ipilimumab & Nivolumab	272 (37.3)
Stage, n (%)	IIC	28 (3.8)
	IV M1a	56 (7.7)
	IV M1b	114 (15.6)
	IV M1c	344 (47.1)
	IV M1d	182 (24.9)
	Missing	6 (0.8)
ECOG performance status, n (%)	0	356 (48.8)
	1	271 (37.1)
	2-4	73 (10.0)
	Missing	30 (4.1)
Brain metastases, n (%)	absent	497 (68.1)
	asymptomatic	94 (12.9)
	symptomatic	88 (12.1)
	missing	51 (7.0)
Liver metastases, n (%)	absent	471 (64.5)
	present	224 (30.7)
	missing	35 (4.8)
LDH, n (%)	normal	459 (62.9)
	1-2x ULN	199 (27.3)
	>2x ULN	62 (8.5)
	missing	10 (1.4)
Number of affected organs, n (%)	<3	432 (59.2)
	≥ 3	298 (40.8)
Clinical benefit, n (%)	benefit	435 (59.6)
	no benefit	295 (40.4)
Objective response, n (%)	response	373 (51.1)
	no response	357 (48.9)

**Table 4.1:** Characteristics of included patients

clinical model achieved an AUROC of 0.635 [95% CI 0.592-0.678], and the combination model an AUROC of 0.635 [95% CI 0.595-0.676]. Differences in AUROC between the clinical and combination model were not statistically significant (Supplementary Figures 3). There was no evidence of poor fit in the three models (Hosmer-Lemeshow  $p > 0.113$ ). The 95% interval of predicted probabilities was 0.51-0.63 for the deep learning model, 0.28-0.77 for the clinical model and 0.49-0.72 for the combination model. Results were similar for prediction of objective

response (Supplementary Figure 4 and 5), and in treatment subgroups (Supplementary Figure 6-9).

#### 4.4.4 Interpretability analysis

Figure 5 shows the t-SNE embedding of the final layer of one of the fitted models (outer fold ‘Amsterdam UMC’, inner fold 3). The t-SNE analysis shows that the deep learning model learns to detect a lesion’s organ location (Figure 4.5A). Especially for liver and lung lesions, the predicted probability of lesion benefit is lower and higher, respectively (Figure 4.5D). However, there is a large overlap between benefitting and non-benefitting lesions (Figure 4.5B). Supplementary Figures 10 and 11 show the same analysis for different outer and inner folds. In line with these findings, Figure 4.6 shows that the patient level predictions of the deep learning model are significantly correlated with four out of five of the clinical predictors (Kruskal-Wallis  $p < 0.020$ ). Furthermore, lesion level predictions are weakly but significantly correlated with lesion volume ( $r = -0.28$ ,  $p < 0.0001$ ).

## 4.5 Discussion

A deep learning model on CT imaging of lesions had a significant but clinically limited predictive value for predicting response to checkpoint inhibitors in patients with advanced melanoma. Despite the substantial dataset size and extensive hyperparameter tuning, the achieved level of discrimination was limited. This result, combined with earlier findings on handcrafted radiomics, indicates that CT imaging of melanoma lesions at baseline holds limited information about treatment outcomes. Other studies have demonstrated that using on-treatment scans yields substantially better predictive performance, but on-treatment prediction is clinically far less relevant: most toxicity occurs in the first three months [35], and conventional follow-up measurements can already accurately predict long-term outcomes [36].

Addition of this deep learning model to clinical predictors did not improve predictive value. The difference in discrimination between both models was marginal. This was despite the large sample size and the cross-validation setup, which leverages every patient for independent validation. Furthermore, the range of predicted

## 4.5. Discussion

---

probabilities was wider for the clinical model.

This overlap in predictive value is likely to stem from the fact that the deep learning model learns information which is already encoded in the clinical model. The most plausible explanation is that the model encodes a lesion’s size and organ location, which may subsequently be correlated with stage and tumor load and therefore LDH, ECOG performance status and number of affected organs. This is in line with our earlier findings using a handcrafted radiomics approach [20].

The present work has important implications for future research. First, the overlap in predictive information between the clinical and deep learning model shows that it is essential to assess the added value of an imaging-based model over known predictors. In practice, however, this is rarely done [25]. Second, the present work suggests that previous results on imaging-based prediction of checkpoint inhibitor outcomes may be overoptimistic. Published results are almost exclusively positive, but numerous concerns exist regarding study size and quality [25]. The fact that these positive results are not confirmed in a large, multicenter dataset curated specifically for this purpose nuances this optimism.

The strengths of the present work are the large sample size and multicenter design. The training of deep learning models requires a substantial dataset size due to the large number of trainable parameters. To our knowledge, we have collected the largest dataset to date. Furthermore, the multicenter design allows for the evaluation of the generalizability of the model to new centers, which was a limitation of most previous studies. This, in combination with the cross-validation setup, adds significantly to the strength of the presented analysis.

This study has two potential limitations. First, a large group of patients was excluded due to unavailability of a contrast-enhanced baseline CT scan. Our hypothesis for the small difference in disease stage between in- and excluded patients is that patients with more advanced disease are more likely to present to medical oncology directly, instead of being referred after an FDG-PET CT scan has been performed. The risk of selection bias is limited however, as absolute differences in characteristics between in- and excluded patients are small. Second, performance of the deep learning model could in theory improve with the inclusion of more than five lesions per patient. However, we believe this is unlikely to change the conclusion, as a sensitivity analysis in a subset of the data did not show a difference in performance when more lesions were included. Furthermore, more than



## Chapter 4. Deep learning on CT scans to predict checkpoint inhibitor treatment outcomes in advanced melanoma

---

half of patients have at most five lesions.

In conclusion, a deep learning model based on baseline CT imaging of melanoma lesions had limited value for predicting checkpoint inhibitor treatment outcomes. Furthermore, this approach was unable to add information over a clinical model. The predictive value of the deep learning model was very comparable to a radiomics model, indicating that the predefined features of a handcrafted radiomics approach are not the limiting factor. Instead, the limited predictive power suggests a lack of predictive information regarding checkpoint inhibitor response in the single-energy CT images of melanoma lesions. Future research may investigate spectral CT imaging, or body composition metrics extracted from baseline CT imaging. Furthermore, research in other modalities remains necessary to move towards accurate baseline predictions of treatment response.

### 4.6 Supplementary materials

Supplementary materials are available through

<https://www.medrxiv.org/content/medrxiv/early/2023/07/27/2023.07.25.23293133/DC1/embed/media-1.docx?download=true>

### 4.7 References

- [1] van Zeijl MCT, Haanen JBAG, Wouters MWJM, de Wreede LC, Jochems A, Aarts MJB, et al. Real-world Outcomes of First-line Anti-PD-1 Therapy for Advanced Melanoma: A Nationwide Population-based Study. *J Immunother* 2020;43:256–64. <https://doi.org/10.1097/CJI.0000000000000334>
- [2] Korn EL, Liu P-Y, Lee SJ, Chapman J-AW, Niedzwiecki D, Suman VJ, et al. Meta-Analysis of Phase II Cooperative Group Trials in Metastatic Stage IV Melanoma to Determine Progression-Free and Overall Survival Benchmarks for Future Phase II Trials. *J Clin Oncol* 2008;26:527–34. <https://doi.org/10.1200/JCO.2007.12.7837>
- [3] Asher N, Ben-Betzalel G, Lev-Ari S, Shapira-Frommer R, Steinberg-Silman Y, Gochman N, et al. Real World Outcomes of Ipilimumab and Nivolumab in Patients with Metastatic Melanoma. *Cancers* 2020;12:2329. <https://doi.org/10.3390/cancers12082329>
- [4] Leeneman B, Uyl-de Groot CA, Aarts MJB, van Akkooi ACJ, van den Berkmortel FWPJ, van den Eertwegh AJM, et al. Healthcare Costs of Metastatic Cutaneous Melanoma in the Era of Immunotherapeutic and Targeted Drugs. *Cancers* 2020;12:E1003. <https://doi.org/10.3390/cancers12041003>
- [5] Verma V, Sprave T, Haque W, Simone CB, Chang JY, Welsh JW, et al. A systematic review of the cost and cost-effectiveness studies of immune checkpoint inhibitors. *J Immunother Cancer* 2018;6:128. <https://doi.org/10.1186/s40425-018-0442-7>
- [6] Wolchok JD, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Long-Term Outcomes With Nivolumab Plus Ipilimumab or Nivolumab Alone Versus Ipilimumab in Patients With Advanced Melanoma. *J Clin Oncol* 2022;40:127–37. <https://doi.org/10.1200/JCO.21.02229>
- [7] Silva IP da, Ahmed T, McQuade JL, Nebhan CA, Park JJ, Versluis JM, et al. Clinical Models to Define Response and Survival With Anti-PD-1 Antibodies Alone or Combined With Ipilimumab in Metastatic Melanoma. *J Clin Oncol* 2022. <https://doi.org/10.1200/JCO.21.01701>

## 4.7. References

---

- [8] Morrison C, Pabla S, Conroy JM, Nesline MK, Glenn ST, Dressman D, et al. Predicting response to checkpoint inhibitors in melanoma beyond PD-L1 and mutational burden. *J Immunother Cancer* 2018;6:32. <https://doi.org/10.1186/s40425-018-0344-8>
- [9] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278:563–77. <https://doi.org/10.1148/radiol.2015151169>
- [10] Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities. *IEEE Signal Process Mag* 2019;36:132–60. <https://doi.org/10.1109/NSP.2019.2900993>
- [11] Park C, Na KJ, Choi H, Ock C-Y, Ha S, Kim M, et al. Tumor immune profiles noninvasively estimated by FDG PET with deep learning correlate with immunotherapy response in lung adenocarcinoma. *Theranostics* 2020;10:10838–48. <https://doi.org/10.7150/thno.50283>
- [12] He B, Dong D, She Y, Zhou C, Fang M, Zhu Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. *J Immunother Cancer* 2020;8. <https://doi.org/10.1136/jitc-2020-000550>
- [13] Mu W, Jiang L, Zhang JY, Shi Y, Gray JE, Tunali I, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun* 2020;11. <https://doi.org/10.1038/s41467-020-19116-x>
- [14] Tian P, He B, Mu W, Liu K, Liu L, Zeng H, et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. *Theranostics* 2021;11:2098–107. <https://doi.org/10.7150/thno.48027>
- [15] Rundo F, Bersanelli M, Urzia V, Friedlaender A, Cantale O, Calcara G, et al. Three-Dimensional Deep Noninvasive Radiomics for the Prediction of Disease Control in Patients With Metastatic Urothelial Carcinoma treated With Immunotherapy. *Clin Genitourin Cancer* 2021;19:396–404. <https://doi.org/10.1016/j.clgc.2021.03.012>
- [16] Rundo F, Banna GL, Prezzavento L, Trenta F, Conoci S, Battiato S. 3D Non-Local Neural Network: A Non-Invasive Biomarker for Immunotherapy Treatment Outcome Prediction. Case-Study: Metastatic Urothelial Carcinoma. *J Imaging* 2020;6. <https://doi.org/10.3390/jimaging6120133>
- [17] Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol Off J Eur Soc Med Oncol* 2019;30:998–1004. <https://doi.org/10.1093/annonc/mdz108>
- [18] Brendlin AS, Peisen F, Almansour H, Afat S, Eigentler T, Amaral T, et al. A Machine learning model trained on dual-energy CT radiomics significantly improves immunotherapy response prediction for patients with stage IV melanoma. *J Immunother Cancer* 2021;9. <https://doi.org/10.1136/jitc-2021-003261>
- [19] Peisen F, Hänsch A, Hering A, Brendlin AS, Afat S, Nikolaou K, et al. Combination of Whole-Body Baseline CT Radiomics and Clinical Parameters to Predict Response and Survival in a Stage-IV Melanoma Cohort Undergoing Immunotherapy. *Cancers* 2022;14:2992. <https://doi.org/10.3390/cancers14122992>
- [20] Maat LS ter, Duin IAJ van, Elias SG, Leiner T, Verhoeff JJC, Arntz ER a. N, et al. CT radiomics to predict checkpoint inhibitors treatment outcomes in patients with advanced cutaneous melanoma 2022:2022.12.19.22283574. <https://doi.org/10.1101/2022.12.19.22283574>
- [21] Xia X, Gong J, Hao W, Yang T, Lin Y, Wang S, et al. Comparison and Fusion of Deep Learning and Radiomics Features of Ground-Glass Nodules to Predict the Invasiveness Risk of Stage-I Lung Adenocarcinomas in CT Scan. *Front Oncol* 2020;10.
- [22] Fradet G, Ayde R, Bottois H, El Harchaoui M, Khaled W, Drapé J-L, et al. Prediction of lipomatous soft tissue malignancy on MRI: comparison between machine learning applied to radiomics and deep learning. *Eur Radiol Exp* 2022;6:41. <https://doi.org/10.1186/s41747-022-00295-9>
- [23] Li X, Yang L, Jiao X. Comparison of Traditional Radiomics, Deep Learning Radiomics and Fusion Methods for Axillary Lymph Node Metastasis Prediction in Breast Cancer. *Acad Radiol* 2022. <https://doi.org/10.1016/j.acra.2022.10.015>
- [24] Castillo T. JM, Arif M, Starmans MPA, Niessen WJ, Bangma CH, Schoots IG, et al. Classification of Clinically Significant Prostate Cancer on Multi-Parametric MRI: A Validation Study Comparing Deep Learning and Radiomics. *Cancers* 2022;14:12. <https://doi.org/10.3390/cancers14010012>

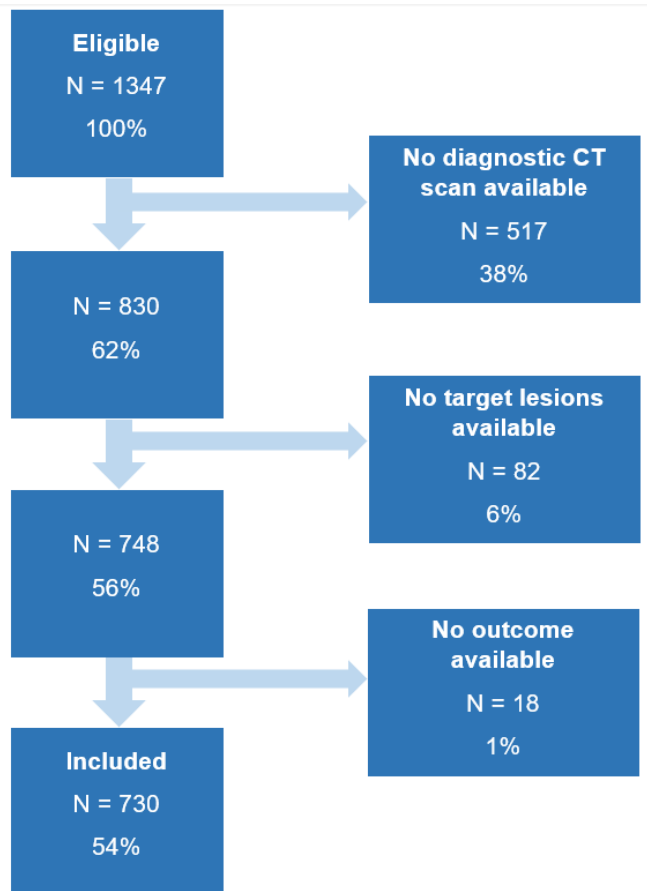
## Chapter 4. Deep learning on CT scans to predict checkpoint inhibitor treatment outcomes in advanced melanoma

---

- [25 ] ter Maat LS, van Duin IAJ, Elias SG, van Diest PJ, Pluim JPW, Verhoeff JJC, et al. Imaging to predict checkpoint inhibitor outcomes in cancer. A systematic review. *Eur J Cancer* 2022;175:60–76. <https://doi.org/10.1016/j.ejca.2022.07.034>
- [26 ] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47. <https://doi.org/10.1016/j.ejca.2008.10.026>
- [27 ] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition, 2016, p. 770–8.
- [28 ] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks, 2018, p. 7132–41.
- [29 ] Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proc. 36th Int. Conf. Mach. Learn.*, PMLR; 2019, p. 6105–14.
- [30 ] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated Residual Transformations for Deep Neural Networks, 2017, p. 1492–500.
- [31 ] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale 2021. <https://doi.org/10.48550/arXiv.2010.11929>
- [32 ] Maurício J, Domingues I, Bernardino J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. *Appl Sci* 2023;13:5521. <https://doi.org/10.3390/app13095521>
- [33 ] van Zeijl MCT, de Wreede LC, van den Eertwegh AJM, Wouters MWJM, Jochems A, Schouwenburg MG, et al. Survival outcomes of patients with advanced melanoma from 2013 to 2017: Results of a nationwide population-based registry. *Eur J Cancer* 2021;144:242–51. <https://doi.org/10.1016/j.ejca.2020.11.028>
- [34 ] LeDell E, Petersen M, van der Laan M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron J Stat* 2015;9:1583–607. <https://doi.org/10.1214/15-EJS1035>
- [35 ] Larkin J, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med* 2019;381:1535–46. <https://doi.org/10.1056/NEJMoa1910836>
- [36 ] Robert C, Long GV, Brady B, Dutriaux C, Giacomo AMD, Mortier L, et al. Five-Year Outcomes With Nivolumab in Patients With Wild-Type BRAF Advanced Melanoma. *J Clin Oncol* 2020. <https://doi.org/10.1200/JCO.20.00995>

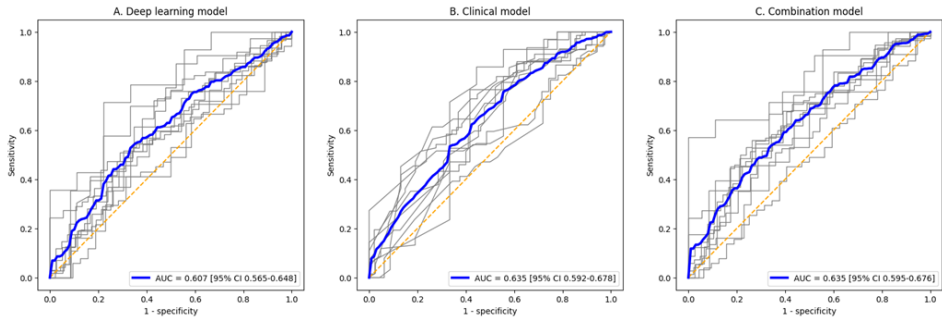
## 4.7. References

---

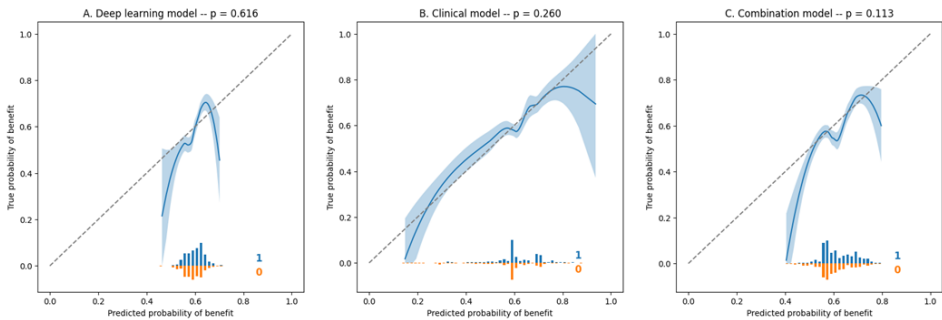


**Figure 4.2:** Flowchart of the inclusion process

## Chapter 4. Deep learning on CT scans to predict checkpoint inhibitor treatment outcomes in advanced melanoma



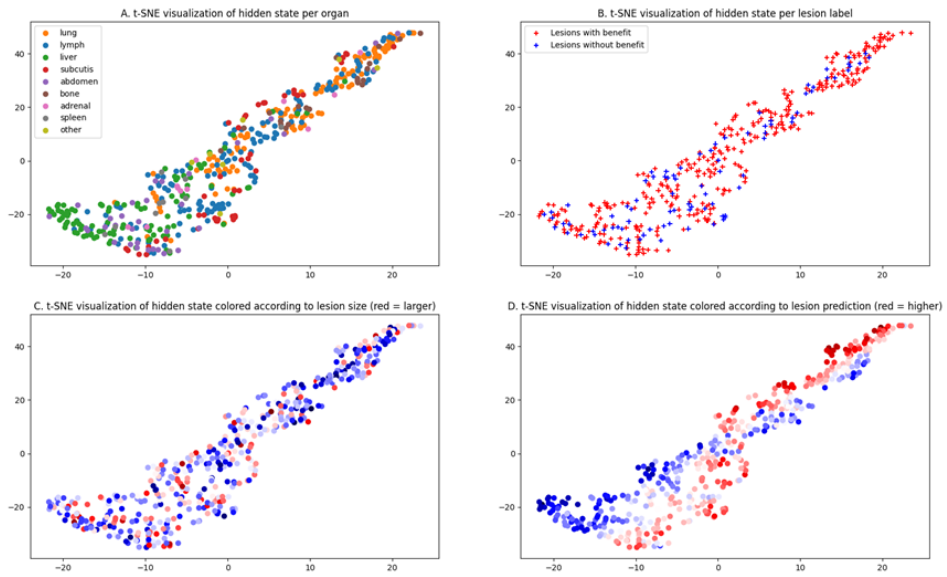
**Figure 4.3:** Receiver operator characteristic (ROC) curves for (A) the deep learning model, (B) the baseline clinical model and (C) the combination model for predicting clinical benefit on a patient level. Curves of the individual folds/validation centers are shown in gray; the average ROC curve is shown in blue. Corresponding areas under the curve (AUC) are supplied in the legend. The orange line corresponds to the line of random performance.



**Figure 4.4:** Locally estimated scatterplot smoothing (LOESS) fitted calibration curves with corresponding 95% confidence interval for (A) the deep learning model, (B) the clinical model and (C) the combination model for predicting clinical benefit on a patient level. The dashed line indicates the line of perfect calibration. Histograms of individual predictions, split for patients with (blue) and without (orange) benefit, are shown below the curves. The p-value for the Hosmer-Lemeshow test for goodness of fit is shown in the plot title.

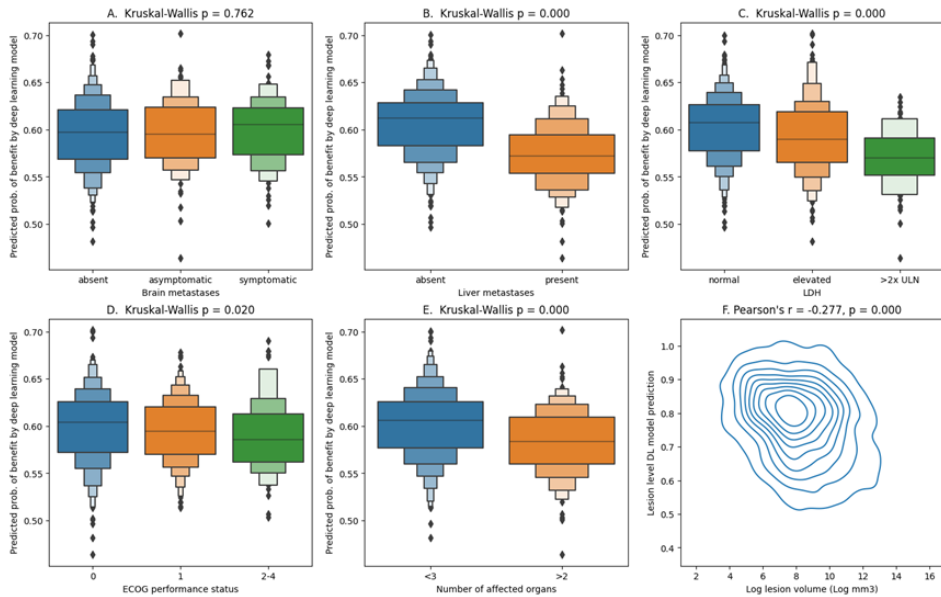
## 4.7. References

---



**Figure 4.5:** Based on the training data, the deep learning model learns to map every lesion to a point in space where, intuitively, similar lesions are closer together. This mapping is visualized in this figure in 2D using t-SNE. Every point corresponds to a single lesion. Relative distance indicates how similar lesions are according to the model; absolute location is not informative in this figure. Lesions are colored in the four different plots to show how the information learned by the model corresponds with information about the lesion. (A) Lesions located in different organs are clustered together, indicating that the deep learning model detects the lesion’s location. (B) There is no clear separation of lesions with and without benefit, indicating that the model cannot accurately discriminate between lesion treatment outcomes. (C) Although some clusters of large and small lesions can be seen, lesion size appears to be less determining for the model’s output than location. (D) Overall, predicted probability of benefit is lower in lesions marked as liver lesions in Figure 4A, and higher in lung lesions.

## Chapter 4. Deep learning on CT scans to predict checkpoint inhibitor treatment outcomes in advanced melanoma



**Figure 4.6:** (A-E) Boxplots of known clinical predictors with the output of the deep learning model for predicting clinical benefit per patient. P-values of the Kruskal-Wallis test for difference in distribution are given in the plot titles. (A) The output of the deep learning model is not significantly different for patients with or without brain metastases. (B-E) The output of the deep learning model is significantly different for patients with and without liver metastases (B), with varying levels of LDH (C), different ECOG performance status (D) and with less than 3 and 3 or more affected organs (E). (F) Kernel-density estimate plot of log-transformed lesion volume versus the lesion-level prediction of the deep learning model. The output of the deep learning model per lesion is significantly lower in larger lesions.

## 4.7. References

---



## Chapter 5

# Body composition and checkpoint inhibitor treatment outcomes in advanced melanoma: a multicenter cohort study

L.S. Ter Maat MD, I.A.J. Van Duin MD, R.J. Verheijden MSc, P. Moeskops PhD, J.J.C. Verhoeff MD PhD, S.G. Elias MD PhD, W.A.C. van Amsterdam MD PhD, F.H. Burgers MD, F.W.P.J. Van den Berkmortel MD PhD, M.J. Boers-Sonderen MD PhD, M.F. Boomsma MD PhD, J.W. De Groot MD PhD, J.B.A.G. Haanen MD PhD, G.A.P. Hospers MD PhD, D. Piersma MD PhD, G. Vreugdenhil MD PhD, H.M. Westgeest MD PhD, E. Kapiteijn MD PhD, M. Labots MD PhD, W.B. Veldhuis MD PhD, P.J. Van Diest MD PhD, P.A. De Jong MD PhD, J.P.W. Pluim PhD, T. Leiner MD PhD, M. Veta PhD, K.P.M. Suijkerbuijk MD PhD

*Submitted*

## 5.1. Abstract

---

# 5.1 Abstract

### 5.1.1 Introduction

The association of body composition with checkpoint inhibitor outcomes in melanoma is a matter of ongoing debate. In this study, we aim to add to previous evidence by investigating body mass index (BMI) alongside CT derived body composition metrics in the largest cohort to date.

### 5.1.2 Methods

Patients treated with first-line anti-PD1 ± anti-CTLA4 for advanced melanoma were retrospectively identified from 11 melanoma reference centers in The Netherlands. Age, sex, Eastern Cooperative Oncology Group performance status, serum lactate dehydrogenase, presence of brain and liver metastases, number of affected organs and BMI at baseline were extracted from electronic patient files. From baseline CT scans, five body composition metrics were automatically extracted: skeletal muscle index, skeletal muscle density, skeletal muscle gauge, subcutaneous adipose tissue index and visceral adipose tissue index. All predictors were correlated in uni- and multivariable analysis to progression-free, overall and melanoma-specific survival (PFS, OS and MSS) using Cox proportional hazards models.

### 5.1.3 Results

A total of 1471 eligible patients were included. Median PFS and OS were 8.8 and 34.8 months, respectively. A significantly worse PFS was observed in underweight patients (multivariable HR=1.87, 95% CI 1.14–3.07). Furthermore, better OS was observed in patients with higher skeletal muscle density (multivariable HR=0.91, 95% CI 0.83-0.99) and gauge (multivariable HR=0.88, 95% CI 0.84-0.996), and a worse OS with higher visceral adipose tissue index (multivariable HR=1.13, 95% CI 1.04-1.22). No association with survival outcomes was found for overweightness or obesity and survival outcomes, or for subcutaneous adipose tissue.

### **5.1.4 Discussion**

Our findings suggest that underweight BMI is associated with worse PFS, whereas higher skeletal muscle density and lower visceral adipose tissue index were associated with better OS. These associations were independent of previously identified predictors, including sex, age, performance status and extent of disease. No significant association between higher BMI and survival outcomes was observed.

## **5.2 Introduction**

The introduction of checkpoint inhibitors has revolutionized advanced melanoma care. The prognosis for advanced melanoma was historically very poor, with a 1-year overall survival of less than 25% [1]. In contrast, patients treated in the CheckMate 067 trial with anti-programmed cell death 1 (anti-PD1) had a 6.5-year overall survival rate of 43%. Patients treated with both anti-PD1 and anti-cytotoxic T-lymphocyte associated protein-4 (anti-CTLA4) antibodies even had a 6.5-year overall survival rate of 57% [2].

However, many open questions remain about how checkpoint inhibitors interact with tumor and host. Both anti-CTLA4 and anti-PD1 antibodies block proteins that inhibit immune response, which leads to increased immune activity against the tumor [3]. Although some mechanisms of primary resistance have been identified [4], it is not fully understood why some patients progress during treatment while others do not.

One such open question is the association between obesity and checkpoint inhibitor treatment outcomes. On the one hand, several pan-cancer meta-analyses published in 2020 and 2021 reported better survival outcomes in patients with obesity compared to patients with normal body mass index (BMI) [5–7]. This association, dubbed the “obesity paradox”, was also found to be significant in the subgroup of studies on patients with melanoma [6,7]. On the other hand, an updated meta-analysis by Rocuzzo et al. (2023) in melanoma concluded that the prognostic value of BMI could not be confirmed due to the limited available evidence [8]. This indicates that the topic of obesity and checkpoint inhibitor treatment outcomes is an area of ongoing research where more high-quality evidence is needed.

### 5.3. Methods

---

In addition to BMI, previous works investigated computed tomography (CT) derived body composition metrics. These metrics include the amount and density of skeletal muscle and the amount of subcutaneous and adipose tissue [9]. Due to advances in deep learning for automatic image analysis, this category of predictors has become increasingly prominent in research in recent years [10,11]. The advantage of these metrics is that they can more accurately capture a patient’s body composition, whereas BMI may misrepresent patients with high muscle mass and cannot distinguish between patients with high visceral or subcutaneous adipose tissue. Previous studies on these metrics, however, reported differing results and have some methodological limitations, most notably a limited sample size [12].

Several causal mechanisms have been proposed for explaining associations between body composition and checkpoint inhibitor outcomes. First, a more aggressive disease may affect both body composition (e.g., through weight loss) and outcomes. Second, patients with a worse physical condition, as reflected in body composition metrics, may succumb more quickly to their disease. Third, body composition may modulate the efficacy of checkpoint inhibitor therapy. For example, an increased efficacy of anti-PD(L)1 therapy was observed in obese mice compared to mice with normal weight [13]. Furthermore, increased PD-1 expression was noted in obese patients with melanoma [13].

Research into these causal mechanisms, however, is hindered by the controversy surrounding the association between body composition and checkpoint inhibitor treatment outcomes. This work therefore aimed to contribute to the existing evidence on this topic by presenting the largest cohort to our knowledge to date. Additionally, we aimed to provide a more fine-grained picture of body composition by evaluating CT derived metrics alongside BMI.

## 5.3 Methods

### 5.3.1 Patient selection

Patients were eligible if they were (i) over 18 years of age, (ii) treated for unresectable stage IIIC or stage IV cutaneous melanoma with (iii) first-line anti-PD1 with or without CTLA4 inhibition (iv) between January 1st, 2016, and February 1st, 2023. Patients were excluded if (i) no baseline CT scan was available, (ii) no

## Chapter 5. Body composition and checkpoint inhibitor treatment outcomes in advanced melanoma: a multicenter cohort study

---

transverse slice of the third lumbar vertebrae was in the field of view of the scan, (iii) metal artefacts were present at the L3 level or (iv) patient height or weight at baseline were unavailable. Eligible patients from eleven melanoma treatment centers in the Netherlands (Amphia Breda, Amsterdam UMC, Isala Zwolle, Leiden University Medical Center, Maxima MC, Medisch Spectrum Twente, Netherlands Cancer Institute, Radboudumc, University Medical Center Groningen, University Medical Center Utrecht, Zuyderland) were identified using high-quality registry data. This study was deemed not subject to Medical Research Involving Human Subjects Act according to Dutch regulations by the Medical Ethics Committee; informed consent was waived.

### 5.3.2 BMI and clinical predictors

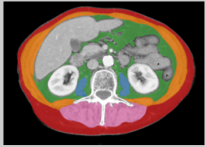
Height and weight at baseline were extracted from electronic patient files and were used to calculate BMI. In addition, several previously identified clinical predictors of checkpoint inhibitor treatment outcomes in advanced melanoma were extracted. These were (i) Eastern Cooperative Oncology Group (ECOG) performance status, (ii) level of lactate dehydrogenase (LDH), presence of (iii) brain and (iv) liver metastases and (v) number of affected organs [14–17] (categories are shown in Figure 5.1).

### 5.3.3 CT body composition metrics extraction

Metrics were obtained using Quantib Body Composition version 0.2.1, a dedicated deep learning segmentation algorithm that has proven to achieve high correspondence to manual segmentations in previous studies [18–20]. First, all baseline CT scans were resampled to a slice thickness of 5mm. Subsequently, the slice in the middle of the third lumbar vertebra [21] was automatically selected using a convolutional neural network. On the five consecutive slices centered around this selected slice, the following compartments were automatically segmented using a second convolutional neural network: psoas, abdominal and long spine muscles (together making up the skeletal muscles), subcutaneous adipose tissue and visceral adipose tissue. All segmentations were manually reviewed and corrected where necessary. Based on these segmentations, five commonly used metrics [9,22–24] were calculated using the definitions in Table 1: skeletal muscle index (SMI),

### 5.3. Methods

skeletal muscle density (SMD), skeletal muscle gauge (SMG), subcutaneous adipose tissue index (SATI) and visceral adipose tissue index (VATI). All metrics were normalized to zero mean and unit standard deviation (SD) to facilitate interpretation. Since skeletal muscle density and gauge differed significantly between patients who underwent a contrast-enhanced CT scan versus those who underwent a non-contrast CT scan, SMD and SMG were normalized separately for both groups.

Predictor	Levels/definition	Example of segmentation and extracted features																								
<b>Brain metastases</b>	<ul style="list-style-type: none"> <li>Absent</li> <li>Present, <i>asymptomatic</i></li> <li>Present, <i>symptomatic</i></li> </ul>	 <table border="1" data-bbox="899 797 1044 1006"> <thead> <tr> <th>Region</th> <th>Area (cm²)</th> <th>%</th> <th>HU</th> </tr> </thead> <tbody> <tr> <td>Psoas</td> <td>9.6</td> <td>2.0%</td> <td>36.5</td> </tr> <tr> <td>Abdominal</td> <td>69.2</td> <td>14.5%</td> <td>6.6</td> </tr> <tr> <td>Long spine</td> <td>41.3</td> <td>8.6%</td> <td>11.5</td> </tr> <tr> <td>Subcutaneous</td> <td>88.0</td> <td>18.4%</td> <td>-78.5</td> </tr> <tr> <td>Visceral</td> <td>68.8</td> <td>14.4%</td> <td>-56.1</td> </tr> </tbody> </table>	Region	Area (cm²)	%	HU	Psoas	9.6	2.0%	36.5	Abdominal	69.2	14.5%	6.6	Long spine	41.3	8.6%	11.5	Subcutaneous	88.0	18.4%	-78.5	Visceral	68.8	14.4%	-56.1
Region	Area (cm²)		%	HU																						
Psoas	9.6		2.0%	36.5																						
Abdominal	69.2		14.5%	6.6																						
Long spine	41.3		8.6%	11.5																						
Subcutaneous	88.0		18.4%	-78.5																						
Visceral	68.8		14.4%	-56.1																						
<b>Liver metastases</b>	<ul style="list-style-type: none"> <li>Absent</li> <li>Present</li> </ul>																									
<b>LDH</b>	<ul style="list-style-type: none"> <li>Normal</li> <li>Elevated</li> <li>&gt;2x ULN</li> </ul>																									
<b>ECOG performance status</b>	<ul style="list-style-type: none"> <li>0</li> <li>1</li> <li>2-4</li> </ul>																									
<b>Number of affected organs</b>	<ul style="list-style-type: none"> <li>&lt;3</li> <li>&gt;2</li> </ul>																									
<b>Body Mass Index (continuous)</b>	$= \frac{\text{weight (kg)}}{\text{height (m)}^2}$																									
<b>Body Mass Index (categorical)</b>	<ul style="list-style-type: none"> <li>Underweight (BMI &lt; 18.5)</li> <li>Normal (18.5 &lt; BMI &lt; 25)</li> <li>Overweight (25 &lt; BMI &lt; 30)</li> <li>Obese (BMI &gt; 30)</li> </ul>																									
<b>Skeletal Muscle Index</b>	$= \frac{\text{skeletal muscle cross sectional area (cm}^2\text{)}}{\text{height (m)}}$																									
<b>Skeletal Muscle Density</b>	Mean skeletal muscle density in Hounsfield Units																									
<b>Skeletal Muscle Gauge</b>	$= \text{SMI} \cdot \text{SMD}$																									
<b>Subcutaneous Adipose Tissue Index</b>	$= \frac{\text{subcutaneous adipose tissue cross sectional area (cm}^2\text{)}}{\text{height (m)}}$																									
<b>Visceral Adipose Tissue Index</b>	$= \frac{\text{visceral adipose tissue cross sectional area (cm}^2\text{)}}{\text{height (m)}}$																									

CT=computed tomography, ECOG=Eastern Cooperative Oncology Group, LDH=lactate dehydrogenase, ULN=upper limit of normal, BMI=body mass index, SMI=skeletal muscle index, SMD=skeletal muscle density

Figure 5.1: Definition of included predictors and evaluated models

#### 5.3.4 Outcome definition

The primary endpoints were progression-free survival (PFS) and overall survival (OS). PFS was defined as the time from the start of treatment to progression or death; OS was defined as time from the start of treatment to death due to any cause. The secondary outcome was melanoma-specific survival (MSS), defined as the time from the start of treatment to death from melanoma. Patients not reaching the endpoint were right-censored at the date of the last contact, or when a different treatment was initiated.

### **5.3.5 Statistical analysis**

Correlation among body composition variables was assessed using Pearson's correlation coefficient. The association between body composition metrics and outcomes were assessed using uni- and multivariable Cox proportional hazards models. In multivariable analyses, a separate model was constructed for every body composition metric, combined with previously identified clinical factors (ECOG performance status, level of LDH, presence of brain and liver metastases and number of affected organs). BMI was assessed as a categorical variable, using the established cut-offs for underweight (<18.5), normal (between 18.5 and 25), overweight (between 25 and 30) and obese (>30). In addition, all variables were modelled using restricted cubic splines with three knots to account for non-linear effects. Multiple imputation was performed using the MICE R package with 21 imputations. Subgroup analyses were conducted for patients treated with monotherapy (anti-PD1) and combination therapy (anti-PD1 + anti-CTLA4), and for patients who underwent a contrast-enhanced and non-contrast CT scan. Unless stated otherwise, 95% confidence intervals are displayed.

## **5.4 Results**

### **5.4.1 Patient characteristics**

Out of 1944 eligible patients, 1471 patients (76%) were included (Supplementary Figure 1). Characteristics of the included patients are shown in Table 5.1; these characteristics were similar to those of excluded patients (Supplementary Table 1). Median PFS and OS were 9.1 and 38.1 months, respectively. Median MSS was not reached. The subgroups of patients treated with anti-PD1 monotherapy and anti-PD1 plus anti-CTLA-4 combination therapy consisted of 942 (64%) and 529 (36%) patients, respectively. Subgroups of patients who underwent non-contrast CT (in combination with 18-fluorodeoxyglucose positron emission tomography) versus contrast-enhanced consisted of 611 and 860 patients, respectively. Characteristics of patients in subgroups are shown in Supplementary Tables 2 and 3.

## 5.4. Results

n		1471
Age, mean (SD)		65.1 (13.0)
Sex, n (%)	Female	579 (39.4)
	Male	892 (60.6)
Therapy, n (%)	Anti-PD1	942 (64.0)
	Ipilimumab & Nivolumab	529 (36.0)
Scan type, n (%)	Contrast-enhanced	860 (58.5)
	No contrast	611 (41.5)
Stage, n (%)	IIIC	131 (8.9)
	IV M1a	130 (8.8)
	IV M1b	217 (14.8)
	IV M1c	639 (43.4)
	IV M1d	344 (23.4)
	missing	10 (0.7)
ECOG performance status, n (%)	0	798 (54.2)
	1	489 (33.2)
	2-4	110 (7.5)
	missing	74 (5.0)
Brain metastases, n (%)	absent	952 (64.7)
	asymptomatic	212 (14.4)
	symptomatic	132 (9.0)
	missing	175 (11.9)
Liver metastases, n (%)	absent	939 (63.8)
	present	379 (25.8)
	missing	153 (10.4)
LDH, n (%)	normal	1013 (68.9)
	1-2x ULN	330 (22.4)
	>2x ULN	110 (7.5)
	missing	18 (1.2)
Number of affected organs, n (%)	<3	886 (60.2)
	>2	585 (39.8)
Body Mass Index, n (%)	underweight	21 (1.4)
	normal	604 (41.1)
	overweight	586 (39.8)
	obese	260 (17.7)
Skeletal Muscle Index, median [Q1,Q3]		91.0 [78.8,102.4]
Skeletal Muscle Density, median [Q1,Q3]		19.5 [8.2,28.6]
Skeletal Muscle Gauge, median [Q1,Q3]		1685.4 [720.7,2629.5]
Subcutaneous Adipose Tissue Index, median [Q1,Q3]		91.2 [66.1,125.5]
Visceral Adipose Tissue Index, median [Q1,Q3]		83.1 [46.0,129.2]
Median overall survival (months)		38.1
Median progression-free survival (months)		9.1

**Table 5.1:** Characteristics of included patients

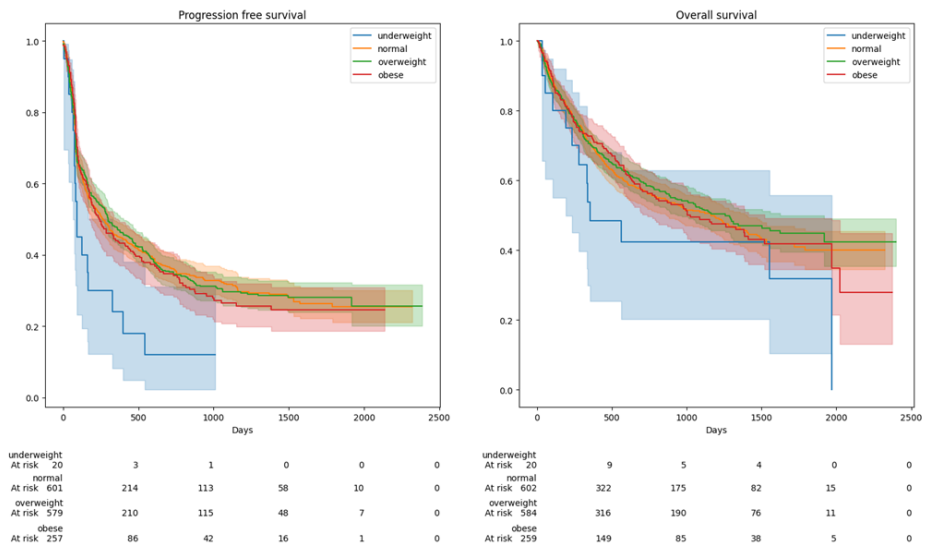
### 5.4.2 Body mass index

Out of 1471 patients, 21 (1.4%) were underweight, 604 (41.1%) had normal BMI, 586 (39.8%) were overweight and 260 (17.7%) were obese. Underweight patients had significantly worse PFS than patients with normal weight in both uni- and multivariable analysis (multivariable HR=1.87 95% CI 1.14-3.07, Table 5.2, Figure 5.2). A similar, but statistically nonsignificant association was observed for OS



## Chapter 5. Body composition and checkpoint inhibitor treatment outcomes in advanced melanoma: a multicenter cohort study

(multivariable HR=1.57, 95% CI 0.89-2.77, Table 5.3, Figure 5.2). Underweight patients had more advanced disease, worse ECOG performance status, higher levels of LDH at baseline and were less likely to receive combination therapy (Supplementary Table 4). OS and PFS were not significantly different in overweight or obese patients when compared to normal BMI. No significant associations with OS and PFS were observed when BMI was analyzed using restricted cubic splines (Supplementary Figures 2-3). Results were comparable in the performed subgroup analyses (Supplementary Tables 6-13).



**Figure 5.2:** Kaplan-Meier curves for progression free and overall survival according to BMI subgroup

### 5.4.3 CT derived body composition metrics

All body composition metrics were significantly correlated with each other (Supplementary Table 5). Of note is the negative correlation between skeletal muscle index and density ( $r = -0.14$ ). Significant associations with outcomes were observed for three of the five CT derived body composition metrics. First, higher skeletal muscle density was associated with better OS (multivariable HR=0.91 per SD increase, 95% CI 0.83-0.99, Table 5.3) and MSS (multivariable HR=0.90 per

## 5.5. Discussion

---

SD increase, 95% CI 0.81-0.999, Table 5.4). Second, higher skeletal muscle gauge was associated with better OS (multivariable HR=0.91 per SD increase, 95% CI 0.86-0.996, Table 3). Third, higher visceral adipose tissue index was associated with worse OS (multivariable HR=1.13 per SD increase, 95% CI 1.04-1.22, Table 5.3), with similar but statistically non-significant trends for PFS (multivariable HR=1.07 per SD increase, 95% CI 1.00-1.15, Table 5.2) and MSS (multivariable HR=1.10 per SD increase, 95% CI 0.997-1.21, Table 5.4). No significant associations were observed between skeletal muscle index or subcutaneous adipose tissue index and survival outcomes. Results were similar in subgroups of patients who underwent contrast-enhanced and non-contrast CT scans, in subgroups of patients treated with anti-PD1 and combination therapy (Supplementary Tables 6-13). When analyzing CT derived body composition metrics using restricted cubic splines, similar directions of effect were observed (Supplementary Figures 2-3).

## 5.5 Discussion

The contributions of this work are threefold. First, we demonstrate significantly worse PFS in patients who are underweight. Second, we find no evidence for an association between obesity and better outcomes. Third, we show that higher skeletal muscle density and gauge, and lower visceral adipose tissue index are associated with improved survival.

PFS was significantly worse in underweight patients. Surprisingly, this association was significant in multivariable analysis despite the association between underweight BMI and other poor baseline characteristics. Although this result must be interpreted with care due to the small numbers (N=21) in the underweight group, it may indicate that the prognosis of this group of patients is even worse than is to be expected based on their stage of disease, performance status and level of LDH. Potential explanations for this association are a confounding effect of tumor aggressiveness, and an increased vulnerability to complications due to reduced physical reserves.

We found no association between obesity and better treatment outcomes, when measured as BMI, or as visceral or subcutaneous adipose tissue index. In contrast, we observed worse survival in patients with more visceral adipose tissue, the type of

## Chapter 5. Body composition and checkpoint inhibitor treatment outcomes in advanced melanoma: a multicenter cohort study

		Univariable			Multivariable*		
		HR**	95% CI	p-value	HR**	95% CI	p-value
Body Mass Index	underweight	1.826	1.123 - 2.969	0.015	1.868	1.136 - 3.069	0.014
	normal	1.000			1.000		
	overweight	0.987	0.858 - 1.135	0.853	1.004	0.872 - 1.156	0.953
	obese	1.035	0.867 - 1.236	0.702	1.113	0.930 - 1.331	0.243
Skeletal Muscle Index		1.028	0.966 - 1.094	0.386	1.069	0.988 - 1.156	0.098
Skeletal Muscle Density		0.970	0.911 - 1.033	0.340	0.983	0.914 - 1.058	0.655
Skeletal Muscle Gauge		0.986	0.926 - 1.049	0.650	1.007	0.933 - 1.086	0.868
Subcutaneous Adipose Tissue Index		0.960	0.902 - 1.022	0.201	0.985	0.923 - 1.052	0.661
Visceral Adipose Tissue Index		1.065	1.001 - 1.132	0.045	1.070	1.000 - 1.146	0.051

**Table 5.2:** Univariable and multivariable Cox proportional hazards models for progression free survival. Abbreviations: HR=Hazard Rate Ratio, CI=Confidence Interval. \*Corrected for age, sex, serum lactate dehydrogenase, presence of brain metastases (absent vs. asymptomatic vs. symptomatic) and liver metastases, Eastern Cooperative Oncology group performance status and number of affected organs. \*\*Hazard rate ratios for skeletal muscle index, density and gauge, and subcutaneous and visceral adipose tissue index are provided per standard deviation increase.

		Univariable			Multivariable*		
		HR**	95% CI	p-value	HR**	95% CI	p-value
Body Mass Index	underweight	1.487	0.853 - 2.594	0.162	1.569	0.889 - 2.768	0.121
	normal	1.000			1.000		
	overweight	0.947	0.801 - 1.120	0.523	0.995	0.840 - 1.179	0.958
	obese	0.997	0.808 - 1.229	0.977	1.174	0.948 - 1.453	0.141
Skeletal Muscle Index		1.018	0.946 - 1.096	0.628	1.064	0.968 - 1.170	0.198
Skeletal Muscle Density		0.861	0.801 - 0.925	0.000	0.906	0.830 - 0.988	0.026
Skeletal Muscle Gauge		0.870	0.810 - 0.935	0.000	0.912	0.835 - 0.996	0.040
Subcutaneous Adipose Tissue Index		0.937	0.868 - 1.011	0.095	1.017	0.940 - 1.100	0.678
Visceral Adipose Tissue Index		1.138	1.060 - 1.222	0.000	1.126	1.038 - 1.221	0.004

**Table 5.3:** Univariable and multivariable Cox proportional hazards models for overall survival. Abbreviations: HR=Hazard Rate Ratio, CI=Confidence Interval. \*Corrected for age, sex, serum lactate dehydrogenase, presence of brain metastases (absent vs. asymptomatic vs. symptomatic) and liver metastases, Eastern Cooperative Oncology group performance status and number of affected organs. \*\*Hazard rate ratios for skeletal muscle index, density and gauge, and subcutaneous and visceral adipose tissue index are provided per standard deviation increase.

fat most associated with inflammation [25]. The other metrics that reflect obesity, namely subcutaneous adipose tissue index and higher BMI, were not associated with any of the investigated outcomes. These findings are in line with the meta-analysis by Rocuzzo et al. [8], which found no significant association between higher BMI and survival outcomes in melanoma. This meta-analysis thereby differs in its conclusion from earlier meta-analyses, a fact which can be explained by the inclusion of studies which were not yet published during these earlier analyses.

Better survival was observed in patients with higher skeletal muscle density

## 5.5. Discussion

		Univariable			Multivariable*		
		HR**	95% CI	p-value	HR**	95% CI	p-value
Body Mass Index	underweight	1.352	0.694 - 2.634	0.376	1.372	0.696 - 2.707	0.362
	normal	1.000			1.000		
	overweight	0.925	0.761 - 1.125	0.436	0.977	0.802 - 1.190	0.815
	obese	0.907	0.706 - 1.166	0.448	1.092	0.846 - 1.409	0.500
Skeletal Muscle Index		1.031	0.946 - 1.123	0.490	1.112	0.996 - 1.240	0.059
Skeletal Muscle Density		0.913	0.837 - 0.995	0.039	0.902	0.814 - 0.999	0.049
Skeletal Muscle Gauge		0.925	0.849 - 1.008	0.075	0.921	0.829 - 1.023	0.124
Subcutaneous Adipose Tissue Index		0.937	0.856 - 1.025	0.153	1.017	0.928 - 1.115	0.722
Visceral Adipose Tissue Index		1.078	0.990 - 1.174	0.085	1.098	0.997 - 1.210	0.059

**Table 5.4:** Univariable and multivariable Cox proportional hazards models for melanoma specific survival. Abbreviations: HR=Hazard Rate Ratio, CI=Confidence Interval. \*Corrected for age, sex, serum lactate dehydrogenase, presence of brain metastases (absent vs. asymptomatic vs. symptomatic) and liver metastases, Eastern Cooperative Oncology group performance status and number of affected organs. \*\*Hazard rate ratios for skeletal muscle index, density and gauge, and subcutaneous and visceral adipose tissue index are provided per standard deviation increase.

and gauge, and lower adipose tissue index. There are multiple explanations for the results. On the one hand, it could be that these metrics are general prognostic indicators irrespective of treatment. This interpretation is supported by the fact that the associations were stronger for overall survival than for PFS and MSS. On the other hand, it could be that body composition influences the effect of checkpoint inhibitor treatment. A proposed mechanism is that visceral adipose tissue dysregulates the body's immune response, leading to worse treatment effects [26,27]. Future research, however, is needed to confirm this association and to determine the underlying causal mechanisms.

This study contributes to previous evidence in two important ways. First, it adds the largest cohort collected on this topic to date and thereby strengthens the conclusion of the meta-analysis by Rocuzzo et al. [8] regarding obesity. Second, it provides a more fine-grained view of body composition through the use of CT derived body composition metrics. This is particularly relevant in the case of visceral adipose tissue, where our findings suggest a negative association with survival, rather than a positive one as was suggested by earlier findings on BMI. A limitation is the exclusion of otherwise eligible patients due to unavailable data. Approximately 25% of eligible patients were excluded due to lack of required data. We argue, however, that the risk of selection bias is limited, as differences in patient characteristics between included and excluded patients were small. Furthermore, the correction of skeletal muscle density for the presence of

## Chapter 5. Body composition and checkpoint inhibitor treatment outcomes in advanced melanoma: a multicenter cohort study

---

contrast is likely to be imperfect. This correction assumes that the mean and standard deviation of the true skeletal muscle density is the same for patients who underwent contrast-enhanced and no-contrast baseline scans. This may not be the case, given the difference in patient characteristics between the two groups. Given the consistent results in the subgroup analyses, we think it is unlikely that this imperfect correction would have significantly influenced the results.

In conclusion, underweight BMI, more visceral adipose tissue and lower skeletal muscle density are associated with worse outcomes in ICI treated advanced melanoma patients, independent of known predictors. The significance of the associations in multivariable analysis indicates that the information provided by body composition metrics is not fully captured by previously identified predictors, such as ECOG performance status. Outcomes were not significantly different in overweight and obese patients, as compared with those with normal BMI. This finding is in accordance with a recent meta-analysis on this topic. Our work contributes to previous research by presenting the largest cohort to date and by providing detailed data on body composition through CT derived metrics.

## 5.6 Supplementary Materials

Supplementary materials are available through

<https://www.medrxiv.org/content/medrxiv/early/2024/03/02/2024.03.01.24303607/DC1/embed/media-1.pdf?download=true>

## 5.7 References

- [1] Korn EL, Liu P-Y, Lee SJ, Chapman J-AW, Niedzwiecki D, Suman VJ, et al. Meta-Analysis of Phase II Cooperative Group Trials in Metastatic Stage IV Melanoma to Determine Progression-Free and Overall Survival Benchmarks for Future Phase II Trials. *J Clin Oncol* 2008;26:527–34. <https://doi.org/10.1200/JCO.2007.12.7837>
- [2] Wolchok JD, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Long-Term Outcomes With Nivolumab Plus Ipilimumab or Nivolumab Alone Versus Ipilimumab in Patients With Advanced Melanoma. *J Clin Oncol* 2022;40:127–37. <https://doi.org/10.1200/jco.21.02229>
- [3] Carlino MS, Larkin J, Long GV. Immune checkpoint inhibitors in melanoma. *The Lancet* 2021;398:1002–14. [https://doi.org/10.1016/S0140-6736\(21\)01206-X](https://doi.org/10.1016/S0140-6736(21)01206-X)
- [4] Blank CU, Haanen JB, Ribas A, Schumacher TN. The “cancer immunogram.” *Science* 2016;352:658–60. <https://doi.org/10.1126/science.aaf2834>
- [5] An Y, Wu Z, Wang N, Yang Z, Li Y, Xu B, et al. Association between body mass index and survival outcomes for cancer patients treated with immune checkpoint inhibitors: a systematic review and meta-analysis. *J Transl Med* 2020;18:235. <https://doi.org/10.1186/s12967-020-02404-x>

## 5.7. References

---

- [6] Chen H, Wang D, Zhong Q, Tao Y, Zhou Y, Shi Y. Pretreatment body mass index and clinical outcomes in cancer patients following immune checkpoint inhibitors: a systematic review and meta-analysis. *Cancer Immunol Immunother* 2020;69:2413–24. <https://doi.org/10.1007/s00262-020-02680-y>
- [7] Nie R-C, Chen G-M, Wang Y, Yuan S-Q, Zhou J, Duan J-L, et al. Association Between Body Mass Index and Survival Outcomes In Patients Treated With Immune Checkpoint Inhibitors: Meta-analyses of Individual Patient Data. *J Immunother Hagerstown Md* 1997 2021;44:371–5. <https://doi.org/10.1097/CJI.0000000000000389>
- [8] Rocuzzo G, Moirano G, Fava P, Maule M, Ribero S, Quaglino P. Obesity and immune-checkpoint inhibitors in advanced melanoma: A meta-analysis of survival outcomes from clinical studies. *Semin Cancer Biol* 2023;91:27–34. <https://doi.org/10.1016/j.semcancer.2023.02.010>
- [9] Young AC, Quach HT, Song H, Davis EJ, Moslehi JJ, Ye F, et al. Impact of body composition on outcomes from anti-PD1 +/- anti-CTLA-4 treatment in melanoma. *J Immunother Cancer* 2020;8:e000821. <https://doi.org/10.1136/jitc-2020-000821>
- [10] Weston AD, Korfiatis P, Kline TL, Philbrick KA, Kostandy P, Sakinis T, et al. Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. *Radiology* 2019;290:669–79. <https://doi.org/10.1148/radiol.2018181432>
- [11] Tolonen A, Pakarinen T, Sassi A, Kyt   J, Cancino W, Rinta-Kiikka I, et al. Methodology, clinical applications, and future directions of body composition analysis using computed tomography (CT) images: A review. *Eur J Radiol* 2021;145:109943. <https://doi.org/10.1016/j.ejrad.2021.109943>
- [12] ter Maat LS, van Duin IAJ, Elias SG, van Diest PJ, Pluim JPW, Verhoeff JJC, et al. Imaging to predict checkpoint inhibitor outcomes in cancer. A systematic review. *Eur J Cancer* 2022;175:60–76. <https://doi.org/10.1016/j.ejca.2022.07.034>
- [13] Wang Z, Aguilar EG, Luna JI, Dunai C, Khuat LT, Le CT, et al. Paradoxical effects of obesity on T cell function during tumor progression and PD-1 checkpoint blockade. *Nat Med* 2019;25:141–51. <https://doi.org/10.1038/s41591-018-0221-5>
- [14] van Zeijl MCT, Haanen JBAG, Wouters MWJM, de Wreede LC, Jochems A, Aarts MJB, et al. Real-world Outcomes of First-line Anti-PD-1 Therapy for Advanced Melanoma: A Nationwide Population-based Study. *J Immunother* 2020;43:256–64. <https://doi.org/10.1097/CJI.0000000000000334>
- [15] Silva IP da, Ahmed T, McQuade JL, Nebhan CA, Park JJ, Versluis JM, et al. Clinical Models to Define Response and Survival With Anti-PD-1 Antibodies Alone or Combined With Ipilimumab in Metastatic Melanoma. *J Clin Oncol* 2022. <https://doi.org/10.1200/JCO.21.01701>
- [16] van Zeijl MCT, de Wreede LC, van den Eertwegh AJM, Wouters MWJM, Jochems A, Schouwenburg MG, et al. Survival outcomes of patients with advanced melanoma from 2013 to 2017: Results of a nationwide population-based registry. *Eur J Cancer* 2021;144:242–51. <https://doi.org/10.1016/j.ejca.2020.11.028>
- [17] van Not OJ, de Meza MM, van den Eertwegh AJM, Haanen JB, Blank CU, Aarts MJB, et al. Response to immune checkpoint inhibitors in acral melanoma: A nationwide cohort study. *Eur J Cancer* 2022;167:70–80. <https://doi.org/10.1016/j.ejca.2022.02.026>
- [18] Moeskops P, Vos B de, Veldhuis WB, Jong PA de, I  gum I, Leiner T. Automatic quantification of body composition at L3 vertebra level with convolutional neural networks. *ECR 2020 EPOS 2020*. [https://epos.myesr.org/poster/esr/ecr2020/C-09334\(accessedJune8,2023\)](https://epos.myesr.org/poster/esr/ecr2020/C-09334(accessedJune8,2023))
- [19] de Jong DJ, Veldhuis WB, Wessels FJ, de Vos B, Moeskops P, Kok M. Towards Personalised Contrast Injection: Artificial-Intelligence-Derived Body Composition and Liver Enhancement in Computed Tomography. *J Pers Med* 2021;11:159. <https://doi.org/10.3390/jpm11030159>
- [20] Van Erck D, Moeskops P, Schoufour JD, Weijs PJM, Scholte Op Reimer WJM, Van Mourik MS, et al. Evaluation of a Fully Automatic Deep Learning-Based Method for the Measurement of Psoas Muscle Area. *Front Nutr* 2022;9.
- [21] Elhakim T, Trinh K, Mansur A, Bridge C, Daye D. Role of Machine Learning-Based CT Body Composition in Risk Prediction and Prognostication: Current State and Future Directions. *Diagnostics* 2023;13:968. <https://doi.org/10.3390/diagnostics13050968>
- [22] Shachar SS, Deal AM, Weinberg M, Nyrop KA, Williams GR, Nishijima TF, et al. Skeletal Muscle Measures as Predictors of Toxicity, Hospitalization, and Survival in Patients with Metastatic Breast Cancer Receiving Taxane-Based Chemotherapy. *Clin Cancer Res* 2017;23:658–65. <https://doi.org/10.1158/1078-0432.CCR-16-0940>

## Chapter 5. Body composition and checkpoint inhibitor treatment outcomes in advanced melanoma: a multicenter cohort study

---

- [23 ] Ebadi M, Martin L, Ghosh S, Field CJ, Lehner R, Baracos VE, et al. Subcutaneous adiposity is an independent predictor of mortality in cancer patients. *Br J Cancer* 2017;117:148–55. <https://doi.org/10.1038/bjc.2017.149>
- [24 ] Martin L, Birdsell L, MacDonald N, Reiman T, Clandinin MT, McCargar LJ, et al. Cancer Cachexia in the Age of Obesity: Skeletal Muscle Depletion Is a Powerful Prognostic Factor, Independent of Body Mass Index. *J Clin Oncol* 2013;31:1539–47. <https://doi.org/10.1200/JCO.2012.45.2722>
- [25 ] Fontana L, Eagon JC, Trujillo ME, Scherer PE, Klein S. Visceral Fat Adipokine Secretion Is Associated With Systemic Inflammation in Obese Humans. *Diabetes* 2007;56:1010–3. <https://doi.org/10.2337/db06-1656>
- [26 ] Ringel AE, Drijvers JM, Baker GJ, Catozzi A, García-Cañaveras JC, Gassaway BM, et al. Obesity Shapes Metabolism in the Tumor Microenvironment to Suppress Anti-Tumor Immunity. *Cell* 2020;183:1848-1866.e26. <https://doi.org/10.1016/j.cell.2020.11.009>
- [27 ] Farag KI, Makkouk A, Norian LA. Re-Evaluating the Effects of Obesity on Cancer Immunotherapy Outcomes in Renal Cancer: What Do We Really Know? *Front Immunol* 2021;12.

## 5.7. References

---



## Chapter 6

# Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study

L.S. Ter Maat MD, M. Schuiveling MD, W.A.M. Blokx MD PhD, P.J. Van Diest MD PhD, J.P.W. Pluim PhD, K.P.M. Suijkerbuijk MD PhD, M. Veta PhD

*In preparation*

## 6.1. Abstract

---

# 6.1 Abstract

### 6.1.1 Introduction

Checkpoint inhibitors have significantly improved the prognosis for advanced melanoma patients, but these treatments are costly and associated with severe toxicity. Accurate pre-treatment predictors of response are still lacking, and histopathology material may hold potential for improving prediction accuracy.

### 6.1.2 Methods

This study utilized deep learning models to analyse H&E histopathology images of both primary tumors and metastases from patients with advanced melanoma undergoing checkpoint inhibitor treatment. The study included adult patients diagnosed with unresectable stage IIIC or IV cutaneous melanoma, treated with first-line anti-PD1 ± anti-CTLA4 therapy. The deep learning model involved splitting of histopathology images into patches, feature extraction using a pre-trained neural network, and classification using a Transformer model. In addition to predicting clinical benefit, performance was evaluated for classifying lymphocytic infiltration, a known predictor of response visible on histopathology images. Model evaluation was performed through 5-fold cross-validation and performance assessment using AUROC.

### 6.1.3 Results

A total of 716 patients from five centers were included, with 471 primary and 516 metastatic samples. For predicting clinical benefit, the model achieved an AUROC of 0.50 [0.43 – 0.55] on primary tissue samples and 0.54 [0.44 - 0.61] on metastatic tissue samples. For classifying lymphocytic infiltration, the model achieved an AUC 0.68 [0.62 - 0.74] in primary, and 0.66 [0.62 - 0.71] in metastatic samples.

### 6.1.4 Discussion

The results of this preliminary study indicate that checkpoint inhibitor treatment outcome prediction based on melanoma H&E images using deep learning is a challenging task. Further research should investigate the model’s performance for

## Chapter 6. Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study

---

classifying lymphocytic infiltration, which is lower than may be expected based on the reported interobserver variability for this metric. Furthermore, future efforts should expand the dataset, evaluate multitarget learning and experiment with specifically defining a region of interest on a sample.

### 6.2 Introduction

Checkpoint inhibitor treatment has revolutionized the treatment of advanced melanoma. Before the introduction of checkpoint inhibitor drugs, the prognosis of this group of patients was poor due to the limited efficacy of available treatment options [1]. In contrast, 6.5-year overall survival rates are as high as 49% for ipilimumab + nivolumab in trial settings [2]. The introduction of checkpoint inhibitors therefore represents one of the most significant advances in oncological care of the past decade.

This treatment works by mobilizing the immune system against tumor cells. Checkpoint inhibitor drugs in melanoma work by blocking either the cytotoxic T-lymphocyte associated protein 4 (CTLA4), or the Programmed cell Death-1 (PD1) proteins. In a healthy setting, these proteins serve to downregulate immune activity. Some tumors are able to hijack this mechanism and thereby evade immune response. Blocking the CTLA4 and PD1 pathways can thereby reactivate an immune response against the tumor [3].

While very successful, checkpoint inhibitor treatment is also associated with high costs and severe toxicity. Anti-PD1 medication was the most expensive intramural medication given in The Netherlands, with a total cost of 327 million EUR [4]. Advanced melanoma and non-small cell lung cancer make up the majority of indications of this treatment [5,6]. In addition, severe immune-related adverse events occur in 60% of patients treated with anti-PD1 + anti-CTLA4 treatment [2]. These adverse events both negatively impact quality of life in treated patients [7] and put an additional burden on the health care system [8].

Despite intensive research, accurate pretreatment predictors of response are still lacking. Finding such a predictor is a significant challenge, as is exemplified by the expression of PD1 in tumors: even when this very target of anti-PD1 therapy is absent, tumors may still respond [9]. Other efforts, investigating for example tumor mutational burden [10], radiomics [11] and gut microbiome [12] have not

## 6.2. Introduction

---

yet resulted in predictors that can accurately guide clinical decisions. Predictors that are used in daily practice are clinical characteristics, such as which organs are affected, neurological symptoms of brain metastases and ECOG performance status [13]. Although these factors can stratify patients into groups with significant differences in outcomes, further research is necessary to reach the goal of precision medicine in this setting.

Routinely collected histopathology material could be valuable for predicting checkpoint inhibitor outcomes. Histopathology material provides a detailed view of, among others, which cells are present in and around a tumor, their visual appearance, and the spatial relationship between different cell types. It may thereby provide information on several characteristics that could influence checkpoint inhibitor efficacy. Potential examples are peritumoral stroma that prevents immune response or the number of tumor-infiltrating lymphocytes in the metastases, which were recently shown to be correlated with checkpoint inhibitor treatment outcomes [14]. In addition to showing potential as a source of valuable information, histopathology material of the primary tumor would be ideal as a predictor since it is obtained in every patient to make the initial diagnosis. Predictors based on histopathology would therefore require little change to clinical care.

Methods that use histopathology material as a predictor roughly fall into two categories. The first category consists of manually defined features, such as the local abundance of a certain cell type. An example of this is the degree of presence of tumor infiltrating lymphocytes. These manually defined features can either be assessed manually, as is done with the Clark score in the case of tumor infiltrating lymphocytes [15], or automatically, using recent techniques in image analysis to give a quantitative estimate of cell abundance and localization [16]. The second category of predictors consists of end-to-end deep learning models. In contrast to the first category, these models are not based on manually predefined features. Instead, informative features are learned based on the provided data. The advantage of these models is therefore that they are much more flexible in what information can be extracted. This flexibility, however, also poses a major challenge. This is because histopathology images contain an enormous amount of information. A successful model must therefore be able to find the useful signal without overfitting to the abundant noise. A dataset of sufficient size could mitigate this problem, but collecting such a dataset is often not feasible. Deep learning mod-

## **Chapter 6. Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study**

---

els for histopathology material therefore rely heavily on dimensionality reduction through feature extraction [17]: a collection of techniques to summarize data while retaining as much useful information as possible.

Limited evidence exists on the value of deep learning on histopathology material for predicting checkpoint inhibitor outcomes in advanced melanoma. Johannet et al. [18] report on a deep learning model trained on 121 patients treated with anti-CTLA4, anti-PD1, or both. Their model was able to stratify patients into a high- or low-risk group, with significantly better progression-free survival (PFS) in the low-risk group when evaluated in an independent cohort of 30 patients. Hu et al. [19] obtained an area under the receiver operator characteristic curve (AUROC) of 0.778 for predicting anti-PD1 response in an independent test set of 54 melanoma patients. These findings are promising, but the small sample size of these studies precludes strong conclusions. Furthermore, the methods used in these works have been superseded by recent developments in deep learning in histopathology [20,21].

The present study aims to provide high-quality evidence on the predictive value of an end-to-end deep learning approach on histopathology. For this purpose, we have collected the largest dataset to date of histopathology material of patients treated with checkpoint inhibitors for advanced melanoma. Furthermore, our method incorporates state-of-the-art techniques in histopathology image classification [21].

## **6.3 Methods**

### **6.3.1 Patient selection**

Eligible patients were selected based on the following inclusion criteria: (i) over 18 years of age, (ii) diagnosed with unresectable stage IIIC or stage IV cutaneous melanoma, (iii) treated with first-line anti-PD1 with or without anti-CTLA4 therapy and (iv) a start date of therapy after January 1st, 2016. Patients were excluded in the case of unavailability of both a primary and metastatic melanoma tissue sample. Eligible patients were retrospectively selected through high-quality registry data from eight participating centers (Amphia Breda, Isala Zwolle, Leiden University Medical Center, Maxima Medical Center, Medisch Spectrum Twente, Radboudumc, University Medical Center Groningen, Amsterdam University Med-

### 6.3. Methods

---

ical Center). The Medical Ethics Committee deemed this study not subject to the Medical Research Involving Human Subjects Act according to Dutch regulations; informed consent was waived.

#### 6.3.2 Data collection

Hematoxylin and eosin-stained slides of primary and metastatic tumors were obtained for all included patients through the Dutch National Tissue Archive Portal, and subsequently scanned with a Nanozoomer XR C12000-21/-22 (Hamamatsu Photonics, Hamamatsu, Shizuoka, Japan) at 40 $\times$  magnification and a resolution of 0.22 micrometers per pixel. A single representative primary and metastatic sample was selected for every patient. For the primary slides, we selected the most likely primary tumor based on Breslow thickness and location relative to positive lymph nodes. For the metastases, the most recent sample before treatment initiation was selected.

#### 6.3.3 Outcome definition

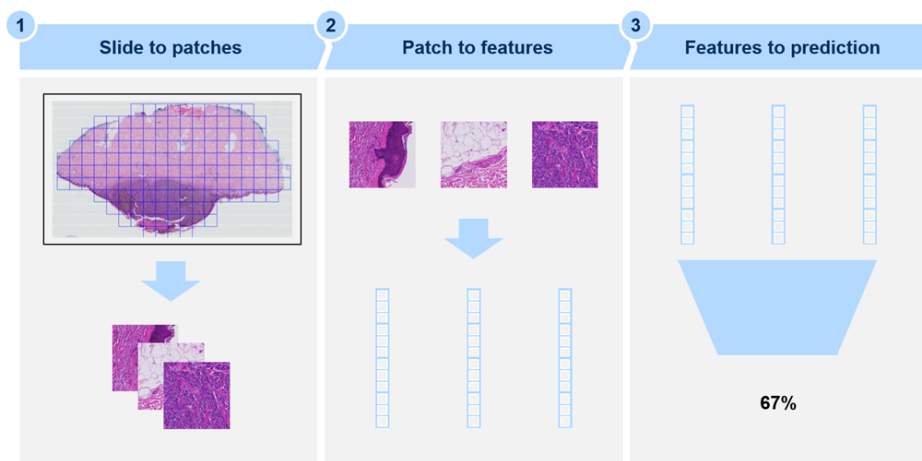
The primary outcome was clinical benefit, which was defined as stable disease (per RECIST 1.1 [22]) for a minimum of six months, or partial or complete response at any point during follow-up. The secondary outcome was objective response, defined as a partial or complete response at any point during follow-up. Several other prediction targets were explored to evaluate the performance of the deep learning methodology. These included (i) distinguishing primary from metastatic samples, (ii) classifying BRAF mutation status (wild-type versus any BRAF mutation) and (iii) classifying infiltration with TILs (absent vs. non-brisk/brisk as per the Clark score [15]). The performance of the model for BRAF mutation status was contrasted with that of known clinicopathological characteristics correlated with BRAF mutation status that are discernable on H&E material, namely location, subtype, ulceration and solar elastosis [23–25].

#### 6.3.4 Deep learning model

As stated above, one of the most important challenges of deep learning-based classification of histopathology material is the abundance of information from which

## Chapter 6. Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study

the useful signal must be filtered. In addition, there are the technical challenges of the large size of the images (with uncompressed sizes of more than 40 gigabytes per whole-slide image), which is also highly variable between images. Our method addresses these challenges in three steps (Figure 6.1). In the first step, the foreground of the image is segmented and divided into square, non-overlapping patches. The foreground of the image is defined as the part of the image that contains tissue; a segmentation of this foreground is obtained through traditional image processing techniques (Supplementary Methods). Through this step, the image is divided into  $N$  patches; the number  $N$  will vary between slides. Second, a pretrained neural network is used to extract features for every patch. The aim of this step is to summarize large images as a much smaller feature vector that retains the information present in the patches. For every slide, this step yields an output of  $N$  feature vectors. Third, a deep learning classifier is trained to predict the probability of clinical benefit based on the pre-extracted feature vectors belonging to a single slide. Training was terminated after 100 epochs, or when no improvement was observed in 10 successive epochs.



**Figure 6.1:** Graphical overview of methodology. Treatment outcome predictions for every histopathology whole slide image are generated in three steps. In step 1, the whole-slide image is split into non-overlapping square patches. In step 2, every patch is summarized as a feature vector by a pretrained feature extractor. In step 3, the features of all patches of a single slide are combined to make a single prediction, in this case the probability of clinical benefit from treatment.

## 6.4. Methods

---

### 6.3.5 Hyperparameter exploration

Hyperparameters are arbitrary design choices that may influence model performance. Relevant hyperparameters were systematically explored with a fixed train-test split to arrive at an optimal model. These included the choice of feature extractor, level of magnification during feature extraction and classifier. For the feature extractor, four models were considered: (i) a ResNet-50 model pretrained on the ImageNet dataset; (ii) PLIP, a Vision Transformer pretrained using contrastive learning on pathology images collected through crowd platforms such as medical Twitter [26]; (iii) patch-level and (iv) region-level HIPT [21], the smaller and larger scale models, respectively, of a hierarchical Vision Transformer model pretrained using the self-supervised DINO methodology [27] on a large dataset of publicly available histopathology images [28]. Three levels of magnification were explored, namely 20x, 5x and 1.5x; for the HIPT feature extractors, only the 20x level was used as the model was pretrained at this level. For the classifier, two models were investigated: (i) an attention pooling classifier [29] and (ii) a Transformer model [30]. Binary cross entropy loss was used in combination with the Adam optimizer with a fixed learning rate of  $2e-4$ ; weight decay and dropout were set to  $1e-5$  and 0.25 respectively. Training was terminated after 100 epochs, or when no improvement was observed in 10 successive epochs.

### 6.3.6 Model evaluation

The deep learning model was evaluated using a 5-fold cross validation. In every fold, the available data was divided into a non-overlapping training, validation and test set (60%, 20% and 20%, respectively). Test sets did not overlap between folds. The model was fitted on the training set. The tuning set was used to monitor training performance, terminate training if results did not improve for 10 epochs and select the best model. This model was subsequently evaluated on the test set. Performance was evaluated using the receiver operator characteristics (ROC) curve and corresponding area under the curve (AUROC). The cvAUC R-package was used to calculate 95% confidence intervals for the mean AUROC across folds [31]. Separate models were trained for predicting outcomes on primary and metastatic samples.



## 6.4 Results

### 6.4.1 Patient characteristics

A total of 1944 eligible patients were identified, of which 716 patients (37%) were included. In these 716 patients, a total of 471 primary and 516 metastatic samples were available. Characteristics of included patients are shown in Table 6.1. The proportion of patients with clinical benefit and objective response was 62% and 52%, respectively.

		Overall	Primary available	Metastasis available
n		716	471	516
Age, mean (SD)		65.7 (12.5)	66.1 (12.5)	64.7 (12.0)
Sex, n (%)	Female	251 (35.1)	164 (34.8)	171 (33.1)
	Male	465 (64.9)	307 (65.2)	345 (66.9)
Therapy, n (%)	Anti-PD1	448 (62.6)	305 (64.8)	308 (59.7)
	Ipilimumab & Nivolumab	268 (37.4)	166 (35.2)	208 (40.3)
Stage, n (%)	IIIC	53 (7.4)	33 (7.0)	36 (7.0)
	IV M1a	55 (7.7)	32 (6.8)	39 (7.6)
	IV M1b	96 (13.4)	69 (14.6)	67 (13.0)
	IV M1c	339 (47.3)	235 (49.9)	235 (45.5)
	IV M1d	168 (23.5)	99 (21.0)	136 (26.4)
	missing	5 (0.7)	3 (0.6)	3 (0.6)
ECOG performance status, n (%)	0	349 (48.7)	230 (48.8)	248 (48.1)
	1	291 (40.6)	195 (41.4)	217 (42.1)
	2-4	52 (7.3)	35 (7.4)	35 (6.8)
	missing	24 (3.4)	11 (2.3)	16 (3.1)
Brain metastases, n (%)	absent	466 (65.1)	323 (68.6)	321 (62.2)
	asymptomatic	100 (14.0)	64 (13.6)	78 (15.1)
	symptomatic	68 (9.5)	35 (7.4)	58 (11.2)
	missing	82 (11.5)	49 (10.4)	59 (11.4)
Liver metastases, n (%)	absent	446 (62.3)	287 (60.9)	322 (62.4)
	present	210 (29.3)	146 (31.0)	153 (29.7)
	missing	60 (8.4)	38 (8.1)	41 (7.9)
LDH, n (%)	normal	462 (64.5)	308 (65.4)	325 (63.0)
	1-2x ULN	190 (26.5)	119 (25.3)	143 (27.7)
	>2x ULN	54 (7.5)	39 (8.3)	39 (7.6)
	missing	10 (1.4)	5 (1.1)	9 (1.7)
Number of affected organs, n (%)	<3	384 (53.6)	251 (53.3)	278 (53.9)
	>2	332 (46.4)	220 (46.7)	238 (46.1)
Clinical benefit, n (%)	no	270 (37.7)	178 (37.8)	198 (38.4)
	yes	446 (62.3)	293 (62.2)	318 (61.6)
Objective response, n (%)	no	323 (45.1)	214 (45.4)	235 (45.5)
	yes	393 (54.9)	257 (54.6)	281 (54.5)

**Table 6.1:** Characteristics of included patients. Abbreviations: SD=standard deviation, ECOG=Eastern Cooperative Oncology Group, LDH=lactate dehydrogenase, ULN=upper limit of normal

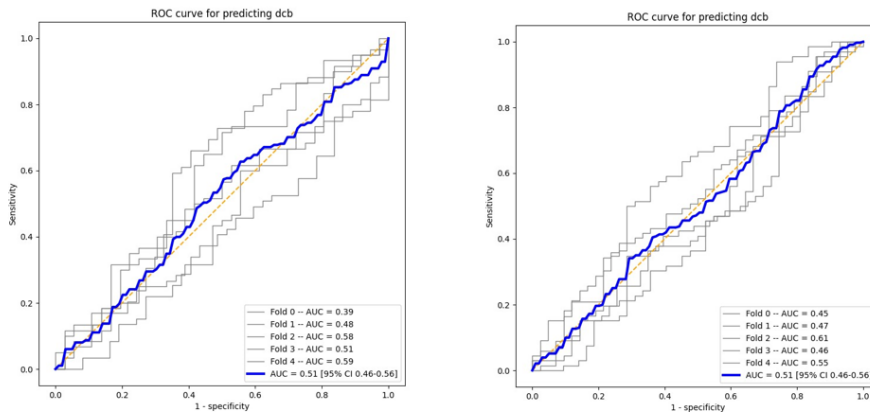
## 6.4. Results

### 6.4.2 Hyperparameter tuning

Initial experiments showed small differences in performance for different feature extractors, magnification levels and classifiers. Overall, the PLIP and HIPT models outperformed the ImageNet feature-extractor for most prediction targets. Based on these results, the region-level HIPT feature extractor at 20x magnification was selected in combination with a Transformer classifier.

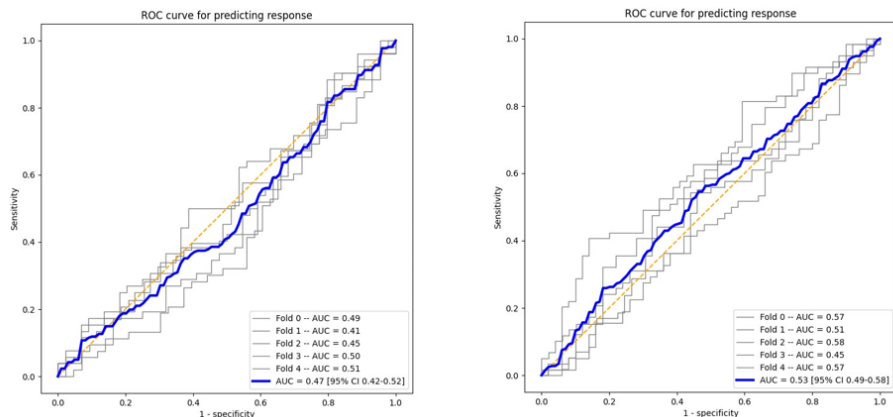
### 6.4.3 Treatment outcome prediction

For predicting clinical benefit based on primary tissue samples, the model achieved an AUROC of 0.51 [0.46 – 0.56] (Figure 6.2). For predicting objective response, the AUROC was 0.47 [0.42-0.52] (Figure 6.3). The deep learning model based on metastatic tissue samples reached an AUROC of 0.51 [0.46 - 0.56] for predicting clinical benefit, and an AUROC of 0.53 [0.49-0.58]. Heatmaps showing model attention are provided in the Supplementary Figures 1. No clear patterns were observed in the patches that were given the most attention.



**Figure 6.2:** Receiver operator characteristic curve for predicting clinical benefit using primary (left) and metastatic (right) samples

## Chapter 6. Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study



**Figure 6.3:** Receiver operator characteristic curve for predicting objective response using primary (left) and metastatic (right) samples

### 6.4.4 Other prediction targets

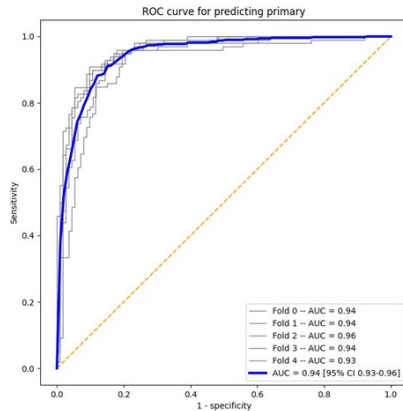
For distinguishing primary from metastatic samples, the model achieved an AU-ROC of 0.94 [0.93-0.96] (Figure 6.4). Examples of cases which were misclassified by the model are shown in Supplementary Figures 2; most cases were either primary samples with little to no epidermis visible, or samples of cutaneous metastases. The model achieved an AUROC of 0.63 [0.58-0.68] for classifying BRAF mutation status in primary samples. By comparison, a logistic regression based on primary subtype, location and ulceration reached an AUROC of 0.61 [0.56-0.66] (Figure 6.5, where sensitivity and specificity of solar elastosis are also shown). For predicting lymphocytic infiltration, the model reached an AUROC of 0.68 [0.62-0.74] and 0.66 [0.62-0.71] in primary and metastatic samples, respectively (Figure 6.5).

## 6.5 Discussion

The results of this preliminary study indicate that predicting checkpoint inhibitor treatment outcomes using deep learning on H&E histopathology images is a challenging task. We did not reach a performance significantly better than random using either primary or metastatic samples, despite the use of methodology which is both diverse and proven to reach state-of-the-art results on histopathology clas-

## 6.5. Discussion

---



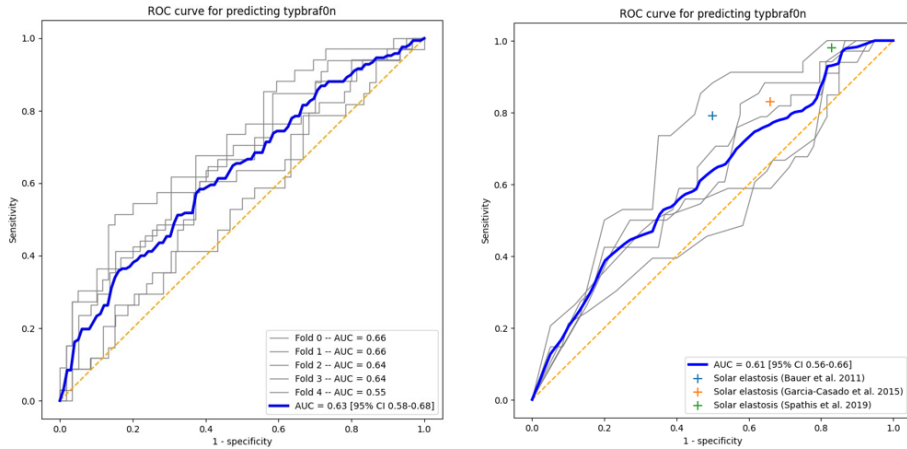
**Figure 6.4:** Receiver operator characteristic curve for distinguishing primary from metastatic samples

sification, and the largest dataset for this purpose to date. Several open questions, however, remain to be answered.

Such an open issue is the random performance of the model, although these images are known to hold information that is associated with checkpoint inhibitor treatment outcomes. For one, lymphocytic infiltration in metastases is discernible from H&E images and was previously shown to be positively associated with better response. In addition, BRAF mutation status is associated with better response in patients receiving anti-PD1 + anti-CTLA4 combination therapy [32]. Histopathological characteristics associated with BRAF mutation status should therefore hold some predictive power. A potential explanation for why this information is not learned by the model is that it must be picked up among the vast number of features available. This is made even more difficult since the association of both lymphocytic infiltration and BRAF mutation status with treatment outcomes is modest [14,32].

Furthermore, the use of pretrained feature extractors raises the question whether all useful information is captured. With current hardware capabilities, simultaneous training of feature extractor and classifier is practically infeasible due to the enormous memory requirements. Feature extraction is therefore a necessary step, but this will always result in the loss of some information. The feature ex-

## Chapter 6. Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study



**Figure 6.5:** Receiver operator characteristic curve for predicting BRAF mutation status in primary samples through the end-to-end deep learning model (left) and based on known clinicopathological characteristics (right)

tractors used in this work are trained using state-of-the-art methodologies that are designed to maximize the retention of relevant information, as evidenced by their high performance for on other classification datasets [27,33]. This does not guarantee, however, that all relevant information is retained. An open question is therefore how results may change with simultaneous training of feature extractor and classifier, or with the development of new method for feature extractor pretraining.

In addition, our method does not approach human performance for all evaluated tasks. For distinguishing primary from metastatic samples, performance is as expected: a high AUC of 0.94 with misclassifications in metastatic metastases, or primary cases where epidermis is absent. Furthermore, the performance of the deep learning model is comparable to that of clinicopathological characteristics for classifying BRAF mutation status. However, the performance for classifying degree of lymphocytic infiltration is significantly lower than may be expected based on the interobserver variability for human observers [34]. This is an important observation, especially since lymphocytic infiltration in metastases is the pathological characteristic that is most strongly associated with checkpoint inhibitor treatment response [14]. The limitations of the model in reaching the expected

## 6.5. Discussion

---

performance raises the question what other potentially informative features are missed.

To address these open questions, three research directions could be considered for further exploration. First, multitarget learning should be explored. This technique adds one or several auxiliary prediction targets alongside the target of interest. Since model weights are shared among these prediction targets, this enables the model to extract features for an easier target, which may be subsequently leveraged for the more difficult target. Clearly, the auxiliary targets should be informative for the prediction task of interest. In this case, these targets could be BRAF mutations status and degree of lymphocytic infiltration.

Second, the lower performance of the model for classifying lymphocytic infiltration should be investigated. This can be approached in several ways. For one, it would be insightful to show if the utilized feature extractors encode information that is associated with the number of TILs in a single patch. Furthermore, evaluating model performance on similar tasks in other datasets could provide an indication as to whether this diminished performance is specific for this combination of task and dataset, or if this result is consistent across datasets. Pursuing this direction would also be valuable for the field of computational pathology in general, as it would give insight into the kind of information that is available through pretrained feature extractors.

Third, the use of more targeted regions of interest, as opposed to the use of all foreground patches, can be explored. As the foreground includes all the material visible on the slide, this means that information from healthy or damaged tissue parts is also available to the model. These tissue parts are unlikely to contain information that is associated with response to checkpoint inhibitor treatment, and therefore probably only dilute the signal that may be present. This can be addressed by segmenting the areas which are likely to be the most informative, such as the area of vital tumor tissue, or the interface between tumor and host tissue.

In conclusion, this preliminary study describes a deep learning model based on H&E histopathology images, trained to predict checkpoint inhibitor outcomes. In both primary and metastatic samples, the model did not perform significantly better than random. Further work remains to be done, namely (i) further expansion of the dataset, (ii) use of multitarget learning, (iii) investigation of the

## Chapter 6. Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study

---

reason for the relatively poor performance for classifying lymphocytic infiltration of the presented method and (iv) use of more specific regions of interest. For now, the present work indicates that predicting checkpoint inhibitor outcomes based on H&E images is a challenging task.

### 6.6 Supplementary Materials

Supplementary Materials are available through:

<https://drive.google.com/file/d/1rVY93rArrNi61JggOfDoNPN3Wp4RFkcN/view?usp=sharing>

### 6.7 References

- [1] Korn EL, Liu P-Y, Lee SJ, Chapman J-AW, Niedzwiecki D, Suman VJ, et al. Meta-Analysis of Phase II Cooperative Group Trials in Metastatic Stage IV Melanoma to Determine Progression-Free and Overall Survival Benchmarks for Future Phase II Trials. *J Clin Oncol* 2008;26:527–34. <https://doi.org/10.1200/JCO.2007.12.7837>
- [2] Wolchok JD, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Long-Term Outcomes With Nivolumab Plus Ipilimumab or Nivolumab Alone Versus Ipilimumab in Patients With Advanced Melanoma. *J Clin Oncol* 2022;40:127–37. <https://doi.org/10.1200/jco.21.02229>
- [3] Sharma P, Allison JP. Dissecting the mechanisms of immune checkpoint therapy. *Nat Rev Immunol* 2020;20:75–6. <https://doi.org/10.1038/s41577-020-0275-8>
- [4] Zaken M van A. Bijlage 2: Uitgaven per geneesmiddel - Publicatie - Rijksoverheid.nl 2023. <https://www.rijksoverheid.nl/documenten/publicaties/2023/03/28/bijlage-2-uitgaven-2021-per-geneesmiddel> (accessed October 20, 2023)
- [5] Nederland Z. Horizonscan geneesmiddelen - Pembrolizumab n.d. <https://www.horizonscangeneesmiddelen.nl/geneesmiddelen/pembrolizumab-oncologie-en-hematologie-longkanker/versie1?lang=nl> (accessed October 20, 2023)
- [6] Nederland Z. Pembrolizumab n.d. <https://www.horizonscangeneesmiddelen.nl/geneesmiddelen/pembrolizumab-oncologie-huidkanker/versie4> (accessed January 23, 2024)
- [7] Schulz TU, Zierold S, Sachse MM, Pesch G, Tomsitz D, Schilbach K, et al. Persistent immune-related adverse events after cessation of checkpoint inhibitor therapy: Prevalence and impact on patients' health-related quality of life. *Eur J Cancer* 2022;176:88–99. <https://doi.org/10.1016/j.ejca.2022.08.029>
- [8] George S, Bell EJ, Zheng Y, Kim R, White J, Devgan G, et al. The Impact of Adverse Events on Health Care Resource Utilization, Costs, and Mortality Among Patients Treated with Immune Checkpoint Inhibitors. *The Oncologist* 2021;26:e1205–15. <https://doi.org/10.1002/onco.13812>
- [9] Morrison C, Pabla S, Conroy JM, Nesline MK, Glenn ST, Dressman D, et al. Predicting response to checkpoint inhibitors in melanoma beyond PD-L1 and mutational burden. *J Immunother Cancer* 2018;6:32. <https://doi.org/10.1186/s40425-018-0344-8>
- [10] Ning B, Liu Y, Wang M, Li Y, Xu T, Wei Y. The Predictive Value of Tumor Mutation Burden on Clinical Efficacy of Immune Checkpoint Inhibitors in Melanoma: A Systematic Review and Meta-Analysis. *Front Pharmacol* 2022;13.
- [11] ter Maat LS, van Duin IAJ, Elias SG, Leiner T, Verhoeff JJC, Arntz ERAN, et al. CT radiomics compared to a clinical model for predicting checkpoint inhibitor treatment outcomes in patients with advanced melanoma. *Eur J Cancer* 2023;185:167–77. <https://doi.org/10.1016/j.ejca.2023.02.017>
- [12] Zitvogel L, Ma Y, Raouf D, Kroemer G, Gajewski TF. The microbiome in cancer immunotherapy: Diagnostic tools and therapeutic strategies. *Science* 2018;359:1366–70. <https://doi.org/10.1126/science.aar6918>

## 6.7. References

---

- [13] Silva IP da, Ahmed T, McQuade JL, Nebhan CA, Park JJ, Versluis JM, et al. Clinical Models to Define Response and Survival With Anti-PD-1 Antibodies Alone or Combined With Ipilimumab in Metastatic Melanoma. *J Clin Oncol* 2022. <https://doi.org/10.1200/JCO.21.01701>
- [14] Duin IAJ van, Schuiveling M, Maat LS ter, Amsterdam WAC van, Berkmortel F van den, Boers-Sonderen M, et al. Baseline tumor-infiltrating lymphocyte patterns and response to immune checkpoint inhibition in metastatic cutaneous melanoma 2023:2023.11.27.23299053. <https://doi.org/10.1101/2023.11.27.23299053>
- [15] Clark WH, Elder DE, Guerry D, Braitman LE, Trock BJ, Schultz D, et al. Model predicting survival in stage I melanoma based on tumor progression. *J Natl Cancer Inst* 1989;81:1893–904. <https://doi.org/10.1093/jnci/81.24.1893>
- [16] Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images - ScienceDirect n.d. <https://tinyurl.com/3vnh7j7u>(accessedJanuary23,2024)
- [17] Cascianelli S, Bello-Cerezo R, Bianconi F, Fravolini ML, Belal M, Palumbo B, et al. Dimensionality Reduction Strategies for CNN-Based Classification of Histopathological Images. In: De Pietro G, Gallo L, Howlett RJ, Jain LC, editors. *Intell. Interact. Syst. Serv.* 2017, Cham: Springer International Publishing; 2018, p21–30. [https://doi.org/10.1007/978-3-319-59480-4\\_3](https://doi.org/10.1007/978-3-319-59480-4_3)
- [18] Johannet P, Coudray N, Donnelly DM, Jour G, Ila-Bochaca I, Xia Y, et al. Using Machine Learning Algorithms to Predict Immunotherapy Response in Patients with Advanced Melanoma. *Chin Cancer Res Off J Am Assoc Cancer Res* 2021;27:131–40. <https://doi.org/10.1158/1078-0432.CCR-20-2415>
- [19] Hu J, Cui C, Yang W, Huang L, Yu R, Liu S, et al. Using deep learning to predict anti-PD-1 response in melanoma and lung cancer patients from histopathology images. *Transl Oncol* 2021;14:100921. <https://doi.org/10.1016/j.tranon.2020.100921>
- [20] Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021;5:555–70. <https://doi.org/10.1038/s41551-020-00682-w>
- [21] Chen RJ, Chen C, Li Y, Chen TY, Trister AD, Krishnan RG, et al. Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning, 2022, p. 16144–55.
- [22] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47. <https://doi.org/10.1016/j.ejca.2008.10.026>
- [23] García-Casado Z, Traves V, Bañuls J, Niveiro M, Gimeno-Carpio E, Jimenez-Sanchez AI, et al. BRAF, NRAS and MC1R status in a prospective series of primary cutaneous melanoma. *Br J Dermatol* 2015;172:1128–31. <https://doi.org/10.1111/bjd.13521>
- [24] Bauer J, Büttner P, Murali R, Okamoto I, Kolaitis NA, Landi MT, et al. BRAF mutations in cutaneous melanoma are independently associated with age, anatomic site of the primary tumor, and the degree of solar elastosis at the primary tumor site. *Pigment Cell Melanoma Res* 2011;24:345–51. <https://doi.org/10.1111/j.1755-148X.2011.00837.x>
- [25] Spathis A, Katoulis AC, Damaskou V, Liakou AI, Kottaridi C, Leventakou D, et al. BRAF Mutation Status in Primary, Recurrent, and Metastatic Malignant Melanoma and Its Relation to Histopathological Parameters. *Dermatol Pract Concept* 2019;9:54–62. <https://doi.org/10.5826/dpc.0901a13>
- [26] Huang Z, Bianchi F, Yuksekgonul M, Montine TJ, Zou J. A visual–language foundation model for pathology image analysis using medical Twitter. *Nat Med* 2023;29:2307–16. <https://doi.org/10.1038/s41591-023-02504-3>
- [27] Caron M, Touvron H, Misra I, Jégou H, Mairal J, Bojanowski P, et al. Emerging Properties in Self-Supervised Vision Transformers 2021. <https://doi.org/10.48550/arXiv.2104.14294>
- [28] Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge n.d. <https://www.termedia.pl/Review-The-Cancer-Genome-Atlas-TCGA-an-immeasurable-source-of-knowledge,77,24047,0,1.html>(accessedFebruary12,2024)
- [29] Ilse M, Tomczak J, Welling M. Attention-based Deep Multiple Instance Learning. *Proc. 35th Int. Conf. Mach. Learn., PMLR*; 2018, p. 2127–36.
- [30] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. *Adv. Neural Inf. Process. Syst.*, vol. 30, Curran Associates, Inc.; 2017.



## Chapter 6. Deep learning on histopathology to predict checkpoint inhibitor outcomes in advanced melanoma: a preliminary study

---

- [31 ] LeDell E, Petersen M, van der Laan M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron J Stat* 2015;9:1583–607. <https://doi.org/10.1214/15-EJS1035>
- [32 ] van Not OJ, Blokx WAM, van den Eertwegh AJM, de Meza MM, Haanen JB, Blank CU, et al. BRAF and NRAS Mutation Status and Response to Checkpoint Inhibition in Advanced Melanoma. *JCO Precis Oncol* 2022:e2200018. <https://doi.org/10.1200/P0.22.00018>
- [33 ] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning Transferable Visual Models From Natural Language Supervision 2021. <https://doi.org/10.48550/arXiv.2103.00020>
- [34 ] Busam KJ, Antonescu CR, Marghoob AA, Nehal KS, Sachs DL, Shia J, et al. Histologic Classification of Tumor-Infiltrating Lymphocytes in Primary Cutaneous Malignant Melanoma: A Study of Interobserver Agreement. *Am J Clin Pathol* 2001;115:856–60. <https://doi.org/10.1309/G6EK-Y6EH-OLGY-6D6F>

## 6.7. References

---

# Chapter 7

## General discussion

### 7.1 Overview

The main research question of this thesis is as follows: what is the added value of CT- and histopathology-based predictors over known clinical predictors for checkpoint inhibitor outcomes in advanced melanoma?

As discussed in chapter 1, the motivation for finding accurate predictors in this treatment setting is to prevent unnecessary toxicity and high costs. A predictor must have a high negative predictive value for it to influence clinical decision making, since the potential benefits from checkpoint inhibitor therapy are very large. Although some clinical predictors have been identified, further research is needed to improve the accuracy of available models.

This thesis therefore explores predictors derived from pretreatment CT scans and histopathology material, which were categorized as follows:

1. Readily obtainable characteristics, such as tumor burden, location and avidity on FDG-PET imaging.
2. Machine learning models based on radiomics features of lesions on CT imaging.
3. Deep learning models that use the raw volume of lesions on CT imaging as input.

## 7.2. Summary of findings

---

4. Body composition metrics, such as skeletal muscle density or the amount of visceral adipose tissue.
5. Deep learning models based on H&E histopathology images of both primary and metastatic samples.

This chapter will discuss for every category of predictors the most important findings. Subsequently, it will put these findings into context and outline the limitations, main conclusions and recommendations for clinical practice and future research.

## 7.2 Summary of findings

When considering image-derived predictors, higher tumor burden, and presence of liver and symptomatic brain metastases were shown to be associated with worse outcomes. The studies reviewed in chapter 2 consistently supported these predictors. Of these predictors, the presence of brain and liver metastases has been integrated into a multivariable model by Da Silva et al., demonstrating the added value of these predictors [1]. However, with an AUROC of 0.67, their model leaves room for further improvement. For tumor burden, the added value over known predictors is less well established: serum lactate dehydrogenase (LDH) was previously shown to be correlated with tumor burden [2], and since the included studies did not evaluate tumor burden alongside LDH, the added value of tumor burden as assessed on radiological imaging remains to be determined.

The added prognostic value of markers derived from 18F-FDG PET/CT imaging is not well supported by previously published studies. Reviewed studies that reported on standardized uptake value, which reflects tumor metabolism, showed conflicting findings. Results on metabolic tumor volume were in line with those on tumor burden: higher volume was associated with worse outcomes, but the added value over metrics such as LDH was not investigated. Lastly, the combined results on total lesion glycolysis are not conclusive on the prognostic value of this marker.

Although a radiomics model reached a significant predictive performance, adding this model to previously identified clinical predictors did not yield improvement. The radiomics model presented in chapter 3 reached a modest AUROC of 0.61. Although statistically significant, this performance is considerably lower than the

AUROC of 0.77 reported in earlier work [3]. Furthermore, chapter 3 shows that the information learned by the radiomics model overlaps with known predictors, namely presence of liver metastases and tumor burden. This can explain why combining the radiomics model with clinical predictors did not improve over a model of clinical predictors alone.

A deep learning approach also reached a significant predictive performance, but did not improve over traditional radiomics for predicting outcomes. The hypothesis that the added flexibility of a deep learning model can improve performance for this purpose, must therefore be rejected. As shown in chapter 4, both models reached very similar accuracies, and the same relation with known predictors was observed: the output of the deep learning model correlated with the presence of liver metastases and overall tumor burden, and adding the output to a clinical model did not improve results.

Lower skeletal muscle density and higher visceral adipose tissue index are associated with worse overall survival. For skeletal muscle density, the findings of the study presented in chapter 5 and those of previous works agree. Regarding visceral adipose tissue index, the results of previous works varied. Chapter 5, however, showed a clear correlation between larger amounts of visceral adipose tissue and worse outcomes. These associations were independent of previously identified predictors. For the purpose of predicting individual patient outcomes, however, it must be noted that the absolute effect sizes are modest. Although body composition metrics may improve clinical models, it is therefore unlikely that these metrics alone will increase performance enough to significantly impact decision making regarding treatment.

Obesity, when measured as BMI, was not found to be associated with outcomes. The findings presented in chapter 5 therefore argue against the obesity paradox, which has been postulated by others, hypothesizing better outcomes in obese patients. In fact, the observed association between more visceral adipose tissue and worse outcomes even suggests an opposite effect. The results are in line with those of the two previous largest studies which investigated the association between BMI and checkpoint inhibitor treatment outcomes in advanced melanoma.

A deep learning model trained on whole-slide HE histopathology samples did not achieve a significant predictive value (chapter 6). This was true for both

### 7.3. Strengths and limitations

---

primary and metastatic samples. These results should be regarded as preliminary, as improvements to the method can be made and additional data can be added. Furthermore, a previous study showed that the degree of lymphocytic infiltration, manually scored on HE images, is predictive of checkpoint inhibitor outcomes [4]. Results may therefore improve, and a deep learning model based on HE samples may prove to add to existing predictors. The current negative results, however, suggest that predicting treatment outcomes based on this data is a challenging task.

In short, several predictors that add to existing ones were identified, but these are not yet accurate enough to definitively guide treatment decisions. From the predictors investigated in this thesis, most predictive value is added by the presence of liver and symptomatic brain metastases and, to a lesser extent, by skeletal muscle density and visceral adipose tissue index. Even if the association between these factors and outcomes reflects a general prognostic relation rather than one specific to checkpoint inhibitor treatment, they may nonetheless be useful in informing clinical decisions. In order to truly reduce unnecessary costs and toxicity, however, further improvements are necessary.

### 7.3 Strengths and limitations

The research presented in this thesis has four key strengths. First, every study included in this thesis has a substantial sample size. In fact, the collected datasets are all significantly larger than those of previous works. This is especially important for the research on machine and deep learning methods: as these methods are very flexible in the relationships that they can learn, a large dataset is vital to ensure that this flexibility is leveraged. Second, the presented data is collected from multiple centers. Although this may have lead to a more heterogeneous dataset and therefore more difficult prediction task, this provides a more realistic of performance when implemented in practice. Third, the machine and deep learning models of chapters 3, 4 and 6 are all evaluated using a cross validation method. This method of evaluation is much less likely to produce outlier results than a fixed train-test split, the method that is used in most prior works. Fourth, all tested predictors are evaluated alongside known predictors. This is perhaps the most important strength of the presented research, since this is what makes it possible

to draw conclusions about their real value in clinical practice.

The first limitation is that results may change with even more data. This is an important discussion point since the performance of machine and deep learning models is typically better in larger datasets. An extreme case of this is ChatGPT, where the massive upscaling of dataset size and training resources had led to unprecedented performance in natural language processing [5]. Could a similar brute-force approach also work for predicting checkpoint inhibitor outcomes? This question cannot be answered based on the research in this thesis, but three things should be noted here. The first is that in the case of ChatGPT, the model has the information necessary to perform its task. We know this, because humans can also perform the this “chat-bot” task. In contrast, human experts cannot predict response based on baseline CT scans or histopathology material. This indicates that the information available in these sources of data is likely to be a limiting factor for performance. Second, the models in chapter 3 and 4 learn only the information that is already present in the clinical models, but nothing more. This suggests that any additional information is at least much harder to extract than what is currently learned. Third, collecting the data necessary for a brute-force approach will require extensive resources. These costs must be carefully weighed against the expected benefits, especially given the limited predictive value of the presented models thus far.

The second limitation is that results may change with more advanced modelling techniques. This limitation is important for the same reasons as the previous: better, yet to be discovered models could potentially improve results and change the negative conclusions of chapters 3, 4 and 6 to positive ones. Again, the research in this thesis cannot exclude this possibility. And in fact, incremental improvements in the field of image classification are seen every year. It is important to note, however, that these improvements seem to be subject to the law of diminishing returns: the introduction of convolutional neural networks to image classification was a major paradigm shift [6], whereas subsequent developments were iterative refinements [7]. Furthermore, recent developments are becoming increasingly dependent on large datasets. Illustrative in this respect is the introduction of Vision Transformers: although Transformers formed a paradigm shift in natural language processing alike that of CNNs in computer vision, their application to computer vision tasks resulted in modest gains only in large datasets [8]. In summary, there

## 7.4. Interpretation

---

is a possibility that new techniques may significantly change results, but this would be unexpected based on the current trend of improvements.

Furthermore, the retrospective, observational and uncontrolled design of the presented research must be acknowledged. This has two important implications for the interpretation of the results. The first is that included patients were those for which the decision to give treatment was already made. This makes the studied population slightly different from the population of interest, namely patients in which the decision to give checkpoint inhibitor treatment is yet to be made. For example, clinicians may advise against treatment in patients with very poor prognosis, potentially leading to a lower proportion of these patients being included. Reported findings are therefore not guaranteed to generalize, and further prospective validation is required to investigate this. The second is that it is not possible to distinguish between general prognostic factors and predictors specific to checkpoint inhibitor efficacy. Making this distinction would require a control group of patients not treated with checkpoint inhibitors, which would not be ethical to attain. This means that the predictors investigated in this thesis cannot be used to decide between treatments; they can only be used to identify patients with a poor prognosis regardless of treatment.

## 7.4 Interpretation

What makes prediction of checkpoint inhibitor outcomes such a challenging task? At this point a comparison with other forecasting challenges may be useful, such as human population growth, economics, epidemics or meteorology. These forecasting challenges are approached by considering a system that is made up of a state (e.g. the number of active COVID-19 cases) and a set of laws that describe how this state will evolve through time (e.g. exponential growth). Predictions are then made by stepping forward through time, while updating the state based on the governing laws [9,10].

Making accurate predictions is more difficult in some problems than in others. Predictions for population growth, for example, are typically accurate [11]. Factors that contribute to this are the relatively small number of relevant variables that form the state (i.e. the number of persons per age category and birth rate), and the relatively straightforward dynamics. Contrast this with meteorology, the sci-



ence of weather predictions. The governing physical laws can be precisely defined, but the number of relevant variables is huge: temperature, pressure and humidity, among others, must be known with a high spatial resolution [12]. Another key difference is that small differences in initial conditions can lead to drastically different outcomes, a phenomenon that is colloquially known as the ‘butterfly effect’ [13]. This last factor, especially, makes accurate predictions challenging to make, as small measurement errors are typically amplified [14].

Predicting checkpoint inhibitor outcomes has several factors that make it a challenging prediction problem. First, the exact governing dynamics of cancer immunology remain elusive, in part because the immune system does not lend itself well to controlled, repeatable experiments [15]. Second, there are many relevant variables that influence the behavior of the system, such as the amount and location of many different types of cells and molecules. Adding to this is the fact that these variables are only partially available through blood samples, radiological scans and tissue biopsies. Third, studies on the immune system suggest a similar susceptibility to initial conditions as observed in meteorology [16–20]. These challenges manifest themselves not only in the setting of predicting checkpoint inhibitor treatment outcomes, but also in, for example, predicting flares in patients with auto-immune disease [21].

These factors may explain the results described in this thesis. Machine and deep learning models are immensely powerful in the relationships between input data and corresponding labels. In making predictions, however, they are limited by the information that is provided to them. The complexity of cancer immunology, along with the huge number of relevant variables, make it plausible that not enough information is captured by diagnostic modalities to make accurate predictions of the course of disease for an individual patient.

Furthermore, these factors suggest that a detailed understanding of the governing dynamics and high-resolution measurements are necessary to achieve the goal of accurate, individualized predictions. Clearly, this is much more challenging in cancer immunology than in meteorology, where the most important governing laws of meteorology have been known for centuries and real-time measurements of the relevant variables are widely available. Nonetheless, previous research has resulted in major breakthroughs, of which checkpoint inhibitor treatment is a prime example. Future research that focuses on more accurate understanding and mea-

## 7.5. Recommendations

---

surements may thereby not only facilitate better predictions, but also new treatment options. This is not to say that a data-driven approach cannot be of value. The results presented in this thesis, however, suggest that using these methods may be of more value when employed alongside others, working together towards a better understanding of the dynamics of cancer immunology. And in fact, the large number of relevant variables in cancer immunology may make data-driven methods invaluable for their analysis.

## 7.5 Recommendations

Based on the findings presented in this thesis, several recommendations for future research and clinical practice may be made. First, the presented findings give no reason to deviate from regular dietary and exercise recommendation in patients treated with checkpoint inhibitors for advanced melanoma. No conclusions can be drawn about a causal relationship between diet or exercise with treatment outcomes. However, the observational evidence of chapter 5 suggests a positive association between higher muscle density and less visceral adipose tissue with better outcomes. In addition, no association between higher body mass index and better outcomes was supported by both chapter 5 and a recent meta-analysis on this topic. Regular dietary and exercise recommendations are generally considered to be safe, and the presented findings do not suggest otherwise.

Second, future research should work towards accurate understanding and measurements in cancer immunology, potentially leveraging data-driven techniques. This will require the development and use of new research methodologies. A concrete example of this is the use of spatial transcriptomics [22], a cutting-edge technology that provides insight into which genes are expressed at a specific location in a tissue section. These patterns of localized gene expression could prove to be predictive of checkpoint inhibitor outcomes. In addition, this technique can provide a granular view into the spatial organization of a tumor and its surrounding tissue in terms of its microenvironment, which may significantly contribute to our understanding of cancer immunology.

Third, future research into data-driven techniques for clinical outcome predictions should investigate their added value over known predictors. Although this has been recognized before, the results of chapter 3 and 4 underline its importance.

This is especially true in settings where simple image-derived characteristics, such as size or organ location, are known to be predictive, as a machine or deep learning model can trivially learn to detect these characteristics.

## 7.6 References

- [1 ] Silva IP da, Ahmed T, McQuade JL, Nebhan CA, Park JJ, Versluis JM, et al. Clinical Models to Define Response and Survival With Anti-PD-1 Antibodies Alone or Combined With Ipilimumab in Metastatic Melanoma. *J Clin Oncol* 2022. <https://doi.org/10.1200/JCO.21.01701>
- [2 ] Deckers EA, Kruijff S, Brouwers AH, van der Steen K, Hoekstra HJ, Thompson JF, et al. The association between active tumor volume, total lesion glycolysis and levels of S-100B and LDH in stage IV melanoma patients. *Eur J Surg Oncol* 2020;46:2147–53. <https://doi.org/10.1016/j.ejso.2020.07.011>
- [3 ] Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Cálin AM, Delli Pizzi A, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol Off J Eur Soc Med Oncol* 2019;30:998–1004. <https://doi.org/10.1093/annonc/mdz108>
- [4 ] Duin IAJ van, Schuiveling M, Maat LS ter, Amsterdam WAC van, Berkmortel F van den, Boers-Sonderen M, et al. Baseline tumor-infiltrating lymphocyte patterns and response to immune checkpoint inhibition in metastatic cutaneous melanoma 2023:2023.11.27.23299053. <https://doi.org/10.1101/2023.11.27.23299053>
- [5 ] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language Models are Few-Shot Learners. *Adv Neural Inf Process Syst* 2020;33:1877–901.
- [6 ] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86:2278–324. <https://doi.org/10.1109/5.726791>
- [7 ] Li Z, Liu F, Yang W, Peng S, Zhou J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans Neural Netw Learn Syst* 2022;33:6999–7019. <https://doi.org/10.1109/TNNLS.2021.3084827>
- [8 ] Lu Z, Xie H, Liu C, Zhang Y. Bridging the Gap Between Vision Transformers and Convolutional Neural Networks on Small Datasets. *Adv Neural Inf Process Syst* 2022;35:14663–77.
- [9 ] Bender EA. *An Introduction to Mathematical Modeling*. Courier Corporation; 2000.
- [10 ] Thompson JMT, Stewart HB, Turner R. *Nonlinear Dynamics and Chaos*. *Comput Phys* 1990;4:562–3. <https://doi.org/10.1063/1.4822949>
- [11 ] Ritchie H, Roser M. *The UN has made population projections for more than 50 years – how accurate have they been?* *Our World Data* 2023.
- [12 ] *An Introduction to Dynamic Meteorology, Volume 88 - 5th Edition* — Elsevier Shop n.d. [https://shop.elsevier.com/books/an-introduction-to-dynamic-meteorology/holton/978-0-12-384866-6\(accessedFebruary20,2024\)](https://shop.elsevier.com/books/an-introduction-to-dynamic-meteorology/holton/978-0-12-384866-6(accessedFebruary20,2024))
- [13 ] Lorenz EN. Deterministic Nonperiodic Flow. *J Atmospheric Sci* 1963;20:130–41. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
- [14 ] Firth WJ. Chaos—predicting the unpredictable. *BMJ* 1991;303:1565–8.
- [15 ] *Tumor Immunology and Immunotherapy*. Oxford University Press; 2014. <https://doi.org/10.1093/med/9780199676866.001.0001>
- [16 ] Akaishi T, Takahashi T, Nakashima I. Chaos theory for clinical manifestations in multiple sclerosis. *Med Hypotheses* 2018;115:87–93. <https://doi.org/10.1016/j.mehy.2018.04.004>
- [17 ] Allilou M, Firouznia M, Patil P, Bera K, Gilkeson R, Rajiah P, et al. Chaos-based fractal radiomic features of nodule vasculature predicts response to immunotherapy on non-contrast lung CT. *J Immunother Cancer* 2018;6. <https://doi.org/10.1186/s40425-018-0423-x>

## 7.6. References

---

- [18 ] Kumar S, Kumar A, Samet B, Gómez-Aguilar JF, Osman MS. A chaos study of tumor and effector cells in fractional tumor-immune model for cancer treatment. *Chaos Solitons Fractals* 2020;141:110321. <https://doi.org/10.1016/j.chaos.2020.110321>
- [19 ] Mayer H, Zaenker KS, an der Heiden U. A basic mathematical model of the immune response. *Chaos Interdiscip J Nonlinear Sci* 1995;5:155–61. <https://doi.org/10.1063/1.166098>
- [20 ] Sharma V. The Application of Chaos Theory and Fractal Mathematics to the Study of Cancer Evolution: Placing Metabolism and Immunity Centre Stage. *Med Res Arch* 2016;4.
- [21 ] Gensous N, Marti A, Barnetche T, Blanco P, Lazaro E, Seneschal J, et al. Predictive biological markers of systemic lupus erythematosus flares: a systematic literature review. *Arthritis Res Ther* 2017;19:238. <https://doi.org/10.1186/s13075-017-1442-6>
- [22 ] Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;353:78–82. <https://doi.org/10.1126/science.aaf2403>



## 7.6. References

---

# Chapter 8

## Summary

This thesis investigates the predictive value of computed tomography (CT) and histopathology based predictors for outcomes of checkpoint inhibitor treatment in patients with advanced melanoma. The overarching goal is to identify accurate predictors based on routine diagnostic modalities that can guide clinical decisions, thereby mitigating unnecessary toxicity and healthcare costs. The presented research is part of the PREMIUM study, a multicenter effort of eleven melanoma treatment centers in The Netherlands.

Chapter 1 introduces the role of checkpoint inhibitor treatments in managing advanced melanoma, emphasizing the significant improvements in patient survival rates. It also highlights the challenge of predicting treatment outcomes due to the variable response among patients, serious toxicity and high costs associated with these treatments. Predictors explored include radiological characteristics, body composition metrics, and models based on CT scans and histopathology material, focusing on their added value over known predictors.

Chapter 2 presents a systematic review of previous literature on imaging biomarkers for response and survival in checkpoint inhibitor treatments across cancer types. To this end, 119 studies with a total of 15,580 patients were included and critically appraised. It found evidence supporting several imaging biomarkers, such as tumor burden and liver metastases, but notes the limited added value of baseline FDG-PET parameters. Radiomics and radioactive drug labeling are highlighted as promising yet methodologically challenging approaches.

## 8.0.

---

We assessed the predictive value of CT radiomics in chapter 3. These radiomics are visual characteristics such size, roundness and texture, which together capture the appearance of a metastatic lesion on CT imaging. Pretreatment CT imaging of 620 patients was collected, visible lesions were manually segmented and total of 9370 features per lesion were extracted. Lastly, a machine learning model was trained to predict clinical benefit based on the extracted features. While this model achieved moderate predictive accuracy (AUROC = 0.61), it did not significantly outperform a clinical model (AUROC = 0.65) based on the presence of brain and liver metastases, performance status and serum lactate dehydrogenase. An overlap in information between radiomics and clinical predictors can explain why radiomics does not provide additional value for predicting checkpoint inhibitor outcomes in melanoma.

Chapter 4 investigates deep learning models trained on CT images of metastatic lesions for predicting checkpoint inhibitor treatment outcomes. In contrast to the method used in chapter 3, no predefined features are extracted. Instead, a deep learning model is trained to predict treatment outcomes using the raw CT image of the lesion. The hypothesis explored in this chapter was that this method will improve over the method in chapter 3, because it is not limited by the choice of extracted features. Baseline CT scans from 730 patients were collected, lesions were located and a deep learning model was trained on lesion volumes to predict treatment outcomes. Results were similar to those of chapter 3: the model achieved a significant predictive value (AUC = 0.61) but did not improve over a model of known clinical predictors (AUC = 0.64). Again, an overlap in learned information appears to explain this lack of improvement.

Chapter 5 explores the association between body composition metrics derived from CT scans and treatment outcomes. For 1471 patients, data was collected on pretreatment body mass index, in addition to the CT-derived metrics skeletal muscle index and density, and subcutaneous and visceral adipose tissue. The association of these metrics with overall and progression-free survival was investigated using Cox proportional hazards models. Results suggest that underweight BMI, lower skeletal muscle density and higher visceral adipose tissue index associated with worse outcomes, which is independent of known clinical predictors.

The preliminary study presented in chapter 6 investigates deep learning to analyze histopathology images for predicting treatment outcomes. The investi-



gated deep learning model works by splitting the whole-slide histopathology image into square, non-overlapping patches, extracting features from these patches using pretrained deep learning models and training a classifier to predict treatment outcomes based on these features. From 716 patients, a total of 471 primary and 516 metastatic samples were collected. This model did not perform significantly better than random for predicting treatment outcomes in both primary (AUC = 0.50) and metastatic samples (AUC = 0.54). Although further research is required, this preliminary result suggests that predicting checkpoint inhibitor treatment outcomes based on histopathology imaging is a challenging task.

The thesis concludes by synthesizing the findings from the various explored predictors, acknowledging the complexity of accurately predicting checkpoint inhibitor outcomes. The discussion emphasizes the challenging nature of predicting immune response to cancer and suggests that advancements in understanding cancer immunology and high-resolution measurements are crucial for future improvements in prediction accuracy. Recommendations for future research include leveraging new methodologies like spatial transcriptomics and ensuring data-driven techniques are evaluated against known predictors. Despite identifying several predictors that add to existing knowledge, the thesis acknowledges that current predictive accuracy is insufficient to withhold checkpoint inhibitor treatment, although it can be used to inform shared decision making.

8.0.

---

# Chapter 9

## Dutch Summary (Nederlandse Samenvatting)

Dit proefschrift onderzoekt de voorspellende waarde van op computertomografie (CT) en histopathologie gebaseerde voorspellers voor de uitkomsten van patiënten die worden behandeld met immuuntherapie (checkpointremmers) voor een gevorderd melanoom. De beoordeling van het histopathologische materiaal (microscopische beelden van tumorweefsel) en CT-scan maken onderdeel uit van het aanvullend onderzoek dat standaard wordt verricht om de diagnose te stellen en uitgebreidheid van de ziekte vast te stellen. Het doel is om nauwkeurige voorspellers te identificeren op basis van deze databronnen, zodat onnodige bijwerkingen en zorgkosten kunnen worden voorkomen. Dit onderzoek werd verricht in het kader van de PREMIUM studie, een samenwerking van elf behandelcentra voor melanoom in Nederland.

Hoofdstuk 1 introduceert de grote verbeteringen in de overlevingskansen van patiënten sinds de ontdekking van immuuntherapie. Het benadrukt echter ook de uitdagingen, namelijk de wisselende uitkomsten van therapie, ernstige bijwerkingen en hoge zorgkosten die met deze behandeling gepaard gaan. In dit proefschrift onderzoeken we radiologische kenmerken, lichaamssamenstelling en modellen gebaseerd op CT-scans en histopathologisch materiaal, waarbij de nadruk ligt op hun toegevoegde waarde ten opzichte van bekende voorspellers.

Hoofdstuk 2 geeft een systematisch overzicht van radiologische kenmerken die

## 9.0.

---

de respons en overleving voorspellen bij behandelingen met checkpointremmers voor alle soorten kanker. In totaal werden 119 studies met 15,580 patiënten geïnccludeerd. In deze studies werd gevonden dat sommige kenmerken op scans, zoals het totale volume van de uitzaaiingen en de aanwezigheid van uitzaaiingen in de lever enige voorspellende waarde hebben. FDG-PET-parameters. Radiomics en radioactieve labeling van geneesmiddelen worden benoemd als veelbelovende maar methodologisch uitdagende benaderingen.

In hoofdstuk 3 onderzochten we de voorspellende waarde van CT-radiomics in hoofdstuk 3. Radiomics zijn visuele kenmerken zoals grootte, vorm en textuur, die samen het uiterlijk van een uitzaaiing op CT-beelden samenvatten. CT-beelden van 620 patiënten werden verzameld, zichtbare laesies werden handmatig ingekleurd en in totaal 9370 kenmerken per laesie werden berekend. Tot slot werd een model getraind om behandeluitkomsten te voorspellen op basis van deze berekende kenmerken. Hoewel dit model enige voorspellende waarde bereikt, voorspelt dit model niet significant beter dan een model gebaseerd op voorspellers uit het medisch dossier van de patiënt, zoals de aanwezigheid van hersen- en leveruitzaaiingen, de conditie van de patiënt en de hoogte van lactaat dehydrogenase in het bloed. De overlap in informatie tussen radiomics en klinische voorspellers kan verklaren waarom radiomics geen extra waarde bieden voor het voorspellen van de uitkomsten van checkpointremmers bij melanoom.

Hoofdstuk 4 onderzoekt deep learning modellen die zijn getraind op CT-beelden van uitzaaiingen om behandelresultaten van checkpointremming te voorspellen. In tegenstelling tot de methode die in hoofdstuk 3 wordt gebruikt, worden er hierbij geen vooraf gedefinieerde kenmerken geëxtraheerd. In plaats daarvan wordt een model getraind om de behandelresultaten te voorspellen aan de hand van de ruwe CT-beelden van de laesie. In dit hoofdstuk onderzochten we of deze methode een verbetering oplevert ten opzichte van de methode in hoofdstuk 3, omdat deze niet wordt beperkt door de keuze van de berekende kenmerken. Er werden baseline CT-scans van 730 patiënten verzameld, laesies werden gelokaliseerd en een deep learning-model werd getraind op laesievolumes om de behandelresultaten te voorspellen. De resultaten waren vergelijkbaar met die uit hoofdstuk 3: het model behaalde een significante voorspellende waarde ( $AUC = 0.61$ ), maar verbeterde de voorspelling niet ten opzichte van een model met bekende klinische voorspellers ( $AUC = 0.64$ ). Opnieuw leek een overlap in geleerde informatie dit gebrek aan

verbetering te verklaren.

Hoofdstuk 5 onderzoekt de associatie tussen lichaamssamenstelling, bepaald op basis van CT-scans, en behandelresultaten. Voor 1471 patiënten werd data verzameld over BMI en verschillende maten van lichaamssamenstelling op basis van CT-beelden, namelijk skeletspierindex en -dichtheid, en de hoeveelheid subcutaan en visceraal vetweefsel. De associatie van deze statistieken met de algehele en progressievrije overleving werd onderzocht met behulp van Cox proportional hazards modellen. De resultaten suggereren dat patiënten met een te lage BMI, een lagere skeletspierdichtheid en/of een hogere viscerale vetweefselindex minder vaak goed effect van immuuntherapie hebben. Deze voorspellende waarde lijkt onafhankelijk te zijn van bekende voorspellers.

Het voorlopige onderzoek dat in hoofdstuk 6 wordt gepresenteerd, onderzoekt deep learning op microscopische beelden van tumorweefsel om behandeluitkomsten van checkpointremming te voorspellen. Het onderzochte model werkt door het microscopische beeld op te delen in vierkante, niet-overlappende patches, kenmerken uit deze patches te extraheren door middel van vooraf getrainde deep learning modellen en een model te trainen dat behandelresultaten kan voorspellen op basis van deze kenmerken. Van 716 patiënten werden in totaal 471 pathologiebeelden van het oorspronkelijke melanoom en 516 beelden van uitzaaiingen verzameld. Zowel het model gebaseerd op primaire melanomen (AUC = 0.50) als dat waarbij beelden van uitzaaiingen werden gebruikt (AUC = 0.54) had geen voorspellende waarde. Hoewel verder onderzoek nodig is, suggereert dit voorlopige resultaat dat het voorspellen van behandelresultaten van immuuntherapie op basis van histopathologische beelden een uitdagende taak is.

Het proefschrift sluit af met het samenvatten van de bevindingen voor de verschillende onderzochte voorspellers. Deze resultaten benadrukken de complexiteit van de immuunrespons op kanker en suggereren dat verder onderzoek zich moet richten op een gedetailleerder begrip van kankerimmunologie om voorspellingen te verbeteren. Aanbevelingen voor toekomstig onderzoek zijn onder meer het benutten van nieuwe methodes zoals “spatial transcriptomics”, waarbij data-gedreven technieken veel waarde kunnen toevoegen. Concluderend kunnen voorspellers uit het medisch dossier samen met informatie over lichaamssamenstelling op CT iets zeggen over de kans op respons op immuuntherapie. De combinatie van deze factoren is echter op dit moment nog niet genoeg om patiënten op basis hiervan be-

## 9.0.

---

handeling te onthouden. Wel kunnen ze gebruikt worden in de in de gezamenlijke besluitvorming van patiënt en arts.

## Chapter 10

# Acknowledgements (Dankwoord)

Geachte prof. dr. J.P.W. Pluim, beste **Josien**, ik wil je hartelijk bedanken voor de hulp, steun en het vertrouwen dat ik de afgelopen drie jaar van je heb mogen ontvangen. Ik heb erg veel bewondering voor je scherpzinnigheid en betrouwbaarheid, en heb veel kunnen leren van je altijd waardevolle feedback.

Geachte prof. dr. P.J. van Diest, beste **Paul**, weinig mensen hebben het charisma en de overredingskracht die jij bezit. Deze eigenschappen van jou waren onmisbaar in het mogelijk maken van dit grote multicenter project. Dank hiervoor, en daarnaast voor je begeleiding, steun en feedback de afgelopen drie jaar.

Geachte prof. dr. K.P.M. Suijkerbuijk, beste **Karijn**, ik heb veel bewondering voor wat je (nu al) in je carrière bereikt hebt, maar nog meer voor dat je, ondanks je indrukwekkende prestaties, bescheiden bent gebleven en altijd openstaat voor de ideeën van anderen. Dit is wat mij betreft een van je grootste krachten, en zorgde ervoor dat ik met veel plezier in jouw team gewerkt heb. Dankjewel hiervoor, en voor alles wat ik van je heb mogen leren.

Dear dr. Veta, dear **Mitko**, your relaxed and positive attitude and your sense of humor have made it a pleasure to work together with you these past three

## 10.0.

---

years. You never let down an opportunity to help me out. I could not have hoped for a better copromotor.

Geachte leden van de **beoordelingscommissie**, dr. K.G.A. Gilhuijs, prof. dr. G.A.P. Hospers, prof. dr. J. van der Laak, prof. dr. M.G.E.H. Lam, prof. dr. ir. J. de Ridder, hartelijk dank voor het beoordelen van dit proefschrift.

Dear **colleagues from the Image Sciences Institute**, dear Bea, Cyrano, Djenifer, Giulia, Iris, Ishaan, Maria, Mark, Mathijs, Max, Myrthe, Ryan, Ryanne, Saeed and Tessa, your personalities are very diverse, but you all have three things in common: love for fun, openness to everyone who is new and the right amount of nerdiness. These things have made me feel very much at home in the ISI these past years. Thank you for all the fond memories.

Beste **collega's van de melanoom- en CRC-research groepen**, beste Dania, Emma, Ingrid, Karel, Lidwien, Marinde, Manja, Mick, Olivier, Mirjam, Patricia, Renee, Rik, Ruben, Sietske, Suzanna en Wouter, zonder jullie waren de afgelopen jaren niet even leuk geweest. Dank jullie allemaal voor de vele ontspannen momenten die ik met jullie heb mogen delen, en ook voor de kritische blik op mijn presentaties en werk.

**Belle en Mark**, ongekende PREMIUM legends. Samen hebben we de afgelopen jaren de kar getrokken voor het PREMIUM project en ons een weg door de ziekenhuisbureaucratie genavigeerd op zoek naar onze data. Het was me een waar genoegen om met jullie samen te werken, en zonder onze gezamenlijke inspanning was dit proefschrift er niet geweest. Dankjulliewel.

**Robin en Rogier**, sehr geehrte Kollegen en vrienden, dank dat jullie me willen bijstaan als paranymfen. Robin, door een gelukkige speling van het lot had ik het voorrecht om tegelijkertijd met jou mijn PhD te mogen doen. Dank voor alle lol die ik samen met je heb gehad de afgelopen jaren, en voor de fantastische vriend die je bent. Let's go baby. Rogier, al sinds het begin der tijden in 2014 vind ik je een topgozer, zowel op de ontspannen als serieuze momenten die we samen hebben meegemaakt. Dank dat je altijd voor me klaarstaat als dat nodig is.



## Chapter 10. Acknowledgements (Dankwoord)

---

Lieve **Marit, Frank, Jara** en **Gerwin**, ik had mij geen betere schoonfamilie kunnen wensen. Dankjewel voor de warmte en gastvrijheid waarmee jullie mij vanaf dag 1 hebben ontvangen, en voor hoe jullie altijd voor me klaarstaan.

Lieve **Laurie** en **Josina**, als zussen hebben jullie een speciale plek in m'n hart. Laurie, ik kijk naar je op in hoeveel je nu al bereikt hebt hoe authentiek je altijd bent. Josina, ik heb ontzettend veel bewondering voor hoe jij je door uitdagingen in je leven heen hebt geslagen. Het maakt me heel blij om te zien waar je nu staat. **Josco**, dankjewel voor de steun en toeverlaat die je bent voor mijn zus.

Lieve **papa** en **mama**, zeker nu ik zelf vader ben, wordt mij nog meer dan voorheen duidelijk hoeveel jullie voor mij hebben gedaan en over hebben gehad. Dank voor alles wat jullie voor me betekend hebben en nog steeds betekenen.

Lieve **Kobe**, ik kan me niet voorstellen dat je ooit het proefschrift van je vader zou willen doorlezen. Maar voor het geval dat je het tot het dankwoord gered hebt, dan wil ik dat je weet dat jij en je moeder het mooiste zijn wat me ooit is overkomen.

Lieve **Kyla**, ik kom woorden tekort om te beschrijven hoeveel jij voor mij betekent. Je bent een fantastisch vrouw, moeder en partner. Ik ben zo ontzettend blij dat ik jou ontmoet heb en dat ik mijn leven met je mag delen. Dankjewel voor al je steun, hulp, advies en geduld van de afgelopen jaren, maar bovenal voor de liefde die ik elke dag weer van je krijg.

**10.0.**

---

# Chapter 11

## Curriculum Vitae

Rens ter Maat was born on April 26th, 1995 in Woerden, The Netherlands. He attended the Minkema College in Woerden, graduating cum laude in 2011. At the age of 19, he completed a Bachelor's degree in Artificial Intelligence at Utrecht University. Subsequently, he enrolled into med school, finishing his training as a doctor in 2021 with a GPA of 4.0. During his his studies, he completed an internship at a Google-founded Silicon Valley tech company that develops a new surgical robot. Furthermore, he won the national and international competition for the simulated resuscitation of a critically ill patients, and was the chairman of the Red Cross Studentdesk Utrecht.

In 2021, he started his PhD research at the Image Sciences Institute under the supervision of prof. Plum, prof. Van Diest, prof. Suijkerbuijk and dr. Veta. During his research, he supervised two master students and two bachelor students during their research internships. He will continue pursuing his passion for conducting research and solving complex problems as a Senior Mathware Engineer at Sioux Technologies, Eindhoven.